Revenue-Based Call Admission Control for MPEG-4 Wireless Videoconference Traffic

Aggelos Lazaris[1] and Polychronis Koutsakis[2]
[1]Department of Electronic and Computer Engineering, Technical University of Crete, Greece
E-mail: alazaris@telecom.tuc.gr
[2]Department of Electrical and Computer Engineering, McMaster University, Canada
E-mail: polk@ece.mcmaster.ca

Abstract

In this work, we first introduce a new traffic model for medium quality MPEG-4 videoconference traffic and then proceed to use it in the implementation of a new Call Admission Control scheme. Our scheme makes decisions on the acceptance/rejection of a new video call not only based on the predicted capacity that users will consume, but also on the possible revenue gained for the provider when degrading current users in order to accommodate new ones. Via an extensive simulation study our scheme is shown to provide excellent Quality of Service (QoS) to wireless videoconference users.

1. Introduction

Emerging wireless networks will need to accommodate significant loads of traffic related to real-time video services, and especially videoconference traffic [1]. The QoS requirements of video users are particularly strict. The reason is that video packet transmission delays and the subsequent video packet dropping when the delay exceeds an upper bound result in the viewer's annoyance. Call Admission Control (CAC) is a strategy used to limit the number of call connections into the network in order to reduce network congestion, therefore enabling the system to provide the desired QoS to newly incoming as well as existing calls. In wireless cellular networks the traffic conditions in the cells can change very quickly due to user mobility; also, when mobile users change their point of attachment (handoff), the end-to-end path may be changed while they still expect to receive the same QoS. An efficient CAC mechanism should be able to cope with this strict user requirement.

In recent work [2] we have proposed such an efficient CAC scheme for high quality MPEG-4 videoconference traffic. Like many CAC schemes in the literature, our scheme adopted the idea of a probabilistic service, as described in [4]. This type of service does not provide for the worst-case scenario, but instead guarantees a bound on the rate of lost/delayed packets based on statistical characterization of the traffic. However, our scheme did not adopt the standard method of implementation for this service type which is the use of an "equivalent bandwidth" estimation, larger than the average rate but less than the peak rate of the sources. Although widely used, "equivalent bandwidth"-based schemes are known to significantly overestimate the sources' actual bandwidth requirements and therefore to provide quite conservative CAC schemes, which fail to use efficiently all the available bandwidth [7, 9]. Instead, we proposed the use of our recent modeling approach [16] for traffic originating from MPEG-4 videoconference sources, in order to design a new CAC scheme for wireless cellular networks which uses the traffic parameters which the video source either declares at call setup or has agreed on in its contract with the wireless provider, in order to precompute a large number of traffic scenarios for its decision-making.

In the present work we introduce a new model for medium quality (MQ) MPEG-4 videoconference traces. We then use the accuracy of our modeling approach both for medium and high quality (in [16]) MPEG-4 videoconference traffic, in order to propose a new CAC scheme for wireless cellular networks which makes decisions not only based on the system's ability to accommodate newly arriving users in terms of capacity (as in [2]), but also on the profit that can be made by the provider if existing users are degraded in order for new video calls to be accepted.

2. MQ MPEG-4 Videoconference Traffic Model

In this work, we study the medium quality version of three different long sequences of MPEG-4 encoded videos from [11]. The difference between the medium and the high quality encoding of the movies lies in the quantization parameters used in each case, in [11]. The three traces ("Office Cam", "Lecture Room Cam", "Boulevard Bio") are movies with low motion. We have investigated the possibility of modeling the traces with a number of well-known distributions (gamma,

978-1-4244-2100-8/08/$25.00 ©2008 IEEE

lognormal, log-logistic, exponential, geometric, Weibull, Pearson V). Our results (derived with the use of Q-Q plots [10], Kolmogorov-Smirnov (KS) tests [10] and Kullback-Leibler (KL) tests [8]) have shown that, similarly to our work in [16] on modeling high quality MPEG-4 videoconference traffic, the best fit among these distributions for modeling a single movie is achieved for all traces examined with the use of the Pearson type V distribution. The data for each trace consists of a sequence of the number of cells per video frame. The length of the videos varies from 45 to 60 minutes. Table 1 presents the trace statistics for each trace, in the Medium Quality (MQ) columns. Table 1 also contains the trace statistics for the high quality version of the same movies, as both versions will be needed in our work on an efficient CAC scheme. *The sets of parameters of the traces comprise the "modes" adopted by videoconference users in our study. This will be further explained in Section 5.*

However, although the Pearson V was shown to be the better fit among all distributions, the degree of goodness-of-fit for the Pearson V varied significantly, and even in the cases of a quite good fit, the fit was not perfectly accurate. This was expected, as the gross differences in the number of bits required to represent *I, P* and *B* frames impose a degree of periodicity on MPEG-encoded streams, based on the cyclic GOP formats. Any model which purports to reflect the frame-by-frame correlations of an MPEG-encoded video stream must account for GOP cyclicity, otherwise the model could produce biased estimates of cell loss rate for a network with some given traffic policing mechanism [3]. Hence, we proceeded to study the frame size distribution for each of the three different video frame types (*I, P, B*), in the same way we studied the frame size distribution for the whole trace. The Pearson V distribution once again provided the best fitting results for all types of video frames' sequences, and the modeling results were much improved in comparison with those of modeling the trace as a whole. We present, indicatively, the results from our KS-tests for the *I, P* and *B* frames of the lecture trace in Figures 1-3. The results show that the Pearson V distribution is the best fit, as it has the smallest maximum vertical deviation from all the distributions. Similar results were deducted by all our statistical tests. The goal of our work in this Section is to build a model which, based on the good but not perfect fit of the Pearson V distribution for modeling a single movie, will accurately capture the behavior of *multiplexed medium quality MPEG-4 videoconference movies* from VBR coders.

A Discrete Autoregressive model of order *p*, denoted as DAR(*p*) [14], generates a stationary sequence of discrete random variables with an arbitrary probability distribution and with an autocorrelation structure similar to that of an Autoregressive model. DAR(1) is a special case of a DAR(p) process. We build *for each video frame type* a model based only on four physically meaningful parameters, i.e., the mean, peak, variance and the lag-1 autocorrelation coefficient ρ of the offered traffic (which is typically very high for videoconference sources), which are either known at call set-up time or can be measured without introducing much complexity in the network. We proceeded with testing our models statistically (with the same methods used for single traces) in order to study whether it produces a good fit for the trace superposition. The accurate fits in our results have shown that the superposition of the actual traces can be modeled well by a respective superposition of data produced by our modeling approach. In Figure 4, we have plotted the 0.01-, 0.2-, 0.03-,… quantiles of the actual office camera trace versus the respective quantiles of the DAR(1) model for the superposition of 30 traces' P frames. The values in both axes are in packets. As shown in the Figure, the points of the Q-Q plot fall either very close or completely along the 45-degree reference line (which would correspond to a perfect match of the actual trace quantiles), with the exception of the first and last 3% quantile (left and right-hand tail), for which the DAR(1) model greatly overestimates the probability of frames with a very small or very large, respectively, number of cells. The very good fit (shown from all our results, which are similar in nature to those in [16] and omitted here due to space limitations) shows that the superposition of the actual traces can be accurately modeled by a respective superposition of data produced by the DAR(1) model.

3. Channel Error Model

Errors in the wireless channel due to noise typically occur over relatively short bursts and are highly correlated in successive slots, but uncorrelated over long time windows. In our simulation study we adopt a channel error model similar to the widely studied Gilbert-Elliot [12] model, where the channel switches between a "good state" and a "bad state", but with the modification that the good state is not always error-free, neither the bad state always with errors. In our model, we consider a case where the error probability in the bad state is in the order of 10^4 times larger than the error probability in the bad state (this ratio is taken from one set of parameters in [5]). The parameters of our model are: P_r (good) =0.99994, P_r (good-bad) =0.000006, P_r (bad-good) =0.1, Error probability (good state) =$0.4*10^{-4}$, Error probability (bad state) =0.7. The total probability of a transmission error occurring is equal to: P_r (good)*Error prob (good) + P_r (bad)*Error prob (bad) $=8.2*10^{-5}$. That is, the total probability of a transmission error is only slightly smaller than the maximum acceptable video packet dropping probability of 10^{-4} [13] which we consider in this work for real-time videoconference traffic, making the need for very efficient call admission control imperative.

4. Revenue-Based Call Admission Control

The precomputation of traffic scenarios, along with the online simulation, was made in [2] based on the traffic parameters declared by the video sources at call setup. These parameters are used for the "identification" of the source as a user adopting a specific "mode". In order to explain what a "mode" is, we first note that a logical assumption for next generation wireless networks is that videoconference users will be allowed to adopt one of a few specific "modes", each corresponding to a set of traffic parameters. Therefore, we used in our work each user's declared set of parameters in order to examine the respective precomputed traffic scenario, based on our MPEG-4 model for a source with such a set of parameters. This approach is especially plausible for wireless videoconference traffic, as the number of variations between source bandwidth requirements is naturally restricted by the type of application (a much larger pool of "modes" would have to be used in the case of regular video traffic). In [15] we have shown that our scheme works equally well for H.263 videoconference traffic and excels, again, in comparison to the equivalent bandwidth approach. We denote in the present work as "modes" for the MPEG-4 videoconference users the sets of traffic parameters presented in Table 1. Hence, we use six "modes" for MPEG-4 videoconference traffic, one high and one medium quality "mode" for each one of the three traces. High-paying users adopt the high quality (HQ) modes, due to the increased bandwidth these modes offer.

One parameter not included in our study in [2] was that in a real-life scenario, the decision of admitting or rejecting a new call in the network will be made by the provider not only based on the capacity needed to accommodate the call, but also on the revenue that the admission of the new call will provide. That is, if the admission of a new call (and the subsequent increase in bandwidth utilization) can only be made with the degradation of a higher-paying customer who enjoys higher QoS, the CAC module should compute whether this is a profitable decision. For this reason, in our new CAC scheme we not only adopt the idea of [2] for precomputation and online computation of various traffic scenarios and implement it for MPEG-4 traffic, but we also assign "revenue weights" to each one of the six MPEG-4 "modes", thereby differentiating them into different service classes. These weights are shown in Table 1 and are assigned in an ad-hoc manner here, without loss of generality, based on the traffic parameters of each "mode". Therefore, users adopting the high quality "Boulevard Bio" MPEG-4 videoconference mode are the ones demanding the highest QoS and paying respectively for it, followed by users adopting the high quality "office" mode and the high quality "lecture" mode; users adopting the MQ versions of the traces are the low-paying users. Users choose one of the six "modes" with a probability when they enter the system (in Section 5 we will discuss our results when altering this probability). The assignment of the weights is not linear, since there are significant differences among the traffic parameters of each "mode". We consider that 50% of the users of the high quality "office" and high quality "lecture" modes can accept degradation to a lower quality mode. By "lower quality mode" we refer to the MQ mode of the same movie. Users who have adopted the high quality "Boulevard Bio" mode are considered to be the highest paying users, therefore we assume that only a small percentage of them (20%) accept degradation. The choice of the percentages of users who accept degradation is again used here indicatively, and a change would not alter the nature of our results.

Our CAC scheme uses the traffic models presented in Section 2 for high quality and in [16] for medium quality MPEG-4 traffic, in order to *precompute* a number of traffic scenarios. Naturally, not all traffic scenarios can be precomputed, due to the very large number of all possible traffic loads; still, with the use of an adequate number of precomputed scenarios and our accurate video model, when a non-precomputed traffic load occurs in the system an online simulation can be conducted relatively quickly by our system in order to compute the "deviation" between the bandwidth needed currently and the "closest" (in terms of the synthesis of modes) precomputed traffic scenario. This new traffic scenario will then be added into the CAC scheme's database of precomputed scenarios. As already explained, our scheme does not make its decision based only on the maximization of system capacity, as in [2] and in most CAC schemes in the literature, but also on the maximization of provider revenue. Therefore, the current revenue R is computed as $R = \sum_i Ni * Wi$

where N_i is the total number of video users of "mode" i, and W_i is the revenue from each user of "mode" i (shown in the last columns of Table 1). Then our proposed CAC algorithm proceeds with the following steps, at the arrival of a new user request (either from within the picocell or from handoff). The system first checks whether it can be accommodated in terms of the total bandwidth which will be needed when the user is multiplexed with the existing users in the system. If this is not possible, the algorithm attempts to degrade the user, if the user accepts degradation. The rationale behind this decision is that the arrival of a new user should cause the minimum possible number of degradations, and hence irritation, to users who are already in the system, therefore it is preferable that the new user is accepted with degradation. One point which needs to be stressed here is that in most of the relevant works in the literature (including our work in [2, 15]), it is commonly accepted that handoff calls have absolute priority in obtaining an equal amount of channel bandwidth as the one they were occupying in their previous picocell location, i.e., handoff calls are not expected to endure any quality degradation, as this would lead to user dissatisfaction. We take a different approach in this work. It is indeed crucial for a handoff user to not experience call dropping when moving from one picocell to the next, as this would lead to significant user irritation (call dropping is much more irritating than the blocking of the call of a new user who attempts to transmit). However, if the mobile user experiences, during handoff, a degradation for which he has agreed in his contract, this should not be a cause for user irritation and therefore is allowed in our algorithm. If after degradation (of either a new or a handoff videoconference call) the acceptance of the call is still not possible, the CAC scheme checks all possibilities of degrading users of the same or lesser priority of the new call in order to accommodate it. If such a possibility exists and the call comes from handoff, it is accepted. If, however, it is a new call originating from within the picocell, it will be accommodated *only if* its acceptance will lead to higher revenue; otherwise, even if the total bandwidth that will be used with the acceptance of the new call is larger than the bandwidth previous used, there is no reason to degrade a significant number of users (and cause them even a slight irritation) if the provider will receive no extra revenue. In the case that the new call does not accept any degradation, the attempt is still made to degrade lesser or equal priority users who are already in the system, and a new call from within the picocell is again accepted only if it leads to higher revenue.

It needs to be stressed that it is actually not necessary to adopt the same approach (i.e., of basing the CAC scheme on traffic models) for all types of flows in the wireless network. If an accurate model exists for video traffic, which is the most bursty type of traffic in the network, the remaining types of flows (e.g., voice and data flows) could be admitted based simply on their declared mean rate, or with any other of the many efficient approaches proposed in the literature.

5. Results and Discussion

Fourth generation mobile data transmission rates are planned to be up to 20 Mbps, therefore in this work we study a channel of this rate. The maximum allowed transmission delay for the video packets of a Video Frame (VF) is equal to the time before the arrival of the next VF, with packets being dropped when the deadline is reached (the interframe period in MPEG-4 encoded movies is 40 ms). Our scheme is evaluated in 12 different scenarios versus the actual traffic generated by the real video traces, under handoff loads ranging from 5% to 15% of the total traffic (a handoff call can belong to any of the "modes" with equal probability). The first scenario is one in which the six "modes" in our study are used with equal probability. When studying this scenario, we found that the maximum number of users that the system could accommodate, without violating the strict QoS requirement of 0.01% maximum video packet dropping, was 51. As it will be shown in Table 3, this corresponds to less than 75% utilization of the channel capacity. The burstiness of video traffic is responsible for the system's inability to accommodate more sources without violating their QoS requirements. In each one of the other traffic scenarios studied (scenarios 2-12), we have considered various combinations of the cases where each one of the six modes is selected by users with one of the probabilities: 10%, 20%, 30%, 40% and the total number of users present in the system is 51 (we have chosen to keep the maximum number of traces equal to the maximum that can be achieved in scenario 1). The percentages shown in Table 2 refer to the initial "mode" with which a user enters the network; this mode may change due to degradation, based on the user's contract.

Each simulation point is the result of an average of 10 independent runs, each simulating one hour of network operation. Table 3 presents in its first column the bandwidth which is actually needed by the video traces, in its second column the estimated bandwidth that the traces will need based on our DAR(1) modeling approach, and in its third column the bandwidth that is utilized with the use of our CAC scheme. Two significant conclusions can be drawn from the Table. The first is that the estimation provided by our mechanism yields an overestimation of the actual bandwidth requirements of the superposed sources (the reasons for this are explained in [16]); still, this overestimation is small and ranges in all simulated scenarios from a minimum of 3.3% to a maximum of 7.92%. The average overestimation provided by our scheme is 5.04% over all the studied scenarios, which is acceptable, especially given that a small overestimation of the actual bandwidth requirements of video traffic is usually preferable, in order for the system to cope with the bursty nature of video users. The second conclusion has to do with the efficiency of our CAC scheme. By comparing the three respective columns for 5%, 10% or 15% it is clear that the actual bandwidth utilized when our revenue-based CAC scheme is enforced is smaller than the bandwidth needed by the real traces, and hence also smaller than the estimated bandwidth with the use of the DAR(1) model. The reason is that some of the MQ videoconference calls are rejected from the system in order to achieve higher revenue for the provider; hence, with the use of the DAR(1) modeling approach we reserve slightly more bandwidth than actually needed, and then, with the use of the CAC scheme we hinder a number of MQ users from accessing the system in order to keep HQ users continuously content with the service they are receiving (i.e., they seldom need to be degraded). In all the studied scenarios, the maximum system throughput is achieved in the case of 15% handoff traffic. This is expected, as in this case a larger number of video traces are accommodated by the system, by the gradual degradation of HQ users to medium video quality. The maximum throughput achieved is 76.55%, in Scenario 3.

We further investigate our mechanism's performance in the results presented in Figure 5, where we present in the y axis both the estimation provided by the DAR model and the actual bandwidth usage from our revenue-based CAC scheme; we also indicatively present the estimation provided by the use of the equivalent bandwidth approach from [6]. All the above are presented versus the normalized real system utilization; this indicates the actual traffic load generated by the traces, normalized to the channel capacity, e.g., a traffic load equal to 40% represents 40% of the 20 Mbps uplink capacity, i.e., 8 Mbps system throughput (these loads have been created with different combinations of probabilities for the six "modes" under study, and the results presented are the average over all the combinations used). As shown in the Figure, the equivalent bandwidth estimation significantly overestimates the actual traffic load in all cases. Finally, and most importantly, the Figure shows once again that the use of our CAC scheme leads to a slight underallocation to the videoconference users, in comparison to their offered load. The reason is that some of the MQ videoconference calls are rejected from the system in order to achieve higher revenue for the provider. As the offered load increases, this underallocation increases as well, in order not to allow low-paying users to fill the channel capacity at the cost of degrading high-paying ones.

Acknowledgment

This work has been funded, in part, by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] M. Etoh and T. Yoshimura, "Advances in Wireless Video Delivery", *Proceedings of the IEEE*, Vol. 93, No. 1, 2005.

[2] P. Koutsakis, "Efficient Call Admission Control for MPEG-4 Wireless Videoconference Traffic", *in Proc. of the IEEE ICCCN 2007*.

[3] M. Frey, S. Ngyuyen-Quang, "A Gamma-Based Framework for Modeling Variable-Rate Video Sources: The GOP GBAR Model", *IEEE/ACM Transactions on Networking,* Vol. 8, No. 6, 2000.

[4] H. Zhang and E. W. Knightly, "Providing end-to-end statistical performance guarantee with bounding interval dependent stochastic models", *in Proc. of the ACM SIGMETRICS*, 1994, Nashville, USA.

[5] J. G. Kim and M. Krunz, "Effective Bandwidth in Wireless ATM networks", *in Proc. of the 4th ACM/IEEE MOBICOM* 1998.

[6] W. Verbiest, L. Pinoo and B. Voeten, "The impact of the ATM concept on video coding", *IEEE Journal on Selected Areas in Communications,* Vol. SAC-6, 1988, pp. 1623-1632.

[7] R. Guerin, H. Ahmadi and M. Naghsineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks", *IEEE Journal on Selected Areas in Communications,* Vol. 9, No. 7, 1991.

[8] K. P. Burnham and D. R. Anderson, "Model Selection and Multi-Model Inference", Springer-Verlag, 2002.

[9] J. Q.-J. Chak and W. Zhuang, "Capacity Analysis for Connection Admission Control in Indoor Multimedia CDMA Wireless Communications", *Wireless Personal Communications*, Vol. 12, 2000, pp. 269-282.

[10] A. M. Law and W. D. Kelton, "Simulation Modeling & Analysis", 2nd Ed., McGraw Hill Inc., 1991.

[11] F. H. P. Fitzek, M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation", *IEEE Network,* Vol. 15, No. 6, 2001.

[12] E. N. Gilbert, "Capacity of a burst-noise channel", *Bell Systems Technical Journal,* Vol.39, 1960.

[13] D. A. Dyson and Z. J. Haas, "A Dynamic Packet Reservation Multiple Access Scheme for Wireless ATM", *Mobile Networks and Applications (MONET) Journal,* Vol. 4, No. 2, pp. 87-99, 1999.

[14] P. A. Jacobs and P. A. W. Lewis, "Time series generated by mixtures", *Journal of Time Series Analysis*, Vol. 4, No. 1, 1983.

[15] S. Chatziperis, P. Koutsakis and M. Paterakis, "A New Call Admission Control Mechanism for Multimedia Traffic over Next Generation Wireless Cellular Networks", *IEEE Transactions on Mobile Computing,* Vol. 7, No. 1, 2008.

[16] A. Lazaris, P. Koutsakis and M. Paterakis, "A New Model for Video Traffic Originating from Multiplexed MPEG-4 Videoconference Streams", *Performance Evaluation*, Vol. 65, No. 1, 2008, pp. 51-70.

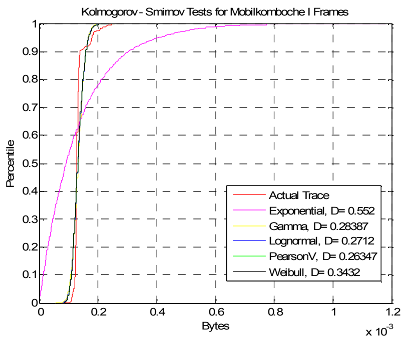

Figure 1. K-S test for the Lecture movie I frames.

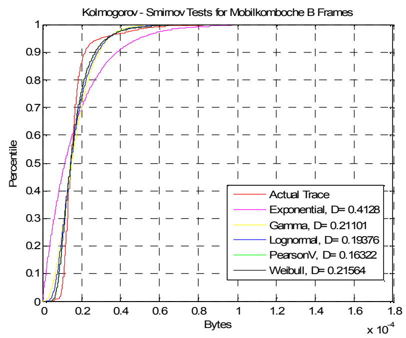

Figure 2. K-S test for the Lecture movie B frames.

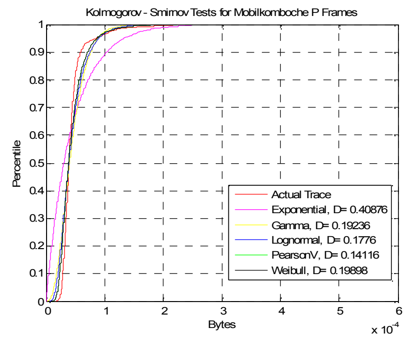

Figure 3. K-S test for the Lecture movie P frames.

Movie	Mean (Mbps)		Peak (Mbps)		Standard Deviation (Mbps)		Revenue Weight	
	HQ	MQ	HQ	MQ	HQ	MQ	HQ	MQ
Office	0.4	0.11	2	1	0.434	0.253	6	2
Lecture	0.21	0.058	1.5	0.69	0.182	0.094	4	1
Boulevard Bio	0.65	0.19	2.6	1.3	0.368	0.197	8	3

Table 1. Statistics for the High and Medium Quality versions of the video traces.

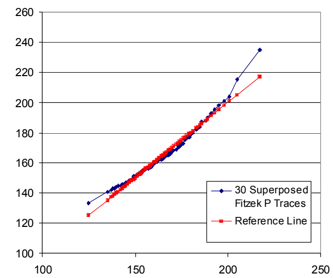

Figure 4. Q-Q plot of DAR(1) model versus the actual office camera trace for the P frames of 30 superposed sources.

Scenario	Office		Lecture		Boulevard Bio	
	HQ	MQ	HQ	MQ	HQ	MQ
1	16.67	16.67	16.67	16.67	16.67	16.67
2	10	30	20	10	20	20
3	20	10	20	10	20	20
4	30	10	20	20	10	10
5	10	20	10	10	10	40
6	40	20	10	10	10	10
7	20	10	30	20	10	10
8	10	30	10	10	20	20
9	10	10	10	20	20	30
10	10	10	40	10	20	10
11	20	20	20	20	10	10
12	10	30	10	20	20	10

Table 2. Traffic Scenarios – Percentages of HQ and MQ "modes".

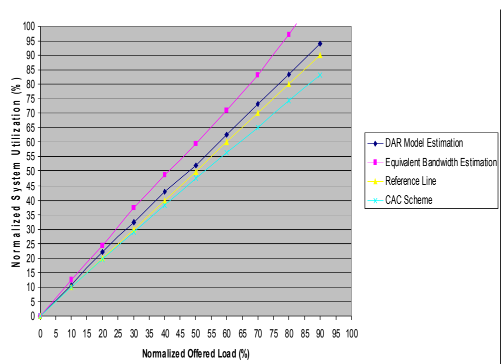

Figure 5. Capacity Utilization with the CAC scheme, 10% handoff.

Scenario	Real Traces-Bandwidth (Mbps) under various handoff loads (%)			DAR Model- Bandwidth (Mbps) under various handoff loads (%)			Actual Bandwidth Used with the CAC Scheme (Mbps) under various handoff loads (%)		
	5%	10%	15%	5%	10%	15%	5%	10%	15%
1	14.55	14.81	14.98	15.26	15.84	16.27	14.27	14.42	14.52
2	11.48	11.72	12.14	12.05	12.52	12.96	11.24	11.35	11.70
3	15.44	15.85	16.19	16.07	16.45	16.81	14.75	15.03	15.31
4	14.23	14.51	15.20	14.96	15.21	15.82	13.67	13.88	14.28
5	11.87	12.23	12.63	12.28	12.93	13.44	11.73	11.99	12.34
6	14.93	15.19	15.61	15.47	15.95	16.47	14.12	14.39	14.65
7	13.07	13.28	13.49	13.72	13.97	14.64	12.42	12.63	12.85
8	13.78	14.01	14.42	14.25	14.78	15.19	13.59	13.78	14.20
9	13.83	14.12	14.64	14.34	14.96	15.35	13.61	13.93	14.36
10	14.71	15.21	15.62	15.29	15.94	16.52	14.03	14.48	14.81
11	12.30	12.62	13.17	12.76	13.39	13.91	11.92	12.16	12.64
12	13.06	13.57	14.07	13.67	14.25	14.88	12.85	13.21	13.72

Table 3. Estimations of the required bandwidth and bandwidth utilization with the CAC scheme.