INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

1-Bit Schemes for Service Discrimination in the Internet: Analysis and Evaluation

Martin May   Jean-Chrysostome Bolot   Christophe Diot   Alain Jean-Marie

N° 3238

1 Août 1997

THÈME 1

Rapport de recherche

ISSN 0249-6399

1-Bit Schemes for Service Discrimination in the Internet: Analysis and Evaluation

Martin May Jean-Chrysostome Bolot Christophe Diot Alain Jean-Marie

Thème 1 — Réseaux et systèmes
Projets Rodeo et Mistral

Rapport de recherche n° 3238 — 1 Août 1997 — 21 pages

Abstract: Schemes using a single-bit priority field in IP packets have recently been proposed as a low-cost (both in terms of implementation and architectural changes from the current Internet architecture) way to augment the single class best effort service model of the current Internet to include some kind of service discrimination. Such schemes appear attractive, however it is not clear yet what kind of service model they would provide to applications. We examine this in the paper. Specifically, we describe and solve analytic models of two 1-bit schemes recently proposed by Clark and Crowcroft; we obtain expression for performance measures that characterize the service provided to tagged packets, the service provided to non-tagged packets, and the prevalence of denial of service (i.e. the percentage of tagged packets that do not get the better service). We use these expressions, as well as simulations and experiments from actual implementations on a testbed at INRIA, to illustrate the benefits and shortcomings of the schemes. We also discuss implications of our results such as how these schemes can be used to transmit layered data, and how they would interact with tariffing schemes.

Key-words: Internet, integrated services networks, priority mechanisms, analytic modeling kw

(Résumé : tsvp)

Unité de recherche INRIA Sophia Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA ANTIPOLIS Cedex (France)
Téléphone : 04 93 65 77 77 – international : +33 4 93 65 77 77 – Fax : 04 93 65 77 65 – international : +33 4 93 65 77 65

Schémas de Type “1-Bit” pour l’Intégration de Services dans l’Internet: Analyse et Evaluation

Résumé : Des mécanismes basés sur l’utilisation d’un seul bit de priorité dans l’en-tête de paquets IP ont été proposés pour améliorer le service “du mieux possible” offert par l’Internet actuel, ou plus exactement pour offrir des services discriminés, ce qui en pratique veut dire offrir en plus du service actuel un autre type de service de meilleure qualité. Ces mécanismes ont l’avantage d’être simples, mais il n’est pas encore clair quelle qualité de service ils vont permettre d’offrir, et comment ils seront implémentés en pratique. Nous examinons ces deux questions dans le papier.

Nous présentons des modèles analytiques ce mécanismes de type “1-bit”, et obtenons des résultats pour diverses mesures de performance, en particulier le délai des paquets haute priorité, le délai des paquets basse priorité, et la fraction de paquets haute priorité qui sont en fait servis en basse priorité. Nous utilisons ces résultats, en plus de simulations et de mesures expérimentales obtenues sur un réseau de test à l’INRIA, pour illustrer les bénéfices et limites de ces mécanismes.

Mots-clé : Internet, réseaux à services intégrés, priorités, modèles analytiques

1 Introduction

There has been a major effort these past few years aimed at augmenting the single class best effort service of the current Internet to include services offering a variety of performance guarantees. Providing such services requires i) that one define the desired services, and ii) that the admission control, scheduling, and/or signaling mechanisms required to provide the appropriate services be defined and implemented in the Internet. Much of this work has been (and is still being) carried out in IETF working groups. The Intserv (Integrated Services) working group has identified a number of desirable services [15]. The ISSLL (Integrated Services over Specific Lower Layers) working group is specifying admission control and scheduling algorithms for a variety of underlying network technologies [16]. Finally, the RSVP (Resource Reservation Protocol) working group has specified a state setting and signaling protocol [1].

An integrated Internet providing a variety of services ranging from the best effort to the deterministic guarantee services is a worthy goal. However, reaching this goal is not so easy. This is illustrated for example by observing that some of the IETF groups mentioned above have been active for quite some time, yet the Internet still only provides a single class best effort service, with other services not expected to be widely deployed for a while. This is caused in part by the difficulties associated with identifying the desirable services among all possible services (we note that the IntServ working group eventually singled out two specific services, namely the guaranteed service GS [25] and the controlled load CL service [29] for initial deployment and experience gathering in the Internet). Furthermore, many services turn out to require complex associated admission control and scheduling mechanisms, which in turn implies a relatively complex interface between the user/application and the network. For example, the interfaces for the SBR (Statistical Bit Rate) or the CL services require that the user specify parameters such as burst length, while even technologically minded users know little about peak rates (all this is even excluding the parameters related to tariffing). Finally, there are concerns about the costs associated with the wide deployment of complex scheduling and signaling protocols, and more generally with scalability issues when moving from a stateless to a stateful network architecture.

This has led researchers to propose "simple" schemes for providing services that extend (even slightly) beyond best effort, where "simple" above applies to architectural, user interface, and implementation complexity. Of course, simple is still more costly than the current stateless single class FIFO scheduling. However, the hope is to obtain at least some of the benefits of a versatile multiservice network (in particular some kind of service discrimination) for a small cost. For example, some recently proposed schemes do away with signaling and admission control altogether.

Examples of such schemes include the Random Early Detection (RED) scheme proposed by Floyd et al. [10], and "1-bit" schemes recently proposed by Clark [4] and Crowcroft [5]. The idea of RED is to monitor queue activity in routers, and drop packets selectively from connections depending on connection behavior. The initial purpose of RED appeared to be the provision of a "fairer" best-effort Internet, with in particular protection from malicious connections obtained by punishing (i.e. preferentially dropping packets from) connections that behave in an overly aggressive fashion. However, appropriate drop policies could provide more general kinds of service discrimination between connections in a RED router. Refer to [10] for a detailed description and [10, 8] for thorough analyses of RED.

One-bit schemes explicitly attempt to introduce service discrimination between different connections in the Internet. The idea there is to include a 1-bit flag in packets, which if set to 1 indicates that the packet should receive preferential treatment. Clearly, this idea is not new since it amounts to using in the Internet environment concepts such as the drop preference bit (DE bit) used in Frame Relay, or the cell loss priority bit (CLP bit) used in ATM. However, the bit can be used to provide different kinds of service discrimination. In [5], it is argued that a guaranteed or expected low delay would benefit applications that most suffer from the current state of the Internet, namely interactive applications such as interactive multimedia or DIS-like applications [22], and that tagging should be used to discriminate between delay sensitive and non delay sensitive applications. In [4], it is argued that a guaranteed or at least expected throughput is the one quality of service (QoS) feature of interest to most applications, and that tagging packets (i.e. setting their bits to 1) should be used to provide such guarantee.

In any case, 1-bit schemes are attractive because they appear to be simple to implement (for example DE-bit discrimination has been available for quite some time in commercial Frame Relay products). Furthermore, they seem to be able to provide service discrimination even in the absence of "heavyweight" stateful architecture with connection admission control, scheduling, and signaling. However, it is not clear in that case exactly what kind of service is provided to applications.

These are precisely the issues examined in this paper. Specifically, in Section 2, we briefly review the 1-bit schemes proposed in [5, 4], and we describe possible ways to implement them. In Section 3, we describe and analyze analytic models of such schemes. We obtain analytic expression for performance measures that characterize the service provided to tagged packets, the service provided to non-tagged packets, and the prevalence of denial of service (i.e. the percentage of tagged packets that do not get the better service). We use these expressions, as well as simulations and experiments from an actual implementation of the scheme on a testbed at INRIA, to illustrate the benefits and shortcomings of the schemes. In Section 4, we discuss implications of our results such as how well an integrated services Internet can be used to transmit layered or hierarchically encoded data, and to transfer TCP traffic. Section 5 concludes the paper.

2 Recent Proposals for 1-Bit Schemes

The 1-bit schemes have been proposed as low cost schemes to provide service discrimination in the Internet. They do not require signaling protocols, relying instead on simple monitoring at the edge of the network, and on relatively simple scheduling in routers, as shown in the figure below. Essentially, these scheduling algorithms can be thought of as managing two queues. Packets with the service bit (which we also refer to as the priority bit) set to 0 go to the low priority, best-effort queue; packets with the priority bit set to 1 go to the high

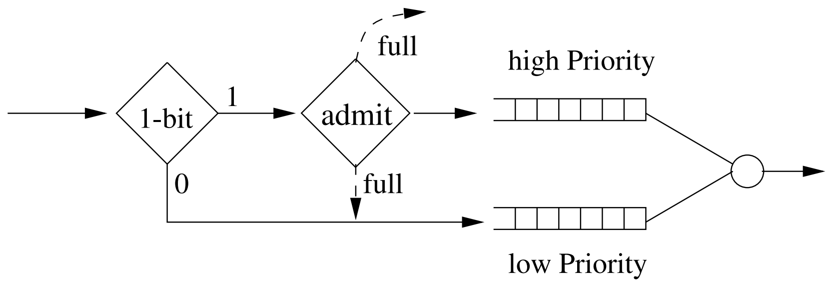

Figure 1: Service discrimination with 1-bit scheduling

priority queue, up to the maximum capacity of this queue; when this queue is full, packets with priority bit 1 go directly to the best-effort queue or are dropped, depending on the traffic management scheme.

The goal, of course, is that to a low deployment cost of 1-bit schemes would correspond a significant increase in the panoply of services provided to users. However, the descriptions in [5] and [4] do not quantify the type of service one can expect when using the proposed schemes (we investigate this in Section 3). In any case, the main idea appears to offer i) at least some kind of service discrimination between best effort and "better" than best effort service, and ii) a "better" service somewhat similar to the IETF CL service. This service can be expressed informally as "x% of the packets with priority bit 1 receive a priority service". For the delay bit, this means that x% of the packets with priority bit set to 1 will suffer a delay close to the minimum delay; for the throughput bit, this means that x% of the bandwidth sent with the priority bit set to 1 will reach the destination.

To obtain harder guarantees, the 1-bit schemes must be augmented with measurement based admission control (packets are then admitted on an evaluation of the flow size), or with static queue configuration (the high priority queue is designed to affect each flow a static share of the queue capacity). In both cases, a flow identifier (such as IPv6 flow label) is used to identify packets belonging to the same flow.

However, service discrimination even without guarantees appears to be of interest in a number of cases, such as when used i) by adaptive applications, ii) by applications with layered coding of data, or iii) by applications with specific delay or throughput constraints. Regarding point i), rate adaptive applications, for example, would benefit from the 1-bit throughput service, and adapt the rate at which they send tagged packets depending on the observed loss rate of tagged packets so as to make sure that a constant and large fraction of tagged packets gets priority service, even in the presence of congestion in routers. Regarding point ii), applications with layered coding of data have often been cited as being particularly well suited to networks that offer two services classes. Recall that with layered coding, source data is encoded into a number of layers or (sub)bands that can be combined to reconstruct a signal that gets closer to the original signal as the number of combined layers increases. In practice, layered encoding schemes are organized so that some layers (in particular the base layer) are more important than other layers, and might in fact be mandatory to reconstruct a coherent signal. This is for example the case with most layered video coders such as MPEG, layered DCT, or wavelet coders [14, 20]. Using such schemes can be done efficiently over a network that is able to discriminate between packets and provide packets that carry information from the "important" layers a better service class (i.e. here with the service bit set to 1), and packets from the other layers the best effort service. Regarding point iii), applications such as DIS (Distributed Interactive Simulation) or distributed gaming do require that some packets such as collision packets (which indicate when a collision occurs and which parties are involved, e.g. a missile hitting a target) suffer minimum delay on the way to their destinations [13, 22]. However, they do not require much in terms in bandwidth; thus, the 1-bit scheme for delay sensitive data would be very well suited to this. State update packets that describe the current status of application objects are usually redundant and they do not need a high priority service as they can be dead-reckoned when missing.

Next, we briefly review the schemes proposed by Clark [4] and Crowcroft [5]. We refer to Clark's scheme as the 1-bit scheme for throughput sensitive traffic, and to Crowcroft's scheme as the 1-bit scheme for delay sensitive traffic.

2.1 The 1-Bit Scheme for Throughput Sensitive Traffic

Dave Clark argues in [4] that i) a guaranteed or at least expected throughput is the one quality of service (QoS) feature of interest to most applications, and thus that ii) there is a need for a mechanism that directly reflects users desires in term of a transfer time. Regarding point i), the idea is that users/applications know the size of data to be transferred and the desired delivery time, which then can be used to define a minimum transfer rate. This rate then becomes the service objective for the transfer. One way to achieve this objective is to define at the source a service profile, which defines how packets should be sent so as to meet the rate objective, and to incorporate a traffic meter. The traffic meter monitors the transfer in progress and flags each packet of the data stream. The flag is set to 1 if packets are sent according to the profile (i.e. at a rate that conforms to the expected rate); we then refer to such packets as tagged packets. The flag is set to 0 otherwise; we then refer to such packets as non tagged packets.

Implementing the traffic meter depends on how the profile has been defined. For example, with a profile that specifies mean rate and maximum burst length, the meter can be implemented using a leaky bucket filter. Packets that conform to a leaky bucket output are tagged, others are not tagged. The tag information is then used by the intermediate routers in case of congestion. Non tagged packets are preferentially selected to receive a congestion pushback notification, which typically means that they are dropped. In practice, there are three ways to implement such selective drop scheme, namely

- A threshold mechanism: When buffer occupancy reaches a given threshold M, arriving non tagged packets are discarded; tagged packets are still admitted in the queue as long as the buffer is not full.

- RED combined with a threshold scheme: In the original RED scheme, the drop probability depends on the queue length and the time elapsed since the last packet was dropped; we could extend this mechanism by increasing the drop probability of the non tagged packets.
- A push-out mechanism: An arriving tagged packet may enter the saturated queue provided a non tagged packet is already awaiting transmission. Then, one of the non tagged packets is discarded and the tagged packet joins the queue. Non tagged packets cannot enter a saturated queue and are dropped.

The threshold strategy and the modified RED mechanisms differ from the push-out mechanism in that tagged packets can be discarded while non tagged packets are still in the buffer.

The proposal in [4] argues that, since the goal is to obtain an expected rather than a guaranteed throughput, the scheme above does not require complex signaling and admission control mechanisms. Instead the network should provide information about the actual usage across the network links to prevent the user from too optimistic expectations. Thus, the only building blocks required for implementation are the traffic meter at the host (in combination with the mechanisms to determine the traffic profile), the 2-class scheduling in the routers, and a tariffing scheme, which we will get back to in Section 4.

2.2 The 1-Bit Scheme for Delay Sensitive Traffic

The scheme proposed by Crowcroft in [5] is in a way the dual of that of Clark, with throughput replaced by delay. The argument there is that i) a guaranteed or at least expected low delay is a quality of service (QoS) feature, and thus that ii) there is a need for a mechanism that directly reflects users desires in term of a delay. Regarding point i), [5] claims that there is an increasing demand from applications which need a faster carriage for at least a fraction of the packets sent. We briefly described earlier a few such applications, in particular DIS applications.

The scheme in [5] also relies on a flagging mechanism. However, in contrast to the throughput scheme above, there is no general algorithm to tag the packets. Instead, the application itself (or the user) has to decide which packets are important and therefore should be tagged. Therefore, it is important both for quality monitoring by the application, and for tariffing purposes, to monitor the rate or fraction of high priority packets in the data flow, and, if necessary, mark the packets for a low priority service.

Implementing this 1-bit delay scheme in routers cannot be done using the same mechanism as that above for the throughput scheme since the waiting time of the high priority packets is not influenced. Therefore we consider another system in which two outgoing queues are accessed by the tagged and non tagged packets. The queue for tagged packets (high priority queue) has limited capacity K; the queue for non tagged packets (low priority queue) has a buffer capacity N. The scheduling mechanism is then implemented as follows: If there is a packet in the high priority queue, this packet is queued first. Only if the high priority queue is empty is the low priority queue served.

In the following section we define a model for the two proposed traffic schemes. The model for the throughput sensitive traffic corresponds to the push-out mechanism described in section 2.1. The delay sensitive model follows the mechanisms described above.

3 Analytic Modeling and Evaluation

This section gives a detailed performance evaluation of the two service discrimination schemes introduced in [4] and [5]. For both schemes we assume that the input traffic is Poisson and that the buffer space is limited. The model for the scheme for throughput sensitive traffic is valid for general independent service times. The model for the scheme for delay sensitive traffic is analyzed with a M/M/1/K queueing system.

The arrival process to the queue can be approximated by a memoryless process if a large number of independent traffic sources is assumed and each source contributes a small fraction to the total load. The service discipline of all queues is FIFO.

3.1 The 1-Bit Scheme for Throughput Sensitive Traffic

We mentioned earlier that the 1-bit scheme for throughput sensitive traffic pretty much amounts to applying to the Internet environment concepts such as the cell loss priority bit (CLP bit) used in ATM. Thus, it is not surprising that models of routers (or rather switches) handling tagged traffic have been examined. Consider for example an implementation of tagged traffic using the pushout scheme mentioned above. An arriving tagged packet may enter a saturated queue provided that a non tagged packet is already awaiting transmission. This packet is then discarded and the tagged packet joins the queue. If the queue contains only tagged packets, the arriving packet is discarded. A model of a queue using such a scheduling scheme is shown below.

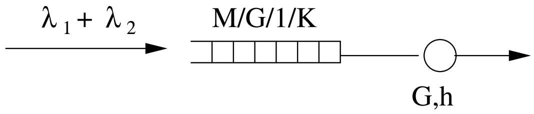

Figure 2: A queue with priority discarding mechanism

Under the assumption that the input traffics of tagged and non tagged packets are Poisson, this model has been examined in detail and solved in [18, 24]. In particular, expressions are given for the fraction of tagged packets that are not accepted in the high priority queue. We refer to the references for details.

3.2 The 1-Bit Scheme for Delay Sensitive Traffic

We are not aware of models that would fit exactly the 1-bit scheme for delay sensitive traffic. Thus, we focus on this latter scheme in the rest of this section. We develop and solve a model to analyze its performance next.

For the delay analysis, we consider a model where two buffers are accessed by high and low priority packets: a high priority buffer with a limited buffer size of K, and a unlimited buffer for the low priority packets. Whenever the high priority queue is not empty, these packet are transmitted. The packets in the low priority queue are transmitted only if there are no high priority packets in the high priority queue. When the high priority queue is full and there are high priority packets arriving, these are discarded immediately (see Figure 3).

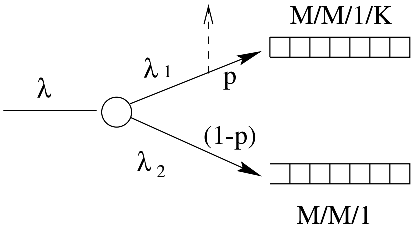

Figure 3: A two-queue priority system

We shall assume that the input stream is Poisson with rate λ, and that the arriving packets have high priority (class 1) with probability p, and low priority (class 2) with probability $1-p$. Alternately, we may consider that the input streams of the high and low buffers are independent Poisson processes with rates $\lambda_1 = p\lambda$ and $\lambda_2 = (1-p)\lambda$. The parameter p is the fraction of high priority packets. Let $\rho = \lambda/\mu$, and let $\rho_1 = \lambda_1/\mu$ be the load factor of the high priority queue.

Our objective is to determine R_i, the response time of a class-i customer accepted in the system. We shall assume for simplicity that the service discipline is *preemptive*, and that the service times are exponential with parameter μ.

For the high priority customers we find that R_1 is the response time of customers in a $M/M/1/K$ queue with arrival rate λ_1. Let N_1 be the stationary number of customers in this queue. By Little's law and well known results [17], we have

$$ER_1 = \frac{EN_1}{\lambda_1} = \frac{1}{\mu - \lambda_1} \; \frac{1-(K+1)\rho_1^K + K\rho_1^{K+1}}{1-\rho_1^{K+1}} \; . \tag{1}$$

We also obtain the loss probability of the high priority class:

$$\pi_1 \;=\; \rho_1^K \; \frac{(1-\rho_1)}{1-\rho_1^{K+1}} \; .$$

This allows to find the *stability* condition of the system:

$$\mu \;>\; \lambda_2 + \lambda_1 \pi_1 \;=\; \lambda(1-p(1-\pi_1)) \; . \tag{2}$$

From this condition follows the maximum admissible offered load (p being fixed), or, for a fixed offered load, the minimal fraction p of packets that must have high priority.

The response time of the low priority customers is more difficult to obtain that R_1. We develop below bounds based on the analysis of [27] and [21]. Note that in any case, the response time of each customer of class 2 is less than what it would be if K were replaced by $K' > K$. This is clear because then there would be more high priority customers accepted and the preemption periods would be longer. In particular, if $K = \infty$, our model reduces to a two-class $M/M/1$. The Laplace transform of the waiting time in this queue (actually, in the $M/GI/1$ queue) is known from [27], for both preemptive and non-preemptive priorities. In our case, we obtain the bound

$$ER_2 \;\leq\; ER_2^\infty \;:=\; \frac{1}{\mu}\,\frac{1}{1-\rho_1}\,\frac{1}{1-\rho}\;. \tag{3}$$

A finer bound is possible. Following [27], we argue that the response time R_2 is the sum of χ_2, the workload encountered by this customer upon arrival (including its own service time, because the discipline is preemptive), plus as many preemption periods as there were arrivals of high priority customers during the execution of this workload. In symbols, we have:

$$R_2 \;=\; \chi_2 \;+\; \sum_{j=1}^{A(\chi_2)} B_j. \tag{4}$$

Because of the finite capacity, the distribution of χ_2, $A(\chi_2)$ or of the lengths B_j are not easy to obtain. However, χ_2 is less than R_2^∞, the stationary response time in the $M/M/1$ queue, which is known to be exponential with parameter $\mu - \lambda_1$. Then, $A(\chi_2)$ is less than the number of arrivals of the Poisson process of rate λ_1 in the interval $[0, R_2^\infty]$. Finally, the preemption periods are less than B_K, the length of the busy period in the $M/M/1/K$ queue. According to [21], the numbers $b_K := EB_K$ are given by the recurrence:

$$b_n \;=\; (b_{n-1} - \sum_{j=1}^{n-1} b_j p_{n-j}) \;/\; p_0, \tag{5}$$

$$p_k \;=\; \int_0^\infty (\lambda_1 t)^k / k! e^{-\lambda_1 t} d\sigma(t) \;=\; \frac{\rho_1}{(1+\rho_1)^{k+1}}\;, \tag{6}$$

starting from $b_0 = 1/\mu$. Combining all this, we obtain from (4) the bound:

$$\begin{aligned} ER_2 \;&\leq\; E\chi_2\;(1 \;+\; \lambda b_K) \\ &\leq\; ER_2^\infty\;(1 \;+\; \lambda b_K)\;. \end{aligned} \tag{7}$$

In this expression, ER_2^∞ is given in (3). The sharpness of this bound is discussed in the next section (see Figure 5). The rigorous justification of (7) uses the methods of stochastic ordering [26]. Finally, note that this approach may be applied to the *non-preemptive* case, and that it should be possible to generalize it to generally distributed service times. Indeed, the results of [27, 21] are not limited to exponential service times, and they provide results on the Laplace transforms of the distributions, not solely on averages. This offers the possibility of analyzing queue length and jitter *distributions* as explained in [24, p. 432]. The details will be omitted here.

We can now use the means of the response time of both queues to evaluate the jitter experienced by both classes of packets. The jitter can be evaluated by, for instance, the difference between the delivery times of two consecutive packets of different priorities, hence:

$$Jitter = E((a_2 - a_1) + R_2 - R_1) = \frac{1}{\lambda_2} + ER_2 - ER_1 \ . \tag{8}$$

This jitter is of interest when we think of a multi-layer applications sending data with different priorities. When this jitter gets greater than the play-back value of the stream, the hierarchical encoding, used together with service discrimination is useless or even harmful. We discuss this problem in section 4. The combination of (1) and the bound (7) yields an upper bound on the jitter (8).

3.3 Analytic and Simulation Results

In this section we present the results of the analytical calculations and of the simulation results we collected. Our evaluation focuses on the scheme for delay sensitive traffic, since schemes similar to that for throughput sensitive traffic have been studied before.

All the simulations are done with the queueing simulation tool QNAP [23]. The experimental part of the validation is done on a local testbed at INRIA, which includes a pool of PCs running a modified Linux version 2.0.25. The traffic-generating applications we used to measure the loss and delay characteristics were developed at INRIA too. We use these applications to exmaine the delay and drop probability as a function of the amount of high priority traffic. Specifically, to evaluate the resulting traffic behavior of the two service discrimination schemes, we measure the following characteristics:

- The per packet delay for the high priority and the low priority packets.
- The loss probability for the high priority and the low priority packets.
- The mean jitter between packets of the two service classes.

We first take a look at the delay characteristics. Figure 4 shows the simulation results for two buffer sizes of $K = 10$ and $K = 20$. The model is that of Section 3.2. For this experiment, μ is set to 1.0, the total offered load λ varies from 0.67 to 0.95 on the y-axis. The load parameter p varies from 0 to 1 on the x-axis. The jagged aspect of the simulation curves at high load is due to the classical difficulty of obtaining statistical convergence.

We make two observations from Figure 4:

- The maximum delay experienced with high priority service, not surprisingly, is always smaller than the delay experienced with low priority. This first observation confirms

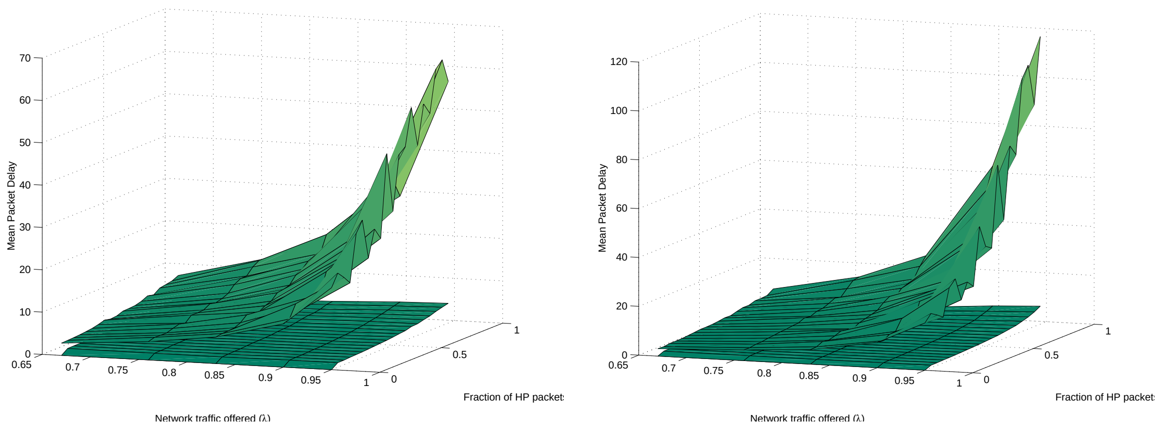

Figure 4: Mean delay for different values of λ and p, for $K = 10$ (left) and $K = 20$ (right)

the 1-bit discrimination works efficiently, despite its simplicity. In the case of high priority packets, the delay remains very close to the minimum network delay (which corresponds to 1 on the vertical axis with our model). With low priority, it is much higher.

- The increase of the delay for increasing values of λ is more important for the low priority service. For a small fraction of high priority packets, tagged packets all experience the same delay (independent of the value of λ) while for low priority packets, the delay varies by almost an order of magnitude for increasing values of λ over the range we consider. When the fraction of high priority packets is high, the delay variation in the high priority queue remains very small (2 times the minimum delay), while that for low priority varies by up to two orders of magnitude.

We now discuss the sharpness of the bound derived in section 3.2, in order to decide if it can be used for an accurate prediction of response times and jitter. Figure 5 shows the comparison of the bounds computed using (7) and of the simulations already shown in Figure 4. The graph on the left is for $K = 10$, the one on the right for $K = 20$. The five different curves correspond to different values of λ, namely 0.67, 0.76, 0.83, 0.90 and 0.95 from bottom to top. The curves of Figure 5 clearly show that the bound (7) constitutes an excellent approximation, except perhaps when λ and p are both close to 1, in which case the effect of the finite capacity K becomes measurable.

The bound being very fast to compute, it is possible to envision the online control of the system (*e.g.* best choice of K or p) based on some criterion involving λ, ER_1, ER_2 and/or π_1.

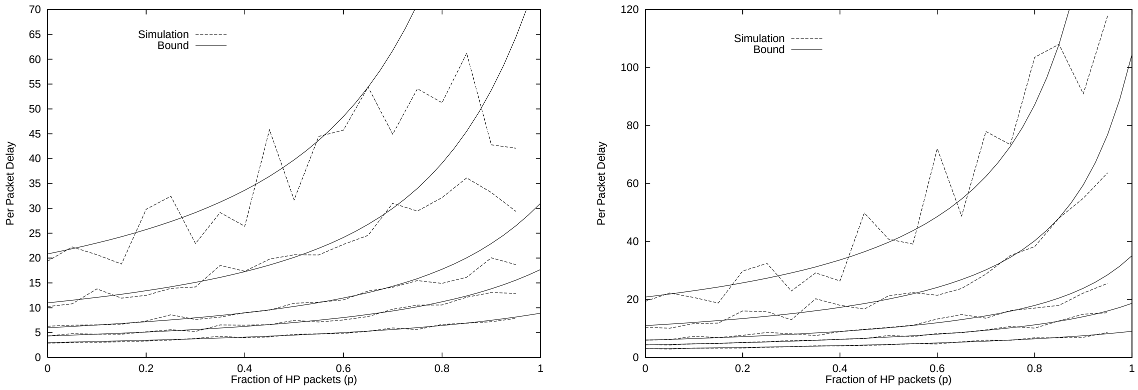

Figure 5: Comparison of simulation and analytical results for $K = 10$ and $K = 20$

We have examined above measures (specifically the mean delay) that characterize the service provided to each class of service. We next examine measures that characterize in a way the "amount of discrimination" between the two classes. These include the jitter and the fraction of high priority packets that do not get high priority service. We define the jitter to be the difference between the mean response time of the high and the low priority packets. This measure is of particular interest when transmitting data flows using different service classes for the different layers [2] since it indicates how much the destination has to wait before it can reconstruct the original signal from the different portions of the signal carried in each flow. Figure 6 shows the mean jitter for different values of λ and p, obtained for $K = 10$ (left) and $K = 20$ (right).

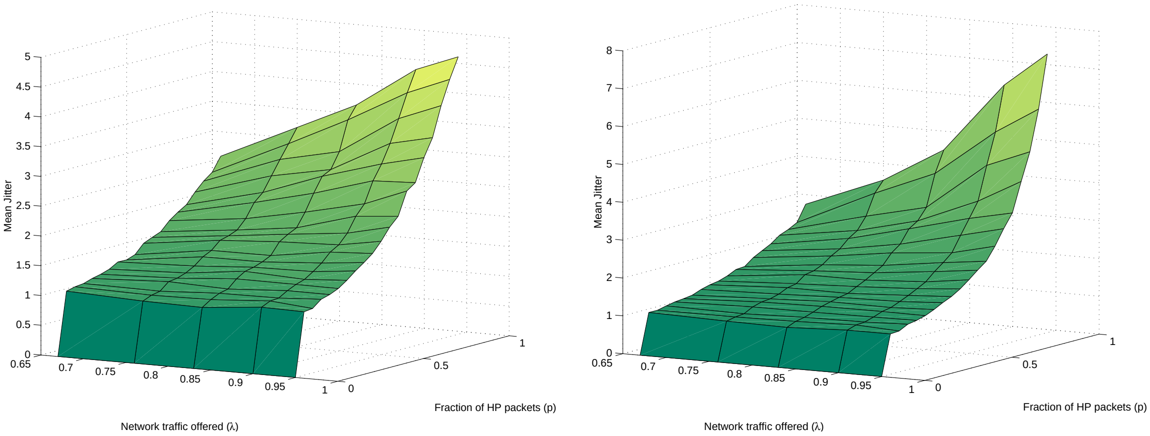

Figure 6: Mean jitter for increasing fraction of high priority packets

We observe a striking increase of the jitter when the fraction of high priority packets exceeds 40%. When λ increases, the jitter rapidly increases to become one or two orders of magnitude larger than the minimum transmission delay. This result suggests that in a moderately or highly loaded network, the resynchronization of the flows sent over different classes of service links would take so much time as to make the whole operation unattractive. This point is discussed further in Section 4.

Figure 7 shows the loss probability, i.e. the probability that a high priority packet does not get high priority service, for different values of p and λ, obtained for different values of high priority load $\rho = 0.6$ and $\rho = 2$. For this tests we use a configuration of $K = 10$ and $K = 20$ for the high priority queue (upper surface in the figures) and a fixed buffer size of 100 for the low priority queue (lower surface).

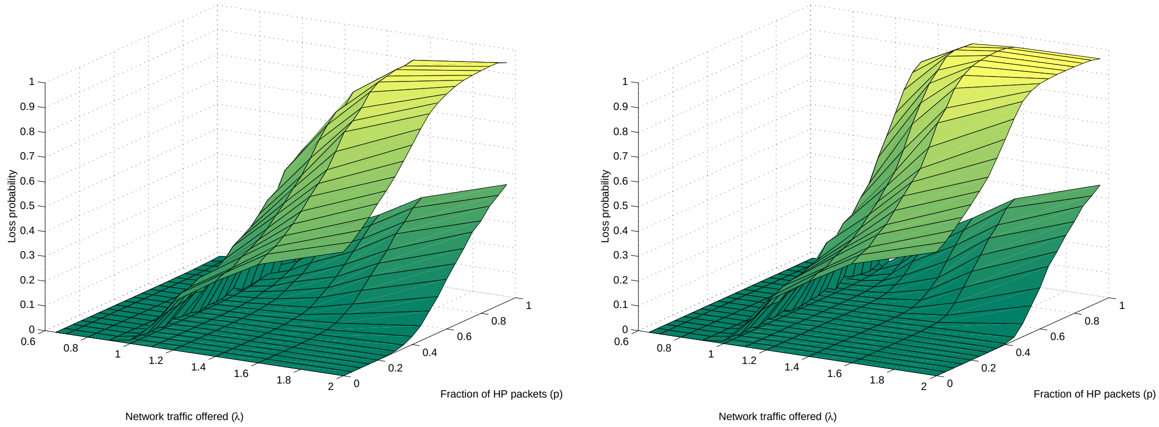

Figure 7: Loss probability for different values of λ and p, with network loads $\lambda_1 = 0.6$ and $\lambda_1 = 2.0$

The figures show that with fewer than 40% of high priority packets, all the high priority packets are transmitted with the high priority service. When the proportion of high priority packets increases beyond 40%, and for a heavily congested network (offered load greater than 1), the loss rate stays below 45%. Results here clearly depend on the Poisson assumption for input traffic. Given this assumption, we find that even with heavy congestion, the majority of high priority packets does get high priority service.

3.4 Experimental Results

We have implemented the 1-bit schemes in a testbed at INRIA to gain experience with implementation complexity, and to obtain experimental results that would complement the analytic and simulation results above. The testbed involves a number of PCs running Linux 2.0.25 and acting as routers, in which we have changed the FIFO scheduling discipline to replace it with 1-bit based priority disciplines. For this paper, we have used the simple setup shown in Figure 8. The first source PC generates the traffic under test; the second source

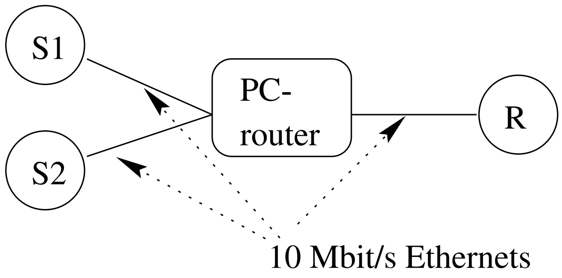

Figure 8: Network setup

PC generates cross traffic. Both sources generate Poisson streams. The PCs are connected with 10 Mbit/s Ethernets. They exchange UDP packets with a fixed packet size equal to 1250 bytes, and a mean send rate of 100 packets/s (compared to the Ethernet capacity equal to 1000 packets/s).

Figure 9 shows the evolutions of the delay experienced by tagged and non tagged packets. The source sends packets at the the maximum rate of the router, with 80% of the packets being high priority packets. The graph nicely shows the impact of tagging: low priority packets wait for service until the end of a busy period caused by high priority packets; if the duration of the busy period is long, the delay suffered by low priority packets can be extremely high, and so is the jitter between the two service classes.

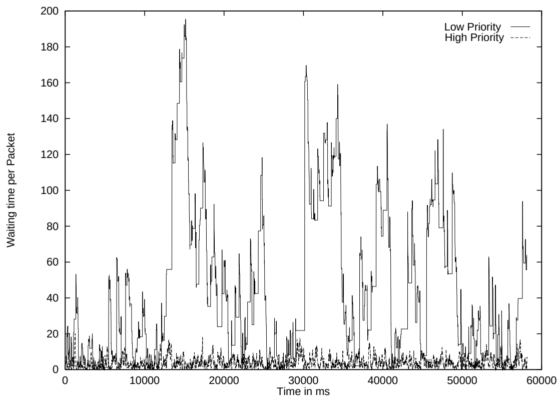

Figure 9: Evolutions of packet delay in the low and high priority queues

Figure 10 shows the fraction of lost packets measured in time intervals of length 100 ms. In both experiments the router is heavily overloaded, with the send rate being 25% higher than the router capacity. The left graph shows losses when 50% of the packets are high priority packets; the right graph shows losses when 80% of the packets are high priority packets. In all cases, the high priority queue has size $K = 10$, the low priority queue has size $N = 100$.

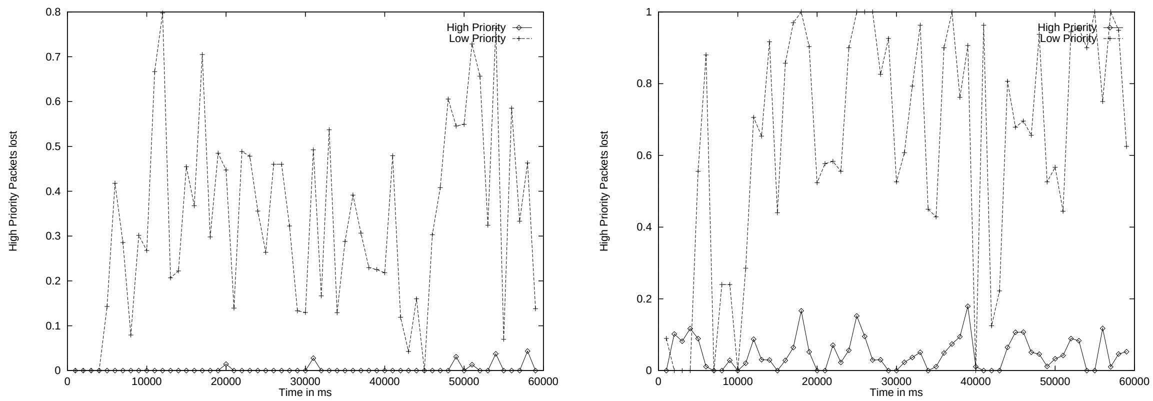

Figure 10: Evolutions of the fraction of high and low priority packets lost

The curves show that the experimental behavior is consistent with the simulation results; while losses are rare with a moderate fraction of high priority packets, their number increases when this fraction increases. The results of the dropping statistics confirm, that with a fixed rate of high priority packets and no bursts, the high priority queue does not drop any packet.

4 Discussion

In this section, we examine further a few issues that appear of interest given the results above. For reasons of space, we focus on results related to jitter, on 1-bit schemes and the ALF approach, and on 1-bit schemes and adaptive applications.

- The jitter between the two classes of service

  We examined above the mean jitter between high and low priority packets, i.e. between the different service classes. We have seen that this jitter is high when the network load is moderate or high. Consider then the case when source data is encoded using layered coding. Several researchers have claimed [2] that applications with layered coding of data are particularly well suited to networks that offer different service classes. This is because layered coding schemes encode data into layers, some of which typically are more important than others (for example, cosine transform coding of video data as done in H.261 or MPEG produces frequency coefficients, and the human eye is more sensitive to low frequency coefficients than to high frequency coefficients). The idea then is to send the more important data (for example the low frequency coefficients in the example above) using the class of service that offers the better performance guarantees. At the destination, the signal is reconstructed at time t from the layers received up to time t; the reconstructed signal gets closer to the original signal as data from more and more layers is received.

However, our results show that when the network is moderately or highly loaded, i.e. precisely when service discrimination is important to users, jitter is high meaning that the destination user has to wait for a long time before the data has been received from all the layers (including layers sent over the high and low priority classes) to reconstruct the original signal. Thus, if the application is time constrained, it will have the time to wait for low priority layers, and the question then is whether sending these layers was worth it in the first place.

The situation is different if data in different layers do not depend from each other, and/or if timing constraints are loose. Then service discrimination between layers could be very useful, as well as be cheaper to the user since sending only the more important layers over the high priority service presumably will translate into a lower cost, while keeping overall quality high. This is the case, for example, with image transfer in the Web.

Our results on jitter also show that in the general case of a multihop path, a high priority packet which does not get high priority service should be either dropped or forced to become a low priority packet. We have not examined this more general case, and this is clearly an item for future work.

- 1-bit schemes and ALF

  Application Level Framing (ALF) is an architectural design principle for developing high performance applications [3]. ALF relies in part on the ability of applications and protocols to process, control, and transmit packets independently one from the other. These packets are called ADUs (Application Data Units).

  We note that ALF-based design of applications improves the efficiency and the interest of 1-bit service discrimination schemes for two reasons. The first reason has to do with the fact that ALF prohibits segmentation. In case of segmentation, the different parts of a single high priority packet could be processed with different classes of service. Then, the segments directed to the low priority queue can be either lost or delayed. In both cases, our results show that the jitter experienced at the destination is such that the high priority service would become useless. The second reason has to do with the fact that in an ALF application, each type of ADU can be processed with "tailored" protocols in order to maximize the benefit of the 1-bit approach. For example, in case of redundant transmission of ADUs, the number of redundancies can be different for the high priority ADUs and for the low priority ones. As ADU are designed to be processed independently (in particular out of sequence), the receiving application can process ADUs upon reception, not waiting for other ADUs transmitted with the lower priority service. This decreases the amount of time when the destination is blocked waiting for late packets/segments, and thus provides end users with increased performance (refer to [6] for a detailed analysis).

- Pricing and adaptation mechanisms

  Providing service discrimination in the Internet requires that users have incentives to choose the appropriate service. One way to provide such incentives is via pricing. Now, 1-bit schemes are attractive because it turns out that they can be coupled easily with pricing schemes. Consider for example the 1-bit scheme for throughput sensitive traffic. The pricing scheme can be based on the rate at which tagged packets are sent, and on the offered service, as follows. The cost for a data transmission consists of two fees. The first is a fixed connection fee, the consumer has to pay for the connectivity. The second fee is variable and depends on the rate of high priority traffic and the actual network behavior.

  We now describe an adaptation mechanism which allows us to adapt the fraction of high priority traffic to the level of network congestion. To do this, we define a cost function, which measures the cost of congestion that an increasing high priority traffic rate λ_1 imposes on the other users. For our two traffic management schemes we define two different cost functions: for the throughput sensitive scheme, the cost function depends on the loss rate of low priority packets; for the delay sensitive scheme, the cost function depends on the delay suffered by the low priority packets. In both cases the cost function depends on the send rate of high priority packets, and we denote it by $C(\lambda_1)$.

  Our pricing mechanism is then defined as follows: When the network in uncongested, the user does not use the tagging option for the packets. In case of congestion, the user may decide to send a fraction of his packets tagged for the high priority service. This can be done by pushing a 'increase service' button, which indicates that the user is willing to increase the price it is paying to get improved service. Our mechanism will then set the user's amount to θ_j credits (θ_j is the credit user j is willing to pay). A feedback mechanism, like the *expectation* proposed by Clark, then returns the feedback from the network in the form of the costs defined above, and a rate adaptation algorithm adapts the fraction of high priority packets using this cost information. The algorithm uses the inverse of the cost function $C^{-1}(\lambda_1)$ and the amount of credits the user is willing to pay. If the user is not satisfied with the resulting high priority rate, it can increase its credits and therefore the high priority rate. Conversely, when network congestion decreases the user can decrease its credits and thus reduce the corresponding high priority packet rate.

  This mechanism conforms to a auction strategy, where competitive users bid for their transmission rates. With the cost-function based on the low priority delay (or losses), the high priority will not be overloaded and losses may only occur during the adaptation phase. When all users using the same algorithms, the fraction of "reserved" bandwidth of the paying user corresponds to the proportion of θ_j/θ where $\theta = \sum_{j=1}^{M} \theta_j$ and M denotes the number of users. Note that the algorithm will not lead to a starvation of the low priority traffic. Low priority packets prevented from being sent are a a sign of insufficient network capacity, not of a misbehaved traffic source. Thus, we see that a combination of 1-bit scheme and an appropriate pricing scheme can provide users with bandwidth guarantees (by increasing his credits whenever the user rate tends to decrease), even in the absence of control and signalling schemes.

5 Conclusion

In this paper, we provide an analysis of the 1-bit delay service discrimination schemes recently proposed by Clark and Crowcroft. The contributions of this paper, in regard to 1-bit schemes are:

- Propose a model to evaluate the performance of the 1-bit delay schemes.
- Describe and quantify the offered service in both delay and throughput schemes.
- Use the analytic and simulation results to examine more general issues, such as the transmission of time constrained layered coded data over multiple service class links.

This is still preliminary work, and much remains to be done. Issues of particular interest include

- Combination of the 1-bit approach with RED

  We have seen that 1-bit schemes can be used to provide service discrimination. However, we have not examined the issue of providing some kind of fairness between flows of a same service class. One way to do this would be to use RED-like schemes.

- Interaction with the control mechanism of TCP
- More accurate evaluation of the service provided to individual users

  The analysis in this paper examines a model in which the input flow into the router is a stream that really is the aggregation of many streams generated by individual users/connections. It would be interesting to have a model in which we could quantify the impact of service discrimination on a *per connection* basis, as opposed to per aggregate of connections basis as is done in this paper. This would allow a more precise characterization of the service provided to individual users.

Acknowledgements

The authors thank Benoit Brodard for his help with the testbed setup.

References

[1] Braden, R., et al., “RSVP: A New Resource ReSerVation Protocol”, *IEEE Network*, September 1993. See also http://www.ietf.org/html.charters/rsvp-charter.html

[2] Campbell, A., Coulson, G., "QoS Adaptive Transports: Delivering Scalable Media to the Desktop", *IEEE Multimedia*, March 1997.

[3] Clark, D. D., Tennenhouse D. L., "Architectural considerations for a new generation of protocols", *Proc. ACM Sigcomm'90.*

[4] Clark, D. D., "Adding Service Discrimination to the Internet", Sept. 1995. http://ana-www.lcs.mit.edu/anaweb/papers.html

[5] Crowcroft, J., "All you need is just 1 bit", keynote presentation, *IFIP Conf. on Protocols for High Speed Networks*, Sophia Antipolis, Oct. 1996. http://www.cs.ucl.ac.uk/staff/jon/hipparch/dollarbit/

[6] Diot, C., Gagnon, F., "Impact of Out-of-Sequence Processing on Data Transmission Performance", *INRIA Research Report RR-3216*, INRIA Sophia Antipolis (France), July 1997.

[7] Doshi, B. ,Heffes, H., "Overload Performance of Several Processor Queueing Disciplines for the M/M/1 Queue", *IEEE Transactions on Communications*, vol. 34, no. 6, June 1986.

[8] Fall, K., Floyd, S., "Router mechanisms to support end to end congestion control", LBNL technical report, Feb. 1997.

[9] Ferrari, D., Verma D., "A scheme for real-time channel establishment in wide area networks", *IEEE Journal on Selected Areas in Communication*, Vol. 8, 368-379, April 1990.

[10] Floyd, S. and Jacobson, V., "Random Early Detection Gateways for Congestion Avoidance", *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 397-413, August 1993.

[11] Floyd, S., "Link-sharing and Resource Management Models for Packet Networks", *ACM/IEEE Transactions on Networking*, vol. 3, no. 4, pp. 365-386, August 1995.

[12] Floyd, S., CBQ web page, *URL http://www-nrg.ee.lbl.gov/floyd/cbq.html*

[13] Gautier L., Diot C., "MiMaze: A distributed game over the Internet", *INRIA research report* in preparation. Draft available at http://www.inria.fr/rodeo/MiMaze/

[14] Gharavi H., "Multilayer subband-based video coding", *IEEE Trans. on Communications*, vol. 39, no 9, pp. 1288-1291, Sept. 1991.

[15] Integrated Services working group, http://www.ietf.org/html.charters/intserv-charter.html

[16] Integrated Services over Specific Link Layers working group http://www.ietf.org/html.charters/issll-charter.html

[17] Kleinrock, L., *Queueing Systems*, vol. 1, J. Wiley & sons, 1975.

[18] Kröner, H., Hébuterne, G., Boyer, P. and Gravey, A., "Priority Management in ATM Switching Nodes", *IEEE Journal on Selected Areas in Networking*, vol. 9, no. 3, pp. 418–427, April 1991.

[19] May, M., Diot, C. "An experimental Implementation of Traffic Control for IP networks", *Proceedings of HPCS'97*, June 1997.

[20] McCanne S., Vetterli M., Jacobson V., "Low-complexity video coding for receiver-driven layered multicast", to appear in *IEEE Journal on Selected Areas in Communications*, 1997.

[21] Miller, L. W., "A Note on the Busy Period of an M/G/1 Finite Queue", *Op. Res. 23* (1975), pp. 1179–1182.

[22] Pullen, J., Myjak, M., Bouwens, C., "Limitations of Internet protocol suite for distributed simulation in the large multicast environment", Internet draft draft-ietf-lsma-limitations-01.txt, expires 9/97.

[23] *QNAP2 version 9.3 User's Guide*, SIMULOG, St Quentin-en-Yvelines, France, July 1996.

[24] Roberts, J., Mocci, U., Virtamo, O. (eds), *Broadband Network Teletraffic*, Lecture Notes in Computer Science, vol. 1155, Springer, 1996.

[25] Shenker, S., Partridge, C., and Guerin, R., "Specification of guaranteed quality of service", *Internet Draft*, Integrated Services WG, IETF, November 1996.

[26] Stoyan, D., *Comparison Methods for Queues and Other Stochastic Models*. English translation (D.J. Daley editor), J. Wiley and Sons, New York, 1984.

[27] Takács, L., "Priority Queues", *Op. Res. 12* (1964), pp. 63–74.

[28] Wolff, R.W., "Poisson arrivals see time averages", *Operations Research*, vol. 30, no. 2, pp. 223-231, April 1982.

[29] Wroclawski, J., Shenker, S., "Specification of the controlled load network element service", *Internet Draft*, Integrated Services WG, IETF, November 1996.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399