Distributed connection acceptance control for a connectionless network *

R. J. Gibbens and F. P. Kelly[a]

[a] Statistical Laboratory, University of Cambridge,
16 Mill Lane, Cambridge, CB2 1SB, UK.

We describe how a packet network with a simple pricing mechanism and no connection acceptance control may be used to carry a telephony-like service with low packet loss and some call blocking. The packet network uses packet marking to indicate congestion and end-systems are charged a fixed small amount per mark received. The end-systems are thus provided with the information and the incentive to use the packet network efficiently. We demonstrate that algorithms embedded in the end-systems are able to synthesize a telephony-like service by blocking calls at times when the load in the packet network is high.

1. INTRODUCTION

Proposals for an Internet with differentiated services are usually based on a small number of service classes with well defined quality and prices. However the approach has drawbacks (for a recent discussion, see [2,5]). In particular, the development of ATM traffic classes [17] has illustrated some of the difficulties of defining service categories before the applications that might use these categories have been invented or have become widespread, difficulties compounded by the requirement that associated pricing schemes be incentive-compatible for users ([14,18]).

A body of work is now emerging that takes a radically different approach to differentiated services ([5,9,12,13]). Its premise is that a simple packet network may be able to support an arbitrarily differentiated set of services by conveying information on congestion from the network to intelligent end-nodes, which themselves determine what should be their demands on the packet network. There would then be no need for large buffers or priority queues within the network, or for connection acceptance control at the border of the network. This paper is intended to contribute a further step in the exploration of this approach.

Gibbens and Kelly [9] have described how simple marking mechanisms may be used to distribute resource allocation decisions from the resources within a packet network to the end-systems connected to the network. The essential idea is that by appropriately marking packets at resources approaching overload, and by charging a fixed small amount for each mark received, end-systems can be provided with the necessary information and the correct incentive to use the packet network efficiently. This was illustrated in [9] for a variety of scenarios involving both elastic users able to smoothly adapt their demands in accordance with network load, and file transfers wishing to be completed as soon as possible subject to cost constraints.

*This is a preprint of a paper to be presented at the 16th International Teletraffic Congress in Edinburgh, June 1999. For pointers to related work see `http://www.statslab.cam.ac.uk/~frank/dcac.html`.

In this paper we develop further some of the consequences of this approach for a more complex scenario involving gateways interposed between the packet network and users. We consider an architecture for the provision of a telephony-like service over a packet network. One way to run such a network would be to have a centralized controller deciding whether or not to accept a new call, perhaps on the basis of the number of calls already in progress or of measurements of load on the packet network (see, for example, [8,10,11]). The aim of the centralized controller would be to assure reasonably low packet loss rates for accepted calls by occasionally blocking call requests. But is it necessary to have a centralized controller with its inherent drawbacks, or can the connection acceptance decisions be distributed?

In this paper we investigate whether connection acceptance control may be distributed to end-systems, with the packet network unaware even that the packets it carries come from higher-level entities identified as calls. In the scenario we describe, end-systems, called gateways, receive marks from the packet network and take decisions about whether or not to block arriving calls. We demonstrate that a gateway is indeed able to provide a telephony-like service, with low packet loss and some call blocking. Moreover, a gateway can improve its call blocking probability, if it is prepared to accept a higher packet marking rate from the network. Thus a gateway may effectively synthesize Quality of Service, in this case perceived by users in terms of call blocking, from a network providing simply packet transport.

The organization of the paper is as follows. In Section 2 we describe the queueing model which we use as a caricature of a transport network carrying packetized voice traffic. The model is familiar to teletraffic theorists, and our analytical approximations are taken from [16]. In Section 2 we also outline some strategies by which packets may be marked, to indicate congestion. Packet marking strategies are familiar to flow control specialists, and are an active research area ([3,6,7,9,12,15]). In Section 3 we present a simple model of a gateway, and show that gateways are able to effect a distributed connection acceptance control for the network. We compare their performance with the performance achieved by a centralized controller, roughly summarized by Erlang's formula. In Section 4 we consider heterogeneous gateways, and show how the call blocking probability achieved by a gateway is influenced by its willingness to pay. In Section 5 we briefly consider the case of a gateway which handles only an occasional call and is able quickly to probe the network before deciding to accept the call. Section 6 concludes.

2. MODELS OF PACKET MARKING

In this section we describe a simple queueing model, which we use to represent a packet network, and some possible marking mechanisms. (The underlying theory described in [9,12] generalizes to networks of many queues, but in this paper we confine attention to the simplest case.)

Consider a queue operating in discrete time with fixed size packets and with a finite buffer able to hold at most K packets. Thus the queue size Q_t evolves according to the recursion

$$Q_{t+1} = [Q_t + A_t - 1]_0^K \tag{1}$$

where A_t is the number of packets that arrive in slot t and $[a]_b^c = (a \wedge c) \vee b$. The packet arrival process is generated by n independent sources, each representing a call. Each source alternates between 'on' and 'off' periods, spending a proportion of time α in the 'on' state. Within 'on' periods the source produces a periodic stream of packets, with period D.

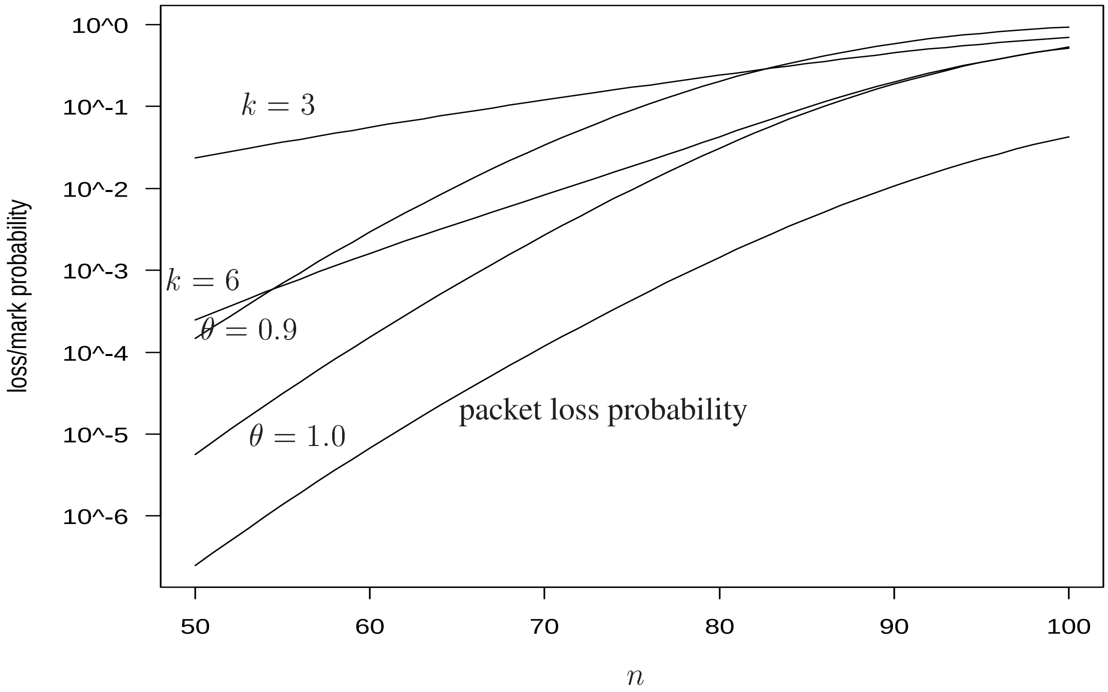

Figure 1. Loss and marking probabilities. $D = 50$, $\alpha = 0.5$, $K = 10$. The curves labelled $k = 3, 6$ correspond to strategies that mark all packets arriving to find a buffer containing k or more packets. The curves labelled $\theta = 0.9, 1.0$ correspond to strategies that mark from the detection of congestion until a buffer empties. For comparison, the packet loss probability is also included. This Figure is computed from analytical approximations described in the Appendix.

Queueing models of the above form have been the subject of intensive research over the last two decades: an excellent review and synthesis is provided by Roberts [16]. Using results outlined in the Appendix we can approximately calculate the packet loss probability for given values of the various parameters. In Figure 1 we illustrate the case $\alpha = 0.5$, $D = 50$, $K = 10$. For values of $n \leq 83$ the major cause of packet loss is near simultaneous packet arrivals from different sources that overwhelm the finite buffer. For values of $n \geq 84$ the possibility that more than 50 sources may be 'on', thus exceeding the service rate of the queue, is the major cause of packet loss. These two distinct causes of packet loss are called *packet-scale* and *burst-scale* effects respectively. The busy period preceding packet-scale loss is typically of the order $D/2$, that is half the period, while the busy period preceding burst-scale loss is typically several periods long.

Next suppose that the resource marks some of the packets passing through it, in accordance with a marking strategy. For example suppose that packets are marked when they arrive at a buffer containing k $(k < K)$ or more packets. The curves labelled $k = 3, 6$ in Figure 1 show the approximate packet marking probabilities.

In [9] an alternative marking strategy is described, that attempts to approximate the marking probability to the *shadow price* at the queue. Essentially the shadow price at the queue is the probability that an additional packet arriving at the queue will cause an extra packet loss either directly, through the loss of that packet itself, or indirectly, through the loss of an extra packet before the buffer next empties. The marking strategy is as follows: if a packet is lost, mark every packet leaving the buffer until the buffer next empties. An analytical approximation for the packet marking probabilities is labelled $\theta = 1.0$ in Figure 1.

Next we describe a variant on the above marking strategy designed to increase the proportion of marked packets and to provide early warning of congestion. Suppose the resource computes a virtual queue size, Q_t^θ, according to the recursion

$$Q_{t+1}^\theta = [Q_t^\theta + A_t - \theta]_0^{\theta K}, \tag{2}$$

and marks every packet served by the (real) queue from any time t' when $Q_{t'}^\theta + A_{t'} - \theta > \theta K$ until Q_t^θ next reaches zero. The virtual queue thus has a fraction θ of the service capacity of the real queue, and a fraction θ of the buffer of the real queue. The approximate packet marking probabilities under this strategy are shown for the cases $\theta = 0.9, 1.0$ in Figure 1. Observe that, as expected, a higher proportion of packets are marked. A fuller rationale for this marking strategy would involve a discussion of the *many sources* large deviations regime and its application to the detection of overload (see [3,4]). We simply note that the strategy is intended to be more robust to errors in the statistical characterization of arrival processes than strategies based simply on queue length in the real queue.

In Figure 2 we show how the marking probabilities of the different schemes vary with packet loss probability. This figure is computed from the data shown in Figure 1, and thus the various curves are implicitly parametrized by the number of calls in progress, n. In our later simulations we use the marking strategy corresponding to the parameter $\theta = 0.9$. For this strategy the marking probability rises quickly in the region of loss probability 10^{-4} to 10^{-3}, the region we focus upon in our later simulations.

3. GATEWAYS

In this section we consider the call arrival process. We suppose that there are G gateways, and that each gateway receives a Poisson stream at rate ν/G of call requests. The gateway must decide to accept or reject a call; if the call is accepted, it has a (independent) random holding time with mean h, during which it acts as one of the sources described in Section 2. The gateway receives the marks generated by its calls, and must decide whether or not to accept an arriving call according to its recent experience of marking probabilities.

Suppose a gateway estimates marking probabilities, P_t, by the following moving average method with gain parameter γ. Let

$$P_{t+1} = P_t + \gamma(M_t - P_t T_t) \tag{3}$$

where T_t is the number of packets transmitted by the gateway in slot t, and M_t is the number of marks received by the gateway in slot t. The gateway uses its estimate of the current marking probability to decide upon admission or otherwise of a call request. We suppose gateway j accepts a call request if its current estimate, P_t, is below its acceptance threshold A_j, or if gateway j has no calls currently in progress. (We allow a gateway to accept a call when it has no calls in progress in order that it may probe the network: without this possibility, a gateway

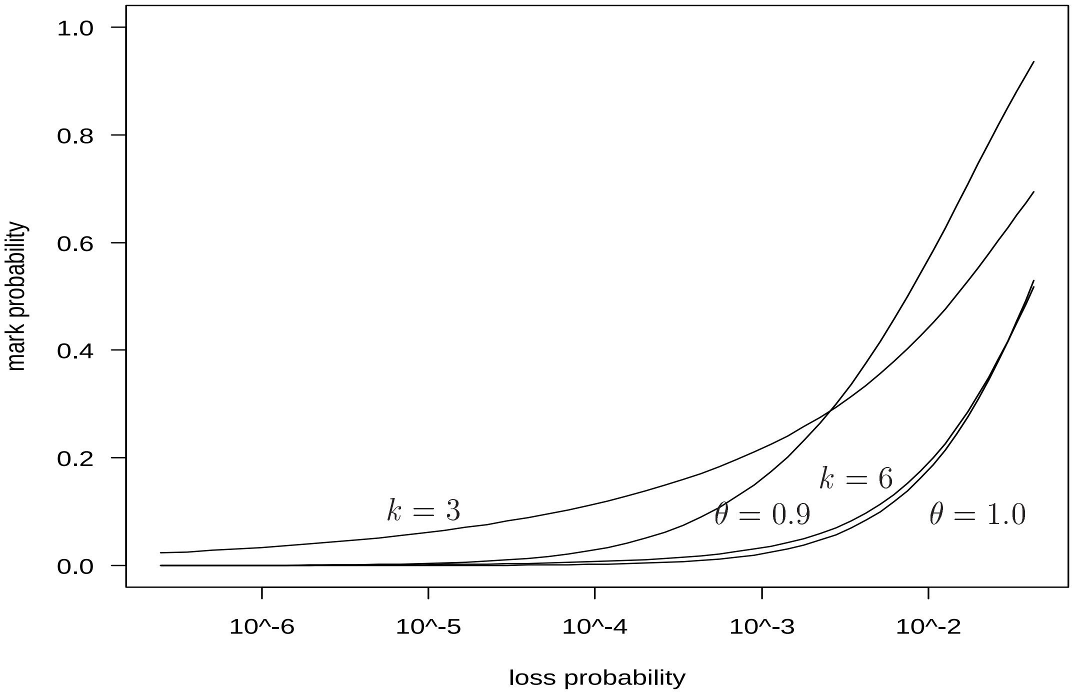

Figure 2. Marking probabilities vs loss. $D = 50$, $\alpha = 0.5$, $K = 10$. The marking strategy labelled $\theta = 0.9$ is chosen for later simulations. Observe that its curve rises steeply over the range of loss probabilities from 10^{-4} to 10^{-3}. This Figure is based on analytical approximations described in the Appendix.

could be locked out.) We choose $G = 10$, $\gamma = 1/1000$ corresponding to an averaging, by the iteration (3), over about 1000 packets, and set the thresholds, $A_j = 0.1$ for $j = 1, 2, \ldots, 10$.

We have simulated the system, with the parameters $\alpha = 0.5$, $D = 50$, $K = 10$. We take a slot length to be 0.5 ms, and 'on' and 'off' periods to be geometrically distributed with mean 1000 slots, or 500 ms. The mean number of packets transmitted during the 'on' period of a source is then 20. We take call holding times to be geometrically distributed with mean 400,000 slots, or 200 seconds. These numbers were chosen to be broadly comparable with those used in studies of both ATM and IP networks (see [16, Section 8.14], [1] respectively). We take the delay between a marked packet leaving the queue and knowledge of the mark reaching the source to be 100 slots, or 50ms (corresponding to a round trip of 15,000km at the speed of light).

Since Poisson arrivals see time averages, the probability an arriving call is blocked can be estimated by calculating the proportion of time a gateway's estimate, P_t, is below the gateway's threshold. This is shown in Table 1 for a range of values of ν, the offered load. (The offered load is measured in Erlangs, defined as the expected number of call arrivals in one mean holding time.) Table 1 also shows packet loss probabilities and packet marking rates. For rough comparison, Figure 3 compares blocking probabilities with Erlang's formula for a capacity of 70. (Table 1 and the solid line of Figure 3, and later tables and figures, describe the results of simulations.) From Figure 1 we deduce that a centralized controller which allows at most 80

Table 1
Packet loss probabilities and packet marking rates as the offered load, ν, increases from 50 to 100 Erlangs. Observe the rapid rise in the packet marking rate as the packet loss probability increases.

Offered load	$\nu = 50$	$\nu = 60$	$\nu = 70$	$\nu = 80$	$\nu = 90$	$\nu = 100$
Packet loss	0.00034%	0.014%	0.019%	0.024%	0.037%	0.044%
Packets marked	0.29%	2.2%	3.8%	5.7%	7.8%	8.6%

simultaneous calls in progress could hold packet loss probabilities below about 10^{-3}. Our distributed connection acceptance control achieves these packet loss probabilities or better for a relatively small capacity reduction, from 80 to 70.

4. HETEROGENEOUS GATEWAYS

When the offered traffic, ν, is higher, the blocking probability shown in Figure 3 is, naturally, higher. A gateway may well be prepared to pay more in order that it achieve a better quality of service (as measured here by blocking probability), even when the load on the network is high. In this Section we consider differential thresholds, $A_j = 0.02j$, $j = 1, 2, \dots, 10$.

The top panels in Figure 4 show the results of a simulation with a total load of $\nu = 70$ Erlangs (so that the expected number of call arrivals in a mean holding time is 70). The bottom panels correspond to a total load of $\nu = 100$ Erlangs.

Since Poisson arrivals see time averages, the probability a gateway blocks an offered call is determined by calculating the probability the gateway's estimate of P_t is above that gateway's threshold. This is shown, as a function of the number of calls in progress, as the right hand panels in Figure 4. Note that the blocking probabilities are strictly ranked (in decreasing order) by the gateway thresholds, A_j.

When $\nu = 70$ the gateways with higher thresholds block very few arriving calls; essentially they only block when the total number of calls in progress is dangerously high. When $\nu = 100$ the gateways with lower thresholds block most arriving calls, while the gateways with higher thresholds still manage to accept most arriving calls. Essentially, gateways with low thresholds only allow access to the network when there is relatively little load upon the network. The two highest curves in the bottom right panel of Figure 4 fluctuate, since they are estimated from little data: the number of calls in progress when $\nu = 100$ is rarely below 70 (as shown by the bottom left panel).

In Table 2 we show the achieved utilization, packet loss and proportion of packets marked when $\nu = 70$ and 100. Observe that the average cost of using the network (the proportion of packets marked) is about 2.5 times higher when $\nu = 100$ than when $\nu = 70$. Figure 5 shows the blocking probabilities for various gateways under the traffic loads $\nu = 70$ and $\nu = 100$. Observe again that the gateways with higher thresholds have low blocking probabilities: the gateways with lower thresholds block more traffic, especially when the network is overloaded ($\nu = 100$).

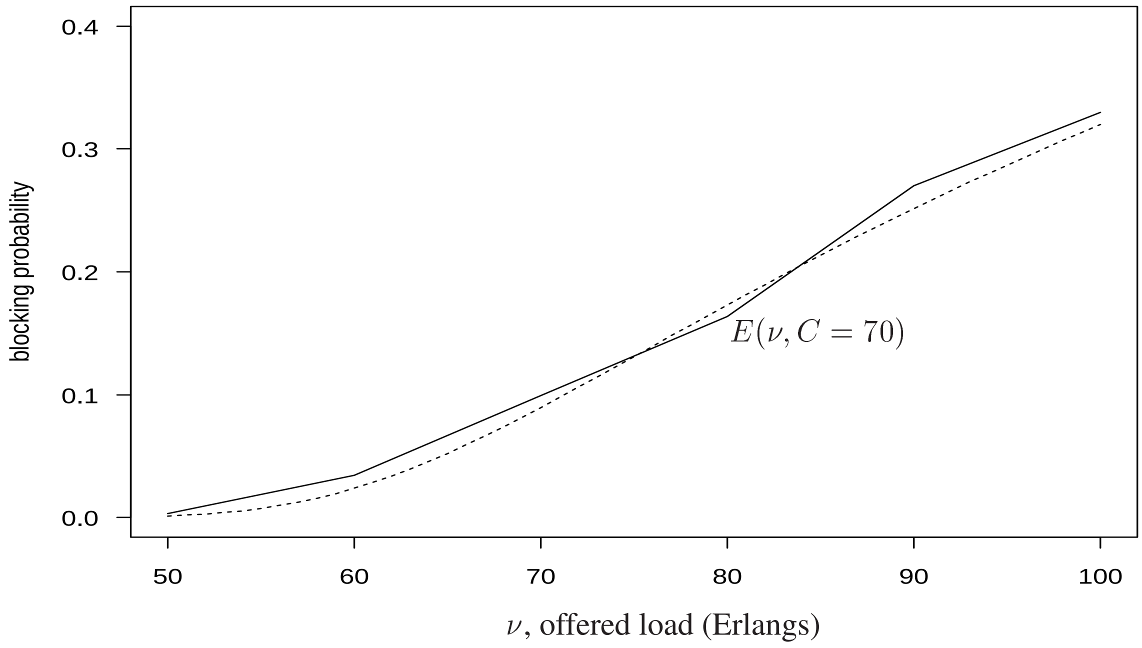

Figure 3. Blocking probabilities compared with Erlang's formula. The dotted line shows the function $E(\nu, C = 70)$.

5. PROBING USERS

The gateways studied in the previous sections each handle a substantial stream of traffic. In this Section we briefly consider the case of a gateway or user which handles only an occasional call, but is able quickly to probe the network before deciding to accept the call.

Consider the example of Section 4 with 10 gateways with differentiated thresholds, A_j, together with 5 additional users not attached to any such gateway. Instead each of these additional users probes the network for a short period whenever a new connection for that user is attempted. During the probing period of 200 time slots the user sends a periodic stream of packets one every 10 time slots. At the end of the probing period the call is established if the total number of marks received during the probing period is less than a threshold value, a_j, for prob-

Table 2
Performance measures for two scenarios with total offered loads of $\nu = 70$ and $\nu = 100$ and heterogeneous gateways. Utilization is defined as the traffic intensity of the queue server.

Offered load	$\nu = 70$	$\nu = 100$
Utilization	71%	75%
Packet loss	0.012%	0.032%
Packets marked	3.3%	7.4%

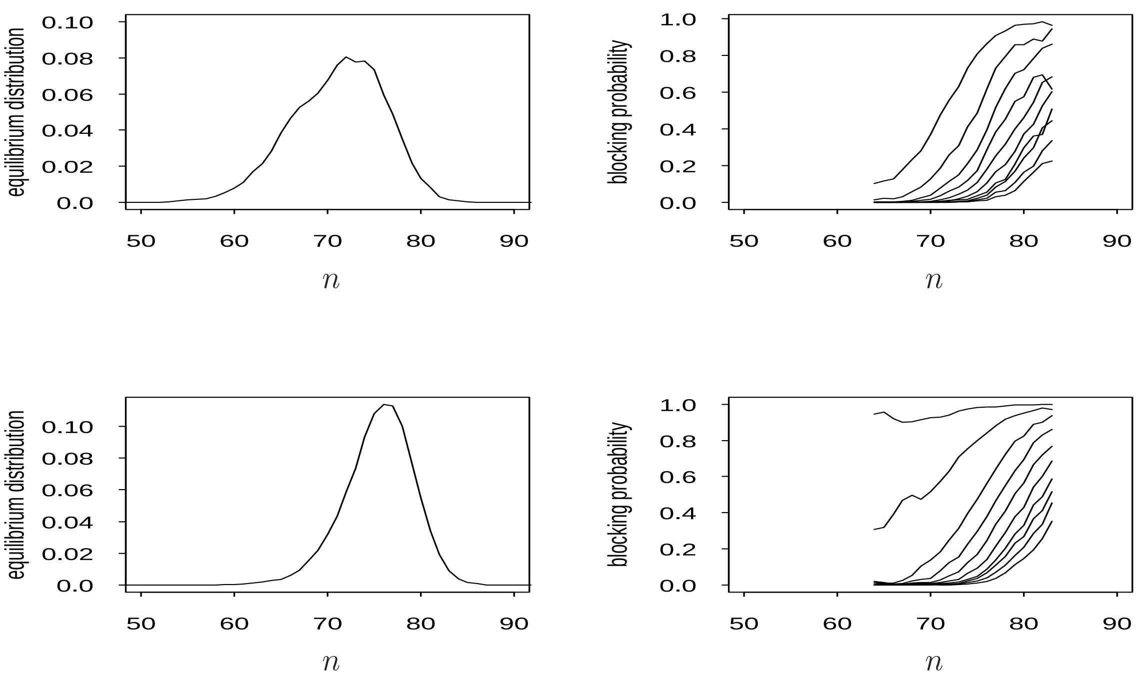

Figure 4. Equilibrium distribution and blocking probabilities in the case $D = 50$, $\alpha = 0.5$, $K = 10$, $\gamma = 1/1000$, $\theta = 0.9$ and 10 gateways with varying thresholds equally spaced between 0.02 and 0.2. The top panels correspond to an offered load of 70 Erlangs, the bottom panels an offered load of 100 Erlangs. This Figure is computed from a simulation, as described in Section 4.

ing user $j = 1, \ldots, 5$. If the number of marks received equals or exceeds the probing user's threshold, a_j, then the connection is rejected. (Note that because the round-trip time is 100 slots, a decision is taken after only ten packets have had the possibility of their marks being returned.) After a call attempt is rejected or a call is completed, the user waits for a geometrically distributed period with mean 400,000 slots before attempting another call. A call produces a stream of packets in the same manner as in earlier Sections.

Table 3 shows the measured blocking probabilities and average charges for the choice of threshold values $a_j = j$ for $j = 1, \ldots, 5$. The results were obtained from a scenario where the offered load of the gateways varied over the range $\nu = 40, 50, 60, 70, 80$, spending an equal amount of time at each value. Notice that as the threshold, a_j, is increased the user is more successful at making connections but will broadly speaking pay a higher price. Essentially the lower thresholds tend to block calls when the offered load, ν, is high, and this leads to lower average charges.

Including the round-trip delay of 100 slots, the strategies described take 200 slots, or 100ms, to reach a decision. Generally a speech call will comprise a two-way connection: co-ordination of the acceptance decision between two end-systems may add perhaps another one or two round-trip delays. There is clear trade-off between the time allowed for probing and the accuracy of

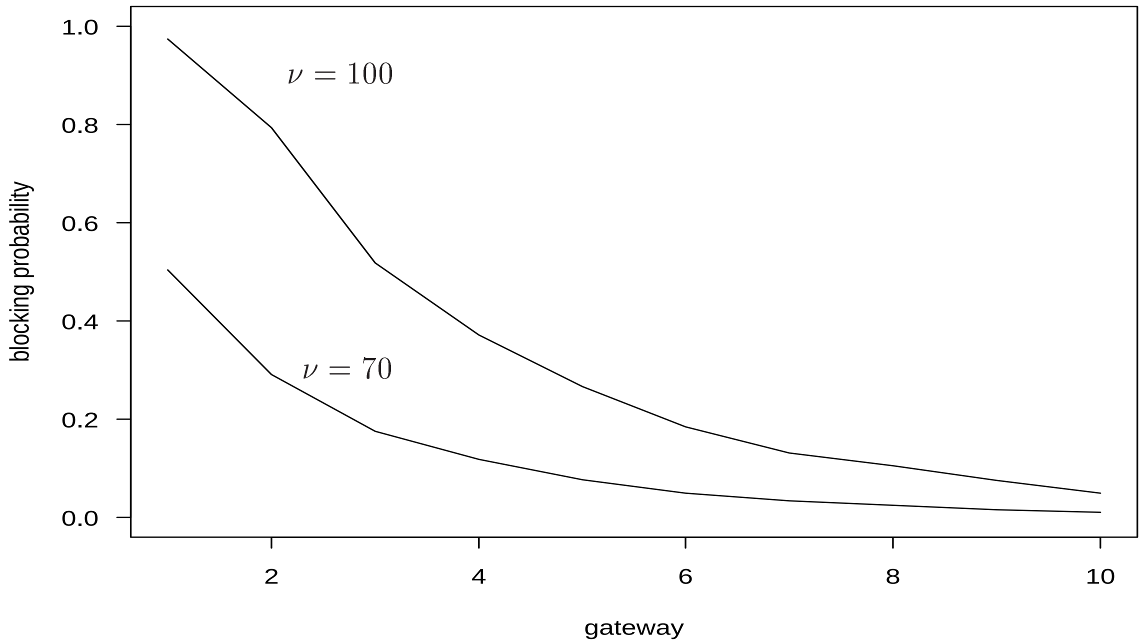

Figure 5. Blocking probabilities for different gateways with $\nu = 70$ and $\nu = 100$.

Table 3
Blocking probabilities and average charges for probing users with the choice of threshold values $a_j = j$ for $j = 1, \dots, 5$.

Threshold, a_j	1	2	3	4	5
Blocking	58.9%	32.7%	18.4%	11.0%	10.4%
Packets marked	2.5%	3.8%	4.4%	4.7%	4.7%

any estimation procedures, but the very simple strategies of this section suggest that a connection set-up time of half a second would be ample.

6. CONCLUSIONS

We have seen that a gateway may provide a telephony-like service, with low packet loss and some call blocking, from a network providing packet transport. We expect that different gateway strategies will be of some interest, and may be explored in the context of the distributed game described in [13]. It is important to note that the role of the gateway is informational rather than physical: the aim is to infer expected prices from observations on the network. Thus the gateway itself may be embedded in software and distributed over users.

We have defined Quality of Service, as perceived by users, in terms of call blocking. More general definitions are, of course, possible. For example, the heterogeneous gateways of Section 4 may be differentiated not just by call blocking, but also by the packet rate allocated to an accepted call. Or the strategy of Section 5 mght choose to allocate an accepted call a high or moderate packet rate depending on the proportion of probe packets marked. Quality of Service for a user might then depend on the probability an accepted call has high or moderate speech quality, as well as the probability a call is accepted.

A user may also require a telephony-like service to have a fixed, known charge per unit time for an accepted call. But to provide such a service a gateway need only choose the charge per unit time and the threshold value so that on average the gateway is profitable: the gateway operator would then undertake the various risks associated with predicting traffic conditions. More generally, we think it possible that simple marking and pricing schemes may be sufficient to allow gateways to synthesize Quality of Service, however defined by the user, from an underlying packet network.

ACKNOWLEDGEMENTS

Richard Gibbens was supported by a Royal Society University Research Fellowship and additional support was provided by the EPSRC under grant GR/M09551.

REFERENCES

1. S. Bajaj, L. Breslau, and S. Shenker. Is service priority useful in networks?, 1998. http://www.cs.wpi.edu/∼sigmet98/bajaj.ps.
2. D. D. Clark. Adding service discrimination to the Internet. *Telecommunications Policy*, 20, 1996. http://ana-www.lcs.mit.edu/anaweb/abstracts/TPRC2-0.html.
3. C. Courcoubetis, G. Kesidis, A. Ridder, J. Walrand, and R. R. Weber. Admission control and routing in ATM networks using inferences from measured buffer occupancy. *IEEE Transactions on Communications*, 43:1778–1784, 1995.
4. C. Courcoubetis and R. Weber. Buffer overflow asymptotics for a buffer handling many traffic sources. *J. Appl. Prob.*, 33:886–903, 1996.
5. J. Crowcroft and P. Oechslin. Differentiated end-to-end Internet services using a weighted proportionally fair sharing TCP. *ACM Computer Communications Review*, 28:53–67, 1998.
6. S. Floyd. TCP and Explicit Congestion Notification. *ACM Computer Communications Review*, 24:10–23, 1994. http://www-nrg.ee.lbl.gov/floyd/ecn.html.
7. S. Floyd and V. Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Trans. Networking*, 1:397–413, 1993. ftp://ftp.ee.lbl.gov/papers/early.pdf.
8. R. J. Gibbens and F.P. Kelly. Measurement-based connection admission control. In *15th International Teletraffic Congress*, volume 2b, pages 879–888. Elsevier, 1997.
9. R. J. Gibbens and F.P. Kelly. Resource pricing and the evolution of congestion control, 1998. http://www.statslab.cam.ac.uk/∼frank/evol.html.
10. R. J. Gibbens, F.P. Kelly, and P. B. Key. A decision-theoretic approach to call admission control in ATM networks. *IEEE Journal on Selected Areas in Communications, special issue on Advances in the Fundamentals of Networking*, 13(6):1101–1114, 1995.
11. M. Grossglauser and D. N. C. Tse. A framework for robust measurement-based admission control. In *Proc. ACM Sigcomm97*, 1997. Cannes, France.
12. F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998. http://www.statslab.cam.ac.uk/∼frank/rate.html.

13. P. Key and D. McAuley. Differential QoS and pricing in networks: where flow control meets game theory. *IEE Proceedings Software*, 146(2), 1999.
14. M. Mandjes and N. van Foreest. Guidelines for pricing in an integrated services network. Submitted to ITC16, 1998.
15. K. K. Ramakrishnan and R. Jain. A binary feedback scheme for congestion avoidance in computer networks. *ACM Trans. Computer Systems*, 8:158–181, 1990.
16. J. W. Roberts, editor. *COST224 Performance evaluation and design of multiservice networks*. Commission of the European Communities, October 1992. Final Report.
17. International Telecommunications Union. Recommendation I371: Traffic control and congestion control in B-ISDN. Geneva, 1996.
18. D. Walker, F. P. Kelly, and J. Solomon. Tariffing in the new IP/ATM environment. *Telecommunications Policy*, 21:283–295, 1997.

APPENDIX

The analytical approximations we employ are described in detail in [16, Sections 6.2 and 8.2]. Here we briefly outline the form of the approximations.

Consider a queue operating in slotted time with fixed size packets and with a finite buffer able to hold at most K packets. Each of n independent sources produces bursty on/off traffic; the proportion of time a source is in the ‘on’ state is α. Within ‘on’ periods the source produces a periodic stream of packets, with period D.

We approximate the rate of packet loss by

$$R_{\text{loss}}(K, D; n) = R_{\text{packet-scale}}(K, D; n) + R_{\text{burst-scale}}(K, D; n) \tag{4}$$

where

$$R_{\text{packet-scale}}(K, D; n) = \sum_{m=1}^{n \wedge D} \binom{n}{m} \alpha^m (1-\alpha)^{n-m} Q_D^m(K) \tag{5}$$

and

$$R_{\text{burst-scale}}(K, D; n) = \sum_{m=n\wedge D+1}^{n} \binom{n}{m} \alpha^m (1-\alpha)^{n-m} \frac{m-D}{D}. \tag{6}$$

Here $Q_D^m(K)$ is the known exact solution (given by equation (6.2.2) of [16]) for the tail probability of an $m * D/D/1$ queue,

$$Q_D^m(K) = \sum_{K<n\leq m} \binom{m}{n} \left(\frac{n-K}{D}\right)^n \left(1 - \frac{n-K}{D}\right)^{m-n} \frac{D-m+K}{D-n+K}. \tag{7}$$

Equation (5) approximates the packet- or cell-scale loss rate while the number of sources in the ‘on’ state, n, satisfies $n \leq D$. When $n > D$, the traffic intensity at the queue is greater than one; equation (6) approximates the corresponding burst-scale loss rate. The overall loss rate (4) is just the sum of these two components, and the proportion of packets lost is correspondingly approximated by

$$\frac{D}{n\alpha} R_{\text{loss}}(K, D; n). \tag{8}$$

This function is plotted, marked 'packet loss probability', in Figure 1. Packet-scale loss is the major contributor to the packet loss probability for $n \leq 83$, while burst-scale loss is the major contributor for $n \geq 84$.

Suppose next that packets are marked when they arrive at a buffer already containing k $(k < K)$ or more packets. Then the packet marking rate may be similarly approximated by

$$R_{\text{mark}}(k, D; n) = R_{\text{packet-scale}}(k, D; n) + \sum_{m=n \wedge D+1}^{n} \binom{n}{m} \alpha^m (1-\alpha)^{n-m}.$$

The proportion of packets marked is correspondingly approximated by

$$\frac{D}{n\alpha} R_{\text{mark}}(k, D; n);$$

this expression is plotted, labelled $k = 3, 6$, in Figure 1.

Next we estimate the shadow price of a packet (for background on shadow prices in this context see [9,12]). An additional source, increasing the number of sources from n to $n+1$, will increase the packet loss rate from $R_{\text{loss}}(K, D; n)$ to $R_{\text{loss}}(K, D; n+1)$. Since this additional source produces on average α/D packets per slot, the shadow price of each single packet is approximately

$$\frac{D}{\alpha} \big(R_{\text{loss}}(K, D; n+1) - R_{\text{loss}}(K, D; n) \big). \tag{9}$$

This function is plotted, labelled '$\theta = 1.0$', in Figure 1. More generally, the functions labelled θ in Figure 1 are given by replacing K, D by θK, θD respectively in expression (9).