Multiple Class Selective Discard Under a Long-Range Dependent Process

N.L.S Fonseca e M.J. Ferreira

Relatório Técnico IC 98-05

Abril de 1998

Multiple Class Selective Discard Under a Long-Range Dependent Process

Nelson L. S. Fonseca and Marcelo J. Ferreira
State University of Campinas
Institute of Computing
Brazil
{nfonseca, mjf}@dcc.unicamp.br

Abstract

Providing diverse QoS requirements in multimedia networks is a challenging task. Under a long-range dependent process, the loss rate does not decrease considerably as we increase the buffer size. In this paper, we investigate the advantages of adopting a multi-priority selective discard mechanisms over traditional two-priority mechanisms under a long range dependent process.

I) INTRODUCTION

Several studies [1]-[2] have claimed that different types of network traffic, *e.g*. local area network traffic (LAN), can be accurately modeled by a self-similar process. A self-similar process is able to capture the long-range dependence (LRD) phenomenon exhibited by this traffic. Moreover, series of simulation and analytical studies [3]-[6] demonstrated that this phenomenon might have a pervasive effect on queueing performance, *i.e.*, there is clear evidence that it can potentially cause massive cell losses in ATM networks. In fact, Norros [5] and Duffield [6] showed that the buffer overflow probability for an ATM queueing system with fractional Brownian arrivals follows a Weibull distribution. Furthermore, this queueing system suffers from the buffer inefficacy phenomenon [4], [7], *i.e.*, by just increasing the buffer size, one is not able to decrease the buffer overflow probability considerably.

Different multimedia applications have diverse loss requirements. For instance, a telephone conversation may afford a loss rate of 10^{-3}, whereas an MPEG video transmission may tolerate a loss rate of the order of 10^{-9}. Coping with different loss requirements is a challenging task. Selective discard is a congestion control mechanism aimed at enabling the network to deal with diverse loss requirements. In a selective discard mechanism cells are discarded in an overflow situation according to their priority level [8].

Selective discard has been studied in the past few years. However, only recently, has it been investigated under long-range processes. In [9], we analyzed selective discard with two priority levels under a long-range dependent process. We found out that the complete sharing with push-out buffer policy is clearly worth adopting, while complete sharing with guaranteed queue minimum buffer policy may be not. It was also evident that the choice of push-out policies has a significant impact on loss performances.

A multiple class selective discard mechanism furnishes more than two priority levels. Besides providing a higher number of distinct bearer, a multiple class mechanism demands less buffer space to support diverse QoS than does a two-priority mechanism. In other words, with a higher number of priority classes we do not need to guarantee loss rate bounds lower than the required. Additionally, we may carry higher loads than does a two-priority mechanism. Multiple class selective discard under a short-range dependent process (SRD) is not as attractive as it is under a long-range dependent process. In fact, under a SRD process as we increase the buffer size, we decrease the loss rate significantly [10].

In this paper, we investigate the advantages of adopting a multi-priority selective discard mechanisms under a long range dependent process. We consider a complete sharing queue under a fBm, and evaluate the impact of the offered load, variance and Hurst parameter on the per class loss rate and on the loss gap length.

This paper is organized as follows: Section II introduces the buffer inefficacy phenomenon. Sections III and IV describe selective discard mechanism and selective discard under a long range dependent process, respectively. Section V compares a multiple-priority mechanism with a two-priority mechanisms. Finally, conclusions are drawn in section VI.

II) THE BUFFER INEFFICACY PHENOMENON

The buffer inefficacy phenomenon is the queueing phenomenon in which by just increasing the buffer size, we are not able to decrease the buffer overflow probability considerably. This phenomenon has been reported earlier by several other studies [11]-[13]. In this section, we present a very intuitive explanation for it and show that it is of particular importance when the traffic source exhibits long-range dependencies.

We model an ATM node as a deterministic

queueing system with constant departure rate given by *c* and finite buffer size given by *b*. The input traffic is given by the stochastic process *A(t)* with mean input rate $\bar{a} < c$. It defines the aggregate number of cell arrivals up to time $t, t \geq 0$. Assume that the buffer overflow occurs at time *t* so that we can write $A(t) = ct + b$. Moreover, $A(t)/t \geq c + b/t$.

By the law of large numbers, the average arrival rate $A(t)/t$ converges to its mean $\bar{a}$. Therefore, the probability that it exceeds the term $(c + b/t)$ decreases with *t*: $P(A(t)/t \geq c + b/t) = \psi(t)$

In other words, $\Psi(t)$ is a decreasing function with time. The buffer inefficacy phenomenon occurs, if the buffer overflow probability given by $\Psi(t)$ decays *slowly* with *t*, *i.e.* if $\Psi(t)$ is non-negligible for large *t*. In this case, since *t* is large, the term (b/t) is negligible. Therefore, even if we increase the buffer size, we are not able to increase the term (b/t) significantly in order to decrease the cell loss probability. Intuitively, this phenomenon occurs if the arrival process is able to transmit at *high* rates for very long periods of time, *i.e.* if it converges slowly to its mean. We show that a LRD source can transmit at *high* rates for very long periods of time. Following Norros' work [5], assume that the arrival process $A_H(t)$ is a fractional Brownian motion (fBm) process given by $A_H(t) = \bar{a}t + \sigma Z(t)$ where $\bar{a} > 0$ is the mean input rate, $\sigma > 0$ is the standard deviation, $H \in [1/2, 1)$ is the self-similar (Hurst) parameter and *Z(t)* is a normalized fractional Brownian motion. When *H=1/2*, we have the special case of the ordinary Brownian motion. The probability that over a time interval of length *t* the source $A_H(t)$ can overcome the potential service *ct* and further exceed a buffer level *b* is given by:

$$P\left(A_H(t) \geq ct + b\right) = P(\bar{a}t + \sigma Z(t) > ct + b)$$

$$= P\left(Z(t) > \frac{t(c - \bar{a}) + b}{\sigma}\right)$$

By the self-similarity property $Z(t) = t^H Z(1)$ we have:

$$P\left(Z(1) > \frac{t(c - \bar{a}) + b}{\sigma t^H}\right) = \bar{\Phi}\left(\frac{t(c - \bar{a}) + b}{\sigma t^H}\right)$$

where $\bar{\Phi}(y) = P(Z(1) > y)$ is the residual distribution function of the standard Gaussian distribution. In fact, using the approximation:

$$\bar{\Phi}(y) \approx (2\Pi)^{-1/2}(1 + y)^{-1} exp\left(\frac{-y^2}{2}\right) \approx exp\left(\frac{-y^2}{2}\right)$$

we obtain:

$$P\left(A_H(t) > ct + b\right)$$

$$= \bar{\Phi}\left(\frac{t(c - \bar{a}) + b}{\sigma t^H}\right) \approx exp\left(-\frac{1}{2}g(t)^2\right)$$

$$= exp\left(-\frac{1}{2}\left(\frac{t(c - \bar{a}) + b}{\sigma t^H}\right)^2\right) \qquad (1)$$

We compute Equation 1 for two sources with same mean, standard deviation and Hurst parameter *H=0.50* and *H=0.85* respectively. We choose the link bandwidth so that the link utilization given by is *50%*. Figure 1 shows the results. We can see that the probability of buffer overflow for the LRD source decays very slowly with time. Therefore, increasing the buffer size is not enough to accommodate the strong low-frequency component of this source in order to avoid cell losses. On the other hand, ψ decays very fast in the case of uncorrelated arrivals (Brownian motion) [12]-[13].

III) SELECTIVE DISCARD MECHANISM

In a selective discard mechanism cells are discarded in overflow situations according to their priority level. A selective discard mechanism is completely specified by a buffer policy and by a push-out policy. While a buffer organization policy defines which buffer slot can be occupied by which cell, a push-out policy chooses a cell to be discarded among the cells with lowest priority.

Although the loss rate (the ratio between the number of lost cells and the total number of transmitted cells) is a meaningful and measurable parameter, it is an average value which does not entirely describe the loss process entirely. The number of cells consecutively lost (loss gap) gives a more detailed description of the loss process. For a certain value of loss rate, cells may be lost in several different ways. For instance, for a loss rate of 0.25, we may lose one cell out of every four cells or we may lose a quarter of the total number of cells in a role. Depending on the signal recovery procedure at the receiver side, the length of the loss gap may have different impact on user's perceived QoS.

Our investigation considers the Complete Sharing (CS) with Push out buffer policy. In a CS policy if a high priority cell finds the buffer full it drops an enqueued low priority cell. CS is a loss conserving discipline. In a loss conserving discipline, (fixed size) cells are lost only in overflow situations. In other words, a loss-conserving discipline always admits a cell into the buffer if there is available space. Loss-conserving queues are of special interest because they minimize the overall cell loss, and consequently maximizes the throughput.

A push-out policy selects a cell to be dropped among cells with lowest priority. The most common policies are Last-In-First-Drop (LIFD), First-In-First-Drop (FIFD), Random selection (RAND) and Modified-FIFD (M-FIFD) [14]. The modified-FIFD policy always drops the oldest low priority cells to make room for an arriving cell irrespective of its class. Dropping low priority cells at different positions define distinct queue distribution over a certain period of time and consequently may differently affect the perceived QoS.

An in-depth view of selective discard mechanism may be found in [8].

IV) SELECTIVE DISCARD UNDER A LONG-RANGE DEPENDENT PROCESS

In [9], we investigated the extent to which selective discard under a long-range dependent process can provide differentiated QoS. We analyzed complete sharing with push out and complete sharing with guaranteed queue minimum (CSGQM) buffer policies. Similarly to CS, in a CSGQM, low priority cells are discarded during buffer overflows, however, we guaranteed a minimum queue length to the low priority cells in an overflow scenario.

We verified that under CS and for Hurst parameter lower than 0.75, the high priority loss rate decreases considerably as the buffer size increases. Although for higher values of the Hurst parameter, the high priority loss rate decreases very slowly, CS produces differentiated per class loss rates. Conversely, for guaranteed queue length higher than 20% of the total buffer size, the per class priority loss rate given by CSGQM becomes indistinguishable.

The choice of push-out policy impacts significantly loss performances. M-FIFD may produce high priority loss rates three orders of magnitude lower than LIFD. On the other hand, LIFD may give maximum loss gap size which are typically half of the maximum loss gap size given by M-FIFD.

We also notice that the per class loss rate and loss gap size are highly sensitive to the Hurst parameter, as well as to the variance.

V) NUMERICAL EXAMPLES

To assess the effectiveness of multiple class selective discard under long-range dependent process, we simulate a four-priority level CS queue with constant service time and with a fractal Brownian motion input process. The fractal Brownian motion is generated according to Mandelbrot and van Ness's procedure, *i.e.*, the fBm is obtained by superposing high and low frequency Markov processes. Moreover, we consider that the priority level of a cell is independent of the priority level of other cells in the flow. We obtained 95% confidence intervals via the independent replication method.

In Figure 2, we display the per class loss rate as a function of the buffer size. We notice that in a four-level priority queue we are able to offer distinct loss rates. Moreover, as the buffer size increases we differentiate even more the loss rates. If we had a two-priority queue instead of a four-priority queue, we either would have to carry classes 2 and 3 with a higher loss rate or we would need to have a much larger buffer size. For instance, we would need a buffer space larger than 15000 slots to provide a loss rate of 10^{-6} instead of 700 buffer slots as required in a four-level queue.

We observe that the per class loss decay rate are highly influenced by the proportion of cells with highest priority. In Figure 3, we show the same scenario as is in Figure 2 except that the proportion of cells with highest priority (P_1) is 0.55 instead of 0.4. The change in the decay rate is noticeable when compared to Figure 2. Therefore, to provide diverse loss rates we need to keep the proportion of cells carried with highest priority low. Moreover, the decay rate is also affected by the Hurst parameter. Figure 4 shows that as we increase the Hurst parameter from 0.75 (Figure 4.a) to 0.8 (Figure 4.b) the decay rate of classes 2 and 3 decreases. In other words, as the Hurst parameter increases, so does the overall loss rate. As a consequence, we are able to decrease only the highest priority loss rate.

Providing diverse loss rate bounds imply that we can satisfy loss requirements with much less buffers. In a two-priority mechanism, we need to aggregate applications with diverse QoS requirements and transport the aggregated load according to the QoS requirements of the most stringent aggregating application. For instance, consider a four-priority level queue with buffer size 100, under offered load of 0.8, in which the proportion of cells in each priority class from the highest to the lowest is respectively: 0.55, 0.05, 0.15 and 0.25. In this scenario, it is possible to furnish 10^{-7} and 10^{-5} loss rates for the two highest priority classes. In a two-priority level queue, under the same traffic condition we would need 1200 buffer slots.

Additionally, the buffer demand is strongly affected by the Hurst parameter as well as by the variance of the input process. In the previous example, if the variance were 1.2 instead of 0.5 we would be able to furnish loss rates of 10^{-5} and 10^{-3} to the two highest classes in a four-priority queue, whereas we would need 2500 buffers in a two-priority level queue.

In Figure 5, we plot the per class loss rate as a function of the input process variance. We notice that even for higher values of the variance, a multiple class selective discard mechanism provides distinct per class loss rate. This is quite different from a two-priority level mechanism. In the latter, as the variance increases both high and low priority loss rates converge to the same value.

Another advantage of a multi-priority mechanism over a two-priority mechanism is that we can have higher loads in a multi-priority system In fact, we can increase the offered load by increasing the load with less stringent QoS. For instance, in a two-priority queue with buffer size 500 and offered load of 0.65 (H = 0.75, P_1 = 0.7 and P_2 = 0.3), if we introduce more two priority levels, we are able to increase the offered load up to 0.9 by distributing the additional load among the two new lowest priority classes.

In Figure 6 we show the per class loss rate given by M-FIFD and by LIFD. We notice that for large buffer size the same trend found for the high priority loss rate in two-priority queue emerged for the highest priority class in a four-priority queue. The push out policy impacts more significantly the highest priority loss rate.

Although useful, the loss rate is an average value which does not fully describe the loss process. On the other hand, the length of the loss gap provides more information about the loss process. Using the average gap length to compare different policies may be misleading given that most of the loss gap are of small size which brings the mean to a value which hides relevant information. Therefore, we focus our discussion on the gap length distribution and especially on the maximum values of the gap length. In the next figure we display a set of distributions associated with one replication run of our simulation experiment. We show the set of distributions which best represents the average pattern observed among all replications.

Figures 7 and 8 show the loss gap length for class 2, 3 and 4 produced by M-FIFD and by LIFD, respectively. In a four-priority queue, the loss gap is distributed among the several priority classes. Therefore, the maximum loss gap length of the lowest priority class in a four-priority queue is lower than the maximum loss gap length in a two-priority queue. In this specific example, the corresponding maximum loss gap produced by M-FIFD and by LIFD in a two-priority queue are 1800 and 1200, respectively. The maximum loss gap length produced by M-FIFD and by LIFD in a four-priority queue does not differ as much as they do in a two-priority queue. However, the frequency of long loss gap under M-FIFD is much higher than it is under LIFD.

The Hurst parameter has a significant impact on the loss gap length. If we consider the same traffic scenario of Figure 7 but with H = 0.75 the maximum loss gap length would be half the value found in Figure 7.The lowest priority loss gap in a four-priority queue is less influenced than the low priority loss gap in a two-priority queue since the lost cells are distributed among other priority classes. For instance, for the same traffic scenario and for the same Hurst parameter the low priority loss gap in a two-priority queue grows three times as much its original value.

VI) CONCLUSIONS

In this paper we investigated the advantages of adopting a multiple-class mechanism over a two-priority mechanism. We showed that in a multi-priority queue, we can carry the same load with less buffer. In addition, we can carry higher loads than in a two-priority queue by increasing the load of less stringent applications. Moreover a multi-priority scheme has flexibility to deal with traffic stream with high variance.

If on the one hand, M-FIFD produces the lowest highest priority loss rate among all policies, on the other hand, the maximum loss gap length given by LIFD is in the same order of the maximum loss gap given other policies. Therefore, M-FIFD seems to be the most appropriate push-out policy to be adopted in an ATM multiplexer. Insofar as results given by FIFD are close to results given by M-FIFD and FIFD is less complex to implement (it requires less buffer shifting than M-FIFD), FIFD is a more attractive choice.

While multiple-priority mechanisms does not seem to be not so attractive under short range dependent process they are very appealing under long-range dependent process, since they offer differentiated service even when massive loss occurs.

VII) REFERENCES

[1] W. Leland, M. Taqqu, W. Willinger and D. Wilson, “On the Self-Similar Nature of Ethernet Traffic (Extended Version)”, *IEEE/ACM Transaction on Networking*, vol 2, no 1, pp. 1-15, February 1994.

[2] V. Paxson and S. Floyd, “Wide-Area Traffic: The Failure of Poisson Modeling”, In *Proc. of ACM SIGCOMM*, 1994.

[3] G. Mayor and J.Silvester, “A Trace-Driven Simu-

lation of an ATM Queueing System with Real Network Traffic”, *Proc. of IEEE ICCCN*, 1996.
[4] G. Mayor and J. Silvester, “An ATM Queueing System with a Fractional Brownian Noise Arrival Process”, *Proc. of IEEE ICC*, 1996.
[5] I. Norros, “A Storage Model with Self-Similar Input”, Queueing Systems 16, 1994.
[6] N. Duffield, J. T. Lewis, N. O’Connell, R. Russel, F. Toomey, “Predicting Quality of Service with Long-range Fluctuations”, In *Proc of IEEE ICC’95*, pp. 473-477, 1995.
[7] H. D. Sheng and S. Q. Li,”Spectral Analysis of Packet Loss Rate at a Statistical Multiplexer for Multimedia Services", *IEEE/ACM Transactions on Networking*, January 1994.
[8] R. Choen, N.L.S. Fonseca and M. Zukerman, “Traffic Management and Network Dimensioning”, To appear in Multimedia Communications Networks, M Tatipamula and B. Kasnabish Editores, Artech House, 1997 .
[9] N.L.S. Fonseca, M.J. Ferreira and G.S. Mayor, “On the Effectiveness of Loss-Conserving Discipline Under a Long-Range Dependent Process”, in Proc of Globecom’97, pp. 1102-1107, 1997
[10] N. L. S. Fonseca e J. A. Silvester, “Estimating the Loss Probabilty in a Multiplexer Loaded with Multipriority MMPP Streams”, in *Proc of IEEE ICC’93* , pg. 1037-1041, 1993.
[11] G. Mayor and J.Silvester, “Time Scale Anaysis of an ATM Queueing System With Long-Range Dependent Traffic”, to appear in *Proc. of INFOCOM’97*, 1997.
[12] R. G. Addie, M. Zukerman and T. Neame, “Fractal Traffic: Measurements, Modelling and Performance Evaluation", In *Proc. of IEEE INFOCOM* 1995.
[13] B. Ryu and A. Elwalid,”The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities", *In Proc. of ACM SIGCOMM*, 1996.
[14] N. L. S. Fonseca e J. A. Silvester, “A Comparison of Push-out Policies in an ATM Multiplexer”, In *Proc. of IEEE Pacific Rim Conference on Comp. Comm and Signal Proc.*, pg. 338-341, 1993

Acknowledgments

This work was partially sponsored by CNPq, CNPq Protem Almadem and CAPES. We would like to thank CENAPAD-SP for the computational support.

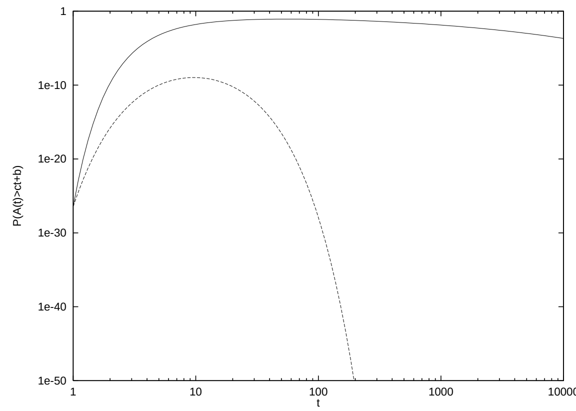

Figure 1: The Buffer Inefficacy Phenomenon

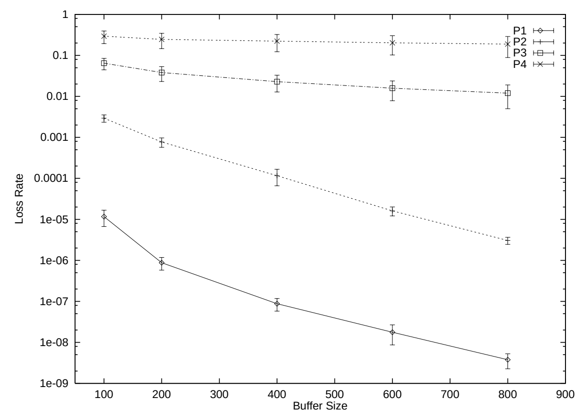

Figure 2: Per class Loss Rate *x* Buffer Size for ρ = 0.8 σ =1.0 H=0.85, P_1=0.4, P_2=0.1, P_3=0.2 and P_4 = 0.3,

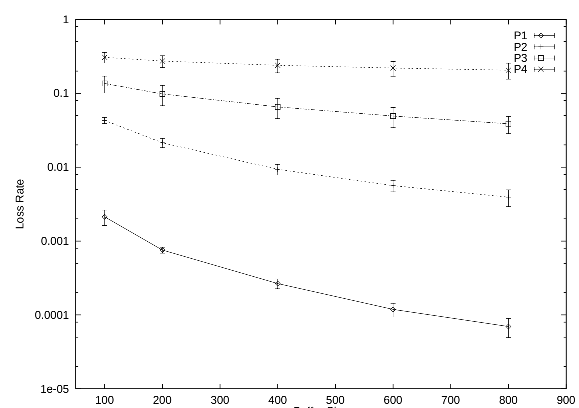

Figure 3: Per class Loss Rate *x* Buffer Size for ρ = 0.8 σ =1.0 H=0.85, P_1=0.55, P_2=0.05, P_3=0.2 and P_4 = 0.25

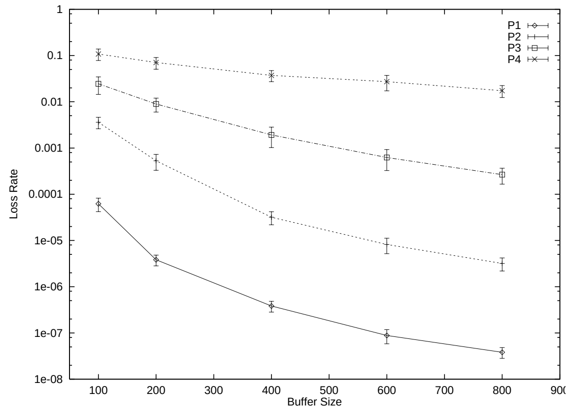

Figure 4a: H = 0.75

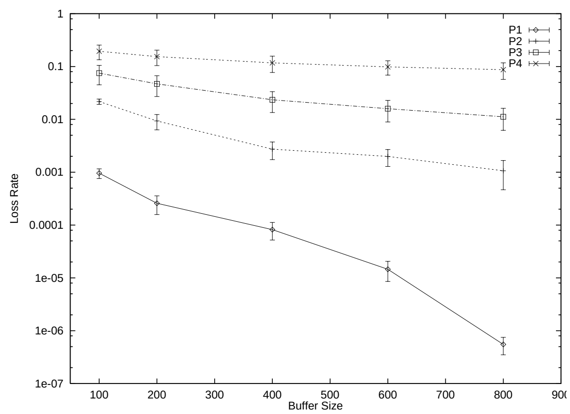

Figure 4b: H = 0.8

Figure 4: Per class Loss Rate *x* Buffer Size for ρ = 0.8 σ =1.0, P_1=0.6, P_2=0.1, P_3=0.1 and P_4 = 0.2

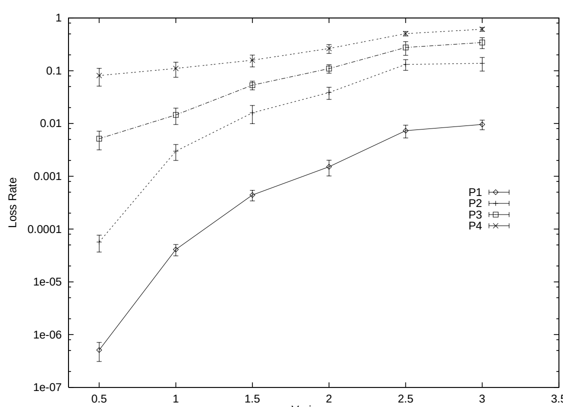

Figure 5: Loss Rate *x* Variance for ρ = 0.8, H = 0.75, Buffer Size = 400, P_1 = 0.7, P_2 = 0.05, P_3 = 0.1 and P_4 = 0.15

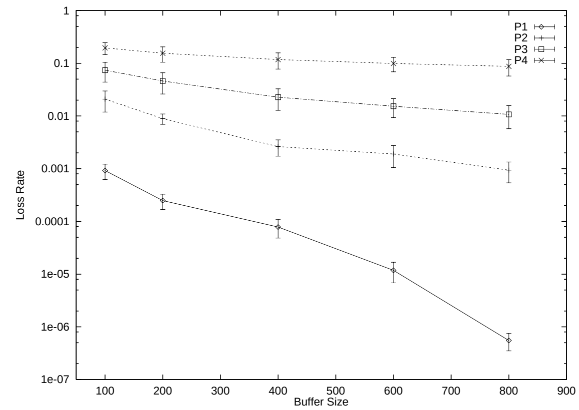

Figure 6a: M-FIFD

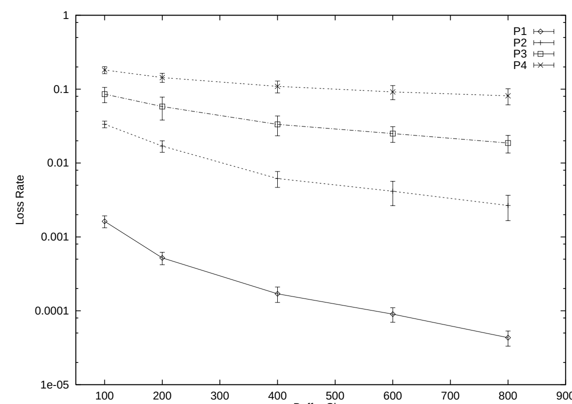

Figure 6b: LIFD

Figure 6: Loss Rate *x* Buffer Size for Different Push-out Policies, ρ=0.8, σ=1, P_1=0.6, P_2=0.05,P_3=0.15, P_4=0.2

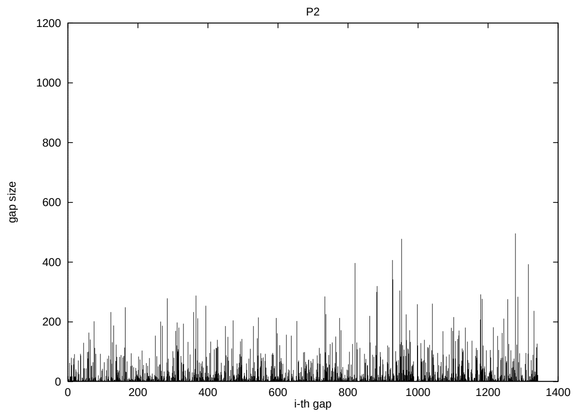

Figure 7.a: P_2

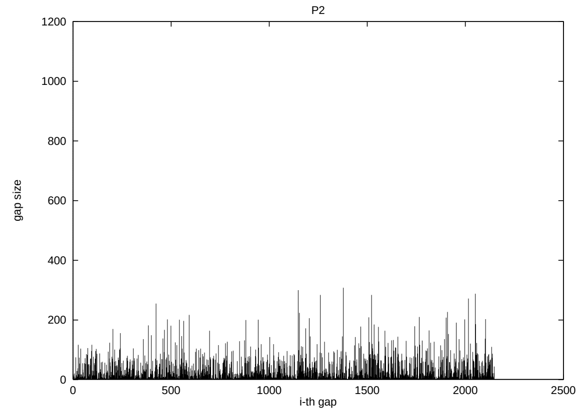

Figure 8.a: P_2

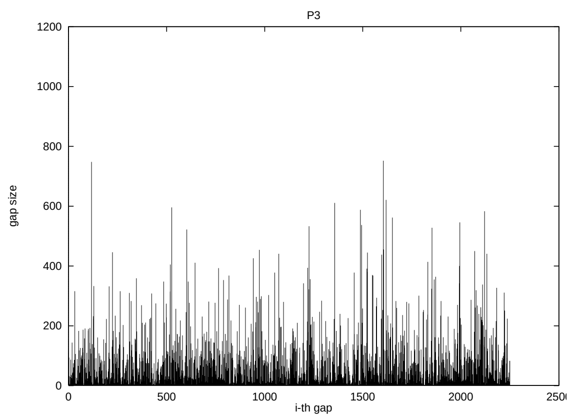

Figure 7.b: P_3

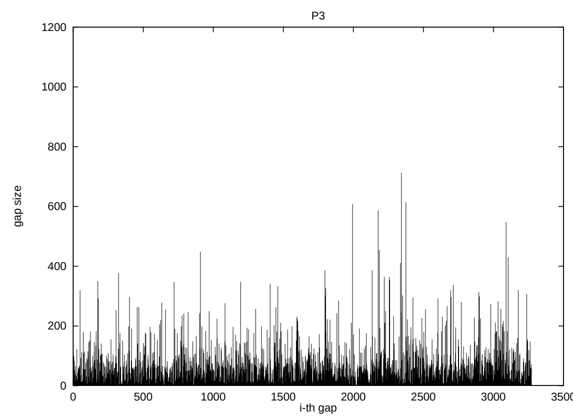

Figure 8.b: P_3

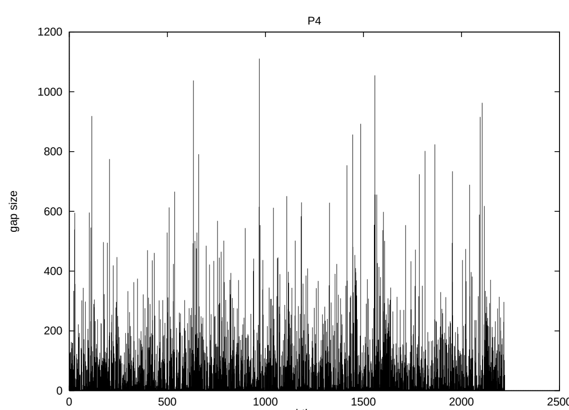

Figure 7.c: P_4

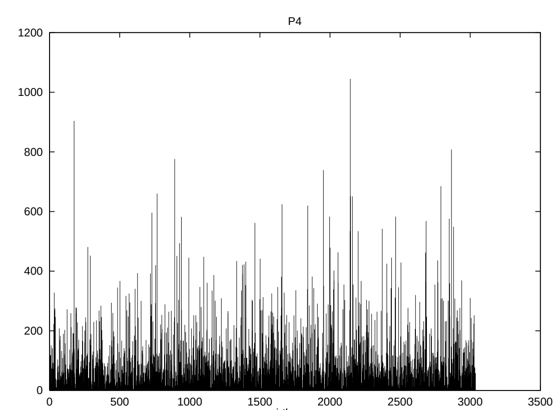

Figure 8.c: P_4

Figure 7: Loss Gap Size *x* i^{th} Gap for M-FIFD, priority classes 2, 3 and 4, $\rho = 0.8$, $\sigma = 1$, Buffer Size = 400, $P_1 = 0.7$, $P_2 = 0.05$, $P_3 = 0.1$ and $P_4 = 0.15$

Figure 8: Loss Gap Size *x* i^{th} Gap for LIFD, priority classes 2, 3 and 4, $\rho = 0.8$, $\sigma = 1$, Buffer Size = 400, $P_1 = 0.7$, $P_2 = 0.05$, $P_3 = 0.1$ and $P_4 = 0.15$