

Transport Area Working Group  
Internet-Draft  
Intended status: Informational  
Expires: September 6, 2007

B. Briscoe  
BT & UCL  
March 05, 2007

Flow Rate Fairness: Dismantling a Religion  
draft-briscoe-tsvarea-fair-01.pdf

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on September 6, 2007.

Copyright Notice

Copyright (C) The IETF Trust (2007).

Abstract

Resource allocation and accountability have been major unresolved problems with the Internet ever since its inception. The reason we never resolve these issues is a broken idea of what the problem is. The applied research and standards communities are using completely unrealistic and impractical fairness criteria. The resulting mechanisms don't even allocate the right thing and they don't allocate it between the right entities. We explain as bluntly as we can that thinking about fairness mechanisms like TCP in terms of

sharing out flow rates has no intellectual heritage from any concept of fairness in philosophy or social science, or indeed real life. Comparing flow rates should never again be used for claims of fairness in production networks. Instead, we should judge fairness mechanisms on how they share out the `cost' of each user's actions on others.

Summary of Changes (to be removed by the RFC Editor on Publication)

Full diffs created using the rfcdiff tool are available at  
<<http://www.cs.ucl.ac.uk/staff/B.Briscoe/pubs.html#rateFairDis>>

From -00 to -01:

Toned down the polemic.

Added Section 8 "Critiques of Specific Schemes", adding subsections on TCP and WFQ to previously disparate material on max-min fairness, TFRC & router-based fairness approaches like XCP, which have been shortened and clarified. Also added subsections on TCP-style fairness wrt RTT and packet size that has been copied by other transports.

Added substantial new Section 9 "Implications for the RFC Series". Added to the introduction the importance of the issue and the general implications.

Created an expanded and clarified new subsection Section 5.2 "Comparing Costs" from text previously at the end of Section 5.1 "Something to Integrate the Allocations"

Added quote on flow granularity from RFC2309 & RFC2914.

Clarified and expanded Section 5.3.2 "Enforcing Cost Fairness".

Various clarifications throughout.

Table of Contents

- 1. Introduction . . . . . 4
- 2. Requirements notation . . . . . 6
- 3. Fair Allocation of What Among What? . . . . . 7
  - 3.1. Structure of Document . . . . . 7
- 4. Cost, not Benefit . . . . . 8
- 5. Economic Entities not Flows . . . . . 11
  - 5.1. Something to Integrate the Allocations . . . . . 11
  - 5.2. Comparing Costs . . . . . 13
  - 5.3. Enforcement of Fairness . . . . . 15
    - 5.3.1. Cheating with Whitewashed or Split Flow Identities . . 15
    - 5.3.2. Enforcing Cost Fairness . . . . . 17
- 6. Fairness between Fairnesses . . . . . 19
- 7. The Seminal Literature . . . . . 21
- 8. Critiques of Specific Schemes . . . . . 24
  - 8.1. Max-min flow rate fairness . . . . . 24
  - 8.2. TCP . . . . . 24
  - 8.3. TFRC . . . . . 24
  - 8.4. RTT and Fairness . . . . . 25
  - 8.5. Packet Size and Fairness . . . . . 26
  - 8.6. XCP and router-based fairness schemes . . . . . 26
  - 8.7. WFQ . . . . . 27
- 9. Implications for the RFC Series . . . . . 27
- 10. IANA Considerations . . . . . 30
- 11. Security Considerations . . . . . 30
- 12. Conclusions . . . . . 30
- 13. Acknowledgements . . . . . 32
- 14. Comments Solicited . . . . . 32
- 15. References . . . . . 32
  - 15.1. Normative References . . . . . 32
  - 15.2. Informative References . . . . . 33
- Author's Address . . . . . 39
- Intellectual Property and Copyright Statements . . . . . 40

## 1. Introduction

"But he has nothing on at all."

The Emperor's New Clothes, Hans Christian Andersen

This document is deliberately destructive. It sets out to destroy an ideology that is blocking progress---the idea that fairness between multiplexed packet traffic can be achieved by controlling relative flow rates alone. Flow rate fairness was the goal behind fair resource allocation in widely deployed protocols like weighted fair queuing (WFQ), TCP congestion control and TCP-friendly rate control [WFQ][RFC2581][RFC3448]. But it is actually just unsubstantiated dogma to say that equal flow rates are fair. This is why resource allocation and accountability keep reappearing on every list of requirements for the Internet architecture (e.g. [NewArchReq]), but never get solved. Obscured by this broken idea, we wouldn't know a good solution from a bad one.

Controlling relative flow rates alone is a completely impractical way of going about the problem. To be realistic for large-scale Internet deployment, relative flow rates should be the outcome of another fairness mechanism, not the mechanism itself. That other mechanism should share out the `cost' of one user's actions on others---how much each user's transfers restrict other transfers, given capacity constraints. Then flow rates will depend on a deeper level of fairness that has so far remained unnamed in the literature, but is best termed `cost fairness'.

It really is only the idea of flow rate fairness that needs destroying---nearly everything we've engineered can remain. The Internet architecture needs some minor additions, but otherwise it is largely already suited to cost fairness.

The metric required to arbitrate cost fairness is simply volume of congestion, that is congestion times the bit rate of each user causing it, taken over time. In engineering terms, for each user it can be measured very easily as the amount of data the user sent that was dropped. Or with explicit congestion notification (ECN [RFC3168]) the amount of each user's data to have been congestion marked. Importantly, unlike flow rates, this metric integrates easily and correctly across different flows on different paths and across time.

What we call cost fairness has been in the literature for nearly a decade, but it hasn't been put so bluntly before. We were moved to spell it out unambiguously (and avoiding maths), because this isn't just some dry academic fairness debate that might change allocation

percentages somewhere in the third decimal place. The outcomes due to flow rate fairness that we see on the Internet today are hopelessly unlike the outcomes that would result from cost fairness.

Not that the outcomes we see are the deliberate intent of flow rate fairness. They are the random result of an absence of fairness control, because flow rate fairness isn't even capable of reasoning about questions like, "How many flows is it fair to start between two endpoints? or over different routes?" or, "What rate is fair for a flow that has been running longer than another?".

Resource allocation and accountability are two issues that reappear on every list of requirements for a new Internet architecture [NewArchReq]. We could have started filling this architectural vacuum a decade ago, but architecture not only requires foundational ideas, it also requires consensus. In 1997, the basis of the dominant consensus was completely undermined, but it didn't even notice.

While everyone prevaricates, novel p2p applications have started to thoroughly exploit this architectural vacuum with no guilt or shame, by just running more flows for longer. Application developers assume, and they have been led to assume, that fairness is dealt with by TCP at the transport layer. In response some ISPs are deploying kludges like volume caps or throttling specific applications using deep packet inspection. Innocent experimental probing has turned into an arms race. The p2p community's early concern for the good of the Internet is being set aside, aided and abetted by commercial concerns, in pursuit of a more pressing battle against the ISPs that are fighting back. Bystanders sharing the same capacity are suffering heavy collateral damage.

This trend has spread beyond the p2p community. There is now no shame in opening multiple TCP connections, or offering VoIP or video streaming software without any congestion control.

Whether the prevailing notion of flow rate fairness has been the root cause or not, there will certainly be no solution until the networking community gets its head out of the sand and understands how unrealistic its view is, and how important this issue is. Certainly fairness is not a question of technical function---any allocation `works'. But getting it hopelessly wrong badly skews the outcome of conflicts between the vested interests of real businesses and real people.

But isn't it a basic article of faith that multiple views of fairness should be able to co-exist, the choice depending on policy? Absolutely correct---and we shall return to how this can be done

later. But that doesn't mean we have to give the time of day to any random idea of fairness.

Fair allocation of rates between flows was used in good faith as a guiding principle, but it isn't based on any respected definition of fairness from philosophy or the social sciences. It has just gradually become the way things are done in networking. But it's actually self-referential dogma. Or put more bluntly, bonkers.

We expect to be fair to people, groups of people, institutions, companies---things the security community would call `principals'. But a flow is merely an information transfer between two applications. Where does the argument come from that information transfers should have equal rights with each other? It's equivalent to claiming food rations are fair because the boxes are all the same size, irrespective of how many boxes each person gets or how often they get them.

Because flows don't deserve rights in real life, it is not surprising that two loopholes the size of barn doors appear when trying to allocate rate fairly to flows in a non-co-operative environment. If at every instant a resource is shared among the flows competing for a share, any real-world entity can gain by i) creating more flows than anyone else, and ii) keeping them going longer than anyone else.

We appeal to the networking community to quietly set aside rate fairness between flows. It had its time, but now it has been shown to be unfounded, unrealistic and impractical. Papers and standards proposals using it should be rejected. No-one should ever have to cater for it in future Internet protocols---it places arbitrary requirements on the system that can't be met and wouldn't achieve any meaningful sort of fairness even if they could be met.

Alternatively, someone should write a defence of flow rate fairness. Continuing to use flow rate fairness as the dominant ideology, without rebutting Kelly's seminal 1997 paper that undermined it, just leaves the Internet community divided into religious sects, making a mockery of the scientific process towards consensus.

## 2. Requirements notation

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

### 3. Fair Allocation of What Among What?

The issue with flow rate fairness is far more basic than whether allocations should be max-min, proportional or whatever. Flow rate fairness doesn't even allocate the correct thing. And it doesn't allocate it among the correct entities either. At this most basic level we will contrast the two main contending views:

- o Allocate rate among flows (flow rate fairness)
- o Allocate congestion cost among the bits sent by economic entities (cost fairness)

When cost fairness was proposed, it stated its case in terms of the dominant belief system---flow rate fairness. Unfortunately, this meant that the dominant belief system didn't notice it had been struck an intellectual death blow. Its believers carried on regardless and it remains dominant today.

As a result, one sees talk of weighted proportional fairness in the same context as proportional, max-min or min-max fairness as if they are all members of the same set. They are not. Weighted proportional fairness has an extra weight parameter  $w$  that all the

others lack. With weighted proportional fairness, the interesting bit is what regulates users in their choice of  $w$ . Otherwise, it would hardly be a useful definition of fairness to say it is fair for flow A to go  $w$  times as fast as flow B, if the user behind flow A has free choice of  $w$ .

In fact, it turns out that the interesting bit is nothing to do with flows, or their weights. For internetworking the only interesting definition of fairness depends on the allocation of cost among the bits sent by economic entities, regardless of which flows the bits are in. A user's choice of  $w$  then depends on that.

#### 3.1. Structure of Document

The body of this document is structured around our question, "Fair allocation of what among what?":

- o Section 4 answers the "...of what...?" question, explaining why fair allocation of costs is a sufficient and realistic form of fairness, and allocation of rate is not. A sub-section also explains why TCP fair rate control (TFRC) causes greater congestion costs than TCP, because it wrongly tries to achieve equality of flow rate.

- o Section 5 answers the "...among what?" question, explaining why fairness among flows can only be myopic whereas fairness among economic entities naturally spans all flows and spans time. Also we show fairness among flows is hard, if not impossible, to enforce while enforcing fairness among economic entities seems practical.

Having debunked the dominant ideology of flow rate fairness, and replaced it with cost fairness, in Section 6 we discuss how other forms of fairness can be asserted locally. Then, before we draw conclusions, Section 7 maps the progression of seminal ideas in the literature on which this memo is based and Section 8 outlines concrete criticisms of specific fairness schemes: max-min flow rate fairness, TCP, TFRC, WFQ and XCP as well as discussions of dependence on RTT and packet size. Finally, Section 9 surveys the implications on the RFC series if we are to dismantle all traces of flow rate fairness. A FAQ Web page [FairFAQ] is planned to answer some frequently asked questions that didn't fit easily into the main flow of this document.

#### 4. Cost, not Benefit

The issues of fair allocation of resources comes under the domain of political economy, with philosophy reasoning about our judgements. In Section 6 we will discuss how different fairness policies can co-exist. But to answer our question, "Fair allocation of what?" we start from the premise used in economics (and life) that fairness concerns comparing benefits, costs or both.

The benefit of a data transfer can be assumed to increase with flow rate, but the shape and size of the function relating the two (the utility function) is unknown, subjective and private to each user. Flow rate itself is an extremely inadequate measure for comparing benefits: user benefit per bit rate might be ten orders of magnitude different for different types of flow (e.g. SMS and video). So different applications might derive completely different benefits from equal flow rates and equal benefits might be derived from very different flow rates.

Turning to the cost of a data transfer across a network, flow rate alone is not the measure of that either. Cost is also dependent on the level of congestion on the path. This is counter-intuitive for some people so we shall explain a little further. Once a network has been provisioned at a certain size, it doesn't cost a network operator any more whether a user sends more data or not. But if the network becomes congested, each user restricts every other user, which can be interpreted as a cost \_to all\_---an externality in

economic terms. For any level of congestion, Kelly showed [wPropFair] that the system is optimal if the blame for congestion is attributed among all the users causing it, in proportion to their bit rates. That's exactly what routers are designed to do anyway. During congestion, a queue randomly distributes the losses so all flows see about the same loss rate (or ECN marking rate); if a flow has twice the bit rate of another it should see twice the losses. In this respect random early detection (RED [RFC2309]) is slightly fairer than drop tail, but to a first order approximation they both meet this criterion.

So in networking, the cost of one flow's behaviour depends on the congestion volume it causes which is the product of its instantaneous flow rate and congestion on its path, integrated over time. For instance, if two users are sending at 200kbps and 300kbps into a 450kbps line for 0.5s, congestion is  $(200+300-450)/(200+300) = 10\%$  so the congestion volume each causes is  $200k \times 10\% \times 0.5 = 10kb$  and  $15kb$  respectively.

So cost depends not only on flow rate, but on congestion as well. Typically congestion might be in the fractions of a percent but it varies from zero to tens of percent. So, flow rate can never alone serve as a measure of cost.

To summarise so far, flow rate is a hopelessly incorrect proxy both for benefit and for cost. Even if the intent was to equalise benefits, equalising flow-rates wouldn't achieve it. Even if the intent was to equalise costs, equalising flow-rates wouldn't achieve it.

But actually a realistic resource allocation mechanism only needs to concern itself with costs. If we set aside political economy for a moment and use pure microeconomics, we can use a competitive market to arbitrate fairness, which handles the benefits side, as we shall now explain. Then once we have a feasible, scalable system that at least implements one defined form of fairness, we will show how to build other forms of fairness within that.

In life, as long as people cover the cost of their actions, it is generally considered fair enough. If one person enjoys a hot shower more than their neighbour enjoys the toast they made with equal units of electricity, no-one expects the one who enjoyed the shower to have to pay more. If someone makes more of their lot in life than another, some complain it's not fair, but most call this envy, not unfairness.

Market economics works on the same premise (unsurprisingly given life and market economics are closely related). In fact, a market adjusts

supply to meet demand so that benefits are as fairly distributed as is consistent with the pre-existing inequalities in life. But this fairness between benefits is an outcome and it is only met as long as a market mechanism makes people accountable for the costs of their actions (and various market failures are avoided).

We deliberately say 'make people accountable' to avoid the phrase 'make people pay', because users tend to prefer flat rate subscription for Internet access. So, ISPs want to be able to limit the congestion costs they are able to cause (Section 5.3.2), not just charge them for whatever unlimited costs they cause.

If we do make users truly accountable for the cost of the congestion they cause, a form of fairness between flow rates emerges automatically. As everyone increases the rate of each of their flows, congestion rises. As congestion rises, everyone pays due regard to the share of the cost attributed to them. So, each individual will want their congestion control algorithm to continuously adjust its rate to maximise their net utility--benefit minus cost. Kelly [wPropFair] shows that even if each user keeps their utility function private but we model all the different users

by an arbitrary weight that scales their utility function relative to others, users will allocate themselves flow rates so that the cost they cause will equal the weight they choose--weighted proportional fairness.

But such a flow rate allocation is not the measure of fairness, it is merely a possible outcome caused by cost fairness, given some assumptions about how to model the shape of users' private utility functions. Enforcing underlying cost fairness is in itself a sufficient form of fairness. We repeat: the resulting relative flow rates are not the measure of fairness.

Most importantly, Kelly proved cost fairness would lead everyone to maximise their combined aggregate utility across the whole Internet. In other words, if anyone was allocated more and someone else less, the outcome would be worse as a whole. This is why cost fairness is so important, as other forms of fairness cannot be better, unless some major flaw is found in Kelly's assumptions. Kelly et al also proved that, even though relative flow rates would likely be very different from those seen today, the Internet would remain stable given reasonable constraints and assumptions [wPropStab].

While on the subject of assumptions, we should add that the benefit of a real-time application depends on jitter, not just transfer rate. But simple scaling arguments show that it will be possible for network operators to minimise congestion delay as networks increase in capacity ([SelfMan] S.2), an argument supported by recent research

showing that router buffers are often significantly oversized [BufSizUp].

On the other hand, no-one has even tried to claim that flow rate equality achieves any fairness objective. It has just been asserted as an arbitrary engineer's dogma. This is why flow rate fairness is so open to criticism as unrealistic---having no basis in any recognised form of fairness in real life, science or philosophy.

Proponents of flow-rate fairness might be forgiven for aiming for an 'unrealistic' form of fairness if a 'realistic' form was difficult to implement in practice. In fact, it is flow rate fairness that is completely impractical to enforce (Section 5.3.1). The reason we are resurrecting cost fairness is that we believe there are now much more practical ways to enforce it---ways that are built around existing Internet congestion control but, unlike Kelly's, they don't require all ISPs to change their retail model to congestion charging (Section 5.3.2).

But how would users "allocate themselves flow rates in proportion to the share of the cost that they cause"? If they were made accountable for congestion, they would install a version of TCP with a weight parameter (for example MultTCP [MultTCP]), at least for TCP-based applications. Of course, most users wouldn't want the fuss of weighting each individual flow. But if they chose to set policies on average for large classes of flows (or to accept the defaults set by application developers), the resulting suboptimal outcome for themselves would be their own private choice to trade optimality against hassle. The underlying fairness criterion would still be met: that people should be accountable for the costs they cause to others.

In contrast, with flow-rate fairness, two flows may cause orders of magnitude different costs to others (for instance if one has been running orders of magnitude longer) by running at equal rates. Nowhere do we find any justification for the dogma that flow rates must be equal to be fair. Nowhere do we find any rebuttal of Kelly's destruction of flow rate fairness, even after nine years.

## 5. Economic Entities not Flows

### 5.1. Something to Integrate the Allocations

Imagine loaves of bread are regularly delivered to a famine-struck refugee camp. Each time a loaf is brought out, a queue forms and the loaf is divided equally among those in the queue. If the individuals who appear in each queue are always different, except for one who

always appears in every queue, would it still be fair to share each loaf equally among those in each queue?

This example shows that realistic fairness policies must depend on an individual's history. But if that isn't a convincing argument, it doesn't have to be. We don't have to show that fairness policies must depend on history, only that realistic ones probably will. So a fairness mechanism that claims to support commercially realistic fairness policies must be structured to hold individual history without destroying scalability. And here, 'individual' means some real-world entity with an economic existence, not a flow.

Router-based flow rate fairness mechanisms tend to have to be myopic. To be otherwise would seem to require holding the history of most Internet connected individuals on most routers, because a flow from nearly any individual in the world might appear at nearly any router. So instead, router-based schemes tend to share out flow rate at each instant without regard to individual history---and unfortunately without regard to commercial reality.

Instead of arbitrating fairness on routers, fairness can be and already is arbitrated where state can be held scalably---at the endpoints where the congestion costs of each individual are already collected together. One reason for our frustration with the networking community's focus on flow rate fairness is that the TCP/IP-based architecture of the Internet already has a structure very close to that required to arbitrate fairness based on the costs that individuals cause, rather than on flow rates.

Congested routers generate cost signals (losses or ECN marks) that are carried to the transport causing the congestion, piggy-backed in the packet stream either as gaps in the transport stream or as ECN marks. These congestion signals are already fed back to the sending transport by nearly all transport protocols. And congestion control algorithms like TCP already adapt their flow rates in response to congestion. So all we would need to change would be to use a weighted TCP algorithm [MultTCP] (or equivalent for inelastic applications) that could weight itself under the control of a process overarching all the flows of one user, which would take into account the user's cost history across all flows.

Of course, there is no incentive for anyone to voluntarily subject themselves to such fairness (nonetheless, they already subject themselves to TCP which voluntarily halves its rate whenever it senses congestion). But as we shall see in Section 5.3.1, policing fairness between individuals (and between networks) at their point of attachment to the Internet has already been solved, whereas getting every router to police fairness between every individual connected to

the Internet is a pipedream, because it would be extremely complicated for routers to have to know about individuals globally.

## 5.2. Comparing Costs

So, how come one attachment point can arbitrate fairness between everyone on the Internet when it only knows about locally attached individuals? Do we have to add some fully connected mesh of coordination messages between every endpoint in the world? The answer is no, because, in a very subtle sense, we already have such a mesh. Fairness at one endpoint is kept in line with all the others by the commonly aligned discipline of `_cost_` throughout the globe. Cost in any part of the world has an exchange value with cost in any other part, because, wherever there's an Internet attachment, there's a connection with the global economy.

Different types of users (heavy users, light users, servers, server farms, companies) will want to be able to cause different volumes of congestion. As long as congestion can be equated to cost, it can be related to the amount each user has paid for their attachment to the Internet. Even if some localised authority asserts a non-economic variant of fairness between some sub-set of users (e.g. in a university or corporation), the authority as a whole will still align its understanding of cost with that of the global economy (see Section 6).

To be able to compare costs globally, we cannot merely talk of volume of congestion as a cost to other users without calibrating it---without specifying how it relates to monetary cost. In a competitive market, the monetary cost that should be assigned to congestion volume turns out to be the marginal cost of the capacity needed to alleviate the congestion [PrCong] (see FAQ [FairFAQ] for details)

The term 'marginal' cost is used in economics for the slope of the curve of cost against capacity. To take a toy example, imagine a 10Gbps interface card costs \$1,000 and the cost follows a rough square root law so that a 20Gbps interface card will cost about \$1,400 ( $\sqrt{2}$ x the cost for 2x the capacity). Even though the average cost of the 10Gbps card is \$100 per Gbps, the marginal cost is only \$50 per Gbps, implying an 11Gbps card (if cards could be upgraded with such fine granularity) would cost about \$1,050.

Note that when we say that the cost of congestion equates to the marginal cost of capacity, we are not introducing any additional cost; we are merely categorising cost into sub-divisions. So, an existing flat fee Internet charge should be considered to consist of parts to cover:

- o operational (non-capacity) costs;
- o capacity upgrade costs to alleviate congestion (the \$50/Gbps);
- o the balance of the cost of capacity ( $\$100 - \$50 = \$50/\text{Gbps}$ ).

The distinction between the last two is important, because the cost of capacity is traditionally shared out in proportion to access link capacity. But different users with the same access link capacity can cause hugely different volumes of congestion across time and across all the Internet links they regularly use, so it is fair to share out the upgrade cost part in proportion to congestion caused, not access link capacity.

Once a cost is assigned to congestion that equates to the cost of alleviating it, users will only cause congestion if they want extra capacity enough to be willing to pay its cost. Of course, there will be no need to be too precise about that rule. Perhaps some people might be allowed to get more than they pay for and others less. Perhaps some people will be prepared to pay for what others get, and so on.

But, in a system the size of the Internet, there has to be some handle to arbitrate how much cost some users cause to others. Flow rate fairness comes nowhere near being up to the job. It just isn't realistic to create a system the size of the Internet and define fairness within the system without reference to fairness outside the system---in the real world where everyone grudgingly accepts that fairness usually means "you get what you pay for".

Note that we use the phrase "you get what you pay for" not just "you pay for what you get". In Kelly's original formulation, users had to pay for the congestion they caused, which was unlikely to be taken up commercially. But the reason we are revitalising Kelly's work is that recent advances (Section 5.3.2) should allow ISPs to keep their popular flat fee pricing packages by aiming to ensure that users cannot cause more congestion costs than their flat fee pays for.

The engineering details of all these commercially realistic systems don't have to concern the research or standards communities in networking. It is sufficient to design protocols so that congestion costs can be integrated into one simple counter across different flows and across time for some higher layer to use, so that senders can be made accountable for the congestion they cause. Systems and protocols intended for Internet deployment do not have to always realise the sort of fairness over time that we find around us in the real world, but they must be able to.

This subtle connection with the global economy at every Internet attachment point ensures that there is no need for some system to decide how far back the history of each individual's costs should still be taken into account. Once the cost that one entity causes to others (integrated over time and over all its flows) has been suffered by that entity itself, it can be forgotten. Just like the costs for all the other benefits everyone assimilates in their daily lives. And the concept of a customer account also naturally ensures that a user cannot escape accountability merely by roaming or mobility.

Finally, note well that this `ISP' and `customer' terminology doesn't preclude peer-to-peer creations that arbitrate fair use of the resources of a self-provided community network [ArchP2pEcon].

### 5.3. Enforcement of Fairness

#### 5.3.1. Cheating with Whitewashed or Split Flow Identities

In the real world of deployed networks, if it is easy to cheat the fairness mechanism to get an unfair allocation, it's hardly a useful fairness mechanism. All known flow rate fairness mechanisms are wide open to cheating. The network community cannot continue in denial of this glaring inconsistency if we claim to be designing commercially realistic protocols.

For instance, if I am the customer of a system giving max-min flow rate allocations, it is in my interest to split the identities of my flows into lots of little flows until they are all less than the minimum allocation. Then the system will dance to my tune and reduce the allocations of everyone else in order to increase all the allocations of my little flows. The more I split my traffic down across more and more identifiers, the larger share of the resource all my flows taken together will get.

If a history-based fairness mechanism (Section 5.1) believes it should allocate fewer resources to one flow identifier that it considers has already been given enough, it is trivially easy for the source behind that identifier to create a new identifier with a whitewashed reputation for its traffic.

And it's no good imagining that a router will be able to tell which flow IDs are actually all from the same entity (either in the security sense or the economic sense), because routers have to arbitrate between flows emanating from networks many domains away. They cannot be expected to know which sets of flow identifiers should be treated as a single entity. Flows between a pair of IP addresses may even be attributable to more than one entity, for instance, an IP

address may be shared by many hundreds of accounts on a Web or e-mail hosting site or behind a NAT.

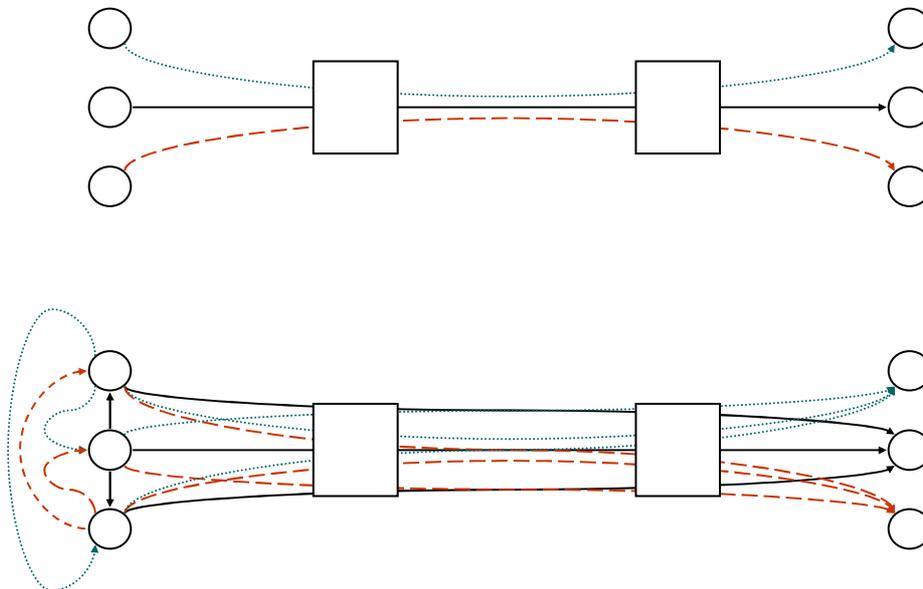


Figure 1: Splitting flow identifiers to cheat against flow rate fairness.

Bottleneck policers [pBox],[XCHOKe],[AFD], suffer from the same inherent problem. They look for a flow ID at a bottleneck that is consuming much more bit rate than other flows in order to police use of TCP. But anyone can cheat by simply running multiple TCP flows. If the policer looks for cheating pairs of source-destination IP addresses, without regard to port numbers, a pair of corresponding nodes can still cheat by creating extra flows from spoofed source addresses after telling each other out of band where to send acknowledgements (or just using error correcting coding, not acks).

Alternatively, pairs of corresponding nodes can collude to share parts of each other's flows. For instance, if the three pairs of nodes in Figure 1 are trying to communicate, the senders can act as stepping stones for each other so that their three (n) flows appear as nine (n^2) across the bottleneck link in the middle. In effect, they have created a routing overlay, much like BitTorrent file-sharing software does. If one pair of naive nodes competes for this bottleneck against n pairs of nodes adopting this strategy, it will get about n times smaller share than each of the other pairs, assuming n is large.

These inherent problems with how to define flow granularity were

understood when recommendations on active queue management (AQM) were made [RFC2309], also quoted in the IETF's best current practice on congestion control [RFC2914]. The problem was known to be particularly acute in the context of the above bottleneck policer ideas, which were current at the time. But the answer was left open "We would guess that the source/destination host pair gives the most appropriate granularity in many circumstances. The granularity of flows for congestion management is, at least in part, a policy question that needs to be addressed in the wider IETF community."

Given identifiers can generally be freely created in cyberspace, it is well-known that they shouldn't be relied on for resource allocation (or more generally for negative reputation) [FrRideP2p],[CheapPseud]. Kelly [wPropFair] chose cost-based fairness (his term was `pricing per unit share') because it was immune to this problem---it allocates cost to bits not to flows and hence doesn't rely on any cyber-identifiers.

In summary, once one accepts that fairness should be based on concepts from social science, fairness can only be meaningful between entities with real-world identities---humans, organisations, institutions, businesses. Otherwise two entities can claim to have arbitrarily many flows between them, making fairness between flows completely meaningless.

### 5.3.2. Enforcing Cost Fairness

If enforcing flow rate fairness is impractical, is enforcing cost fairness any more achievable? Happily, the Internet's architecture is already suited to carrying the right cost information for cost fairness mechanisms to be enforced in a non-co-operative environment.

Kelly's stated motivation for his focus on pricing was so that the system would be applicable to a non-co-operative environment. In 1999, Gibbens and Kelly went further, pointing out [Evol\_cc] that ECN [RFC3168] provided an ideal basis on which to base cost fairness. The idea was simply for network operators to ECN mark traffic at congested routers without regard to flows, then to apply a price to the volume of traffic carrying ECN marks, which would make the transport endpoints accountable for the congestion they caused.

However, understandably, the idea of Internet retailers charging their end-customers directly for congestion met strong resistance. Customers are known to be highly averse to unpredictable charges for services ([PMP] S.5) so duration charging for each Internet flow was unlikely to replace flat monthly charging.

Many threw out the baby with the bath water, associating Kelly's

theoretical work solely with its suggested pricing model. But over the ensuing years, an active research community has sought to keep the underlying theory but wrapped around with a more realistic incarnation.

Indeed the recent proposal called re-ECN [Re-TCP] claims to do just that. Here the discussion is confined to whether the economic structure and functional effect on the network service that re-ECN aspires to is valid. If it is, the research agenda should be focused on producing that outcome, even if re-ECN itself isn't the answer. (Readers tempted to game re-ECN shouldn't rely on the brief description here; rather they should use the full spec above, which, as of early 2007, documents one outstanding vulnerability and defences against other known attacks.)

Re-ECN aims not to constrain retail pricing, requiring no change to typical flat rate Internet contracts. But it enables addition of a policer that can limit the volume of congestion a customer's sent traffic causes over, say, a moving month. Thus, if endpoint congestion control (e.g. MultTCP [MultTCP]) doesn't voluntarily act fairly the network ingress can force it to. It is expected that various styles of policing (including none) will evolve through market selection.

Although Gibbens & Kelly rightly identified that standard ECN reveals the necessary information for cost-based fairness, it doesn't reveal it in the right place for network layer policing---at the `_sender's_` network attachment. In the current TCP/IP architecture, congestion information emerges from the end of a forward data path, which is the last point in the feedback loop that any network operator can reliably intercept it---the wrong end for policing the sender.

Re-ECN preserves IP's datagram model, but makes delivery conditional on the sender `pre-loading' packet streams with enough `credit' to remain non-negative despite being decremented by congestion experienced along the path. It should then be in `_both_` the endpoints' interests for the sender to use a pattern of feedback where the sender re-inserts the feedback from each congestion event into the next sent packet as a `credit' (re-feedback [Re-fb]). It should also be in the sender's interest to start every flow slowly and with some initial `credit' while it establishes the path's congestion level.

Like Kelly's original proposal, re-ECN uses ECN routers (and receivers) unchanged to ensure the cost of congestion is communicated to each transport causing it, precisely in proportion to their bit rates, without any per-flow processing in the network. But, unlike Kelly, sources not receivers are held responsible and the network

cannot raise unsolicited charges without the sender deliberately marking packets itself.

Re-ECN also aims to ensure cost-fairness between whole networks. Because the congestion level in every stream of packets decrements towards zero, at an inter-domain border both neighbouring networks can count the bulk volume of congestion that the passing packets are causing downstream of the border. If the downstream neighbour charges the upstream neighbour this cost (complementing fixed capacity charges), the upstream network should in turn want to ensure its upstream users (or networks) are accountable for their share of these costs arriving from their borders.

Each network could choose to share out its downstream costs between its upstream customers by some other fairness policy than cost

(including absence of policy, which ensures incremental deployment). So, on the grander scale, re-ECN aims to ensure that networks have to be fair to each other, and that different fairness policies can co-exist, which is the subject of the next section.

## 6. Fairness between Fairnesses

A social anthropologist would be able to give numerous examples of tribes and societies holding differing opinions on fairness. But, we must also recognise that societal views of fairness are heavily influenced by the fairness that a market would produce [SovJstce]. Just as gravity pre-dated Newton, the invisible hand of the (maturing) market had been allocating resources in society long before Adam Smith noticed, particularly where the larger picture of trade between societies was concerned. Equality is sometimes considered fair for life's essentials, but in life few expect to get an equal share of every cake for nothing. As a society, we accept that a reasonably competitive market mechanism does produce a 'realistic' form of fairness; a form of fairness that people grudgingly accept they have to live with, where the buyer gets no more than she pays for, at a competitive price that reflects the effort expended by the seller.

However, monarchs, governments, charities and so on have also been stamping their own view of fairness on this backdrop, sometimes less equal sometimes more. But even if different allocation schemes are chosen locally, perhaps taking account of social inequality, on a global scale arbitration between local views on fairness has largely been through market economics---we are not asking anyone to judge whether this is good or bad, it just is. The Internet should at least be able to cope with the world as it is (as well as how it might be). This doesn't imply we believe that economic forces are

somehow above policy control. Rather, we observe that market forces (aside from wars) have been the default global resource allocation mechanism over many centuries. In the Greco-Roman civilisations, in the Buddhist, Confucian and later in the Islamic world, trade was a necessary but not central aspect of life. And over the last two decades, Western civilisations have been going through a phase of 'economics imperialism', where attempting to exert policy control over economics is even viewed as counter-productive.

However, we must not assume the current globalisation trend [Saul05] heralds the end of history. The Internet should be able to reflect the shifting of societal forces as different local fairness regimes come and go---'design for tussle' [Tussle]. On the whole, interworking of resource allocation between most parts of the Internet must be able to be based on market economics, but it should be possible to apply other fairness criteria locally. For instance, a University might choose to allocate network resources to each student equally rather than by how much their parents can afford. But the network resources one whole University gets relative to another institution depend on how much each pays their service provider.

With arbitration of fairness at the network edge, these enclaves where local fairness prevails can be virtual overlays; they need not align with physical network boundaries. A distance-learning University or company with a mobile sales-force could buy quotas from different networks and redistribute the aggregate among its members using its own view of fairness. Or whole countries might arrange to subsidise a minimum universal service obligation for Internet usage, but still, the country as a whole would be expected to pay its way in the world.

On the other hand, in market-led countries, commercial ISPs might solely allocate resources proportionate to customer subscriptions. Local pockets of heterogeneity will exist, from computer clubs to NATO, but the overall fabric of resource allocation gluing all these pockets together at the (inter)network layer is likely to be market-based.

This is what we mean by 'realistic'---fitting the commercial reality of a global market economy. We are fully aware that the power of market economics can be stretched too far; controlling aspects of society where economic assumptions break down (prompting Samuelson to describe Friedman as "...somebody who had learned how to spell banana but didn't know where to stop" [Swed90]). But we are not advocating that one religion should replace another---market economics replacing flow rate fairness. However, in the case of Internet resource allocation, it must at least be possible to use market economics,

despite its known failings, given it is currently the most appropriate tool for managing conflicting demands on resources from any part of the globe.

A market is meant to optimise allocations in the face of conflicts of self-interest. If we want to assert other fairness regimes, we must recognise this acts against self-interest. If we don't understand how to overcome self-interest, its invisible hand will force its will on us some other way, distorting our attempts to work against it. This is why the loop holes in flow rate fairness are being so thoroughly exploited.

And this is our point. A market mechanism has to be designed. A weak design will be exploited mercilessly. The designs behind flow rate fairness are worse than weak. They are not even aware that, as resource allocation mechanisms, they should be able to meet the stringent requirements of a good market mechanism, such as forgery-resistant 'currency', information symmetry, internalisation of externalities and so forth.

If we did wish to promote the cause of equality, equalising flow rates would in no way achieve our ends. In fact, it would only promote the cause of selfishness and malice, because flows don't equate to people, so its broken logic can be thoroughly exploited. Only by providing a bullet-proof mechanism to arbitrate self-interest, can we then move on to allocate resources locally in other ways.

## 7. The Seminal Literature

For a rigorous tutorial on the various form of fairness, the reader is referred to Le Boudec [ccFairTut].

Max-min flow rate fairness has a long history in networking, with research to find distributed (router-based) max-min algorithms starting in 1980 [DeMaxMin] and Nagle proposing a novel approach in 1985 [RFC0970]. All these early 'fair queuing' algorithms gave equal rights to each source. In 1989, to solve the problem of some sources deserving more rate than others, the authors of 'weighted fair queuing' (WFQ) proposed that per-source destination pair would be a better model of the size of different sources. It was admitted that a source could deny service to other sources by faking transfers with numerous destinations, but a reasonable trade-off between efficiency and security was required [WFQ]. Recently, an approach called Justice [Jstce] has proposed a return to (weighted) per source fair queuing, but with configurable link weights throughout the network. However, all these 'fair queuing' approaches allocate bit rate as

their measure of fairness.

TCP congestion control was also introduced in the late 1980s [TCPcc], based on the assumption that it would be fair if flow rates through a single bottleneck converged on equality.

In 1991, Mazumdar *et al* [UtilFair] pointed out that there was nothing special about max-min fair rate allocation, and that other *ad hoc* definitions of fairness perhaps based on ratios of individual demands would be no less valid. Instead Mazumdar *et al* advocated that it would be precise to base a definition of fairness on game theory, specifically the Nash bargaining solution. This resulted in proportional fairness, but still using the rate allocated to flows as the measure of fairness.

In 1997, Kelly considered that Mazumdar's use of co-operative game theory was unlikely to be relevant to public networks where fairness would have to be enforced. Instead he introduced *weighted* proportional fairness [wPropFair], which finally broke the link between fairness and flow rates. However, the break in tradition wasn't obvious because the new form of fairness could easily be expressed in terms of flow rates, essentially using the weight of a flow as a *'fiddle-factor'*.

Kelly showed that all a network had to do to achieve fairness in its economic sense (cost fairness) was to share the cost of congestion among bits (not flows). Then, as long as the network made users experience the cost of their bits, users could choose any size flows they wished. But their choice would be regulated by their own trade off between how much they valued bit rate and the charge for congestion.

Kelly's fairness with respect to bit rate per unit charge could also be (and was) framed in terms of fairness between flows by allowing the user an arbitrary choice of weight per flow. But Kelly pointed

out that a flow could be divided into sub-flows without changing the overall rate allocation to all the sub-flows taken together; the user merely had to imagine that the weight she assigned to one flow could be subdivided proportionately into its sub-flows.

Kelly's work built on MacKie-Mason & Varian's seminal paper on the economics of networks from 1995, "Pricing Congestible Network Resources" [PrCong]. This work explained the dual role of congestion costs in controlling demand and regulating supply, in welfare maximising, competitive and monopoly markets.

In his 1997 paper, Kelly framed cost fairness in terms of weighted proportional fairness of flow rates in order to relate to an ATM

technology context. With ATM's flow-based user-network interface, users had to declare the weight they chose for their flows to the network. But by 1998 Kelly *et al* applied this work [wPropStab] to an Internet setting where flows were not part of the user's interface with the network, so flow weights could become a purely private device, internal to the user's rate control algorithm. Nonetheless, the *outcome* at the flow level was still weighted proportional fairness, and the underlying fairness that produced this outcome was still based solely on sharing the cost of congestion among bits.

Back in 1995, Shenker had identified two main types of network traffic: elastic and inelastic, distinguished respectively by their concave and sigmoid utility functions [FundUtil]. Whatever the utility function, Kelly teaches us that covering congestion costs is sufficient to achieve fairness. But then the outcome (in terms of flow rates) depends on the type of utility function:

- o Weighted proportionally fair flow rates will be the outcome for elastic traffic streaming;
- o Inelastic traffic flows hit a discontinuity once congestion rises beyond a certain level, at which point each is better off with zero rate, leading to a need for some form of admission control, whether self-admission control or arbitrated by the network [DCAC]. This was the theoretical backing to the IETF working group on doing admission control using pre-congestion notification (PCN), which is in the process of being chartered [PCNcharter].
- o Key & Massoulie identified a third major class of network traffic where utility is derived solely from the duration required to complete transfer of a fixed volume of data [UtilFile]. They proposed that, if cost fairness applied, self-interested congestion control would toggle between full line rate and zero (with occasional probes). Such behaviour alone can destabilise the network, but it can be stabilised by mixing with streaming traffic [FairIntgr]. Research on the second order incentives necessary to encourage stability continues. Policing rather than pricing congestion is one way to safeguard everyone's common interest in stability.

Since these seminal papers in the late 1990s, theoretical refinement has continued, but the main thrust of research has been to find more realistic and practical ways of applying the insights, a process which is now bearing fruit (see Section 5.3.2).

## 8. Critiques of Specific Schemes

### 8.1. Max-min flow rate fairness

In 1997, Kelly demonstrated [wPropFair] that realistic users would not choose max-min flow rate fairness if they were accountable for the congestion they caused to others. Users would only have chosen max-min if they valued bit rate with an unrealistically extreme set of utility functions that were all identical and that all valued low bit rate infinitesimally less than high bit rate. To spell Kelly's result out even more bluntly, max-min fair rate allocation would only be considered fair if everyone valued bit rate in a really weird way: that is, they all valued very low bit rate hardly any less than very high bit rate and they all valued bit rate exactly the same as each other. (Note that max-min could be meaningful if allocating something like utility among users, but not rate among flows.)

### 8.2. TCP

TCP's congestion avoidance [RFC2581] leads to a form of fairness similar to cost fairness, except it is myopic, only being concerned with each instant in time and with each flow, as explained in Section 5. To be cost fair each user would have to take account of costs across time and across flows, and weight each class of TCP flow according to its importance to them, as can be done with MultTCP [MultTCP].

### 8.3. TFRC

An algorithm that converges on the same flow rate as TCP at equilibrium is called TCP-friendly. It can only claim to be TCP-compatible if it also exhibits the same dynamics as the TCP specification [RFC2581]. Certain streaming applications won't work unless they are allowed a more sluggish response to congestion than TCP's, so researchers invented TCP-friendly rate control (TFRC [RFC3448]) to define fair use of the network in competition with TCP-compatible flows.

`TCP-friendly' congestion control currently has proposed standard status in the IETF [RFC3448], and it is incorporated into one of the congestion control profiles of the new datagram congestion control protocol (DCCP [RFC4342]) that is also a proposed standard.

Given TFRC aims to emulate TCP, by far its most significant fairness problems are those it shares with TCP as just mentioned. However, even if we set aside this myopia in time and within flows, TFRC exhibits an extra fairness problem because its design was based wholly on the broken idea that it is fair for a TCP-friendly flow to

get the same rate as a TCP-compatible flow.

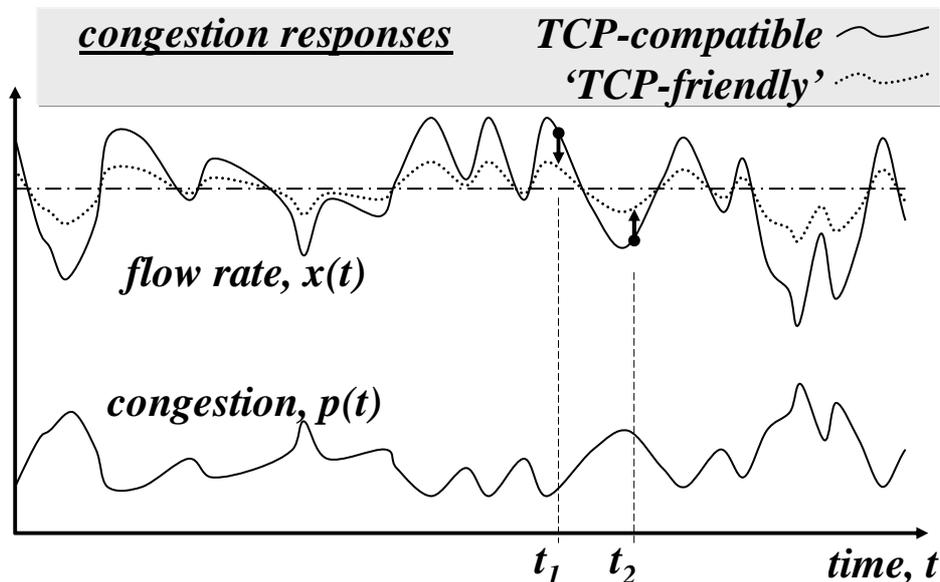


Figure 2: Schematic showing `TCP-friendly' flows cause more congestion than TCP. A TCP-friendly flow is smoother than a TCP-compatible one but with the same mean rate if measured over long enough time. Therefore at times of high congestion ( $t_2$ ) it uses more bandwidth than TCP while at times of low congestion ( $t_1$ ) it uses less.

To explain, we need to remember that both congestion and flow rate vary over time. A more nimble congestion response like TCP's can mirror changing congestion fairly faithfully. It reduces its rate quickly during periods of higher congestion and increases again more quickly whenever congestion falls. In Figure 2 the resulting schematic plots of congestion and flow rate are shown as mirror images of each other. A more sluggish rate response is not as good at tracking the fast-changing congestion process. So the sluggish flow more often uses higher bandwidth when congestion is high, and more often uses lower bandwidth when congestion is low, causing more volume of congestion on average. Giving more during times of plenty doesn't compensate for taking it back during times of scarcity.

#### 8.4. RTT and Fairness

TCP, and congestion controls such as SCTP that inherit from it, converge on a rate that is inversely proportional to round trip time (RTT). This is due to TCP's original design goal of avoiding adding more than one segment to the data in flight each RTT.

Congestion controls certainly have to take RTT delay in the feedback loop into account to ensure stability, but RTT is not in itself a factor that affects fairness. In fact, once a sender is accountable for the congestion it causes, it will be in its own interests to be more cautious on longer RTT paths, as it has proportionately more data in flight so it risks causing more congestion before it can react.

It is perfectly possible to design a robust congestion control that responds more slowly to changes on longer paths, but still converges to the same rate as it would with a shorter RTT. FAST TCP [FAST] is an example of such a congestion control. Siris's weighted window-based congestion controller [WindowPropFair] also has dynamics that are sensitive to RTT, while converging on a bit-rate that is independent of RTT.

Broadly the extra risk of causing congestion with larger RTTs is usually sufficient to encourage behaviour that leads to stability. However, this gross generalisation needs to be couched in assumptions and constraints that are beyond the scope of this memo (and beyond my ability to keep up with the literature).

#### 8.5. Packet Size and Fairness

The issue of how to take packet size into account and whether packet size should be adjusted for in the network (AQM algorithm) or in the transport (rate control algorithm) will be covered in [BytePktMark].

#### 8.6. XCP and router-based fairness schemes

This document has focused on the fairness ideas we see in the production networks around us today. However, our most pressing concern is that these broken ideas also pervade the community working on replacing the Internet architecture. It is well-known that TCP congestion control is running out of dynamic range and many proposals for replacements that can take advantage of higher link capacities by accelerating faster have been put forward. XCP was the first of a family of router-based hi-speed congestion control mechanism, but it is particularly of interest because it claims to allow different fairness criteria to be configured.

However, XCP fairness is based on the myopic flow-rate-based view that we have so roundly criticised in this document. For instance, XCP claims to be able to achieve a weighted proportional fair rate allocation[XCP] S.6 by adding a weight field to each packet, but it glosses over how anyone could regulate each user's choice of the weight. If we compare weighted fair XCP with Kelly's original ATM-based weighted proportional fairness, the weight parameter advises

network equipment on what allocation it should give each flow, but there is no direct congestion information in the XCP protocol that could be used at the ingress to make each source accountable for its choice of weight.

Further, we believe it will be necessary to be able to apply different fairness criteria to different subsets of users of a network and subsets across an internetwork as outlined in Section 6. We cannot immediately see how this would be feasible with router-based approaches like XCP, where routers would seem to have to share information on the history of each user with potentially every other router in the world (as explained in Section 5.1).

A combination of XCP's protocol fields could yield approximate congestion information to integrate each sender's congestion cost history at the access network close to the sender. This would allow the user's choice of weight to be regulated and enable different forms of fairness to be asserted locally. But one then has to question whether it would be simpler for the end system to do the rate control, given it has to give routers all the information they need to arbitrate fairness between flows anyway.

#### 8.7. WFQ

Weighed fair queuing aims to isolate the capacity that a flow receives from excessive load applied by other flows, while at the same time ensuring the router's capacity is fully utilised. WFQ allocates capacity per-flow not per-user, so it is vulnerable to the flow ID splitting games described in Section 5.3.1 and it only controls fairness over flow lifetimes, not over user history. A comparison of cost fairness against WFQ (both as originally defined and as sold commercially) would be interesting given features of the two approaches overlap even though they don't have the same goals. But this subject would require a dedicated paper.

### 9. Implications for the RFC Series

This document points out that the question of cost-fairness between congestion controls sits above the transport layer as a policy concern. Applications would then exert policy control over congestion control in transport protocols (e.g. by setting a weight). This implies that the IETF is not (and never has been) the arbiter of cost-fairness between its protocols, but it should still be responsible for their stability and perhaps their efficiency. This would seem to have wide-ranging implications on the current approach to congestion control standardisation throughout the IETF's RFC series.

RFCs on congestion control fall into the following categories with respect to who is mandated (or encouraged) to do what:

- o Those that specify a congestion control algorithm as a building block without specifying where it should be used (e.g. TFRC [RFC3448]);
- o Those that specify the implementation of congestion control for a specific transport which often draw on building block congestion controls such as TFRC above or TCP (e.g. TCP [RFC2581], SCTP [RFC2960], the DCCP CCIDs [RFC4341][RFC4342] and the RTP profiles such as that for RTP/AVP [RFC3551] and RTP/AVPF with earlier feedback [RFC4585] as well as a number of experimental unicast and multicast protocols);
- o Those that specify that hosts must implement a particular transport (e.g. the 'Requirements for Internet Hosts' [RFC1122]);
- o Those that specify what hosts must do if they implement certain congestion control enhancements (e.g. the 'Congestion Manager' [RFC3124]);
- o Those that specify that applications must implement safe congestion control behaviour (e.g. HTTP/1.1 [RFC2616] and RTP [RFC3550]);
- o Those that specify the meaning of congestion notifications and how buffer implementations should generate them (e.g. recommendations on AQM [RFC2309] and explicit congestion notification [RFC3168]);
- o Those that specify best current practice, guidelines and principles for implementers of congestion control (e.g. the 'Gateway Congestion Control Survey' [RFC1254], recommendations on AQM [RFC2309], 'Congestion Control Principles' [RFC2914], 'General Architectural and Policy Considerations' [RFC3426] and IAB Concerns Regarding Congestion Control for Voice Traffic [RFC3714]);
- o Those that recommend how new transport protocols should interact with existing ones (e.g. recommendations on AQM [RFC2309], Criteria for Evaluating Reliable Multicast Transports [RFC2357] and 'Congestion Control Principles' [RFC2914]).

Generally, the RFC series standardises congestion control by specifying what implementations of a particular transport protocol should or must do in response to congestion events. RFCs generally avoid mandating what users should do, or what networks should allow, which are considered policy concerns. For instance, a TCP

implementation must comply with the congestion control in RFC2581 to be able to claim it is standard TCP, but the RFCs haven't told applications that they must use TCP and they certainly haven't told users that they must only use applications that use TCP (or a TCP-fair alternative).

Therefore, a move to an emphasis on policy control over congestion control will not require changes to the RFCs that specify the implementation of non-policy-based congestion control for specific transports, or congestion control building blocks. These will stand as implementations that can be used by applications that do not desire policy control. Similarly, mandating that a particular transport must be implemented on all hosts, only mandates that it must be available, not that applications must use it.

The RFCs that specify that applications (HTTP/1.1 and RTP) must implement safe congestion control behaviour are sufficiently broadly stated that they are still meaningful after a shift of the congestion control goal-posts.

The RFCs that define congestion notification (RED [RFC2309] and ECN [RFC3168]) are critical standards for cost-fairness and they are already in line with what is required (except for the uncertainty in RFC2309 over byte-mode packet marking, which will be addressed in [BytePktMark]).

The RFCs that specify best current practice, guidelines and principles generally give excellent advice on congestion control.

However, we will have to deal with the RFCs that recommended that applications should use congestion control that results in a flow rate similar to that TCP would achieve under the same conditions, specifically [RFC2309][RFC2357] and [RFC2914]. For instance RFC2357 says, "Note that congestion control mechanisms that operate on the network more aggressively than TCP will face a great burden of proof that they don't threaten network stability."

These RFCs were written in good faith based on the idea that the IETF is responsible for fairness between flow rates, but this memo has now shown that there is nothing at all special about flow rates that happen to be equal when the number of flows from one user and flow durations are considered. We can safely assume that the IETF certainly does not believe it should have any control over the duration of flows, or whether a user should open different flows across different parts of the Internet at different times.

Therefore we will have to update this guidance on fairness to take account of the desires of users and of networks for a fairer outcome

than we have at present. This guidance will also have to address the concerns of the users of transports that implement currently standardised variants of flow-rate fairness.

Some of these `legacy' flows would use more resources and others less if they were under policy control:

- o A future network that protects careful users from aggressive users might well curtail some legacy flows from over-aggressive users (e.g. they might be using application that open many TCP connections that transfer for very long durations).
- o Those legacy flows that use less than they would under policy control seem to be of concern, because they will receive a smaller share of capacity than they would if other flows were not policy-controlled. However, they can upgrade to use policy control if they choose, and they have an incentive to do so. The network will appear more congested than it used to for these flows, but they should still function ok, given they were designed to work over a best efforts service.

Nonetheless, we need to discuss this issue further and reach community agreement on how best to handle the transition towards the more rigorous form of fairness introduced in this memo.

#### 10. IANA Considerations

This document includes no request to IANA.

#### 11. Security Considerations

The whole of Section 5.3 discusses how there are no known ways of enforcing flow rate fairness securely in a non-co-operative environment like the current Internet, whereas practical, secure solutions have been proposed for enforcing cost-fairness.

#### 12. Conclusions

The outstanding barrier to realistic resource allocation for the Internet is purely religious. In much of the networking community you have to put fairness in terms of flow rates, otherwise your work is `obviously' irrelevant. At minimum, you are an outcast, if not a heretic. But actually basing fairness on flow rates is a false god---it has no grounding in philosophy, science, or for that matter `commercial reality'.

It is a classic case of a hegemony where those living within the box have been unaware of the existence of the box, let alone the world outside the box. This memo was written from frustration that no-one inside the box believed that voices outside the box should be listened to. We expect complaints about the blunt style of this document, but it seemed the only way forward was to force the issue, by making the box look ridiculous in its own terms.

Cost fairness was derived from economic concepts of fairness back in 1997. Flow rate fairness had been used in good faith as a guiding principle, but when it is seen through the wider angle of this economic analysis it is clearly broken, even on its own terms. The criticism is far more damning than merely whether allocations are fair. Both the thing being allocated (rate) and what it is allocated among (flows) now appear completely daft---both unrealistic and impractical. However, the Internet community continued to judge

fairness using flow rates, apparently unaware that this approach had been shown to have no intellectual basis. In fact, flow rate fairness algorithms are myopic in both space and time---they are completely unable to control fairness at all, because they don't adjust depending on how many flows users create and for how long.

To be clear, this accusation applies to the so-called 'fairness' that emerges from the TCP algorithm and the various fair queuing algorithms used in production networks. And, more worryingly, this broken idea of flow rate fairness has carried over into the community working on replacing the Internet architecture.

In real life, fairness generally concerns costs or benefits. Flow rate doesn't come anywhere near being a good model of either. User benefit per bit rate might be ten orders of magnitude different for different types of flow. And cost depends on the product of bit rate with congestion, which is very variable and nothing like bit rate alone.

Worse, there is no evidence whatsoever that fairness between flows relates in any way to fairness between any real-world entities that one would expect to treat fairly, such as people or organisations. If fairness is defined between flows, users can just create more flows to get a larger allocation. Worse still, fairness between flows is only defined instantaneously, which bears no relation to real-world fairness over time. Once the idea of fairness based on integrating costs over time is understood, we cannot see any reason to take any form of instantaneous per-flow rate fairness seriously, ever again---whether max-min or TCP.

Even if a system is being designed somehow isolated from the economy, where costs will never have to relate to real economic costs, we

cannot see why anyone would adopt these forms of fairness that so badly relate to real-life fairness. For instance, how can people still be designing schemes to achieve max-min flow rate fairness years after Kelly's proof that users would have to value bit rate in a really weird way in order for max-min fairness to be desirable?

In contrast, cost fairness promises realistic solutions to all these issues. Further, it seems more tractable to enforce, unlike flow rate fairness, which seems inherently broken in this respect. We believe cost fairness is a coherent way forward with all the technical barriers overcome, or close to being overcome. This is where the research & standards agenda should be focused.

If anyone with aspirations to scientific credentials still wants to cling to flow rate fairness, they must justify their preposterous position with reference to some previously respected fairness notions in philosophy or social science. In this memo, we have shown how the whole ideology is unlikely to be up to such rigor.

### 13. Acknowledgements

Thanks are due to Scott Shenker for persuading me to write this, Louise Burness for insight into why people think the way they do, Ben Strulo for giving a better way of expressing it and Marc Wennink and Damon Wischik for challenging the ideas. Also thanks to Arnaud Jacquet, Jon Crowcroft, Frank Kelly and Peter Key for their useful reviews. And thanks to Michael Welzl and Wes Eddy for their excellent survey of Congestion Control in the RFC Series [I-D.irtf-iccr-g-cc-rfcs], on which Section 9 is based. However, the author alone shoulders the blame for any offence caused by the bluntness of style.

### 14. Comments Solicited

Comments and questions are encouraged and very welcome. They can be addressed to the IETF Transport Area mailing list <tsv-area@ietf.org>, and/or to the authors.

### 15. References

#### 15.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

- [RFC2309] Braden, B., Clark, D., Crowcroft, J., Davie, B., Deering, S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G., Partridge, C., Peterson, L., Ramakrishnan, K., Shenker, S., Wroclawski, J., and L. Zhang, "Recommendations on Queue Management and Congestion Avoidance in the Internet", RFC 2309, April 1998.
- [RFC2357] Mankin, A., Romanov, A., Bradner, S., and V. Paxson, "IETF Criteria for Evaluating Reliable Multicast Transport and Application Protocols", RFC 2357, June 1998.
- [RFC2914] Floyd, S., "Congestion Control Principles", BCP 41, RFC 2914, September 2000.
- [RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, September 2001.

## 15.2. Informative References

- [AFD] Pan, R., Breslau, L., Prabhaker, B., and S. Shenker, "Approximate Fairness Through Differential Dropping", CCR 33(2)23--40, April 2003.
- [ArchP2pEcon] Strulo, B., Smith, A., and J. Farr, "An Architecture for Peer-to-Peer Economies", Proc. 3rd Int'l Conf. On Peer-to-Peer Computing (P2P 2003) pp208--209, 2003, <<http://csdl.computer.org/comp/proceedings/p2p/2003/2023/00/20230208.pdf>>.
- [BufSizUp] Ganjali, Y. and N. McKeown, "Update on Buffer Sizing in Internet Routers", ACM SIGCOMM CCR 36, October 2006, <<http://www.sigcomm.org/ccr/drupal/?q=node/72>>.
- [BytePktMark] Briscoe, B., "Byte and Packet Congestion Notification", draft-briscoe-tsvwg-byte-pkt-mark-00 (work in progress), March 2007.
- (to appear---work in progress)
- [CheapPseud] Friedman, E. and P. Resnick, "The Social Cost of Cheap Pseudonyms", Journal of Economics and Management Strategy 10(2)173--199, 1998.

- [DCAC] Gibbens, R. and F. Kelly, "Distributed connection acceptance control for a connectionless network", Proc. International Teletraffic Congress (ITC16) pp941--952, 1999, <<http://www.statslab.cam.ac.uk/~frank/dcac.html>>.
- [DeMaxMin] Jaffe, J., "A Decentralized, "Optimal", Multiple-User, Flow Control Algorithm", Proc. Fifth Int'l. Conf. On Computer Communications pp839--844, October 1980.
- [Evol\_cc] Gibbens, R. and F. Kelly, "Resource pricing and the evolution of congestion control", Automatica 35(12)1969--1985, December 1999, <<http://www.statslab.cam.ac.uk/~frank/evol.html>>.
- [FAST] Jin, C., Wei, D., and S. Low, "FAST TCP: Motivation, Architecture, Algorithms, and Performance", Proc. IEEE Conference on Computer Communications (Infocom'04) , March 2004, <[http://www.ieee-infocom.org/2004/Papers/52\\_2.PDF](http://www.ieee-infocom.org/2004/Papers/52_2.PDF)>.
- [FairFAQ] Briscoe, B., "Cost Fairness FAQ", Web page , October 2006, <<http://www.cs.ucl.ac.uk/staff/B.Briscoe/projects/2020comms/fairfaq.html>>.
- [FairIntgr] Key, P., Massoulie, L., Bain, A., and F. Kelly, "Fair Internet traffic integration: network flow models and analysis", Annales des Telecommunications 59 pp1338--1352, 2004, <<http://citeseer.ist.psu.edu/641158.html>>.
- [FrRideP2p] Feldman, M., Papadimitriou, C., Chuang, J., and I. Stoica, "FreeRiding and Whitewashing in Peer-to-Peer Systems", Proc. Workshop on Practice and Theory of Incentives in Networked Systems (PINS'04) pp228--236, 2004, <<http://doi.acm.org/10.1145/1016527.1016539>>.
- [FundUtil] Shenker, S., "Fundamental Design Issues for the Future Internet", IEEE Journal on Selected Areas in Communications 13(7)1176--1188, 1995, <<http://citeseer.ist.psu.edu/shenker95fundamental.html>>.
- [I-D.irtf-iccr-g-cc-rfcs] Welzl, M. and W. Eddy, "Congestion Control in the RFC Series", draft-irtf-iccr-g-cc-rfcs-00 (work in progress), October 2006.

- [Jstce] Eriksson, J., Faloutsos, M., and S. Krishnamurthy, "Justice: Flexible and Enforceable Per-Source Bandwidth Allocation", Proc. Networking pp1206--1218, 2005, <[http://www.cs.ucr.edu/~krish/jakob\\_networking05.pdf](http://www.cs.ucr.edu/~krish/jakob_networking05.pdf)>.
- [MultTCP] Crowcroft, J. and Ph. Oechslin, "Differentiated End to End Internet Services using a Weighted Proportional Fair Sharing TCP", CCR 28(3) 53--69, July 1998, <<http://www.cs.ucl.ac.uk/staff/J.Crowcroft/hipparch/pricing.html>>.
- [NewArchReq] Braden, R., Clark, D., Shenker, S., and J. Wroclawski, "Developing a Next-Generation Internet Architecture", DARPA white paper , July 2000, <<http://www.isi.edu/newarch/DOCUMENTS/WhitePaper.pdf>>.
- [PCNcharter] IESG Secretariat, "WG Review: Congestion and Pre-Congestion Notification (pcn)", Email archive , Feb 2007, <<http://www1.ietf.org/mail-archive/web/ietf-announce/current/msg03418.html>>.
- [PMP] Odlyzko, A., "A modest proposal for preventing Internet congestion", AT&T technical report TR 97.35.1, September 1997, <<http://www.dtc.umn.edu/~odlyzko/doc/modest.proposal.pdf>>.
- [PrCong] MacKie-Mason, J. and H. Varian, "Pricing Congestible Network Resources", IEEE Journal on Selected Areas in Communications, `Advances in the Fundamentals of Networking' 13(7)1141--1149, 1995, <<http://www.sims.berkeley.edu/~hal/Papers/pricing-congestible.pdf>>.
- [RFC0970] Nagle, J., "On packet switches with infinite storage", RFC 970, December 1985.
- [RFC1122] Braden, R., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, October 1989.
- [RFC1254] Mankin, A. and K. Ramakrishnan, "Gateway Congestion Control Survey", RFC 1254, August 1991.
- [RFC2581] Allman, M., Paxson, V., and W. Stevens, "TCP Congestion Control", RFC 2581, April 1999.
- [RFC2616] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext

Transfer Protocol -- HTTP/1.1", RFC 2616, June 1999.

- [RFC2960] Stewart, R., Xie, Q., Morneault, K., Sharp, C., Schwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Zhang, L., and V. Paxson, "Stream Control Transmission Protocol", RFC 2960, October 2000.
- [RFC3124] Balakrishnan, H. and S. Seshan, "The Congestion Manager", RFC 3124, June 2001.
- [RFC3426] Floyd, S., "General Architectural and Policy Considerations", RFC 3426, November 2002.
- [RFC3448] Handley, M., Floyd, S., Padhye, J., and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification", RFC 3448, January 2003.
- [RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.
- [RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.
- [RFC3714] Floyd, S. and J. Kempf, "IAB Concerns Regarding Congestion Control for Voice Traffic in the Internet", RFC 3714, March 2004.
- [RFC4341] Floyd, S. and E. Kohler, "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 2: TCP-like Congestion Control", RFC 4341, March 2006.
- [RFC4342] Floyd, S., Kohler, E., and J. Padhye, "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC)", RFC 4342, March 2006.
- [RFC4585] Ott, J., Wenger, S., Sato, N., Burmeister, C., and J. Rey, "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)", RFC 4585, July 2006.
- [Re-TCP] Briscoe, B., Jacquet, A., Salvatori, A., and M. Koyabi, "Re-ECN: Adding Accountability for Causing Congestion to TCP/IP", draft-briscoe-tsvwg-re-ecn-tcp-02 (work in progress), June 2006.

- [Re-fb] Briscoe, B., Jacquet, A., Di Cairano-Gilfedder, C., Salvatori, A., Soppera, A., and M. Koyabe, "Policing Congestion Response in an Internetwork Using Re-Feedback", ACM SIGCOMM CCR 35(4)277--288, August 2005, <<http://www.acm.org/sigs/sigcomm/sigcomm2005/techprog.html#session8>>.
- [Saul05] Saul, J., "The Collapse of Globalism and the Reinvention of the World", Pub: Viking, Canada ISBN: 0-670-06367-3, 2005.
- [SelfMan] Kelly, F., "Models for a Self-Managed Internet", Philosophical Transactions of the Royal Society 358(1773)2335--2348, August 2000, <<http://www.statslab.cam.ac.uk/~frank/smi.html>>.
- [SovJstce] Siderenko, S., "Characteristics of Perceptions of Social Justice in the Contemporary USSR", Chapter in: Transitional Agendas: Working Papers from the Summer School for Soviet Sociologists Ch3 pp41--45, 1991, <<http://lucy.ukc.ac.uk/csacpub/russian/siderenko.html>>.
- [Swed90] Swedberg, R., "Economics and Sociology. Redefining their Boundaries: Conversations with Economists and Sociologists", Pub: Princeton University Press , 1990.
- [TCPcc] Jacobson, V., "Congestion Avoidance and Control", Proc. ACM SIGCOMM'88 Symposium, Computer Communication Review 18(4)314--329, August 1988, <<ftp://ftp.ee.lbl.gov/papers/congavoid.ps.Z>>.
- [Tussle] Clark, D., Sollins, K., Wroclawski, J., and R. Braden, "Tussle in Cyberspace: Defining Tomorrow's Internet", ACM SIGCOMM CCR 32(4)347--356, October 2002, <<http://www.acm.org/sigcomm/sigcomm2002/papers/tussle.pdf>>.
- [UtilFair] Mazumdar, R., Mason, L., and C. Douligeris, "Fairness in Network Optimal Flow Control: Optimality of Product Forms", IEEE Transactions on Communications 39(5)775--782, May 1991, <<http://dionysos.cs.unipi.gr/~cdoulig/Fairness%20in%20network%20optimal%20flow%20control%20optimality%20of%20product%20forms.pdf>>.
- [UtilFile] Key, P. and L. Massoulie, "User policies in a network

implementing congestion pricing", Proc. Workshop on Internet Service Quality and Economics , December 1999, <<http://www.marengoresearch.com/isqe/index.htm>>.

- [WFQ] Demers, A., Keshav, S., and S. Shenker, "Analysis and Simulation of a Fair-Queueing Algorithm", ACM SIGCOMM CCR 19(4)1--12, September 1989, <<http://portal.acm.org/citation.cfm?id=75248>>.
- [WindowPropFair] Siris, V., "Service Differentiation and Performance of Weighted Window-Based Congestion Control and Packet Marking Algorithms in ECN Networks", Computer Communications 26(4) 314--326, 2002, <[http://www.ics.forth.gr/netgroup/publications/weighted\\_window\\_control.html](http://www.ics.forth.gr/netgroup/publications/weighted_window_control.html)>.
- [XCHOKe] Chhabra, P., Chuig, S., Goel, A., John, A., Kumar, A., Saran, H., and R. Shorey, "XCHOKe: Malicious Source Control for Congestion Avoidance at Internet Gateways", Proceedings of IEEE International Conference on Network Protocols (ICNP-02) , November 2002, <<http://www.cc.gatech.edu/~akumar/xchoke.pdf>>.
- [XCP] Katabi, D., Handley, M., and C. Rohrs, "Congestion Control for High Bandwidth-Delay Product Networks", ACM SIGCOMM CCR 32(4)89--102, October 2002, <<http://www.ana.lcs.mit.edu/dina/XCP/>>.
- [ccFairTut] Le Boudec, J-Y., "Rate Adaptation, Congestion Control and Fairness: A Tutorial", Web page , November 2005, <[http://icalwww.epfl.ch/PS\\_files/LEB3132.pdf](http://icalwww.epfl.ch/PS_files/LEB3132.pdf)>.
- [pBox] Floyd, S. and K. Fall, "Promoting the Use of End-to-End Congestion Control in the Internet", IEEE/ACM Transactions on Networking 7(4) 458--472, August 1999, <<http://www.aciri.org/floyd/end2end-paper.html>>.
- [wPropFair] Kelly, F., "Charging and Rate Control for Elastic Traffic", European Transactions on Telecommunications 8 pp33--37, 1997, <<http://www.statslab.cam.ac.uk/~frank/elastic.html>>.
- [wPropStab] Kelly, F., Maulloo, A., and D. Tan, "Rate control for communication networks: shadow prices, proportional

fairness and stability", Journal of the Operational  
Research Society 49(3) 237--252, 1998,  
<<http://www.statslab.cam.ac.uk/~frank/rate.html>>.

Author's Address

Bob Briscoe  
BT & UCL  
B54/77, Adastral Park  
Martlesham Heath  
Ipswich IP5 3RE  
UK

Phone: +44 1473 645196

Email: [bob.briscoe@bt.com](mailto:bob.briscoe@bt.com)

URI: <http://www.cs.ucl.ac.uk/staff/B.Briscoe/>

## Full Copyright Statement

Copyright (C) The IETF Trust (2007).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

## Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at [ietf-ipr@ietf.org](mailto:ietf-ipr@ietf.org).

## Acknowledgments

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA). This document was produced using xml2rfc v1.32 (of <http://xml.resource.org/>) from a source in RFC-2629 XML format.