Herding Cats†: - Modelling QoS for Internet Applications

By Jon Crowcroft jon@cs.ucl.ac.uk,
http://www.cs.ucl.ac.uk/staff/jon

Department of Computer Science
University College London
Gower Street
London WC1E 6BT

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-organising 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, RTP 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 organisation 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, QoS, Performance Engineering

1. Introduction

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

The paper is organised 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 ensue; finally we summarise the paper.

---

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

2. Traffic Matrix

The Internet is growing very rapidly. It would appear that in some places, anecdotally (NLANR) 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 VBNS) are deploying capacity and technology at an incredible rate. In some cases, they are testing state of the art opto-electronic routers five years ahead of predictions for research output would have predicted only 2 years ago.

Traffic types are also varying as new applications such as streamed multimedia emerge; for example, Internet traffic radio growth is roughly as WWW 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, DVMRP, 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 optimisations. However, they can have dramatic effects on network loading. Currently, downloads from Microsoft's Seattle web site saturates their 1.2Gbps 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 accommodate by a 10Kbps 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) QoS Myths

Work on QoS in 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 WFQ 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 equation:

$$d = \frac{b}{g} + (N-1)\frac{P1max}{g} + N\frac{P2max}{r} \qquad (3.1)$$

Table 1. *Terms in the Parekh TCP Equation*

d ; worst case delay
b ; token bucket size
g ; allocated rate
N ; hopcount
$P1max$; this flow maximum packet size z
$P2max$; networks maximum packet size
r ; link speed

One example of this is illustrated in figure 1, which plots the capacity needed for typical 64kbps PCM coded voice calls packetised using 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 - audio, video, gaming and data applications can all be configured to corresponding loosely to some user perceived quality. However, this correspondence has to be characterised 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 artifact of the network design as well as host computer software. Why should the user be concerned with choosing or setting these?

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 optimising 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 their network - this may undermine their competitive edge with other providers.

(b) Reality

In the real world today, we have three basic application types: TCP based ones such as WWW, 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 controlled 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 the Internet RTCP protocol, and loss as in Kouvelas' self organising transcoder.

(128000 + (10-1)*(y*8))/(0.1 - (10*65536/x)) ⸻
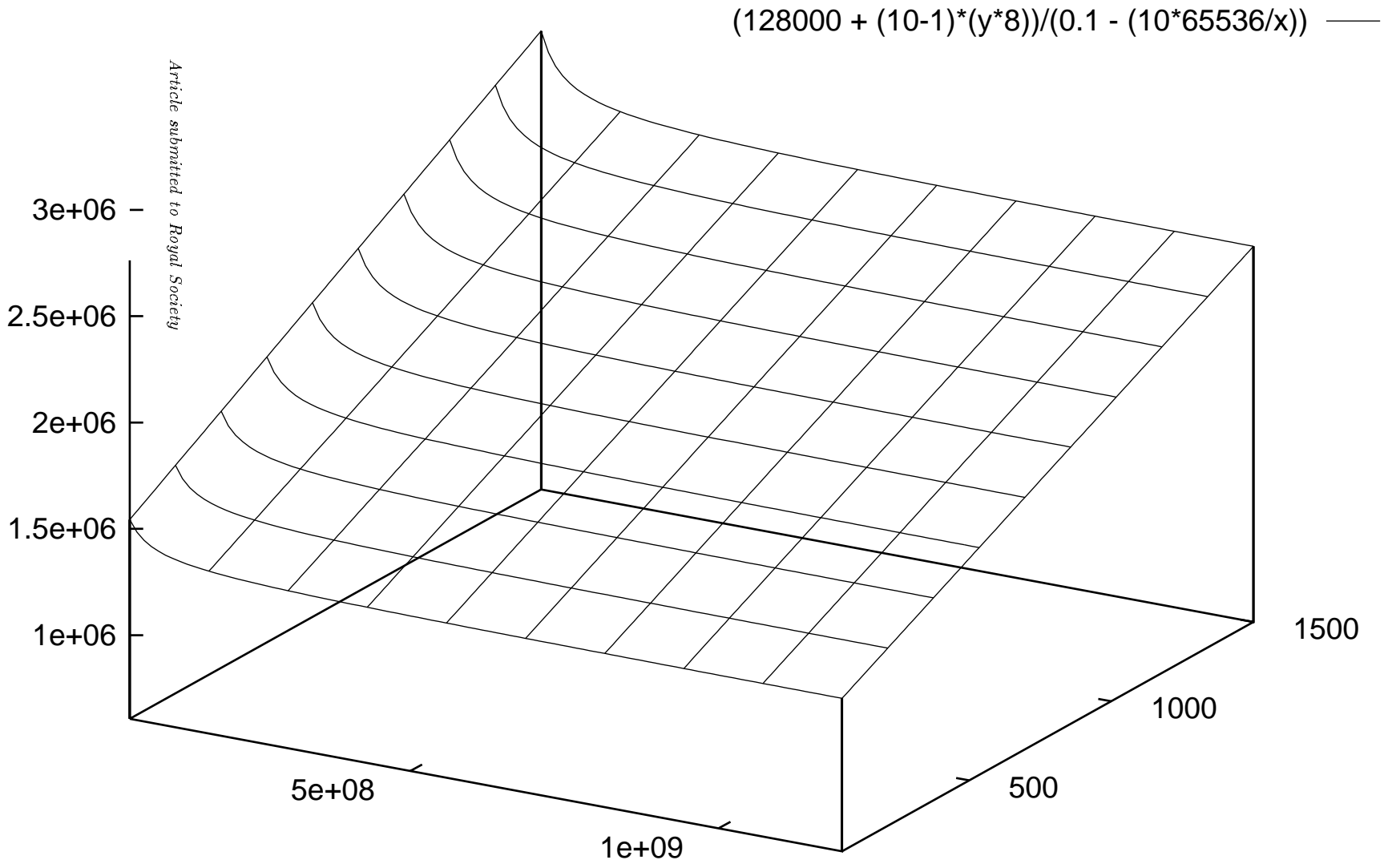
Figure 1. Guaranteed Bounded Delay and Utilisation

Table 2. *Terms in the Padhye TCP Equation*

W_m ; Maximum advertised receive window
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

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

$$B = min(\frac{W_m}{RTT}, \frac{1}{RTT\sqrt{\frac{2bp}{3}} + T_0 min(1, 3\sqrt{\frac{3bp}{8}})p(1 + 32p^2)}) \qquad (3.2)$$

This equation is for long lived transfers. There is more recent work 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 e.g. 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 utilisation 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.

Reliable multicast protocols (for games, share dealing and software and news distribution) such as the family of protocols using Handley and Floyd's "TCP Friendly Multicast Congestion Control" use similar adaption techniques to that which TCP employs. However, one additional level of complexity is that these protocols also have self-organising 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 1998 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.

Table 3. *Terms in the Erlang Equation*

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

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.

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

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 3 + 3 + 3 model is not appropriate to a rapid evolution of service sites in a network. Some researchers have proposed adding call admission for TCP, or RM or RTP flows.

Another problem with this naive approach is that the signalling and blocking themselves form a type of traffic - this traffic mist be congestion controlled, or re-call 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.

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, Bouch (1998, 1999) and Gibbens et al. One exception is the INDEX project at Berkeley 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 - push 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 "turbulance".

7. Aggregation

It is often suggested that wile 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 tend to interfere with attempts to aggregate flows that are destined between "similar" locations.

8. Summary and Conclusions

We've discussed feedback in Internet Protocols. There are lots of mechanisms, which, in the end, provide push back is all the way to the user (whether browser, Web Server, or streaming server). The challenge for the modeller is really in tackling large scale system design that needs to have the human as part of the model. This seems to be rather complex, and beyond current theory. In the meantime, the Internet Engineering Task Force continues to build build more and more adaptive systems. Other authors such as Paxson & Floyd, Nichols, and Crowcroft 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 thermo-dynamical systems. Finally, one very speculative economic hypothesis we might form (and is common dogma in the Telecommunications Industry) is that in am evolutionary system which exhibits rapid growth, it may be better to maximise revenue from new services than to optimise for stable services.

References

Paxson, V., & Floyd, S., "Why We Don't Know How To Simulate The Internet", Proceedings of the 1997 Winter Simulation Conference, December 1997. (see http://ftp.ee.lbl.gov/floyd/papers.html)

K Nichols, Cisco Systems, "Adding Feedback to Internet Simulations" unpublished draft/personal communication.

J. Padhye, V. Firoiu, D. Towsley, J. Kurose, "Modelling TCP Throughput: A Simple Model and its Empirical Validation", Proceedings of Sigcomm'98, Vancouver, CA, September 1998.

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

I.Kouvelas, V.Hardman, J.Crowcroft Self Organising Transcoders submitted to NOSSDAV (see `http://www.cs.ucl.ac.uk/staff/I.Kouvelas/publications/sot.ps.gz`).

Philippe Oechslin, J Crowcroft Weighted Proportional Fairness and Pricing for TCP submitted to ACM CCR (see `ftp://cs.ucl.ac.uk/darpa/multcp.ps`).

Mark Handley's Mbone Measurements and nam animation thereof, See `http://north.east.isi.edu/mbonemon/`

See `http://www-mash.cs.berkeley.edu/ns/` for the details about NS.

Abhey Parekh, "Generalized Processor Sharing", PhD Thesis, MIT Laboratory for Decision Systems, MIT 1993.

S. R. Caceres, N. Duffield, J. Horowitz, D. Towsley, T. Bu "Multicast-Based Inference of Network-Internal Characteristics: Accuracy of Packet Loss Estimation" . IEEE Infocom '99, (New York, USA, March 1999).

Jonathan Rosenberg and Henning Schulzrinne, "Timer reconsideration for enhanced RTP scalability," in Proceedings of the Conference on Computer Communications (IEEE Infocom), (San Francisco, California), March/April 1998.

S.Floyd, K.Fall, "Router Mechanisms to Support End-to-End Congestion Control", Technical report, `ftp://ftp.ee.lbl.gov/papers/collapse.ps`.

V.Jacobson, "Congestion Avoidance and Control", *ACM SIGCOMM'88*, August 1988, Stanford, CA, pp.314-329.

J.C. Lin, S.Paul, "RMTP: A Reliable Multicast Transport Protocol", IEEE INFOCOM '96, March 1996, pp.1414-1424.
Available as `ftp://gwen.cs.purdue.edu/pub/lin/rmtp.ps.Z`

Chris Greenhalgh, Steve Benford, Adrian Bullock, Nico Kuijpers and Kurt Donkers, "Predicting Network Traffic for Collaborative Virtual Environments", Computer Networks and ISDN Systems, 30 (19 1677-1685. First appeared in Proc. TNC'98, TERENA Networking Conference, Dresden, 5-8 October 1998.

Roberts, J. Call Admission Control for TCP ITC 99, Edinburgh, Scotland, June 1999.

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 International Symposium on Integrated Network Management (IM'99), Boston 24-28 May 1999

Cowley SJ "Of timing, turn-taking, and conversations", Journal of Psycholinguistic Research, 1998, Vol.27, No.5, pp.541-571

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.

HYbrid Monitoring and Network Simulations - HYMNS Jon Crowcroft University college London, HP Labs, Bristol, Isidor Kouvelas and Lorenzo Vicisano,UCL `http://www-iri.hpl.hp.com/`

Resource pricing and the evolution of congestion control R.J. Gibbens and F. P. Kelly. Automatica 35 (1999). `http://www.statslab.cam.ac.uk/ frank/evol.html`

National Laboratory for Applied Network Research at `http://www.nlanr.net`

VBNS Network Topology at `http://www.vbns.net/logical.html`

SuperJANET Pricing, see `bill.ja.net`