On the Effectiveness of Loss-Conserving Disciplines Under a Long-Range Dependent Process

N.L.S Fonseca, M.J. Ferreira e G. S. Mayor

Relatório Técnico IC 97-15

Outubro de 1997

On the Effectiveness of Loss-conserving Disciplines under a Long-range Dependent Process

N.L.S. Fonseca[1], M. J. Ferreira[1] and G. S. Mayor[2]

1 - State University of Campinas
Institute of Computing
P.O.Box 6176
13083-970 Campinas SP
Brazil
e-mail: nfonseca@dcc.unicamp.br

and

2 - University of Southern California
Electrical Engineering-Systems Department
Los Angeles, CA90089-2561
U.S.A.

Abstract

In this paper, we investigate the effectiveness of selective discard in a multiplexer subject to a long-range dependent process. We consider loss conserving disciplines, and we evaluate the impact of the input process traffic characteristics such as the Hurst parameter and variance into the per class loss rate and the loss gap. We found out that the Complete Sharing discipline is clearly worth adopting whereas Complete Sharing with Guaranteed Queue Minimum may be not. Furthermore, we show that the choice of push out policy may impact significantly the perceived QoS.

I) Introduction

Several studies [1]-[4] have claimed that different types of network traffic, *e.g*. local area network traffic (LAN), can be accurately modeled by a self-similar process. Even though the self-similar nature of network traffic has not been fully understood, a self-similar process is able to capture the long-range dependence (LRD) phenomenon exhibited by this traffic. Moreover, a series of simulation and analytical studies [5]-[10] 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 [8] and Duffield [11] 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 [7], [12], *i.e.*, by just increasing the buffer size we are not able to decrease the buffer overflow probability considerably.

Different multimedia applications have diverse loss requirements. For instance, a telephone conversation my 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 [13], [14]. In a selective discard mechanism cells are discarded in an overflow situation according to their priority level. Selective discard has been studied in the past few years. However, most of these studies considered short-range dependent processes.

In this paper, we investigate the effectiveness of selective discard mechanism in a multiplexer subject to a long-range dependent process. We consider loss conserving disciplines, and we evaluate the impact of the input process traffic characteristics such as the Hurst parameter and variance into the per class loss rate and the loss gap length [15]. We found out that the Complete Sharing discipline is clearly worth adopting, whereas Complete Sharing with Guaranteed Queue Minimum [16] may be not. Furthermore, we show that the choice of push out policy may significantly impact the perceived QoS. Our study is a first step in understanding end-to-end loss phenomenon in networks driven by realistic LRD processes [17], [18].

This paper is organized as follows: Section II introduces the Fractional Brownian Motion process. Section III explains the buffer inefficacy phenomenon. Section IV briefly introduces selective discard mechanisms. Section V describes the traffic generator used in this paper. In Section VI the main findings are discussed. Finally, conclusions are drawn in section VII.

II) Fractional Brownian Motion

The ordinary Brownian motion, *B(t)*, describes the movement of a particle in a liquid subjected to collisions and other forces [19]. It is a real random function with independent Gaussian increments such that

$$E[B(t+s) - B(t)] = 0$$

$$Var[B(t+s) - B(t)] = \sigma|s|$$

Mandelbrot defines fractional Brownian motion (fBm) as being the moving average of *dB(t)* in which past increments of *B(t)* are weighted by the kernel $(t\text{-}s)^{h - 1/2}$ [20].

Definition:

Let *H* be such that $0 < H < 1$. The fBm is defined as the Weyl's fractional integral of *B(t)*.

$$B_H(t) = \frac{1}{\Gamma(H+1/2)}\int_{-\infty}^{0}\left((t-s)^{H-1/2} - (-s)^{H-1/2}\right)dB(s) + \int_{0}^{t}(t-s)^{H-1/2}dB(s)$$

This equation leads to the ordinary Brownian motion if $H = 1/2$. Its self-similar property is based on the fact that $B_H(\rho s)$ is identical in distribution to $\rho^h * B_H(s)$. The increments of the fBm, Y_j form a stationary sequence called fractional Gaussian noise (fGn):

$$Y_j = B_H(j+1) - B_H(j), j = \ldots, -1, 0, 1, \ldots$$

We should note that these increments are not independent unless you have pure Brownian motion, *i.e.,* $H = 1/2$. Moreover, Hurst law states that $Var[B_H(t+s) - B_H(t)] = \sigma s^{2H}$, *i.e.,* a fBm arrival model is also able to capture the inherent high *variability* exhibited by real network traffic.

III) 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 [7],[12], [21]. 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 dependence.

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 [8], 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:

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

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:

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

we obtain:

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

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).

We also find a minimizer $t' \in \operatorname{arginf}_{t>0} g(t)$ so that the overflow probability is maximized. This corresponds to a likely time scale on which overflow occurs in this system. Therefore, t' is given by $t' = \frac{bH}{(c-\bar{a})(1-H)}$

The same result was derived independently by Addie in [24] and Ryu in [25]. We note that since the probability of buffer overflow for the LRD source decreases slowly with time, the probability of buffer overflow at time t' is not an accurate estimate for the buffer overflow probability at steady state. In other words, if we estimate the buffer overflow probability in a queueing system driven by a LRD source by the overflow probability at time t' we might not capture the heavy tail exhibited by Equation 1 leading to inaccurate results.

IV) Selective Discard Mechanism

Due to the bursty nature of multimedia sources, cell loss always occurs in networks based on statistical multiplexing, such as ATM networks. The only way to avoid loss is to allocate bandwidth based on a source peak rate which, obviously, makes statistical multiplexing ineffective. The QoS requirements of an application are usually translated into two main performance metrics: the loss rate (the ratio between the lost cells and the total number of transmitted cells) and the length of the gap loss (the number of cells consecutively lost). Different multimedia applications have different loss requirements. Coping with different requirements is a challenging task. Selective discard is a mechanism aimed at enabling the network to deal with diverse loss requirements.

In a selective discard mechanism cells are discarded in overflow situations according to their priority level.

Although the loss rate is a meaningful and measurable parameter, it is an average value and it does not entirely describe the loss process. 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.

Selective discard is related only to the management of the buffer space and not to the transmission order. A selective discard mechanism is completely specified by a buffer organization policy and by a push-out policy. A buffer organization policy defines which buffer slot may be occupied by which cell, a push-out policy chooses a cell to be discarded among the cells with the lowest priority.

Our investigation considers two buffer policies: Complete Sharing (CS) with Push out and Complete Sharing with Guaranteed Queue Minimum (CSGQM). In a CS policy low priority cells are pushed out from the buffer in an overflow situation. Similarly to CS, in a CSGQM, low priority cells are discarded during buffer overflows, however, we guarantee a minimum queue length to the low priority cells in an overflow scenario. Both CS and CSGQM are 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. An example of non-work-conserving discipline is Partial Sharing in which low priority cells can occupy up to a certain buffer position. Loss-conserving queues are of special interest because they minimize the overall cell loss and consequently maximizes the throughput.

The priority level of a cell can be defined either statically or dynamically. If statically assigned, the priority level indicates the importance of a cell. Cells which carry important information in a traffic stream, such as MPEG I frames cells, should have high priority. Moreover, if all cells of an application have the same priority level, they indicate the loss requirements of that application. Dynamically assigned priorities are usually related to the discrepancy between the negotiated and the actual transmission pattern of a connection, *i.e.*, a policing mechanisms marks as low priority cells whose transmission time corresponds to violation periods (periods in which the transmission characteristics of a connection do not conform to the values negotiated at the connection admission time) [26].

In order to capture the priority pattern of dynamically assigned priorities, we consider that the probability of a cell belonging to a certain priority class (priority probability) depends on the priority of the previous cell in the flow. After a period with no-arrivals, the priority probability of the first arriving cell is independent of any other cell in the flow. Thus, the “memory” of the priority classification exists between two no-arrival periods. Let us define:

P (high / no) - is the probability that the first cell after a period with no arrivals be high priority;

P (high / high) - is the probability of a cell being high priority given that the previous one was also high priority;

P (low / low) - is the probability of a cell being low priority given that the previous one was also low priority;

We define a correlation measure γ as *P (high / high)* + *P (low / low)* - 1. If *P (high / no)* = *P (high / high)* = 1 - *P (low / low)* (and consequently $\gamma = 0$) we have independent priorities. A positive value of γ indicates that cells of at least one of the classes tend to be agglutinated in bursts, meanwhile a negative value of γ shows that for at least one class there is no tendency of burst formation. The maximum value of γ (=1) happens when *P (high / high)* = *P (low / low)* = 1 and corresponds to the situation in which we have whole bursts of just one priority level.

A push-out policy selects a cell to be dropped among the low priority cell. The most common policies are Last-In-First-Drop (LIFD), First-In-First-Drop (FIFD), Random selection (RAND) and Modified-FIFD (M-FIFD). The modified-FIFD [27] 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 different queue distribution over a certain period of time and consequently may differently affect the perceived QoS. Let us give an example (Figure 2). Let us suppose that there are two low priority cells in the queue, one at the tail and the other at the head of the queue. An incoming high priority cell finds the buffer full. According to the LIFD policy, the low priority cell at the tail of the queue will be dropped, and according to the FIFD policy the one at the head of the queue will be the one dropped. Sometime later, after several high priority arrivals, the buffer is full again and a high priority arrival occurs. In the LIFD, the cell at the head of the queue at the previous high priority arrival (which was not dropped) has already been transmitted, and consequently the high priority cell is lost. For FIFD, however, the low priority cell at the tail of the queue at the previous high priority arrival was not dropped and can now be dropped to make space for the new high priority cell. M-FIFD brings the difference between FIFD and LIFD to its maximum because M-FIFD tries to concentrate the low priority cells at the tail of the queue and consequently maximize the likelihood that a low priority cell be dropped. Consequently, M-FIFD minimizes the high priority loss rate [27].

An in-depth view of selective discard mechanism can be found in [13].

V) A Fractional Gaussian Noise Generator

There are several algorithms for generating a fBm trace [28],[29]. More recently, new methods were developed. Huang [9] proposes a simulation method based on importance sampling, Pruthi [31] uses nonlinear chaotic maps and Lau [30] uses a random midpoint displacement algorithm. In this work we decide to use an algorithm proposed originally by Mandelbrot and improved latter by Chi [32]. It generates a discrete-time fGn sample. Our decision is based on the following factors: i) this algorithm is based on a sum of Markov processes, therefore our future work will investigate possible numerical solutions for this system, ii) the fBm samples are composed of a sum of a low and a high-frequency processes, hence it is suitable to investigate the contributions of each frequency component to the system performance, iii) it is a relative fast algorithm, it takes about 4 minutes to generate a million sample points on a Pentium 90 machine.

Discrete fractional Gaussian noise processes normalized to have zero mean and unit variance are Gaussian random processes having the autocovariance function [18]:

$$C(s;H) = E\left[B_H(t+s+1) - B_H(t+s)\right]\left[B_H(t+1) - B_H(t)\right] =$$

$$2^{-1}\left[|s+12|^{2H} - 2|s|^{2H} + |s-1|^{2H}\right]$$

The method proposed by Chi [16, 17] approximates the fGn by a sum of a low-frequency and a high-frequency Markov process. The resulting aggregate process has an autocovariance function very similar to [1]. We should point out that the autocorrelation of a single Markov process decays exponentially while the fGn autocorrelation decays linearly. Therefore, we need a sum of independent Markov processes in order to obtain this same correlation structure. An arbitrary value of autocorrelation, approximately *1/3*, is chosen as a threshold value to separate the low and high frequency components. Consider:

$$X_f(t) = X_L(t) + X_H(t)$$

where X_L and X_H denote low and high frequency terms respectively. X_f is the aggregate process. For large lags (s) the covariance of $X_L(t)$, $C_L(s;H)$, is approximately:

$$C_L(s;H) = H(2H-1)s^{2H-2}$$

Instead of approximating the low-frequency term by a single Markov process, a weighted sum of N standardized Markov-Gauss processes is used.

$$X_L(t) = \sum_{n=0}^{N} \left(W_n\right)^{1/2} M^{(n)}(t)$$

The Markov-Gauss, $M^{(n)}(t)$, are assumed to be pairwise uncorrelated. The objective is to calculate W_n so that the covariance of $X_L(t)$ matches $C_L(s;H)$. Therefore, after some calculation we get:

$$W_n = \frac{H(2H-1)}{\Gamma(3-2H)}\left(B^{1-H} - B^{H-1}\right)B^{2(H-1)n}$$

for $0 \le n \le N$, where $\Gamma(3-2H)$ is the gamma function.

The deficiency due to the approximation is $D(s;H) = C(s;H) - C_L(s;H)$.

We use another Markov-Gauss process to make-up the high-frequency deficiency [7]. This process should have variance $D(0;H)$ and lag one covariance $D(1;H)$ so that the covariance of the sum of the low frequency and high frequency process will have a self-similar structure [1]. The approximation of the fGn is given by:

$$X_f(t) = X_L(t) + X_H(t)$$
$$C_f(s) = C_L(s) + C_H(s)$$

VI) Numerical Examples

To evaluate the efficacy of selective discard mechanisms under LRD processes, we analyze a queue fed by a fbm which is generated by the procedure described in section V. We consider queues with Complete Sharing and with Complete Sharing with Guaranteed Queue Minimum.

We evaluate the impact of traffic parameters such as mean, variance and the Hurst parameter on the per class loss rate and loss gap length. Furthermore, we also investigated the influence of queue parameters such as buffer size, level of protection (level of guaranteed queue minimum) and push-out policy on the mentioned performances.

Figure 3 shows the loss rate as a function of the buffer size for different values of the Hurst parameter *H* considering a CS queue with FIFD push out policy (if not mentioned otherwise, FIFD is the push out policy for all examples in this section). The mean arrival rate (ρ) the variance (σ^2), P (high / high) and γ for this figure are 0.8, 1.0, 0.7, and 0.0, respectively (note that when $\gamma = 0.0$ we have *P (high / high) = P (high / no) = 1 - P (low / low)*). We notice that for high values of the Hurst parameter ($H > 0.8$) the high priority loss rate is almost insensitive to the buffer size. Therefore, introducing selective discard has a minor impact for this range of the Hurst parameter. For lower values of *H*, the high loss rate decreases as we increase the buffer size. For instance, we are able to reduce the loss rate in this specific example of an order of magnitude every 200 buffer slots. The low priority loss rate is in the order of the total loss rate. Figure 4 and 5 shows the high priority and the low priority loss rate for the CSGQM discipline. We notice that for a protection level of 10% (of the total buffer size) CSGQM gives almost the same results of the results given by CS. For a level of protection of 20%, we are still able to offer differentiated services. However, a level of protection of 30% high and low priority classes have almost the same loss rate, *i.e.*, we completely eliminate the advantage of selective discard. Furthermore, we point out that guaranteeing a minimum buffer space for the low priority class slightly decreases its loss rate. In other words, under LRD processes, CSGQM increases the high loss rate and does not decrease the low priority loss rate significantly.

In Figure 6 we evaluate the loss rate as a function of the variance for different values of the Hurst parameter considering a CS queue. As we increase the variance, we also increase the high priority loss rate irrespective of the Hurst parameter value since the buffer overflow probability given by Equation (1) increases with the variance. We point out that we are able to offer a selective service for high variance values only for low values of the Hurst parameter. Figure 7 and 8 shows the high and low priority loss rate for the CSGQM discipline, respectively. We note that for a protection level of 10% CSGQM gives similar results to CS. As we increase the protection level to 20% for higher values of the variance there is almost no difference between the high and the low priority loss rate specially for high values of the Hurst parameter. For low values of *H* the difference is not highly significant (two orders of magnitude). For a level of protection of 30% the two loss rates have almost the same value. Moreover, for the low priority class, the protection level has no influence on the loss rate for high variance values. This lack of sensitiveness to the level of protection may be due to the fact that most of the low priority loss may occur in long bursts.

In Figure 9 we analyze the relationship between the Hurst parameter and the priority correlation of a flow. In Figure 9.a we consider different values of γ by varying *P (high / high)*. We observe that the high priority loss rate for higher values of *P (high / high)* is always greater than for lower values of *P (high / high)* because for higher values of *P (high / high)* we have longer burst of high priority cells, and consequently we have longer high priority bursts in overflow situations. As we increase the Hurst parameter, we increase the overall loss rate and the difference between the loss rate given by a high value of *P (high / high)* and a low value of *P (high / high)* decreases. For high values of the Hurst parameter (H = 0.85) the difference in the high priority loss rate for *P (high / high)* values of 0.8 and 0.6 can be of two orders of magnitude, whereas for lower values of H, the difference is of three order of magnitude. In Figure 9.b we increase γ by increasing *P (low / low)*. Again, as we increase H, we increase the overall loss rate, and consequently the high priority loss rate. For higher values of *P(low/low)* we have lower values of the high priority loss rate since we increase the probability that a high priority cell finds long enqueued bursts of low priority cells in overflow situations. The difference on the high priority loss rate given by a *P (low / low)* of 0.8 and by a *P (low / low)* of 0.5 can be of four orders of magnitude. Similar behavior was observed for CSGQM except that as we increase the level of protection we increase the high priority loss rate.

To evaluate how the introduction of different push out policies may affect the per class performance in a network with LRD traffic we compare the Last-in-First-Drop, Random selection, First-in-First-Drop and Modified-First-in-First-Drop disciplines. As explained in section IV, distinct queue distribution produced by different push out policies may mainly influence the high priority loss rate and the low priority loss gap length. We initially discuss the relationship between the push out policies and the high priority loss rates.

In Figure 10.a we plot the high priority loss rate as a function of the buffer size. It is evident that the push out policy has a major impact on the high priority loss rate. The difference between the results given by LIFD and the results given by M-FIFD can be as high as three orders of magnitude for large buffer size and as low as one order of magnitude for small buffer sizes. We also observe that FIFD and M-FIFD give similar results. In Figure 10.b we show results for CSGQM with a protection level of 20%. It is possible to reduce the high loss rate by an order of magnitude when we use M-FIFD. This advantage disappear with higher degree of protection.

When investigating the dependence of the high priority loss rate on the variance, we verify that for low values of the variance it is possible to obtain a difference of an order of magnitude. However, it is almost non-existing for high variance value.

Figure 11 indicates that for short burst of high priority cells (*P (high / high)* = 0.6) it is possible to have a difference of two orders of magnitude while for long bursts (*P (high / high)* = 0.8), there is no significant difference in the result produced by these policies. This happens because with longer bursts of high priority cells we lose a higher percentage of high priority cells in overflow situations. Finally, we observe that the difference of the high priority loss rate produced by these policies is almost constant if we vary the offered load maintaining constant the other traffic parameters (Figure 12). In summary, we can state that the P_{high} *(LIFD)* > P_{high} *(RAND)* > P_{high} *(FIFD)* > P_{high} *(M-FIFD)* where P_{high} is the high priority loss rate, and that FIFID and M-FIFD produce quite similar results.

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 specially on the maximum values of the gap length. In the following 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 the replications.

Figure 13 shows the LIFD, FIFD and M-FIFD distributions for a buffer size of 100. We notice that the maximum gap length of FIFD and M-FIFD are twice the value of LIFD maximum value. The maximum value for RAND (not shown) is close to the LIFD maximum value. We notice that by increasing the buffer size the difference between the maximum value decreases. M-FIFD always gives the highest maximum value followed closely by FIFD and then by RAND and finally by LIFD. For instance, for a buffer size of 800 the maximum gap length for LIFD, RAND, FIFD and M-FIFD are 730, 750, 870 and 990, respectively. LIFD always gives the lowest maximum value because an eventual transmission of a cell close to the head of the queue breaks down a long loss burst. Conversely, M-FIFD always tries to concentrate the low priority cells at the tail of the queue increasing their chance of being dropped, and consequently gives the highest maximum value. Figure 14 shows the same scenario for CSGQM with 20% of buffer protection. We notice that guaranteeing a certain percentage of the buffer space for the low priority cells brings down the maximum gap length for all policies except for M-FIFD. As the buffer size increases, the maximum gap length also increases, however, to lower values when compared to CS. For instance, the maximum length with a buffer size of 800 for LIFD, RAND, FIFD and M-FIFD are 130, 150, 400 and 500, respectively. We observe that the low priority gap length is highly sensitive to the Hurst parameter. Figure 15a and 15.b show the M-FIFD distributions for $H = 0.75$ and $H = 0.8$, respectively. We found out that the FIFD (~ 1800) and the M-FIFD (1800) maximum values for $H = 0.8$ are almost three times the value for $H = 0.7$ (~ 600). For LIFD and RAND disciplines the maximum gap length value for $H = 0.8$ (~500) is almost twice the maximum value for $H = 0.75$ (~ 250).

The offered load obviously affects the low priority gap length. For a CS queue and an arrival rate of 0.75 the maximum gap length for LIFD, RAND, FIFD and M-FIFD are 400, 400, 450 and 500, respectively, whereas they are 500, 600, 800 and 850 for an offered load of 0.85. For a CSGQM queue with 20% of buffer protection and same traffic parameters these values are 50, 70, 140 and 250 for an offered load of 0.75 and 70, 90, 400 and 500 for an offered load of 0.85.

The gap length is also highly sensitive to the variance of the input process. For a CS queue and variance of 1.0 the maximum length of LIFD, RAND, FIFD and M-FIFD are 100, 120, 500 and 1000, respectively, whereas for a variance value of 2.5 they are 4500, 4700, 5300 and 9000. For CSGQM these values are 100, 120, 470 and 950 for a variance value of 1.0 and 150, 190, 1400 and 5000 for a variance of value 2.5. These values clearly indicates the benefits of buffer protection for the low priority class.

VI) Conclusions

In this paper, we investigated the effectiveness of selective discard mechanisms based on Complete Sharing with Push out and Complete Sharing with Guaranteed Queue Minimum in providing differentiated services under long-range dependent processes.

Results based on the assumption of having a FIFD push out policy indicated that the Hurst parameter has a major influence on the efficacy of selective discard mechanism based on the CS discipline. For high values of the Hurst parameter ($H > 0.8$) there is a slight decrease on the high priority loss rate even for large buffers (> 800). Conversely, for low values of the Hurst parameter, the high priority loss rate decreases as we increase the buffer size. For queues based on CSGQM, the high priority loss rate significantly increases with a level of protection of 20%. Additionally, no significant impact on the low priority loss rate was observed. These findings indicate that selective discard based on CS is worth adopting for low values of the Hurst parameter (< 0.8) while it does not make sense at all for the CSGQM discipline.

We also pointed out that the push out policy has a great impact on the results. The M-FIFD discipline minimizes the high priority loss rate and it can give a loss rate three order of magnitude lower than does LIFD. Results based on FIFD are very close to those given by M-FIFD which reinforce the conclusions drawn above. FIFD is less complex to implement since it requires less buffer shifting than M-FIFD does. Thus, FIFD is an attractive option for implementing in an ATM switch. Regarding the maximum gap length, we verified that M-FIFD can give a value which may be three times higher than the maximum loss gap given by LIFD. We also observe that the buffer protection mechanism (CSGQM) can reduce significantly the maximum gap length. Furthermore, the maximum loss gap is very sensitive to the Hurst parameter. Finally, we mention that findings for long-range dependent processes are quite different from results for short-range dependent processes [15].

VI) 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] M. Garrett and W. Willinger , “Analysis Modeling and Generation of Self-Similar VBR Video Traffic”. In *Proc. of ACM SIGCOMM*, 1994.

[3] J.Beran et al., “Long-Range Dependence in Variable-Bit-Rate Video Traffic”. *IEEE Transactions on Communications*, 1995.

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

[5] G. Mayor and J.Silvester, “A Trace-Driven Simulation of an ATM Queueing System with Real

Network Traffic”, *Proc. of IEEE ICCCN*, 1996.

[6] P. Pruthi and A. Erramilli, “Heavy-Tailed ON/OFF Source Behavior and Self-Similar Traffic”, in *Proc. of IEEE ICC’95*, pp. 445-450, 1995

[7] G. Mayor and J. Silvester, “An ATM Queueing System with a Fractional Brownian Noise Arrival Process”, *Proc. of IEEE ICC*, 1996.

[8] I. Norros, “A Storage Model with Self-Similar Input”, Queueing Systems 16, 1994.

[9] F.Brichet et al., “Heavy Traffic Analysis of a Fluid Queue Fed by On/Off Sources with Long-Range Dependence”, preprint 1995.

[10] N. Likhanov and B. Tsybakov, “Analysis of an ATM Buffer with Self-similar (“Fractal”) Input Traffic, in *Proc of SIGCOMM*, 1995.

[11] 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.

[12] 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.

[13] 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 .

[14] N. L. S. Fonseca e J. A. Silvester, “Estimating the Loss Probabilty in a Multiplexer Loaded with Multi-priority MMPP Streams”, in *Proc of IEEE ICC’93* , pg. 1037-1041, 1993.

[15] 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

[16] P. Landsberg and C. Zukowski. “A novel buffer sharing method: complete sharing subject to guaranteed queue minimums”. *In Proc. of the First International Conference on Computer Communications and Networks*, pages 43--48, 1992.

[17] N. L. S. Fonseca e J. A. Silvester, “An Approximate Model for the Output Process of an ATM Multiplexer with Selective Discard Mechanism”, in *Proc. of IEEE ICC’95*, pag. 783-787, 1995.

[18] N. L. S. Fonseca e J.A. Silvester, “Modelling the Output Process of an ATM Multiplexer with Correlated Priorities”, to appear in *Proc. of IEEE ICC’97*, 1997.

[19] A. Papoulis, “Probability, Random Variables and Stochastic Processes”, A. Papoulis, 3rd edition, McGraw-Hill, Inc., 1978.

[20] B. Mandelbrot and J.Ness, “Fractional Brownian Motions, Fractional Noises and Applications”, *SIAM Review*, October 1968.

[21] S.Q.Li and C.L. Hwang,`”Queue Response to Input Correlation Functions: discrete spectral analysis", *IEEE/ACM Transaction on Networking*, pp. 522-533, October 1993.

[22] G. Mayor and J. Silvester, “Tail Probabilities of an ATM Multiplexer with Long-Range Dependent Traffic”, submitted to *ACM SIGCOMM*.

[23] 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.

[24] R. G. Addie, M. Zukerman and T. Neame, “Fractal Traffic: Measurements, Modelling and Performance Evaluation", In *Proc. of IEEE INFOCOM* 1995.

[25] 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.

[26] J.A. Silvester, N.L.S. Fonseca, G. S. Mayor and S. P. S. Sobral, “The Effecitiveness of Multi-level Policing Mechenisms in ATM Traffic Control, In *Proc. of IEEE International Telecommunications Symposium 96*, pg 98-102, 1996.

[27] L. Tassiulas, Y.C. Hung, and S.S. Panwar. “Optimal Buffer Control during Congestion in a {ATM} Network Node. *IEEE/ACM Transactions on Networking*, 2(4):374--386, Aug 1994.

[28] J. R. Hosking, “Modeling Persistence in Hidrological Time Series Using Fractional Differencing”, *Water Resources Research*, 20(12):1898-1908, 1984.

[29] A.I. McLeod and K. W. Hipel, “Preservation of the Rescaled Adjusted Range: 1. A Reassessment of the Hurst Phenomenom”, *Water Resources Research*, 14(3):491-508, 1978.

[30] W. Lau et al., “Self-Similar Traffic Generation: The Random Midpoint Displaycement Algorithm and Its Properties”, in *Proc of IEEE ICC’95*, pp. 466-470, 1995.

[31] P. Pruthi and A. Erramilli, “Heavy-Tailed ON/OFF Source Behavior and Self-Similar Traffic”, In *Proc of IEEE ICC’95*, pp. 445-450, 1995.

[32] M. Chi, E. Neal and G. Young, “Practical Application of Fractional Brownian Motion and Noise to Synthetic Hydrology” , *Water Resources Research*, Vol. 9, 1523-1533, December 1973.

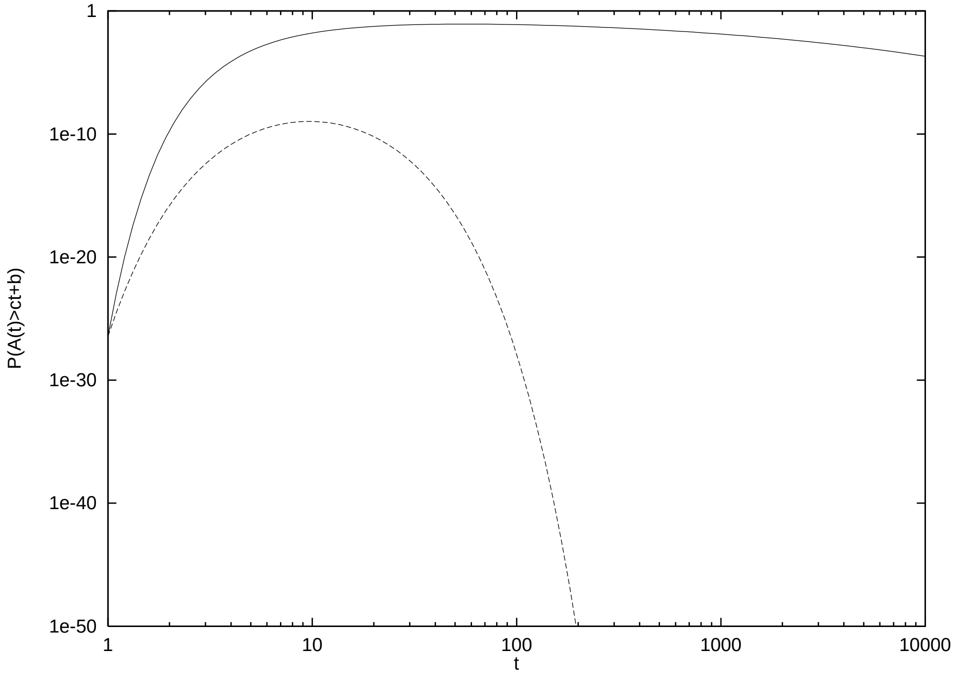


Figure 1: The Buffer Inefficacy Phenomenon.

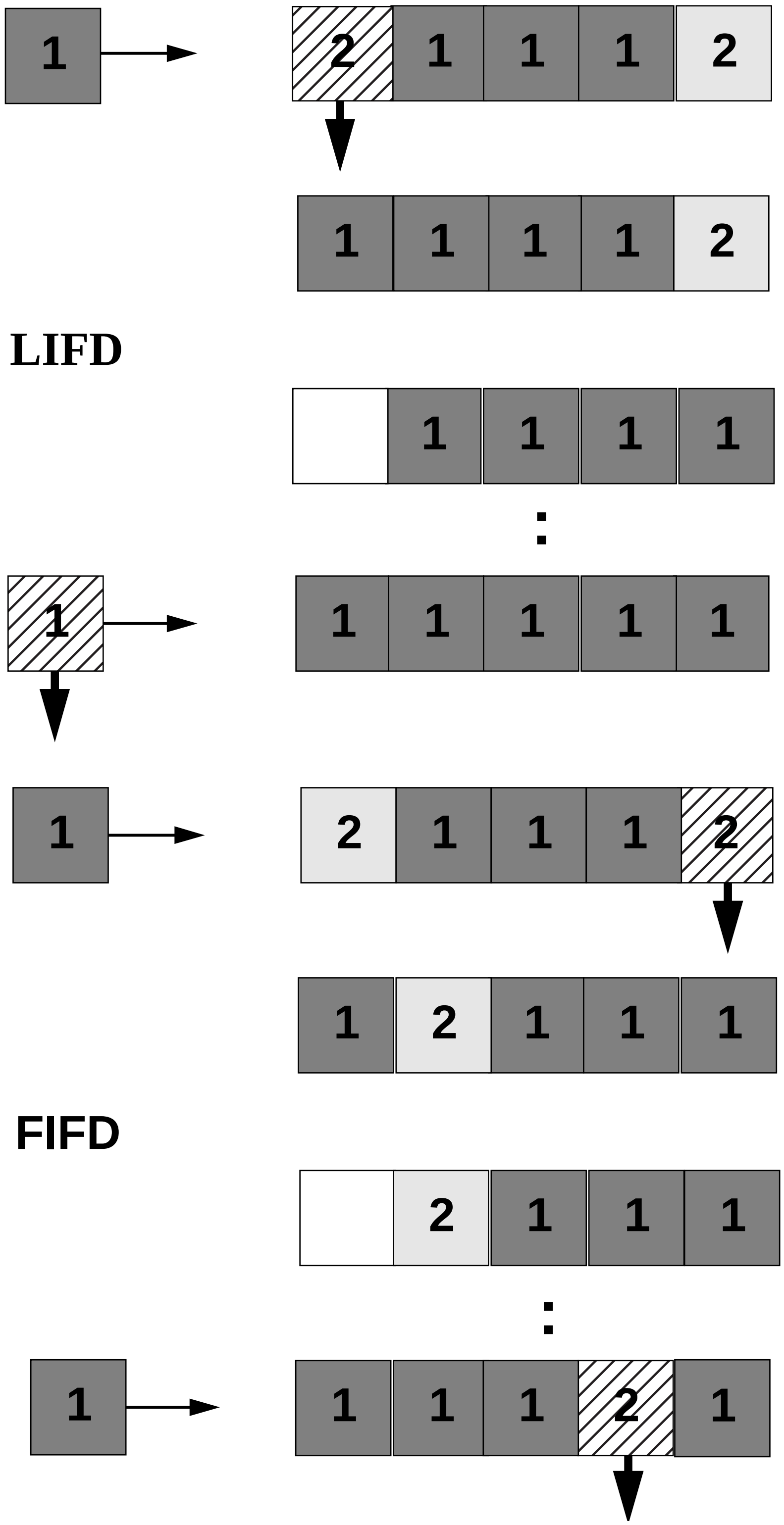
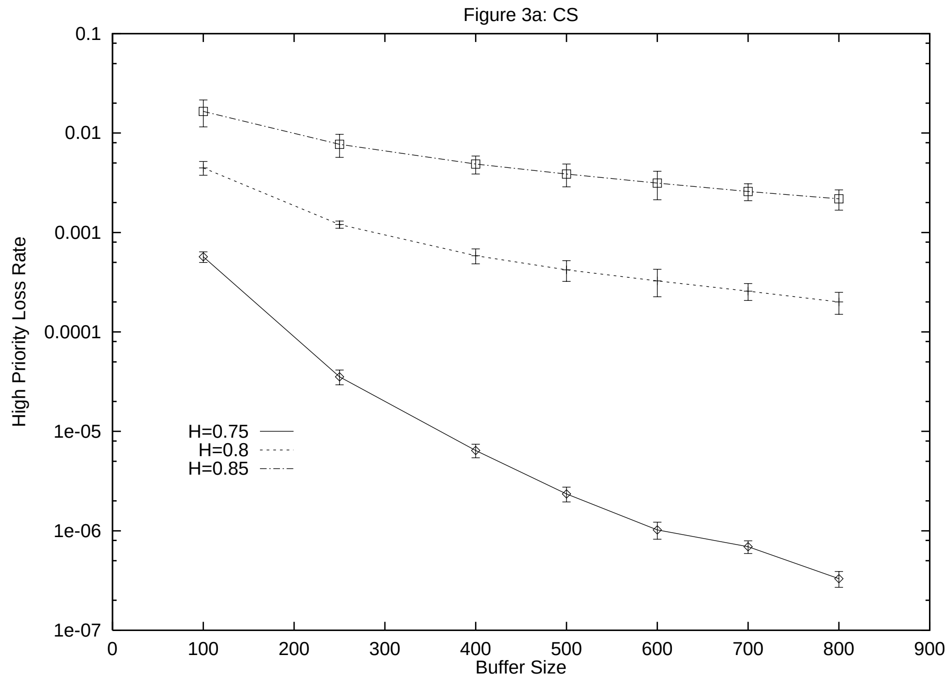


Figure 2: An Example of a Cell Loss due to Different Queue Length Distributions Produced by the LIFD and the FIFD Push out Policies.



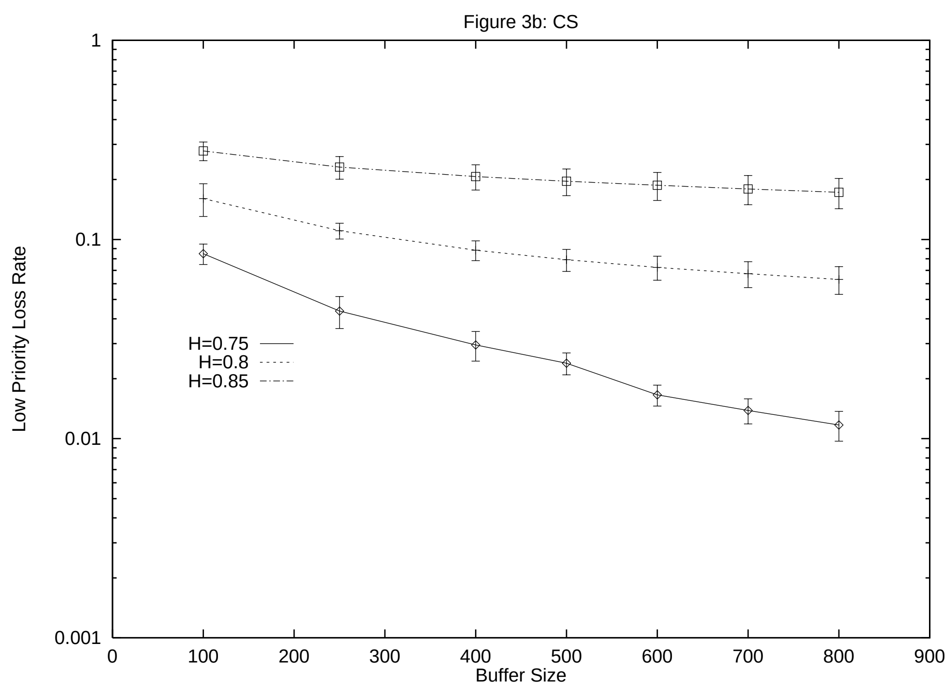


Figure 3: Loss Rate *x* Buffer Size for Different Values of the Hurst Parameter and a CS Queue, $\rho = 0.8$, $\sigma^2 = 1$, *P (high / high)* = 0.7, $\gamma = 0.0$.

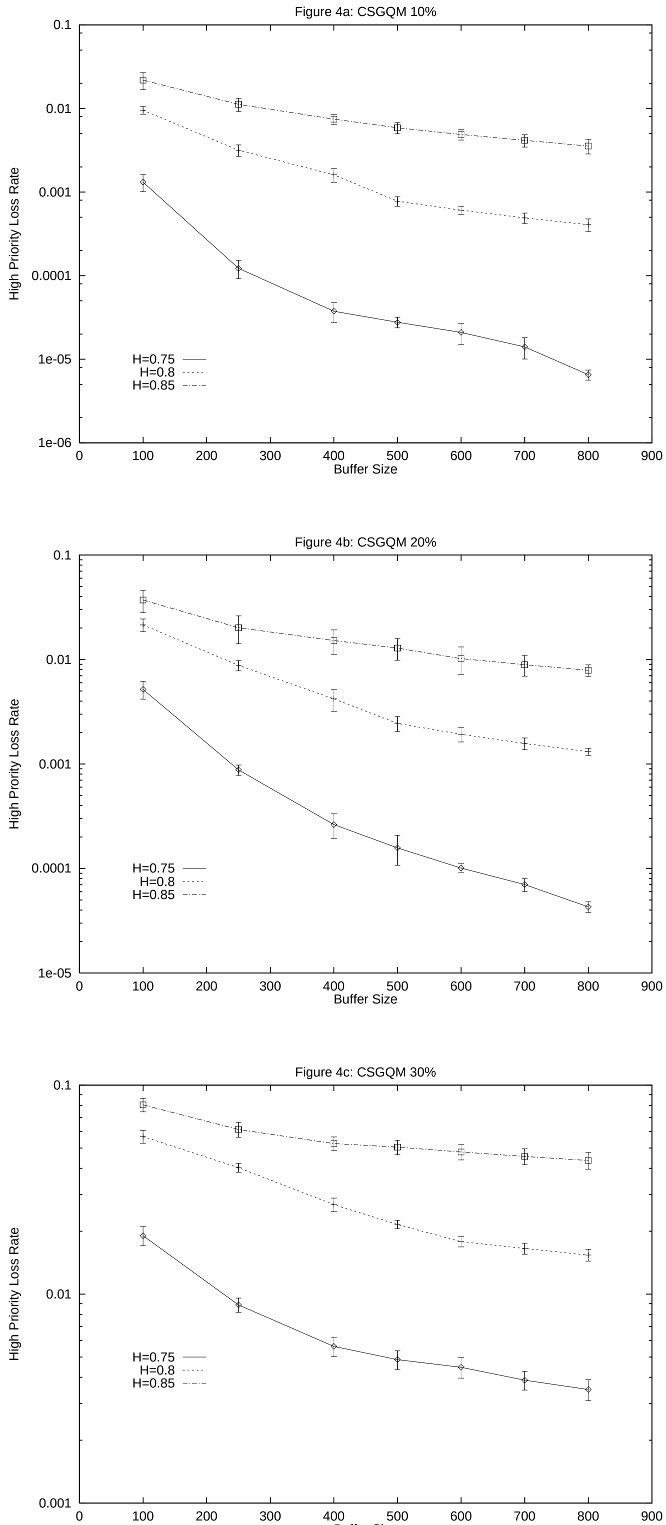


Figure 4:, High Priority Loss Rate *x* Buffer Size for Different values of the Hurst Parameter and a CSGQM Queue with Different Levels of Protection: 10% (Figure 4.a), 20% (Figure4.b) and 30% (Figure 4.c), $\rho = 0.8$, $\sigma^2 = 1$, *P (high / high)* = 0.7, $\gamma = 0.0$.

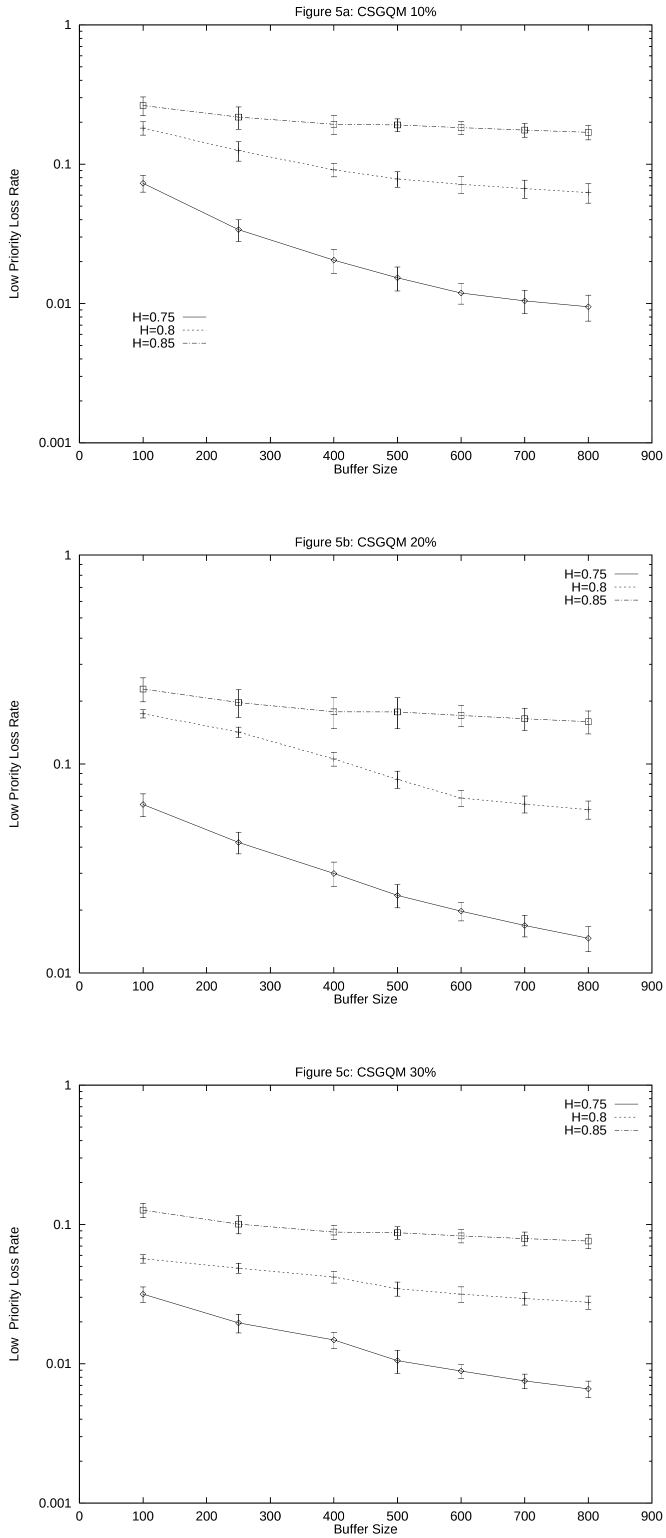
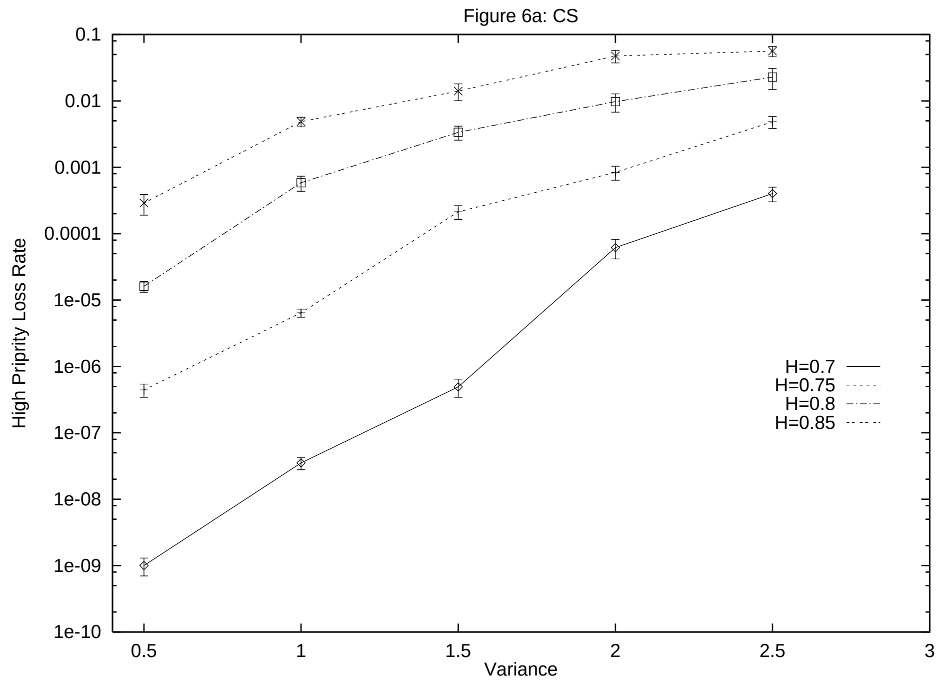


Figure 5:, Low Priority Loss Rate *x* Buffer Size for Different Values of the Hurst Parameter and a CSGQM queue with Different Levels of Protection: 10% (Figure 5.a), 20% (Figure 5.b) and 30% (Figure 5.c), $\rho = 0.8$, $\sigma^2 = 1$, *P (high / high)* = 0.7, $\gamma = 0.0$



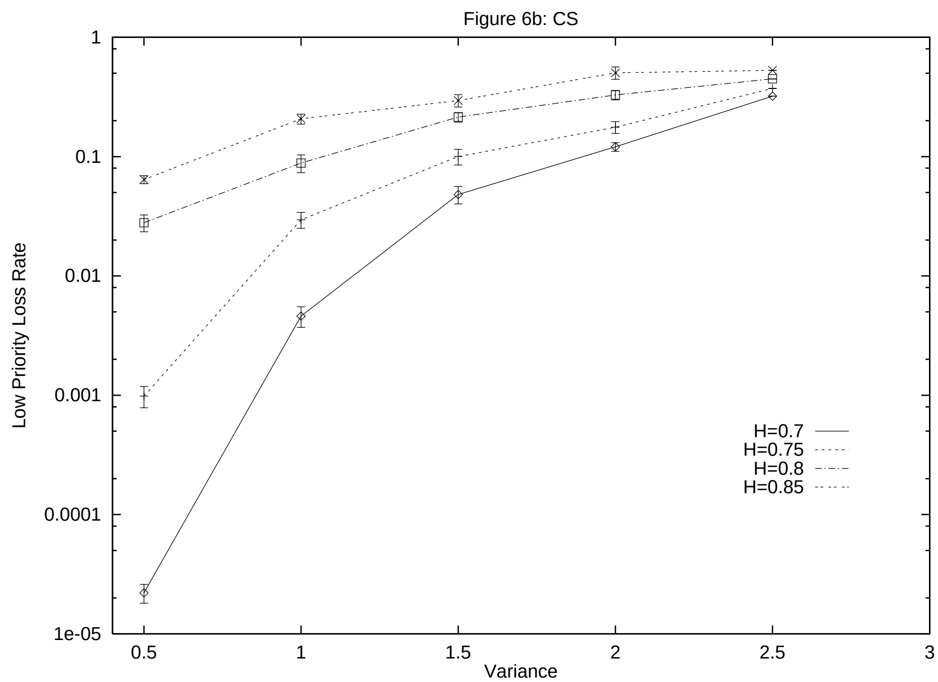


Figure 6: Loss Rate *x* Variance for Different Values of the Hurst Parameter and a CS Queue, Buffer Size = 400, $\rho = 0.8$, *P (high / high)* = 0.7, $\gamma = 0.0$.

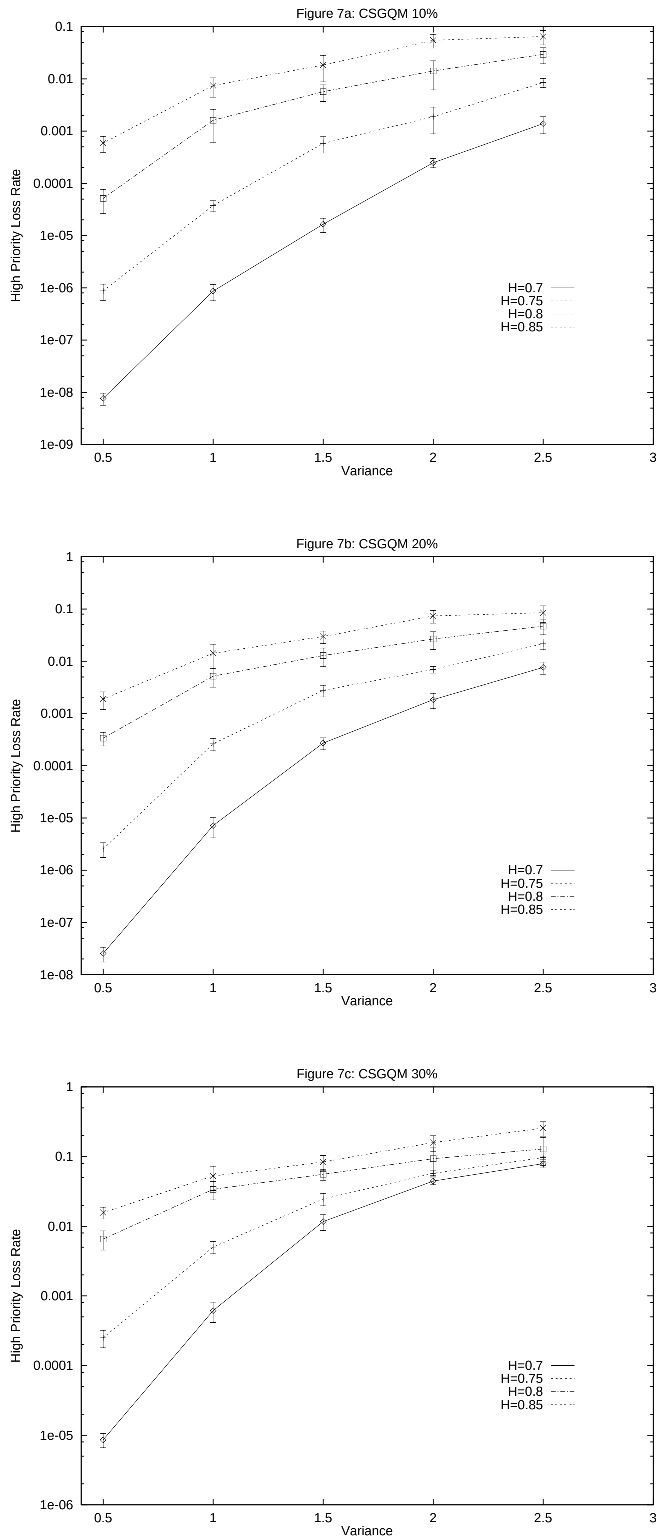
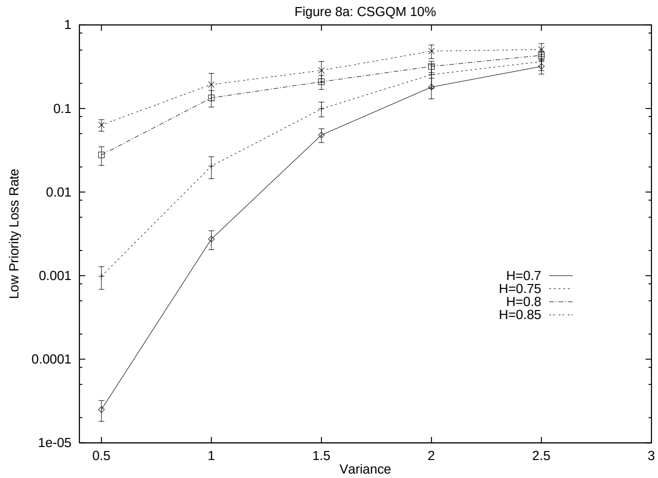
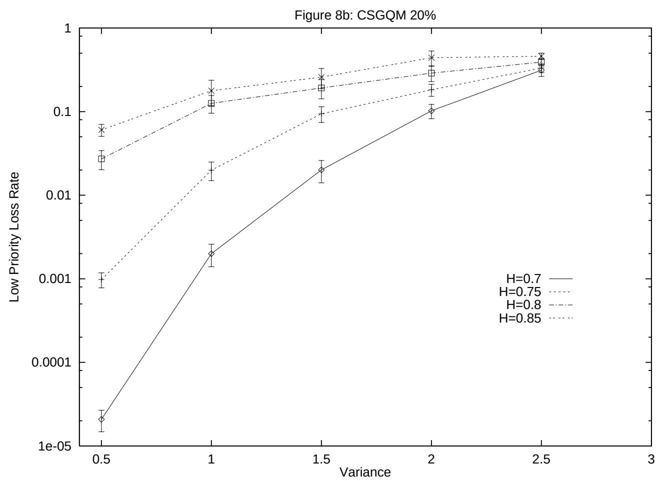


Figure 7: High Priority Loss Rate *x* Variance for Different Values of the Hurst Parameter and a CSGQM Queue with different Levels of Protection: 10% (Figure 7.a), 20% (Figure 7.b) and 30% (Figure 7.c) Buffer Size = 400, ρ = 0.8, *P (high / high)* = 0.7, γ = 0.0.





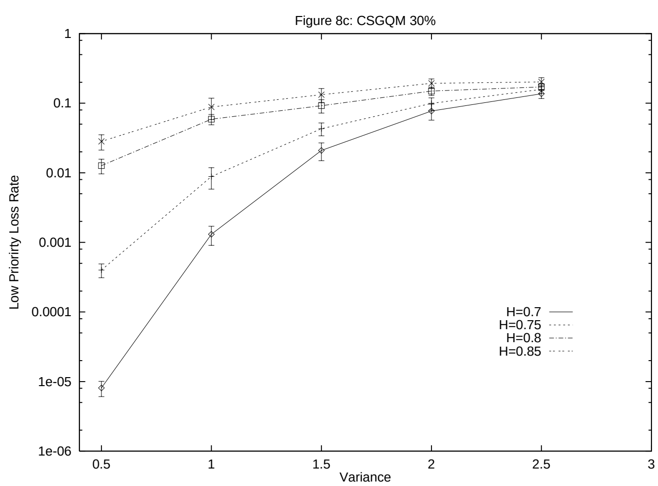
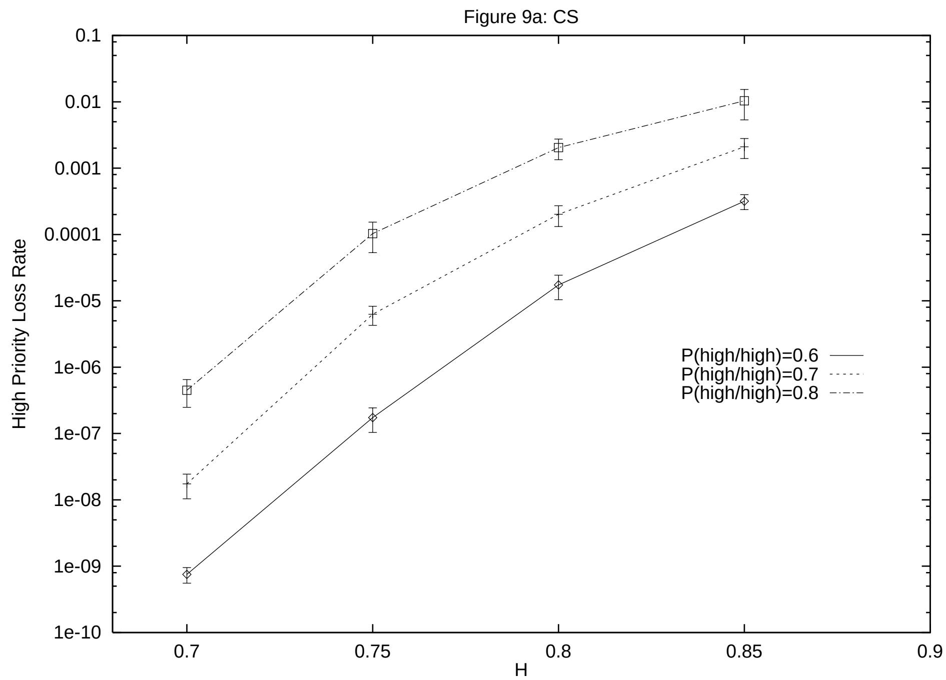


Figure 8: Low Priority Loss Rate *x* Variance for Different Values of the Hurst Parameter and a CSGQM Queue with Different Levels of Protection: 10% (Figure 8.a) , 20% (Figure 8.b) and 30% (Figure 8.c) Buffer Size = 400, $\rho = 0.8$, *P (high / high)* = 0.7, $\gamma = 0.0$.



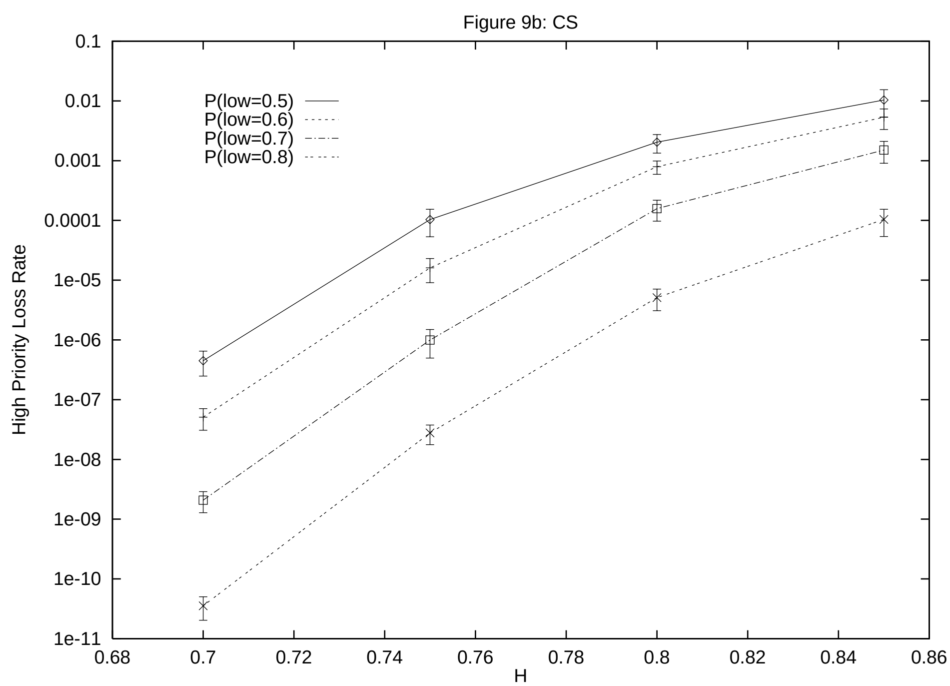
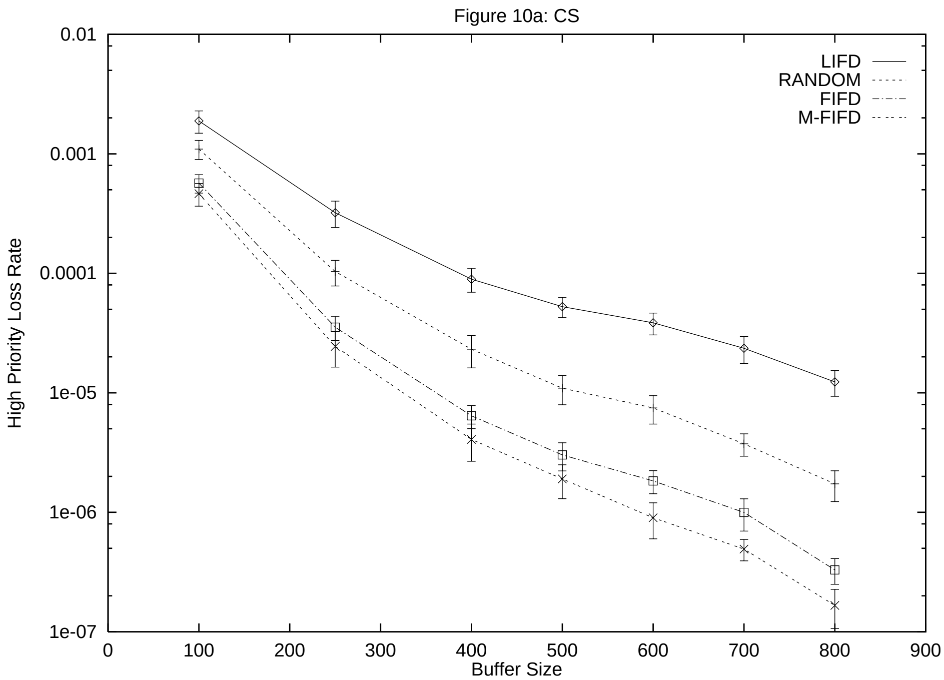


Figure 9: High Priority Loss Rate *x* Hurst Parameter for Different Values of *P(high / high)* and *P(low / low)* = 0.5 (Figure 9.a) and for different values of *P(low / low)* and *P (high / high)* = 0.8 (Figure 9.b) and a CS queue, Buffer Size = 250, ρ = 0.8, σ^2 = 1.



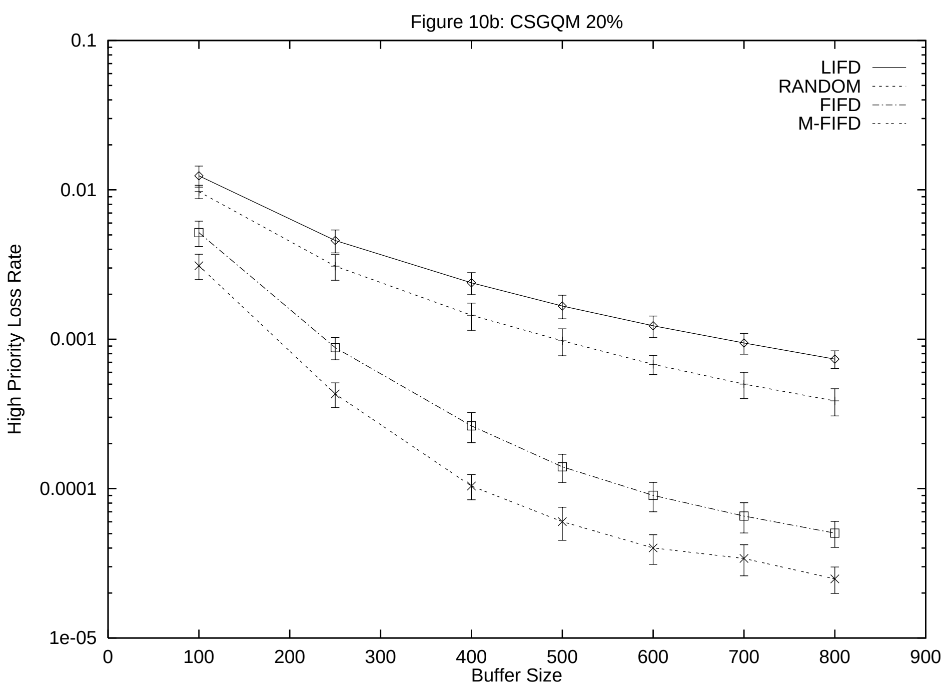
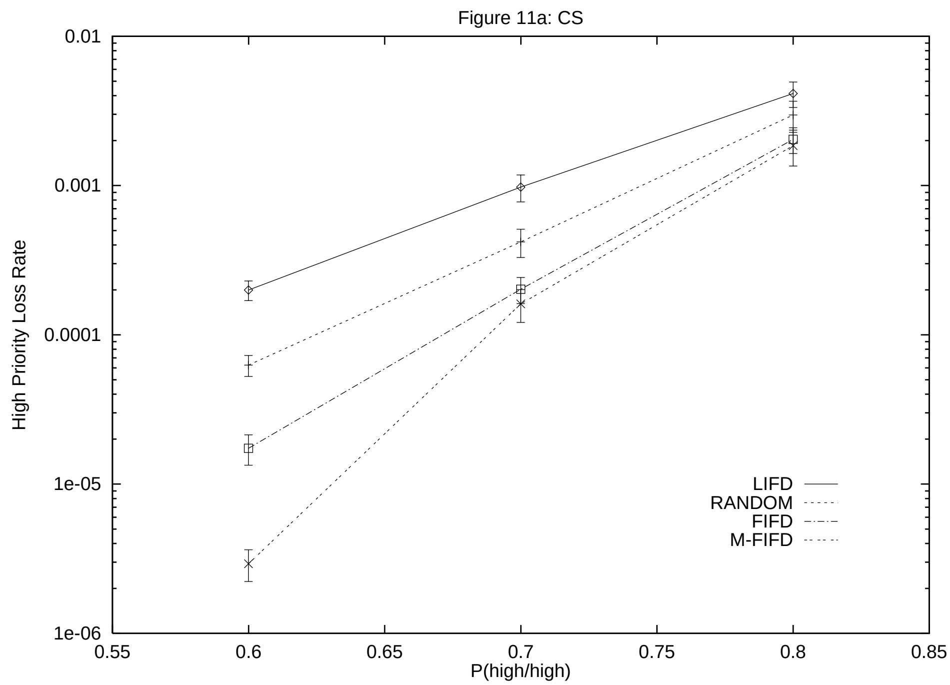


Figure 10: High Priority Loss Rate *x* Buffer Size for Different Push out Policies a CS Queue (Figure 10.a) and a CSGQM (Figure 10.b) Queue with 20% of Protection Level, $H = 0.75$, $\rho = 0.8$, $\sigma^2 = 1$, $P\ (high\ /\ high) = 0.7$, $\gamma = 0.0$.



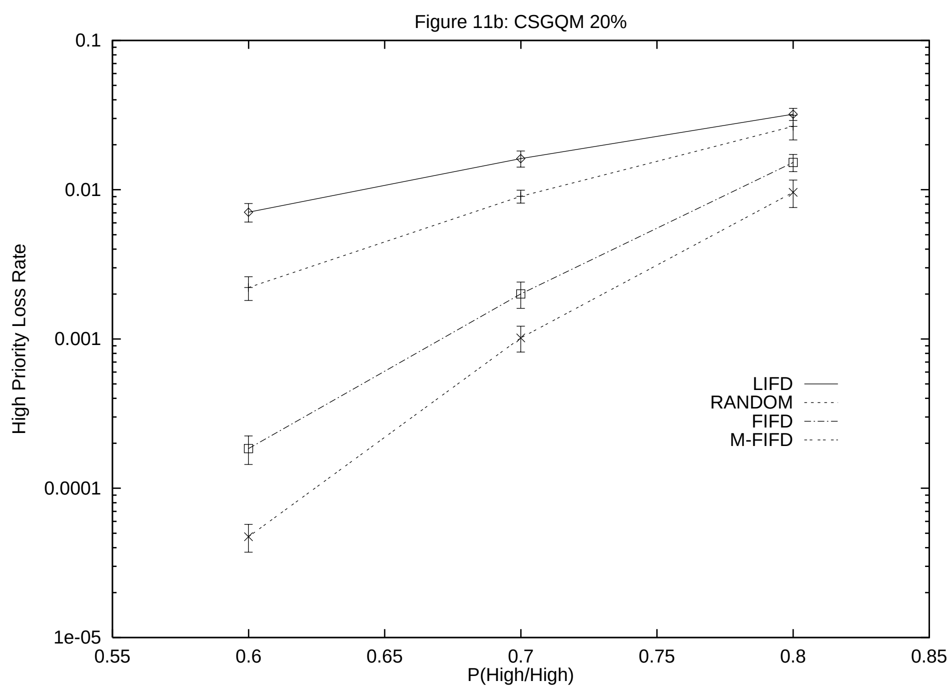
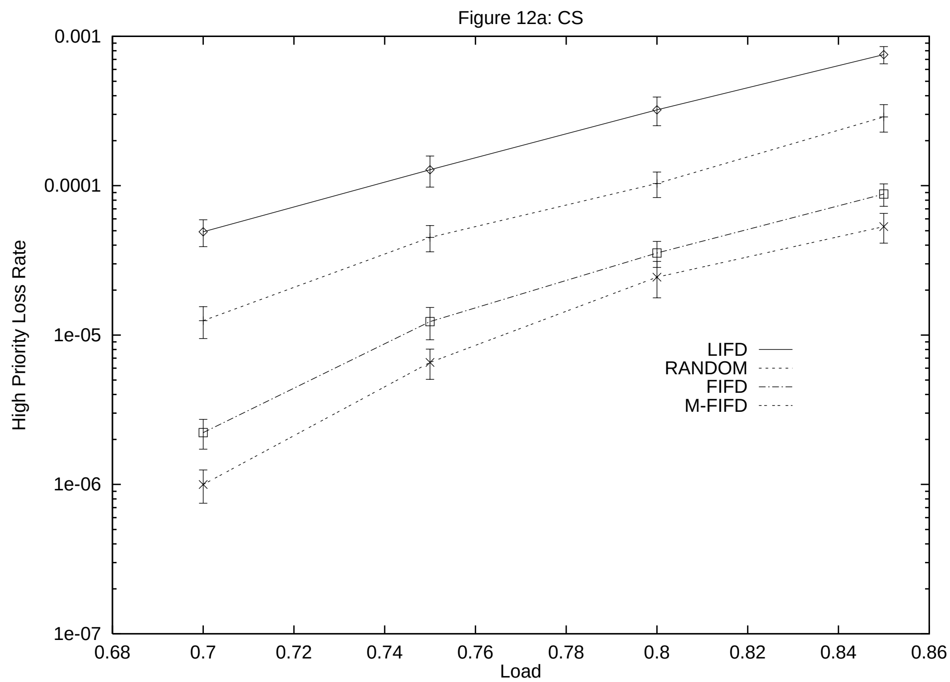


Figure 11: High Priority Loss Rate *x P (high / high)* for Different Push out Policies and a CS Queue (Figure 11.a) and a CSGQM Queue (Figure 11.b), Buffer Size = 400, $H = 0.8$, $\rho = 0.8$, $\sigma^2 = 1$, *P (low / low)* = 0.5.



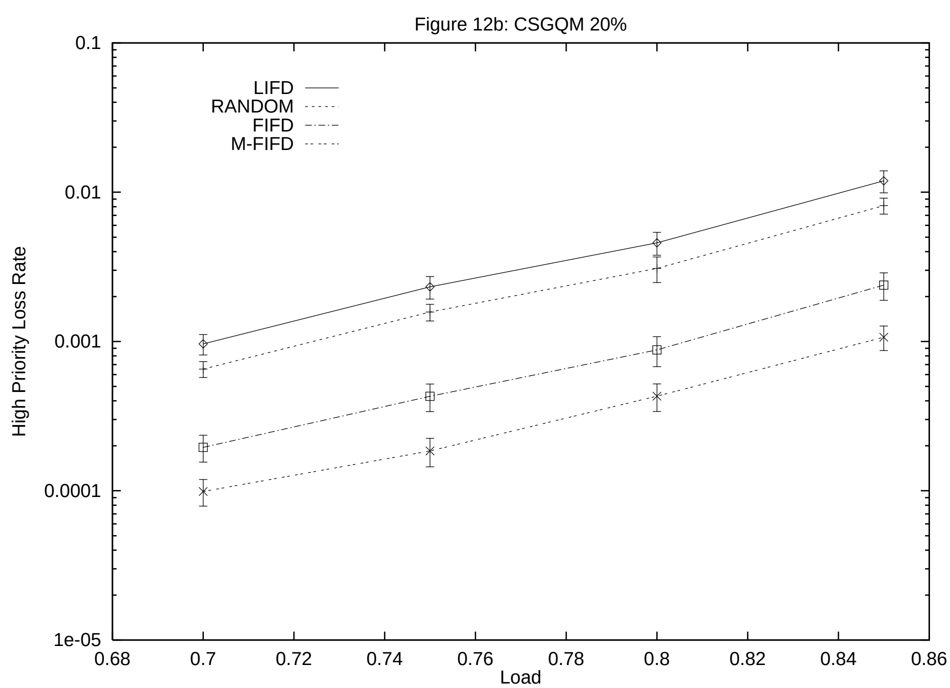


Figure 12: High Priority Loss Rate *x* Offered Load for Different Push out Policies and a CS Queue (Figure 12.a) and a CSGQM with Level of Protection of 20% (Figure 12.b), Buffer SIze = 250, $H = 0.75$, $\sigma^2 = 1$, $P(high / high) = 0.7$, $\gamma = 0.0$.

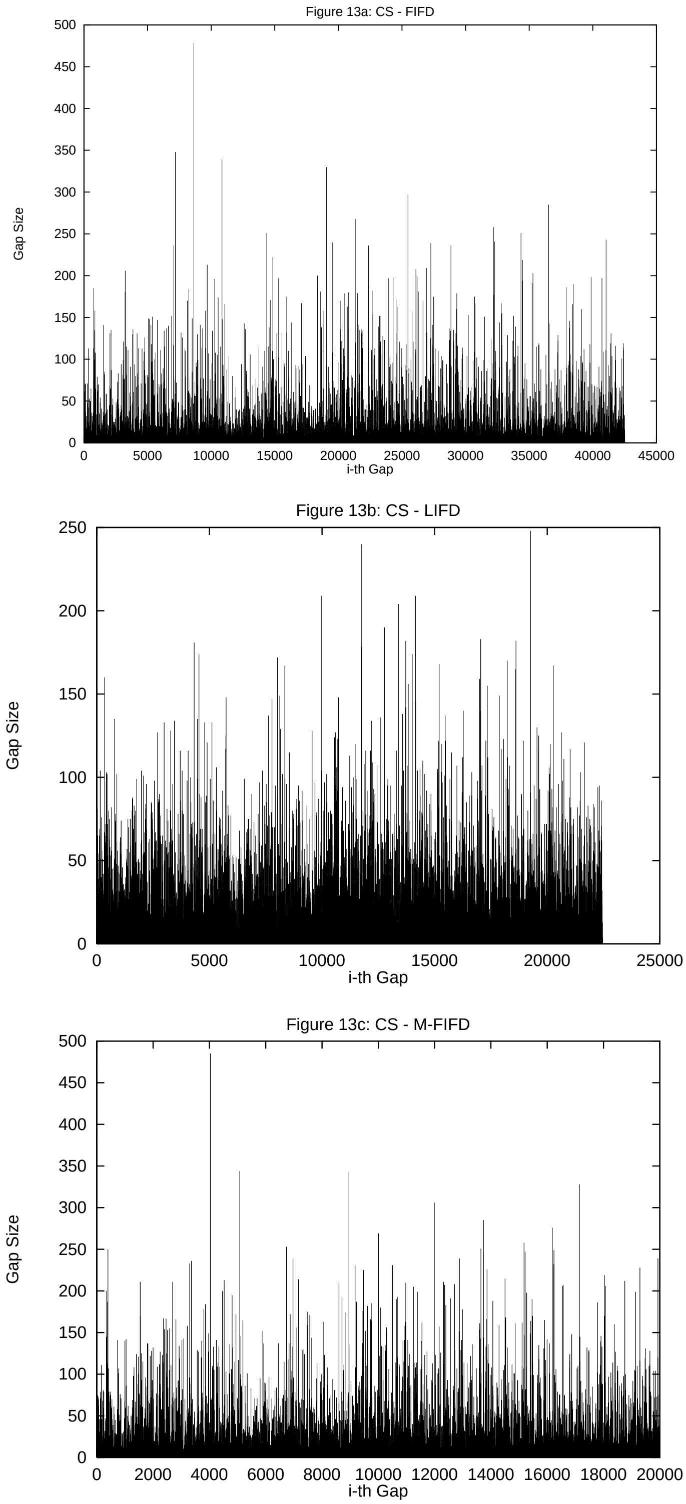
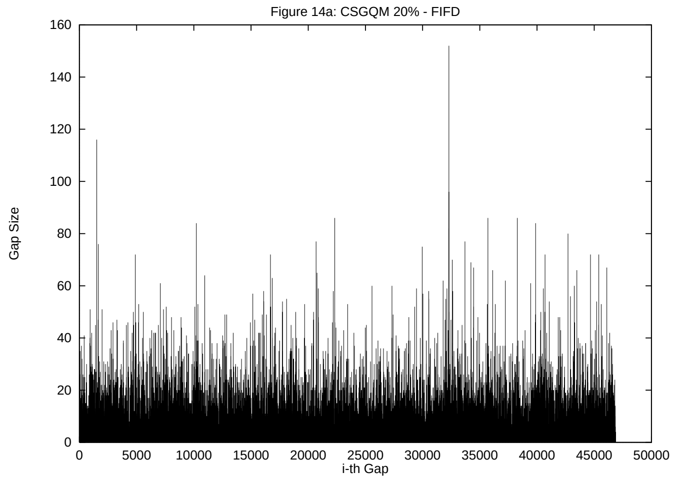
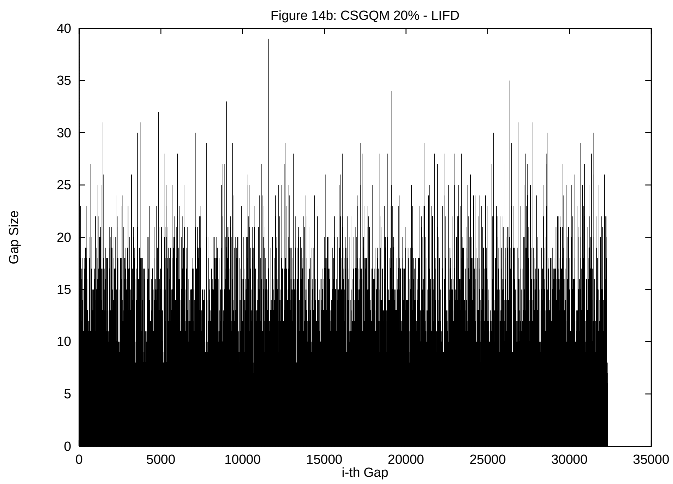


Figure 13: Low Priority Loss Gap Size x ith gap for Different Push ou Policies: FIFD (Figure 13.a), LIFD (Figure 13.b) and M-FIFD (Figure 13.c) for a CS Queue, Buffer Size = 100, H = 0.75 $\rho = 0.8$, $\sigma^2 = 1$, $P(high / high) = 0.7$, $\gamma = 0.0$





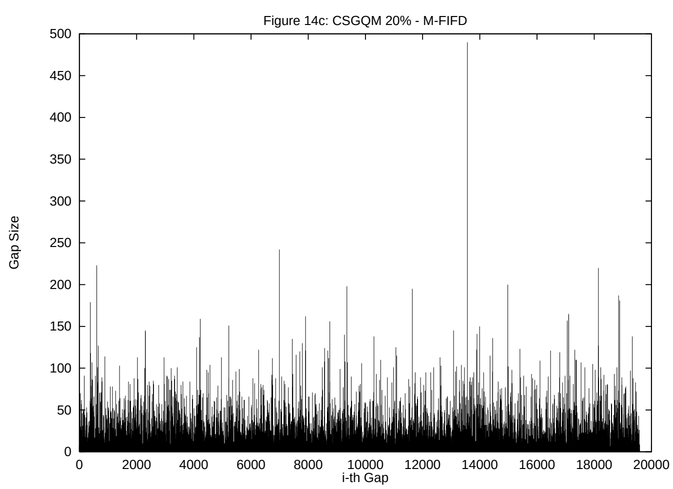
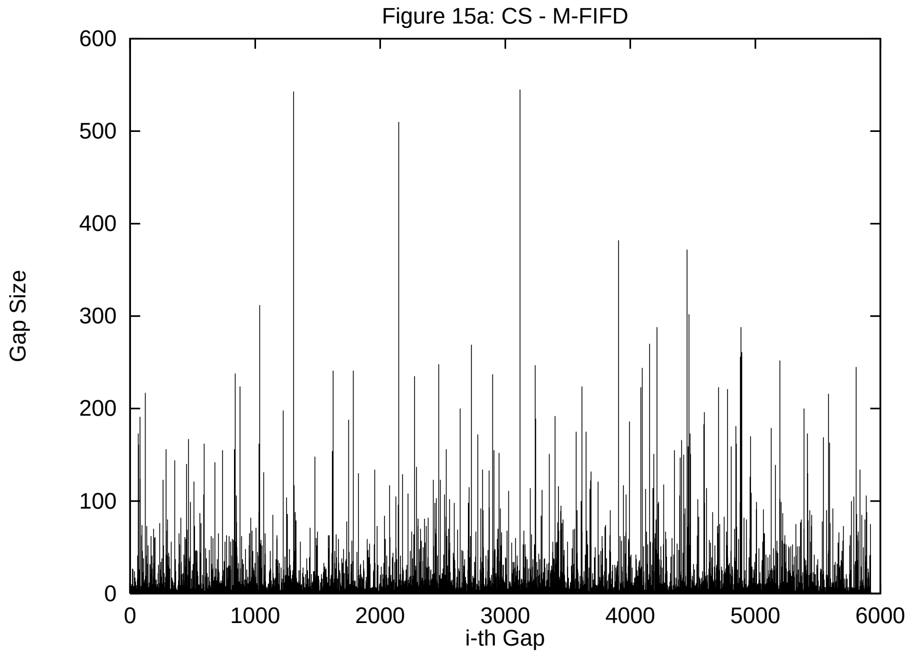


Figure 14: Low Priority Loss Gap Size x ith gap for Different Push ou Policies:FIFD (Figure 14.a), LIFD (Figure 14.b) and M-FIFD (Figure 14.c) for a CSGQM Queue with a Level of Protection of 20%, Buffer Size = 100, $H = 0.75$ $\rho = 0.8$, $\sigma^2 = 1$, $P(high / high) = 0.7$, $\gamma = 0.0$



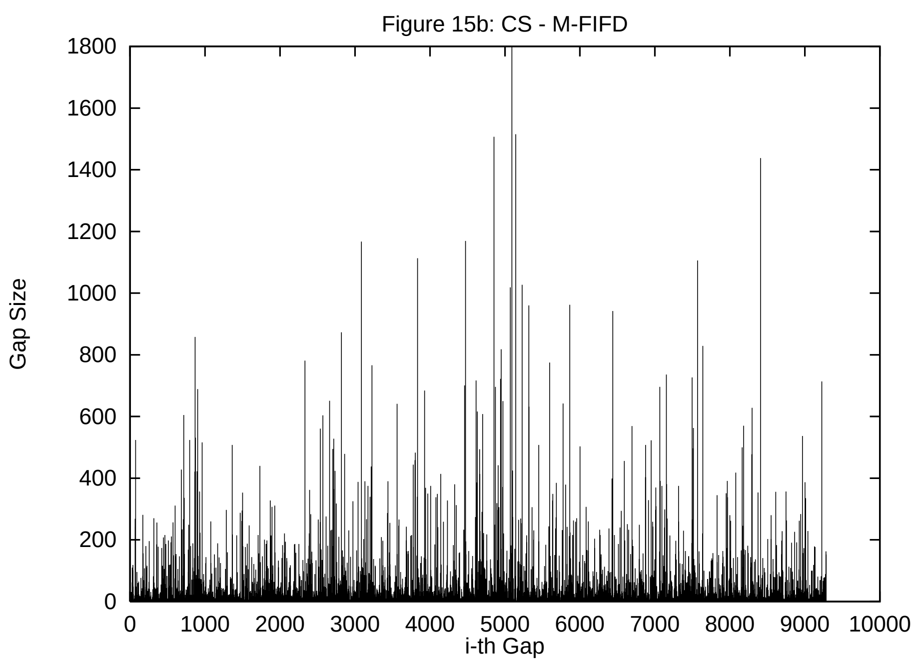


Figure 15: Low Priority Loss Gap Size x ith gap for M-FIFD and Different Values of *H*: H = 0.7 (Figure 15.a) and *H* = 0.8 (Figure 15.b), Buffer Size = 400, ρ =0.8, σ^2=1 *P(high / high)* = 0.7, γ= 1