Performance Analysis of Cell Switching Management Scheme in Wireless Packet Communications

Jongho Bang Sirin Tekinay Nirwan Ansari
New Jersey Center for Wireless Telecommunications
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
University Heights, Newark, NJ 07102

***Abstract* — Third generation wireless networks carry traffic from heterogeneous sources such as voice, data, video or multimedia with a specified quality of service. Efficient traffic management schemes should be developed to provide seamless access to the wireless network. In this paper, we propose a novel cell switching (e.g., packet handoff) scheme to support QoS guarantee for downlink in EGPRS (Enhanced General Packet Radio Service) network. We define the new packets and handoff packets for each type of traffic and devise a new prioritization scheme at the buffer of the base station. We present the procedure to find the optimal thresholds satisfying the QoS requirements. Using the ON-OFF approximation method for the aggregate traffic, we compute the packet loss probability and the packet waiting time. The performance of the proposed scheme is evaluated by simulations and numerical analysis in terms of packet loss probability and packet delay.**

I. INTRODUCTION

The impressive growth of cellular mobile telephony as well as the number of Internet users promises an exciting potential for a market that combines both innovations: cellular wireless data services. Within the next few years, there will be an extensive demand for wireless data services. In particular, high-performance wireless Internet access will be requested by users. The GPRS is a new bearer service for GSM that greatly improves and simplifies wireless access to packet data networks, e.g., to the Internet. GPRS, EDGE (Enhanced Data rate for GSM Evolution), and UMTS (Universal Mobile Telecommunications Services) are all being developed to accommodate data users in wireless networks. EGPRS/EDGE will evolve to third generation (3G) mobile communications while UMTS will make resolution way for third generation mobile communications [1][2].

Traffic management is crucial in wireless networks. At the base station (BS), packets destined for mobile terminals (MTs) are transmitted on the forward channels. It is important to guarantee the QoS for each kind of traffic in the wireless network. Various buffering schemes can be used at the BSs, and packets arriving from a switch will be serviced in several service disciplines across the BS. The maximum radio link throughput is limited and can be expected to be lower than the servicing wired link throughput from the switch. The actual bandwidth allocation must be set according to the wireless link. Packets from the switch, however, can arrive in bursts with a much higher rate than that being serviced over the radio link. This fact explains the requirements of buffering at the BSs. Each burst can cause queueing of packets, and is the main cause of packet loss rate (caused by buffer overflow). General buffer management schemes for congestion control have been proposed; e.g., partial buffer sharing and push-out [3][4][5]. The well-known partial buffer sharing scheme is used in this paper, because different priorities must be given to multi-class traffics.

The focus of this paper is resource allocation for cell switching at BS to satisfy QoS requirements. The QoS requirements are expressed in terms of the packet loss probability and delay. We consider the EGPRS network which is a TDMA-based approach. We propose a technique to define the *new packets* and *handoff packets* for each type of class to give the priority at the BS's buffer; there are two classes of packets for each traffic type. We propose to examine the effect of packet priority scheme at cell switching (e.g., packet tagging and partial buffer sharing scheme). Using the MMPP model for the aggregate ON-OFF traffic streams, we can compute the packet loss probability and the packet waiting time. The performance of the proposed scheme is evaluated by simulations and numerical analysis. We present the procedure to find the optimal buffer thresholds that simultaneously satisfy the QoS requirements for multiple types of classes. The rest of the paper is organized as follows. Section II describes the model; Section III presents packet tagging; Section IV presents cell switching; Section V gives the performance analysis; finally, Section VI presents the conclusions.

II. MODEL DESCRIPTION

There are N types of traffic, labeled $n = 1, 2, 3, ..., N$. Each type of traffic has two priority classes; new packets and handoff packets. These have different priorities at the buffer. There are $I = 2N$ of total priority classes. Buffer size at the BS is assumed to be limited. The number of thresholds at the BS's buffer is the same as the number of total priority classes. The threshold is denoted as S_i for $i = 0, 1, 2, ..., I-1$ where S_{I-1} is the buffer size. All packets from switches are aggregated at the BS and enter the single buffer. They are prioritized and then served. The QoS requirements for each priority class are assumed to be packet loss probability, PLP_i, and packet delay, D_i. The real time traffic such as voice and video has delay requirement and packet loss requirement. The non-real-time

0-7803-7206-9/01/$17.00 © 2001 IEEE

traffic such as data is more tolerant to delay, but has more stringent requirement for packet loss probability.

III. PACKET TAGGING

Each cell, served by a BS, is divided into two zones based on the thresholding the received power from the MT at the BS; zone A_1 and zone A_2. The MT in zone A_1 starts to communicate by sending new call request and sends the handoff call request to the BS when it is across zone A_2 during communication. Therefore, zone A_1 is the area for new call generation and zone A_2 is the area for handoff call generation based on received signal strength. The basic traffic model assumes that the new call origination rate and handoff call rate are uniformly distributed over zone A_1 and zone A_2, respectively. We denote the average rate of new origination in zone A_1 by Λ_n and the average rate of handoff in zone A_2 by Λ_h. Here, call arrivals are assumed to be Poissonian.

We assign handoff packets higher priority over new packets at the BS's buffer. In order to give priority to handoff packets over new packets, new and handoff packets have to be differentiated. In this case, QoS parameters of new packets and handoff packets could be criteria for call acceptance. How can we define new and handoff packets?

New packets and handoff packets can be defined by the generation areas (e.g., zone A_1 and zone A_2). That is, MT tags new packets and handoff packets based on the coverage zone, as shown in Fig. 1; within zone A_1 (e.g., coverage radius r), new packets; within zone A_2 (e.g., coverage between radius R and r), handoff packets.

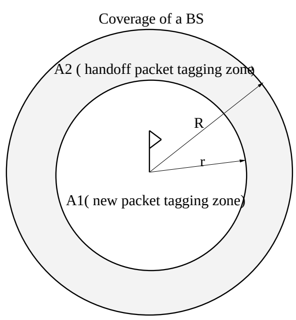

Fig. 1. New and Handoff Packet Tagging Zone

MT knows where it is, based on the received signal strength from the BS. Note that the areas of zones A_1 and A_2 are

$$S_{A_1} = \pi r^2, \qquad S_{A_2} = \pi(R^2 - r^2).$$

It is assumed that a mobile tags handoff packets, rather than new packets, when a call is originated in zone A_2. Let us consider uniform traffic distribution over the service area. A given packet tagged in a cell belongs to zone A_1 with probability p_1 and to zone A_2 with probability p_2, where $p_1 = S_{A_1}/(S_{A_1} + S_{A_2})$ and $p_2 = 1 - p_1$.

IV. CELL SWITCHING

We consider a general partial buffer sharing scheme [3][4][5]. Priority packet discarding is a popular congestion control technique in high-speed networks that allocates network resources more efficiently, thereby making it easier to satisfy QoS requirements of different classes of traffics. In general, loss-sensitive traffic such as data is given priority over loss-tolerant traffic such as voice and video. Real time packets are dropped from a buffer when the buffer occupancy reaches the threshold. In this work, we consider a threshold-based discarding (TBD) scheme.

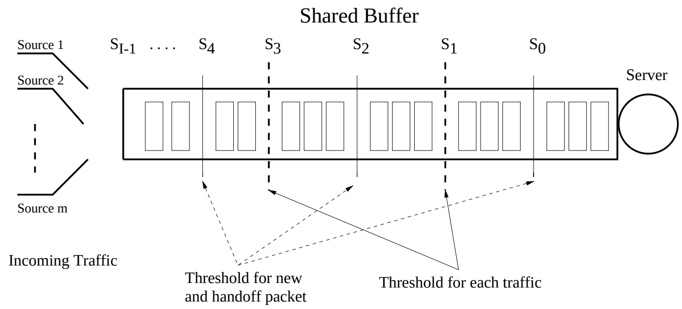

Fig. 2. Threshold-Based Discarding Scheme handling New and Handoff Packets

As shown in Fig. 2, the buffer is partitioned by I thresholds, S_0,...,S_{I-1}, (e.g., S_{I-1} is the buffer size), corresponding to I priority classes. Packets of priority class i can be buffered up to threshold level S_i. Once the buffer level exceeds S_i, arriving packets of class i are dropped. Note that only new arrivals are dropped; class i packets that are already in the buffer are never dropped and are eventually served.

In order to give priority to handoff packets, some buffer space is reserved for handoff packets of each type of traffic. Thresholds S_0 and S_1 are for new packets and handoff packets of traffic type $n = 1$, respectively; the thresholds S_2 and S_3 are for new packet and handoff packet of traffic type $n = 2$, respectively, and so on.

V. PERFORMANCE ANALYSIS

The Markov Modulated Poisson Process (MMPP) has been commonly used for modeling arrival rates of point processes.

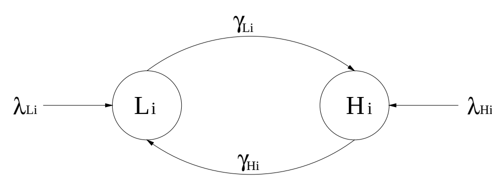

Fig. 3. 2-state MMPP model for class type i

The accuracy of MMPP in modeling an arrival process depends on which statistics of the actual process are used to determine its parameters. In [6][7], the superposition of ON-OFF sources is approximated by means of a 2-state MMPP for each traffic. Four parameters are required to represent the 2-state MMPP source of each traffic, as shown in Fig. 3, where $\gamma_{L_i}(\gamma_{H_i})$ is defined as the mean transition rate out of the Low load i (High load i) state, and $\lambda_{Li}(\lambda_{H_i})$ is the mean arrival rate of the Poisson process in the Low load i (High load i) state for priority class type i.

The stochastic integral technique proposed in [8] is used to obtain loss probabilities for Markov Modulated Arrival (MMA) streams. Individual packet loss expressions are devised using the stochastic integral approach [9]. We adopt these methods to derive the packet loss probability of our model.

A. COMPUTATION OF PACKET LOSS PROBABILITY

We define the limiting probabilities (in the case, $N = 3$). Let $\pi(k, l, m, n, o, p, q)$ $(k = L_0, H_0,\ \ l = L_1, H_1,\ \ m = L_2, H_2,\ \ n = L_3, H_3,\ \ o = L_4 H_4\ \ \ p = L_5, H_5\ \ 0 \leq q \leq S_{I-1})$ be the limiting distribution for the Markov process, where L_i and H_i are the state of superposed arrival for i priority class type, and q is the buffer state, and S_{I-1} is the buffer size. Note that $\sum_l \sum_m \sum_n \sum_o \sum_p \pi(k, l, m, n, o, p, q) = \pi(k, q)$, $\sum_k \sum_m \sum_n \sum_o \sum_p \pi(k, l, m, n, o, p, q) = \pi(l, q)$, $\sum_k \sum_l \sum_n \sum_o \sum_p \pi(k, l, m, n, o, p, q) = \pi(m, q)$ and so on.

We denote $P_i(q)$ as the probability that an arrival from priority class type i sees the system in state q. Using the stochastic integral technique and the limiting probabilities, we obtain

$$\begin{aligned} P_i(q) &= \frac{(\gamma_{L_i} + \gamma_{H_i})(\lambda_{L_i}\pi(L_i, q) + \lambda_{H_i}\pi(H_i, q))}{\lambda_{L_i}\gamma_{H_i} + \lambda_{H_i}\gamma_{L_i}}, \\ & for \quad i = 0, 1, 2, .., I-1. \end{aligned} \tag{1}$$

Packet loss probability, $P_i(loss)$, for each priority class type are following;

$$P_i(loss) = \sum_{q=S_i}^{S_{I-1}} P_i(q), \qquad for \quad i = 0, 1, 2, .., I-1. \tag{2}$$

B. COMPUTATION OF PACKET WAITING TIME

As in the ordinary M/G/1 queueing system, the average number of queue and the mean residual time are considered to find the average packet waiting time. These two random variables indicate the average number of queue and mean residual time seen by an outside observer at a random time [10]. We can also adopt this concept to find the packet waiting time of our system. We define the following parameters.

- N_{Q_i} is defined as the average number of packet in queue seen by a packet from the priority class i.
- R_i is defined as the average residual time seen by a packet from the priority class i.
- r_i is the rate of arrivals from the priority class i.
- $\overline{X}$ is defined as the average packet transmission time.
- $\overline{X^2}$ is defined as the second moment of $\overline{X}$.

From equation (1),

$$N_{Q_i} = \sum_{q=1}^{S_i} (q-1)P_i(q), \qquad for \quad i = 0, 1, 2, .., I-1. \tag{3}$$

From equation (2),

$$\begin{aligned} R_i &= \frac{\{r_i(1 - P_i(loss))\}\overline{X^2}}{2}, \\ & where \quad r_i = \frac{(\lambda_{L_i}\gamma_{H_i} + \lambda_{H_i}\gamma_{L_i})}{\gamma_{L_i} + \gamma_{H_i}}. \end{aligned} \tag{4}$$

Therefore, $W_i(delay)$ denoted by the average packet waiting time is,

$$W_i(delay) = R_i + N_{Q_i}\overline{X}. \tag{5}$$

We use the packet loss probability and the packet delay as criteria for call admission [11] [12]. There are many performance measures for call admission in Wireless Packet Network such as packet loss rate, packet delay time, handoff failure rate, new call blocking rate and so on. Among these measures, the handoff failure rate and new call blocking rate can be assumed to be dependent on the packet loss rate and the packet delay time. Thus, from the QoS guarantee perspective, the packet loss rate and delay are dominant criteria for CAC operation. $n_0, n_1, n_2, ...$ and n_{I-1} is defined as the number of users for priority class type $i = 0$, $i = 1$, $i = 2$, ..., and $i = I-1$, respectively. Call admission for the user of priority class type $i = 0$ is determined by computing the values of $P_{loss}(n_0 + 1, n_1, n_2, ..., n_{I-1})$ and $W_0(delay)$ at this moment, and then comparing with the QoS requirements for priority class type $i = 0$, PLP_0 and D_0. If the value of $P_0(loss)$ is larger than PLP_0 or $W_0(delay)$ is larger than D_0, the connection request is denied because QoS cannot be guaranteed.

C. OPTIMIZING THRESHOLD VALUES

We now turn our attention to the problem of finding the thresholds to satisfy PLP_i and D_i (i.e., packet loss probability requirement and delay requirement for each priority class

i, where $i = 0, 1, 2, ..., I-1$). At first, initial threshold for each priority class is assumed to be arbitrary and small. The thresholds are increased by the program until the QoS requirements are simultaneously satisfied for all priority classes. At each step, packet loss probabilities, delays and normalized difference values (i.e.., ($P_i(loss)$ - PLP_i)/PLP_i) are calculated. When all the QoS requirements are not satisfied, the maximum difference value is found. Then, the threshold for the priority class with the maximum difference value is increased by one. When all the QoS requirements for packet loss probability and delay are satisfied, the search procedure is terminated. Buffer size is decided according to QoS requirements of the highest priority class. The search procedure is as follows:

Step 1) Set $S_i = 1$ for all i

Step 2) Calculate $P_i(loss)$ and $W_i(delay)$ for all i

If $P_i(loss) \leq PLP_i$ and $W_i(delay) \leq D_i$ for all i, then terminate. S_i is the optimal threshold.

Otherwise, $S_k = S_k + 1$, where k= the index i for which ($P_i(loss)$ - PLP_i)/PLP_i is maximum.

Then, repeat Step 2.

VI. SIMULATION MODEL AND RESULTS

We use the ON/OFF model to describe the multi-class sources. In this simulation, 3 types of traffics (i.e., 6 types of priority classes; two priority classes for each traffic) are used. The type 1 traffic with 32$Kbps$ rate is modeled with the parameter values: mean ON period (=1.0 s) and mean OFF period (=1.35 s). The type 2 traffic with 320 $Kbps$ rate is modeled as the superposition of multiple identical ON/OFF source; that is, one source is achieved by the superposition of 15 ON/OFF sources , each characterized by the mean ON period (=33 ms) and the mean OFF period (=67 ms) [13]. The type 3 traffic with 128 $Kbps$ rate is modeled with the parameter values: mean ON period (=0.1 s) and mean OFF period (=0.8 s). We assume that packet length is exponentially distributed with mean 1024 bytes and system capacity is 4.8 $Mbps$ [14].

Computer simulations are conducted to investigate the packet-level performance of the handoff scheme. New arrivals and handoff arrivals follow independent and identical Poisson distribution. The fraction of total traffic due to each traffic type is fixed (e.g, the arrival fraction of each traffic type is 46%, 8% and 46%). Also, the fraction of each traffic due to handoffs is kept fixed while the total offered traffic is varied (e.g., the fraction of handoff packet for each traffic is fixed).

In Fig. 4, packet loss probabilities are plotted as a function of the mean offered load. Note that the simulation results conform to our numerical analysis. The thresholds and buffer size are assumed to be 11, 15, 18, 20, 27 and 29 for priority class type 0, 1, 2, 3, 4, and 5, respectively. In order to get this result, the fraction of new call and handoff call for each traffic is fixed in zone A_1 and A_2. That is, the fraction of new packet and handoff packet (e.g., 50% new packets and 50% handoff packets) is proportional to the ratio of zone A_1 and A_2 (e.g., A_2=A_1). When a call is originated in zone A_2, we assume that a mobile tags handoff packets, rather than new packets.

In Fig. 5, the effect of the fraction of handoff packet for each traffic on packet loss probability is shown. We compare the fraction (e.g., 50% new packets and 50% handoff packets) and the fraction (e.g., 70% new packets and 30% handoff packets). At the fraction of 30% handoff packet, packet loss probabilities is decreased compared to the fraction of 50%, even though the fraction of new packet is increased.

In Fig. 6 and Fig. 7, packet loss probabilities and thresholds are plotted as a function of iteration to show the search for optimal thresholds, respectively. Optimal thresholds satisfying the QoS requirements (e.g., packet loss probability and delay) are obtained at the offered load of 0.8 with the fraction of 50% handoff packets. The packet loss probability requirements are assumed to be 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-10} and 10^{-12} for priority class 0, 1, 2, 3, 4, 5, respectively. The delay requirements are assumed to be 68, 76, 85, 93, 110 and 120 (ms). Since initial threshold values were assumed to be small, at first, packet loss probability for each priority class is much greater than its QoS requirement. As the number of iteration is increased, packet loss probabilities are decreased and thresholds are increased gradually in meeting the QoS requirements. The search procedure starts at thresholds (10, 15, 20, 25, 30, 35) and terminates at thresholds (41, 46, 49, 52, 60, 62) with meeting the QoS requirements.

VII. CONCLUSIONS

In this paper, we consider a novel cell switching scheme to support QoS guarantee in packet-switched wireless cellular networks. The method to examine the effect of packet priority scheme (e.g., partial buffer sharing scheme) at cell switching is proposed. That is, using the packet tagging method, we differentiate packets transmitted in the same cell into new packets and handoff packets, with handoff packet having higher priority. By modeling each type of priority class by a 2-state MMPP, we were able to drive the packet loss probability and the packet waiting time of the respective priority classes using the space priority scheme. The performance of the proposed scheme is evaluated by simulations and numerical analysis in terms of packet loss probability and packet delay. Optimal thresholds at the specific offered load can be obtained through the proposed search procedure.

REFERENCES

[1] S. Faccin *et al*., "GPRS and IS-136 Integration for Flexible Network and Services Evolution," IEEE Personal Communications, pp. 48-54, June 1999.

[2] X. Qiu *et al*., "RLC/MAC Design Alternatives for Supporting Integrated Services over EGPRS," IEEE Personal Communications, pp. 20-33, April 2000.

[3] H. Kroner, G. Hebuterne and P. Boyer, “Priority Management in ATM Switching Nodes,” IEEE JSAC, vol. 9, no. 3, pp. 418-427, Apr. 1991.

[4] M. Krunz, H. Hughes and P. Yegani, “Design and Analysis of a Buffer Management Scheme for Multimedia Traffic with Loss and Delay Priorities,” IEEE Proc. Globecom’94, pp. 1560-1564, Nov. 1994.

[5] D. W. Petr and V. S. Frost, “Nested Threshold Cell Discard for ATM Overload Control: Optimization under Cell Loss Constraints,” IEEE Proc. Infocom’91, pp. 12A.4.1-12A.4.1.10, 1991.

[6] S. Kim, M. Lee and M. Kim, “Σ-Matching Technique for MMPP Modeling of Heterogeneous ON-OFF Sources,” IEEE Proc. Globecom’94, pp. 1090-1094, 1994.

[7] J. Bang, N. Ansari and S. Tekinay, “Selective-delay Push-in Buffering Mechanism for QoS provisioning in ATM Switching Nodes Loaded with ON-OFF Arrival Processes,” The 15th International Conference on Information Networking (ICOIN-15), Beppu, Japan, pp. 799-804, Jan. 2001.

[8] W. A. Rosenkrantz and R. Simha, “Some Theorems on Conditional Pasta: A Stochastic Integral Approach,” Operations Research Letters, vol. 11, pp. 173-177, April 1992.

[9] J. B. Kim, R. Simha and T. Suda, “Analysis of a Finite Buffer Queue with Heterogeneous Markov Modulated Arrival processes: A Study of Traffic Burstiness and Priority packet Discarding,” Computer Networks and ISDN Systems 28 pp. 653-673, 1996.

[10] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, New Jersey, 1992.

[11] M. Naghshineh and M. Schwartz, “Distributed Call Admission Control in Mobile/Wireless Networks,” IEEE JSAC., vol. 14, no. 4, pp. 711-717, May 1996.

[12] S. K. Biswas and B. Sengupta, “Call Admissibility for Multimedia Traffic in Wireless ATM Networks,” IEEE Proc. Infocom’97, pp. 649-657. 1997.

[13] X. Wu, S. Wu and H. Sun, “Dynamic Slot Allocation Multiple Access Protocol for Wireless ATM Networks,” IEEE Proc. Icc’97, pp. 1560-1565, 1997.

[14] GSM 03.60, V5.2.0, “Digital Cellular Telecommunication System (Phase 2); General Packet Radio Service (GPRS); Service Description (Stage 2),” Dec. 1997.

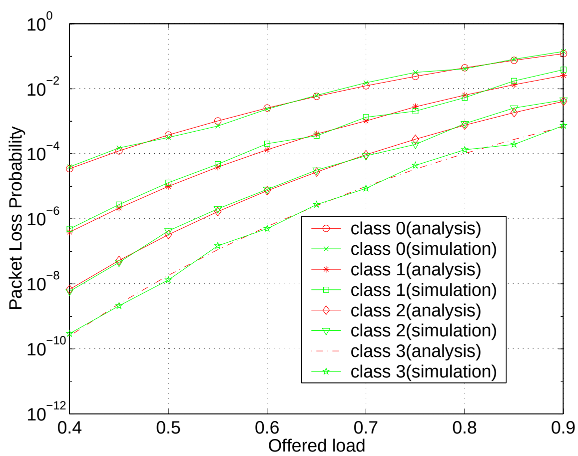

Fig. 4. packet loss probabilities vs. offered load (comparison between analysis and simulation)

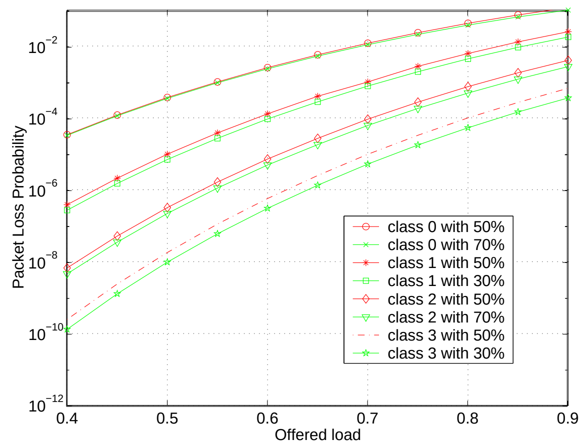

Fig. 5. packet loss probabilities vs. offered load (different fractions of handoff packet)

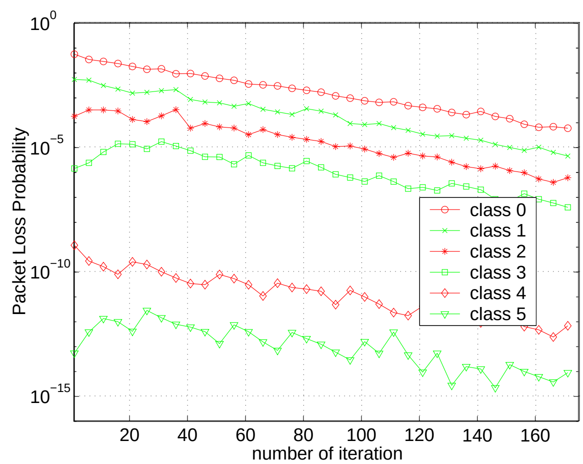

Fig. 6. packet loss probabilities vs. number of iteration (offered load 0.8)

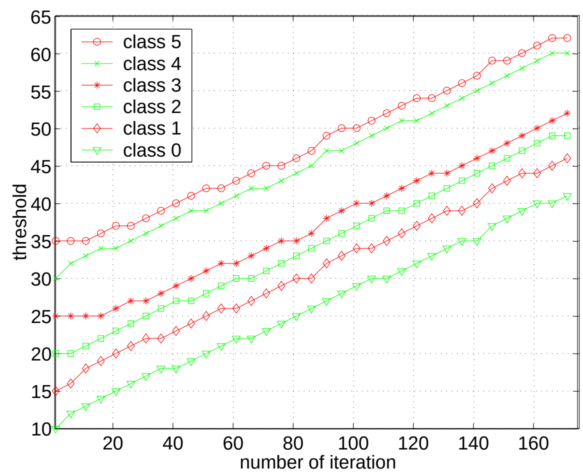

Fig. 7. thresholds vs. number of iteration (offered load 0.8)