Frontmatter, Introduction & Literature Review

extracted from

Freedom with Accountability for Causing Congestion in a Connectionless Internetwork

Bob Briscoe

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London.

Department of Computer Science

University College London

15 May 2009

To Lyn

I, Robert John Briscoe confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____ 15 May 2009

Abstract

This dissertation concerns adding resource accountability to a simplex internetwork such as the Internet, with only necessary but sufficient constraint on freedom. That is, both freedom for applications to evolve new innovative behaviours while still responding responsibly to congestion; and freedom for network providers to structure their pricing in any way, including flat pricing.

The big idea on which the research is built is a novel feedback arrangement termed 're-feedback'. A general form is defined, as well as a specific proposal (re-ECN) to alter the Internet protocol so that self-contained datagrams carry a metric of expected downstream congestion.

Congestion is chosen because of its central economic role as the marginal cost of network usage. The aim is to ensure Internet resource allocation can be controlled either by local policies or by market selection (or indeed local lack of any control).

The current Internet architecture is designed to only reveal path congestion to end-points, not networks. The collective actions of self-interested consumers and providers should drive Internet resource allocations towards maximisation of total social welfare. But without visibility of a cost-metric, network operators are violating the architecture to improve their customer's experience. The resulting fight against the architecture is destroying the Internet's simplicity and ability to evolve.

Although accountability with freedom is the goal, the focus is the congestion metric, and whether an incentive system is possible that assures its integrity as it is passed between parties around the system, despite proposed attacks motivated by self-interest and malice.

This dissertation defines the protocol and canonical examples of accountability mechanisms. Designs are all derived from carefully motivated principles. The resulting system is evaluated by analysis and simulation against the constraints and principles originally set. The mechanisms are proven to be agnostic to specific transport behaviours, but they could not be made flow-ID-oblivious.

Acknowledgements

I am forever indebted to my wife, Lyn, for putting up with me obsessing over this research. And to my youngest son Joe who missed some of the attention his brothers got.

Jon Crowcroft provided excellent supervision and spot-on guidance throughout. Despite switching from UCL to Cambridge he remained dedicated to helping me through, even though he didn't have to, and even though I've taken longer than any of his many PhD students. Thank you, Jon. Also thanks to Stephen Hailes for his wise advice once he took over as my first supervisor, and to Saleem Bhatti and Mark Handley, my second supervisors. And thank you to my examiners, David Clark and Frank Kelly, for their thorough review comments.

I acknowledge the funding and support of BT through Alan Steventon's Steve Wright's and Jonathan Legh-Smith's Strategic Research Programme, and the support of my managers over the duration of this research, Steve Sim and Peter Hovell. I also acknowledge the personal support of BT's CTO Matt Bross.

Specific acknowledgements have been made in each supporting publication, others by reference. Where appropriate, specific contributions have also been identified in the text.

Sébastien Cazalet and Andrea Soppera contributed to the original invention of re-feedback, and the following people have help develop it over the years: Andrea Soppera, Arnaud Jacquet, Toby Moncaster, Carla Di Cairano-Gilfedder, Alessandro Salvatori, Alan Smith, Louise Burness and Martin Koyabe. The simulations in this dissertation were produced by Carla Di Cairano Gilfedder, Alessandro Salvatori and Toby Moncaster.

All the following have given helpful comments: David Songhurst, Ben Strulo, Phil Eardley, Peter Hovell, Gabriele Corliano, Steve Rudkin, Marc Wennink, Nigel Walker, Fabrice Saffre, Cefn Hoile, Steve Wright, Don Clarke, Keith Briggs, John Davey, Nigel Geffen, Pete Willis, John Adams (BT), Sally Floyd, Scott Shenker (ICIR), Joe Babiarz, Kwok Ho-Chan (Nortel), Stephen Hailes, Mark Handley, Adam Greenhalgh, Brad Karp (UCL), David Clark, Bill Lehr, Steve Bauer, Payman Faratin, Sharon Gillett, Liz Maida (MIT) Frank Kelly (Uni Cam) and comments from participants in the CRN/CFP Broadband and DoS-resistant Internet working groups.

These acknowledgements probably miss someone who has helped. I apologise. And finally, I alone am responsible for all the mistakes, overstated claims and any comments about the work of others that may offend.

Contents

List of Figures

List of Tables

Part I

Freedom with Accountability for Causing Congestion in a Connectionless Internetwork

Chapter 1

Introduction

1.1 The Problem

This research concerns the introduction of a resource consumption metric for the datagram internetworking layer, intended to improve the current Internet architecture. The end-to-end design principle [SRC84] advises that removing unnecessarily specific functions is as important as deciding which generic functions to include—necessary but sufficient. Chosen correctly, the internetwork layer should allow communications systems to be built around it that can evolve to meet unforeseen requirements without undue complexity.

This thinking has resulted in the characteristic rudimentary network layer of the Internet that solely delivers datagrams to their destination address. It allows every end-point the freedom to communicate in any way it wants with any other end-point, using any amount of the resource pool in between. But giving all end-points such freedom allows them to conflict with the freedom of others, wherever the capacity of particular resources is insufficient for the total load focused on it.

The problem we address is to include sufficient mechanism in the network layer to transmit a trustworthy resource consumption metric, but no more than the minimum necessary to allow higher layer mechanisms to resolve resource conflicts with a wide range of resource sharing approaches.

A large part of the contribution of this research has been to identify the precise sub-problems that need solving towards this end. Therefore the following problem description becomes a sequence of continually refined sub-problem statements. Perhaps more importantly, in the process it also identifies (non-)problems that were only on the generally accepted research agenda due to unsound reasoning—they were actually huge distractions.

The current Internet architecture allows every data source the freedom to choose whatever sending rate it requires, irrespective of the congestion it may cause. Most application authors choose to use the Internet through the TCP library, which behaves very sociably by reducing its sending rate in response to congestion [Jac88]. However, applications can choose not to reduce their rate in response to congestion, some because they cannot function below a minimum rate (e.g. interactive streaming media) and others through deliberate malice (e.g. flooding attacks). If these applications compete with TCP sources their careless, selfish or malicious behaviour is rewarded further by the TCP sources, which try even harder to alleviate congestion as long as competing sources continue their aggression.

However, even if every application used TCP, or was at least TCP-friendly [FHPW00] (i.e. using roughly the same average bit-rate as a TCP source under similar conditions), although congestion collapse would be avoided, there would be no control over whether resource sharing conflicts were reconciled. TCP certainly provides a safe dynamic (second order) response to congestion, but it is a fallacy that the shares of resources (first order) that TCP allocates are in some way special. That cannot be true because it depends on how much data different users ask TCP to transfer, and how many instances of TCP they use to do it [Bri07b]. A transport protocol alone cannot and should not be expected to share resources fairly, in any sense of the word [BMB08].

Further, it would be a mistake to solve the problem of resource conflicts by forcing every individual application to respond to congestion in a certain way. Curtailing the freedom that an application has to choose whatever sending rate it needs would limit the space for future innovation, stunting the growth of new (and existing) applications such as networked games, flurries of transactional messages or just faster than normal file downloads. Ideally we need to allow freedom within some wider bounds that encourage a generally sociable long-term and short-term sharing of resources, but with allowance for considerable give and take [GK99b].

A more fruitful approach is to view the problem as a need for accountability. We want every application to have the freedom to choose whatever rate is necessary with whatever dynamics. But, where this can restrict the freedom of others, networks need to at least be able to hold users accountable for the consequences of their actions. But, even if some networks don't hold users accountable precisely for the congestion they cause, but may-be for some poor approximation like volume, and even if some networks don't use accountability at all, then the whole system must still work for those who do care about accountability.

Accountability for resource usage was on the original 1988 agenda of requirements for the Internet protocols [Cla88], albeit last of seven in priority order. It was still an unmet requirement in the list for a new Internet architecture (NewArch) in 2000 [BCSW00], though framed as a need for a capacity allocation capability.

Only having solved the problem, do we now truly understand that the ability for networks to associate traffic with the sending user's account is neither a necessary nor sufficient form of accountability for internetworking. Firstly, the problem is one of accountability *for causing congestion*, as traffic itself is not a problem to anyone unless it contributes to congestion.

This is because the minimal accountability necessary for datagram forwarding should concern *cost* of usage. Certainly the Internet architecture should not help reveal other economic information such as consumer value. Consumers try to keep their valuation private and providers try to capture it, so it would be wrong to pre-judge the outcome of this tussle at such a low layer in the architecture. However, if the architecture doesn't reveal true usage costs, no mechanism can ensure that the cost to the consumer tends downwards over time towards the cost of provision. Cost minimisation is a generally accepted goal of all modern societies whatever mechanism is chosen, whether by encouraging competitive markets or by regulating uncompetitive markets, or even by centralised national economic planning.

The marginal cost of usage of a network resource depends entirely on the extra congestion due to the presence of the traffic. So the architecture should reveal congestion.

From the early days of the Internet, end-points were responsible for detecting and controlling congestion. Therefore, end-points place all the information they need for detecting congestion (sequence numbers) in the end-to-end transport layer. But there is no way for all the end-points to co-ordinate themselves sufficiently to hold each end-point accountable for the costs it causes to others, let alone for them to enforce any desired preventative action. Only the operator of a forwarding device can be in the natural position to do either.[1] But networks cannot see this congestion information unless end-points allow them to. It is hard for networks to measure this information reliably, because a gap in a sequence might simply imply a few packets went over a different path. And anyway, if networks did use this information against the interests of end-points, end-points could just encrypt it, or just not send it at all.

As well as which costs to consider, we have to consider who needs to associate the costs with whom. The minimum sufficient accountability requires a forwarding device to be able to associate the expected marginal costs of traffic with the entity directly causing the costs. Although the costs are originally caused by the data sender, each forwarding device directly assists in causing the costs. We now realise that it removes considerable complexity if the congestion is associated with each packet, rather than with the original sender. Then, minimally, accountability can be localised to any trust boundary across which packet traffic flows.

The advantage of making the packet, not just the sender, accountable is that we can then make each party along the forwarding path accountable for forwarding the packet across each trust boundary, localising accountability and enabling aggregation. As the packet crosses each trust boundary, the party on the receiving side can associate the costs in the packet with the party on the sending side of the boundary. Thus, a network forwarding packets on behalf of the sender can be held accountable for *allowing* the sender to cause congestion.

Localisation of accountability avoids any need for globally meaningful identities. Specifically, the validity of the sender's address in the network layer packet header becomes irrelevant for resource accountability. For wireline links this means accountability need only depend ultimately on the security of local physical connectivity.[2] For wireless links and for many virtualised wireline links, accountability will usually also have to depend on identifiers or authentication in link headers, but these need only be trusted local to the link.

The main omission prior to our research was that datagrams could only be held to account for the congestion they caused after the fact—once actual congestion had happened—because datagram transfer is inherently one-way or simplex. Instead, we ensure the sending or forwarding party can be held accountable for its *expectation* of how much congestion it will cause on the rest of the path. Then causes of excessive expected congestion can be curtailed. The problem then becomes one of ensuring that expected congestion is declared honestly, which is the subject of this dissertation.

[1] The operator might also be a consumer, as in an ad hoc or peer-to-peer network.

[2] The term wireline scales to 'wires' at a microscopic level, including data flows crossing process ownership boundaries within virtualised machines (e.g. multi-sender hosts or virtual routers).

We now understand that the problem is how to hold each self-contained datagram accountable for the congestion it expects to cause. Which leads us to have to solve the problem of how to update a packet as it traverses a network, so it always declares the congestion it is likely to cause, but only the likely congestion over the *remainder* of its journey. Laskowski & Chuang have also identified exactly this need to be able to monitor 'rest-of-path' congestion, delay etc. as the major cause of the Internet's economic problems [LC06][3].

By requiring the upstream entity to form an expectation of downstream congestion, it becomes in their interest to monitor recent downstream congestion by soliciting timely feedback. However, it would have been wrong to make feedback a necessary condition for using the Internet—a datagram must be sufficiently self-contained to be delivered alone. So we must not require end-points to depend on feedback about congestion from previous datagrams (although they can use it if they have it). Instead, an upstream entity can simply be conservative in its expectation of congestion if it chooses not to gather feedback (also essential for starting or re-starting a data flow before feedback is available).

Finally, congestion is of course caused by either excess traffic or insufficient capacity. We have so far focused on accountability for sending traffic, not for insufficiently supply capacity—dropping traffic. We believe it would be misguided to try to build a mechanism for networks to be accountable to data senders for specific instances of congestion [AMI+07, LC06]. It is sufficient for network N_B to be accountable to its upstream neighbour N_A both for any congestion within its own network and congestion in downstream networks it chooses to route through. This accountability takes a simple form. If N_B provides N_A with more expensive, more congested paths than other networks, N_A can choose not to use N_B's service. N_A can just not route via N_B, on a path-by-path basis if necessary.[4] So again, the problem is to ensure packets carry downstream congestion information. Then not only can N_B hold N_A accountable for forwarding traffic that causes congestion, but N_A can hold N_B accountable for not having provisioned sufficiently. Again, the problem is that packet networks lack visibility of downstream congestion information.

1.2 Motivation

The problem of improving the sufficiency of datagrams without sacrificing simplicity is an important scientific and engineering endeavour in its own right. But considerable social and economic problems are also at stake.

Firstly, if used as intended, the current Internet architecture allows resource allocations to become extremely sub-optimal relative to the social welfare maximisation that a perfect market would produce. Proving this is not part of the current research. But the intuition has been given above, and the author's complementary work (with co-authors) gives worked examples of how bad resource allocations can be for typical uses of the current Internet [BMB08]. We also try to quantify the problem a little below. We

[3] In a paper published in SIGCOMM'06, articulating the problem we had provided a solution to in the same conference the year before [BJCG+05].

[4] There is an important exception where the terminating network has a monopoly on routes to the destination, which is also part of the problem we address.

can certainly say that current resource allocations are not just slightly out, but hopelessly unlike they would be if a market were allocating resources.

Unconstrained resource sharing can be beneficial in small doses, but if allowed to predominate it can stagnate market growth. If applications that want higher bit rate can help themselves without being held to account during congestion, they can effectively free-ride at the expense of other people's service impairment. Communications infrastructure, particularly the access edges of a network, requires huge levels of investment many years in advance. If free-riding predominates, the risk of investment in new infrastructure becomes too high, because there is no expectation that those most benefiting from the investment can be made to pay the returns on that investment (and usually no-one else will unless Government backing is provided). A downward spiral of declining quality and declining investment results [Gro05].

But there is considerable evidence that investment in networks is not declining. Rather than allow their network to descend into this is spiral, unsurprisingly, ISPs have found other ways to prevent the worst effects of free-riding. With no formal architectural support against free-riding, they have resorted to a hotch-potch of locally invented attempts at improvement.

This is what is happening on the Internet. It is now very common for ISPs to deploy deep packet inspection (DPI) boxes to effectively fight TCP's resource allocations. ISPs identify the application within the payload of each packet flow and throttle those that they *infer* have low value. This violates the Internet architecture. But they are trying to improve their competitive position by pleasing more of their customers more of the time, without spending excessively on capacity. They have to violate the architecture for their businesses to remain viable.

Their 'need' to violate the architecture causes unintended consequences. Those application developers most likely to be hit by throttling are obfuscating their application traffic. Many ISPs are already starting to suspect any encrypted and unidentifiable payload. Anecdotally, there is already some evidence that some applications under threat are starting to imitate the characteristics of other 'business-critical' encrypted traffic. This could lead the ISPs to throttle all unidentifiable traffic or to consider making customers seek permission to send it (possibly for a fee).

Many people don't like companies taking control of their choices, and even those who don't care get angry when ISPs infer their values wrongly. Some ISPs have a vested interest in disadvantaging certain applications or competitor services. So even if an ISP's intentions are honourable (throttling heavy users in the interests of the majority), discriminating against certain packets can be confusable with anti-competitive practice. In the US over the last three years, this has led to politicians getting involved in the details of Internet resource allocation, at which point the possibilities for further unintended consequences expand, and the chance of rational scientific debate worsens.

But what justification is there for saying Internet resource allocation has become extremely sub-optimal? If one considers that a weight could be associated with every data flow (as in weighted proportional fairness [Kel97b]), then the predominance of TCP can loosely be considered as a special case with all the weights set to one. If instead everyone was free to choose their weight, constrained only by

accountability for the congestion they caused, Kelly shows everyone would maximise aggregate social welfare by setting the weights based on their willingness to pay for the bit-rate of each application.

Market studies[5] have shown that value per bit covers a spectrum of about ten orders of magnitude, from messaging (SMS, IM) at the extreme high end to software & video downloads at the bottom, with interactive voice, Web, email, interactive video and music downloads between. Assuming the value of transferring a bit is related to the value of the bit itself, this shows that optimal weights would probably range over many orders of magnitude, so setting all the weights to one is likely to be extremely sub-optimal. Even worse, bulk transfers (a large proportion of traffic on the Internet) would probably be given a very low weight, if users were accountable. But they are currently often given a weight considerably greater than one (by the programmer opening multiple instances of TCP).

Unfortunately, bulk transfers least need a high weight. Even without considering economics, weighting small jobs is the classic way to optimise completion times in scheduling problems with a mix of job sizes [Kle76]. But if the utility of completion times is also considered, when small jobs tend to carry higher utility per unit size, weighted solutions are even more powerful.

If there were accountability for congestion, higher weight would generally be assigned to brief intermittent flows (i.e. flows of fewer bytes interspersed by periods of inactivity) because the extra cost would be easier to sustain over lower activity factors than in larger flows with higher activity factor. And if small data flows go faster they finish sooner, leaving as much capacity on average for the bigger flows over time (modulo inefficiencies due to the greater dynamic range).

1.2.1 Other Motivations

Simpler Quality of Service. Quality of Service (QoS) mechanisms, whether per-session (e.g. Intserv [BCS94]) or bulk (e.g. Diffserv [BBC+98]), have foundered once inter-domain deployment has been attempted (for years they just foundered, full-stop). There seem to be two main problems. One is at the API, the other is the need for considerable operational baggage between networks; to scalably authorise and authenticate, to provision, to monitor contracts and so forth.

The API to QoS seems to be problematic because applications can only ask for something the network knows how to offer, which often isn't really what they want (which in turn would be too complicated to express or even know clearly at design time). The industry has trained its customers to say they want bit-rate, burst size and so forth. But applications (and humans) aren't like that. Once application demands are aggregated, it starts to become easier to express what is wanted, especially in terms of expectations rather than quantitative assurances [Cla95]. But this still leaves an API gap between the application and the aggregated part of the network.

It should be fruitful to look at these QoS problems in a different light. As long as an application is given early warning of impending congestion somewhere on its path, e.g. with explicit congestion notification (ECN [Flo94]), it can take QoS for itself by just not responding as much to approaching congestion as it would otherwise.

Seen like this, the QoS problem becomes one of accountability for causing congestion anywhere

[5] Unfortunately not citable.

on the path. The problem is then not what the application can do—it can always do almost anything. The problem is what the network provider will allow the application to do and how to stop it exceeding these bounds. This becomes a lot easier if each network on the path can see the same information about congestion as the customer's machines. The network directly attached to the consumer can then set limits to the behaviour of the customer as a whole site or household and it can enforce them (or charge for exceeding them). And networks further downstream can do the same recursively against their upstream neighbour networks.

Thus, instead of arranging packets to carry QoS requests to distant networks, the problem can be seen as getting packets to carry congestion information from distant networks to the local one. Importantly, this removes any need to place significance on identifiers in packets.

This approach alone would not be expected to give QoS with strong assurances[6], but it might allow a wide range of expectations to be met without applications having to translate what they think they want into a language that doesn't have the right vocabulary.

The interface between end-point and network or between two networks would be so simple, one could hardly call it a QoS API any more. Only incipient congestion (ECN) information would need to pass across it. But the congestion information would have the additional semantic of cost—for the application to trade off against the benefit it will get from continuing to send bits.

Mitigating Bandwidth Flooding. Mitigating distributed denial of service (DDoS) attacks is another motivation for this research. The security community generally hasn't considered bandwidth flooding as a congestion accountability problem.

But, instead of the victim trying to find where attack packets are coming from, the problem can be seen as ensuring packets reveal expected congestion as they leave the sender. Then the packets headed for a flooding attack should be very obvious to the networks on the way. A high rate stream of packets heading for close to 100% congestion should stand out from everything else, as it would be very unlikely to be a genuine application. And source and networks alike could be held accountable for the congestion cost of the attack, creating strong incentives to remove it [Bri06].

1.3 Road map

The dissertation is in four parts. This first part has explained why freedom with accountability for causing congestion is important for the Internet. It now continues in more depth by surveying the seminal literature in this field followed by the main criticisms of the state of the art that motivated the present research to fill the gaps. With the background to the field explained, we then end this first part by stating the two hypotheses that focus the rest of the dissertation.

It will be more meaningful to give an outline of the approach used in the rest of the dissertation at the end of Part I, in §3.3 after the hypotheses have been introduced.

[6] It can in an edge-to-edge rather than end-to-end architecture [Ear09b], but the API gap opens up again in this case.

Chapter 2

Related Work

In retrospect, reading and thinking deeply about just the following ten or so papers would have been sufficient background for this research. Of course, other sources (extensively referenced throughout this dissertation and in supporting publications) provided necessary background understanding and ideas, as well as many false trails.

2.1 Internet Congestion Control

TCP: In 1988, Jacobson published "*Congestion Avoidance and Control*" [Jac88] to document the collection of algorithms he had implemented to provide congestion control for the transmission control protocol (TCP). Bravely, this was a wholly distributed protocol in which all aspects of resource control—efficiency, stability and fairness—were governed by the collective action of the computers comprising the Internet. Without it, or something like it, it is unlikely the Internet would have ever become widely used. TCP congestion control was produced in response to repeated congestive collapses of the whole Internet in 1986 and 1987. Router-based alternatives were being actively pursued, but Jacobson's distributed solution was such an astonishing improvement on the previous TCP that it was immediately deployed on all 30,000 or so Internet hosts, and has remained the Internet's predominant resource control mechanism to this day.

A colleague[1] recently collected results from 16 traffic characterisation studies conducted at different parts of the Internet (campus, residential and WLAN) between Jan 2003 and May 2006 in an unpublished survey. The proportion of TCP bytes measured in each study clusters around two percentages, 80% and 92%, with a clear mode of 94% Internet bytes controlled by TCP. Two outlier studies found 72% and 98% respectively. There is no significant trend up or down over the years.

Most academic focus has been on the additive increase multiplicative decrease algorithm that TCP's congestion avoidance phase borrowed from Jain *et al* [JRC87]. But probably Jacobson's most important contribution was the balance between the parameters of the initial 'slow-start' phase and the following congestion avoidance phase, which he justified with self-confessed 'hand-waving' in the paper. Internet traffic has a heavy-tailed flow-size distribution, so large numbers of flows either never reach congestion avoidance, or at least send the majority of their bytes in slow start phase. Slow start phase is a tricky

[1] Swadesh Samanta.

period for a flow as it quickly tries to find a fair operating point alongside other traffic. But the majority of bytes (not flows) in all the other traffic are in congestion avoidance phase. So the long flows must react fast enough to losses to allow in brief flows, then they must quickly converge on the new operating point together, then, as the brief flow finishes, the long flow must be able to quickly use up the freed capacity.

ECN: In 1994, Floyd published "*TCP and Explicit Congestion Notification*" (ECN) [Flo94]. It proposed a new field in the Internet protocol (IP) header, which finally reached the first 'Proposed Standard' stage of the Internet Engineering Task Force's (IETF's) standards track nearly seven years later, in 2001 [RFB01].

Prior to ECN, a queue experiencing congestion would discard some packets, then Internet congestion controls like TCP would detect the lost packets as gaps in the sequence numbers of the packet stream. The idea of ECN was to use an explicit marking on packets to indicate the onset of congestion, to try to keep the network at an operating point just below where losses started to be experienced. There is always a possibility that a gap in a packet sequence is merely a symptom of re-ordering, so a transport protocol waits for stronger evidence of a loss (further packet arrivals without filling the gap, or ultimately a timeout) before deciding congestion has really been experienced and slowing its rate. This delay due to uncertainty (which ECN solves) has a disproportionately detrimental effect on the performance of short transfers.

The reason ECN is important to the present research is an unintended but necessary side-effect of its introduction. It makes congestion visible to network devices downstream of the congested link, whereas any discards of packets by upstream devices are difficult if not impossible to monitor within the network. This is because there is no need for a sequence number space at the IP layer. So if the transport or higher layers choose not to reveal their sequence numbers (e.g. by encrypting them), the network cannot detect a gap in them. And even if they are not encrypted, a network element doesn't know whether gaps are due to re-routes or congestion. Readability of the ECN field at the IP layer is a fortunate side-effect of the need for writability of the field at the network layer.

In outline, ECN works as follows. As an ECN-enabled queue in the network starts to grow, it sets the new ECN field to a codepoint termed congestion experienced (CE), with increasing probability the longer the queue. Whenever a CE mark arrives at the receiver it notifies the sender, which can quickly and unambiguously know that congestion has been experienced. The sender is then meant to reduce its rate as if it had detected a drop (e.g. in its congestion avoidance phase TCP would halve its window).

Despite the mention of TCP in the title both of the research paper and the proposed standard, ECN was a change to the network layer's notification of congestion, which then requires any higher layer transport protocol, not just TCP, to be updated in order to understand it. TCP was merely the first transport protocol to be adapted to the new IP. This required some careful attention to backward compatibility to avoid using ECN to signal congestion to legacy transports that only understood loss as a sign of congestion.

Specifically, prior to ECN, the two bits of the ECN field had (nearly) always been left containing `00`

(now termed Not-ECT, a non-ECN-capable transport). So, for packets with the ECN field cleared to zero, even if a queue is ECN-enabled it must use drop rather than ECN to notify congestion. Also a sender-receiver pair must not use ECN unless they have established that they are both capable of understanding it, typically in the capability negotiation during the initial handshaking to start a flow. Then the sender must set the ECN field in every data packet to a non-zero value[2] to indicate to the network that the transport understands ECN (termed ECN-capable transport or ECT).

2.2 Economics of Network Congestion

Two-part congestion pricing: MacKie-Mason & Varian's "*Pricing Congestible Network Resources*" [MMV95] summarises their research in this field. It examines the tension between recovering the cost of capacity through a flat charge or through a variable usage dependent charge. It considers a range of providers available to a user, all buying capacity K [b/s] at cost $c(K)$ [¤/s].[3] It hypothesises a two-part tariff offered to customers i with a fixed subscription price q [¤/s] and a variable usage price p [¤/b] for usage x_i [b/s], such that the rate of charge [¤/s] is $q + px_i$. It considers what choice of p & q would maximise social welfare under a centrally planned economy or under competitive or monopolistic markets. The monopoly case will be set aside in this summary.

Providers vary their prices to maximise profits. Users switch between providers until the price-quality balance suits them. Quality degrades when usage starts to exceed capacity. As a result, both the non-monopoly cases arrive at the same analytical result

$$\frac{\text{usage revenue}}{\text{capacity cost}} = \frac{px_i}{c(K)} = \frac{1}{e},$$

where e is the elasticity of scale of the capacity. Elasticity of scale is solely a property of the shape of the function giving the cost of capacity at the current capacity operating point,

$$e = \frac{\text{average cost}}{\text{marginal cost}} = \frac{c(K)}{K}\frac{1}{c'(K)}.$$

For the present research, the actual result isn't so important as the order of magnitude it implies. Typical elasticities of scale for transmission equipment are of the order of 2 and they approach linear (i.e. 1) as capacity approaches technology limits (unfortunately figures are all from privately published studies of equipment costs, e.g. Lechner [Lec99] and those of Reid). So the usage element of revenue should be about 50% of total revenue—and probably increasing, given no significant cost-saving disruptions in mass transmission technology are on the horizon. This implies that, for the foreseeable future, there will be a significant element of usage pricing in competitive Internet markets, because it holds strong competitive advantage against flat pricing. Note that usage pricing schemes that roughly approximate congestion pricing could be sufficient, such as volume caps at tiered but otherwise flat prices, or with volume limiting at peak periods.[4]

[2] It should use `01` or `10`, but it can also use `11` even though it shouldn't.

[3] The paper concerned usage of general congestible resources. Applying it to specific scenarios like networking was not always natural. So, for our application of the theory to networking, units have been included in brackets to add dimensional precision. ¤ is the symbol for non-specific currency.

[4] To give a current data point, BT's 'up to 8Mb' DSL broadband pricing at Feb 2008 consists of a fixed charge of £4/month and

This formulation also clearly shows that, in a competitive market, congestion pricing will not add to an average customer's charge, rather it will substitute some part of the fixed element with a variable element.

MacKie-Mason & Varian's work also contributed the idea of shadow prices for congestion—borrowed from the classic economics literature and applied to computing and networking problems. The congestion that others experience is a negative side-effect of an individual's usage of a network (a negative externality). Shadow pricing makes an individual internalise this congestion externality. So shadow pricing is a powerful technique for dividing up the Internet's resource allocation problem across all users.

Utility functions: In the same year, Shenker published "*Fundamental Design Issues for the Future Internet*" [She95], which posited that people's utility U for bit-rate x always satiates at high bit-rates, ($\frac{\partial^2 U}{\partial x^2} \leq 0; x \to \infty$) and that utility curves fall into two main classes: elastic and inelastic, being concave ($\frac{\partial^2 U}{\partial x^2} \leq 0$) and convex-concave (sigmoid) respectively. As load increases through a capacity bottleneck, this implies there is no limit to how small a share each user of an elastic application will find useful. But for inelastic applications, there will come a point where higher value for all will result if some users have zero capacity as their share drops below the knee of the sigmoid.

If Shenker's hypothesis is correct, it implies that variable-rate congestion control suits elastic applications, but admission control is preferable for inelastic applications. Shenker also pointed out that typical bit-rate reservation systems of the time were designed as if people's utility was a step function of bit rate, which could be considered as an approximation of a sigmoid. Whereas rate-adaptive codecs would give a better approximation to a more gradually inclined utility curve.

These classes of utility curve had no experimental basis. But, since, we have validated that video utility curves are indeed sigmoid with a wide shallow sloped 'step'. We used carefully designed experiments with users paying real money, but unfortunately the results are only accessible to partners in the M3I project, given they reveal price sensitivity information [HE02]. It is possible that all utility is strictly sigmoid because, to our knowledge, the existence of elastic utility right down to zero bit rate remains unproven. Nonetheless, as long as a network rarely gets so congested that bit rates fall below the knee of typical users' utility curves, it is not cost-effective to introduce the admission control mechanisms for all traffic that some still argue for [MR99, MPCC00]—it is easier to treat the traffic as effectively elastic, which certainly leads to sub-optimal total utility during rare overload episodes, but the total loss of utility over time is probably smaller than the extra it would cost to deploy and operate an admission control mechanism.[5]

---

a variable charge of either £5, £10 or £15/month. Reverse engineering this, the lower two tariffs equate to about £1/GB of volume irrespective of congestion, while the upper, so-called 'unlimited' tariff limits heavy volume users during peak period congestion. Given the fixed element has to cover non-capacity costs as well, these figures imply BT is trying to cover about 80% of its capacity costs from usage revenues. If BT's pricing is rational and if Mackie-Mason & Varian's analysis is broadly correct, this imputes an elasticity of scale figure of perhaps 1.2 for BT's network. Or equivalently BT's network cost, $c(K) \propto K^{0.8}$.

[5] As long as congestion controls handle extreme congestion safely (e.g. TCP's exponential back-off).

Kelly: In 1998, Kelly and others published "*Rate control for communication networks: shadow prices, proportional fairness and stability*" [KMT98], which made advances on many fronts and brought all the previously mentioned research together[6]:

- It applied MacKie-Mason & Varian's shadow pricing to a network, rather than just a single resource. It proved that, where elastic applications compete for bandwidth, the total welfare of everyone using the network can be maximised if the network charges each pair of end-points a shadow price p dependent on the sum of congestion they cause on the path between them.

- It added models of each queue's pricing algorithm and each end-point's rate control algorithm, albeit abstracted and fluid.

- It proved that, given shadow pricing, if application users were modelled with a private willingness to pay per unit time w_i, and private elastic utility modelled by $U_i = w_i \ln x_i$, purely out of self-interest they would have the incentive to weight the rate of their private congestion control algorithm in proportion to their willingness to pay. Specifically, user i would have an incentive to control her rate x_i to converge on $\frac{w_i}{p}$. Kelly proposed users could do this with an equation-based additive increase multiplicative decrease algorithm of the form

$$\Delta x_i = \kappa(w_i - p x_i)\Delta t,$$

  where κ is a gain constant. Later Siris designed and implemented a window-based variant [SCM02].

- It emphasised how the proposed way to distribute the solution preserved the Internet's ability to allow new applications with new congestion control requirements to evolve. Gibbens & Kelly also restated this body of research in a more accessible paper that focused more on the evolvability aspect, "*Resource Pricing and the Evolution of Congestion Control*" [GK99b].

- It proved that such self-interested behaviour would preserve local and global network stability, as long as the gain parameter of everyone's rate control algorithms met certain constraints. Stability was proved assuming instantaneous feedback, but a number of papers later proved stability with propagation delays, each assuming various algorithms and constraints. They are reviewed in [Kel03]. In broad terms they showed the gain constant must be below a certain constraint, which must itself be inversely proportional to round trip time.

- It gave a simple mechanism to implement the proposed scheme, based on explicit congestion notification (ECN) that was in the process of standardisation into IP at the time (now proposed standard status [RFB01]). All an ISP had to do was count the number of bytes in packets arriving marked as having experienced congestion at the receiver and apply a fixed price per marked byte.

[6] Kelly and Voice extended the work to cover end-point congestion-based routing in 2005 [KV05], but the original work made the advances that are most relevant to our points.

Kelly's assumptions seem reasonable, although one continues to cause debate—not over its correctness, but over how soon it will come into play. Kelly uses the scaling arguments outlined in [Kel00, §2] to show that, whichever way that Internet scale increases in the future—whether more flows, longer flows or higher bit-rate flows—as long as scale does indeed continue to increase, congestion delays will become insignificant relative to propagation delays.

Kelly *et al*'s work raised a number of important questions about TCP's congestion control algorithms [Jac88], which dominate congestion control and resource sharing throughout the Internet.

- Firstly it introduced the possibility that the rate towards which a congestion control algorithm converges need not be limited by round trip time (RTT), as long as the algorithm's first order dynamics are limited within a constraint that is inversely proportional to RTT. For instance, with stationary congestion $\bar{p}$, the above rate control algorithm converges on $\bar{x} = \frac{w_i}{\bar{p}}$, which can be independent of the gain, κ and therefore independent of RTT[7]. More recently, FAST TCP [JWL04] has adopted a similar strategy.

- But, much more significantly, the *likely* values that self-interested users would set Kelly's weights to, given congestion pricing, would lead to extremely different capacity shares to those produced by TCP (see §1.2).

However, in the wider Internet community, the message that TCP probably leads to an extremely sub-optimal outcome got lost among the objections to Kelly's proposed means for evolving to the optimal outcome: dynamic congestion pricing.

Simple pricing: Odlyzko's paper "*A modest proposal for preventing Internet congestion*" is more well-known for its main subject, the Internet pricing proposal called Paris Metro Pricing[8] But it also contains a wealth of evidence from numerous other consumer sectors that consumers are highly averse to unpredictable pricing [Odl97, §5].[9] The section is entitled 'The irresistible force runs into the immovable object,' because it seems to be an unescapable fact that the irresistible economic logic of usage-sensitive pricing runs counter to the greater desire of consumers for pricing that is predictable and mentally undemanding. Consumers will pay a premium to not have to continuously work out how to pay less. As a result, as Odlyzko puts it, "…free enterprise companies prefer the socialist method of rationing by queue to that of rationing by price."

Congestion pricing preserves the complete freedom of application logic (under the control of the user) to change its mind at any instant—to increase or decrease spending without seeking the permission of the network. But, consumers must *also* be able to opt not to have complete freedom. Because along with total freedom comes risk—the risk that events outside the consumer's control (the discovery of some desirable information coinciding with high congestion) will tempt them into spending more than they would have wished, in hindsight.

[7] Otherwise an application's attempts to maximise utility can become confused if it doesn't compensate for RTT.

[8] Incidentally, PMP fails in a competitive market [GMS00].

[9] Earlier Barns [Bar89] had provided evidence for a desire for predictable network pricing from the defence sector.

The aim of the M3I project[10] was to produce an architecture that would enable Internet resource sharing to self-manage through a variety of pricing plans that would be able to evolve to take account of the tensions between these immovable consumer pricing preferences, their quality preferences and the irresistible logic of congestion pricing. My own summary of the projects results and their architectural implications, "*Market Managed Multi-service Internet (M3I): Architecture PtI; Principles*" [Bri02a] agreed with Odlyzko's two consumer preferences for pricing (predictable and undemanding) and added a third, transparency, in which the consumer wants to know that they are getting a known quantity of a well-understood good for a known price.

This M3I report includes a summary of how the different pricing scenarios enabled by the M3I architecture resolved all the conflicts between demand control, quality control and pricing preferences to varying extents. By the end of the M3I project, the tensions had been resolved for inelastic traffic at the expense of a little extra complexity at the network edge—a risk broker function between the user's access network and the core[11]. But the tensions remained not fully resolved for elastic traffic.

One could argue (as I did [BDT+00]) that a consumer can buy into congestion pricing but then synthesise her own flat rate pricing by mediating the risk of overspending with her own software agent that keeps congestion charges within a moving window. But, psychologically, this is still not the same as someone else sorting it all out for you. Getting Internet service at minimal cost just isn't important enough to most people who just want to pay a flat-fee and it works. Consequently, ISPs don't want to offer a pricing plan with a footnote saying "As you probably won't like this pricing plan, we also provide free software to make it acceptable."

Further reading: Costas Courcoubetis and Richard Weber, "Pricing Communication Networks" Wiley (2003) [CW03]

2.3 Internetwork Market Structure

Edge-pricing: In 1996, Shenker, Clark, Estrin and Herzog published "*Pricing in Computer Networks: Reshaping the research agenda*" [SCEH96]. It puts forward three main arguments, two of which are outlined here.[12]

Firstly it argues that there is a need to cover more than just marginal costs, so "It is important to allow prices to be based on some approximation of congestion costs, but it is important to not force them to be equal to these congestion costs." This was essentially a precursor to the principle that was better

[10] www.m3i-project.org

[11] The solution is currently being standardised in the IETF congestion and pre-congestion notification (PCN) working group [Ear09b].

[12] The second of the three arguments seems misguided in hindsight. It says that Internet service tries to be generic to all applications, so it is inherently impossible for the network to capture user utility for not having individual packets delayed or dropped, as required for congestion pricing schemes like MacKie-Mason & Varian's 'Smart Market' [MMV93] under discussion at the time. However, the point of the 'Smart Market' proposal is that utility can remain private but then the market mechanism effectively allows *users* (not the network) to sort all the demand into two sets, with utility either above or below the shadow price, in order to limit demand to supply. The argument was perhaps saying that users wouldn't be able to divide their utility down on a packet by packet basis anyway. But, this rather threw out the baby with the bath-water by eliminating the possibility that even a very rough approximation would be better than nothing.

articulated later in 'Tussle in Cyberspace': that researchers shouldn't try to dictate outcomes.

Lastly it argued that the form of pricing wrt. usage was only one aspect of pricing that needed research. Instead it argued that more attention should be given to how contractual relationships should be structured across an internetwork.

The main contribution was a description of a structure called edge-pricing. With edge-pricing, networks levy bulk fees on their neighbours (end-customers and other networks) that all taken together cover a network's costs and profits, but charges don't have to be levied on a flow-by-flow basis. The motivation of edge-pricing is to allow the forms of tariffs to be different on a pairwise basis between neighbours, encouraging evolution of tariffs structures, rather than having to embed a pricing scheme in the architecture.

Information asymmetry: In 2001 Constantiou and Courcoubetis published "*Information Asymmetry Models in the Internet Connectivity Market*" [CC01]. Although it is not a conclusive paper, in that it presents no solutions, it clarifies more precisely than other similar papers which information networks cannot see about the quality of other networks and why this is so corrosive to a successful communications value chain. More recently, Laskowski & Cheung [LC06] also highlighted the same information as the critical missing piece of the Internet, but they did not relate the problem to the economic literature on market failures due to information asymmetry.

When it comes to theoretical understanding of quality issues, basic economic theory is only just in front of the 'science' of computer communications. The detrimental effects of asymmetry of quality information were only first articulated in Akerlof's 1970 paper "*The Market for 'Lemons': Quality, Uncertainty and Market Mechanisms*" [Ake70], which led to him (with others) winning the 2001 Nobel Prize in Economics. Using the example of used car sales, it showed that the salesman's privately held knowledge of which cars were duds ('lemons') drove down the price for used cars across the whole market, because the willingness to pay of consumers would reduce once they took the average risk of buying a dud into account, even if the car in question turned out to be fine. The suppressed market price led in turn to a reduction in the incentive to supply.

One can think of a data sender, or a forwarding network, as contracting with a downstream[13] network to deliver packets. But with one-way datagram technology, the upstream network knows little about the downstream neighbours it contracts with, whereas they know their own traffic loading and distribution, available capacity, resource allocation policies, customer types and interconnection agreements. Similarly, the next network is in a similarly weak position relative to the one after.

Constantiou and Courcoubetis apply the Principal-Agent formulation to model the resulting situation. The Principal-Agent formulation has been developed in economics to model the position of the principal (upstream) and agent (downstream) parties to this contract. By attaching parameterised rewards to any measurable effort of the agent and any measurable outcome for the principal, it is possible to optimise the parameters to design a contract that minimises the negative effects of information asymmetry.

[13] Different fields use the term 'downstream' ambiguously. Communications engineering uses it to mean 'in the direction of data transmission'. In the field of industrial organisation, downstream can also have the sense of 'towards the retail end of a value chain', but that is not the intent here.

Alternatively, it is possible to predict the value of improved measurability of effort or results. As already stated, the paper is inconclusive, but it at least identifies the problem well.

Design for Tussle: Clark and others published "Tussle in Cyberspace: Defining Tomorrow's Internet" in 2002, followed by a clearer journal article in 2005 [CWSB05]. It argues that the architecture of the Internet should allow the major tensions in society and in economics to be resolved at run-time, not design time. It turns this principle into the slogan 'Design for Tussle'.

The M3I architecture mentioned earlier also espoused this principle (it was published in parallel), but Clark *et al* give a far better and more general articulation. The M3I discussion was more specific (but consequently somewhat more concrete), being based on specific examples where the Internet architecture should be changed.

The paper offers further specific design principles, one being particularly relevant here: 'Modularise along tussle boundaries'. In the context of the above two papers on edge-pricing and information asymmetry, one could interpret this as advice to ensure the intended advantage of edge-pricing (independent evolvability of each pair-wise contract) is indeed possible. And to ensure that quality information is visible to both networks at every border.

Towards the end, the paper revisits some of the old design principles of the Internet in the light of the new tussle-related principles. It tries to grapple with the tensions in the end-to-end principle [SRC84] (see §1.1). Although the discussion seems inconclusive, it concludes that "...end-to-end arguments are still valid and powerful, but need a more complex articulation in today's world." We will return to this below as we highlight the main outstanding deficiencies in all the works we have just introduced.

2.4 Critique of Existing Work

TCP: The most pernicious deficiency in existing work has been the false goal of approximately equal flow rates through a bottleneck. The idea that rate equality is a good approximation to 'fair' set in long before Jacobson adopted it for TCP (traceable at least back as far as ATM research in 1980 [Jaf80]), to the extent that he didn't even question it as a reasonable goal. The problem statement of §1.1 has already rehearsed the core arguments that instantaneous flow rate is the wrong metric to be concerned with for fairness, because a) fairness should be between users not flows and b) instantaneous flow rate doesn't take account of the proportion of time that a user (or flow for that matter) is inactive.

My recent paper "*Flow Rate Fairness: Dismantling a Religion*" [Bri07b] published in ACM CCR, is an attempt to explain why Kelly's work shows that flow rate equality through a bottleneck is a nonsensical fairness goal. It is aimed at an audience that requires implications to be spelled out bluntly and one that has an aversion to maths. It carefully builds a case to show that the idea of flow rate fairness is completely unsubstantiated dogma. In contrast Kelly's welfare maximisation is given as an example of a properly defined form of fairness built on the philosophical notion of commutative justice[14].

[14] In 350 B.C.E. Aristotle distinguished two types of justice, distributive and rectifactory (commutative) [Ari25, Book V Chapters 2, 4 & 5]. Distributive justice concerns whether a particular distribution of goods is just but has proved impossible to define convincingly (Rawls [Raw01] comes closest, but still requires all one's preconceptions to be set aside in order to judge a just distribution). Commutative justice concerns whether an action (e.g. a transfer of goods) is just, most often determined by whether

However, the paper's main message is that different forms of fairness should be possible to enforce locally[15], but this will only ever be possible if the Internet architecture as a whole supports the ability to make self-interested individuals or entities (including whole networks) accountable for the costs they cause (or allow to be caused) to others. It therefore advocates the metric of congestion-volume as a prerequisite for different forms of fairness to co-exist. Congestion-volume is defined as a count of all the bytes of dropped or congestion marked data sent by all an individual's flows i over a period of time, T:

$$\int_T \sum_{\forall i} p(t)x(t)_i dt,$$

where $x(t)_i$ is the bit-rate of flow i and $p(t)$ is the congestion it experiences.

"...Dismantling a Religion" was motivated by the extreme unfairness (defined per user and over time) that has resulted on the present Internet in the name of flow-rate equality. But it was particularly motivated by the continued use of friendliness to TCP as a goal for new congestion controls (such as TFRC [FHPW00], XCP [KHR02] and other new high speed congestion controls), which constrains the future solution space completely unnecessarily. Even though it was claimed that an XCP switch could implement different forms of fairness, "...Dismantling a Religion" explained that fairness is a property of the congestion a user causes in a whole network over time, which is not something each switch can ever hope to control by setting the relative rates of just the flows that happen to be passing through it at any particular instant.[16]

More recently, I have published an Internet draft (with others) for the IETF, "*Problem Statement: Transport Protocols Don't Have To Do Fairness*" [BMB08] that justifies the assertion that there is extreme unfairness on the Internet, using numerical examples drawn from Internet measurements. It uses the evidence to argue that the IETF's protocol designs don't, can't and shouldn't have any control over fairness. But instead the IETF should concentrate on a protocol framework to allow fairness to be controlled at run-time (the message of 'Design for Tussle').

---

transfers are entered into voluntarily. It is alternatively termed rectifactory justice because a transfer of value (e.g. goods) in one direction that alters the balance of justice can be rectified by a transfer of value (e.g. money) in the other direction. Welfare maximisation is a result of a continuous sequence of transfers of value that are each commutatively just. If the original distribution of goods was not just (by whatever definition), a series of commutatively just transfers always improves everyone's lot in absolute terms, but it won't necessarily improve distributive justice (e.g. if defined relatively), even though progressive taxation is designed to attempt this. The only way to otherwise improve distributive justice is to somehow define a just distribution then forcibly take from the rich and give to the poor. However, further commutatively just transfers would again diverge from distributive justice, requiring continuous intervention.

[15] Both physically local and locally across a virtual grouping of users.

[16] XCP bears a superficial resemblance to re-feedback in that routers along the path decrement the change in flow rate requested in-band by the source, which is then fed back from receiver to source. However, XCP's structure is more analogous to a dynamic form of RSVP [ZDE+93]. The subtle but important difference from re-feedback is that XCP's metric quantifies the service rate (the primal variable), not the impairment introduced along the path (the dual). Even if the set of all the service rates is combined (e.g. at the customer's attachment point) nothing can be determined about whether that customer's use of the whole network is fair, because there is insufficient information about how much each flow impacts *other* users. In addition, in a non-co-operative setting, the service rate claimed in each XCP packet has to be policed at each border against the service actually provided, which requires per flow processing. This was the issue that killed the scalability of the Integrated Services architecture [BBB+97]. "...Dismantling a Religion" gives fuller discussion of these issues.

The draft accepts that some individuals aren't concerned if the Internet protocols aren't fair, so it aims to show that extreme unfairness leads to other highly detrimental concrete consequences. It uses further numerical examples to show how the inability to prevent free-riding in an architecture (extremely high allocations of congested resource for a minority of users who pay no more than others) leads to significantly higher investment risk. Because the majority will abandon a provider that continues to expect them to share the cost of its investments while receiving only a tiny share of the benefits. Using evidence that investment is still actually continuing, it explains this is because operators are throttling heavy users.

However, operators know that heavy *users* actually represent a mix of light and heavy *usage*. So rather than lose the heavy users' by limiting all their usage indiscriminately, operators are inspecting packet payloads and limiting only applications that they *infer* are causes of heavy congestion. Operators could limit overall traffic for heavy *users* and give them control over limiting their least valuable *usage*, but most users have neither the software nor the inclination to do this, so ISPs keep control themselves.

Users understandably get upset whenever their ISP's inferences are wrong. Also, however honourable the provider's intentions, their discriminatory throttling is easily confusable with anti-competitive discrimination against competitors' services, leading to the recent net neutrality debate.

The goal of flow-rate equality led to a large body of work on policing equal flow rates: Floyd and Fall's penalty box idea [FF99], Stabilized RED (SRED [OLW99]), CHOKe [PPP00], RED with Preference Dropping (RED-PD [MFW01]), Least Recently Used RED (LRU-RED [Red01]), XCHOKe [CCG+02], and Approx. Fair Dropping (AFD [PBPS03])). Because the goal of flow-rate equality is deficient, it has led these works to be triply deficient. Primarily because they are trying to police a flawed goal (per flow not per user, and instantaneous not over time). Secondly because it is easy for flows to circumvent any such policing using multiple flows on multiple paths. And thirdly because flows can simply whitewash their identifiers as soon as they are discovered, because there is no cost to creating new flow IDs.

ECN-based Congestion Pricing: Despite integrating together huge advances on many fronts, ultimately Kelly's work hit practicality problems for two entangled reasons: i) consumer aversion to dynamic congestion pricing and ii) dependence on the asymmetric structure of congestion information in the Internet. The entanglement was explained in "The case against classic feedback" in our main publication so far on re-feedback "*Policing Congestion Response in an Internetwork using Re-feedback*" [BJCG+05] (reproduced and updated slightly in §5.1.1 and outlined below.

Odlyzko's tension between the irresistible economic logic of usage-sensitive pricing and the immovable consumer desire for simpler pricing cannot be side-stepped. It must be possible for a network to ration demand by queue rather than by pricing—to slow down traffic causing congestion rather than delegate this responsibility to consumers under threat of higher charges. Of course any one user's ration will still be able to be sold at the correct congestion price. However, this will be simple, flat congestion pricing, not dynamic.

Only an ingress network, and preferably the first ingress device, can enforce congestion limiting.

But an ingress network cannot see the congestion being caused by the traffic entering the network, unless the congestion happens to be local. The classic feedback structure used by ECN, on which Kelly naturally built, cuts the upstream networks out of the loop. As explained above, ECN reveals information about congestion that was previously hidden, but not to networks upstream of the congestion. Certainly a feedback stream usually returns to the sender, but it is beyond the view of all the intervening networks—in higher layer end-to-end messages that may be encrypted, asymmetrically routed or simply omitted completely.

This is a highly unusual form of information asymmetry, where the buyer holds more information than the seller about the quality of the service. We believe it is this asymmetry that leads to all the Internet's problems of resource control economics. As we discussed in our summary of [BMB08], this asymmetry can lead to heavily suppressed investment. This is the same outcome as for Akerlof's case where the seller holds better information about quality than the buyer. But the chain of logic is the converse to Akerlof's. Nonetheless it has the same underlying structure, in which the market price has to include a premium that averages the risk of uncertainty over each contract.

This unusual information asymmetry is solely because the Internet is simplex at the internetwork layer (one-way information flow).[17] Duplex networks don't seem to exhibit the same economic problems as the Internet because any network can see the quality of the paths into which it is sending traffic by monitoring the feedback returning along each connection and managing traffic accordingly (e.g. ATM traffic management [ITU04]).

Because only the current Internet's classic feedback arrangement was available to Kelly and co-workers, congestion pricing could not be turned into rationing by queue. It might be feasible to throttle traffic at the last egress of the internetwork, based on information emerging from upstream congestion, but only by rationing the congestion *received* by a host. However, this would be a rather odd deal for a consumer to accept as a receiver cannot stop sent traffic from entering the network, filling it with traffic and consuming the receiver's congestion ration.

Given the Internet's feedback structure, the only option available to Kelly was to charge the receiver for congestion received. Then, in order to transfer the correct incentives to the sender, the receiver had to ask the sender to reimburse its congestion costs. This would unfortunately open all receivers to 'denial of funds' attacks, as well as incurring extra transaction costs.

There is a further subtle issue with Kelly's form of congestion pricing. Kelly holds each pair of end-points accountable for *actual* congestion. Whereas MacKie-Mason & Varian's smart market proposal holds the sender accountable for her *bid* if and only if it is greater than the actual congestion price. In both schemes, the charge ends up the same. But the subtle distinction only becomes apparent by thinking at the scale of individual packets. In both schemes the sender only discovers the price after the packet is sent. But in the smart market the sender limits her exposure to the risk of a high price, and if the actual price is higher the packet is discarded—again, rationing by queue rather than by price, but at the

[17] The term connectionless is deliberately avoided because it has a slightly different meaning. For instance multi-protocol label switching (MPLS) is simplex (reverse connections are not associated with forward connections) but not connectionless (connection state is held on network elements).

microscopic scale.

A different potential problem also lurks within Kelly's approach (it actually stems from the placement of utility with respect to instantaneous bit-rate in Shenker [She95]). As Clark had pointed out in 1995 [Cla95] and Shenker *et al* had repeated [SCEH96], the utility of transfers of fixed volume objects will often depend on completion time not instantaneous bit-rate. In 1999, Key & Massoulié pointed out that the two are inversely related because completing earlier stops the congestion costs earlier [KM99]. Therefore, once congestion is above a threshold there seems to be an incentive to drive up bit-rate to the maximum possible. In these cases, Key & Massoulié seem to convincingly argue that there will be no incentive to continuously optimise instantaneous bit-rate against instantaneous congestion. If they are correct, Kelly's results would lose much of their significance, as file transfers with utility from completion time probably comprise a large proportion of elastic Internet traffic. However, Gibbens & Kelly's experiments [GK99b, §3] propose a strategy for optimising instantaneous bit-rate by adapting willingness to pay that does pay off for users transferring fixed volume files.

Despite the importance of file completion time as a metric of value being 'reinvented' recently [DM06], the implications have still not been fully worked out. However Key and others have proved that, in the presence of delays, self-interested rate control will still lead to stability as long as file transfer traffic is mixed sufficiently with other types [KMBK04].

An interesting question is whether it is myopic to solely consider each object transfer in isolation, or whether transferring each object faster necessarily leads to opportunities to transfer more objects. This would imply that fixed volume objects are part of a larger stream with an overall volume that expands with bit-rate, at least when viewed at sufficiently coarse granularity.

Edge-pricing: I developed Shenker *et al*'s edge-pricing further in "*The Direction of Value Flow in Open Multi-Service Connectionless Networks*" [Bri00], a technical report that collects together two previously published papers applied to unicast and multicast [Bri99b, Bri99a]. It questions the proposed close tie between edge-pricing and apportionment of costs between sender and receiver.

The whole reason apportionment of costs between sender and receiver is needed is because different pairs of each have different apportionments of value. If the apportionments of usage cost between sender and receiver are fixed by the network, there will often be cases where the sum of the value they both derive is greater than the sum of their costs, but the value that one alone derives is less than its fixed share of the cost. If the losing party cannot shift some of its charge to the other, the communication won't happen. In the language of industrial organisation, communications is a two-sided market, because at least two buyers are involved in each sale [FW06] (see also §12.1.2).

The Shenker edge-pricing paper argues discursively that edge-prices should embed the chosen apportionment of costs between senders and receivers, whereas "*The Direction of Value Flow...*" argues that different flows will want different apportionments between sender and receiver to match the apportionment of value each derives from the communication. "*The Direction of Value Flow...*" uses a model of internetwork pricing to show that embedding the apportionment of costs between senders and receivers solely in the edge-pricing at the network layer necessarily leads to flow-by-flow charging and

an Internet-wide pricing scheme—exactly what Shenker *et al* were trying to avoid.

"*The Direction of Value Flow...*" outlined an end-to-end clearing function to re-apportion charges between the end-points, where the difference in value apportionment from the default made the transaction cost worth it. The Shenker paper had rejected such a clearing function in a footnote.

"*The Direction of Value Flow...*" further allowed each edge price to be split down into a fixed and a variable charge, and allowed the usage charge to flow in a direction independent of the direction that fixed charges took. This model was termed split-edge pricing.

My later work, co-authored with Rudkin, "Commercial Models for IP Quality of Service Interconnect" [BR05] revisited this whole field in the light of developments like re-feedback. It also added some specific structure to Shenker *et al*'s first point (that charging should merely be based on marginal cost, not equal to it). It reasoned why we can predict that commoditisation to marginal cost will proceed faster in transit (non-access) networks, while access networks will retain a greater ability to extract profits. The reasoning was that, although end-users and software developers might be expected to drive all networking to marginal cost, many end-users do not choose to spend their time minimising their charges (Odlyzko's point again). However, access networks have the motivation and means to aggregate their knowledge of their user's demands but to hide this knowledge from transits. From the viewpoint of transit networks, access networks resemble end-users—recursively. But unlike end-users, access networks have the power of aggregation and the means to use it.

Information asymmetry: Constantiou and Courcoubetis, like other papers on accountability [AMI+07, LC06], put the problem in terms of *network* accountability. But, of course, congestion is the result of too much traffic meeting too little capacity, so it is a question of both network *and* sender accountability.[18]

However, any one source of the traffic is not wholly to blame, because they didn't necessarily know all the others would send at the same time. And the network is not wholly to blame either because traffic can adapt much more quickly to insufficient capacity than capacity can adapt to traffic.

So accountability in both directions needs to be solved. Kelly's work shows how to divide the blame among the traffic—by sharing out instantaneous congestion in proportion to instantaneous bit rate. This can then be integrated over time and each user's contribution can be summed over all queues in the network as in the formula for congestion volume earlier:

$$\int_T \sum_{\forall i} p(t)x(t)_i dt.$$

And, the same information should be used by a network N_A to hold its downstream neighbour N_B accountable for congestion within N_B or in networks beyond, that N_B has chosen to route through towards the destination (its subcontractors). In the short term, this congestion is caused by the decisions of networks like N_A to route their traffic through N_B. But if the congestion persists longer term it implies N_B is not sufficiently provisioning capacity, or it is making poor onward routing decisions into other networks that are insufficiently provisioned.

Tussle: 'Tussle:.' [CWSB05] identifies some of the symptoms of the economic tension within the end-to-end design principle [SRC84] that is central to the Internet's design. But it largely side-steps any

[18] And, as pointed out in §1.1, the problem is simplified further if it is viewed as a *traffic* accountability problem.

challenge to the fundamental economic tension between the principles of 'Design for Tussle' and of 'End-to-end Design'. We believe this tension cannot be fudged to one side with the words "...end-to-end design is still valid but needs a more complex articulation."

The end-to-end principle essentially mandates that the lion's share of the profits from the communications value chain should go to the computing sector. Whereas the message of 'Tussle' is that the Internet architecture should not prejudge the outcome of the continuing competition between the computing and communications sectors. And if the architecture does pre-judge this tussle, the communications sector will choose to serve its own interests rather than comply with the architecture, thus leading to a mess of badly interconnected patches without an architecture—the present reality of the Internet.

If the communications sector were driven to near-zero commodity profits too early, investment capital would move to other less liquid sectors. A sector that is still growing rapidly is, by definition, not a commodity sector. Notwithstanding Odlyzko's points about consumer preferences, shadow pricing of congestion is the end-game that commoditisation will drive towards. But, as well as allowing congestion control to evolve under congestion pricing [GK99b], we have to allow pricing to evolve too. Even if pricing will eventually collapse towards congestion pricing (including congestion limiting), along the way we must allow the market to experiment with other more profitable schemes and services. Consumers, not system designers, are meant to commoditise a market—when they are ready. System designers should merely ensure the architecture would *allow* a shift to commoditisation.

Even if the economics predicts that an outcome (commoditisation) seems inevitable, the architecture shouldn't prejudge how quickly the whole value chain will reach this outcome and it should be able to encompass the structures that might develop on the possibly long road to that outcome.

For instance, many telcos (particularly in the cellular sector) are still wanting to build service-oriented networks, to sell services bundled with basic networking. It might well be that the open Internet model will just steam over them as they hanker after the golden past when they could bundle everything together and lock-in their customers. But it's just as likely their mass customer base might buy into services built on service-oriented networks, then the cellular operators will have resisted the open Internet model. Considerable value would be released [BR05, BOT06] if fixed and cellular networks could converge more closely. Therefore, the lesson from "Tussle" should be that the Internet architecture must encompass service-oriented networks as well as open networks. It's not clear the authors of "Tussle" meant to go that far. But that certainly must be a goal of the present research.

2.5 Conclusions from Reviews

The literature reviewed above builds a picture of the multifaceted problem the present research aims to tackle. It can be pictured as an ancestry diagram in two cascades.

1. The ideas and deficiencies in TCP, ECN, two-part congestion pricing and bit-rate utility were all brought together into one solution by Kelly, along with Kelly's own considerable advances in network traffic modelling.

2. Then the present research brings together the ideas in Kelly with those in simple pricing, edge-

pricing, information asymmetry and tussle to identify and fix a deficiency in the feedback architecture of simplex networks.

Others either discard the power of Kelly's model because it doesn't give simple pricing, or those who identify the information asymmetry problem try to retrofit internal feedback loops within the Internet. Whereas the task we set ourselves is to keep simplex networks fully simplex end-to-end, but convolve necessary and sufficient feedback information into the forward path.

Chapter 3

Hypotheses

Hypothesis 1 (Congestion Signal Integrity)**.** *The incentives of self-interested or malicious economic entities can be aligned to assure the integrity of indications of downstream congestion in the packets of a connectionless simplex internetwork. This can be achieved by only constraining aggregate downstream congestion-volume sent by each economic entity over time, without any dynamic congestion pricing to end-consumers, without any further constraints on transport behaviour and without any further constraints on the agents' freedom to distribute load across the internetwork, or across time.*

Hypothesis 2 (Welfare Maximising Allocation)**.** *With a competitive market and under Assumptions 3.1 & 3.2, incentives of all parties can be aligned so that the system produces the welfare maximising allocation of resources, under all the conditions of Hypothesis 1.*

Assumption 3.1. *Each consumer's demand is small relative to aggregate load on each link.*

Assumption 3.2. *Consumer utility is for bit-rate and the internetwork operates within the concave region of everyone's utility curves or flow admission control prevents anyone operating outside their concave utility range.*

3.1 Clarifications

Can be aligned: An example scalable enforcement mechanism can be defined with acceptably low probability of false hits or false misses.

Scalable: Sub-linear complexity wrt. traffic and network topology characteristics (no. of flows, no. of networks, etc.)

Acceptably low false hits: Losses of the same order as existing losses;

Acceptably low false misses: Where attacks might be successful due to statistical variations, over time the cost of launching failed attacks must be greater than the gains from successful attacks;

Downstream congestion-volume: As defined in §6.2

$$\int \sum_{\forall i} v(t)x(t)_i dt;$$

Constraining aggregate congestion-volume over time: A congestion-volume allowance fed at a constant fill-rate into a bulk token bucket per data-sender by their access network operator, which prevents further congestion being caused below a certain level and also constrains the maximum consumption of allowance;

Sent by each economic entity: The sending end of possibly multiple applications on possibly multiple computers under the control of a single economic entity. It is assumed that economic entities behind the data-receivers may choose to share part of the data-senders' costs, and some data-senders may make sending data conditional on the receiver's contractual commitment to share costs.

Assumption 3.3. *Transaction costs between sender and receiver can be ignored.*

This assumption is invalid, but analysis of how much this higher layer issue affects the welfare maximisation and the mechanisms for sharing costs are left for further research;

Economic entity: A stakeholder with its own motivations, resources and capabilities including end-consumers and network providers (alternatively, economic agent or party);

Welfare maximising: Maximisation of the sum of the utilities of all economic entities.

The system: The combination of economic entities, the internetwork, rate control functions on host computers, congestion notification protocols, and the incentive mechanisms at the network's trust boundaries defined in this dissertation;

Connectionless simplex internetwork: A collection of network domains operated by autonomous economic entities, using only one-way self-contained end-to-end datagrams with no return channels for congestion feedback at the network layer (e.g. not in routing or congestion back-pressure messages).

Self-interested economic entity: Individuals or organisations operating with rational self-interest;

Malicious: Unbounded malice if the entity is an end-consumer or bounded malice if the entity is a network;

Bounded malice: Only willing to exploit amplifying traffic-related vulnerabilities, where the cost to the victim is strictly greater than the cost of the attack;

Unbounded malice: Willing to exploit any traffic-related vulnerabilities to cause harm to others.

Traffic-related vulnerabilities: Vulnerabilities in the re-ECN system, or in related Internet traffic control functions. Information security issues within the payload or pre-existing network security issues (e.g. routing vulnerabilities) are ruled out of scope;

Dynamic congestion pricing: Pricing proportional to congestion caused per bit;

End-consumers: The economic agents behind data-senders and data-receivers;

Further constraints: Constraints other than the aggregate constraint;

Transport behaviour: Increases and decreases in data-rate;

Distribute load across the internetwork: Send to any destination at any data rate;

Distribute load across time: Send less data now and more later or vice versa.

3.2 Significance and Rationale

The congestion signal integrity hypothesis is ambitious. Paraphrasing Popper [Pop63], safe conjectures are not interesting. For me, it is not as important to be correct as it is to be practical, as long as I'm practically correct so there is a possibility of making an impact.

Proving robustness against gaming is an ambitious and ultimately impossible goal, because one cannot know the set of all attacks that might be invented against it. However, one can create an abstract model of the solution, its incentive environment and its information flows, towards proving it has high likelihood of being robust against the attacks we know. This is believed to be a sufficient approach in computer science when proposing systems solutions to large, distributed problems.[1]

The welfare maximising hypothesis has been separated out from the integrity hypothesis. It follows fairly trivially if the first hypothesis holds, by straightforward connection to the arguments in Kelly [KMT98, KV05]. It was felt important to include a case close to the way resources are generally allocated in the world so as to link to the Internet resource allocation motivation for the work. However, it would have been wrong to tie the integrity hypothesis only to this single (albeit important) case. Congestion signal integrity is an architectural building block that would be useful for other ways of allocating resources than a market (reflecting the arguments of 'Design for Tussle').

The practicality conditions are the more challenging and interesting aspects of the integrity hypothesis. These conditions have been carefully chosen because they encapsulate a wider set of practicality constraints.

Dynamic congestion charging not required: We wanted to find a solution that did not present retail network providers with the dilemma of either having to offer an unpopular tariff or not being able to rely on their customers' natural incentives in order to share network resources fairly. We also wanted to avoid the idealistic assumption that players only act rationally, which many proofs of incentive compatibility require. So we replaced congestion pricing with engineering mechanisms that would allow networks to police their customers' responses to congestion whether they were rational or not. Enabling engineered policers, rather than relying on rationality, also protects a player against accidental misconfiguration of its own part of the system.

We know from the start that it is fruitless to align the incentives of some zealot with unbounded

[1] I have also (perhaps deliberately) engineered incentives for others to try to break my solution (by strongly criticising whole fields of other people's work, persistently claiming near-perfection in my own and challenging others to break it!). This has already led to a number of proposed attacks on re-feedback, which have helped my generalisation of possible attacks, and in some cases design changes have been necessary.

malice. However, we believe we may be able to prove our hypothesis if we require only the malice of networks to be bounded (§8.1.2), while we will allow everyone else's malice to have no bounds. This is an ambitious (and therefore interesting) attack model.

No further constraints on transport behaviour: This point aims to ensure accountability is not introduced at the expense of freedom—so that new applications with novel responses to congestion can emerge. Choosing to enforce accountability through network engineering rather than pricing would seem to imply that the congestion behaviours used by today's set of applications will become embedded within every network. But, by constraining our solution to avoid service standardisation between applications and network operators and between operators, we intend to show that new applications would be able to emerge without asking permission. Even if networks do put in certain behaviour constraints, these can be relaxed by bilateral agreement between a customer and the ingress network, without further standardisation effort across other networks downstream, which would effectively block any evolution.

Whether a system allows players the freedom to evolve is notoriously subjective and therefore both easy to prove loosely and difficult to prove conclusively. Whether service standardisation is necessary is perhaps only one aspect of evolvability, but it is at least a provable fact.

Scalability: This constraint aims to ensure that accountability is not introduced at the expense of poor network layer scaling with number of flows, users etc. in the sense used in complexity theory.

3.3 Approach

The vast majority of the rest of the dissertation is aimed at proving Hypothesis 1 (congestion signal integrity). The welfare maximising hypothesis only requires brief treatment at the end.

The following chapters are not only structured around the goal of proving the hypothesis, but at the same time having to introduce the elements of the system in an order that will be readable and interesting, and bring out all the insights learned on the way.

The dissertation proceeds in three passes: i) high level ii) abstraction; iii) concrete, because the parts of the system are interdependent, so it would otherwise have been hard to go into detail on any one part without knowing where it fitted into the whole. Experiments appear next to the aspect of the system that they test, not collected separately nearer the end. Otherwise, they would have become so distanced from the assertion they were trying to prove that the connections would have become tenuous.

Part II (next) does the first two passes. It introduces the re-feedback protocol and its incentive mechanisms in a setting that sets aside the practicalities of deployment on the Internet. In particular, it assumes the protocol can write real numbers of arbitrary precision into packet headers. It then introduces some of the possible uses of re-feedback, taking a broad brush approach, but subjecting some aspects to experiment.

Part III contains the bulk of the recent work. Not only is it grounded in the practicalities of the Internet, but it takes a more principled approach to the design of the components introduced previously.

This allows implementations to be tested against the constraints and principles they were intended to realise.

Finally, part IV ties up the proofs of the hypotheses using the material in the intervening parts. It concludes by enumerating limitations and future research directions, before listing material contributions (papers etc.) and giving concluding remarks.

Appendices are added that describe alternative approaches either deprecated or rejected, to record why they fell short, so others need not tread the same erroneous paths.

Bibliography

[AHCC06] Lachlan L.H. Andrew, Stephen V. Hanly, Sammy Chan, and Tony Cui. Adaptive deterministic packet marking. *IEEE Comm. Letters*, 10(11):790–792, November 2006.

[AK94] P. Almquist and F. Kastenholz. Towards requirements for IP routers. Request for comments 1716, Internet Engineering Task Force, November 1994. (Obsoleted by RFC1812) (Status: informational).

[Ake70] G.A. Akerlof. The market for 'lemons': Quality, uncertainty and market mechanisms. *Quarterly Journal of Economics*, 84:488–500, August 1970.

[AKM04] Guido Appenzeller, Isaac Keslassy, and Nick McKeown. Sizing router buffers. *Proc. ACM SIGCOMM'04, Computer Communication Review*, 34(4), September 2004.

[AKS06] Edward Anderson, Frank Kelly, and Richard Steinberg. A contract and balancing mechanism for sharing capacity in a communication network. *Management Science*, 52:39–53, 2006.

[AMI+07] Katerina Argyraki, Petros Maniatis, Olga Irzak, Subramanian Ashish, and Scott Shenker. Loss and delay accountability for the internet. In *Proc. IEEE International Conference on Network Protocols*. IEEE, October 2007.

[Ari25] Aristotle. *Ethica Nicomachea*. Clarendon Press, Oxford, 1925. Translated by W.D.Ross.

[Arm06] Mark Armstrong. Competition in two-sided markets. *RAND Journal of Economics*, 37(3):668–691, Autumn 2006.

[Bar89] W. Barns. Defense data network usage accounting enhancement approaches. Technical report, The MITRE Corporation, 1989.

[Bau05] Steve Bauer. Incentive misalignment under congestion-based pricing. URL: http://cfp.mit.edu/CFP_WG_WS/BBWG_NOV_2005/Steven_Bauer_11-05.pdf, November 2005.

[BB01] Deepak Bansal and Hari Balakrishnan. Binomial congestion control algorithms. In *Proc. IEEE Conference on Computer Communications (Infocom'01)*, pages 631–640. IEEE, April 2001.

[BB05] Rob Beverly and Steve Bauer. The spoofer project: Inferring the extent of source address filtering on the Internet. In *Proc. Steps to Reducing Unwanted Traffic on the Internet Workshop (SRUTI 2005)*, pages 53–59. USENIX, July 2005.

[BBB+97] F. Baker, B. Braden, S. Bradner, A. Mankin, M. O'Dell, A. Romanow, A. Weinrib, and L. Zhang. Resource ReSerVation protocol (RSVP) — version 1 applicability statement; Some guidelines on deployment. Request for comments 2208, Internet Engineering Task Force, January 1997.

[BBC+98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An architecture for differentiated services. Request for comments 2475, Internet Engineering Task Force, December 1998.

[BC01a] Bob Briscoe and Jon Crowcroft. An open ECN service in the IP layer. Technical Report TR-DVA9-2001-001, BT, February 2001.

[BC01b] Bob Briscoe and Jon Crowcroft. An open ECN service in the IP layer. Internet draft, Internet Engineering Task Force, February 2001. (Expired).

[BCC+98] B. Braden, D. Clark, J. Crowcroft, B. Davie, S. Deering, D. Estrin, S. Floyd, V. Jacobson, G. Minshall, C. Partridge, L. Peterson, K. Ramakrishnan, S. Shenker, J. Wroclawski, and L. Zhang. Recommendations on queue management and congestion avoidance in the Internet. Request for comments 2309, Internet Engineering Task Force, April 1998.

[BCS94] R. Braden, D. Clark, and S. Shenker. Integrated services in the Internet architecture: an overview. Request for comments 1633, Internet Engineering Task Force, June 1994.

[BCSJ04] Bob Briscoe, Sébastien Cazalet, Andrea Soppera, and Arnaud Jacquet. Shared control of networks using re-feedback; an outline. Technical Report TR-CXR9-2004-001, BT, September 2004.

[BCSW00] Robert Braden, David Clark, Scott Shenker, and John Wroclawski. Developing a next-generation Internet architecture. White paper, DARPA, July 2000.

[BDH+03] Bob Briscoe, Vasilios Darlagiannis, Oliver Heckman, Huw Oliver, Vasilios Siris, David Songhurst, and Burkhard Stiller. A market managed multi-service internet (M3I). *Computer Communications*, 26(4):404–414, February 2003.

[BDT+00] Bob Briscoe, Konstantinos Damianakis, Jérôme Tassel, Panayotis Antoniadis, and George Stamoulis. M3I pricing mechanism design; Price reaction. Deliverable 3 Pt II, M3I Eu Vth Framework Project IST-1999-11429, July 2000.

[Bel00] Steve M. Bellovin. ICMP traceback messages. Internet Draft draft-bellovin-itrace-00.txt, Internet Engineering Task Force, March 2000. (Work in progress).

[BES+06] Bob Briscoe, Philip Eardley, David Songhurst, Francois Le Faucheur, Anna Charny, Jozef Babiarz, Kwok-Ho Chan, Stephen Dudley, Georgios Karagiannis, Attila Bader, and Lars Westberg. An edge-to-edge deployment model for pre-congestion notification: Admission control over a DiffServ region. Internet Draft draft-briscoe-tsvwg-cl-architecture-04.txt, Internet Engineering Task Force, October 2006. (Work in progress).

[BFB06] Steve Bauer, Peyman Faratin, and Robert Beverly. Assessing the assumptions underlying mechanism design for the Internet. In *Proc. Workshop on the Economics of Networked Systems (NetEcon06)*, June 2006.

[BJCG+05] Bob Briscoe, Arnaud Jacquet, Carla Di Cairano-Gilfedder, Alessandro Salvatori, Andrea Soppera, and Martin Koyabe. Policing congestion response in an internetwork using re-feedback. *Proc. ACM SIGCOMM'05, Computer Communication Review*, 35(4):277–288, August 2005.

[BJMS09a] Bob Briscoe, Arnaud Jacquet, Toby Moncaster, and Alan Smith. Re-ECN: Adding accountability for causing congestion to TCP/IP. Internet Draft draft-briscoe-tsvwg-re-ecn-tcp-08.txt, Internet Engineering Task Force, March 2009. (Work in progress).

[BJMS09b] Bob Briscoe, Arnaud Jacquet, Toby Moncaster, and Alan Smith. Re-ECN: The motivation for adding congestion accountability to TCP/IP. Internet Draft draft-briscoe-tsvwg-re-ecn-tcp-motivation-01.txt, Internet Engineering Task Force, March 2009. (Work in progress).

[Bla08] Steven Blake. Use of the IPv6 flow label as a transport-layer nonce to defend against off-path spoofing attacks. Internet Draft draft-blake-ipv6-flow-label-nonce-01, Internet Engineering Task Force, November 2008. (Work in Progress).

[BLMR98] John Byers, Michael Luby, Michael Mitzenmacher, and Ashutosh Rege. A digital fountain approach to reliable distribution of bulk data. *Proc. ACM SIGCOMM'98, Computer Communication Review*, 28(4), September 1998.

[BLSS05] Eli Brosh, Galit Lubetzky-Sharon, and Yuval Shavitt. Spatial-temporal analysis of passive TCP measurements. In *Proc. IEEE Conference on Computer Communications (Infocom'05)*, pages 949–959. IEEE, March 2005.

[BMB08] Bob Briscoe, Toby Moncaster, and Lou Burness. Problem statement: Transport protocols don't have to do fairness. Internet Draft draft-briscoe-tsvwg-relax-fairness-01, Internet Engineering Task Force, July 2008. (Work in progress).

[BOT06] Bob Briscoe, Andrew Odlyzko, and Benjamin Tilly. Metcalfe's Law is Wrong. *IEEE Spectrum*, Jul 2006:26–31, July 2006.

[BP87] R. T. Braden and J. Postel. Requirements for Internet gateways. Request for comments 1009, Internet Engineering Task Force, June 1987. (Obsoleted by RFC1812) (Status: historic).

[BR05] Bob Briscoe and Steve Rudkin. Commercial models for IP quality of service interconnect. *BT Technology Journal*, 23(2):171–195, April 2005.

[Bra97] Scott Bradner. Key words for use in RFCs to indicate requirement levels. BCP 14, Internet Engineering Task Force, March 1997. (RFC 2119).

[Bri99a] Bob Briscoe. The direction of value flow in connectionless networks. In *Proc. 1st International COST264 Workshop on Networked Group Communication (NGC'99)*, volume 1736. Springer LNCS, November 1999. (Invited paper).

[Bri99b] Bob Briscoe. The direction of value flow in multi-service connectionless networks. In *Proc. International Conference on Telecommunicatons and E-Commerce (ICTEC'99)*, October 1999.

[Bri00] Bob Briscoe. The direction of value flow in open multi-service connectionless networks. Technical Report TR-NZG12-2000-001, BT, August 2000.

[Bri02a] Bob Briscoe. M3I Architecture PtI: Principles. Deliverable 2 PtI, M3I Eu Vth Framework Project IST-1999-11429, February 2002.

[Bri02b] Bob Briscoe. M3I Architecture PtII: Construction. Deliverable 2 PtII, M3I Eu Vth Framework Project IST-1999-11429, February 2002.

[Bri06] Bob Briscoe. Using self-interest to prevent malice; Fixing the denial of service flaw of the Internet. In *Proc Workshop on the Economics of Securing the Information Infrastructure*, October 2006.

[Bri07a] Bob Briscoe. Fast congestion notification (FCN). Tech report TR-CXR9-2006-003, BT, May 2007. (unpublished work in progress).

[Bri07b] Bob Briscoe. Flow rate fairness: Dismantling a religion. *ACM SIGCOMM Computer Communication Review*, 37(2):63–74, April 2007.

[Bri07c] Bob Briscoe. Flow rate fairness: Dismantling a religion. Internet Draft draft-briscoe-tsvarea-fair-02, Internet Engineering Task Force, July 2007. (Expired).

[Bri08a] Bob Briscoe. Byte and packet congestion notification. Internet Draft draft-ietf-tsvwg-byte-pkt-congest-00.txt, Internet Engineering Task Force, August 2008. (Work in progress).

[Bri08b] Bob Briscoe. A fairer, faster internet protocol. *IEEE Spectrum*, Dec 2008:38–43, December 2008.

[Bri09a] Bob Briscoe. Emulating border flow policing using re-PCN on bulk data. Internet Draft draft-briscoe-re-pcn-border-cheat-03.txt, Internet Engineering Task Force, October 2009. (Work in progress).

[Bri09b] Bob Briscoe. Tunnelling of explicit congestion notification. Internet Draft draft-ietf-tsvwg-ecn-tunnel-06.txt, Internet Engineering Task Force, December 2009. (Work in progress).

[CC01] Ioanna D. Constantiou and Costas A. Courcoubetis. Information asymmetry models in the Internet connectivity market. In *Proc. 4th Internet Economics Workshop*, May 2001.

[CC08] Denis Collange and Jean-Laurent Costeux. Passive estimation of quality of experience. *Journal of Universal Computer Science*, 14(5):625–641, March 2008.

[CCG+02] Parminder Chhabra, Shobhit Chuig, Anurag Goel, Ajita John, Abhishek Kumar, Huzur Saran, and Rajeev Shorey. XCHOKe: Malicious source control for congestion avoidance at Internet gateways. In *Proc. IEEE International Conference on Network Protocols (ICNP'02)*. IEEE, November 2002.

[CF98] David D. Clark and Wenjia Fang. Explicit allocation of best-effort packet delivery service. *IEEE/ACM Transactions on Networking*, 6(4):362–373, August 1998.

[Cla88] David D. Clark. The design philosophy of the DARPA internet protocols. *Proc. ACM SIGCOMM'88, Computer Communication Review*, 18(4):106–114, August 1988.

[Cla95] David D. Clark. A model for cost allocation and pricing in the internet. *Journal of Electronic Publishing*, 1(1&2), January–February 1995.

[Cla96] David D. Clark. Combining sender and receiver payments in the Internet. In G. Rosston and D. Waterman, editors, *Interconnection and the Internet*. Lawrence Erlbaum Associates, Mahwah, NJ, October 1996.

[cla98] kc claffy. The nature of the beast: Recent traffic measurements from an Internet backbone. In *Proc. INET'98*. ISOC, 1998.

[CO98] Jon Crowcroft and Philippe Oechslin. Differentiated end to end Internet services using a weighted proportional fair sharing TCP. *Computer Communication Review*, 28(3):53–69, July 1998.

[CP98] L. Cherkasova and P. Phaal. Session-based admission control — a mechanism for improving performance of commercial web sites. In *Proc. International Workshop on QoS (IWQoS'99)*. IEEE/IFIP, June 1998.

[CW96] Costas Courcoubetis and Richard Weber. Buffer overflow asymptotics for a switch handling many traffic sources. *Journal Applied Probability*, 33:886–903, 1996.

[CW03] Costas Courcoubetis and Richard Weber. *Pricing Communication Networks*. Wiley, 2003.

[CWSB05] David Clark, John Wroclawski, Karen Sollins, and Robert Braden. Tussle in cyberspace: Defining tomorrow's Internet. *IEEE/ACM Transactions on Networking*, 13(3):462–475, June 2005.

[Day07] John Day. *Patterns in Network Architecture: A Return to Fundamentals*. Prentice-Hall, 2007.

[DBT08] Bruce Davie, Bob Briscoe, and June Tay. Explicit congestion marking in MPLS. Request for comments rfc5129.txt, Internet Engineering Task Force, January 2008.

[DKS89] A. Demers, S. Keshav, and S. Shenker. Analysis and simulation of a fair-queueing algorithms. *Computer Communication Review (SIGCOMM'89)*, 19(4):1–12, September 1989.

[DKZSM05] Nandita Dukkipati, Masayoshi Kobayashi, Rui Zhang-Shen, and Nick McKeown. Processor sharing flows in the internet. In *Proc. International Workshop on QoS (IWQoS'05)*, June 2005.

[DM06] Nandita Dukkipati and Nick McKeown. Why flow-completion time is the right metric for congestion control. *ACM SIGCOMM Computer Communication Review*, 36(1):59–62, January 2006.

[DR04] Ian Dobbs and Paul Richards. Innovation and the new regulatory framework for electronic communications in the EU. *European Competition Law Review*, 25(11):716–730, 2004.

[Ear09a] Metering and marking behaviour of PCN-nodes. Internet Draft draft-ietf-pcn-marking-behaviour-05.txt, Internet Engineering Task Force, May 2009. (Work in progress).

[Ear09b] Pre-congestion notification architecture. Internet Draft draft-ietf-pcn-architecture-11.txt, Internet Engineering Task Force, April 2009. (Work in progress).

[Edd07] Wes Eddy. TCP SYN flooding attacks and common mitigations. Request for comments RFC4987, Internet Engineering Task Force, August 2007.

[FAJS07] Sally Floyd, Mark Allman, Amit Jain, and Pasi Sarolahti. Quick-Start for TCP and IP. Request for comments rfc4782.txt, Internet Engineering Task Force, January 2007.

[FF99] Sally Floyd and Kevin Fall. Promoting the use of end-to-end congestion control in the Internet. *IEEE/ACM Transactions on Networking*, 7(4):458–472, August 1999.

[FHPW00] Sally Floyd, Mark Handley, Jitendra Padhye, and Jörg Widmer. Equation-based congestion control for unicast applications. *Proc. ACM SIGCOMM'00, Computer Communication Review*, 30(4):43–56, October 2000.

[FHPW03] Sally Floyd, Mark Handley, Jitendra Padhye, and Jörg Widmer. TCP friendly rate control (TFRC): Protocol specification. Request for comments rfc3448.txt, Internet Engineering Task Force, January 2003.

[FI08] Bryan Ford and Janardhan Iyengar. Breaking up the transport logjam. In *Proc. Workshop on Hot Topics in Networks (HOTNETS VII)*. ACM, October 2008.

[Fin89] G. Finn. A connectionless congestion control algorithms. *ACM SIGCOMM Computer Communication Review*, 19(5), October 1989.

[FJ93] Sally Floyd and Van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, August 1993.

[Flo94] Sally Floyd. TCP and explicit congestion notification. *ACM SIGCOMM Computer Communication Review*, 24(5):10–23, October 1994. (This issue of CCR incorrectly has '1995' on the cover).

[Flo08] Sally Floyd. RED (random early detection) queue management; setting parameters. Web page http://www.icir.org/floyd/red.html#parameters, November 2008. (Last accessed Jan 2009).

[FR98] E. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(2):173–199, 1998.

[FW06] Peyman Faratin and Tom Wilkening. Interconnection discrimination: A two-sided markets perspective. In *Proc. ACM Hot Topics in Networking (HotNets-V)*. ACM, November 2006.

[GK99a] Richard J. Gibbens and Frank P. Kelly. Distributed connection acceptance control for a connectionless network. In *Proc. International Teletraffic Congress (ITC16), Edinburgh*, pages 941–952, 1999.

[GK99b] Richard J. Gibbens and Frank P. Kelly. Resource pricing and the evolution of congestion control. *Automatica*, 35(12):1969–1985, December 1999.

[GK02] Richard J. Gibbens and Frank P. Kelly. On packet marking at priority queues. *IEEE Transactions on Automatic Control*, 47(6):1016–1020, June 2002.

[GKM01] Ayalvadi Ganesh, Peter Key, and Laurent Massoulié. Feedback and bandwidth sharing in networks. In *Proc. 39th Annual Allerton Conference on Communication, Control and Computing*, 2001.

[GM06] Yashar Ganjali and Nick McKeown. Update on buffer sizing in Internet routers. *ACM SIGCOMM Computer Communication Review*, 36, October 2006.

[GMS00] Richard Gibbens, Robin Mason, and Richard Steinberg. Internet service classes under competition. *IEEE Journal on Selected Areas in Communications*, 18(12):2490–2498, 2000.

[GQX+04] David K. Goldenberg, Lili Qiu, Haiyong Xie, Yang Richard Yang, and Yin Zhang. Optimizing cost and performance for multihoming. *Proc. ACM SIGCOMM'04, Computer Communication Review*, 34(4):79–92, October 2004.

[Gro05] Broadband Working Group. The broadband incentive problem. White paper, MIT Communications Futures Programme and Cambridge University Communications Research Network, September 2005.

[HBC05] Peter Hovell, Bob Briscoe, and Gabriele Corlianò. Guaranteed QoS synthesis (GQS): An example of a scalable core IP quality of service solution. *BT Technology Journal*, 23(2), April 2005.

[HE02] David Hands (Ed.). M3I user experiment results. Deliverable 15.2, M3I Eu Vth Framework Project IST-1999-11429, February 2002. (M3I partner access only).

[HG04] Mark Handley and Adam Greenhalgh. Steps towards a DoS-resistant Internet architecture. In *FDNA '04: Proceedings of the ACM SIGCOMM workshop on Future directions in network architecture*, pages 49–56, New York, NY, USA, 2004. ACM Press.

[HGB06] H. Hassan, J.M. Garcia, and C. Bockstal. Aggregate traffic models for voip applications. In *Proc. International Conference on Digital Telecommunications (ICDT'06)*, page 70, 2006.

[HH07] Felipe Huici and Mark Handley. An edge-to-edge filtering architecture against DoS. *Computer Communication Review*, 37(2):39–50, 2007.

[ITU04] Traffic control and congestion control in B-ISDN. Recommendation I.371 (03/04), ITU-T, March 2004.

[Jac88] Van Jacobson. Congestion avoidance and control. *Proc. ACM SIGCOMM'88 Symposium, Computer Communication Review*, 18(4):314–329, August 1988.

[Jaf80] J.M. Jaffe. A decentralized, "optimal", multiple-user, flow control algorithms. In *Proc. Fifth Int'l. Conf. On Computer Communications*, pages 839–844, October 1980.

[Jaf81] J. M. Jaffe. Bottleneck flow control. *IEEE Transactions on Communications*, 29(7):954–962, July 1981.

[JBM08] Arnaud Jacquet, Bob Briscoe, and Toby Moncaster. Policing freedom to use the internet resource pool. In *Proc Workshop on Re-Architecting the Internet (ReArch'08)*. ACM, December 2008.

[JBS05] Arnaud Jacquet, Bob Briscoe, and Alessandro Salvatori. A path-aware rate policer: Design and comparative evaluation. Technical Report TR-CXR9-2005-006, BT, October 2005.

[JRC87] R. Jain, K. Ramakrishnan, and D. Chiu. Congestion avoidance in computer networks with a connectionless network layer. Technical report DEC-TR-506, Digital Equipment Corporation, 1987.

[JWL04] Cheng Jin, David Wei, and Steven Low. FAST TCP: Motivation, architecture, algorithms, performance. In *Proc. IEEE Conference on Computer Communications (Infocomm'04)*. IEEE, March 2004.

[Kel97a] Frank P. Kelly. Charging and accounting for bursty connections. In Lee W. McKnight and Joseph P. Bailey, editors, *Internet Economics*, pages 253–278. MIT Press, 1997.

[Kel97b] Frank P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8:33–37, 1997. A version with a correction by Ramesh Johari and Frank Kelly to distinguish flows with zero weight and using a better structure of proof is available from URL: http://www.statslab.cam.ac.uk/~frank/elastic.html.

[Kel00] Frank P. Kelly. Models for a self-managed Internet. *Philosophical Transactions of the Royal Society*, 358(1773):2335–2348, August 2000.

[Kel03] Frank Kelly. Fairness and stability of end-to-end congestion control. *European Journal of Control*, 9:159–176, 2003.

[KHR02] Dina Katabi, Mark Handley, and Charlie Rohrs. Congestion control for high bandwidth-delay product networks. *Proc. ACM SIGCOMM'02, Computer Communication Review*, 32(4):89–102, October 2002.

[Kle76] Leonard Kleinrock. *Queuing Systems, Vol II: Computer Applications*. Wiley, New York, 1976.

[KM97] D. Kristol and L. Montulli. HTTP state management mechanism. Request for comments 2109, Internet Engineering Task Force, February 1997.

[KM99] Peter Key and Laurent Massoulié. User policies in a network implementing congestion pricing. In *Proc. Workshop on Internet Service Quality and Economics*. MIT, December 1999.

[KMBK04] Peter Key, Laurent Massoulie, Alan Bain, and Frank Kelly. Fair Internet traffic integration: Network flow models and analysis. *Annales des Télécommunications*, 59:1338–1352, 2004.

[KMT98] Frank P. Kelly, Aman K. Maulloo, and David K. H. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, 1998.

[KS01] S. Kunniyur and R. Srikant. Analysis and design of an adaptive virtual queue (AVQ) algorithm for active queue management. *Proc. ACM SIGCOMM'01, Computer Communication Review*, 31(4), October 2001.

[KV05] Frank Kelly and Thomas Voice. Stability of end-to-end algorithms for joint routing and rate control. *ACM SIGCOMM Computer Communication Review*, 35(2):5–12, April 2005.

[LC06] Paul Laskowski and John Chuang. Network monitors and contracting systems: Competition and innovation. *Proc. ACM SIGCOMM'06, Computer Communication Review*, 36(4):183–194, September 2006.

[Lec99] Robert Lechner. Competitive network evolution towards broadband IP. In Bruce Wiltshire, editor, *Proc. BT Alliance Engineering Symposium (AES'99)*, page Session 1A Paper 2. BT Alliance, June 1999. (not publicly accessible).

[LG09] Michael Vittrup Larsen and Fernando Gont. Port randomization. Internet Draft draft-ietf-tsvwg-port-randomization-03, Internet Engineering Task Force, March 2009. (Work in Progress).

[MBJ07] Toby Moncaster, Bob Briscoe, and Arnaud Jacquet. A TCP test to allow senders to identify receiver non-compliance. Internet Draft draft-moncaster-tsvwg-rcv-cheat-02.txt, Internet Engineering Task Force, November 2007. (Work in progress).

[MFW01] Ratul Mahajan, Sally Floyd, and David Wetheral. Controlling high-bandwidth flows at the congested router. In *Proc. IEEE International Conference on Network Protocols (ICNP'01)*, 2001.

[MHR+90] A. Mankin, G. Hollingsworth, G. Reichlen, K. Thompson, R. Wilder, and R. Zahavi. Evaluation of Internet performance — FY89. Technical report MTR-89W00216, MITRE Corporation, February 1990.

[MMV93] Jeffrey K. MacKie-Mason and Hal Varian. Some economics of the Internet. In *Proc. Tenth Michigan Public Utility Conference at Western Michigan University*, March 1993.

[MMV95] Jeffrey K. MacKie-Mason and Hal Varian. Pricing congestible network resources. *IEEE Journal on Selected Areas in Communications, "Advances in the Fundamentals of Networking"*, 13(7):1141–1149, 1995.

[MPCC00] Richard Mortier, I. Pratt, C. Clark, and Simon Crosby. Implicit admission control. *IEEE Journal on Selected Areas in Communications*, 18(12):2629–2639, December 2000.

[MR91] A. Mankin and K. Ramakrishnan. Gateway congestion control survey. Request for comments 1254, Internet Engineering Task Force, July 1991. (Status: informational).

[MR99] Laurent Massoulié and Jim W. Roberts. Arguments in favour of admission control for TCP flows. In *Proc. Int'l Teletraffic Congress (ITC16)*, June 1999.

[MSMO97] Matthew Mathis, Jeffrey Semke, Jamshid Mahdavi, and Teunis Ott. The macroscopic behavior of the TCP congestion avoidance algorithms. *SIGCOMM Comput. Commun. Rev.*, 27(3):67–82, 1997.

[MW00] Jeonghoon Mo and Jean Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, October 2000.

[Nag84] J. Nagle. Congestion control in IP/TCP internetworks. Request for comments 896, Internet Engineering Task Force, January 1984. (Status: unknown).

[Nag85] J. Nagle. On packet switches with infinite storage. Request for comments 970, Internet Engineering Task Force, December 1985. (Status: unknown).

[ns2] Network simulator. URL: http://www.isi.edu/nsnam/ns/.

[Odl97] Andrew Odlyzko. A modest proposal for preventing Internet congestion. Technical report TR 97.35.1, AT&T Research, Florham Park, New Jersey, September 1997.

[OLW99] Teunis J. Ott, T. V. Lakshman, and Larry H. Wong. SRED: Stabilized RED. In *Proc. IEEE Conference on Computer Communications (Infocom'99)*, pages 1346–1355. IEEE, March 1999.

[Orm05] Ralph Orme. British Telecommunications plc's statement about IPR claimed in draft-briscoe-tsvwg-re-ecn-tcp-00.txt. URL: https://datatracker.ietf.org/public/ipr_detail_show.cgi?&ipr_id=651, November 2005.

[Pax99] Vern Paxson. End-to-end Internet packet dynamics. *IEEE/ACM Transactions on Networking*, 7(3):227–292, June 1999.

[PBPS03] Rong Pan, Lee Breslau, Balaji Prabhaker, and Scott Shenker. Approximate fairness through differential dropping. *ACM SIGCOMM Computer Communication Review*, 33(2):23–40, April 2003.

[PFTK98] Jitendra Padhye, V. Firoiu, Don Towsley, and Jim Kurose. Modeling TCP throughput: A simple model and its empirical validation. *Proc. ACM SIGCOMM'98, Computer Communication Review*, 28(4), September 1998.

[PLD04] Antonis Papachristodoulou, Lun Li, and John C. Doyle. Methodological frameworks for large-scale network analysis and design. *ACM SIGCOMM Computer Communication Review*, 34(3):7–20, July 2004.

[Pop63] Karl Popper. *Conjectures and Refutations*. Routledge & Kegan Paul, Abingdon, Oxon, England, 1963.

[Pos81] Jon Postel. Internet control message protocol. Request for comments 0792, Internet Engineering Task Force, September 1981.

[PPP00] R. Pan, B. Prabhakar, and K. Psounis. CHOKe, A stateless active queue management scheme for approximating fair bandwidth allocation. In *Proc. IEEE Conference on Computer Communications (Infocom'00)*. IEEE, March 2000.

[PQW03] Venkata N. Padmanabhan, Lili Qiu, and Helen Wang. Server-based inference of Internet performance. In *Proc. IEEE Conference on Computer Communications (Infocomm'03*. IEEE, April 2003.

[Raw01] John Rawls. *Justice as Fairness: A Restatement*. Harvard University Press, Cambridge, MA, 2001.

[Red01] Smitha A. L. Narasimha Reddy. LRU-RED: An active queue management scheme to contain high bandwidth flows at congested routers. In *Proc Globecomm'01*, November 2001.

[RFB01] K. K. Ramakrishnan, Sally Floyd, and David Black. The addition of explicit congestion notification (ECN) to IP. Request for comments 3168, Internet Engineering Task Force, September 2001.

[RS06] Barath Raghavan and Alex C. Snoeren. Decongestion control. In *Proc. ACM Hot Topics in Networking (HotNets-V)*. ACM, November 2006.

[Sal05] Alessandro Salvatori. Closed loop traffic policing. Master's thesis, Politecnico Torino and Institut Eurécom, September 2005.

[SBS02a] Vasilios A. Siris, Bob Briscoe, and Dave Songhurst. Economic models for resource control in wireless networks. In *Proc. 13th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2002)*. IEEE, September 2002.

[SBS02b] Vasilios A. Siris, Bob Briscoe, and Dave Songhurst. Service differentiation in third generation mobile networks. In *International workshop on Quality of future Internet Services (QofIS'02)*, volume 2511, pages 169–178. COST263, Springer LNCS, October 2002.

[SC02] Computer Science and Telecommunications Board (CSTB). *Broadband; Bringing Home the Bits*. National Academy Press, Washington D.C., 2002.

[SCEH96] Scott Shenker, David Clark, Deborah Estrin, and Shai Herzog. Pricing in computer networks: Reshaping the research agenda. *ACM SIGCOMM Computer Communication Review*, 26(2), April 1996.

[SCM01] Vasilios A. Siris, Costas Courcoubetis, and George Margetis. Service differentiation in ECN networks using weighted window-based congestion control. In *Proc. 2nd International workshop on Quality of future Internet Services (QofIS'01)*. COST263, September 2001.

[SCM02] Vasilios A. Siris, Costas Courcoubetis, and George Margetis. Service differentiation and performance of weighted window-based congestion control and packet marking algorithms in ECN networks. *Computer Communications*, 26(4):314–326, 2002.

[SCWA99] Stefan Savage, Neal Cardwell, David Wetherall, and Tom Anderson. TCP congestion control with a misbehaving receiver. *ACM SIGCOMM Computer Communication Review*, 29(5):71–78, October 1999.

[SEB+06] David J. Songhurst, Phil L. Eardley, Bob Briscoe, Carla Di Cairano Gilfedder, and June Tay. Guaranteed QoS Synthesis for admission control with shared capacity. Technical Report TR-CXR9-2006-001, BT, February 2006.

[She95] Scott Shenker. Fundamental design issues for the future Internet. *IEEE Journal on Selected Areas in Communications*, 13(7):1176–1188, 1995.

[Sir02] Vasilios A. Siris. Resource control for elastic traffic in CDMA networks. In *Proc. ACM International Conference on Mobile Computing and Networks (MobiCom'02)*. ACM, September 2002.

[SPS+02] Alex C. Snoeren, Craig Partridge, Luis A. Sanchez, Christine E. Jones, Fabrice Tchakountio, Beverly Schwartz, Stephen T. Kent, and W. Timothy Strayer. Single-packet IP traceback. *IEEE/ACM Transactions on Networking*, 10(6):721–734, 2002.

[SRC84] Jerome H. Saltzer, David P. Reed, and David D. Clark. End-to-end arguments in system design. *ACM Transactions on Computer Systems*, 2(4):277–288, November 1984. An earlier version appeared in the Second International Conference on Distributed Computing Systems (April, 1981) pages 509–512.

[SWE03] Neil Spring, David Wetherall, and David Ely. Robust explicit congestion notification (ECN) signaling with nonces. Request for comments RFC3540, Internet Engineering Task Force, June 2003. (Status: Experimental).

[Sys02] Cisco Systems. Distributed weighted random early detection. Release Note Cisco IOS Release 11.1 CC and Feature Modules, Cisco Systems, 2002.

[TC04] R.W. Thommes and M.J. Coates. Deterministic packet marking for time-varying congestion price estimation. In *Proc. IEEE Conference on Computer Communications (Infocomm'04)*. IEEE, March 2004.

[Wei91] Mark Weiser. The computer for the 21st Century. *Scientific American*, 265(3):94–104, September 1991.

[WHBB08] Damon Wischik, Mark Handley, and Marcelo Bagnulo Braun. The resource pooling principle. *SIGCOMM Comput. Commun. Rev.*, 38(5):47–52, October 2008.

[Wis07] Damon Wischik. Short messages. In *Proc. Workshop on Networks: Modelling and Control*. Royal Society, September 2007.

[WK08] John Wittgreffe and Kashaf Khan. Orchestrating end-to-end network and resources according to application level service level agreements. *BT Technology Journal*, 26(1):46–57, September 2008.

[Wol82] Ronald W. Wolff. Poisson arrivals see time averages. *Operations Research*, 30(2):223–231, March–April 1982.

[WPSB09] Michael Welzl, Dimitri Papadimitriou, Michael Scharf, and Bob Briscoe. Open research issues in internet congestion control. Internet Draft draft-irtf-iccrg-welzl-congestion-control-open-research-05, Internet Research Task Force, May 2009. (Work in progress).

[XSSK05] Yong Xia, Lakshminarayanan Subramanian, Ion Stoica, and Shivkumar Kalyanaraman. One more bit is enough. *Proc. ACM SIGCOMM'05, Computer Communication Review*, 35(4):37–48, 2005.

[YMKT99] Maya Yajnik, Sue B. Moon, James F. Kurose, and Donald F. Towsley. Measurement and modeling of the temporal dependence in packet loss. In *Proc. IEEE Conference on Computer Communications (Infocom'99)*, pages 345–352. IEEE, 1999.

[YPG00] R. Yavatkar, D. Pendarakis, and R. Guerin. A framework for policy-based admission control. Request for comments 2753, Internet Engineering Task Force, January 2000.

[ZD01] Yin Zhang and Nick Duffield. On the constancy of Internet path properties. In *Proc. 1st SIGCOMM Workshop on Internet Measurement (IMW '01)*, pages 197–211, New York, NY, USA, 2001. ACM.

[ZDE+93] Lixia Zhang, Stephen Deering, Deborah Estrin, Scott Shenker, and Daniel Zappala. RSVP: A new resource ReSerVation protocol. *IEEE Network*, September 1993.