The Gilbert-Elliott Model for Packet Loss in Real Time Services on the Internet

Gerhard Haßlinger and Oliver Hohlfeld *

T-Systems, Deutsche-Telekom-Allee 7, D-64295 Darmstadt, Germany
{gerhard.hasslinger,oliver.hohlfeld}@t-systems.com

Abstract. The estimation of quality for real time services over telecommunication networks requires realistic models in order to generate impairments and failures during transmission. Starting with the classical Gilbert-Elliot model, we derive the second order statistics over arbitrary time scales and fit the parameters to match the packet loss pattern of traffic traces. The results show that simple Markov models are appropriate to capture the observed loss pattern.

1 Introduction

The transfer of real time data over the Internet and channels in heterogeneous packet networks is subject to errors of various types. A packet can be corrupted – and therefore is unusable for a voice or video decoder – due to unrecoverable bit failures. On wireless and mobile links temporary and long lasting reductions in the available capacity frequently occur and even in fixed and wired network sectors packets may be dropped at routers and switches in phases of overload. Most of the Internet traffic is controlled by the TCP protocol, which provides mechanisms for retransmission of lost or corrupted data and for controlling the load on highly loaded links involving FIFO queues with a Tail-Drop or Random Early Detection (RED) [1] policy.

On the other hand, the portion of uncontrolled traffic via the UDP transport protocol has been increasing to a level of 5 - 10% in recent time [2], partly since real time services over IP including voice, video on demand and online gaming are gaining in popularity. The upcoming deployment of IP-TV over VDSL broadband access platforms by the Deutsche Telekom and other Internet service providers will strengthen this trend.

In this work we focus on packet loss on Internet links where most of the traffic is controlled by TCP but with an essential contribution of real time traffic without flow control. Under sufficiently high link load, this causes spontaneous overload peaks causing packet loss. Available traffic traces [2] show, that UDP traffic has a higher variability in the relevant time scales than the total traffic, which at the present stage is dominated by peer-to-peer data exchange.

* The Master's thesis of Oliver Hohlfeld was supported by the project T-V-Model of the Deutsche-Telekom-Laboratories (T-Labs) Berlin. We also thank Rüdiger Geib for the support during the thesis.

In Section 2 we characterise the packet loss pattern observed in traffic traces based on the second order statistics, i.e. the coefficient of variation, in multiple time scales. For individual UDP traffic flows the relevant time scale is extended in comparison to the the complete traffic, since the mean packet interarrival times differ by the ratio of the traffic rate of the flow and the total traffic.

We consider simple Markov processes to be fitted to the observed second order statistics. For the classical 2-state case introduced by Gilbert [3] and Elliot [4] we experienced that the derivation of the second order statistics as given in Section 4 is non-trivial. Despite of the wide spread use of the Gilbert-Elliot model to describe error burst processes and fitting processes with regard to specific short term error pattern, we could not find results and fitting procedures for the second order statistics over longer time frames in the literature. Section 3 summarises classical fitting schemes for the Gilbert-Elliot model and Section 6 considers related work.

In Section 5, a comparison of the model with adapted parameters to the packet loss pattern derived from traffic traces shows that simple Markov processes achieve a fairly close fit to the mean and variances over multiple time scales.

2 Packet Loss Process in Data Transfer over Multiple Time Scales

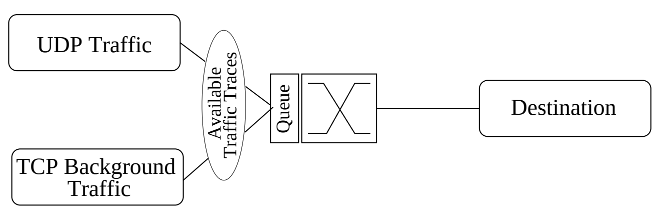


Fig. 1. Measurement topology of the used network; TCP backbone traffic is feed from a trace file along with UDP traffic into a router. The traffic is directed over an bottlenecked link to a destination. The loss rate can be arbitrarily chosen by adjusting the capacity of the outgoing, bottlenecked link.

We consider a scenario of controlled TCP packet flows being superposed with real time traffic over the UDP protocol, which does not provide error recovery and flow control mechanisms. We refer to measurement traces of traffic taken from a 2.5Gb/s interface of a broadband access router of Deutsche Telekom's IP platform, which connects residential ADSL access lines to the backbone. Based on the time stamp and the size of each packet, the variability of the traffic can be observed in any time scale from the accuracy level of the time stamps, from fairly below 1ms up to the 30 minutes length of the traces.

Let Δ be a time frame in this range. Then corresponding traffic rates $R_k^{(\Delta)}$ are determined for successive intervals of length Δ by dividing the sum of the size of all packets arriving in a time interval by its length. From the sequence $R_k^{(\Delta)}$ the mean rate μ_R and the variance $\sigma_R^2(\Delta)$ is computed. The second order statistics is given in this way considering $\sigma_R^2(\Delta)$ over a relevant range of Δ. This statistics is a standard description method for traffic and is equivalent to the autocorrelation function over the considered time scales. Self-similar traffic patterns are usually defined by a specific autocorrelation function, attributed as exactly second order self-similarity in [5].

Table 1 shows the second order statistics for Δ = 1ms, 10ms, 100ms, 1s and 10s for the UDP and the total traffic. The coefficients of variation $c_v(\Delta) = \sigma(\Delta)/\mu$ are observed to be about twice as high for UDP as for the total traffic.

	Mean Rate	$c_v(1ms)$	$c_v(10ms)$	$c_v(100ms)$	$c_v(1s)$	$c_v(10s)$
UDP traffic	$\mu = 50.8Mb/s$	0.3209	0.1220	0.0531	0.0433	0.0394
Total traffic	$\mu = 753.9Mb/s$	0.1689	0.0635	0.0322	0.0259	0.0216

Table 1. Second order statistics for Δ = 1ms, 10ms, 100ms, 1s and 10s for the UDP and the total traffic.

We also adhere to the second order statistics for describing the packet loss process. The traffic traces are at a load level of about 30% and originally do not exhibit packet losses in the considered time scales. However, at higher load, i.e. for reduced capacity < 2.5Gb/s, overload phases occur above some medium load level and we can easily compute the resulting packet loss process corresponding to the trace at any sufficiently high load level. In general, the loss pattern is evaluated for a predefined capacity C (versus load) including a buffer of limited size B, assuming that an arriving packet is lost by tail drop each time when it does not fit into the remaining buffer. The loss pattern obtained in this way are adequate for uncontrolled UDP traffic, but do not regard the TCP retransmission and source rate adaptation. However, the TCP control does not respond on the 1ms, but on essentially larger time scales. We assume that TCP will establish a stabilised non-excessive load level without much data loss and will focus on the UDP traffic portion with regard to TCP background traffic.

Following the previous approach, we obtain the packet loss process from the traces at a predefined load level and calculate its second order statistics. Since the loss rate is monotonously increasing with the load, we can adjust the load in order to approach a considered packet loss rate.

Next, we study simple Markov models again with focus on their second order statistics. The aim is to provide a generator for packet loss pattern to be used in the estimation of the degradation in the Quality of Experience (QoE) for Internet services.

3 Gilbert-Elliot: The Classical 2-State Markov Model for Error Processes

We consider the 2-state Markov approach as introduced by Gilbert [3] and Elliot [4], which is widely used for describing error patterns in transmission channels [6], [7], [8], [9], [10], [11], [12], [13], [14] and for analysing the efficiency of coding for error detection and correction [15]. We follow the usual notation of a good (G) and bad (B) state, each of which may generate errors as independent events at a state dependent error rate $1-k$ in the good and $1-h$ in the bad state, respectively. The model is shown in Figure 2. For applications in data loss processes, we interpret an event as the arrival of a packet and an error as a packet loss. The transition matrix A is given by the two transitions

$$p = P(q_t = B|q_{t-1} = G); \quad r = P(q_t = G|q_{t-1} = B); \quad A = \begin{pmatrix} 1-p & p \\ r & 1-r \end{pmatrix}, \tag{1}$$

where q_t denotes the state at time t.

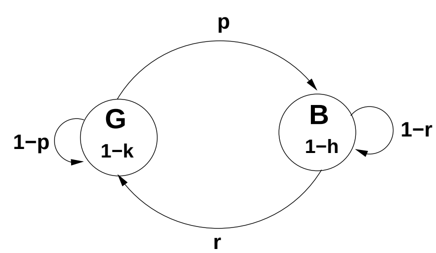


Fig. 2. The Gilbert-Elliott model generating a 2-state Markov modulated failure process.

The stationary state probabilities π_G and π_B exist for $0 < p, r < 1$ [15]

$$\pi_G = \frac{r}{p+r}, \quad \pi_B = \frac{p}{p+r} \tag{2}$$

and the model error rate directly follows from (2)

$$p_{Error} = (1-k)\pi_G + (1-h)\pi_B. \tag{3}$$

In 1960, Gilbert [3] proposed a model to characterise a burst-noise channel. It adds memory to the Binary Symmetric Channel coded into the states of the Markov chain. Gilbert defined the good state to be error-free ($k = 1$). Errors are only generated in the bad state with a probability of $1-h$.

Gilbert suggested to estimate the free model parameters from three measurable instances a, b and c of the packet drop trace $\{E_t\}_{t\in\mathbb{N}}$, where $E_t = 1$ if the packet is dropped and 0 otherwise, and only requires trigram statistics

$$a = P(1), \quad b = P(1|1), \quad c = \frac{P(111)}{P(101) + P(111)}. \tag{4}$$

By knowing a, b and c, the three model parameters can be computed in the following manner

$$1 - r = \frac{ac - b^2}{2ac - b(a + c)}; \quad h = 1 - \frac{b}{1 - r}; \quad p = \frac{ar}{1 - h - a}. \tag{5}$$

Gilbert argues that the c measurement may be avoided by choosing $h = 0.5$ and using $1 - r = 2b$. Furthermore, he showed that the method introduced above can lead to ridiculous parameters ($p, r, h < 0$, or $p, r, h > 1$), if the observation (the trace) is too small. Morgera et al. [16] also conclude that the method proposed by Gilbert is more appropriate for longer traces. In case of shorter observations, better results can be obtained when considering the Gilbert model as Hidden Markov Model and train it by applying the Baum-Welch algorithm [17], [18], [19].

Parameters of an even simplified Gilbert model with $h = 0$ can also be estimated with the method presented by Yajnik et al. [7]

$$p = P(1|0); \quad r = P(0|1). \tag{6}$$

from the frequency of a packet drop being encountered after a correct transmission and vice versa.

A more intuitive parameter estimation technique can be found by considering the Average Burst Error Length (ABEL) and the model error rate (e.g. the average number of packet drops); Equation (3) leads to $p = p_E \cdot r/(h - p_E)$ and r can be set by considering the mean of the geometric distribution of packet loss burst lengths $r = 1/\text{ABEL}$.

Gilbert's model was extended by Elliott in 1963. Elliott's model [4] allows errors to be generated in both states. Table 2 gives an overview of possible simplifications in different 2-state models.

Model	Parameter	Training Complexity	Simplification
Simple Gilbert	p, r	simple	$k = 1$, $h \in \{0, 0.5\}$
Gilbert	p, r, h	medium	$k = 1$
Gilbert-Elliott	p, r, h, k	high	/

Table 2. Comparison of the presented two-state Markov based channel models.

4 Gilbert-Elliot model: Variance of the Error Process over Multiple Time Scales

The second order statistics of the 2-state Markov process can be derived via generating functions. While it is straightforward to compute the distribution function of errors in time frames of length $N+1$ iteratively from the result for length N, explicit expressions are becoming complex already for the 2-state Markov model [20]. To the authors knowledge, explicit expressions for the variance of the number of errors during a time frame of fixed length, i.e. a fixed number of events, are not given in the literature, although there is a large volume of work involving the Gilbert-Elliott model, as partly discussed in the section on related work. However, most of this work is devoted to error detecting and correcting codes and the residual error probabilities of coding schemes, rather than on traffic or packet loss characterisation, where the second order statistics in multiple time scales is a standard description approach [5].

Although Markov models do not exhibit self-similar properties, they have been successfully adapted to self-similar traffic [21] and are still popular since they often lead to simple analytical results. Following this trend, we next derive the variance of the number of errors (i.e. packet drops) for the 2-state Gilbert-Elliott model in time frames with a fixed number of events.

4.1 Generating Functions

Let $G_N(z)$ ($B_N(z)$) denote the generating function for the number of packet drops in a sequence of N packet arrivals, leaving the Markov chain in the last step at state G (B). Iterative relationships can be set up to extend the analysis to longer packet sequences $G_N(z) \rightarrow G_{N+1}(z)$, where state transitions are considered and a possible packet drop at the last arrival is taken into account by a factor $(k+(1-k)z)$ or $(h+(1-h)z)$:

$$G_{N+1}(z) = (1-p)(k+(1-k)z)G_N(z) + r(h+(1-h)z)B_N(z) \qquad (7)$$

$$B_{N+1}(z) = p(k+(1-k)z)G_N(z) + (1-r)(h+(1-h)z)B_N(z) \qquad (8)$$

The above expressions will be initialised by assuming steady state conditions as the starting point:

$$G_0(z) = \frac{r}{p+r}; \quad B_0(z) = \frac{p}{p+r}. \qquad (9)$$

Note that $G_0(z)$ and $B_0(z)$ represent defective distributions, which also holds for $G_N(z)$, $B_N(z)$ $\forall N \in \mathbb{N}$. We finally evaluate complete distributions given by $G_N(z) + B_N(z)$ for the number of packet drops without restriction to a final state, where $G_N(1) + B_N(1) = 1$.

The k-th moment can be derived from the generating function by considering the k-th derivative: $E[X^k] = \left(z\frac{d}{dz}\right)^i G(z)|_{z=1}$ [22], [23]. The first two moments

$E(X), E(X^2)$ are sufficient to derive the second order statistics involving the first and second derivative of the generating functions.

$$\begin{aligned} G'_{N+1}(z) = (1-p) \cdot ((1-k) \cdot G_N(z) + (k + (1-k)z) \cdot G'_N(z)) \\ +r \cdot ((1-h) \cdot B_N(z) + (h + (1-h)z) \cdot B'_N(z)) \end{aligned} \tag{10}$$

$$\begin{aligned} G''_{N+1}(z) = (1-p) \cdot (2(1-k) \cdot G'_N(z) + (k + (1-k)z) \cdot G''_N(z)) \\ +r \cdot (2(1-h) \cdot B'_N(z) + (h + (1-h)z) \cdot B''_N(z)) \end{aligned} \tag{11}$$

$$\begin{aligned} B'_{N+1}(z) = p \cdot ((1-k) \cdot G_N(z) + (k + (1-k)z) \cdot G'_N(z)) \\ +(1-r) \cdot ((1-h) \cdot B_N(z) + (h + (1-h)z) \cdot B'_N(z)) \end{aligned} \tag{12}$$

$$\begin{aligned} B''_{N+1}(z) = p \cdot (2(1-k) \cdot G'_N(z) + (k + (1-k)z) \cdot G''_N(z)) \\ +(1-r) \cdot (2(1-h) \cdot B'_N(z) + (h + (1-h)z) \cdot B''_N(z)) \end{aligned} \tag{13}$$

4.2 Mean Values

The mean values are given by $\mu_N^G = G'_N(1)$ and $\mu_N^B = B'_N(1)$, which leads to the following expressions

$$\mu_{N+1}^G = (1-p) \cdot \left(\frac{(1-k)r}{p+r} + \mu_N^G \right) + r \cdot \left(\frac{(1-h)p}{p+r} + \mu_N^B \right), \tag{14}$$

$$\mu_{N+1}^B = p \cdot \left(\frac{(1-k)r}{p+r} + \mu_N^G \right) + (1-r) \cdot \left(\frac{(1-h)p}{p+r} + \mu_N^B \right). \tag{15}$$

Considering the sum of μ_{N+1}^G and μ_{N+1}^B leads to the expected result of $N+1$ times the mean failure rate in the steady state:

$$\mu_{N+1} = \mu_{N+1}^G + \mu_{N+1}^B = (N+1) \left(\frac{(1-k)r}{p+r} + \frac{(1-h)p}{p+r} \right). \tag{16}$$

To eliminate the reference to the opposite term, μ_{N+1}^B can be rewritten as

$$\begin{aligned} \mu_{N+1}^B &= \frac{pr(1-k)}{p+r} + p \cdot \left(\underbrace{\mu_N^G + \mu_N^B}_{(16)} - \mu_N^B \right) + (1-r) \cdot \left(\frac{p(1-h)}{p+r} + \mu_N^B \right) \\ &= (N+1) \cdot \left(\frac{pr(1-k)}{p+r} + \frac{p^2(1-h)}{p+r} \right) + [1-(p+r)] \cdot \left(\frac{p(1-h)}{p+r} + \mu_N^B \right). \end{aligned} \tag{17}$$

The structure of the above equation is simplified by contracting terms independent of N:

$$\mu_{N+1}^B = (N+1) \cdot \beta_B + \alpha \cdot (\mu_N^B + \gamma_B); \quad \alpha := 1-(p+r); \tag{18}$$

$$\beta_B := \frac{pr(1-k)}{p+r} + \frac{p^2(1-h)}{p+r}; \quad \gamma_B := \frac{(1-h)p}{p+r}. \tag{19}$$

Due to the symmetry of both states G and B, $G_N(z)$ can be obtained from $B_N(z)$ by swapping the parameters $p \leftrightarrow r$ and $h \leftrightarrow k$ and vice versa. Thus, $G_N(p,r,h,k,z) = B_N(r,p,k,h,z)$ and $\mu_N^G(p,r,h,k) = \mu_B^N(r,p,k,h)$.

The recursive relationship (18) is transformed into a direct analytic expression by considering solutions for μ_1^s, μ_2^s, $\cdots$ where $s \in \{G,B\}$ and β_G, γ_G are again determined from β_B, γ_B exploiting the symmetry of the model.

$$\mu_0^s = 0; \quad \mu_1^s = \beta_s + \alpha\gamma_s; \quad \mu_2^s = 2\beta_s + \alpha(\beta_s + \alpha\gamma_s + \gamma_s) = 2\beta_s + \alpha(\beta_s + \gamma_s) + \alpha^2\gamma_s. \tag{20}$$

This suggests the general result, which is proven by induction over N:

$$\begin{aligned}
\mu_N^s &= -\gamma_s + \sum_{j=0}^{N} \alpha^j(\gamma_s + \beta_s\gamma_s(N-j)); \quad s \in \{G,B\} \\
&= \beta_s \frac{N}{1-\alpha} + \left(\gamma_s - \frac{\beta_s}{1-\alpha}\right) \frac{\alpha}{1-\alpha}\left(1-\alpha^N\right), \quad s \in \{G,B\}, \text{for } \alpha \neq 1
\end{aligned} \tag{21}$$

The case $\alpha = 1$, which means $p = r = 0$ implies a reducible and thus non-ergodic Markov chain, which is not relevant for modelling purposes. Considering the sum $\mu_N^B + \mu_N^G$ again leads to Equation (16).

4.3 Explicit Solution for the Variance

Using the mean values, the variance of the number of packet losses in a time frame of size N can be derived as follows

$$\begin{aligned}
G''_{N+1}(1) + B''_{N+1}(1) &= 2(1-k)\mu_N^G + 2(1-h)\mu_N^B + G''_N(1) + B''_N(1) \\
&= 2(1-k)\sum_{i=1}^{N} \mu_i^G + 2(1-h)\sum_{i=1}^{N} \mu_i^B.
\end{aligned} \tag{22}$$

The sum of the mean values yields

$$\begin{aligned}
\sum_{i=1}^{N} \mu_i^s &= \beta_s \frac{i}{1-\alpha} + \left(\gamma_s - \frac{\beta_s}{1-\alpha}\right)\frac{\alpha}{1-\alpha}\left(1-\alpha^i\right), \quad s \in \{G,B\} \\
&= -\beta_s \frac{N(N+1)}{2(1-\alpha)} + \frac{N\alpha}{1-\alpha}\left(\gamma_s - \frac{\beta_s}{1-\alpha}\right) - \left(\gamma_s - \frac{\beta_s}{1-\alpha}\right)\frac{\alpha}{1-\alpha}\frac{\alpha - \alpha^{N+1}}{1-\alpha} \\
&= -\beta_s \frac{N(N+1)}{2(1-\alpha)} + \left(\gamma_s - \frac{\beta_s}{1-\alpha}\right)\frac{\alpha}{1-\alpha}\left(N - \frac{\alpha(1-\alpha^N)}{1-\alpha}\right).
\end{aligned} \tag{23}$$

The obtained explicit solution for $G''_{N+1}(1) + B''_{N+1}(1)$ leads to a term for the variance and the standard deviation $\sigma_N = \sigma_N^B + \sigma_N^G$ using the fact that $G''_N(1) + B''_N(1) = \mu_N^2 + \sigma_N^2 - \mu_N$, as well as for the coefficient of variation $c_v(N) = \sigma_N/\mu_N$.

After putting the equations of the subsection together and reordering and simplifying their components we finally have the following result for the coefficient of variation for the number of errors or packet losses in a sequence of length N generated by the Gilbert-Elliot model:

$$
\begin{aligned}
c_v(N) &= \frac{\sqrt{G_N''(1) + B_N''(1) - \mu_N^2 + \mu_N}}{\mu_N}; \qquad \omega := (1-h)p + (1-k)r \\
&= \frac{1}{\sqrt{N}} \sqrt{\frac{hp+kr}{\omega} + \frac{2pr(1-p-r)(h-k)^2}{\omega^2(p+r)} \left(1 - \frac{1-(1-p-r)^N}{N(p+r)}\right)}.
\end{aligned}
\tag{24}
$$

The solution is comprehensible enough to interpret the influence of the model parameters. Note that the evaluation of the term $1-(1-p-r)^N$ may cause numerical instability for small $p,\ r$, which can be improved by implementing the equivalent form $1-(1-p-r)^N = 1 - e^{\ln(1-p-r)\cdot N}$.

4.4 Simple Cases

In case of $h = k$, both states are indistinguishable and the Markov chain collapses to a single state leading to the simplified result

$$
c_v(N) = \sqrt{\frac{h}{(1-h)N}}. \tag{25}
$$

This corresponds to a binomial distribution $G_N(z) + B_N(z) = [h + (1-h)z]^N$ of independent random packet losses generated by a memoryless process.

If $p + r = 1$, the Markov chain again generates a memoryless process, since the transition probabilities, e.g. to state B, are the same starting from B or G:

$$
P(q_t = B | q_{t-1} = G) = p; \quad P(q_t = B | q_{t-1} = B) = 1 - r = p. \tag{26}
$$

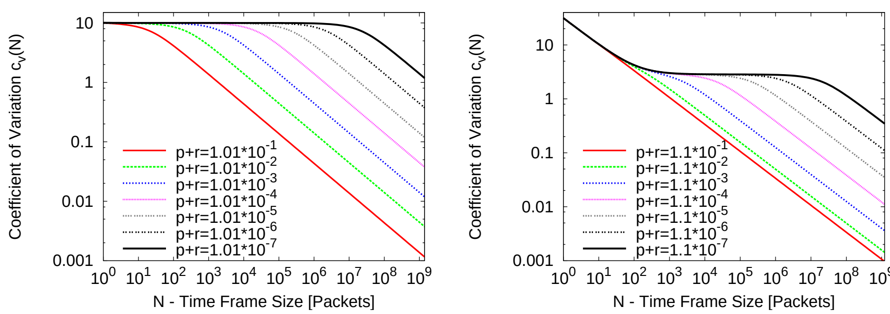
Again, the coefficient of variation is simplified:

$$
c_v(N) = \sqrt{\frac{1}{N} \frac{hp+kr}{(1-h)p + (1-k)r}}. \tag{27}
$$

4.5 Parameter Impact on the Second Order Statistics of the Gilbert-Elliot Model

Based on the analytical result in Equation (24) for $c_v(N)$, further properties of the second order statistics of the Gilbert-Elliot model are immediately visible:

1. Regarding a single packet arrival we obtain $c_v(1) = \sqrt{(hp+kr)/\omega}$ from Equation (24). This expression can be simplified to $c_v(1) = \sqrt{1/p_{Error} - 1}$ using Equation (3). Therefore, the starting point of the curves (cf. Figure 3 and 4) is determined by the mean packet loss rate p_{Error} (3).



(a) $h = 0$, $k = 1$ (b) $h = 0.99$, $k = 0.9999$

Fig. 3. Parameter impact on the second order statistics of the Gilbert-Elliot model.

2. The examples in Figure 3(a) for $h = 0$, i.e. for bursts of subsequent losses in the bad state, are characterised by a first part, which holds the variance on the initial $c_v(1)$ value. The length of the part at constant level is extending with $\lim p + r \;\rightarrow\; 0$, when the ratio p/r and thus the loss rate is preserved at the same time. In particular, the duration of the bad state is geometrical with mean $(1-r)/r$, such that a reduction of r by some factor prolonges the bad state sojourn time by the same factor. As a consequence, the correlation in the modeling process persists over the same time scale.
In this way, the transition point T from the constant level part for $t < T$ into the decreasing part for $t > T$ mainly depends on and thus can be fixed by choice of $p + r$. The decreasing part soon approaches the same slope as is valid for a memoryless process with independent random losses at a given rate, such that $c_v(kN)/c_v(N) \rightarrow \sqrt{k}$.

The precondition $p + r \ll 1/N$ generally simplifies the coefficient of variation to the form:

$$c_v(N) = \sqrt{\frac{hp + kr}{N\omega} + \frac{(N-1)pr(h-k)^2}{N\omega^2}}; \qquad \omega := (1-h)p + (1-k)r.$$

Figure 3(b) shows results, where $p + r$ is reduced by a factor 10 per step as in 3(a), but this time with $h = 0.99$, which means a low loss probability of 1% in the bad state. Again the coefficient of variation can be held at a constant level over multiple time scales for small $p + r$, but this happens essentially below the initial $c_v(1)$ value. In particular, the constant level is reached at $c_v(1)/\sqrt{1-h}$.

5 Evaluation

The evaluation of the trained 2-state Markov models using the coefficient of variation $c_v(N) = \sigma_N/\mu_N$ for two backbone traces with different packet loss rates is shown in Figure 4 and 5. The Poisson process with its assumption of identical and independent distributed packet drops provides a simple lower bound of the coefficients in this evaluation. The model parameters of the simple Gilbert ($h = 0$, $k = 1$) and the Gilbert model have been estimated from the given traces using the traditional methods proposed by Yajnik et al. [7] and Gilbert [3] as introduced in Section 3.

Moreover, the simplified Gilbert, the Gilbert and the Gilbert-Elliott model have been trained by the method introduced in Section 4 over multiple timescales $N \in [1, 10^5]$, as shown in Figure 4 and 5. The model parameters were estimated by fitting the coefficient of variation curve to the one obtained from the corresponding trace using the Levenberg-Marquardt algorithm for numeric optimisation of non-linear functions. Initial trial values for the parameters were estimated from the study of the impact of different model parameters discussed in Section 4.5. In order to avoid confusion with the classical techniques presented in Section 3, results obtained by adapting the model parameters using the coefficient of variation are labelled with H&H in the figures and tables.

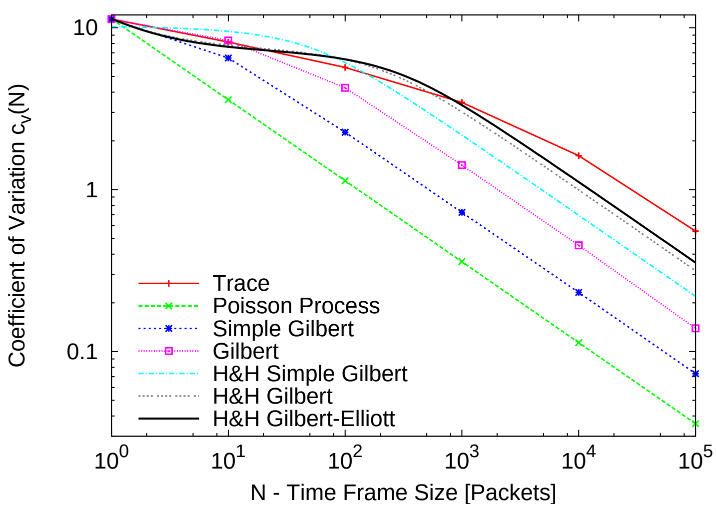


Fig. 4. Evaluation of the trained 2-state Markov models using the coefficient of variation $c_v = \sigma/\mu$ for a backbone trace with a mean packet loss rate of 0.7%.

The distance between different model curves as shown in Figure 4 and 5 and the trace curve is measured by the the Mean Square Error MSE defined by

$$\mathrm{MSE(model)} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{model}_i - \mathrm{trace}_i)^2 \tag{28}$$

A smaller MSE indicates a better fit of the model. The results are shown in Table 4.

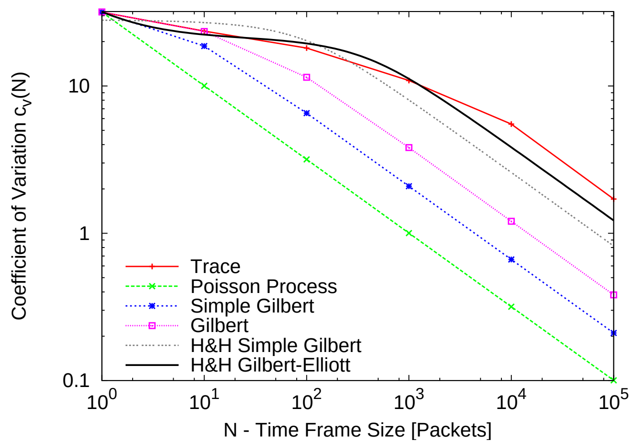


Fig. 5. Evaluation of the trained 2-state Markov models using the coefficient of variation $c_v = \sigma/\mu$ for a backbone trace with a mean packet loss rate of 0.1%.

Trace	Trained Model	p	r	h	k	p_{Error}
A – 0.7 %	H&H Simple Gilbert	0.000401648	0.0414789	0	1	0.95%
	H&H Gilbert	0.000196854	0.0109547	0.563513	1	0.77%
	H&H Elliott	0.000132253	0.00811837	0.559691	0.999372	0.71%
A – 0.1 %	H&H Simple Gilbert	$2.9365 \cdot 10^{-5}$	0.0229754	0	1	0.13%
	H&H Gilbert	$1.3343 \cdot 10^{-5}$	0.00601795	0.555044	1	0.098%
	H&H Elliott	$1.33308 \cdot 10^{-5}$	0.00601795	0.554949	0.999999	0.098%

Table 3. Estimated model parameters for both traces using the proposed moment fitting technique (H&H) and the resulting model mean loss rates. The mean packet loss rate of the first trace is 0.7% and 0.1% in case of the second.

The model parameters resulting from the adaption to the coefficient of variation found in the trace are given in Table 3. This trend confirms the assumption by Gilbert [3] of assigning $h = 0.5$. Furthermore, the model mean packet loss rate can be computed using (3) and matches the mean loss rate of the traces for the Gilbert and the Gilbert-Elliott model.

However, when we look at the distribution of the length of packet losses in a consecutive series, then the classical fitting procedures seem to be in favour, as experienced from first evaluations. This is not unexpected, since they are closer related to error burst lengths whereas the second order statistics is composed of any type of error pattern. The extraction of the most relevant information in measurement traces to be used for the fitting of model parameters with regard to the Quality of Experience aspects (QoE) is still for further study. The relevance of considering bursts surely increases with the observed mean failure burst length observed in the considered traffic flows.

Trace / Model	Simple Gilbert	Gilbert	H&H Simple Gilbert	H&H Gilbert	H&H Elliott
0.7%	4.02	1.3	0.98	0.2	0.18
0.1%	43.42	19.04	7.96	1.09	1.09

Table 4. Mean Square Error (MSE) distance between different trained models and the two traces.

6 Related Work

Cuperman [20] proposes a generating function (29) to calculate the probability of m errors in a code word of length n, $P(m, n)$.

$$H_m(z) = \sum_{n=m}^{\infty} P(m,n) z^n = \frac{p}{p+r} z \left(\frac{1-g(z)}{1-z} \right)^2 [g(z)]^{m-1} \tag{29}$$

$$g(z) = \sum_{k=1}^{\infty} P(0^{k-1}1/1) z^k; \quad 1 \le m \le n$$

However, this generating function cannot be used to derive the second order statistics of the Gilbert-Elliott model, as it is designed to count over infinite window sizes.

Girod et al. [6] found a simplified Gilbert model ($k = 1$, $h = 0$) useful to describe the characteristics of packet losses in Internet connections and to derive an error model for Internet video transmissions on top, as lost packets will affect the perceived quality of the video transmission. Huitika et al. [24] extended the simplified Gilbert model by adapting it to the datagram loss process in the scope of real-time video transmissions, by introducing a third state to describe out-of-order packets. Zhang et al. [8] use a simplified Gilbert model to describe a cell discard model for MPEG video transmissions in ATM networks, where the cell losses are caused by excessive load at ATM multiplexers.

McDougall et al. [25] proposed a 4-state Markov model with a hypergeometrical distribution of the sojourn time in the good and bad state as approximation of an IEEE 802.11 channel. Poikonen et al. [13] [14] compared finite state Markov models, such as the McDougall model, in order to simulate the packet error behaviour of an DVB-H system. The McDougall model and the Markov-based Trace Analysis (MTA) [26] outperformed the Gilbert model, as it was unable to reproduce the variance in burst error lengths. Yajnik et al. [7] point out that the simplified Markov model is suitable if the error gap length of the traces is geometrically distributed, but can be outperformed by considering high-order Markov chains.

Tang et al. [11] used a simplified Gilbert model to create a multicast loss model in IEEE 802.11 channels. Hartwell et al. [12] compared five finite-state Markov models to create a frame loss model for IEEE 802.11 indoor networks and found out that high order models trained by the Baum-Welch algorithm outperformed the Gilbert model. McDougall et al. [10] were able to reproduce the packet error rate and the average burst error length of an IEEE 802.11 channel using the simplified Gilbert model, but failed to replicate the variance in error burst lengths and therefore suggested to use Gamma based state durations, as in [27]. McDougall et al. [10] also suggest that the restriction of geometrically distributed state lengths due to constant transition probabilities in the Gilbert-Elliott model can be overcome and i.e. the Gamma distribution can be used.

7 Conclusion

This work provides a method to adapt the parameter set of a 2-state Markovian error pattern generator to match the second order statistics over multiple time scales. The generating functions approach provides recursive relationships for the distribution of the number of lost packets, which finally leads to an explicit and clearly structured solution for the second order statistics. Special cases of the model as well as the impact of its parameters are discussed. Naturally, fitting procedures based on second order statistics yield a closer match in multiple time scales than classical adaptation schemes, which on the other hand are better in modeling error bursts. Thus it remains partly open, which statistical indicators should be involved in the fitting procedure. However, the proposed approach gives more flexibility to include information from different time scales as is often required for traffic characterization. Several Markov approaches have been proposed providing more states and parameters, which improve the accuracy of the fit to the observed process characteristics on account of more complex adaptation schemes.

References

1. Floyd, S., Jacobson, V.: Random early detection gateways for congestion avoidance. IEEE/ACM Trans. Netw. **1** (1993) 397–413
2. Haßlinger, G., Mende, J., Geib, R., Beckhaus, T., Hartleb, F.: Measurement and characteristics of aggregated traffic in broadband access networks. In Mason, L., Drwiega, T., J.Yan, eds.: 20th International Teletraffic Congress (ITC20), Ottawa, Canada, Springer LNCS 4516 (2007) 998–1010
3. Gilbert, E.N.: Capacity of a Burst-Noise Channel. Bell System Technical Journal **39** (1960) 1253–1265
4. Elliott, E.O.: Estimates of Error Rates for Codes on Burst-Noise Channels. Bell System Technical Journal **42** (1963) 1977–1997
5. Leland, W.E., Taqqu, M.S., Willinger, W., Wilson, D.V.: On the self-similar nature of ethernet traffic. ACM SIGCOMM Computer Communication Review **23** (1993) 183–193
6. Girod, B., Stuhlmüller, K., Link, M., Horn, U.: Packet Loss Resilient Internet Video Streaming. In: Proceedings of SPIE Visual Communications and Image Processing. (1999) 833–844
7. Yajnik, M., Moon, S.B., Kurose, J.F., Towsley, D.F.: Measurement and modeling of the temporal dependence in packet loss. In: IEEE INFOCOM. (1999) 345–352

8. Zhang, L., Chow, D., Ng, C.H.: Cell loss effect on QoS for MPEG video transmission in ATM networks. In: IEEE International Communication Conference (ICC). (1999) 147–151
9. Lemmon, J.J.: Wireless link statistical bit error model. US National Telecommunications and Information Administration (NTIA) Report 02-394 (2002)
10. McDougall, J., Miller, S.: Sensitivity of wireless network simulations to a two-state Markov model channel approximation. In: IEEE Globecom. (2003)
11. Tang, C., McKinley, P.K.: Modeling multicast packet losses in wireless LANs. In: 6th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems (MSWIM), New York, NY, USA, ACM Press (2003) 130–133
12. Hartwell, J.A., Fapojuwo, A.O.: Modeling and characterization of frame loss process in IEEE 802.11 wireless local area networks. In: IEEE Vehicular Technology Conference. (2004) 4481–4485
13. Poikonen, J., Paavola, J.: Comparison of Finite-State Models for Simulating the DVB-H Link Layer Performance. In: 2nd International Symposium on Wireless Communication Systems (IEEE ISWCS). (2005)
14. Poikonen, J., Paavola, J.: Error Models for the Transport Stream Packet Channel in the DVB-H Link Layer. In: IEEE International Conference on Communications (ICC). (2006) 1861–1866
15. Osmann, C.: Bewertung von Codierverfahren für einen störungssicheren Datentransfer. PhD thesis, Universität Duisburg-Essen (1999)
16. Morgera, S.D., Simard, F.: Parameter estimation for a burst-noise channel. In: International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Washington, DC, USA (1991) 1701–1704
17. Turin, W.: Performance Analysis and Modeling of Digital Transmission Systems (Information Technology: Transmission, Processing and Storage). Springer New York, Inc., Secaucus, NJ, USA (2004)
18. Sivaprakasam, S., Shanmugan, K.: An Equivalent Markov Model for Burst Errors in Digital Channels. IEEE Transactions on Communications **43** (1995) 1347–1355
19. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. (1990) 267–296
20. Cuperman, V.: An upper bound for the error probability on the gilbert channel. IEEE Transactions on Communication Technology **17** (1969) 532–535
21. Feldmann, A., Whitt, W.: Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. Performance Evaluation **31** (1998) 245–279
22. Wilf, H.S.: Generatingfunctionology. Academic Press (1992)
23. Grinstead, C.M., Snell, L.J.: Introduction to Probability. American Mathematical Society (1997)
24. Huitika, T., Driessen, P.: Datagram loss model for noninteractive real time streaming video. In: IEEE Pacific Rim Conference on Communications. Computers and Signal Processing (PACRIM). (2003) 756–759
25. McDougall, J., John, J.J., Yu, Y., Miller, S.L.: An improved channel model for mobile and ad-hoc network simulations. In Hamza, M.H., ed.: Communications, Internet, and Information Technology, IASTED/ACTA Press (2004) 352–357
26. Konrad, A., Zhao, B.Y., Joseph, A.D., Ludwig, R.: A Markov-based Channel Model Algorithm for Wireless Networks. Wireless Networks **9** (2003) 189–199
27. Yoma, N., Busso, C., Soto, I.: Packet-loss modelling in IP networks with state-duration constraints. In: IEE Proceedings Communications. Volume 152. (2005)