# Packet Loss in Real-Time Services: Markovian Models Generating QoE Impairments

Oliver Hohlfeld
Multimedia Communications Lab
Darmstadt University of Technology
Merckstr. 25, D-64283 Darmstadt, Germany
hohlfeld@kom.tu-darmstadt.de

Rüdiger Geib, Gerhard Haßlinger
T-Systems
Deutsche-Telekom-Allee 7
D-64295 Darmstadt, Germany
{ruediger.geib, gerhard.hasslinger}@t-systems.com

*Abstract*— **Real-time Internet services are gaining in popularity due to rapid provisioning of broadband access technologies. Delivery of high Quality of Experience (QoE) is important for consumer acceptance of multimedia applications.**

**IP packet level errors affect QoE and the resulting quality degradations have to be taken into account in network operation. We derive the second order statistics of the number of packet losses in finite Markov models over several relevant time scales and adapt them to loss processes visible in wired and wireless transmission channels.**

**Higher order Markov chains offer a large set of parameters to be exploited by complex fitting procedures. We experience that the 2-state Gilbert-Elliott model already captures a wide range of observed loss pattern appropriately and discuss how such models can be used to examine the quality degradations caused by packet losses.**

## I. Introduction

The transfer of real-time data for multimedia services over the Internet and channels in heterogeneous packet networks is subject to errors of various types which will affect the QoS and QoE. 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. Lost information will affect the perceived quality by impairing the multimedia content. The QoE degradation not only depends on the amount of lost packets, but also on the semantic of the lost information at the application layer. In video streams, a lost intra predicted I-Frame that is referenced by subsequent inter predicted P- and B-Frames may cause a much stronger visual impact due to error propagation than a lost inter predicted frame.

In this work, we focus on packet loss on Internet links with most traffic controlled by TCP superposed with a considerable contribution of real-time traffic without flow control. The impact of packet loss on the user's perception of real-time services can be investigated starting from measurement traces of traffic or generated by finite-state stochastic models, which have been adapted to the characteristics observed in the measurement and thus produce statistically similar traces. Using model based generators for loss processes has several advantages:

- the amount of necessary storage capacity is reduced significantly from several gigabyte to a set of model parameters,
- stochastic models usually include a set of parameters with a clear interpretation, which can be adapted to meet the demands of a considered scenario in which the model is used (e.g. a certain packet loss rate) and makes them more flexible than a measurement trace,
- the length of the generated sequence is independent of the measurement trace used for training,
- stochastic models produce random but statistically consistent sequences.

Both, using real data loss traces—e.g. captured in backbone links—and model generated loss traces have their benefits. The main disadvantage of using model generated loss traces is that statistical properties may not fit to the statistical properties of a measured trace as they are likely to be biased by model limitations. As measurement traces show characteristics on multiple time scales, we derive the second order statistics of finite Markov models to be used as a parameter estimation technique to adapt the model to the second order statistics of the amount of packet losses observed in a given traffic trace on multiple time scales by moment matching. We focused on 2-state Markov models in [10]. The present paper gives a generalised view on finite Markov models and discusses how these models can be used in the study of QoE impacts on video streams. 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.

Section II will describe the trace evaluation as the basis for further investigations. The traffic variability and packet loss patterns observed in multiple time scales will be discussed in Section II. Section III will introduce Markov chains as stochastic models to capture statistical properties of the training traces and produce artificial traces as output. Besides the general definition of finite-state Markov models, two commonly used models will be introduced and related work on Markov modelling will be discussed. From the definition of finite Markov chains, the second order statistics on multiple time scales will be derived in Section IV. In Section V, a comparison of different parameter estimation techniques for 2-state Markov models shows that simple Markov processes achieve a fairly close fit to the second order statistics over multiple time scales. Section VI will discuss how the presented models can be used to study degradations in the perceived quality of real-time video.

## II. Traffic Traces and Packet Loss Evaluation

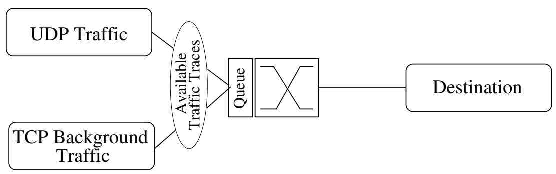



Fig. 1. Measurement topology: TCP backbone traffic is feed from a trace file along with UDP traffic into a router. The traffic is directed over an bottlenecked link to a destination. The loss rate can be arbitrarily chosen by adjusting the capacity of the outgoing, bottlenecked link.

As a starting point to estimate packet loss processes, we consider measurement traces of traffic taken from broadband access routers of Deutsche Telekom's IP platform, which connect residential ADSL access lines to the backbone [11]. Measurement data is available on an aggregation level of 2.5 Gb/s interfaces including a time-stamp and the size for each packet. The load on the links is usually moderate below 50 % such that packet loss is rare or not encountered at all. However, the variability of the traffic is visible on time scales from 1 ms to about half an hour as the length of the traces. Nevertheless, this allows to evaluate overload phases and corresponding loss pattern assuming lower capacity leading to more critical load levels. The evaluation setup is illustrated in Figure 1.

The loss pattern obtained in this way correspond to uncontrolled sources as in the UDP transport protocol (open loop), but do not include retransmissions and source rate adaptation of the TCP congestion control mechanism. TCP flows are dominant at around 90% of the Internet traffic volume together with an increasing portion (5% - 10%) of UDP including real time applications which currently become more and more popular. A response of TCP on packet loss is subject to timeouts and therefore not expected on time scales below 1s. On longer time scales TCP aims at stabilizing the load on a congested link close to a full utilization. In fact, the analysis and simulation of TCP congestion control on packet loss is complex and empirical studies seem missing especially for high speed links with a mixture of thousands of flows in parallel having largely different bandwidths and round trip delays [28].

Since our main focus is on real time video applications via UDP, we refer to the original measurement traces in order to generate realistic packet loss processes, although we are aware that the behaviour may be different on congested links. In general, we consider an arbitrary capacity $C$ such that the load is below the congested region and a tail drop buffer of limited size $B$. Since the loss rate is monotonously decreasing with the assumed capacity $C$, we can adjust $C$ in order to approach a considered packet loss rate.

Aggregated traffic shows characteristics in multiple time scales [11], [17], [36]. Based on the time-stamp and the size of each packet, the variability of the traffic rates can be observed in time scale ranging from the accuracy level of the time-stamps well below 1 ms up to the 30 minutes length of the traces. Let $\Delta$ be a time frame in this range. Then corresponding traffic rates $R_k(\Delta)$ for successive intervals of length $\Delta$ are determined 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 $\mu_R$ and the variance $\sigma_R^2(\Delta)$ are computed. In this way, the second order statistics is given considering $\sigma_R^2(\Delta)$ over a relevant range of $\Delta$. This statistics is a standard description method for traffic and is equivalent to the autocorrelation function over the considered time scales. Long-range dependent traffic patterns are classified as exact or second order self-similar depending on the autocorrelation of the process [17], [34]. The Hurst parameter $H$ characterizes the degree of long-range correlation for self-similar traffic where the coefficient of variation is decreasing in longer time scales depending on $H$: $c_v(\Delta) \stackrel{\text{def}}{=} \sigma(\Delta)/\mu = c{\cdot}\Delta^{(H-1)}$ where $1/2 \le H \le 1$. A Hurst parameter of $H \approx 0.8$ has been found for LAN traffic rates [17].

Table I shows the second order statistics for $\Delta = 1$ ms, 10 ms, 100 ms, 1 s and 10 s for UDP traffic with mean rate $\mu = 50.8$ Mb/s and the total traffic of a trace from a 2.5 Gb/s broadband access link with mean rate $\mu = 753.9$ Mb/s. 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 and $c_v(\Delta)$ does not decrease by constant factors as would be expected for self-similarity. This evaluation technique has been applied to

| | $c_v$(1 ms) | $c_v$(10 ms) | $c_v$(100 ms) | $c_v$(1 s) | $c_v$(10 s) |
|---|---|---|---|---|---|
| UDP | 0.3209 | 0.1220 | 0.0531 | 0.0433 | 0.0394 |
| Total | 0.1689 | 0.0635 | 0.0322 | 0.0259 | 0.0216 |

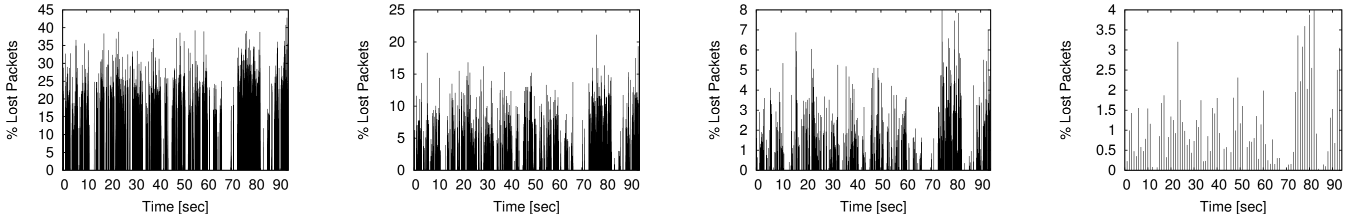
TABLE I
TRAFFIC VARIABILITY FOR UDP AND TOTAL TRAFFIC

the packet loss process obtained from backbone traces. As the traffic rate, also the packet loss process shows characteristic behaviour on multiple time scales

Thus, techniques for describing the variability in traffic rates using second order statistics will be also used for describing the packet loss process. An example in Figure 2 shows different packet loss rates on time scales ranging from 1 ms up to 1 second. Some millisecond intervals show up to 50 % packet loss, which is essentially smoothed for larger intervals.

## III. Stochastic Packet Loss Models

Finite-state Markov chains are widely used to characterize error processes in telecommunication systems and for performance evaluation of coding or other measures for error resilience [25], [27], [31], [35]. In this paper, we will use finite-state models to describe the packet loss process found in backbone measurements (wired channels) and DVB-H traces (wireless channels). A discrete Markov chain with a set of $M$




(a) 1 ms (b) 10 ms (c) 100 ms (d) 1 sec

Fig. 2. Packet Loss in fixed windows on multiple timescales

states $S = \{S_1, S_2, \cdots, S_M\}$ characterises the course of the process with regard to the current state, which may change over time at predefined events, e.g. packet arrivals, based on transition probabilities. Each state is associated with different error or packet loss behaviour. Let $q_t$ denote the current state at event time $t, t \in \mathbb{N}_0$. Then the probabilities $a_{ij}$ to change from state $q_{t-1} = i$ to $q_t = j$ are give in the transition matrix $A$

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{MM} \end{pmatrix}, \tag{1}$$

with coefficients

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i), \quad i \leq i, j \leq M,$$

where

$$a_{ij} \geq 0; \quad \sum_{j=1}^{N} a_{ij} = 1.$$

We restrict our considerations to irreducible and aperiodic Markov chains, where each state can be reached from each other with positive probability after a number of transitions and steady state probabilities $\pi_k$ exist for finding the process to sojourn a state in a long term perspective. The steady state probabilities are invariant with regard to a transition with matrix $A$ and thus can be computed from the system of linear equations:

$$\pi_k = \sum_{j=1}^{M} \pi_j a_{jk} \quad k = 1, \cdots, M; \quad \sum_{j=1}^{M} \pi_j = 1. \tag{2}$$

Finally, we define error or packet loss rates in each state $E = (e_1, ..., e_M);\ 0 \leq e_j \leq 1$ and the output of the process $O(t)$ as a binary sequence $O(t) \in 0, 1$ indicating an error or loss at an event with $O(t) = 1$, whereas $O(t) = 0$ stands for error free events, respectively. Thus we have $P(O(t) = (1 | q_t = S_j) \stackrel{\text{def}}{=} e_j$.

Only in simple cases, e.g. for a 2-state Markov process with $e_1 = 0$ and $e_2 = 1$, the current state can be recovered from the output $S(t) = O(t) + 1$. The term Hidden Markov Models expresses that $O(t)$ in general leaves uncertainty about the corresponding states $S(t)$ of the Markov chain [27]. A Markov process is completely defined by the transition matrix $A$, the state specific error rates $E$ and initial state $S_0$ or, more generally an initial state distribution $\pi_0 = P(S_0 = j)$. We continue with a brief discussion of two- and four state Markov models before the second order statistics of finite Markov models is derived in general.

### *A. Gilbert-Elliot: The Classical 2-State Markov Model for Error Processes*

In 1960, Gilbert [7] proposed a 2-state Markov chain to characterise a burst-noise channel. The usual notation of the Gilbert model distinguishes at first a good (G) and secondly a bad (B) state with different loss rates $e_G < e_B$. Gilbert [7] started with the special case of an error-free good state ($e_G = 0$) and left the extension to include losses generated in both states to Elliott [4]. Dwell times in the states are geometrically distributed with mean $1/p$ for the good and $1/r$ for the bad state, where $p$ and $r$ are the probabilities to change from the good state to the bad and vice versa. The Gilbert-Elliott 2-state Markov approach as depicted in Figure 3 is widely used for describing error patterns in transmission channels [8], [9], [20], [25], [32], [38], [40] and for analysing the efficiency of coding for error detection and correction. For applications in data loss processes, we interpret an event as the arrival of a packet and an error as a packet loss. Thus, the transition matrix and the steady state probabilities are

$$A = \begin{pmatrix} 1-p & p \\ r & 1-r \end{pmatrix}; \quad \pi_G = \frac{r}{p+r}; \quad \pi_B = \frac{p}{p+r}$$

with a total error rate $e = \pi_G e_G + \pi_B e_B = \frac{r e_G + p e_B}{p+r}$.

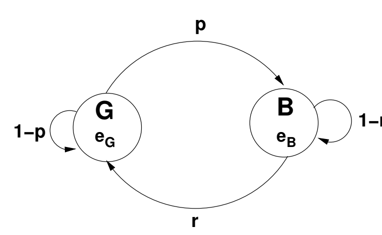



Fig. 3. The 2-state Markov model introduced by Gilbert and Elliott

Gilbert suggested to estimate the three parameters for his special case from measurement instances of a binary error process combining the current output $O(t)$ with two recent values $O(t-1)$ and $O(t-2)$: $a \stackrel{\text{def}}{=} P(1)$ and

$$b \stackrel{\text{def}}{=} P(1|1) = \frac{P(11)}{P(10) + P(11)}, \quad c \stackrel{\text{def}}{=} \frac{P(111)}{P(101) + P(111)}. \tag{3}$$

When the probabilities are computed from the 2-state model with $e_G = 0$ then the parameters are obtained via three equations:

$$r = 1 - \frac{ac - b^2}{2ac - b(a+c)}; \quad e_B = \frac{b}{1-r}; \quad p = \frac{ar}{e_B - a}. \quad (4)$$

Gilbert observed that this way of parameter estimation may lead to invalid parameters outside the range $0 \leq p, r, e_B \leq 1$ especially for small traces. He suggested to avoid the measurement of $c$ by simply choosing $e_B = 0.5 \Rightarrow r = 1 - 2b$. Morgera et al. [21] 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 trained by the Baum-Welch algorithm [31], [35].

Parameters of an even simplified Gilbert model with $e_B = 1$ can be estimated by assigning

$$p = P(1|0); \quad r = P(0|1). \quad (5)$$

### B. 4-State Markov Models

McDougall et al. [19] proposed a 4-state Markov model with two good and bad states to generate a hypergeometrical distribution of the duration of good and bad phases with specific transitions as approximation of an IEEE 802.11 channel. This model is an instance of the general case of partitioned-state models originally studied by Fritchman [6], which uses more than one good and bad state and partitions the state space into a good and bad state group containing $k$ and $M - k$ states, respectively. The McDougall model with its 6 non-redundant parameters $\alpha_g, \alpha_b, \beta_g, \beta_b, p_g$ and $p_b$ is shown in Figure 4 and has been found to perform well also in DVB-H simulations [25]. Due to its hypergeometric distribution of dwell times, it can be considered as extension of the Gilbert-Elliott model. A hypergeometric distribution is a useful approximation for loss bursts with long term correlation, which are often observed on error prone wireless links.

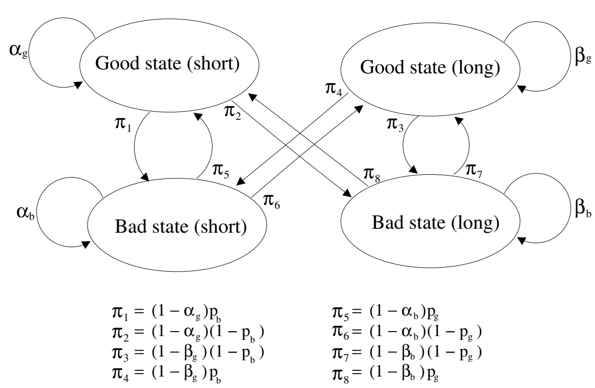



Fig. 4. Four-state Markov model [24]

As stated in [24], the run length distribution for the dwell time in the good and bad state groups is given by

$$f(n) = p(1-\alpha)\alpha^n + (1-p)(1-\beta)\beta^n,$$

which forms a hypergeometric distribution. The model parameters can be adjusted by fitting the distribution of dwell times separately for the good and bad state groups.

The parameter estimation by curve fitting is time intensive and can easily lead to invalid model parameters (transition probabilities $\notin [0, 1]$). Poikonen [25] derived an analytical technique using moment matching. This technique has been further simplified in [26], resulting in a single system of nonlinear equations.

### C. Markov models applied in the literature

Girod et al. [8] found a simple Gilbert model ($e_1 = 0$, $e_2 = 1$) 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. [12] considered real-time video transmissions and extended the simple Gilbert model by adapting it to the datagram loss process by adding a third state to describe out-of-order packets. Zhang et al. [40] use the simple 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.

Yajnik et al. [38] point out that the simple Gilbert model is suitable if the error gap length of the traces is geometrically distributed, but can be outperformed by considering high-order Markov chains. The 4-state Markov model and the Markov-based Trace Analysis [16] outperformed the Gilbert model, as the latter was unable to reproduce the variance in burst error lengths. Gamma distributed state durations can be used to replicate the variance in error burst lengths [20], [39].

### D. Parameter Estimation: Baum-Welch and Other Algorithms

The Baum-Welch algorithm [27] is a standard approach for adjusting the parameters of a finite state Markov model given by $\lambda = \{A, E, \pi\}$ such that the probability of the observation $O(t)$ given the model $\lambda$, $P(O(t)|\lambda)$, is maximised. The Baum-Welch algorithm, as an instance of the expectation-maximisation algorithm [35], will converge to a local maximum using a procedure with iterative updates to improve the parameters. The quality of the solution depends on the initial parameters. There is no guarantee for the estimation to converge to the global maximum.

Alternative techniques to match parameters of a Markov process—especially with regard to second order statistics in multiple time scales—have been successfully applied based on genetic algorithms [15] and using Pronys algorithm [30].

## IV. Variance of the Error Process over Multiple Time Scales

Second-order statistics in multiple time scales are a standard approach to capture and to describe traffic variability including long-range dependencies and self-similarity [17], [34]. Following this trend, we next derive the second-order statistics of the number of packet losses over a range of relevant time frames. Starting with the general solution for finite-state Markov models, we then derive explicit expressions for the Gilbert-Elliott model from the general result. Although Markov models do not exhibit self-similar properties, they have been successfully adapted to self-similar traffic [5], [30] and are still popular since they often lead to simple analytical treatment.

### A. Solution for Finite Markov Models

In order to capture a packet loss process generated by an $M$-state Markov model, we can set up recursive equations for the distribution function of losses in a considered sequence of packets. Let $p_n^{(m)}(k)$ denote the probability of $k$ packets being lost in a sequence of length $n$ generated by a Markov process, which starts in steady state and finally resides at state $m$ ($n \in \mathbb{N}_0$ and $1 \leq m \leq M$).

The probabilities $p_{n+1}^{(m)}(k)$ for sequences of length $n+1$ can be recursively computed from $p_n^{(m)}(k)$, where steady state starting conditions are expressed as $p_0^{(m)}(0) = \pi_m$:

$$p_{n+1}^{(m)}(k) = \sum_{j=1}^{M} p_n^{(j)}(k)\,(1-e_j)\,a_{jm} + p_n^{(j)}(k-1)\,e_j\,a_{jm}.$$

As a standard approach to obtain the mean and the variance of the distributions of packet losses we introduce corresponding generating functions defined as:

$$L_n^{(m)}(z) \stackrel{\text{def}}{=} \sum_{k=0}^{n} p_n^{(m)}(k)\,z^k.$$

A generating function $L_n^{(m)}(z)$ comprises the probabilities $p_n^{(m)}(k)$ for $k = 0, \cdots, n$. The recursive relationship for the probabilities $p_{n+1}^{(m)}(k)$ transfers into the generating function notation as follows:

$$L_{n+1}^{(m)}(z) = \sum_{j=1}^{M} L_n^{(j)}(z)\,(1-e_j+e_j z)\,a_{jm}. \qquad (6)$$

According to steady state starting conditions, we again have $L_0^{(m)}(z) = p_0^{(m)}(0) = \pi_0$. In addition, $L_n^{(m)}(1) = \sum_k p_n^{(m)}(k)$ generally holds for the sum of the probabilities of a distribution by definition of generating functions. Since the considered process stays in steady state, we have $\sum_k p_n^{(m)}(k) = \pi_m$ as the probability to find the process in state $m$ after $n$ packet arrivals. Therefore $L_n^{(m)}(z)$ represent defective distributions with regard to a final state $m$ of the Markov chain, while their sum $L_n(z) = \sum_{m=1}^{M} L_n^{(m)}(z)$ characterises the complete distribution of packet losses during a sequence of $n$ packet arrivals with $L_n(1) = \sum_{m=1}^{M} L_n^{(m)}(1) = \sum_{m=1}^{M} \pi_m = 1$.

*1) Mean values:* Next we compute the mean values $\mu_n^{(m)}$ for the distributions $p_n^{(m)}(k)$ via the first derivative of the generating function using the rule $\mu_n^{(m)} = \frac{d}{dz} L_n^{(m)}(z)|_{z=1}$. The first derivation of equation (6)

$$\tfrac{d}{dz} L_{n+1}^{(m)}(z) = \sum_{j=1}^{M} \Big( \big(\tfrac{d}{dz} L_n^{(j)}(z)\big)(1-e_j+e_j z) + L_n^{(j)}(z) e_j \Big) a_{jm}$$

is evaluated for $z=1$ to obtain the mean values recursively:

$$\mu_{n+1}^{(m)} = \sum_{j=1}^{M} (\mu_n^{(j)} + \pi_j e_j) a_{jm}. \qquad (7)$$

The result is a scheme to compute $\mu_n^{(m)}$ starting from

$$\begin{aligned}
\mu_0^{(m)} &= 0; \\
\mu_1^{(m)} &= \sum_{j=1}^{M} \pi_j e_j a_{jm}; \\
\mu_2^{(m)} &= \sum_{k=1}^{M} (\mu_1^{(k)} + \pi_k e_k) a_{km} \\
&= \sum_{k=1}^{M} \Big( \sum_{j=1}^{M} \pi_j e_j a_{jk} + \pi_k e_k \Big) a_{km} \\
&= \sum_{j=1}^{M} \pi_j e_j \sum_{k=1}^{M} a_{jk} a_{km} + \sum_{k=1}^{M} \pi_k e_k a_{km} \\
&= \sum_{j=1}^{M} \pi_j e_j a_{jm}^{(2)} + \pi_j e_j a_{jm}
\end{aligned}$$

where $a_{jm}^{(k)}$ denote the $k$-step transition probabilities as coefficients of the matrix $A^k$, including $a_{jm}^{(1)} = a_{jm}$. We straightforwardly proceed to the general result

$$\mu_n^{(m)} = \sum_{j=1}^{M} \pi_j e_j \sum_{k=1}^{n} a_{jm}^{(k)}. \qquad (8)$$

To conclude the mean value analysis, we look at the mean values for the complete distribution

$$\begin{aligned}
\sum_{m=1}^{M} \mu_n^{(m)} &= \sum_{j=1}^{M} \pi_j e_j \sum_{k=1}^{n} \sum_{m=1}^{M} a_{jm}^{(k)} \\
&= \sum_{j=1}^{M} \pi_j e_j \sum_{k=1}^{n} 1 = n \sum_{j=1}^{M} \pi_j e_j = ne.
\end{aligned}$$

Since we start and stay in steady state conditions with loss rate $e = \sum_j \pi_j e_j$ per considered packet, it is not surprising to encounter $ne$ losses in the mean for $n$ packets.

*2) Second order statistics:* Next we proceed with the second order statistics of the process to calculate the variance of the number of packet losses as our main focus. Therefore the second derivatives of the generating functions are again evaluated at $z=1$:

$$\frac{d^2}{d^2 z} L_{n+1}^{(m)}(z) = \sum_{j=1}^{M} \Big( \big(\frac{d^2}{d^2 z} L_n^{(m)}(z)\big)(1-e_j+e_j z) + 2\big(\frac{d}{dz} L_n^{(m)}(z)\big) e_j \Big) a_{jm}$$

$$\Rightarrow \quad \nu_{n+1}^{(m)} = \sum_{j=1}^{M} (\nu_n^{(j)} + 2\mu_n^{(j)} e_j) a_{jm}$$

where $\nu_n^{(m)} \stackrel{\text{def}}{=} \frac{d^2}{d^2 z} L_n^{(m)}(z)|_{z=1}$. Summing up over the final state $m$ of the Markov process, we approach the result for the complete distribution:

$$\begin{aligned}\sum_{m=1}^{M} \nu_{n+1}^{(m)} &= \sum_{r=1}^{M} (\nu_n^{(r)} + 2\mu_n^{(r)} e_r) \sum_{m=1}^{M} a_{rm} \\ &= \sum_{r=1}^{M} (\nu_n^{(r)} + 2\mu_n^{(r)} e_r) = \sum_{s=1}^{n} \sum_{r=1}^{M} 2\mu_s^{(r)} e_r \\ &= \sum_{s=1}^{n} \sum_{r=1}^{M} \sum_{j=1}^{M} 2\pi_j e_j \sum_{k=1}^{s} a_{jr}^{(k)} e_r \quad \text{using (8)} \\ &= 2 \sum_{j=1}^{M} \sum_{r=1}^{M} \pi_j e_j e_r \sum_{s=1}^{n} \sum_{k=1}^{s} a_{jr}^{(k)}; \\ \sum_{m=1}^{M} \nu_n^{(m)} &= 2 \sum_{j=1}^{M} \sum_{r=1}^{M} \pi_j e_j e_r \sum_{k=1}^{n-1} (n-k) a_{jr}^{(k)}.\end{aligned}$$

Finally, we obtain the variance—or the coefficient of variation, respectively—of a Markovian loss process for a series of $n$ packets via the general rule for the generating function $X(z)$ of a discrete random variable $X$ with mean $E(X)$ and standard deviation $\sigma(X)$:

$$X''(1) = E(X^2) - E(X) = \sigma^2(X) + \mu^2(X) - \mu(X)$$

With regard to the Markovian loss process within a series of $n$ packets we have

- $X''(1) := \sum_{m=1}^{M} \nu_n^{(m)}$ as the second derivative,
- $\mu(X) := ne$ for the mean number of losses,
- $c_v(n) \stackrel{\text{def}}{=} \sigma(X)/E(X) \;\Rightarrow$

$$\begin{aligned}c_v(n) &= \frac{1}{ne}\sqrt{ne - (ne)^2 + \sum_{m=1}^{M} \nu_n^{(m)}} \qquad (9) \\ &= \frac{1}{ne}\sqrt{ne - (ne)^2 + 2\sum_{j=1}^{M}\sum_{r=1}^{M} \pi_j e_j e_r \sum_{k=1}^{n-1} (n-k) a_{jr}^{(k)}}.\end{aligned}$$

*3) Explicit eigenvalue solution:* The previous result is generally applicable to compute $c_v(n)$ at a moderate computational complexity, including the determination of the $k$-step transition matrices for $k = 1, \cdots, n$ as the main step. In addition, it is well known that the coefficients $a_{jr}^{(k)}$ can be expressed in a direct solution form for the corresponding eigenvalue problem.

In particular, an irreducible transition matrix $A = (a_{jr})$ has $M$ eigenvalues $z_1, \cdots, z_M$ which can be computed as the roots of a characteristic polynomial $A - Iz = 0$, where $I$ denotes the identity matrix. The largest eigenvalue is $z_M = 1$, whereas the others are located inside the complex unit sphere. For each eigenvalue $z_l$ the right and left eigenvectors are denoted as $x_{1l}, \cdots, x_{Ml}$ and $y_{l1}, \cdots, y_{lM}$. Then the $k$-step transition probabilities can be written in the form

$$a_{jr}^{(k)} = \sum_{l=1}^{M} x_{jl} y_{lr} z_l^k \Big/ \sum_{s=1}^{M} x_{sl} y_{ls}.$$

We have tools available to find the roots $z_1, \cdots, z_m$ of the characteristic polynomial even for large degree $M$ [33]. Therefore the latter representation of $a_{jr}^{(k)}$ as a sum of $M$ geometrical terms can be used to resolve the last sum over the length $n$ in equation (9) on account of a sum over $M-1$ eigenvalue terms, where the term for $z_M = 1$ can be ignored. In this way the calculation of $c_v(n)$ is facilitated especially for large $n$ and small number $M$ of states in the Markov process. For a 2-state Markov process the eigenvalue solution reduces to a single real valued geometrical term, which is evaluated and adapted to measurement data in the next subsections.

### B. Solution for 2-State Gilbert-Elliott Models

As a special case, we summarize the explicit solution for the Gilbert-Elliott Markov model as derived in [10]. To the authors knowledge, explicit expressions for the second order statistics of the 2-state Markov model for arbitrary time scales are not given in the literature, although there is a large volume of work involving the Gilbert-Elliott model, as partly discussed in Section III-C. However, most of this work is devoted to error detecting and correcting codes and the residual error rates of coding schemes for different transmission channels, rather than on traffic or packet loss characterisation.

*1) Mean values:* In the previous notation, the recursive formula (7) for the mean number of losses during a series of $n$ packets with final state $G$ ($B$) in the 2-state Markov process are rewritten as

$$\mu_{n+1}^G = (1-p) \cdot \left(\frac{re_G}{p+r} + \mu_n^G\right) + r \cdot \left(\frac{pe_B}{p+r} + \mu_n^B\right),$$

$$\mu_{n+1}^B = p \cdot \left(\frac{re_G}{p+r} + \mu_n^G\right) + (1-r) \cdot \left(\frac{pe_B}{p+r} + \mu_n^B\right).$$

As an explicit solution we obtain

$$\mu_n^B = \beta_B n + \gamma_B \frac{\alpha(1-\alpha^n)}{1-\alpha} \quad \text{where } \alpha \stackrel{\text{def}}{=} 1-(p+r); \qquad (10)$$

$$\beta_B \stackrel{\text{def}}{=} \frac{pre_G + p^2 e_B}{(p+r)^2}; \quad \gamma_B \stackrel{\text{def}}{=} \frac{pr(e_B - e_G)}{(p+r)^2}.$$

The case $\alpha = 1 \;\Leftrightarrow\; p = r = 0$ implies a reducible and thus non-ergodic Markov chain.

Due to the symmetry of both states $G$ and $B$, the result for $\mu_n^G$ has mutually exchanged parameters $p \leftrightarrow r$ and $e_G \leftrightarrow e_B$: $\mu_n^G(p, r, e_G, e_B) = \mu_n^B(r, p, e_B, e_G)$.

*2) Second order statistics:* Again we can start from the result for the coefficient of variation $c_v(n)$ for $M$-state Markov processes given in Equation (9), which is expressed in terms of the mean values $\mu_n^B$ and $\mu_n^G$ using the previously presented explicit solution:

$$\begin{aligned} c_v(n) &= \frac{1}{ne}\sqrt{ne-(ne)^2+\nu_n^{(G)}+\nu_n^{(B)}} \\ &= \frac{1}{ne}\sqrt{ne-(ne)^2+2\sum_{k=1}^{n-1}(\mu_k^{(G)}e_G+\mu_k^{(B)}e_B)} \\ &= \frac{1}{\sqrt{n}}\sqrt{\frac{1}{e}-1+\tau\varphi(n)} \qquad (11) \end{aligned}$$

where $\tau = \frac{2pr(1-p-r)(e_G-e_B)^2}{(p+r)(pe_B+re_G)^2}$

and $\varphi(n) = 1 - \frac{1-(1-p-r)^n}{n(p+r)}$.

The solution is comprehensible enough to interpret the influence of the model parameters.

### *C. Parameter Impact on the Second Order Statistics of the Gilbert-Elliott Model*

Based on the analytical result in Equation (11) for $c_v(n)$, the main properties of the second order statistics of the Gilbert-Elliott model are summarised as follows.

1) Starting point of the $c_v(n)$-curves is $c_v(1) = \sqrt{1/e-1}$ since $\varphi(1) = 0$. $c_v(n)$ only depends on the entire packet loss rate $e = (re_G + pe_B)/(p + r)$.
2) For the asymptotical behaviour we observe $\lim_{n\to\infty}\varphi(n) = 1$, which simplifies the $c_v(n)$-curve $\lim_{n\to\infty} c_v(n) = \sqrt{1/e-1+\tau}/\sqrt{n}$. In the logarithmic representation of Figure 5 the $c_v(n)$-curves therefore approach straight lines with the same slope.
3) Figure 5 shows results, where $p + r$ is again stepwise reduced by a factor 10. In all examples of Figure 5 we keep the ratio $p/r = 0.1$ constant such that $e = 0.001 \Rightarrow c_v(1) = \sqrt{999}$. The curves are characterised by a horizontal part, which holds the variance on the initial $c_v(1)$ value followed by a declining part. The length of the part at constant level depends on $p+r$, i.e. on the intensity of transitions between the states, which is different but fixed for each curve in Figure 5. The sojourn times of the good and bad state are geometrically distributed with mean $1/p$ and $1/r$, respectively. For $\lim_{p+r\to 0}$ the mean holding times of the states are extended on longer times scales. Then the correlation in the modelling process persists over about the same time scale and the transition point from the constant to the declining part of the $c_v(n)$ curve is shifted in the range between $1/p$ and $1/r$.
   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 $c_v(kn)/c_v(n) \to 1/\sqrt{k}$.

It is apparent from the latter example that the $c_v(n)$-curve consists of up to three parts:

1) A decreasing phase where $c_v(n) \approx \sqrt{1/e-1}/\sqrt{n}$ for $n < n_1$ such that $\varphi(n) \approx 0$ and $\tau\varphi(n) \ll 1/e - 1$;
2) A phase at almost constant level for $n_1 \le n < n_2$ where $c_v(n) = \sqrt{1/e-1}/\sqrt{n_1}$ while $\varphi(n)$ is increasing from $\varphi(n_1) \approx 0$ to $\varphi(n_2) \approx 1$;

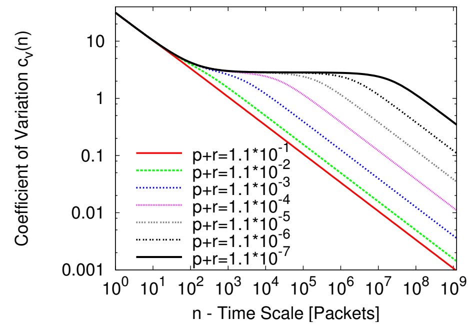



Fig. 5. Parameter impact on the 2. order statistics of the Gilbert-Elliott model

3) A decreasing phase where $c_v(n) = \sqrt{1/e-1+\tau}/\sqrt{n}$ for $n_2 < n$.

Depending on the parameters, the first two phases are not always relevant, as can be seen in some examples of Figure 5.

## V. Evaluation

The following section will discuss the evaluation of the proposed parameter adaption technique based on second order statistics in multiple time scales for 2-state Markov models and compares the results to classical fitting methods.

The evaluation of the trained 2-state Markov models using the coefficient of variation $c_v(n)$ is shown in Figure 6 and 7 for a backbone trace evaluated with two different packet loss rates using the network simulator ns-2 with a RED queue leading to two loss traces. In both plots, the Poisson process provides a linear lower bound for $c_v(n) = c_v(1)/\sqrt{n}$ without any autocorrelation.

The parameters of the Gilbert model and its simplified version for $e_G = 0, e_B = 1$ have been estimated from the given loss traces using the traditional methods shown in Equation (5) and (3-4). A Gilbert-Elliott model has been fitted using the Baum-Welch algorithm for both traces. Due to limitations of the Jahmm library [1], a set of 2436 observations, each containing 10000 packets, have been randomly chosen from the backbone packet loss trace and used as input for the Baum-Welch algorithm. As the mean loss rate in a single observation varies from the mean loss rate of the entire trace (cf. Figure 2), the resulting model parameters did not exactly fit the mean loss rate of the considered trace. We corrected the model to match $e = (pe_G + re_B)/(p + r)$ by modifying $p := p + \Delta$ and $r := r - \Delta$ such that
$((p+\Delta)e_G + (r-\Delta)e_B)/(p+r) := e + (e_G - e_B)\Delta/(p+r)$.
In this way, the loss rate $e$ can be modified in a range between $e_G$ and $e_B$ which is sufficient since usually small corrections apply.

Moreover, the simplified Gilbert model with only two parameters and the Gilbert-Elliott model have been trained based on the second order statistics over multiple time scales $n \in [1, 10^5]$, as shown in Figure 6 and 7. 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 shown in Figure 5.

The distance between different model curves as shown in Figure 6, 7 and 8 and the trace curve is measured by the Mean Square Error (MSE)

$$\mathrm{MSE(model)} = 10^{-5} \sum_{n=1}^{10^5} (c_v^{\mathrm{model}}(n) - c_v^{\mathrm{trace}}(n))^2,$$

where a smaller MSE indicates a better fit. The MSE distance is shown in Table II.

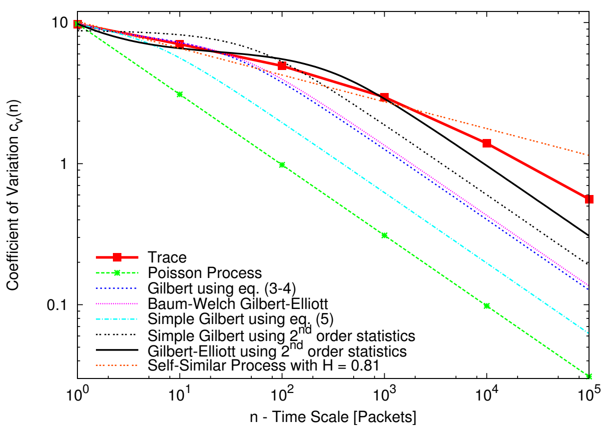



Fig. 6. Evaluation of the trained 2-state Markov models using the coefficient of variation $c_v(n)$ for backbone traffic with a mean packet loss rate of 1%

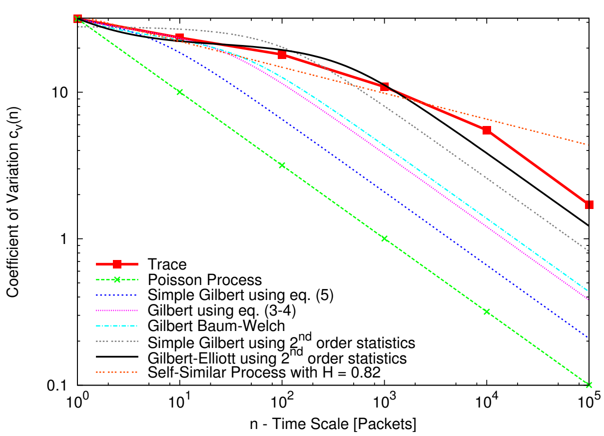



Fig. 7. Evaluation of the trained 2-state Markov models using the coefficient of variation $c_v(n)$ for backbone traffic with a mean packet loss rate of 0.1%

In addition to backbone traces, parameter adaption techniques were evaluated for a DVB-H packet loss trace resulting from a laboratory measurement at University of Turku, leading to a good fit of the Gilbert-Elliott model trained in multiple time-scales as shown in Figure 8.

Considering the coefficient of variation curve as variance-time plot [3], [17], the Hurst parameter can be obtained by

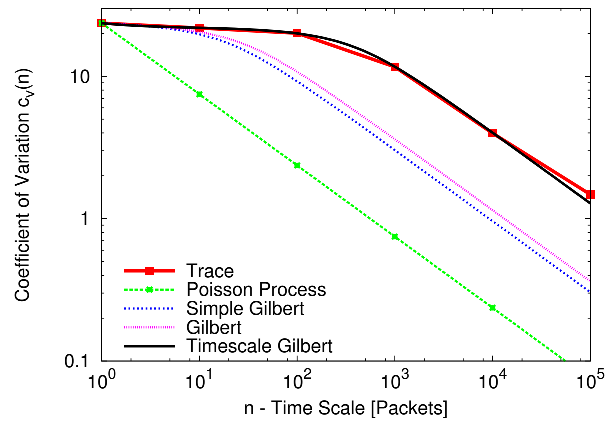



Fig. 8. Evaluation of the trained 2-state Markov models using the coefficient of variation $c_v(n)$ for DVB-H traffic with a mean packet loss rate of 1.8%

fitting $c_v(1) \cdot n^{H-1}$ to the $c_v(n)$ curve for $n \in [1, 10^5]$. For both traces, a Hurst parameter $H \approx 0.8$ has been found, as can be seen in Figure 6 and 7. The loss process of the considered traces shows a faster decay than self-similar processes for a time scale of $> 10^3$ packets, suggesting that Markov models are appropriate to capture this behaviour.

However, when we look at the distribution of the length of packet losses in a 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 can include long-range correlation. 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 bursts surely increases with the observed mean failure burst length in a considered traffic flow.

## VI. Video Quality of Experience

In the following section, we discuss the impacts of packet loss on video quality metrics including related work. The integration of the previous packet layer loss models into a common framework with the video transmission and impairments visible to a viewer are for further study in an ongoing project.

### A. *Impairments at the Video Layer caused by Packed Loss*

We focus mainly on the H.264/AVC video standard (aka MPEG-4 Part 10), written by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG), as used in the IPTV platform of Deutsche Telekom. In video compression, sequences are decomposed into still pictures (*frames*) which are grouped into a *Group of Pictures* (GoP). Frames are segmented into *macro blocks*—usually of size 16x16 pels—and organized into groups of blocks called *slice*. Macro blocks are typically positioned in scan-line order, but due to Flexible Macroblock Ordering in H.264 arbitrary order is also possible. A frame is coded in

| Trace: Mean Loss Rate | Simple Gilbert (eq. 5) | Gilbert (eq. 3-4) | Gilbert (Baum-Welch) | Simple Gilbert (Timescale) | Gilbert-Elliott (Timescale) |
|---|---|---|---|---|---|
| Fig. 6: 1 % | 3.0047 | 0.90945 | 0.78548 | 0.72772 | 0.13490 |
| Fig. 7: 0.1 % | 43.524 | 19.087 | 15.420 | 7.9595 | 1.0929 |
| Fig. 7: 1.8 % | 34.54 | 27.08 | – | – | 0.008 |

TABLE II

MEAN SQUARE ERROR (MSE) DISTANCE BETWEEN TRAINED MODELS AND THE TRACES

one of three modes: *I frames* contain only intra coded macro blocks, *P frames* can contain intra or predicted macro blocks and *B frames* can also contain bidirectional predicted macro blocks. When using prediction, the displacement of a macro block is expressed as motion vector relative to the reference frame.

During the decoding process, the decoder may have to cope with errors due to packet loss, where the concrete error concealment strategy is decoder dependent. When only a small fraction of the slice is corrupted, the decoder often is not aware of the concrete error position and thus may discard the slice. The perceived visual quality of an erroneous received but decoded and displayed slice may be worse than in case of discarding and copying the slice from a previous frame. Information introduced by copying from a previous slice may not always be noticeable by a viewer. An example may be comic strips which usually have a very static background and thus losing a slice that holds only background information is often unnoticed if the lost data is copied from the last slice.

As most currently used encoders are not aware of the underlying transport protocol, slices are typically larger than the payload of a single IP packet. When MPEG2 transport streams are used to packetize the video data into IP packets, slice boundaries are usually inside an IP packet, as MPEG-2 TS considers video data as endless bitstream. Thus, losing such a packet can discard two slices at once.

Losing a parameter set due to packet loss will have a severe effect on the decoding process. Thus, they may be transmitted out-of-band or in-band but periodically, e.g. with each GoP, or prioritized using different service classes in differentiated services.

### B. Markov Models as Basis for Subjective Evaluations

Finite Markov models can be used as a basis for subjective video quality evaluations where subjects are asked to rate given video sequences using the Mean Opinion Score (MOS) as currently conducted at T-Labs Berlin for high-definition (HDTV) video sequences in Internet Protocol Television (IPTV). During the test preparation, H.264/AVC videos are packetized using a MPEG2 transport stream and impaired with a Markov model. The impaired video sequence is rated by the subject during test. The advantage of Markov models is their flexibility to be adapted to arbitrary loss rates and loss burst lengths, such that different impairments can be studied. The result of a subjective evaluation can be a mapping of the packet loss rate to quality degradations in the MOS rating.

### C. Related Work on Analytical Video Frameworks

Whereas empirical evaluations treat video as a black box and typically try to find correlations between link-layer factors and the subjective quality, objective quality metrics can be derived that ideally correlate well with human perception and thus avoid cost intensive empirical evaluations. Quality metrics can be classified into three categories by the required amount of reference information [37]: *Full-reference* (FR) metrics are based on frame-by-frame comparison between a reference video and the video to be evaluated; *No-reference* (NR) metrics have to make assumptions about the video content and distortions, e.g. by evaluating the blockiness of a frame, as a common artifact in block-based compression algorithms such as MPEG; *Reduced-reference* (RR) metrics evaluate the test video based on a subset of features previously extracted from the reference video. Lotfallah et al. [18] propose a RR metric to evaluate the visual quality based on traces which are enriched with frame and motion related information. The proposed metric is based on frame dependencies and models the impact of a packet loss by considering the error propagation of the affected frame, e.g. a lost I-frame will impair the entire Group of Pictures. However, this metric does not consider bursty losses.

The peak signal-to-noise ratio (PSNR) as a NR metric has been found not to correlate well with human perception [23], [41] as it works on a pixel basis only and neglects the complex human perception. Some work extends the PSNR [2], [22].

Based on empirical evaluations, the visibility of packet loss in MPEG-2 video sequences using a decision tree is investigated in [13] and for H.264/AVC coded videos in [14] by assuming a packet loss results in a lost slice and not considering Flexible Macroblock Ordering (FMO). A metric that combines these RR metrics has been proposed in [29].

## VII. CONCLUSION

Quality of Experience aspects are a vital factor in ensuring customer satisfaction in today's network services. In order to study quality degradations in video streams caused by packet loss, a Markovian error pattern generator can be used to simulate physical channels of various type. We derived the second order statistics for the distribution of the number of lost packets over multiple time scales, which in general can be recursively determined for increasing time frames and via explicit terms for the Gilbert-Elliott model. The fitting procedure leads to a closer match in multiple time scales than classical methods. The proposed approach gives more flexibility to include information from different time scales enabling a simple and useful fit for long traces of traffic and packet loss processes. We plan to study the effect on the video frame level for future work, charatersising the impairments at the receiver side depending on the applied video coding scheme.

## Acknowledgement

We would like to thank for the support of Oliver Hohlfeld's Master's thesis by the Deutsche Telekom Laboratories (T-Labs), Berlin within the T-V-Model project and especially the fruitful discussions with Peter List at T-Systems and Alexander Raake and Marie-Neige Garcia at T-Labs. Further thanks to Jussi Poikonen at the University of Turku, Finland, for providing DVB-H packet loss traces.

## References

[1] "Jahmm hmm library," http://www.run.montefiore.ulg.ac.be/~francois/software/jahmm/.

[2] I. Bouazizi, "Estimation of packet loss effects on video quality," in *First IEEE International Symposium on Control, Communications and Signal Processing*, 2004, pp. 91–94.

[3] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," in *SIGMETRICS '96: Proceedings of the 1996 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*. New York, NY, USA: ACM, 1996, pp. 160–169.

[4] E. O. Elliott, "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell System Technical Journal*, vol. 42, pp. 1977–1997, Sep. 1963.

[5] A. Feldmann and W. Whitt, "Fitting mixtures of exponentials to long-tail distributions to analyze network performance models," *Performance Evaluation*, vol. 31, no. 3, pp. 245–279, 1998.

[6] B. D. Fritchman, "A binary channel characterization using partitioned Markov chains," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 221–227, April 1967.

[7] E. N. Gilbert, "Capacity of a Burst-Noise Channel," *Bell System Technical Journal*, vol. 39, pp. 1253–1265, Sep. 1960.

[8] B. Girod, K. Stuhlmüller, M. Link, and U. Horn, "Packet Loss Resilient Internet Video Streaming," in *Proceedings of SPIE Visual Communications and Image Processing*, January 1999, pp. 833–844.

[9] J. A. Hartwell and A. O. Fapojuwo, "Modeling and characterization of frame loss process in IEEE 802.11 wireless local area networks," in *IEEE Vehicular Technology Conference*, 2004, pp. 4481–4485.

[10] G. Haßlinger and O. Hohlfeld, "The Gilbert-Elliott Model for Packet Loss in Real Time Services on the Internet," in *14. GI/ITG Konferenz Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen (MMB)*, Dortmund, Germany, Mar. 2008.

[11] G. Haßlinger, J. Mende, R. Geib, T. Beckhaus, and F. Hartleb, "Measurement and characteristics of aggregated traffic in broadband access networks," in *20th International Teletraffic Congress (ITC20)*, L. Mason, T. Drwiega, and J.Yan, Eds. Ottawa, Canada: Springer LNCS 4516, June 2007, pp. 998–1010.

[12] T. Huitika and P. Driessen, "Datagram loss model for noninteractive real time streaming video," in *IEEE Pacific Rim Conference on Communications. Computers and Signal Processing (PACRIM)*, August 2003, pp. 756–759.

[13] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan, "Modeling packet-loss visibility in MPEG-2 video," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341–355, April 2006.

[14] S. Kanumuri, S. G. Subramanian, P. C. Cosman, and A. R. Reibman, "Predicting H.264 packet loss visibility using a generalized linear model," 2006.

[15] S. Kempken and W. Luther, "Modeling of h.264 high definition video traffic using discrete-time semi-markov processes," in *International Teletraffic Congress*, ser. Lecture Notes in Computer Science, L. Mason, T. Drwiega, and J. Yan, Eds., vol. 4516. Springer, 2007, pp. 42–53.

[16] A. Konrad, B. Y. Zhao, A. D. Joseph, and R. Ludwig, "A Markov-based Channel Model Algorithm for Wireless Networks," *Wireless Networks*, vol. 9, no. 3, pp. 189–199, 2003.

[17] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic," *ACM SIGCOMM Computer Communication Review*, vol. 23, no. 4, pp. 183–193, 1993.

[18] O. A. Lotfallah, M. Reisslein, and S. Panchanathan, "A framework for advanced video traces: evaluating visual quality for video transmission over lossy networks," *EURASIP Journal on Applied Signal Processing*, no. 1, 2006.

[19] J. McDougall, J. J. John, Y. Yu, and S. L. Miller, "An improved channel model for mobile and ad-hoc network simulations." in *Communications, Internet, and Information Technology*, M. H. Hamza, Ed. IASTED/ACTA Press, 2004, pp. 352–357.

[20] J. McDougall and S. Miller, "Sensitivity of wireless network simulations to a two-state Markov model channel approximation," in *IEEE Globecom*, 2003.

[21] S. D. Morgera and F. Simard, "Parameter estimation for a burst-noise channel," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Washington, DC, USA, 1991, pp. 1701–1704.

[22] O. Nemethova, M. Ries, M. Zavodsky, and M. Rupp, "PSNR-based estimation of subjective time-variant video quality for mobiles," in *International Conference MESAQIN on Measurement of Speech, Audio and Video Quality in Networks*, 2006.

[23] E. Ong, X. Yang, W. Lin, Z. Lu, and S. Yao, "Video quality metric for low bitrate compressed videos," in *IEEE International Conference on Image Processing (ICIP)*, vol. 5, 2004, pp. 3531–3534.

[24] J. Poikonen and J. Paavola, "Error Models for the Transport Stream Packet Channel in the DVB-H Link Layer," in *IEEE International Conference on Communications (ICC)*, June 2006, pp. 1861–1866.

[25] J. Poikonen, "Geometric Run Length Packet Channel Models Applied in DVB-H Simulations," in *Proceedings of the IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, September 2006.

[26] ——, "Parametrization of Aggregated Renewal Markov Processes for DVB-H Simulations," in *Proceedings of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2007.

[27] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[28] S. Rank and H.-P. Schwefel, "Transient analysis of red queues: A quantitative analysis of buffer-occupancy fluctuations and relevant time-scales," *Performance Evaluation*, vol. 63, pp. 725–742, 2006.

[29] A. R. Reibman and D. Poole, "Characterizing packet-loss impairments in compressed video," in *IEEE International Conference on Image Processing (ICIP)*, 2007.

[30] P. Salvador, R. Valadas, and A. Pacheco, "Multiscale fitting procedure using markov modulated poisson processes," *Telecommunication Systems*, vol. 23, pp. 123–148, 2003.

[31] S. Sivaprakasam and K. Shanmugan, "An Equivalent Markov Model for Burst Errors in Digital Channels," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 1347–1355, Feb/Mar/Apr 1995.

[32] C. Tang and P. K. McKinley, "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, pp. 130–133.

[33] D. Traczinski, G. Haßlinger, and W. Luther, "Polynomial factorization for servers with semi-markovian workload: performance and numerical aspects of a verified solution technique," *Stochastic Models*, vol. 21, pp. 643–668, 2005.

[34] B. Tsybakov and N. D. Georganas, "On self-similar traffic in atm queues: definitions, overflow probability bound, and cell delay distribution," *IEEE/ACM Trans. Netw.*, vol. 5, no. 3, pp. 397–409, 1997.

[35] W. Turin, *Performance Analysis and Modeling of Digital Transmission Systems (Information Technology: Transmission, Processing and Storage)*. Secaucus, NJ, USA: Springer New York, Inc., 2004.

[36] H. van den Berg, M. Mandjes, R. van de Meent, A. Pras, F. Roijers, and P. Venemans, "Qos-aware bandwidth provisioning for IP backbone links," *Computer Networks*, vol. 58, pp. 631–647, 2006.

[37] S. Winkler, "Video quality and beyond," in *Proc. European Signal Processing Conference*, 2007.

[38] M. Yajnik, S. B. Moon, J. F. Kurose, and D. F. Towsley, "Measurement and modeling of the temporal dependence in packet loss," in *IEEE INFOCOM*, 1999, pp. 345–352.

[39] N. Yoma, C. Busso, and I. Soto, "Packet-loss modelling in IP networks with state-duration constraints," in *IEE Proceedings Communications*, vol. 152, no. 1, February 2005.

[40] L. Zhang, D. Chow, and C. H. Ng, "Cell loss effect on QoS for MPEG video transmission in ATM networks," in *IEEE International Communication Conference (ICC)*, 1999, pp. 147–151.

[41] M. Zink, "Scalable internet video-on-demand systems," Ph.D. dissertation, Darmstadt University of Technology, Jul. 2003.