Internet Loss-Delay Modeling by Use of Input/Output Hidden Markov Models

Pierluigi Salvo Rossi
Dipartimento di Informatica e Sistemistica
Università di Napoli "Federico II"
Email: salvoros@unina.it

Francesco Palmieri
Dipartimento di Ingegneria dell'Informazione
Seconda Università di Napoli
Email: frapalmi@unina.it

Athina P. Petropulu
and Jie Yu
Electrical and Computer Engineering Department
Drexel University
Email: {athina,yujie}@cbis.ece.drexel.edu

Giulio Iannello
Università Campus Bio-Medico di Roma
Email: iannello@unina.it

Abstract—**Performance of real-time applications on end-to-end packet channels are strongly related to losses and temporal delays. Several studies showed that these network features may be correlated and present a certain degree of memory such as bursty losses and delays. The memory and the statistical dependence between losses and temporal delays suggest that the channel may be well modeled by a Hidden Markov Model with appropriate hidden variables that capture the current state of the network. In this paper we propose an Input/Output Hidden Markov Model that, trained with a modified version of the Expectation-Maximization algorithm, shows excellent performance in modeling typical channel behaviors in a set of real packet links. The work extends to case of variable inter-departure time the previous proposed Hidden Markov Model that well characterizes losses and delays of packets from a periodic source.**

I. INTRODUCTION

Gilbert and Elliott works [1][2] on modeling burst-error channels for bit-transmission showed how a simple 2-states Hidden Markov Model (HMM) was effective in characterizing some real communication channels. As in the case of bit-transmission channels, end-to-end packet channels show burst-loss behavior. Jiang and Schulzrinne [8] investigated lossy behavior of packet channels finding that a simple Markov chain is not able to describe appropriately the inter-loss behavior of channels. They also found that delays manifest temporal dependency, i.e. they should not be assumed to be a memoryless phenomenon. Salamatian and Vaton [9] found that an HMM trained with experimental data seems to capture channel loss behavior. Liu, Matta and Crovella [11] used an HMM-based loss-delay modeling in the contest of TCP traffic in order to infer loss nature in hybrid wired/wireless environments. They found that such a kind of modeling can be used to control TCP congestion avoidance mechanism. Similar works have been done by Zorzi [4] and Turin and van Nobelen [6] on wireless fading links.

Fig. 1. Periodic source scenario.

These works suggest that a Bayesian state-conditioned model may be effective in capturing the dynamic behavior of losses and delays on end-to-end packet channels, where channel means all the wired/wireless connection that connect the sender to the final user. The definition of a model capturing jointly losses and delays is highly desirable for designing and evaluating coding strategies, such as Multiple Description Coding (MDC), Forward Error Correction (FEC), Error Concealment (EC). Furthermore, the possibility of learning on-line the model parameters opens the way to design efficient content adaptation services.

A comprehensive model that jointly describes losses and delays has been proposed in [12][14]. It is based on an HMM structure and consider the case of a periodic source with constant packet size, see Fig. 1. A version of the Expectation-Maximization (EM) algorithm [3] for learning the model parameters was also described. It was assumed that the source has constant-packet size and constant inter-departure time, but several works showed that real data traffic present heavy-tail distributions for that parameters [10].

In this paper we want to extend the proposed model to the more general and realistic scenario where the source is not constrained to be periodic. The appropriate learning procedure is derived. A set of simulations based on measurements obtained on real packet links confirms how effective the proposed model can be.

II. INPUT/OUTPUT HIDDEN MARKOV MODELS

An Input/Output Hidden Markov Model (IO-HMM) [5] is a Dynamical Bayesian Network (DBN) that stochastically maps
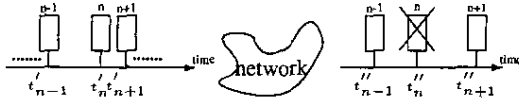
Fig. 2. Variable rate source scenario.

Fig. 3. The proposed model.

an input sequence $\mathbf{u} = [u_1, \ldots, u_L]^T$ into an output sequence $\mathbf{y} = [y_1, \ldots, y_L]^T$. They are in relation by a discrete state variable, the state sequence associated to $\mathbf{u}$ and $\mathbf{y}$ is denoted $\mathbf{x} = [x_1, \ldots, x_L]^T$. It can be viewed as a generalization of a classical HMM, or else as a statistical interpretation of the input-state-output equation that rules a dynamic system.

An IO-HMM is characterized by the set of parameters $\lambda = \{\mathbf{A}(u), \mathbf{B}(y)\}$, where

$$A_{i,j}(u) = Pr(x_n = s_j | x_{n-1} = s_i, u_n = u) , \qquad (1)$$

$$B_i(y) = Pr(y_n = y | x_n = s_i) . \qquad (2)$$

The learning problem for this kind of structure is very similar to Baum-Welch [3] algorithm for HMM and is based on recursive computation of appropriate Forward and Backward variables. The Forward and Backward variables for an IO-HMM are defined respectively as,

$$\alpha_n(i) = Pr(y_1, \ldots, y_n, x_n = s_i | u_1, \ldots, u_n, \lambda) , \qquad (3)$$

$$\beta_n(j) = Pr(y_{n+1}, \ldots, y_L | x_n = s_j, u_{n+1}, \ldots, u_L, \lambda) . \qquad (4)$$

They can be recursively computed by use of the following formulas,

$$\alpha_{n+1}(j) = \sum_{i=1}^{N} \alpha_n(i) A_{i,j}(u_{n+1}) B_j(y_{n+1}) , \qquad (5)$$

$$\beta_{n-1}(i) = \sum_{j=1}^{N} A_{i,j}(u_n) \beta_n(j) B_j(y_n) . \qquad (6)$$

The re-estimation formulas for the learning algorithm are based on the following quantities,

$$\begin{aligned}
\xi_n(i,j) &= Pr(x_n = s_i, x_{n+1} = s_j | \mathbf{y}, \mathbf{u}, \lambda) \\
&= \frac{\alpha_n(i) A_{i,j}(u_{n+1}) \beta_{n+1}(j) B_j(y_{n+1})}{Pr(\mathbf{y} | \mathbf{u}, \lambda)} , \qquad (7)
\end{aligned}$$

$$\gamma_n(i) = Pr(x_n = s_i | \mathbf{y}, \mathbf{u}, \lambda) = \frac{\alpha_n(i) \beta_n(i)}{Pr(\mathbf{y} | \mathbf{u}, \lambda)} , \qquad (8)$$

where

$$Pr(\mathbf{y} | \mathbf{u}, \lambda) = \sum_{i=1}^{N} \alpha_n(i) \beta_n(i) =$$

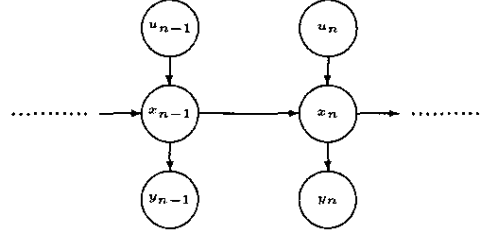$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_n(i) A_{i,j}(u_{n+1}) \beta_{n+1}(j) B_j(y_{n+1}) . \qquad (9)$$

III. THE MODEL

Our reference scheme is shown in Fig. 2, where a source traffic with constant packet size of N_b bits and variable inter-departure time is considered. Transmitted packets are numbered, $n = 1, 2, \ldots$. The network randomly cancels and delays packets according to current congestion. t'_n, t''_n, Δ_n and τ_n denote the departure time, the arrival time, the inter-departure time and the accumulated delay of the n-th packet, respectively,

$$\Delta_n = t'_n - t'_{n-1} , \qquad (10)$$

$$\tau_n = t''_n - t'_n . \qquad (11)$$

HMM framework showed accurate performance in modeling channel losses and delays in the case of periodic source, but some modifications for the model structure are needed to take into account the presence of the new variable Δ_n. Such a variable, representing the inter-departure time, characterizes the variable-rate behavior of the source.

IO-HMM's seem to be the natural extension to achieve the objective of modeling loss-delay behavior on a real communication channel with a variable-rate packet source.

Fig. 3 shows the proposed IO-HMM structure:

- u_n is the input of the model at time t'_n. It is function of the inter-departure time of the n-th packet, $u_n = f_K(\Delta_n)$, where $f_K(.)$ is a function such that $t \in (0, +\infty) \rightarrow f_K(t) \in \{1, 2, \ldots, K\}$,

- x_n represents the channel state at time t'_n distinguishing one among N levels of congestion $\{s_1, \ldots, s_N\}$,

$$A_{i,j,k} = Pr(x_n = s_j | x_{n-1} = s_i, u_n = k) , \qquad (12)$$

- y_n is a hybrid variable jointly describing loss or delay associated to the n-th packet, it is characterized by the following state-conditioned probability density function

$$B_i(y) = p_i \delta(t+1) + (1 - p_i) \frac{(y/\vartheta_i)^{\gamma_i - 1} e^{-(y/\vartheta_i)}}{\vartheta_i \Gamma(\gamma_i)} , \qquad (13)$$

where $\begin{cases} \tau_n = y_n & y_n > 0 \\ n^{th} \text{ packet is lost} & y_n = -1 \end{cases}$,

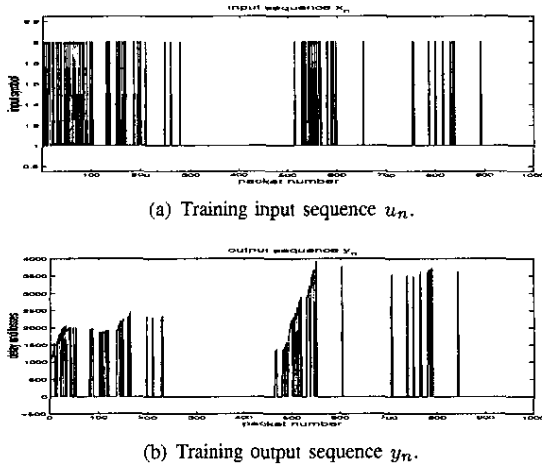that means that delays in state s_i are Gamma-distributed [7] with parameters γ_i and ϑ_i.

471

(a) Training input sequence u_n.

(b) Training output sequence y_n.

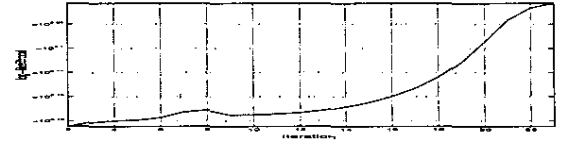Fig. 4. The training set for the IO-HMM.

Fig. 5. Trend of the log-likelihood during the learning procedure.

TABLE I

LOSS PROBABILITY BEFORE AND AFTER THE LEARNING PROCEDURE COMPARED TO STATISTICS OF THE TRAINING SEQUENCE.

	P_{loss}
training sequence	0.873
trained model	0.863

Though IO-HMM can have both discrete or continuous input variable, we consider the case of discrete input for the sake of simplicity. If the input variable is continuous, the transition matrix $A_{i,j}(u)$ is a parametrized matrix usually managed by use of the softmax functions [5]. The learning problem is much simpler if we consider a discrete input variable, in this case $A(u)$ is a 3—dimensional matrix, where $A_{i,j,k} = A_{i,j}(u = k)$. This is the reason behind introducing the trick of considering as input variable a function of the inter-departure time.

The single iteration of the learning procedure for this structure is based on the following formulas,

$$\hat{A}_{i,j,k} = \frac{\sum_{\substack{n=1 \\ u_{n+1}=u}}^{L-1} \xi_n(i,j)}{\sum_{\substack{n=1 \\ u_{n+1}=u}}^{L-1} \gamma_n(i)} , \qquad (14)$$

$$\hat{p}_i = \frac{\sum_{n=1}^{L} \rho_n(i)\beta_n(i)}{\sum_{n=1}^{L} \gamma_n(i)} , \qquad (15)$$

$$\hat{\gamma}_i \hat{\vartheta}_i = \frac{\sum_{n=1}^{L} \rho_n(i)\beta_n(i)y_n}{\sum_{n=1}^{L} \rho_n(i)\beta_n(i)} , \qquad (16)$$

$$\hat{\gamma}_i \hat{\vartheta}_i^2 = \frac{\sum_{n=1}^{L} \rho_n(i)\beta_n(i)(y_n - \gamma_i\vartheta_i)^2}{\sum_{n=1}^{L} \rho_n(i)\beta_n(i)} , \qquad (17)$$

where

$$\rho_n(j) = \sum_{i=1}^{N} \alpha_{n-1}(i)A_{i,j,u_n}p_j \left.\frac{\partial b_j(t)}{\partial p_j}\right|_{t=y_n} , \qquad (18)$$

and where $\alpha_n(i)$, $\beta_n(j)$, $\xi_n(i,j)$ and $\gamma_n(i)$, have been previously defined in Eqs. (3), (4), (7) and (8) respectively.

IV. EXPERIMENTAL RESULTS

Measures of losses and delays have been performed on real Internet channels using the software Distributed-Internet Traffic Generator (D-ITG) [13]. D-ITG was used to generate synthetic UDP traffic with constant packet size and Pareto-distributed inter-departure time, send it on real Internet paths,

and measure packet delays and losses at destination. We do not address the problem of packet-delay estimation assuming that the accuracy of measurements is acceptable. A small portion of the sequences was used as the training sequence to learn model parameters. The performance of the trained model was tested based on the remaining portions of sequences.

The experiment reported here is not to be considered exhaustive, but it suggests a very typical scenario in which losses and delays show the usual "burstiness". It was performed in March 2004 on the path between *Dipartimento di Informatica e Sistemistica, Università di Napoli "Federico II"* and *Dipartimento di Ingegneria dell'Informazione, Seconda Università di Napoli* in which the packet size is $N_b = 8000$ $bits$, while the inter-departure time is Pareto distributed with $\alpha_\Delta = 1.5$ and $\beta_\Delta = 1.5$ ms, $R = 1.6$ $Mbps$, (α_Δ and β_Δ denote the shape and scale parameters of Pareto distribution, while R is the average bit-rate). The input and observable sequences are composed of 41000 samples. The first $L = 1000$ samples were considered as the input and output sequence for the training. We used a model with $N = 2$ states, and $K = 2$ input symbols, more precisely we considered

$$f_2(\Delta_n) = \begin{cases} 1 & \Delta_n < \alpha_\Delta\beta_\Delta/(\alpha_\Delta - 1) \\ 2 & \Delta_n \geq \alpha_\Delta\beta_\Delta/(\alpha_\Delta - 1) \end{cases} . \qquad (19)$$

The input and output sequence used for the training are shown in Fig. 4. It can be noted how the channel behavior is strongly influenced by the input variable. Long sequences of $u_n = 1$, i.e. of high local bit-rate induce a very lossy behavior to the channel. This dependance strengthens the introduction of the input variable related to inter-departure time.

Fig. 5 shows the trend of log-likelihood evolution during our EM learning procedure. Fig. 6 and Table I show the result of the training with respect to delay and loss modeling, respectively. Loss probability and delay distribution (the continuous term of output variable) for the trained model are compared to the statistics of the training sequence. The histogram of the delays of the training sequence in Fig. 6 clearly shows a multi-modal behavior of the channel. The model uses its two states to discriminate the channel modes. The trained model
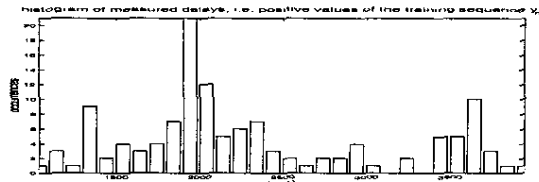
(a) Histogram of measured delays, i.e. of the positive value of training sequence y_n. The bin size is 50 ms.

(b) Continuous term of pdf of a 2—state model after learning procedure.

Fig. 6.   Example of delays statistics learning for a 2—state model.
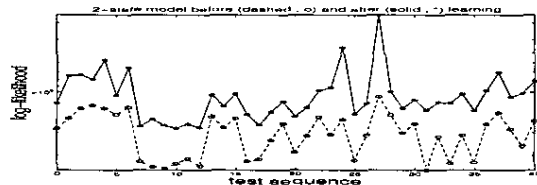

Fig. 7.   Log-likelihood for the test set of the trained and starting models.

shows to capture channel statistics.

Furthermore the trained model captures channel statistic that hold for future behavior of the channel itself. The 41000 input and output sequence samples were grouped in block of L samples. Each block-pair constitutes a test sequence (the first obviously is the training sequence). For each block-pair the likelihood of the trained and starting models was evaluated. Fig. 7 shows how the trained model has an almost stable behavior in modeling channel statistics.

The trained model was also considered as a channel simulator in terms of losses and delays. Fig. 8 show the auto-correlation of sequences generated by the trained and the starting models with respect to the auto-correlation of the training sequence. Fig. 9 show the throughput of the trained and the starting models with respect to the throughput of the channel during the training sequence. The throughput is intended as the fraction of packets that are delivered within a maximum allowed delay.

V.   CONCLUSION

In this paper we have proposed an IO-HMM whose objective is to model end-to-end packet channel behavior, jointly capturing losses and delays characteristics. The proposed model generalizes the HMM description of real channels introducing a joint stochastic modeling of losses and delays and is not restricted to the case of a periodic source. Preliminary results are very encouraging, as the IO-HMM is able to capture
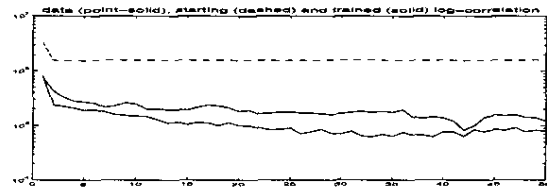

Fig. 8.   Correlation for the trained and starting models with respect to the training sequence.
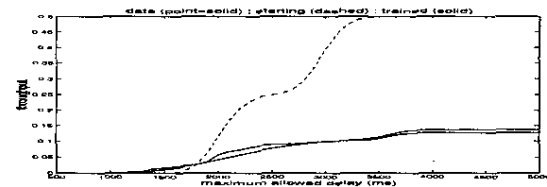

Fig. 9.   Throughput for the trained and starting models with respect to the training sequence.

losses and delays characteristics of the network with reference to the local bit-rate of the source.

REFERENCES

[1] E.N. Gilbert, *Capacity of a burst-noise channel*. Bell System Technical Journal, Vol. 39, pp. 1253-1265, Sept. 1960.

[2] E.O. Elliott, *Estimates of error-rate for codes on burst-noise channels*. Bell System Technical Journal, Vol. 42, pp. 1977-1997, Sept. 1963.

[3] L.R. Rabiner, *A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*. Proceedings of the IEEE, Vol. 77(2), pp. 257-285, Feb. 1989.

[4] M. Zorzi, R.R. Rao, and L.B. Milstein, *On the Accuracy of a First-Order Markov Model for Data Block Transmission on Fading Channels*. IEEE International Conference on Personal Communications, pp. 211-215, Nov. 1995.

[5] Y. Bengio, and P. Frasconi, *Input-Output HMM's for Sequence Processing*. IEEE Transactions on Neural Networks, Vol. 7(5), pp. 1231-1249, Sept. 1996.

[6] W. Turin, and R. van Nobelen, *Hidden Markov Modeling of Flat Fading Channels*. IEEE Journal on Selected Areas on Communications, Vol. 16(9), pp. 1809-1817, Dec. 1998.

[7] V. Paxson, *End-to-End Internet Packet Dynamics*. IEEE Transactions on Networking, Vol. 7(3), pp. 277-292, June 1999.

[8] W. Jiang, and H. Schulzrinne, *Modeling of Packet Loss and Delay and Their Effect on Real-Time Multimedia Service Quality*. 10th International Workshop on Network and Operating System Support for Digital Audio and Video, June 2000.

[9] K. Salamatian, and S. Vaton, *Hidden Markov Modeling for Network Communication Channels*. ACM Sigmetrics/Performance, Vol. 29, pp. 92-101, 2001.

[10] X. Yang, and A.P. Petropulu, *The Extended Alternating Fractal Renewal Processing for Modeling Traffic in High-Speed Communination Networks*. IEEE Transactions on Signal Processing, Vol. 49(7), pp. 1349-1363, July 2001.

[11] J. Liu, I. Matta, and M. Crovella. *End-to-End Inference of Loss Nature in a Hybrid Wired/Wireless Environment*. Technical Report, Boston University, Mar. 2002.

[12] P. Salvo Rossi, G. Romano, F. Palmieri, and G. Iannello, *A Hidden Markov Model for Internet Channels*. 3rd IEEE International Symposium on Signal Processing and Information Technology, Dec. 2003.

[13] S. Avallone, A. Pescapé, and G. Ventre, *Analysis and Experimentation of Internet Traffic Generator*. International Conference on Next Generation Teletraffic and Wired/Wireless Advanced Networking, Feb. 2004.

[14] P. Salvo Rossi, F. Palmieri, and G. Iannello, *HMM-based Monitoring of End-to-End Packet Channels*. 7th IEEE International Conference on High Speed Network and Multimedia Communications, June 2004.

473