A Framework for the Analysis of Adaptive Voice over IP

C. Casetti*, J.C. De Martin†, M. Meo*

* Dipartimento di Elettronica
† IRITI-CNR
Politecnico di Torino
Corso Duca degli Abruzzi 24, I-10129 Torino, Italy
E-mail: {casetti,demartin,michela}@polito.it

Abstract— **In this paper, we present a framework for the analysis of a set of adaptive variable-bit-rate voice sources in a packet network. The instantaneous bit rate of each source is determined by an end-to-end control mechanism that, based on measurements of packet delay and loss rate, selects the rate that best matches current network conditions. Several such algorithms can be analyzed with the proposed framework, which consists of a detailed Markovian model of the source behavior and of an approximate description of the interaction between the sources and the underlying network. The model of the source takes into account time intervals during which a connection is active as well as intervals of inactivity; within a given conversation, it also models on/off (speech/silence) periods. The interaction of a source with the rest of the system is derived through an iterative procedure that evaluates the feedback that a source receives from the network. A case study presenting the results relative to an adaptive system transmitting at bit rates typical of widely used speech coding standards (64 kb/s, 13 kb/s and 8 kb/s) illustrates the proposed framework.**

I. Introduction

In recent years, a large amount of interest has been focused on multimedia communications over IP networks. Voice communication has received special attention and the possibility to offer telephony services alternative to the standard PSTN-based ones has often been considered.

The main factors that determine such interest include: the widespread adoption of IP networks for data transmission at all levels of industry and academia; the astounding expansion of the Web; new speech coding standards providing good or even toll quality at unprecedented low bit-rates. Besides, IP telephony has the potential to offer a wide range of value-added services, such as high-quality audio and integration with web-based applications, white-boards, or other multimedia tools.

Although the issue of voice transmission over packet networks is hardly new (see, for instance, [1]), the specific problem of how to best deliver voice over IP networks is relatively novel. New protocols specifically designed to aid the transmission of real-time data over IP networks (e.g., RTP, [2]) offer new tools. However, several issues still need attention. For instance, there is no speech coding standard specifically designed for Voice over IP applications. Variable bit-rate speech coding solutions, very apt in a packet-switching context, have not been much investigated. Techniques to cope with packet loss have been mainly derived from other contexts, such as wireless applications, without fully exploiting the potential of solutions specifically thought for packet transmission over IP-based networks, as, instead, attempted in [3], [4]. And perhaps even more conspicuously, there are few results about the overall performance of Internet telephony systems (see [5] for a discussion of implications of Internet loss and delay characteristics for the deployment of IP telephony).

This paper tries to outline a methodology for the design of a specific class of Internet Telephony solutions, that is, solutions based on variable bit-rate speech coders that adapt to network conditions. We might call this approach adaptive Voice over IP, or AVoIP for short. In this context, we present a framework to address system-level questions, i.e., how should an application-level adaptive solution degrade the speech quality when the network is congested? What performance metrics should drive the algorithm? Where should thresholds be positioned?

The paper is organized as follows: Section II describes the requirements of a variable-bit-rate IP telephony application and proposes a simple adaptive algorithm; Section III further details the algorithm and introduces an analytical methodology to describe the behavior of variable bit-rate voice sources subject to a rate control algorithm; the performance indices that can be derived from our model are briefly outlined in Section IV, and applied to a case study in Section V. Section VI concludes the paper.

II. Network-driven end-to-end control over variable bit-rate voice sources

Let us begin by illustrating a plausible scenario for an adaptive IP telephony application. We are focusing on a situation where the large majority of traffic loading the network belongs to IP phone calls. Also, we are assuming that these applications run over a standard RTP/UDP/IP protocol stack. Since the UDP transport stack does not implement any end-to-end flow control, the application itself should determine whether the network is congested, and take appropriate action.

The main goal of the algorithm, as can be guessed, is reducing the load on the network when queue build-ups occur. If a large part of network traffic is composed of variable-bit-rate-regulated voice traffic, and the algorithm regulating each source is engineered so as to react to specific "warnings" such as increasing delay or sudden packet loss, then a control strategy may successfully reduce the occurrence of congestion.

The topic of variable-rate adaptive schemes for multimedia streams has been addressed by several researchers in the past. In [6], Cox and Crochiere foreshadow adaptive schemes for variable-bit-rate coders by analyzing the behavior of a buffer control mechanism that has sources react to buffer overflows and underflows. The idea of an end-to-end feedback control is already present in the work by Bially et al. [7], where network-driven voice sources based on embedded speech coding schemes are studied, although the involvement of intermediate nodes is still required. It should be noted that detecting the onset of congestion at an intermediate node is not a viable solution in our case, since it would require that all routers have this capability and that an ad-hoc signaling mechanism be set up (such as ABR congestion control in ATM). An end-to-end form of control is, therefore, desirable. Such a solution was studied by Bolot and Turletti who, in [8], proposed and tested an adaptive video coder based on the state of the network. A voice-oriented approach can be found in the works by Busse et al. [9] and Sisalem and Schulzrinne [10], who employ a sender-based packet loss rate estimation scheme to drive the transmission rate of the real-time source; the latter, in particular, also attempts to have the adaptive real-time source behave fairly towards concurring TCP connections. An analytical model for an end-to-end feedback control mechanism for variable bit-rate coders was developed by Yin and Hluchyj in [11]. The latter work is similar in spirit to the content of this paper.

Generally speaking, the feedback can be expressed as aggregate rate, traffic delay or loss rate at the receiver. Using the aggregate rate is hardly feasible since the receiver has no perception of it. However, a rise in the aggregate rate, close to the network capacity, can be reflected by increasing delays seen at the receiving end. In turn, a delay increase is often followed by packet loss within the next few round-trip times. Therefore, being able to detect changes in the delay measure patterns, and acting on such indication, can prevent packet loss, provided that a sufficient number of sources apply the same strategy.

If, from a preliminary analysis of the network, we assume to know the relation between the delay that packets experience and the aggregate traffic of the network, we can derive an indication of the speed with which sources react to the end-to-end control. The possibility to introduce a relationship of this kind is one of the most appealing features of our methodology, as will be explained below. Indeed, the effectiveness of the control is strongly related to the delay-load function. The control is mostly needed under heavy load conditions, and in these conditions the packet delays are high, and thus the reactivity is low.

An alternative control scheme might rely on a receiver-based measure of packet losses. However, the loss of a single packet, or of batches of packets, is an indication of a severe network congestion: any action taken following this notification may be belated, thus leading to a prolonged congestion.

To summarize, the rate control algorithm described in this paper does not rely on explicit congestion indications by either the intermediate nodes or by receivers. Rather, it uses the information carried by cyclic RTCP receiver reports to let the source know the state of the on-going connection. The fundamental idea is that the source coder should gracefully reduce its rate when the delay has been observed to have considerably increased above a ‘high-mark threshold’. It should switch to higher rates if the delay decreases below a ‘low-mark threshold’. We will refer to this control scheme as “network-driven”, thus acknowledging the role that the state of the network plays in regulating the transmission rate.

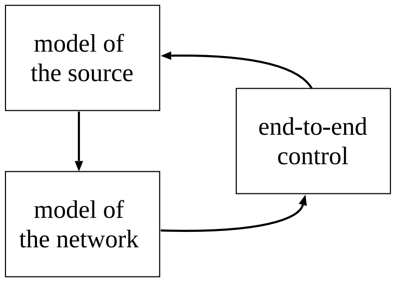

Fig. 1. Steps of the proposed procedure

Clearly, since we are looking at an IP telephony scenario, both parties can be considered source and destination. In order to simplify the description, we will decouple these interactions and only consider a communication between a generic source and its destination.

While we are aware that our model cannot conceivably describe all PDU-level interactions that take place inside the network, our end-to-end application-level approach masks the details of the underlying layers, leaving their representation to the “signaling” information conveyed by receiver reports (or lack thereof).

III. Analytical approach

In this Section, we describe an analytical approach to the performance evaluation of network-driven voice sources operating in the environment described in the previous Section.

The basic ideas underlying the approach are sketched in Fig. 1. The procedure is iterative and it is articulated in the following steps. First, we analyze a source in isolation from the rest of the system. The presence of the other sources and of the network is taken into account in an approximate way, as explained below. Then, from the results of this analysis, we evaluate the behavior of a network which comprises several sources. Finally, we examine the impact that the end-to-end control mechanism has on a single source. We employ this evaluation to update the description of the source in isolation and then we repeat the analysis. The procedure stops when a given convergence condition is reached.

The basic assumptions are:

- every source is modeled in isolation and is statistically independent from the other sources;
- the interaction of the sources with the rest of the network is perceived through the feedback information conveyed back to each source;
- the model dynamics are determined by a number of *driving stochastic processes*:
  - new call generation;
  - call termination;
  - source bit-rate increase;
  - source bit-rate decrease.

In the next subsection, we detail the model we have outlined above.

A. Model of the source in isolation

In order to model the voice source we need to consider two different aspects governing its behavior. The first aspect refers to the user's dynamics, the second one is related to the end-to-end control mechanism that depends on the working conditions of the network.

According to commonly accepted models of user's behavior [12], [13], [11], a conversation between two users can be seen as a sequence of alternating intervals of silence and talk-spurts. Adopting this model of user's dynamics, we identify three types of behaviors for the voice source:

- the source is *idle* when not engaged in a conversation;
- the source is *talking* when, during a conversation, the user is talking, hence the source is transmitting;
- the source is *silent* when, during a conversation, the user is silent; the source is transmitting only signalling information.

We make the following assumptions about the processes involved:

- the time between the setup of two consecutive conversations is a random variable (r.v.) with negative exponential distribution with parameter λ;
- the call holding time is a r.v. with negative exponential distribution with parameter μ;
- the user's talking period is a r.v. with negative exponential distribution with parameter β;
- the user's silent period is a r.v. with negative exponential distribution with parameter α;
- signalling information is not modeled.

We derive a three-state continuous-time Markov chain, with idle state I, talking state t and silent state s. When in the idle state, a new connection may be originated by the user itself, or by a remote user. The chain can move from the idle state I either to talking state t (the user is calling), or to silent state s (the user is answering a call); both transitions have rate $\lambda/2$.

We now enrich the description of the source behavior by introducing the possibility of operating at variable bit rates. We assume that, when a call is in progress, the speech signal can be encoded at one of N different bit rates: let B_N be the highest and B_1 be the lowest rate. Driven by the feedback messages received by the end-to-end control, the source changes the transmission rate. We assume that, at connection setup, the source transmits at the highest bit rate, though different assumption may of course be done. Based upon delay and loss measurements carried out at the receiver, and relayed back to the source, each of the parties may decide to alter the generation rate, as will be explained below.

Introducing variable bit rate mechanisms in the behavior of the voice source, the states t and s can be expanded and the Markov chain is as in Fig. 2.

The idle state is labeled I, the other states are labeled by a pair of values (l, m) where $l \in \mathcal{L} = \{1, 2, \cdots, N\}$ is the label of the current transmission rate B_l, and $m \in \mathcal{M} = \{t, s\}$ is the type of activity of the user, who can be either talking or silent. The state space $\mathcal{S}$

$$\mathcal{S} = \{I\} \cup \{\mathcal{L} \times \mathcal{M}\} \tag{1}$$

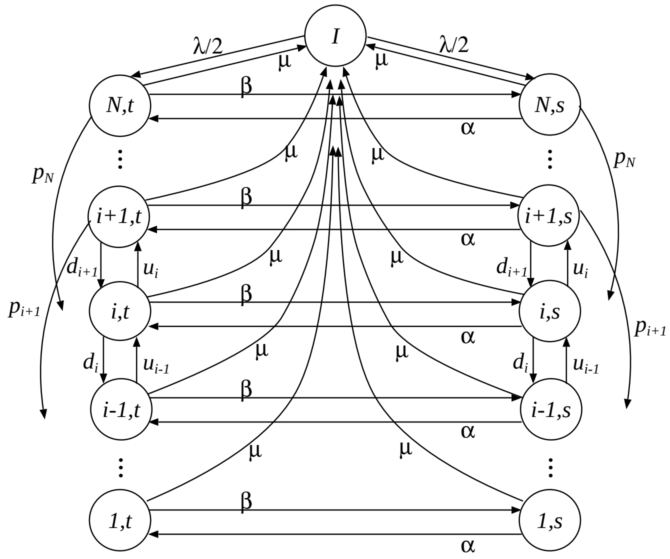

Fig. 2. Continuous-time Markov chain model of the variable bit rate source

has cardinality $2N + 1$. In order to simplify the notation we number the states $i \in \mathcal{S}$ from 0 up to $S = 2 \cdot N$; state 0 denotes the idle state.

From the flow balance equations we derive the equilibrium state probabilities of the Markov chain:

$$\underline{\Pi} = \{\pi_j\} \quad \text{with} \quad j \in \mathcal{S} \tag{2}$$

For the present time, let us assume that we know the transition rates d_i and u_i according to which the chain moves from states $(i, \cdot)$ to $(i-1, \cdot)$ and to $(i+1, \cdot)$. We will later discuss how we can derive the value of d_i and u_i, and what feedback metrics can be used to trigger the transitions.

B. Model of the network

We now outline the model of a network with K sources exhibiting the behavior described so far.

A detailed model of the network requires an ad-hoc description of the state of each voice source. However, the description of a realistic scenario with a large value of K would result in an overwhelming state space, thus making the analysis intractable. As previously mentioned, instead of keeping track of the state of all sources, we assume the statistical independence of the sources.

In our model, the state $\underline{n}$ of a network composed of K sources behaving according to the model described above is the vector:

$$\underline{n} = (u_0, u_1, \cdots, u_S)$$

where u_i represents the number of sources in state i, $i \in \mathcal{S}$.

The state space for the network is:

$$\mathcal{N} = \{\underline{n} = (u_0, u_1, \cdots, u_S) \,|\, \sum_{i=0}^{S} u_i = K\}$$

Given the assumption of statistical independence of the sources, the probability that the network is in state $\underline{n}$ =

$(u_0, u_1, \cdots u_S)$ is derived from the steady-state probability of the model describing a source in isolation, as follows:

$$P(\underline{n}) = \begin{pmatrix} K \\ u_0 \end{pmatrix} \begin{pmatrix} K - u_0 \\ u_1 \end{pmatrix} \cdots \begin{pmatrix} K - \sum_{i=0}^{S-1} u_i \\ u_S \end{pmatrix} \cdot \pi_0^{u_0} \pi_1^{u_1} \cdots \pi_S^{u_S}$$

In the latter expression, $P(\underline{n})$ is evaluated as the probability of all the possible combinations of having u_0 out of K sources in state I, u_1 sources in state 1 and so on.

The aggregate generated traffic $T(\underline{n})$ when the network is in state $\underline{n}$ is given by:

$$T(\underline{n}) = \sum_{j=0}^{S} f(j) B_j u_j$$

where $f(j)$ is equal to 1 if j is a talking state and 0 otherwise.

We can further characterize the network model by taking into account the delay experienced by packets, as suggested in Section II. We recognize two main contributions to the packet delay D. The first one, which we may call the propagation delay, D_p, is constant with respect to the load; it can be seen as the delay perceived when the network is unloaded, and it also includes the processing delay at both ends. The second contribution is the queueing delay, D_q, which represents the time a packet spends in the buffers of the nodes and depends on the traffic in the network. We can then write the delay experienced by the packets when the system is in state $\underline{n}$ as:

$$D(\underline{n}) = F(T(\underline{n})) = D_p + D_q(T(\underline{n})) \tag{3}$$

C. Interaction between a source and the network

The source-network interaction is modeled through the end-to-end control of the bit rate of the sources, aimed at limiting network congestion and thus packet delay.

In our assumptions, when the measured delay is above a given threshold D_h, the source is required to lower its bit rate; when, on the contrary, the measured delay is below a given threshold D_l, the source can increase its bit rate in order to provide better speech quality of service.

The probability that the network is operating in critical conditions for which the delay that packets experience is over the threshold D_h is:

$$P_h = \sum_{\underline{n} \in \mathcal{S}_h} P(\underline{n}) \tag{4}$$

where

$$\mathcal{S}_h = \{\underline{n} = (u_0, u_1, \cdots, u_S) | D(\underline{n}) > D_h\}$$

is the set of states for which the aggregate rate is such that the packet delay is above D_h. Observe that $\mathcal{S}_h$ is a subset of the state space $\mathcal{N}$ of the network, $\mathcal{S}_h \subset \mathcal{N}$. Similarly, the probability that the network is operating in underload conditions, below the threshold D_l, is:

$$P_l = \sum_{\underline{n} \in \mathcal{S}_l} P(\underline{n}) \tag{5}$$

where

$$\mathcal{S}_l = \{\underline{n} = (u_0, u_1, \cdots, u_S) | D(\underline{n}) < D_l\}$$

As previously mentioned, the probabilities P_h and P_l determine the source's choice of increasing or decreasing the bit rate. We can now derive the values of the transition rates d_i and u_i of the source model (see Fig. 2).

The source reaction to critical conditions of the network depends upon the round trip time: assuming homogeneous traffic, the round trip time is twice the packet delay in state $\underline{n}$, $2 \cdot D(\underline{n})$. Evaluating the reaction rate as the inverse of the reaction time, we can derive the rates d_i and u_i from the weighted sums:

$$d_i = \sum_{\underline{n} \in \mathcal{S}_h} \frac{P(\underline{n})}{2 \cdot D(\underline{n})} \qquad u_i = \sum_{\underline{n} \in \mathcal{S}_l} \frac{P(\underline{n})}{2 \cdot D(\underline{n})} \tag{6}$$

Other end-to-end control mechanisms can be easily analyzed simply modifying the way d_i and u_i are computed. For example, for the control mentioned in Section II which requires that sources decrease their transmission rate only in case of packet losses, d_i is computed from (6) using $\mathcal{S}_b$ instead of $\mathcal{S}_h$, where:

$$\mathcal{S}_b = \{\underline{n} = (u_0, u_1, \cdots, u_S) | T(\underline{n}) > B_c\} \tag{7}$$

is the set of states in which the sources generate more traffic than the network capacity B_c. Some results comparing the latter scheme with the above delay-based approach will be shown in Section V.

D. Model solution

We now have all the elements needed to find a solution: the model of a source in isolation, the model of the whole network and the mechanism describing the source-network interaction.

The solution is derived by an iterative procedure since the transition rates d_i and u_i are, in fact, a function of the model solution itself.

The iterative procedure stops when a given convergence condition is reached. If $u_i^{(j)}$ and $d_i^{(j)}$ are the values of the transition rates at the $j-$th step of the procedure, it is required that:

$$err_i = max \left(\frac{|u_i^{(j)} - u_i^{(j-1)}|}{u_i^{(j)}}, \frac{|d_i^{(j)} - d_i^{(j-1)}|}{d_i^{(j)}} \right) < \epsilon \;\; \forall\, i \in \mathcal{S}$$

where ϵ is the maximum acceptable error.

IV. Performance evaluation

Several interesting performance indices can be computed from the model.

We define the packet loss probability, P_c, as the fraction of packets that are lost due to the aggregate traffic being greater than the network capacity:

$$P_c = \frac{\sum_{\underline{n} \in \mathcal{N}} P(\underline{n}) max\,(0, T(\underline{n}) - B_c)}{\sum_{\underline{n} \in \mathcal{N}} T(\underline{n}) P(\underline{n})}$$

We can also compute the probability P_b of being in one of the states where the system is losing packets and the users are thus receiving a deteriorated quality of service:

$$P_b = \sum_{\underline{n} \in \mathcal{S}_b} P(\underline{n})$$

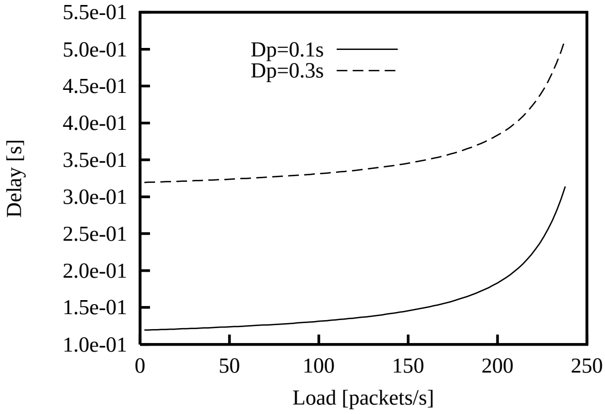

Fig. 3. Packet delay versus total load in the network for the two considered cases: D_p =0.1 s and D_p =0.3 s

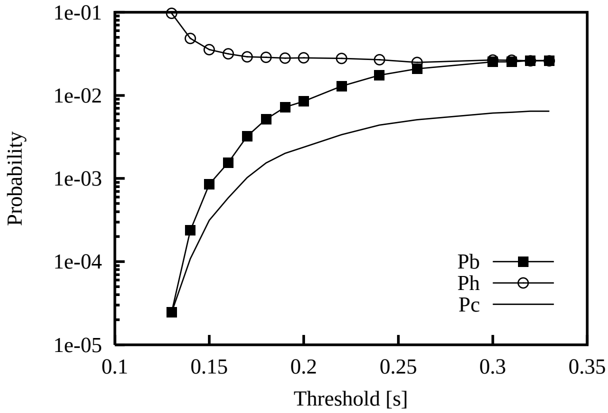

Fig. 4. Probabilities P_b and P_h versus D_h (left plots); packet loss probability P_c versus D_h (right plots)

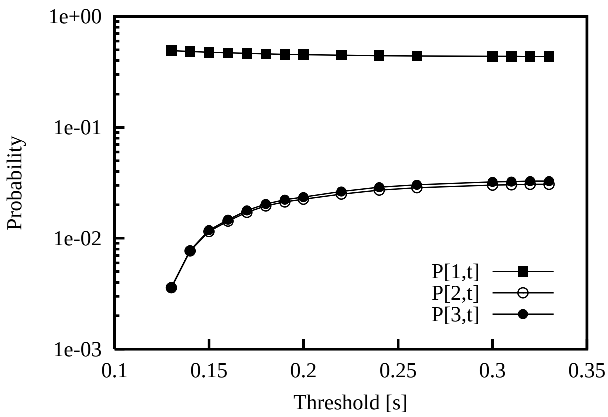

Fig. 5. Steady-state probabilities of a source, $\pi_{(1,t)}$, $\pi_{(2,t)}$, $\pi_{(3,t)}$, versus D_h

where $\mathcal{S}_b$ is given in (7).

The probabilities P_h and P_l that the system is above the threshold D_h or below D_l are given in (4) and (5).

Other interesting measures are the probabilities that a source is silent or is generating at rate B_i. These probabilities are given by the steady-state probability vector Π.

These indices are used next to evaluate a case study.

V. Case study

In this Section we show some results about the performance evaluation of a network in which sources can switch between three bit-rates, corresponding to widely used telephone bandwidth speech coding standards:

1. 8 kb/s (e.g., ITU-T G.729)
2. 13 kb/s (e.g., ETSI GSM-EFR)
3. 64 kb/s (e.g., ITU-T G.711)

A hypothetical variable bit-rate system operating at such bit rates can be viewed as a system that always delivers “toll quality,” but with different levels of complexity, delay and robustness. The goal of this Section is not a complete analysis of the considered system, but, rather, an indication that interesting performance evaluation indices can be derived through the proposed approach.

The average duration of a call is 180 s (the standard assumption in telephony); the average user’s talking and silence periods are respectively equal to 1.8 s and 1.2 s, as suggested in [12], [13].

We computed the delay-load relationship expressed by (3) in the following way. The propagation delay D_p is assumed to be a constant value either equal to 0.1 s or to 0.3 s, while the queueing delay is derived from the analysis of a network with five $M/M/1/B$ queues in a linear topology. The queue capacity, B, is assumed to be equal to 50 packets. Packets are 1000 bytes long and the transmission rate of the link (unless otherwise stated) is assumed to be equal to 2 Mb/s, or, equivalently, 250 packets per second.

In Fig. 3, we plot the total delay versus the load for the two considered configurations of the system, corresponding to $D_p = 0.1$ s and $D_p = 0.3$ s. Results shown below in Figs. 4–7 refer to $D_p = 0.1$ s, while comparisons between the cases $D_p = 0.1$ s and $D_p = 0.3$ s are shown in Fig. 8.

Fig. 4 shows the probability P_b that the network is in “lossy” states (i.e. the network is congested), the probability P_h to be in states where the delay is higher than the threshold D_h and the packet loss probability P_c. The three performance indices are plotted versus increasing values of D_h. Since the control is more tighter and reactive for smaller values of D_h, the loss probability increases as D_h increases.

Fig. 5 reports the probability that a source is in one of the talking states: $(1,t)$, $(2,t)$ and $(3,t)$. For the considered configuration, since the system is under high load conditions, the probability of being in $(1,t)$ (i.e. 8 kb/s) is much higher than the others.

The probabilities P_b and P_h versus the network capacity (expressed in packets per second) are reported in Fig. 6. The threshold D_h is fixed and it is assumed to be equal to 0.2 s. As expected, the probabilities decrease for increasing values of the capacity.

Fig. 7 justifies the choice of a delay-based congestion control scheme: when compared to a loss-based algorithm (“no D_h” curve) such as the one outlined in Section III, both versions of the delay-based algorithms (with $D_h = 0.20\ s$ and $D_h = 0.25\ s$) exhibit a lower loss probability.

Finally, in Fig. 8 we compare the performance (in terms of P_h and P_b) of the two configurations of the system for which the curves of delay versus load are shown in Fig. 3. P_h and P_b are

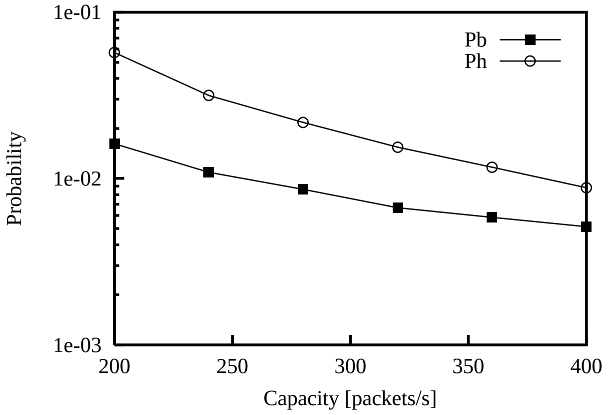

Fig. 6. Probabilities P_b and P_h versus the network capacity

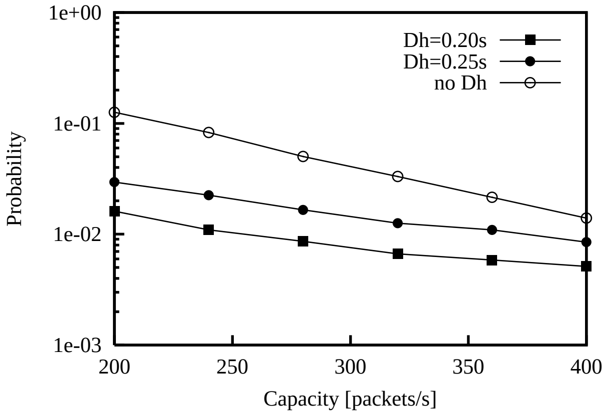

Fig. 7. Packet loss probabilities versus the network capacity, for different delay threshold and for a loss-based control algorithm

computed for increasing values of the threshold D_h. Since the two configurations have different propagation delays D_p, and thus different ranges of D_h, a fair comparison is achieved by expressing D_h as a function of both D_p and D_q^{MAX}, the maximum queueing delay. Therefore, we defined $D_h = D_p + \tau \cdot D_q^{\mathrm{MAX}}$, and plotted values of the queueing delay factor τ on the x-axis of Fig. 8. Two observations are in order. First of all, the loss probability is smaller for $D_p = 0.1\ s$, since the control is tighter for smaller values of delay. Secondly, the “critical” values for τ are below 0.2 (which sets the threshold at the propagation delay plus 20% of the maximum queueing delay). Values above 0.2 yield a marginal gain in terms of probability of being in lossy or over-the-threshold states.

VI. Conclusions

We proposed a framework for the analysis of adaptive, variable-bit-rate sources operating in a packet network environment; their coding rate is assumed to be adjusted according to an end-to-end control mechanism based on measurements of the packet delay. The flexibility of the proposed framework makes the analysis of the performance of various control algorithms possible.

The proposed methodology consists in the separate analysis of a source in isolation and of the working conditions of the network. The source model is detailed and the interaction with the rest of the system is represented by means of the approximate description of the feedback that the source receives from the network.

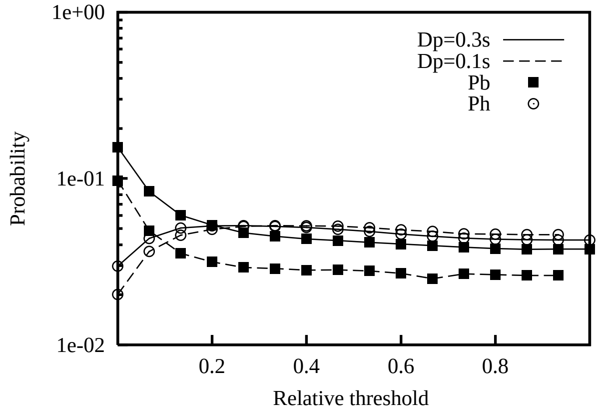

Fig. 8. Comparison of P_h and P_b versus D_h (expressed as fraction of the range of values which D_h can assume) for D_p equal to 0.1 s and 0.3 s

The framework is illustrated considering a specific end-to-end control mechanism applied to a variable bit-rate system operating at three typical telephone-bandwidth bit rates—64, 13 and 8 kb/s. The analysis is simple and efficient, so that the framework comes across as a flexible tool to evaluate the performance of systems whose complete and detailed analysis is difficult and time-consuming. It can be useful in the preliminary phases of study of complex systems, for comparisons between different control algorithms or for tuning the parameters of the algorithm.

References

[1] Gold. Digital Speech Networks. *Proceedings of the IEEE*, 65(12), December 1977.

[2] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. Rtp: A Transport Protocol for Real-Time Applications. RFC 1889, January 1996.

[3] Jean-C. Bolot and A. Vega-Garcia. The Case for FEC-Based Error Control for Packet Audio in the Internet. *ACM Multimedia Systems*, 1997.

[4] M. Podolsky, C. Romer, and S. McCanne. Simulation of FEC-Based Error Control for Packet Audio in the Internet. In *Proceedings of IEEE INFOCOM'98*, San Francisco, Ca, USA, March/April 1998.

[5] T.J. Kostas, M.S. Borella, I. Sidhu, and G.M. Schuster. Real-time Voice Over Packet-Switched Networks. *IEEE Network*, 12(1):18–27, January 1998.

[6] R.V. Cox and R.E. Crochiere. Multiple User Variable Rate Coding for TASI and Packet Transmission Systems. *IEEE Transactions on Communications*, 28(3):334–344, March 1980.

[7] T. Bially and B. Gold. A Technique for Adaptive Voice Flow Control in Integrated Packet Networks. *IEEE Transactions on Communications*, 28(3):334–344, March 1980.

[8] J.C. Bolot and T. Turletti. A Rate Control Mechanism for Packet Video in the Internet. In *Proceedings of IEEE INFOCOM'94*, Toronto, Canada, June 1994.

[9] I. Busse, B. Deffner, and H. Schulzrinne. Dynamic QoS Control of Multimedia Applications based on RTP. *Computer Communications*, 1996.

[10] D Sisalem and H. Schulzrinne. The Loss-Delay Adjustment Algorithm: A TCP-Friendly Adaptation Scheme. In *Proceedings of the International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, Cambridge, England, July 1998.

[11] N. Yin and M.G. Hluchyj. A Dynamic Rate Control Mechanism for Source Coded Traffic in a Fast Packet Network. *IEEE Journal on Selected Areas in Communications*, 9(7):1003–1012, September 1991.

[12] P.T. Brady. A Model for Generating On-off Speech Patterns in Two-way Conversations. *Bell Systems Techn. Journal*, 48:2445–2472, September 1969.

[13] J.N. Daigle and J.D Langford. Models for Analysis of Packet Voice Communications Systems. *IEEE Journal on Selected Areas in Communications*, 9(7):1003–1012, September 1991.