

On the Effects of the Packet Size Distribution on FEC Performance

György Dán* Viktória Fodor Gunnar Karlsson

*KTH, Royal Institute of Technology,
Department of Signals, Sensors and Systems*

Abstract

For multimedia traffic like VBR video, knowledge of the average loss probability is not sufficient to determine the impact of loss on the perceived visual quality and on the possible ways of improving it, for example by forward error correction (FEC) and error concealment. In this paper we investigate how the packet size distribution affects the packet loss process, i.e. the probability of consecutive losses and the distribution of the number of packets lost in a block of packets and the related FEC performance. We present an exact mathematical model for the loss process of an $MMPP + MMPP/E_r/1/K$ queue and compare the results of the model to simulations performed with various other packet size distributions (PSDs), among others, the measured PSD from an Internet backbone. The results show that analytical models of the PSD matching the first three moments (mean, variance and skewness) of the empirical PSD can be used to evaluate the performance of FEC in real networks. We conclude that the exponential PSD, though it is not a worst case scenario, is a good approximation for the PSD of today's Internet to evaluate FEC performance. We also conclude that the packet size distribution affects the packet loss process and thus the efficiency of FEC mainly in access networks where a single multimedia stream might affect the multiplexing behavior. We evaluate how the PSD affects the accuracy of the widely used Gilbert model to calculate FEC performance and conclude that the Gilbert model can capture loss correlations better if the CoV of the PSD is high.

⁰ This work has been supported in part by E-NEXT and by the Swedish Foundation for Strategic Research under the program Affordable Wireless Services and Infrastructures.

* Corresponding author, Address: Osquldas väg 10, 10044 Stockholm, Sweden, Tel.: +46 8 790 4253, Fax: +46 8 752 6548

Email addresses: gyuri@kth.se (György Dán), viktorias3.kth.se (Viktória Fodor), gk@s3.kth.se (Gunnar Karlsson).

1 Introduction

For flow-type multimedia communications, as opposed to elastic traffic, the average packet loss is not the only measure of interest. The burstiness of the loss process, the number of losses in a block of packets, has a great impact both on the user-perceived visual quality and on the possible ways of improving it, for example by error concealment and forward error correction.

Forward error correction (FEC) is an attractive means to decrease the loss probability experienced by delay sensitive traffic, such as real-time multimedia, when ARQ schemes can not be used to recover losses due to strict delay constraints. There are two main directions of FEC design to recover from packet losses. One solution, proposed by the IETF and implemented in Internet audio tools like Rat [1] and Freephone [2] is to add a redundant copy of the original packet to one of the subsequent packets [3]. In the case of packet loss the information is regained from the redundant copy. The other set of solutions use media-independent block coding schemes based on algebraic coding, e.g. Reed-Solomon coding [4]. The error correcting capability of RS codes with k data packets and c redundant packets is c if data is lost. While Reed-Solomon codes are typically used to correct bit errors, they can be used to recover lost packets via block interleaving as described in [5]. Given a block of k packets, the packets are prefixed by their lengths in bytes, and packets shorter than the longest one in the block are padded by zeros. Reed-Solomon coding is applied to the i^{th} symbol (typically byte) of each packet (in total k symbols) to form the i^{th} symbols of the c redundant packets. Packets are then transmitted one-by-one, without the padding zeros. The loss of a packet appears as the loss of a symbol in a block of $c + k$ symbols at the receiver and can be corrected as long as the number of lost packets is no more than c . The performance of both FEC schemes depends on the burstiness of the loss process: the performance of the first, media-dependent FEC scheme depends on the probability of consecutive packet losses; the capability of the second, media-independent FEC scheme to recover from losses depends on the distribution of the number of packets lost in a block.

The burstiness of the loss process in the network can be influenced by three factors, the burstiness of the stream traversing the network, the burstiness of the background traffic and the packet size distribution. The effects of the burstiness of the stream traversing the network and the background traffic have been investigated before. In [6] the authors performed simulations to study the efficiency of FEC for recovery of lost packets. They concluded that as long as the ratio of streams using FEC is low, FEC can decrease the uncorrected loss probability significantly. In [4] the authors used an analytical model to evaluate how the burstiness, the block length and the number of sources employing FEC influences the efficiency of FEC, concluding that the burstiness has a significant effect on the efficiency of FEC. The authors in [7] performed simulations to investigate the trade-off between traffic shaping and

FEC to decrease the uncorrected loss probability. They concluded that for longer delays the joint use of shaping and FEC gives the best performance. The authors studied the effects of the burstiness of the background traffic on the efficiency of FEC as well, and concluded that its effects are moderate. In [8] the authors used the Gilbert channel model [9] and evaluated the efficiency of FEC as a function of the burstiness of the loss process, concluding that a burstier loss process decreases the efficiency of FEC.

The effects of the packet size distribution (PSD) are not clear however, since previous research concentrated on the exponential and deterministic packet size distributions only. The deterministic packet size distribution was motivated by the ATM standard in the last decade, while the exponential packet size distribution was motivated by the analytical tractability of the resulting models. Nevertheless, the PSD in the network can vary on the short term due to changes in the ongoing traffic and on the long term as new applications and protocols emerge. As individual applications cannot control the PSD in the network, it is important to know how the PSD will affect their performance, for example, how much gain can an application expect from FEC given a certain measured end-to-end average loss probability.

It is well known that in an $M/G/1$ queue the average number of customers is direct proportional to the coefficient of variation (CoV) of the service time distribution, as given by the Pollaczek-Khintchine formula [10]. For the finite capacity $M/G/1$ queue there is no closed form formula to calculate the packet loss probability [11,12], though we know from experience that a lower CoV of the service time distribution yields lower average loss probability. It is however unclear how the distribution of the service time affects the loss process in a finite queue and thus how the potential of using FEC changes.

In this paper we present a queuing model to analyze the packet loss process of a bursty source, for example VBR video, multiplexed with background traffic in a single multiplexer. The multiplexer is modeled as a single server with a finite queue. We model the bursty source and the background traffic by Markov-modulated Poisson processes (MMPP) and consider Erlang- r distributed packet sizes. We investigate the effects of the network PSD on the packet loss process and the efficiency of FEC based on the queuing model presented here, a similar model with deterministic packet sizes and simulations. In particular, we compare the analytical results with Erlang- r (exponential, as a special case) and deterministic PSDs to those of simulations performed with general PSDs, among them the measured PSD of an Internet backbone [13].

The paper is organized as follows. Section 2 gives an overview of the previous work on the modeling of the loss process of a single server queue. In Section 3 we describe our model used for calculating the loss probabilities in a block of packets and the consecutive loss probability. In Section 4 we evaluate the effects of the PSD on the packet loss process in various scenarios. We consider constant average

load in Subsection 4.1, constant average loss probability in Subsection 4.2, and we isolate the effect of the PSD from other factors in Subsection 4.3. In Section 5 we evaluate how the packet size distribution affects the accuracy of the Gilbert model to capture the correlation between losses. We conclude our work in Section 6.

2 Related work

In [14], Cidon et al. presented an exact analysis of the packet loss process in an $M/M/1/K$ queue, that is the probability of losing j packets in a block of n packets, and showed that the distribution of losses may be bursty compared to the assumption of independence. They also considered a discrete time system fed with a Bernoulli arrival process describing the behavior of an ATM multiplexer. Gurewitz et al. presented explicit expressions for the above quantities of interest for the $M/M/1/K$ queue in [15]. In [16], Altman et al. obtained the multidimensional generating function of the probability of j losses in a block of n packets and gave an easy-to-calculate asymptotic result under the condition that $n \leq K + j + 1$.

Schulzrinne et al. [17] derived the conditional loss probability (CLP) for the $N * IPP/D/1/K$ queue and showed that the CLP can be orders of magnitude higher than the loss probability. In [4] Kawahara et al. used an interrupted Bernoulli process to analyze the performance of FEC in a cell switched environment. The loss process of the $MMPP/D/1/K$ queue was analyzed in [18] and the results compared to a queue with exponential packet size distribution.

Models with general service time distribution have been proposed for calculating various measures of queuing performance [19,20], but not to analyze the loss process. Though models with exponential and deterministic PSDs are available, a thorough analysis of the effects of the PSD on the packet loss process has not yet been done.

3 Model description

In this section we first present the model used to calculate the probability of losses in a block, then we show how it can be used to calculate the consecutive loss probability.

Flows traversing large networks like the Internet cross several routers before reaching their destination. However, most of the losses in a flow occur in the router having the smallest available bandwidth along the transmission path, so that one may model the series of routers with a single router, the bottleneck [21,22].

We consequently model the network with a single queue with Erlang- r distributed packet sizes having average transmission time $1/\mu$. The Erlang- r distribution is the distribution of the sum of r independent identically distributed random variables each having an exponential distribution. By increasing r to infinity the variance of the Erlang- r distribution goes to zero, and thus the distribution becomes deterministic.

Packets arrive to the system from two sources, two Markov-modulated Poisson processes (MMPP), representing the tagged source ($MMPP_a$) and the background traffic ($MMPP_s$) respectively. The packets are stored in a buffer that can host up to K packets, and are served according to a FIFO policy.

It is well known that compressed multimedia, like VBR video, exhibits a self-similar nature [23]. In [24] Robert and Le Boudec used a discrete time MMPP to fit the mean and the Hurst parameter of pseudo self-similar traffic. In [25] Andersen and Nielsen used the superposition of two-state MMPP's to model second-order self-similar behavior over several timescales. Yoshihara et al. used the superposition of 2-state interrupted Poisson processes (IPPs) and a Poisson process to model self-similar traffic in [26] and compared the behavior of the resulting $MMPP/D/1$ queue with simulations. They found that the approximation works well under heavy load conditions and gives a tight upper bound on the queue length. Klemm et al. [27] used the batch Markovian arrival process for aggregated traffic modeling in IP networks, and showed the effectiveness of the model in terms of queueing behavior and statistical properties of the traffic. Ryu and Elwalid [28] showed that short term correlations have dominant influence on the network performance under realistic scenarios of buffer sizes for real-time traffic. Based on these previous works we argue that the MMPP may be a practical model to derive approximate results for the queueing behavior of long range dependent traffic such as real-time VBR video, especially in the case of small buffer sizes [29].

Our assumption on the background traffic is justified by recent results indicating that Internet traffic can be approximated by a non-stationary Poisson process [30]. According to the results the change free intervals are well above 150 ms, the ITU's G.114 recommendation for end-to-end delay for real-time applications. The authors in [31] used the superposition of MMPPs to model self-similar traffic over several timescales, and achieved good results in terms of queueing behavior. These empirical results are consistent with recent theoretical results [32].

3.1 Probability of losses in a block

In the following we describe the calculation of the probability of losses in a block. Every n consecutive packets from the tagged source form a block, and we are interested in the probability distribution of the number of lost packets in a block in

the steady state of the system. Throughout this section we use notations similar to those in [14].

We assume that the sources feeding the system are independent. $MMPP_a$ is described by the infinitesimal generator matrix Q_a with elements r_{bc}^a , ($b, c \in \mathcal{B} = (1 \dots B)$) and the arrival rate matrix $\Lambda_a = \text{diag}\{\lambda_1^a, \dots, \lambda_B^a\}$, where λ_b^a is the average arrival rate while the underlying Markov chain is in state b [33]. $MMPP_s$ is described by the infinitesimal generator matrix Q_s with elements r_{uv}^s , ($u, v \in \mathcal{U} = (1 \dots U)$) and the arrival rate matrix $\Lambda_s = \text{diag}\{\lambda_1^s, \dots, \lambda_U^s\}$, where λ_u^s is the average arrival rate while the underlying Markov chain is in state u . Let us denote the joint state space of the two MMPP's with the set of ordered pairs $\mathcal{B}\mathcal{U} = (b, u)$ of cardinality $B \times U$.

The superposition of the two sources can be described by a single MMPP with arrival rate matrix $\hat{\Lambda} = \Lambda_a \oplus \Lambda_s$, and infinitesimal generator $\hat{Q} = Q_a \oplus Q_s$, where \oplus is the Kronecker sum [33,34]. Both $\hat{\Lambda}$ and \hat{Q} are square matrices of size $L = B \times U$ and we denote the state of the superposed MMPP with $l \in \mathcal{L} = (1 \dots L)$. Due to the special structure of $\hat{\Lambda}$, $l = (b-1) \times U + u$ is a one-to-one and onto mapping from $\mathcal{B}\mathcal{U}$ to \mathcal{L} , with the property that the arrival intensity in state $l \in \mathcal{L}$ of the superposed MMPP is $\hat{\lambda}_l = \lambda_b^a + \lambda_u^s$ for $l = (b-1) \times U + u$, $b \in \mathcal{B}, u \in \mathcal{U}$. Since the mapping is invertible, we can calculate $\hat{\lambda}_l^a$, the arrival intensity of $MMPP_a$ in state l of the superposed MMPP, as $\hat{\lambda}_l^a = \lambda_{[l/U]+1}^a$.

Each packet in the queue corresponds to r exponential stages, and thus the state space of the queue is $\{0, \dots, rK\} \times \{1, \dots, L\}$. Our purpose is to calculate the probability of j losses in a block of n packets $P(j, n)$, $n \geq 1$, $0 \leq j \leq n$. We define the probability $P_{i,l}^a(j, n)$, $0 \leq i \leq rK$, $l \in \mathcal{L}$, $n \geq 1$, $0 \leq j \leq n$ as the probability of j losses in a block of n packets, given that the remaining number of exponential stages in the system is i just before the arrival of the first packet in the block from the tagged source and the first packet of the block is generated while the superposed $MMPP$ is in state l . As the first packet in the block is arbitrary,

$$P(j, n) = \sum_{l=1}^L \sum_{i=0}^{rK} \Pi(i, l) P_{i,l}^a(j, n). \quad (1)$$

$\Pi(i, l)$, the steady state distribution of the exponential stages in the queue as seen by an arriving packet from the tagged source can be derived from the steady state distribution of the $MMPP/E_r/1/K$ queue as

$$\Pi(i, l) = \frac{\pi(i, l) \hat{\lambda}_l^a}{\sum_{l=1}^L \hat{\lambda}_l \sum_{i=0}^{rK} \pi(i, l)}, \quad (2)$$

where $\pi(i, l)$ is the steady state distribution of the $MMPP/E_r/1/K$ queue and $\hat{\lambda}_l^a$ is

the arrival intensity of $MMPP_a$ in state l of the superposed MMPP.

The probabilities $P_{i,l}^a(j, n)$ can be derived according to the following recursion. The recursion is initiated for $n = 1$ with the following relations

$$P_{i,l}^a(j, 1) = \begin{cases} 1 & j = 0 \\ 0 & j \geq 1 \end{cases} \quad i \leq r(K-1),$$

$$P_{i,l}^a(j, 1) = \begin{cases} 0 & j = 0, j \geq 2 \\ 1 & j = 1 \end{cases} \quad r(K-1) < i. \quad (3)$$

We denote the probability that a packet arriving in state m of the superposed MMPP comes from the tagged source with $p_m = \frac{\lambda_b}{\lambda_b + \lambda_u}$ and the probability of the complement event with $\bar{p}_m = \frac{\lambda_u}{\lambda_b + \lambda_u}$, where $m = (b-1) \times U + u$. Thus for $n \geq 2$ the following equations hold.

$$P_{i,l}^a(j, n) = \sum_{m=1}^L \sum_{k=0}^{i+r} Q_{i+r}^{l,m}(k) \{p_m P_{i+r-k,m}^a(j, n-1) + \bar{p}_m P_{i+r-k,m}^s(j, n-1)\} \quad (4)$$

for $0 \leq i \leq r(K-1)$, and for $r(K-1) < i$

$$P_{i,l}^a(j, n) = \sum_{m=1}^L \sum_{k=0}^i Q_i^{l,m}(k) \{p_m P_{i-k,m}^a(j-1, n-1) + \bar{p}_m P_{i-k,m}^s(j-1, n-1)\} \quad (5)$$

$P_{i,l}^s(j, n)$ is given by

$$P_{i,l}^s(j, n) = \sum_{m=1}^L \sum_{k=0}^{i+r} Q_{i+r}^{l,m}(k) \{p_m P_{i+r-k,m}^a(j, n) + \bar{p}_m P_{i+r-k,m}^s(j, n)\}, \quad (6)$$

for $0 \leq i \leq r(K-1)$, and for $r(K-1) < i$

$$P_{i,l}^s(j, n) = \sum_{m=1}^L \sum_{k=0}^i Q_i^{l,m}(k) \{p_m P_{i-k,m}^a(j, n) + \bar{p}_m P_{i-k,m}^s(j, n)\}. \quad (7)$$

The probability $P_{i,l}^s(j, n)$, $0 \leq i \leq rK$, $l \in \mathcal{L}$, $n \geq 1$, $0 \leq j \leq n$ is the probability of j losses in a block of n packets, given that the remaining number of exponential stages in the system is i just before the arrival of a packet from the background traffic and the superposed $MMPP$ is in state l . $Q_i^{l,m}(k)$ denotes the joint probability of that the next arrival will be while the superposed $MMPP$ is in state m and that k

exponential stages out of i will be completed before the next arrival from the joint arrival process given that the last arrival was in state l of the superposed $MMPP$. A way to calculate $Q_i^{l,m}(k)$ is shown in the Appendix.

The procedure of computing $P_{i,l}^a(j,n)$ is as follows. First we calculate $P_{i,l}^a(j,1), i = 0 \dots rK$ from the initial conditions (3). Then in iteration k we first calculate $P_{i,l}^s(j,k), k = 1 \dots n-1$ using equations (6) and (7) and the probabilities $P_{i,l}^a(j,k)$, which have been calculated during iteration $k-1$. Then we calculate $P_{i,l}^a(j,k+1)$ using equations (4) and (5).

In the special case when the background traffic is a Poisson process with arrival intensity λ , we have that $U = 1, L = B, \hat{\Lambda} = \Lambda_a + \lambda I$ and $\hat{Q} = Q_a$, where I is an identity matrix of size B . The mapping from \mathcal{BU} to \mathcal{L} becomes $l = b$, and thus $p_m = \frac{\lambda_m}{\lambda_m + \lambda}$ as presented in [35]. We will use the resulting $MMPP + M/E_r/1/K$ model in Section 4.

3.2 Consecutive loss probability

Now we turn to calculate the probability of consecutive losses. We define two sets of states, α and ω as the set of states of the queue where arriving packets can enter the system and where arriving packets are discarded respectively. Then $\alpha = \{0 \dots r(K-1)\}$ and $\omega = \{r(K-1) + 1 \dots rK\}$. Let us denote by A_i the event that the first packet in a block arrives to the system when the remaining number of exponential stages in the system is i , and we define the event $A_\omega = \cup_{i \in \omega} A_i$. Similarly, we denote with A_l the event that the first packet in the block was generated in state l of the superposed $MMPP$. Using these notations the consecutive loss probability, i.e. the conditional probability that a packet arriving to the system from the tagged source is lost, given that the previous packet from the tagged source was lost is given as

$$\begin{aligned}
p_{\omega|\omega} &= P(2,2|A_\omega) = \frac{P(2,2 \cap A_\omega)}{P(A_\omega)} = \frac{\sum_{i \in \omega} P(2,2 \cap A_i)}{P(A_\omega)} = \\
&= \frac{\sum_{l \in \mathcal{L}} \sum_{i \in \omega} P(2,2 \cap A_i \cap A_l)}{P(A_\omega)} = \frac{\sum_{l \in \mathcal{L}} \sum_{i \in \omega} P_{i,l}^a(2,2) \Pi(i,l)}{P(A_\omega)} \\
&= \frac{\sum_{l \in \mathcal{L}} \sum_{i \in \alpha \cup \omega} P_{i,l}^a(2,2) \Pi(i,l)}{P(A_\omega)} = \frac{P(2,2)}{P(1,1)}, \tag{8}
\end{aligned}$$

since the first packet is arbitrary and $P_{i,l}^a(2,2) = 0$ for $i \in \alpha$. The probabilities $p_{\alpha|\alpha}$, $p_{\omega|\alpha}$ and $p_{\alpha|\omega}$ can be defined similarly and calculated as $p_{\alpha|\omega} = 1 - p_{\omega|\alpha}$, $p_{\omega|\alpha} = p_{\alpha|\omega} p_\omega / (1 - p_\omega)$ and $p_{\alpha|\alpha} = 1 - p_{\omega|\alpha}$, where p_ω is the average loss probability.

4 Performance analysis

In this section we show results obtained with the $MMPP + M/E_r/1/K$ model described in Section 3, the $MMPP + M/D/1/K$ model described in [18] and simulations. The average packet length of both the tagged and the background traffic is set to 454 bytes, which is the mean packet size measured on an Internet backbone [13]. Note that increasing the average packet length is equivalent to decreasing the link speed, and thus the particular fixed value of the average packet length does not limit the generality of the results presented here. The PDF, CoV (σ/m) and skewness ($\sum (X - m)^3/\sigma^3$) parameters of the twelve considered PSDs are shown in Table 1. The G1 distribution is the measured PSD on a 2.5 Gbps Internet backbone link as given by the Sprint IP Monitoring project [13]. The considered link speeds are 10 Mbps, 22.5 Mbps and 45 Mbps. The maximum queuing delay is set to around 1.5 ms in all cases, resulting in buffer sizes from 5 to 20 packets depending on the link speed. Both in the analytical models and in the simulations we consider a 3 state MMPP as the tagged source, with an average bitrate of 540 kbps, arrival intensities $\lambda_1^a = 116/s$, $\lambda_2^a = 274/s$, $\lambda_3^a = 931/s$ and transition rates $r_{12}^a = 0.12594$, $r_{21}^a = 0.25$, $r_{23}^a = 1.97$, $r_{32}^a = 2$. These values were derived from an MPEG-4 encoded video trace by matching the average arrival intensities in the three states of the MMPP with the average frame size of the I,P and B frames.

For the background traffic we use a Poisson process. This assumption is valid if there are many sources sharing the same link, as the traffic generated by a large number of sources tends to Poisson as the load increases due to statistical multiplexing [36]. We believe that this assumption will not influence our results with respect to the effects of the PSD, on the other hand it keeps the number of parameters of the model low, and thus eases readability. The simulations were performed in ns-2, the simulation time was between 40 thousand and 400 thousand seconds (5-50 million packets from the tagged source), the margin of error of the simulations was below 5 percent at a 95 percent confidence level.

We use three measures to compare the packet loss process. The first measure is the consecutive packet loss probability, denoted by $p_{\omega|\omega}$ and calculated according to (8). The consecutive packet loss probability has an influence on the efficiency of media-dependent FEC schemes proposed for real-time audio [3,37]. The lower the value of the consecutive packet loss probability, the more effective are the media-dependent FEC schemes, as shown in [38].

The second measure is the Kullback-Leibler distance [39] of the distributions of the number of packets lost in a block. The Kullback-Leibler distance is a commonly used measure of closeness, defined for two distributions as

$$d(p_1, p_2) = \sum_{j=0}^n P_1(j, n) \log_2 \frac{P_1(j, n)}{P_2(j, n)}, \quad (9)$$

Distribution	CoV	Skewness	PDF	Notation
General 1	1.2	1.07	$b(x)$ taken from [13], see Figure 1	G1
General 2	1.2	1.07	$b(x) = 0.74N(127, 20) + 0.26N(1366, 20)$	G2
Phase type	1.2	1.07	$b(x) = 0.54E(5, 26) + 0.46E(5, 956)$	G3
Exponential	1	2	$E(1, 454)$	M*
General 4	1	$\sqrt{2}$	$b(x) = 0.79N(219, 1) + 0.21N(1331, 1)$	G4
General 5	$1/\sqrt{2}$	2	$b(x) = 0.85N(321, 1) + 0.15N(1229, 1)$	G5
Erlang-2	$1/\sqrt{2}$	$\sqrt{2}$	$E(2, 454)$	E2*
General 6	$1/\sqrt{2}$	$\sqrt{0.4}$	$b(x) = 0.65N(219, 1) + 0.35N(892, 1)$	G6
General 7	$\sqrt{0.1}$	$\sqrt{2}$	$b(x) = 0.79N(379, 1) + 0.21N(731, 1)$	G7
Erlang-10	$\sqrt{0.1}$	$\sqrt{0.4}$	$E(10, 454)$	E10*
General 8	$\sqrt{0.1}$	0	$b(x) = 0.5N(310, 1) + 0.5N(598, 1)$	G8
Deterministic	0	0	$b(x) = \delta_{454}(x)$	D*

Table 1

Considered packet size distributions: coefficient of variation, skewness, PDF and notation in the figures. $N(m, \sigma)$ denotes a normal distribution with mean m and variance σ^2 . $E(r, 1/\mu)$ denotes an r -stage Erlang distribution with mean $1/\mu$. Results for PSDs marked with a * are obtained with the models, the rest with simulations.

The Kullback-Leibler distance is the same as the relative entropy of p_1 with respect to p_2 . It is not a true metric, as it is not symmetric and does not satisfy the triangle inequality, but it is always non-negative and equals zero only if $p_1 = p_2$.

The third measure is based on the gain that can be achieved by using FEC. We denote an FEC scheme using Reed-Solomon coding with k data packets and c redundant packets by $FEC(k, c+k)$. Given the probabilities $P(j, n)$ the uncorrected loss probability for an $FEC(k, c+k)$ scheme can be calculated as

$$p_{\omega}^{k, c+k} = \frac{1}{c+k} \sum_{j=c+1}^{c+k} jP(j, c+k). \quad (10)$$

Based on the uncorrected packet loss probability we define the FEC gain as the ratio of the average loss probability without the use of FEC and the uncorrected loss probability when using FEC

$$f(k, c+k) = p_{\omega} / p_{\omega}^{k, c+k}. \quad (11)$$

4.1 Constant average load case

In this subsection we investigate the effects of the PSD on the packet loss process and the efficiency of FEC as a function of the average load in the network. Figure 2 shows the uncorrected packet loss probability without error correction (denoted by FEC(1,1)), for FEC(10,11) and for FEC(20,22) on a 10 Mbps link for the G1,G2, G3 (which have the same first three moments), M and D distributions. Figures 3 and 4 show the same results on a 22.5 Mbps and a 45 Mbps link. The figures show that results obtained with the G1, G2 and G3 distributions are practically the same (the difference is less than 5%). This indicates that by matching the first three moments of a distribution one can derive accurate results in terms of average loss probability and FEC gain even for very low loss probabilities. In the following we will only use the G1 distribution out of these three distributions.

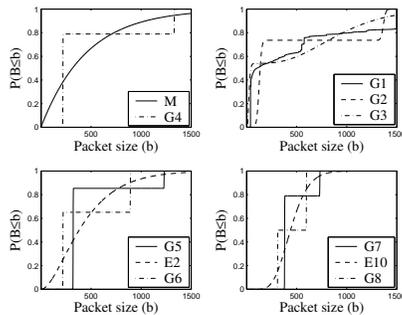


Fig. 1. Cumulative density functions of the considered packet size distributions.

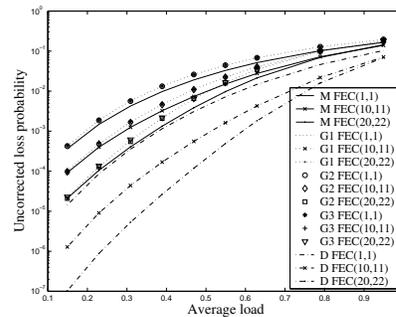


Fig. 2. Average loss probability with and without FEC vs average load on a 10 Mbps link.

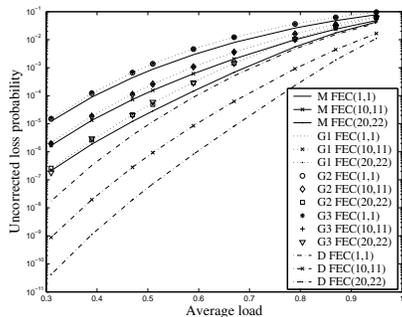


Fig. 3. Average loss probability with and without FEC vs average load on a 22.5 Mbps link.

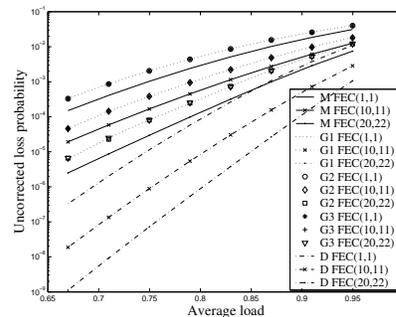


Fig. 4. Average loss probability with and without FEC vs average load on a 45 Mbps link.

Figures 5 and 6 show the Kullback-Leibler distance obtained with different PSDs on a 10 Mbps link for $P(j, 11)$ and $P(j, 22)$ respectively. Figures 7 and 8 show the FEC gain for the same scenarios. Comparing the figures we conclude that FEC(10,11) and FEC(20,22) are qualitatively similar, and thus in the following we will only show figures for FEC(20,22) for brevity. Figure 9 shows the conditional packet loss probability obtained with different PSDs on a 10 Mbps link. Comparing

results obtained with PSDs having the same CoV but different skewness we can see that even though the skewness has an effect on the packet loss process (especially at low loss probabilities), the CoV of the PSD has the biggest impact on both the conditional loss probability and the efficiency of FEC. As shown in the figures, a lower CoV value yields a less correlated loss process.

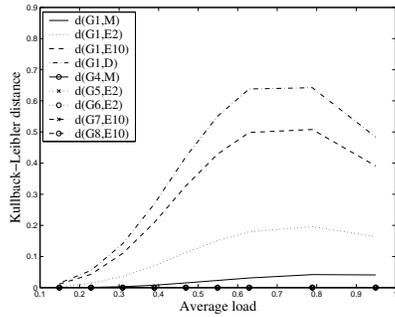


Fig. 5. Kullback-Leibler distance vs average load for $P(j,11)$ on a 10 Mbps link

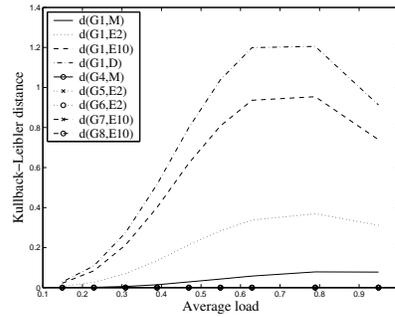


Fig. 6. Kullback-Leibler distance vs average load for $P(j,22)$ on a 10 Mbps link

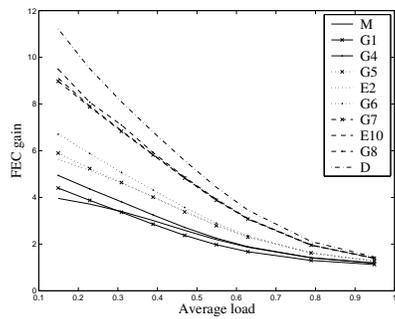


Fig. 7. FEC gain vs average load for $FEC(10,11)$ on a 10 Mbps link

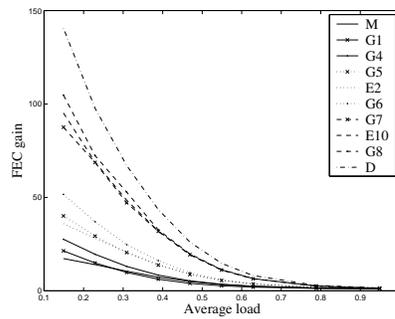


Fig. 8. FEC gain vs average load for $FEC(20,22)$ on a 10 Mbps link

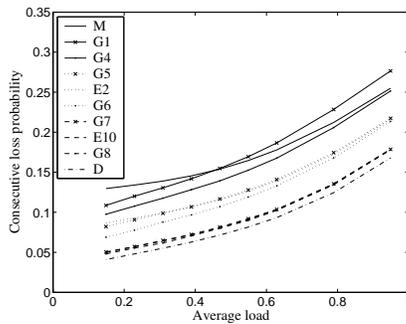


Fig. 9. Conditional loss probability vs average load on a 10 Mbps link

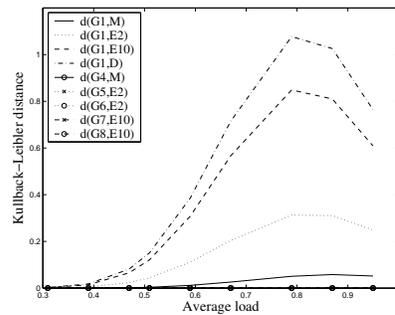


Fig. 10. Kullback-Leibler distance vs average load for $P(j,22)$ on a 22.5 Mbps link

We draw the same conclusion by examining Figures 10, 11, 13 and 14 which show the Kullback-Leibler distance and the FEC gain on a 22 Mbps and a 45 Mbps link as a function of the average load for $P(j,22)$ and $FEC(20,22)$ respectively. The same conclusion can be drawn from Figure 12, which shows the consecutive packet loss probability on a 22 Mbps link. Thus, analytically tractable PSD models (for example phase-type, which includes both the Erlang and the hyper-exponential

distributions as special cases, and has an extensive literature [40–43]) can be used to derive approximate results for FEC performance by matching the first two, and accurate results by matching the first three moments of the empirical PSD. Furthermore as the packet size and the CoV of the PSD in the network is bounded from above, one can estimate a lower bound on the achievable gain of using FEC independent of the packet size distribution in the network. Comparing Figures 8, 11 and 14 shows that the ratio of the FEC gains with different PSDs decreases slightly as the link speed increases. We draw the same conclusion regarding the conditional loss probabilities comparing Figures 9 and 12.

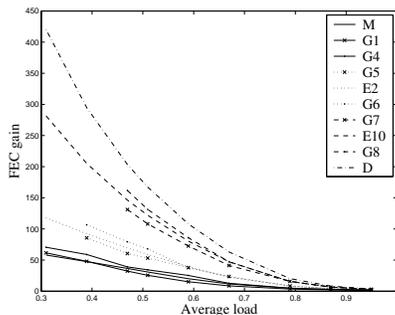


Fig. 11. FEC gain vs average load for FEC(20,22) on a 22.5 Mbps link

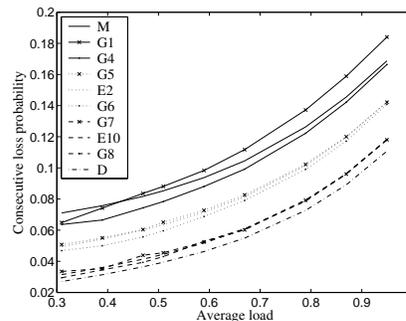


Fig. 12. Conditional loss probability vs average load on a 22.5 Mbps link

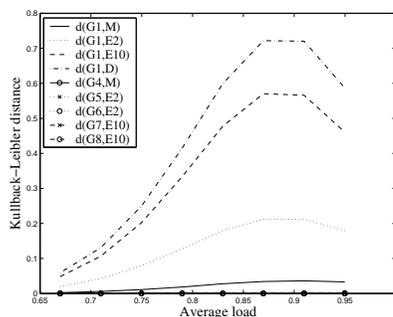


Fig. 13. Kullback-Leibler distance vs average load for $P(j,22)$ on a 45 Mbps link

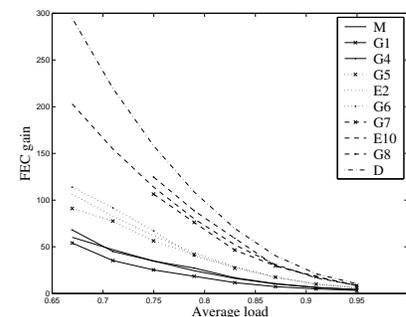


Fig. 14. FEC gain vs average load for FEC(20,22) on a 45 Mbps link

Though for some networks the exponential PSD might fit, it is clear from the results that it does not represent a worst case scenario if the average packet size is not equal to the center of the domain of the PSD and thus the CoV of the PSD can exceed one. Nevertheless, the exponential PSD is a good approximation for the considered empirical PSD G1, and for other empirical PSDs to be found at [13]. This finding justifies the assumption of exponential service time distribution in earlier works on the efficiency of FEC [3,14–16,44].

The difference between the results obtained at a particular average load with distributions having different CoV values is significant, up to one order of magnitude in terms of FEC gain in the considered scenarios and a factor of three in terms of conditional packet loss probability, a lower CoV value yielding a less bursty loss

process. The difference however is partly due to the different average loss probabilities. We eliminate the effects of the average loss probability in the following subsection.

4.2 Constant average packet loss case

In this subsection we consider results with different PSDs as a function of the average loss probability. This enables us to investigate what an application (unaware of the network PSD) can expect from the different FEC solutions given that it experiences a certain end-to-end average packet loss probability. In order to be able to compare the packet loss process at a certain average loss probability we take the results from simulations with the G1 PSD and increase the background traffic of the mathematical models to match the average packet loss probability given by the simulations.

Figure 15 shows the Kullback-Leibler distance between the results obtained with the different distributions as a function of the average loss probability on a 10 Mbps link for $P(j,22)$. The figure shows that the distance between the results obtained with different distributions decreased significantly (three orders of magnitude). Figure 16 shows the FEC gain for the same scenario. The effects of the PSD are significantly smaller compared to Figure 8.

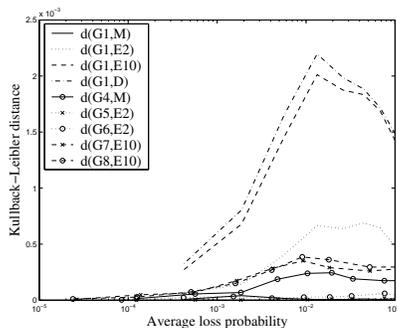


Fig. 15. Kullback-Leibler distance vs average loss probability for $P(j,22)$ on a 10 Mbps link

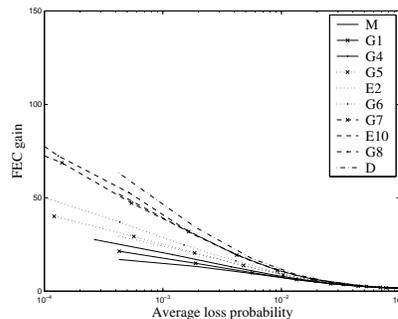


Fig. 16. FEC gain vs average loss probability for $FEC(20,22)$ on a 10 Mbps link

Figures 17 and 18 show the FEC gain on a 22.5 Mbps and a 45 Mbps link respectively. Comparing the figures we can see that the difference between results with different PSDs in terms of FEC gain decreases as the link speed increases (from 10 Mbps to 45 Mbps). More specifically, the efficiency of FEC in the case of exponential PSD increases, while in the case of deterministic PSD remains approximately unchanged at a given loss probability as the link speed increases. The reason for this is that the higher the link speed the less the background traffic has to be changed to keep the average loss probability constant, and thus the change in the level of statistical multiplexing decreases. Figures 19, 20 and 21 show the condi-

tional loss probability on a 10 Mbps, 22.5 Mbps and 45 Mbps link respectively, and lead to the same conclusions, i.e. the distance between the results obtained with different distributions is lower compared to Figs. 9 and 12, and the difference between results obtained with the different PSDs decreases as the link speed increases. The results for the conditional loss probability support those for the efficiency of FEC, a lower CoV results in lower conditional loss probability at a given average loss probability, which in turn makes FEC more efficient, since losses are spread out more evenly.

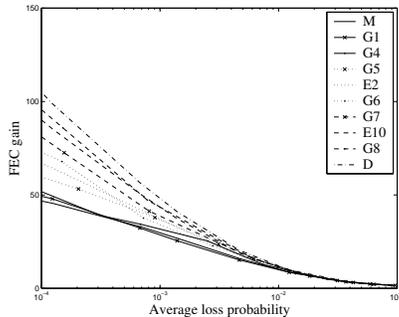


Fig. 17. FEC gain vs average loss probability for FEC(20,22) on a 22.5 Mbps link

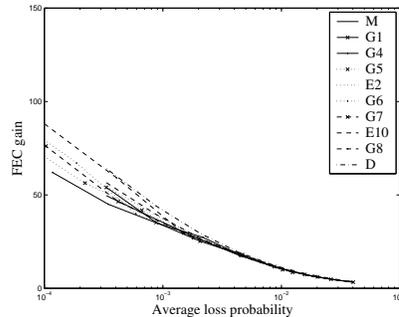


Fig. 18. FEC gain vs average loss probability for FEC(20,22) on a 45 Mbps link

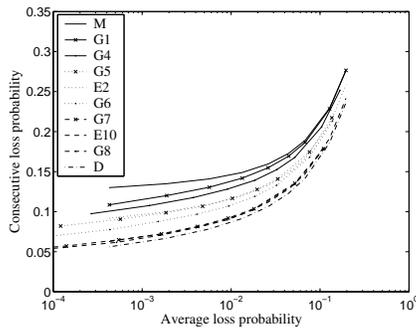


Fig. 19. Conditional loss probability vs average loss probability on a 10 Mbps link

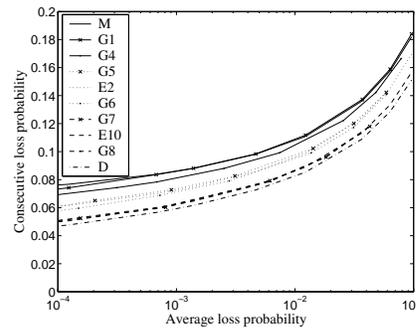


Fig. 20. Conditional loss probability vs average loss probability on a 22.5 Mbps link

The observed difference in FEC gain and conditional loss probability can be due to the difference in the level of statistical multiplexing (the background traffic intensity was increased to maintain the average loss probability constant and as a result the packet loss process became more independent) and to the difference between the packet size distributions.

4.3 Isolating the effects of the packet size distribution

In this subsection we separate the effects of the level of statistical multiplexing and the PSD. We do it by changing the arrival intensity of both the background traffic and the tagged stream in the mathematical models in order to match the average loss probability given by the simulations with the G1 PSD, thus we keep the level

of statistical multiplexing constant (doing so is equivalent to matching the average loss probability through decreasing the link speed). Figure 22 shows the Kullback-Leibler distance as a function of the average loss probability on a 10 Mbps link for $P(j,22)$. Comparing this to Figure 15 we can see a further significant decrease in the distance of the distributions. The same effect can be seen in Figure 23, which shows the FEC gain on a 10 Mbps link as a function of the average loss probability for FEC(20,22). Figure 24 shows the conditional loss probability on a 10 Mbps link. Comparing it to Figure 19 shows a further decrease in the difference of the results with different PSDs.

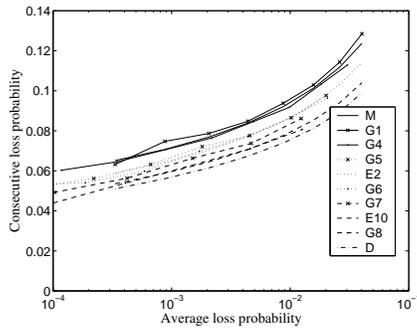


Fig. 21. Conditional loss probability vs average loss probability on a 45 Mbps link

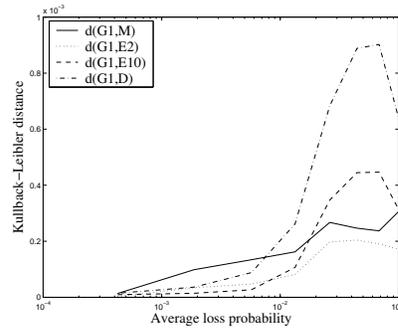


Fig. 22. Kullback-Leibler distance vs average loss probability for $P(j,22)$ on a 10 Mbps link (same level of statistical multiplexing)

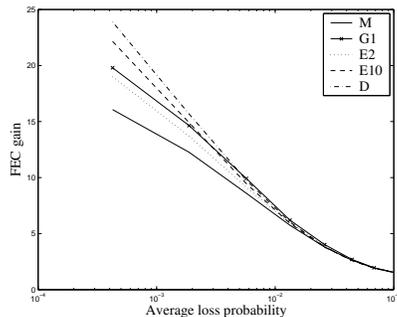


Fig. 23. FEC gain vs average loss probability for FEC(20,22) on a 10 Mbps link (same level of statistical multiplexing)

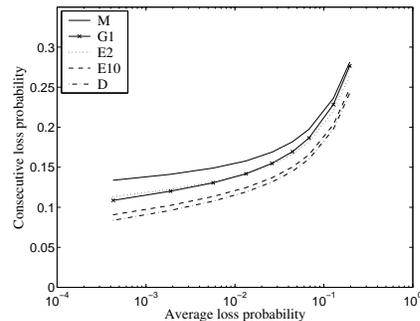


Fig. 24. Conditional loss probability vs average loss probability on a 10 Mbps link (same level of statistical multiplexing)

Thus the difference in the FEC gain and the conditional loss probability considering a fixed average load (Subsection 4.1) is mainly due to the different average loss probabilities and in a lower extent to the different PSD. Similarly, the difference in the FEC gain and the conditional loss probability considering a fixed average loss (Subsection 4.2) is mainly due to the different levels of statistical multiplexing and in a lower extent to the different PSD. This is in accordance with the observation in Section 4.2 that the difference between the results with different PSDs decreases as the link speed increases.

5 On the accuracy of the Gilbert model

The Gilbert model [9] is a widely used model for the design and evaluation of error control solutions for channels with correlated losses. As proposed in [9], it is a two state time-discrete Hidden Markov Model (HMM), where state 0 corresponds to the good state of the channel, state 1 to the bad state of the channel, and packets are lost with probability h when the channel is in the bad state. Let us denote the transition probability from state 0 to state 1 by p and from state 1 to state 0 by q . A special case of this model, where the value of the parameter h is taken to be one, is commonly used in the literature to describe packet loss correlations [45–47]. With this restriction, the Gilbert model becomes a two state time-discrete Markov model. Estimating the parameters of the two state Markov model is much easier than those of the HMM, which is the reason for this common simplification of the original model.

In the following we briefly evaluate how the packet size distribution influences the accuracy of the Gilbert model to predict FEC performance. For this end we first calculate the probability $p_{\omega|\omega}$ from the simulations or using the model presented in Section 3 respectively for given network conditions, i.e. average load and packet size distribution. The parameters of the Gilbert model are then set to match p_{ω} , $p_{\alpha|\omega}$ and $p_{\omega|\alpha}$ as

$$p = \frac{p_{\omega|\alpha}(p_{\omega} - p_{\omega|\alpha})}{p_{\omega}(1 - p_{\omega|\alpha} - p_{\alpha|\omega})},$$

$$q = \frac{p_{\omega|\alpha} - p_{\omega}p}{p_{\omega}}, \quad h = \frac{p_{\omega|\alpha}}{p}.$$

The above parametrization ensures $h = 1$, and is the maximum likelihood estimator for the commonly used two state Markov model as shown in [48]. Using the Gilbert model the probability of j losses in a block of n packets, $P_G(j, n)$ can be calculated [49], and can be used to calculate the uncorrected loss probability for an RS(k,c+k) scheme

$$p_{G,\omega}^{k,c+k} = \frac{1}{c+k} \sum_{j=c+1}^{c+k} j P_G(j, c+k), \quad (12)$$

and the corresponding FEC gain

$$f_G(k, c+k) = p_{\omega} / p_{G,\omega}^{k,c+k}. \quad (13)$$

To evaluate the effects of the PSD on the accuracy of the Gilbert model we define the ratio $\phi_G = f(k, c+k) / f_G(k, c+k)$, the ratio of the FEC gain as calculated from

simulations or using the model presented in Section 3 and as calculated with the Gilbert model. The further the value of $\phi_G(k, c+k)$ from 1 the higher the error induced by the Gilbert model, e.g. a value lower than one indicates that using the Gilbert model one overestimates the efficiency of FEC.

Figs. 25 and 26 show $\phi_G(10, 11)$ and $\phi_G(20, 22)$ respectively as a function of the average load on a 10 Mbps link for five packet size distributions corresponding to Subsection 4.1. The figures show that the Gilbert model is more accurate for packet size distributions with a high CoV. Similar conclusions can be drawn from Figs. 27 and 28, which show the same scenario for a link of 22.5 Mbps and 45 Mbps respectively. As it can be seen in the figures, below a certain average load the accuracy of the Gilbert model starts increasing, as losses become more and more independent.

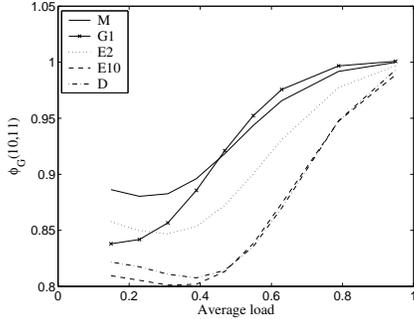


Fig. 25. $\phi_G(10, 11)$ vs average load on a 10 Mbps link

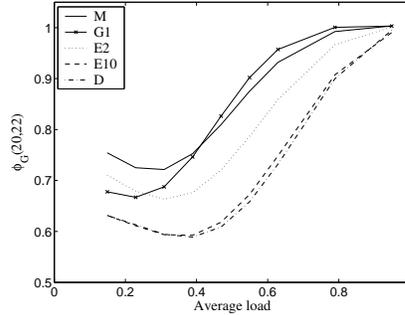


Fig. 26. $\phi_G(20, 22)$ vs average load on a 10 Mbps link

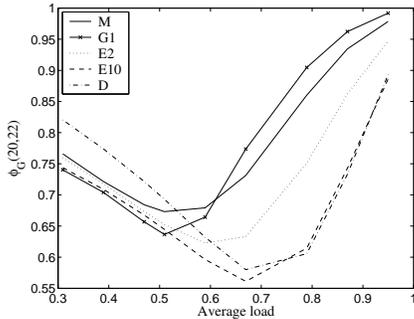


Fig. 27. $\phi_G(20, 22)$ vs average load on a 22.5 Mbps link

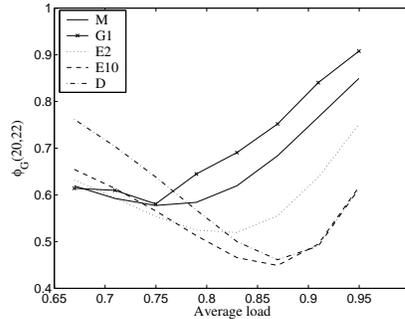


Fig. 28. $\phi_G(20, 22)$ vs average load on a 45 Mbps link

Figure 29 shows $\phi_G(20, 22)$ as a function of the average loss probability on a 10 Mbps link corresponding to Subsection 4.2. We observe that the difference between results with the different PSDs does not decrease significantly compared to Figure 26. Figs. 30 and 31 show the same scenario for links of 22.5 Mbps and 45 Mbps. As it can be seen in both figures, the Gilbert model performs worst when the average loss probability is around 10^{-3} . This explains the shape of the curves in Figs. 27 and 28. Comparing Figs. 29, 30 and 31 we also see that the accuracy of the Gilbert model decreases at a particular loss probability as the link speed increases. Figure 32 shows $\phi_G(20, 22)$ as a function of the average loss probability on a 10 Mbps link corresponding to Subsection 4.3. Comparing this to Figure 29

we see that the difference between the results obtained with different distributions increases. This phenomenon is due to that, in order to match the average loss probability, the load of the scenarios with a PSD with low CoV has to be increased as explained in Subsection 4.3. Increasing the load results in the increasing importance of long term correlations between losses and the Gilbert model fails to capture those correlations. Thus we conclude that the Gilbert model can be used best to model the loss process of a queue with a PSD with a high CoV, for example exponential distribution.

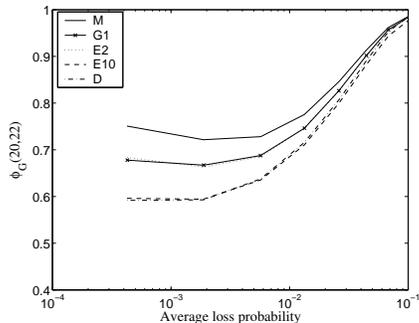


Fig. 29. $\phi_G(20,22)$ vs average loss probability on a 10 Mbps link

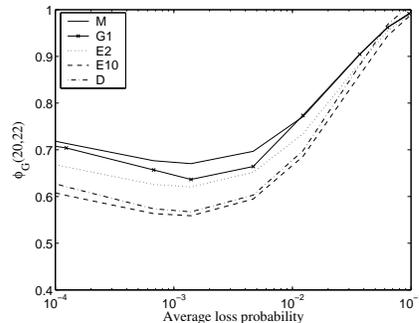


Fig. 30. $\phi_G(20,22)$ vs average loss probability on a 22.5 Mbps link

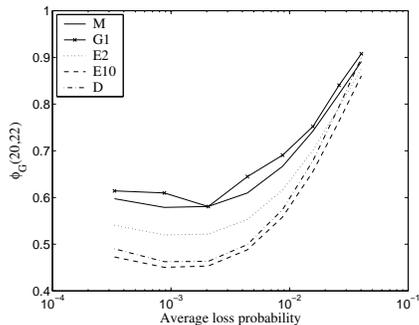


Fig. 31. $\phi_G(20,22)$ vs average loss probability on a 45 Mbps link

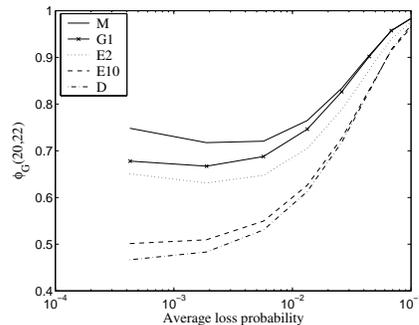


Fig. 32. $\phi_G(20,22)$ vs average loss probability on a 10 Mbps link (same level of statistical multiplexing)

6 Conclusion

In this paper we investigated the effects of the packet size distribution on the packet loss process and the related FEC performance in a single server queue with a finite buffer. We presented a mathematical model for the analysis of the packet loss process of the $MMPP + MMPP/E_r/1/K$ queue and compared the results of simulations and mathematical models in different scenarios. Our results show that analytical models of the PSD matching the first three moments of the empirical PSD can be used to evaluate the performance of the different FEC solutions in real networks, while the exponential PSD is a reasonable approximation for the PSD of

today's Internet to evaluate FEC performance. Nevertheless the exponential PSD is not a worst case scenario, the PSD in today's networks has a higher CoV and thus shows slightly worse queuing performance. However, as the CoV of the packet size distribution in a real network is bounded from above, one can give a lower bound on the efficiency of FEC and thus predict its performance. The results show that the effects of the packet size distribution decrease as the link speed increases if one considers a particular average packet loss probability. Thus at a given average loss probability the actual network PSD does not influence the efficiency of FEC on a backbone link, however it has a big influence on it in access networks. At the same time applications can have a bigger influence on the packet size distribution in access networks and thus have an impact on the packet loss process of their traffic. Comparing results obtained with the Gilbert model to the analytical results shows that the Gilbert model can best capture the correlations in a queue with a packet size distribution with high CoV, e.g. the exponential distribution. The results presented here can serve as a basis for future research on the performance of end-to-end error control and facilitate the use of FEC in tomorrow's applications.

7 Appendix

The probability $Q_i^{l,m}(k)$ denotes the joint conditional probability that between two arrivals from the joint arrival stream there are k exponential service stage completions out of i and the state of the MMPP at the moment of the next arrival is m given that at the time of the last arrival the MMPP was in state l as defined in Section 3. To calculate $Q_i^{l,m}(k)$ we will first calculate $f_{l,m}(t)$ the joint probability density function of the interarrival-time distribution of the joint arrival process and that the next arrival will occur in state m of the MMPP given that the last arrival was in state l of the MMPP.

Using the notations introduced in Section 3 the Laplace transform $f_{l,m}^*(s)$ of $f_{l,m}(t)$ is given by [33]

$$f_{l,m}^*(s) = \mathcal{L} \left\{ e^{(\hat{Q}-\hat{\Lambda})x} \hat{\Lambda} \right\} = (sI - \hat{Q} + \hat{\Lambda})^{-1} \hat{\Lambda} = \frac{Adj(sI - \hat{Q} + \hat{\Lambda}) \hat{\Lambda}}{det(sI - \hat{Q} + \hat{\Lambda})}, \quad (1)$$

where $Adj(\cdot)$ is the adjoint of (\cdot) . The inverse Laplace transform of (1) can be calculated by finding the roots of $det(sI - \hat{Q} + \hat{\Lambda})$ and performing partial fraction decomposition. If $det(sI - \hat{Q} + \hat{\Lambda})$ has no repeated roots (i.e. $\hat{\lambda}_l \neq \hat{\lambda}_m$ if $l \neq m$), then $f_{l,m}^*(s)$ can be written as

$$f_{l,m}^*(s) = \sum_{j=1}^L \frac{B_j^{l,m}}{s + \beta_j}, \quad (2)$$

where β_j are the roots of $t(s) = \det[sI - \hat{Q} + \hat{\Lambda}]$, i.e. $t(s) = \prod_{j=1}^L (s + \beta_j)$, and $B_j^{l,m}$ are the coefficients resulting from the partial fraction decomposition. The roots of $t(s)$ can be found analytically as long as $L \leq 4$ and numerically for $L > 4$. Performing the inverse Laplace transform we get

$$f_{l,m}(t) = \sum_{j=1}^L B_j^{l,m} e^{-\beta_j t}. \quad (.3)$$

Now we turn to the calculation of $Q_i^{l,m}(k)$. We denote with $P^{l,m}(k)$ the joint probability of having k Poisson arrivals with intensity $r\mu$ between two arrivals from the MMPP and the next arrival from the MMPP coming in state c of the MMPP given that the last arrival came in state b . The z-transform $P^{l,m}(z)$ of $P^{l,m}(k)$ is given by [10]

$$P^{l,m}(z) = \sum_{k=0}^{\infty} \left(\int_0^{\infty} \frac{(r\mu t)^k}{k!} e^{-r\mu t} f_{l,m}(t) dt \right) z^k = f_{l,m}^*(r\mu - r\mu z). \quad (.4)$$

Thus we can express $P^{l,m}(k)$ from (.4) by performing the inverse z-transform after the substitution. Using the notation $\alpha_j = 1 + \beta_j/(r\mu)$ and $A_j^{l,m} = B_j^{l,m}/(r\mu\alpha_j)$ we get

$$P^{l,m}(k) = \sum_{j=1}^L A_j^{l,m} \frac{1}{\alpha_j^k}. \quad (.5)$$

Given the probability $P^{l,m}(k)$ one can express $Q_i^{l,m}(k)$ as

$$\begin{aligned} Q_i^{l,m}(k) &= P^{l,m}(k) & \text{if } k < i \\ Q_i^{l,m}(k) &= \sum_{j=i}^{\infty} P^{l,m}(j) & \text{if } k = i, \end{aligned} \quad (.6)$$

so that we have

$$Q_i^{l,m}(k) = \begin{cases} \sum_{j=1}^L A_j^{l,m} \left(\frac{1}{\alpha_j}\right)^k & 0 \leq k < i \\ \sum_{j=1}^L \frac{A_j^{l,m}}{1 - 1/\alpha_j} \left(\frac{1}{\alpha_j}\right)^i & k = i. \end{cases} \quad (.7)$$

References

- [1] The Multimedia Integrated Conferencing for European Researchers (MICE) Project, "RAT: Robust Audio Tool." <http://www-mice.cs.ucl.ac.uk/multimedia/software/rat>.

- [2] A. V. García and S. Fosse-Parisis, “Freephone audio tool.” <http://www-sop.inria.fr/rodeo/fphone>.
- [3] P. Dube and E. Altman, “Utility analysis of simple FEC schemes for VoIP,” in *Proc. of IFIP Networking 2002*, pp. 226–239, May 2002.
- [4] K. Kawahara, K. Kumazoe, T. Takine, and Y. Oie, “Forward error correction in ATM networks: An analysis of cell loss distribution in a block,” in *Proc. of IEEE INFOCOM*, pp. 1150–1159, June 1994.
- [5] M. Luby, L. Vicisano, J. Gemmell, L. Rizzo, M. Handley, and J. Crowcroft, “The use of forward error correction (FEC) in reliable multicast,” RFC 3453, December 2002.
- [6] E. Biersack, “Performance evaluation of forward error correction in ATM networks,” in *Proc. of ACM SIGCOMM*, pp. 248–257, August 1992.
- [7] G. Dán and V. Fodor, “Quality differentiation with source shaping and forward error correction,” in *Proc. of MIPS’03*, pp. 222–233, November 2003.
- [8] P. Frossard, “FEC performances in multimedia streaming,” *IEEE Commun. Lett.*, vol. 5, no. 3, pp. 122–124, 2001.
- [9] E. Gilbert, “Capacity of a burst-noise channel,” *Bell Syst. Tech. J.*, vol. 69, pp. 1253–1265, September 1960.
- [10] L. Kleinrock, *Queueing Systems*, vol. I. Wiley, New York, 1975.
- [11] E. Altman, C. Barakat, and V. Ramos, “On the utility of FEC mechanisms for audio applications,” in *Proc. of Quality of Future Internet Services, LNCS 2156*, pp. 45–56, 2001.
- [12] J. W. Cohen, *The Single Server Queue*. North-Holland Publishing, Amsterdam, 1969.
- [13] Sprint IP Monitoring Project, “<http://ipmon.sprint.com/>.”
- [14] I. Cidon, A. Khamisy, and M. Sidi, “Analysis of packet loss processes in high speed networks,” *IEEE Trans. Inform. Theory*, vol. 39, no. 1, pp. 98–108, 1993.
- [15] O. Gurewitz, M. Sidi, and I. Cidon, “The ballot theorem strikes again: Packet loss process distribution,” *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2595–2599, 2000.
- [16] E. Altman and A. Jean-Marie, “Loss probabilities for messages with redundant packets feeding a finite buffer,” *IEEE J. Select. Areas Commun.*, vol. 16, no. 5, pp. 779–787, 1998.
- [17] H. Schulzrinne, J. Kurose, and D. Towsley, “Loss correlation for queues with bursty input streams,” in *Proc. of IEEE ICC*, pp. 219–224, 1992.
- [18] G. Dán, V. Fodor, and G. Karlsson, “Analysis of the packet loss process for multimedia traffic,” in *Proc. of the 12th International Conf. on Telecommunication Systems, Modeling and Analysis*, July 2004.

- [19] H. Heffes and D. M. Lucantoni, "A markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE J. Select. Areas Commun.*, vol. 4, no. 6, pp. 856–868, 1986.
- [20] C. Blondia, "The N/G/1 finite capacity queue," *Commun. Statist. - Stochastic Models*, vol. 5, no. 2, pp. 273–294, 1989.
- [21] J. C. Bolot, "End-to-end packet delay and loss behavior in the Internet," in *Proc. of ACM SIGCOMM*, pp. 289–298, September 1993.
- [22] O. J. Boxma, "Sojourn times in cyclic queues - the influence of the slowest server," *Computer Performance and Reliability*, pp. 13–24, 1988.
- [23] J. Beran, R. Sherman, M. Taquq, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1566–1579, 1995.
- [24] S. Robert and J. Le Boudec, "New models for pseudo self-similar traffic," *Performance Evaluation*, vol. 30, no. 1–2, pp. 57–68, 1997.
- [25] A. T. Andersen and B. F. Nielsen, "A markovian approach for modeling packet traffic with long-range dependence," *IEEE J. Select. Areas Commun.*, vol. 16, no. 5, pp. 719–732, 1998.
- [26] T. Yoshihara, S. Kasahara, and Y. Takahashi, "Practical time-scale fitting of self-similar traffic with markov-modulated poisson process," *Telecommunication Systems*, vol. 17, no. 1-2, pp. 185–211, 2001.
- [27] A. Klemm, C. Lindemann, and M. Lohmann, "Modeling IP traffic using the batch Markovian arrival process," *Performance Evaluation*, vol. 54, no. 2, pp. 149–173, 1993.
- [28] 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*, pp. 3–14, 1996.
- [29] P. Skelly, M. Schwartz, and S. Dixit, "A histogram-based model for video traffic behavior in an ATM multiplexer," *IEEE/ACM Trans. Networking*, vol. 1, no. 4, 1993.
- [30] T. Karagiannis, M. Molle, and M. Faloutsos, "A nonstationary poisson view of Internet traffic," in *Proc. of IEEE INFOCOM*, pp. 1–12, March 2004.
- [31] P. Salvador, R. Valadas, and A. Pacheco, "Multiscale fitting procedure using Markov-modulated Poisson processes," *Telecommunication Systems*, vol. 23, no. 1-2, pp. 123–148, 2003.
- [32] R. Gaigalas and I. Kaj, "Convergence of scaled renewal processes and a packet arrival model," *Journal of the Bernoulli Society for Mathematical Statistics and Probability*, vol. 9, no. 4, pp. 671–703, 2003.
- [33] W. Fischer and K. Meier-Hellstern, "The markov-modulated poisson process MMPP cookbook," *Performance Evaluation*, vol. 18, no. 2, pp. 149–171, 1992.

- [34] R. Bellman, *Introduction to Matrix Analysis*. McGraw-Hill, New York, 1960.
- [35] G. Dán, V. Fodor, and G. Karlsson, "Packet size distribution: an aside?," in *Proc. of QoS-IP'05*, pp. 75–87, February 2005.
- [36] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "Internet traffic tends toward poisson and independent as the load increases," in *Nonlinear Estimation and Classification*, Springer, 2002.
- [37] J. C. Bolot, S. Fosse-Parisis, and D. Towsley, "Adaptive FEC-based error control for Internet telephony," in *Proc. of IEEE INFOCOM*, pp. 1453–1460, 1999.
- [38] G. Dán, V. Fodor, and G. Karlsson, "Are multiple descriptions better than one?," in *Proc. of IFIP Networking 2005*, pp. 684–696, May 2005.
- [39] S. Kullback, *Information Theory and Statistics*. Wiley, New York, 1959.
- [40] L. Le Ny and B. Sericola, "Transient analysis of the BMAP/PH/1 queue," *I.J. of Simulation*, vol. 3, no. 3-4, pp. 4–15, 2003.
- [41] W. Whitt, "Approximating a point process by a renewal process, I: Two basic methods," *Operations Research*, vol. 30, no. 1, pp. 125–147, 1982.
- [42] M. Neuts, *Matrix Geometric Solutions in Stochastic Models*. John Hopkins University Press, 1981.
- [43] E. P. C. Kao, "Using state reduction for computing steady state probabilities of queues of GI/PH/1 types," *ORSA J. Comput.*, vol. 3, no. 3, pp. 231–240, 1991.
- [44] O. Ait-Hellal, E. Altman, A. Jean-Marie, and I. A. Kurkova, "On loss probabilities in presence of redundant packets and several traffic sources," *Performance Evaluation*, vol. 36-37, no. 1–4, pp. 485–518, 1999.
- [45] M. Yajnik, S. Moon, J. Kurose, and D. Towsley, "Measurement and modelling of the temporal dependence in packet loss," in *IEEE INFOCOM*, pp. 345–352, 1999.
- [46] W. Jiang and H. Schulzrinne, "Modeling of packet loss and delay and their effect on real-time multimedia service quality," in *Proc. of NOSSDAV*, 2000.
- [47] P. Frossard and O. Verscheure, "Joint source/FEC rate selection for quality-optimal MPEG-2 video delivery," *IEEE Trans. Image Processing*, vol. 10, no. 12, pp. 1815–1825, 2001.
- [48] P. Billingsley, *Statistical Inference for Markov Processes*. Chicago University Press, 1961.
- [49] E. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, September 1963.