Herding cats:? modelling Quality of Service for Internet applications

Jon Crowcroft

Phil. Trans. R. Soc. Lond. A 2000 **358**, 2209-2215
doi: 10.1098/rsta.2000.0642

10.1098/rsta.2000.0642

THE ROYAL SOCIETY

Herding cats:† modelling Quality of Service for Internet applications

By Jon Crowcroft

Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK (jon@cs.ucl.ac.uk)

The art and science of tele-traffic modelling is quite mature. On the other hand, Internet traffic seems to defy all attempts to capture its essence in simple models. This is not so surprising when we consider that the Internet consists of a large number of self-organizing systems, each evolved almost independently, which is quite a different way to construct a network than the ground-up design associated with telecommunications. IP routing, TCP congestion control, Relative Transport Protocol playout and loss adaption, Web caching and load balancing, and user behaviour are all involved in a system of massive complexity.

In this paper, we survey some of these mechanisms and some of the attempts to bring this unruly bunch of schemes into a more coherent whole. We argue that these attempts are misguided, and that the strength of the Internet design is in the loose organization of these components. As the commercial investors turn their eyes on the Internet with a view to pricing, we argue that they should take extreme care not to propose mechanisms that kill the goose that lays the golden egg.

Keywords: Internet; modelling; Quality of Service; performance engineering

1. Introduction

In this paper, we look at some of the problems that confront the network modeller when studying applications traffic in the Internet.

The paper is organized as follows. In the next section we look at the evolution of the traffic matrix and the network topology; following that, we look at typical Internet traffic source behaviour; we then look at the evolution of congestion control and adaption; this is followed by an overview of proposals for integrated and differentiated services for the Internet; a brief discussion of proposals for pricing and aggregation ensues; finally we summarize the paper.

2. Traffic matrix

The Internet is growing very rapidly. It would appear that in some places, anecdotally (Network Laboratory for Applied Network Research) it has been reported that the traffic from so-called 'hot spots' doubles every 14 weeks.

To accommodate this, the backbone providers (e.g. the Very-high-speed Backbone Network Service) are deploying capacity and technology at an incredible rate. In

† 'At Group L, Stoffel oversees six first-rate programmers, a managerial challenge roughly comparable to herding cats.' *The Washington Post Magazine*, 9 June 1985.

Table 1. *Terms in the Parekh TCP equation*

d	worst case delay
b	token bucket size
g	allocated rate
N	hopcount
P_{F}	this flow maximum packet size
P_{N}	networks maximum packet size
r	link speed

some cases, testing of opto-electronic routers is already three years in advance of predictions from recent research.

Traffic types are also varying as new applications such as streamed multimedia emerge; for example, Internet traffic radio growth is roughly as the World Wide Web was five years ago. New traffic has new source models and new distribution models, and to accommodate this, the network link deployment, capacity provisioning, and very routing algorithms themselves are evolving on a yearly basis. In 1997 there was only one multicast routing protocol, the Distance Vector Multicast Routing Protocol, deployed. Now there are at least three, which form different topologies as the transmitters and receivers start and stop dynamically.

Application level mechanisms such as Web mirroring and caching started as minor optimizations. However, they can have dramatic effects on network loading. Currently, downloads from Microsoft's Seattle Web site saturates their 1.2 Gb s^{-1} Internet access line. It is estimated (personal communication from Jim Gemmell at Microsoft Research) that if this site were properly mirrored around the world, the total aggregate from the master site could be accommodated by a 10 kb s^{-1} line.

Such factors are just examples of the ways in which the Internet taxes the resources and imagination of the modeller of traffic distribution.

3. IP: source description

Looking at a more microscopic level at the behaviour of a single source, we find that the complexity of the Internet reveals itself again. Researchers have been trying to push the jack-in-the-box back into its box by proposing Integrated and Differentiated services mechanisms which entail traffic source specification, admission tests, policing, and possible shaping. These are unlikely to see deployment for some time. We look briefly at these proposals, then look at what the real traffic currently looks like.

(*a*) *Quality of Service myths*

Work on Quality of Service (QoS) of the Internet was dominated in the 1990s by the classical telecommunications approach: define the source model, observe the traffic matrix, and provide quality assurance by constraining the sources that fit the network resources provisioned. How much bandwidth must be reserved at a weighted fair queuing server to guarantee a particular end-to-end delay bound? One solution for a bound on this can be obtained by inverting Parekh and Gallager's well-known

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

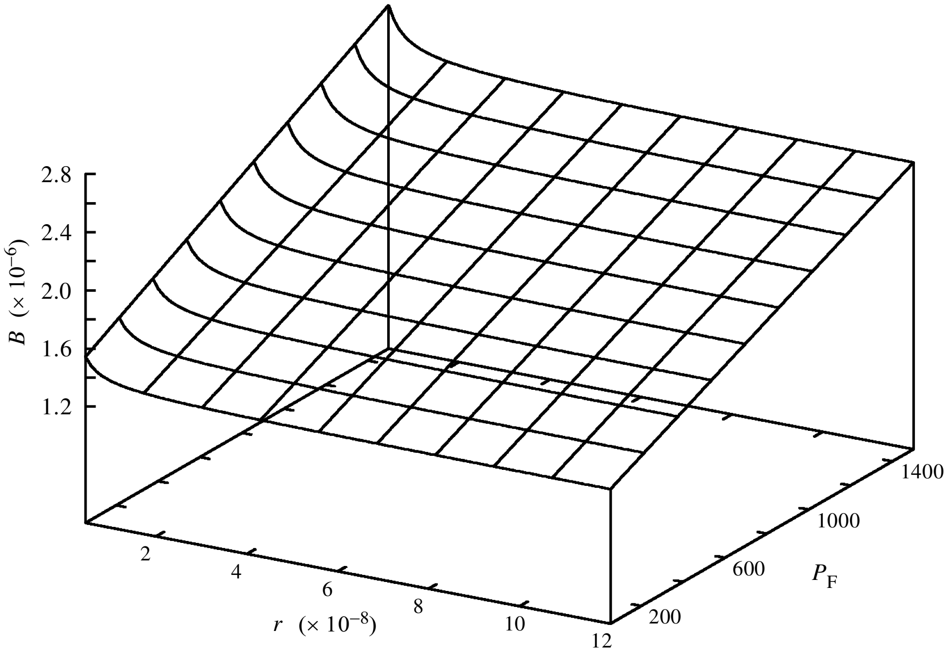

Figure 1. Guaranteed bounded delay and utilization
$((128\,000 + 10^{-1}(y \times 8))/(0.1 - (10 \times 65\,536/x)))$.

equation (Parekh 1993):

$$d = \frac{b}{g} + (N-1)\frac{P_{\mathrm{F}}}{g} + N\frac{P_{\mathrm{N}}}{r}. \tag{3.1}$$

One example of this is illustrated in figure 1, which plots the capacity needed for typical 64 kb s^{-1} pulse code modulated coded voice calls packetized using Relative Transport Protocol (RTP) and a range of typical Internet line speeds, a delay bound as needed for interactive voice, and the over-allocation needed. Note the presence of hop count as part of the equation.

The problems with this equation-based source specification are many and various. Some include the following.

Bucket rate. An application can choose to send at many different rates. For example, audio, video, gaming and data applications can all be configured to correspond loosely to some user perceived quality. However, this correspondence has to be characterized for each and every application, and for a wide range of values of the data rate. As new coding and compression techniques emerge rapidly, this means that it is very hard for the user to choose a meaningful number here.

Burstiness. Just as the rate is hard to assess, the burstiness is even harder; this is also dependent on content as well as implementation.

Packet size. The packet size, and other parameters of note are an artefact of the network design as well as host computer software. Why should the user be concerned with choosing or setting these?

Table 2. *Terms in the Padhye TCP equation*

W_{m}	maximum advertised receive window
T_0	initial timeout value
RTT	the round trip time
b	the number of packets acknowledged by 1 ACK
p	the mean packet loss probability
B	the throughput achieved by a TCP flow

One problem with this type of source model is that it is over-specified. Another, more serious, problem is that it appears to be oriented more towards optimizing for the network provider, not the user. Lastly, asking the user to state parameters, including network-specific ones, requires the provider to reveal internal factors about its network; this may undermine its competitive edge with other providers.

(*b*) *Reality*

In the real world today, we have three basic application types: TCP-based ones such as the World Wide Web, RTP-based ones such as Real Audio, and Reliable Multicast ones. All of these adapt to network conditions, employing a family of congestion avoidance algorithms loosely oriented around the same results for stable delayed-feedback control systems.

RTP/UDP applications adjust to delay variation and to packet loss; up to 50% of packets may be lost without compromising voice comprehension in some tools. This means that even if we were to offer QoS, the source does not need a hard bound; it can choose a trade-off between delay, as in Schulzrinne's work on the Internet RTCP protocol, and loss as in the self-organizing transcoder of Kouvelas *et al.* (1998).

Looking at some detail at TCP, latest theory and measurement by Padhye *et al.* (1998), shows how it really behaves.

$$B = \min\left(\frac{W_{\mathrm{m}}}{\mathrm{RTT}}, \frac{1}{\mathrm{RTT}\sqrt{(\frac{2}{3}bp)}} + T_0 \min(1, 3\sqrt{(\tfrac{3}{8}bp)})p(1+32p^2)\right). \tag{3.2}$$

This equation is for long-lived transfers. There is more recent work by Cardwell *et al.* (2000) on short-lived transfers (which the majority of the Web usage consists of today). However, just looking at this equation for now, we can see a lot of single instance specific parameters which affect the performance radically. For example, W_{m}, and the range of T_0 and b in implementations could be quite large. Another problem here which we should consider when thinking about pricing is the range of values for RTT; in some theoretical work, RTT is used as a measure of resource use. However, it is actually a real value measured from the path delay, and as such, is suspect when used for comparability of, for example, satellite versus terrestrial hops. Another problem with such proposals is that the loss rate p is considered as congestion feedback, and therefore as a proportional resource utilization indication. However, it depends on the instantaneous load in the queue seen by a single packet, not the average over some period. Loss is also due to interference on wireless networks, and so is a very noisy 'signal'. Proposed Internet replacements for loss such as Explicit Congestion Notification also need specifying with regard to the sampling interval: if a switch measures congestion over some interval, this must be known to the sources that

Table 3. *Terms in the Erlang equation*

B	blocking probability
A	offered load from this call
N	number of calls

receive congestion feedback, out of band, so that they can estimate the significance of some number of packets arriving with the ECN bit set.

Reliable multicast protocols (for games, share dealing and software and news distribution) such as the family of protocols using Handley *et al.*'s (1999) 'TCP Friendly Multicast Congestion Control' use similar adaptation techniques to that which TCP employs. However, one additional level of complexity is that these protocols also have self-organizing repair server mechanisms. These will alter the traffic patterns again.

Similar evolution is occurring in the streamed multimedia traffic flows that emanate from Internet Radio and TV sites, with TCP-like adaption being the apparent goal for Real Audio and Video, the most commonly used product for such services in the Internet.

4. IP: congestion feedback

Instead of offering a prescriptive approach to QoS, the more successful approaches in the Internet to date offer a descriptive one. This is achieved by giving congestion feedback; currently, this is implemented in Jacobson's (1988) scheme, by treating packet loss as such a signal, but in future it is likely to be more efficiently conveyed by Floyd's explicit signals. It is possible to offer a range of QoS behaviours in an adaptive environment. One proposal for doing this was evaluated by Oechslin recently.

5. TCP, RSVP and Erlang: blocking feedback

The alternative to adapting all the flows is to block some. The telephone network offers call blocking for some so that others may get through, and table 3 shows the classical formula used to derive a blocking probability.

$$B = \frac{(A/N!)^N}{\sum_{i=1}^{N}(A^i/i!)}. \tag{5.1}$$

Of course, we must consider the price of a call versus the discontent of being blocked. Such a system is a closed-form solution (also known as a 'self-fulfilling prophecy'). The price is set so that the calls get through to where you want them to go: the informal '3 + 3 + 3' model (often anecdotally cited as the ratio of call to idle, local to long distance, and short to long calls) is not appropriate to a rapid evolution of service sites in a network. Roberts and, separately, R. Mortier & I. Pratt (personal communication, 2000) have proposed adding call admission for TCP, or Reliable Multicast or RTP flows.

Another problem with this naive approach is that the signalling and blocking themselves form a type of traffic; this traffic must be congestion controlled, or recall attempts will flood the network just as effectively as the original data traffic did before a signalling and call admission control scheme was interposed.

Adding call admission to TCP has been suggested (Roberts 1999). However, this will then entail adding congestion control to whatever signalling protocol is used.

6. Pricing and feedback: content stagnation, user discontent

Internet providers have to some extent been competing on price. This has recently reached almost ludicrous extremes, where not only are some providers, such as Freeserve, apparently free, but some even go so far as to offer incentives to users to join their service. In fact, of course, this is a temporary measure based in the artifices of telephony pricing and the access networks. The UK regulator allows any licensed carrier that terminates a call to share the revenue from the call. 'Free' Internet Service Providers have registered as telephone companies so that they can take advantage of the billing infrastructure associated with this facility.

In the long run, we might expect reversion to the crude pay-per-use by leasing by the month by the access line rate. Why is this more tenable that usage (by time, volume and distance) charging in more traditional networks?

I would argue that it is in fact just a more coarse-grained version of the same thing, but that it allows for faster evolution of the network and traffic matrix, as well as having far lower revenue collection overheads.

Work on more detailed specification of how users behave when faced with a more dynamic price is not largely favourable, as found by Cowley (1998) and Bouch and co-workers (Bouch *et al*. 1998; Bouch & Sasse 1999). One exception is the INDEX project at Berkeley, which is reported by Chu & Altmann (this issue) and which shows some user inclination towards 'getting what you pay for'. However, anecdotal reports at a recent workshop by Hoadley (1999) on the SuperJANET volume charging are that user behaviour was *not* significantly affected.

Again, looking at the big picture, pushing the congestion signal right up to the user may help. However, we then need to model user (and societal) behaviour in the feedback loop. The Internet Providers already see a demand for 'turbo click', and it is a small step from there to 'turbo-provider-selection'. These also increase Internet 'turbulence'.

7. Aggregation

It is often suggested that while single sessions or flows of traffic do not exhibit spatial or temporal locality, perhaps aggregations of flows do. Techniques for aggregation abound in addressing, routing, resource allocation and so on. However, the evidence from traffic flows on the VBNS does not support this idea. Any attempt to match a hierarchical assignment of capacity to the traffic matrix finds a poor result, and the evolution of the traffic matrix over time does not appear to be making the match any better. Not only do the hot spots (in terms of sources and sinks of data) move over time, but also, any attempt to fit a straight-jacket around the traffic has to deal with the fact that the interconnect topology is not fixed: currently, the Internet uses a dynamic, opportunistic, single-path-distributed routing scheme that scales very well to accommodate rapid growth of the network, but does not admit of multi-path multi-metric routing at all easily. The dynamics of the routing tends to interfere with attempts to aggregate flows that are destined between 'similar' locations.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF

8. Summary and conclusions

We have discussed feedback in Internet Protocols. There is a wide variety of ways that congestion in the network provides a signal all the way to the user. The challenge for the modeller is really in tackling large-scale system design that needs to have economic modelling. This seems to be rather complex, and beyond current theory. In the meantime, the Internet Engineering Task Force continues to build more and more adaptive systems. Other authors, such as Paxson & Floyd (1997), K. Nichols (1999, unpublished report) and Crowcroft *et al*. (1999), have commented on the difficulty of modelling in such a system. Eventually, the overall system may be more amenable to analysis using techniques for dealing with very large adaptive systems such as the ecology, or thermodynamical systems. Finally, one very speculative economic hypothesis we might form (and is common dogma in the Telecommunications Industry) is that in an evolutionary system which exhibits rapid growth, it may be better to maximize revenue from new services than to optimize for stable services.

Endnote

See http://www-mash.cs.berkeley.edu/ns/ for details about the network simulator.
National Laboratory for Applied Network Research at http://www.nlanr.net.
VBNS Network Topology at http://www.vbns.net/logical.html.

References

Bouch, A. & Sasse, M. A. 1999 It ain't what you charge it's the way that you do it: a user persepctive of network QoS and pricing. *IFIP/IEEE Int. Symp. on Integrated Network Management (IM '99), Boston, MA, 24–28 May 1999.*

Bouch, A., Watson, A. & Sasse, M. A. 1998 QUASS—a tool for measuring the subjective quality of real-time multimedia audio and video. Poster presented at *HCI '98, 1–4 September 1998, Sheffield, England.*

Cardwell, N., Savage, S. & Anderson, T. 2000 Modeling TCP latency. In *IEEE Infocom 2000.* Also available in longer form, see http://www.cs.washington.edu/homes/cardwell/quals/.

Cowley, S. J. 1998 Of timing, turn-taking, and conversations. *J. Psycholing. Res.* **27**, 541–571.

Crowcroft, J., Kouvelas, I. & Vicisano, L. 1999 HYbrid Monitoring and Network Simulations—HYMNS. See http://www-iri.hpl.hp.com/.

Handley, M., Floyd, S. & Whetten, B. 1999 Strawman specification for TCP friendly (reliable) multicast congestion control (TFMCC). IRTF reliable multicast working group.

Hoadley, K. 1999 SuperJANET pricing. See http://bill.ja.net.

Jacobson, V. 1988 Congestion avoidance and control. *Proc. ACM SIGCOMM'88, August 1988, Stanford, CA*, pp. 314–329.

Kouvelas, I., Hardman, V. & Crowcroft J. 1998 Self organising transcoders. *Proc. NOSSDAV 1998*. See ftp://cs.ucl.ac.uk/darpa/sot.ps.gz.

Padhye, J., Firoiu, V., Towsley, D. & Kurose, J. 1998 Modelling TCP throughput: a simple model and its empirical validation. *Proc. SIGCOMM '98, Vancouver, CA, September 1998.*

Parekh, A. 1993 Generalized processor sharing. PhD thesis, MIT Laboratory for Decision Systems, MIT.

Paxson, V. & Floyd, S. 1997 Why we don't know how to simulate the Internet. *Proc. Winter Simulation Conf., December 1997.* See http://ftp.ee.lbl.gov/floyd/papers.html.

Roberts, J. 1999 Call admission control for TCP. *ITC 99, Edinburgh, UK, June 1999.*