Bayesian Modelling for Packet Channels

Pierluigi Salvo Rossi[1], Gianmarco Romano[2],
Francesco Palmieri[2], and Giulio Iannello[1]

[1] Dipartimento di Informatica e Sistemistica, Università di Napoli "Federico II"
Via Claudio 21
80125 Napoli, Italy
{salvoros,iannello}@unina.it
[2] Dipartimento di Ingegneria dell'Informazione, Seconda Università di Napoli
Real Casa dell'Annunziata Via Roma 29
81031 Aversa (CE), Italy
{francesco.palmieri,gianmarco.romano}@unina2.it

Abstract. Performance of real-time applications on network communication 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 modelled by a Dynamic Bayesian Network with an appropriate hidden variable that captures the current state of the network. In this paper we propose a Bayesian model that, trained with a version of the EM-algorithm, seems to be effective in modelling typical channel behaviors.

1 Introduction

Gilbert and Elliott works [1][2] on modelling burst-error channels for bit transmission showed how a simple 2-state Hidden Markov Model (HMM) was effective in characterizing real communication channels. As in the case of bit-transmission channels, end-to-end packet channels show burst-loss behavior.

Jiang and Schulzrinne [9] investigated lossy behavior of packet channels finding that a Markov model 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 as a memoryless phenomenon.

Salamatian and Vaton [10] found that a HMM trained with experimental data seems to capture channel loss behavior. They found that a HMM with 2 to 4 hidden states fits well experimental data.

These works suggest us that a Bayesian model should be effective in capturing the dynamic behavior of losses and delays on end-to-end packet channels. Our objective is to build a comprehensive model that jointly describes losses and delays.

B. Apolloni et al. (Eds.): WIRN VIETRI 2003, LNCS 2859, pp. 218–225, 2003.

2 The Model

Fig.1 shows our reference model with a periodic source traffic with inter-departure period T and fixed packet size. The network randomly cancels and delays packets according to current congestion.

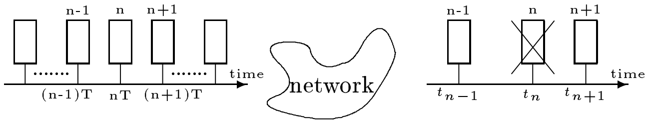


Fig. 1. End-to-end packet channel.

Let us number transmitted packets, $n = 1, 2, \ldots$, and let us denote with t_n and τ_n the arrival time and the accumulated delay of the *n-th* packet respectively, i.e.

$$\tau_n = t_n - nT. \tag{1}$$

We want to model the system in a way which carries information on current congestion that may determine variable loss rates and average delays. Loss phenomenon shows a bursty behavior, i.e. it cannot be thought as a memoryless stochastic process. Moreover, several works [6][8][9] showed that temporal delays seem not to be well modelled as independent identically distributed (iid) random variables, i.e. also delays, as losses, present a certain degree of memory. In real communication networks, losses and delays are strongly correlated; it has been observed [9] that in proximity of a loss, larger delays tend to occur.

Memory presence in phenomena we want to model suggests us to introduce a hidden state variable that stochastically influence losses and delays. State variable is hidden because our knowledge about it comes from observation of loss and delays, and there is no way to access directly to it.

In the following x_n, l_n and τ_n respectively will denote “state”, “loss” and “delay” at time nT,

$$\begin{aligned} x_n &\in \{s_1, s_2, \ldots, s_N\} \\ l_n &\in \{v_1, v_2\} \end{aligned} \tag{2}$$

where s_i is the *i-th* state of the network and v_1 (resp. v_2) means the absence (resp. presence) of a loss.

It should be noted that being in the presence of a loss, the delay has no real value, it can be considered infinite. For easy use we consider,

$$\begin{aligned} \tau_n/\{l_n = v_1\} &\in [0, +\infty) \\ \tau_n/\{l_n = v_2\} &= -1 \end{aligned} \tag{3}$$

Our reference Bayesian model is shown in Fig.2 where the arrows represent statistical dependence among variables. More specifically, the set of parameters characterizing the model is $\Lambda = \{\mathbf{A}, \mathbf{p}, f_1(\tau), f_2(\tau), \ldots, f_N(\tau)\}$, where

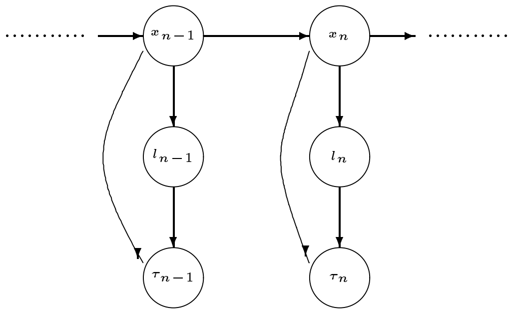


Fig. 2. The Bayesian model for packet channel.

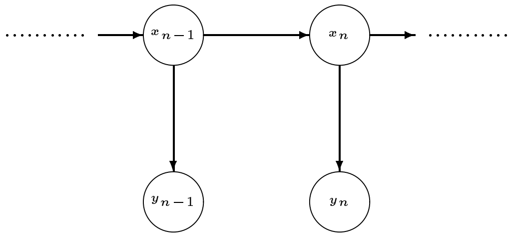


Fig. 3. Hidden Markov Model.

- **A** is the state transition matrix, i.e.

$$a_{ij} = Pr(x_{n+1} = s_j / x_n = s_i) \quad {}_{i \in \{1,2,\ldots,N\}}^{j \in \{1,2,\ldots,N\}} \tag{4}$$

- **p** is the loss probability vector, i.e.

$$\begin{cases} p_i = Pr(l_n = v_1 / x_n = s_i) \\ 1 - p_i = Pr(l_n = v_2 / x_n = s_i) \end{cases} \quad i \in \{1,2,\ldots,N\} \tag{5}$$

- $f_i(\tau)$ is the delay conditional pdf, i.e.

$$Pr(\tau_n > t / x_n = s_i, l_n = v_1) = \int_t^{+\infty} f_i(\tau) d\tau \tag{6}$$

The model can be reduced to a HMM, as seen on Fig.3, with a hidden variable x_n and an observable variable y_n that represents jointly loss and delay as

$$y_n = \begin{cases} \tau_n \ if \ l_n = v_1 \\ -1 \ if \ l_n = v_2 \end{cases} \tag{7}$$

Summarizing:

- x_n is a discrete random variable whose dynamic behavior follows Eq.(4)

– y_n is a hybrid random variable characterized, given $\{x_n = s_i\}$, by the following conditional pdf,

$$b_i(t) = p_i f_i(t) + (1 - p_i)\delta(t + 1) \tag{8}$$

It should be noted, as shown in Fig.4 that y_n is a hybrid variable obtained as a mixture of two components (one continuous, one discrete), when introduced the “trick” of associating $\tau_n = -1$ to a loss in order to have non-overlapping distributions for continuous and discrete components. The continuous component describes network delays behavior in the absence of losses, whereas the discrete component describes losses behavior.

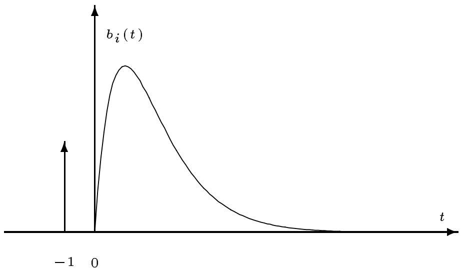


Fig. 4. An example of the conditional pdf $b_i(t)$ for the hybrid variable y_n.

If π is the stationary state probability distribution, i.e.

$$\pi_i = \lim_{n \to \infty} \{Pr(x_n = s_i)\} \tag{9}$$

the loss probability and the average delay of the model are:

$$Pr(loss) = \sum_{i=1}^{N} \pi_i (1 - p_i) \tag{10}$$

$$\overline{delay} = \sum_{i=1}^{N} \pi_i \int_0^{+\infty} t f_i(t) dt \tag{11}$$

3 Learning Parameters of the Model

The Expectation-Maximization algorithm [7] is an optimization procedure searching for a new set of parameters for a stochastic model according to improvements of the likelihood of a given sequence of observable variables. For structures like HMM of Fig.3 this optimization technique reduces to the Forward-Backward algorithm [3][4][5] studied for discrete and continuous observable variables with a broad class of allowed conditional pdf.

More specifically, given a sequence of observable variables $\mathbf{y} = (y_1, y_2, \ldots, y_K)^T$, and a set of parameters $\lambda = \{\mathbf{A}, \mathbf{p}, \mu\}$, where

$$\mu_i = E[\tau_n / \{x_n = s_i\}] = \int t f_i(t) dt \tag{12}$$

the update $\hat{\lambda} = \{\hat{\mathbf{A}}, \hat{\mathbf{p}}, \hat{\mu}\}$ of λ follows the recursions

$$\hat{a}_{ij} = \frac{\sum_{k=1}^{K-1} \alpha_k(i) a_{ij} b_j(y_{k+1}) \beta_{k+1}(j)}{\sum_{k=1}^{K-1} \alpha_k(i) \beta_k(i)} \tag{13}$$

$$\hat{p}_i = \frac{\sum_{k=1}^{K} \rho_k(i) \beta_k(i)}{\sum_{k=1}^{K-1} \alpha_k(i) \beta_k(i)} \tag{14}$$

$$\hat{\mu}_i = \frac{\sum_{k=1}^{K} \rho_k(i) \beta_k(i) y_k}{\sum_{k=1}^{K-1} \rho_k(i) \beta_k(i)} \tag{15}$$

where

$$\alpha_k(j) = \sum_{i=1}^{N} \alpha_{k-1}(i) a_{ij} b_j(y_k) \tag{16}$$

$$\beta_k(i) = \sum_{j=1}^{N} a_{ij} b_j(y_{k+1}) \beta_{k+1}(j) \tag{17}$$

are the forward and backward partial likelihood, and where

$$\rho_k(j) = \sum_{i=1}^{N} \alpha_{k-1}(i) a_{ij} p_j \left. \frac{\partial b_j(t)}{\partial p_j} \right|_{t=y_k} \tag{18}$$

The iteratively procedure will reach a local maximum point of the likelihood function,

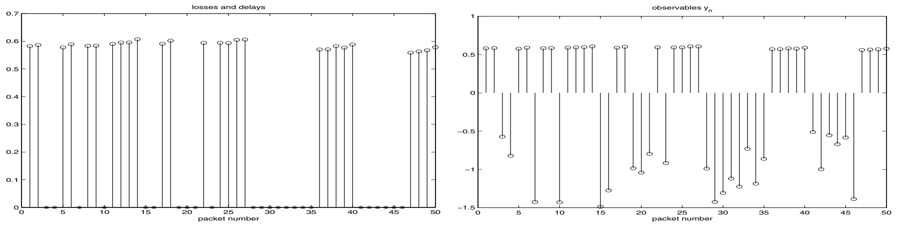
$$L(\mathbf{y}; \lambda) = Pr(\mathbf{y}/\lambda) = \sum_{i=1}^{N} \alpha_K(i) \tag{19}$$

which typically depends on the starting point λ. When necessary, repeated starts with different initial conditions provide the global solution.

The problem of the Dirac-impulse in the conditional pdf (8) was avoided considering a modified function

$$\tilde{b}_i(t) = p_i f_i(t) + (1 - p_i) g(t) \tag{20}$$

where $g(t)$ is any pdf such that $g(t) = 0 \ , \quad \forall \ t \geq 0$, in order to have non-overlapping supports between $f_i(t)$ and $g(t)$. Obviously, while the set $\{f_i(t)\}_{i=1}^N$ will be adjusted by the iterative procedure, $g(t)$ will remain unchanged, as only its area is relevant. This means that losses, in the algorithm have to be randomized according to $g(t)$.



(a)Losses and delays sequence. (b) Corresponding observable sequence.

Fig. 5. Portion of a measured trace on real network.

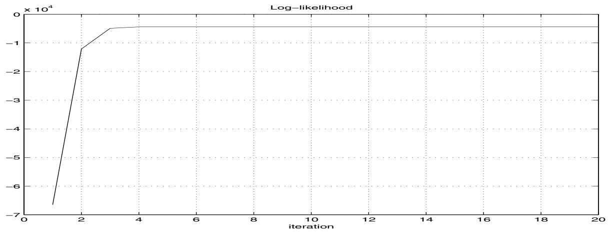


Fig. 6. Log-likelihood trend.

4 Experimental Results

Measures of losses and delays have been performed between the *Dipartimento di Informatica e Sistemistica, Universitá di Napoli "Federico II"*, and the *Dipartimento di Ingegneria dell'Informazione, Seconda Universitá di Napoli*, using the software Internet Traffic Generator (ITG) [12].

ITG, a new version of Mtools [11], can generate both traffic at transport layer and "layer 4-7". It implements both TCP and UDP traffic generation according to several statistical distributions both for inter-departure times and packet sizes. ITG allows simulations of complex traffic sources furnishing information about transmitted and received packets.

The characteristics of generated traffic are: inter-departure period $T = 5 \cdot 10^{-3}$ *sec.* and packet size of 1000 *bytes*, $(bit - rate = 1.6\ Mbps)$.

A typical trace obtained is shown in Fig.5(a), while in Fig.5(b) the corresponding sequence used for learning procedure is shown.

The learning procedure was applied on observable sequences of 500 to 1000 samples, finding in reasonable time (10*sec.*) acceptable estimation of network behavior. Fig.6 shows a typical trend of the log-likelihood obtained in the learning procedure.

Our choice of conditional pdf's for delays was Gamma distributions, as suggested by several works [6][8],

$$f_i(t) = \frac{t^{\gamma_i - 1} e^{-t}}{\Gamma(\gamma_i)} u(t), \qquad (21)$$

while losses was randomized according to a uniform distribution, i.e.

Table 1. Example of parameters learning with a 2-states model.

	$Pr(loss)$	$\overline{delay}$
measured	0.494	543.69 ms
starting model	0.117	131.06 ms
trained model (10$iterations$)	0.495	492.58 ms
trained model (20$iterations$)	0.494	565.81 ms

Table 2. Example of parameters learning with a 3-states model.

	$Pr(loss)$	$\overline{delay}$
measured	0.494	543.69 ms
starting model	0.365	143.12 ms
trained model (10$iterations$)	0.495	492.58 ms
trained model (20$iterations$)	0.494	536.92 ms

$$g(t) = \begin{cases} 1 & t \in [-3/2, -1/2] \\ 0 & t \in R - [-3/2, -1/2] \end{cases} \tag{22}$$

Tabs.1 and 2 show loss probability and average delay of a measured trace used as training sequence, and of a model with 2 and 3 states before and after learning procedure, according to (10),(11).

Actually we are investigating on generalization capability. Preliminary tests show that the model is able to follow the channel behavior until its characteristics can be considered almost stationary.

5 Conclusion and Future Work

In this paper we have proposed a Bayesian Network 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 memory stochastic modelling of delays. Preliminary results are encouraging and future works will be focused on model improvements and coding strategies.

Acknowledgement

This work has been carried out partially under the financial support of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) in the framework of the FIRB Project "Middleware for advanced services over large-scale, wired-wireless distributed systems (WEB-MINDS)".

References

1. Gilbert, E.N.: Capacity of a burst-noise channel. In: Bell System Technical Journal, Vol. 39 (Sept. 1960), 1253–1265

2. Elliott, E.O.: Estimates of error-rate for codes on burst-noise channels. In: Bell System Technical Journal, Vol. 42 (Sept. 1963), 1977–1997
3. Liporace, L.A.: Maximum Likelihood Estimation for Multivariate Observations of Markov Sources. In: IEEE Transactions on Information Theory, Vol. IT-28(5) (Sept. 1982), 729–734
4. Juang, B.H., Levinson, S.E., Sondhi, M.M.: Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains. In: IEEE Transactions on Information Theory, Vol. IT-32(2) (Mar. 1986), 307–309
5. Rabiner, L.R.: A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In: Proceedings of the IEEE, Vol. 77(2) (Feb. 1989), 257–285
6. Bolot, J.C.: Characterizing End-to-End Packet Delay and Loss in the Internet. In: Journal of High-Speed Networks, Vol. 2(3) (Dec. 1993), 305–323
7. Bilmes, J.A.: A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report, ICSI-TR-97-021, University of Berkeley,CA, 1998.
8. Paxson, V.: End-toEnd Internet Packet Dynamics. In: IEEE Transactions on Networking, Vol. 7(3) (June 1999), 277–292
9. Jiang, W., Schulzrinne, H.: Modeling of Packet Loss and Delay and Their Effect on Real-Time Mulrimedia Service Quality. In: 10th International Workshop on Network and Operating System Support for Digital Audio and Video, June 2000.
10. Salamatian, K., Vaton, S.: Hidden Markov Modeling for network communication channels. In: ACM Sigmetrics/Performance, Vol. 29 (2001), 92–101
11. Avallone, S., D'Arienzo, M., Esposito, M., Pescapè, A., Romano, S.P., Ventre, G.: Mtools. In: Networking Column on IEEE Network, Vol. 16(5) (Sept./Oct. 2002), pag.3
12. Avallone, S., Pescapè, A., Ventre, G.: Analysis and Experimentation of Internet Traffic Generator. Submitted to: ACM Workshop on Models, Methods and Tools for Reproducible Network Research