The Evolution of Internet Congestion

Steven Bauer[1], David Clark[2], William Lehr[3]

Massachusetts Institute of Technology

Abstract

This paper discusses the evolution of the congestion controls that govern all Internet traffic. In particular we chronicle and discuss the implications of the fact that the most significant "congestion signals" are increasingly coming from network operators, not the TCP stack. Providers now nudge users into different traffic patterns using a variety of new technical and non-technical means. These 'network-based congestion management' techniques include volume-based limits and active traffic management of best effort traffic. The goal and effect of these techniques differs from the historically coveted flow-rate fairness of TCP, provoking some in the technical and policy community to question the appropriateness of such deviations, and feeding debates over network management and network neutrality. To appropriately evaluate emerging trends in congestion control, it is useful to understand how congestion control has evolved in the Internet, both with respect to its intellectual history within the technical community and with respect to the changing traffic/industry environment in which the Internet operates. This is important because a sophisticated and nuanced view of congestion and its management is necessary for good public policy in this space. We conclude the paper with a discussion of the policy implications and questions raised by the continual evolution of congestion.

1. Introduction

The phenomenal growth of the Internet in terms of the volume of economic activity and traffic it carries over the past fifty years represents a remarkable example of the scalability of the Internet architecture, which in spite of the growth, remains in many respects, fundamentally unchanged. One of the most enduring features of the Internet since the late 1980s has been reliance on TCP's flow-rate fairness, as implemented using Van Jacobson's famous algorithm.[4] The Transmission Control Protocol (TCP) is one of the key Internet protocols and is responsible for managing end-to-end connections across the Internet. Since that time and still today, TCP remains one of the most important congestion control mechanisms in use in the Internet. In light of its success in supporting

[1] Corresponding author: bauer@mit.edu, tel: 617-869-5208.

[2] ddc@csail.mit.edu tel: 617-253-6002

[3] wlehr@mit.edu, tel: 617-258-0630

[4] As we discuss further below, in 1987, Van Jacobson implemented a series of "congestion" control mechanisms that collectively govern how end-hosts adjust their traffic sending rate in response to perceived congestion.

the immense growth in Internet traffic and infrastructure investment, and in the face of the substantial changes in market/industry structure and regulatory policy that have changed the Internet's role in the economy so dramatically since its introduction, it is not surprising that any thought of moving away from TCP "flow-rate" fairness might raise concern.[5]

Given the importance of these issues, it is useful to consider how thinking about congestion control has evolved within the technical community. To some, it may seem that the mere fact that "congestion" occurs is somehow bad, and evidence of a problem that should be addressed. To understand that such a position is naïve, it is useful to remember that Internet congestion is a direct result of the resource "sharing" that is central to notions of why the Internet is so valuable and has been so successful. The Internet provides a shared resource platform that supports interoperability and interconnection for diverse types of applications across heterogeneous networking infrastructures (high and low speed, wired and wireless) on a global basis. When resources are shared, there is the potential that demand for those resources may exceed available supply, requiring some sort of allocation mechanism to address the imbalance and determine which resources get served first. Considered in its broadest context, such allocation mechanisms may be thought of as "congestion control."

In the balance of this paper, we examine the history of congestion control within the Internet technical community, and set it within the context of the changing traffic and industry environment in which Internet congestion is evolving. Briefly, we argue that the growth of the Internet in terms of the diversity of users/uses it is called upon to support, the sheer volume of traffic involved, and the economic value associated with this activity[6] have contributed to making the congestion problem increasingly complex. The emergence of public policy debates over "Network Neutrality" and ISP traffic management practices[7] underscore the wider stakeholder context in which congestion control is now being discussed. What is less well appreciated, perhaps, is the extent to which this complexity is mirrored in the technical challenges posed by the new environment. By highlighting some of these technical issues, we wish to suggest the

[5] See Briscoe (2007), Floyd and Allman (2008), and Mathis (2009) for discussions in the technical community offering divergent perspectives on congestion control in the Internet.

[6] Both in terms of the value users derive from Internet usage and the value of the infrastructure and related resources – including investment – that are required to sustain it.

[7] "Network Neutrality" is the short-hand name used to describe the debate over the need for explicit regulatory controls over how Internet Service Providers (ISPs) manage broadband Internet traffic. The debate was sparked by a paper from Wu (2003) that suggested consideration of a norm limiting broadband discrimination as useful approach to protecting the openness of the Internet. This has subsequently spawned a huge academic and public policy literature. See Lehr, Peha & Wilkie (2007) for a collection of perspectives on this debate. More recently, regulatory authorities have initiated proceedings to examine ISP traffic management practices (e.g., in Canada see http://www.crtc.gc.ca/ENG/archive/2008/pt2008-19.htm, November 2008; and in the U.S. see http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-08-92A1.pdf, January 2008).

value to be gained from further experimentation over how to best collectively evolve congestion control practices for the future health of the Internet.

TCP-based congestion control (as we explain further below) is implemented on the end-hosts at the edge of the Internet and moderates individual flows of traffic.[8] It remains the principal mechanism for managing traffic congestion on the best-effort Internet over short time periods. The transition to the broadband Internet has brought an age of potentially very high and variable data rate demands from edge nodes to support all of the possible activity that may originate across the access link from a broadband connected home or business. When combined with other factors (such as changing expectations of what users ought to be able to do over the Internet), these changes may suggest a greater potential role for network operators in managing Internet congestion over time scales that are much shorter than those associated with re-configuring the network or investing in new capacity.[9] Network operators have introduced a variety of new technical and non-technical strategies for managing short and medium term congestion of best-effort traffic. These traffic management techniques include 1) volume caps that limit the total volume of traffic over different durations of times and in the upstream and or downstream directions, 2) prioritizing subscriber or application traffic based upon factors such as the amount of traffic sent during congested periods or assumptions regarding what subscribers would prefer to be prioritized (such as voice traffic over bulk transfers) and 3) rate limiting traffic classes, such as peer-to-peer traffic, that are believed to significantly contribute to congestion.

We recognize that a public discussion over the implications of changing mechanisms for congestion control is important for the health of the Internet and, potentially, for protecting the openness of the Internet; however, we also believe it would be premature to conclude that we know what the best mechanisms for congestion control are. In this paper, we do not wish to opine on the merits of particular strategies, however our assessment of the present situation leads us to conclude that it would be undesirable to adopt regulatory controls that might serve to effectively enshrine TCP's flow-based fairness as a policy goal at the very time that the technical community is re-evaluating it. In the present paper, we principally wish to help educate the broader Internet community about the evolving technical debates over congestion control, and to suggest research and policy directions for ensuring that the on-going investigation of these issues remains as productive as possible.

The balance of the paper is organized as follows: In Section 2, we discuss the roots the intellectual discussion within the technical community about resource sharing and congestion as two-sides of the same coin. In Section 3, we consider the multiplicity of perspectives that may be used to define what constitutes "congestion." In Section 4 we

[8] TCP only regulates the sending rate for a single flow. It does not keep track of how much data has been sent via the flow over time or whether there are multiple TCP flows running on a single end-node. TCP does not regulate the aggregate demand for network resources from a host except in the sense that each individual TCP flow responds to congestion.

[9] Any role for network operators would complement, not replace, TCP congestion control.

review how Internet traffic, the user environment, and broadband infrastructure have evolved and the implications of these changes for congestion management. In Section 5 we return to the discussion of the intellectual history of explicit congestion control mechanism and delve into greater detail on the implications of TCP flow-based control and the responses of network operators, including ISPs, toward implementing new techniques for managing congestion. In Section 6, we discuss some of the policy implications and questions that these responses are raising. Section 7 concludes.

2. Resource Sharing and Congestion

This paper could have been titled "The Evolution of Internet Sharing". It is because of sharing that congestion exists. One is not possible without the other. The origin of "congestion" in the community's lexicon is interesting because the Internet protocols were largely devised by researchers who started out with a background in operating systems.[10] "Congestion" is not a common part of that community's nomenclature. Instead the operating system terminology focuses on sharing -- "time sharing", "shared memory," "shared libraries," "shared files," et cetera. The very definition of an operating system includes sharing – "an OS is responsible for the management and coordination of activities and the sharing of the resources of the computer." [11] One does not typically talk about a computer becoming congested even though clearly performance bottlenecks are a frequently problem that arises from contention for scarce resources.

Similarly many of the initial documents that provided the intellectual basis for what would become the Internet were infused with the language of sharing. Comparatively little space was devoted to what would occur if demand exceeded supply.[12] The desire to share resources was a central motivation of J.C.R. Licklider, Ivan Sutherland, Bob Taylor and Lawrence Roberts at the Advanced Research Projects Agency (ARPA) at the time they were putting together the network that would evolve into the Internet. They were all convinced of the potential to dramatically lower the cost of computing by networking the time-sharing systems that existed at many universities. The proposed network would allow ARPA-sponsored researchers to share the computers that ARPA was funding and which were in diverse locations.

The origin of "congestion" as a central problem in the Internet may be traced to the work of Leonard Kleinrock.[13] He published the first paper on packet switching theory in July 1961.While he credits the earlier work of Fry[14] in 1928 as offering a unifying view of the previous works on congestion and the work of Feller[15] in 1939 as ushering in the

[10] See Lyon (2004).

[11] See Stallings (2008).

[12] See Licklider (1963).

[13] See Leiner *et al.* (1997).

[14] See Fry (1928).

[15] See Feller (1939).

"modern theory of congestion," Kleinrock was the first to focus on the problems of congestion in large multi-node networks with queuing.

In addition to providing the initial framing of these central questions, Kleinrock established an analytic basis for answering congestion questions. His work is credited with convincing Larry Roberts at ARPA of the theoretical feasibility of communications using packets rather than circuits as a basis for networking computing systems.[16] This was significant because some of the early objections to the feasibility of the Internet centered upon a perceived need for very large packet buffers to handle the uncontrolled loads from end systems.

Since that time, congestion -- congestion control, congestion research, congestion management – has been a central focus of network researchers, protocol standards development, equipment manufacturers, software developers, and network operators. It is this evolution of congestion that this paper examines.

Historically, the sending and receiving rate of applications was primarily governed by TCP congestion control. This is an algorithmic mechanism implemented by the operating systems of senders and receivers that continually probes the network by gradually increasing the sending rate until a packet loss is detected. When a packet loss is detected an inference is made that the loss occurred because of congestion. The sending rate is cut in half and the cycle repeats. This is referred to as a "flow-based" control mechanism because it is implemented on the sending and receiving host or end-nodes of each TCP session or flow. By limiting the traffic offered by the sender when it detects congestion, TCP reduces the load offered to "the network" associated with that flow, thereby alleviating the excess demand condition that results in congestion somewhere downstream in the network. TCP does not know the source or location where packets were dropped, only that losses occurred somewhere along the path from sender to receiver. The distributed implementation by all of the TCP flows collectively is intended to result in sufficient reduction in the rate of traffic being delivered to the network to reduce excess demand for network resources that is the cause of the network being congested – automatically, without requiring explicit intervention or action by network operators.

Of course, the technical details are somewhat more complicated. Over the years, TCP has evolved as various tweaks have been made to improve performance. Indeed, previous examinations exploring the evolution of congestion have focused almost exclusively on the changes to TCP.[17]

As noted earlier, the fundamental cause of the congestion state is too much demand from the collection of users that share the network resources. The pattern of demand and the capacity, architecture, and management of network resources all contribute to

[16] See Leiner *et al.* (1997).

[17] See Peterson (2007). An exception is explicit congestion notification (ECN). ECN has been standardized but is not widely deployed.

determining when, how, and where a network enters a congestion state. Because network operators by their investments and marketing decisions determine the available capacity (supply of network resources) and limit potential per subscriber user demand (by the capacity of the access links ISPs make available), network operators have always played an important role in managing Internet congestion over the medium and long-term. Indeed, a common approach to managing resource sharing is to provision for expected peak demand over some time period, and because many network investments need to be made in relatively large fixed increments and over an investment time horizon that takes months or more, capacity is provisioned in advance of realized demand. Thus, during off-peak periods (which may be measured in periods of hours or days) and over the life of infrastructure investments (which may be measured in periods of months or years), there may be significant amounts of time when the network is over-provisioned relative to offered demand. During such periods, the network may appear to be relatively un-congested.

However, because demand is not smooth and fluctuates stochastically over time at many different time-scales and because the available capacity the Internet varies across the network, congestion events may arise commonly even in a network that may be considered to be generally "over-provisioned."

The new role for network operators that we focus on in this paper is not the role they have and continue to play with respect to the longer time periods over which network provisioning and investment decisions are made, but over shorter time periods that are closer to the short-time focus of TCP congestion control (see Figure 1). In particular, time scales on which these network-based controls often operate are fifteen minutes, one day, or one month. While at the sub-second level the relatively myopic TCP congestion control governs the sending and receiving rates, these new control mechanisms govern the intermediate term contribution of senders and receivers to network congestion.

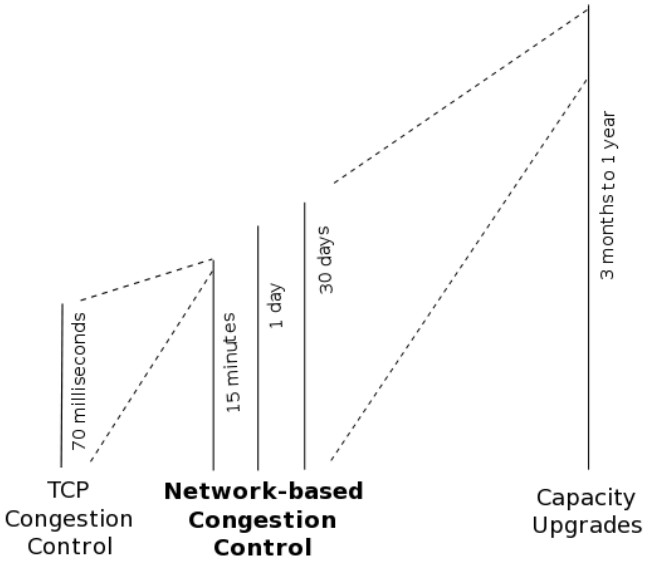

Figure 1: The most significant signals of congestion are now coming from network operators[18] not the TCP stack on the host computer systems that continually probe to discover the current congestion thresholds.

These mechanisms potentially change users' relative share of capacity at bottlenecks in the network from what would have occurred as a result of the distributed computation of end-hosts' TCP algorithms. Instead of achieving an allocation of capacity that is considered "TCP fair" (which we describe in more detail later) a different allocation of resources is produced. How it differs, obviously depends on what is done.

Raising the possibility of the Internet community moving to a different notion of how resources ought to be allocated during periods of congestion at these shorter time scales is a contentious issue. This is, in part, because TCP "flow-based fairness" has been effective for such a long time in protecting the Internet from congestion collapse. Also, the existing TCP mechanism has been vigorously investigated and discussed within the academic and technical community, and is effectively enshrined in standards that have been approved by a diverse community of stakeholders in the IETF.[19] One perspective that appeals to many engineers as well as non-engineers is the old maxim "if it isn't broke, don't fix it!"

To investigate whether the environment to which TCP was designed as a solution has changed sufficiently to warrant a re-examination of TCP fairness and to evaluate the new congestion control strategies being employed by network operators to manage congestion on shorter time scales such as rate limiting applications, volume caps, or selective traffic

[18] Note that by network operators we are not limiting the discussion to for-profit commercial network providers. Non-profit network operators, such as at Universities, sometimes employ network-based congestion controls as well.

[19] As noted early, proposals to refine or tweak TCP and discussions of the relative merits of such reforms have a long history in the technical research literature and in standardization proceedings.

prioritization, it is useful to define what constitutes a "congestion" event more carefully and how this may be changing with the changing Internet.

3. Defining Congestion

If you were to ask a diverse set of technical experts if they could identify the periods of time the network depicted in Figure 2 was congested given the link capacities, queue sizes, and a complete trace of all traffic carried by the network almost all would answer "yes." However, they would not agree on the answer.

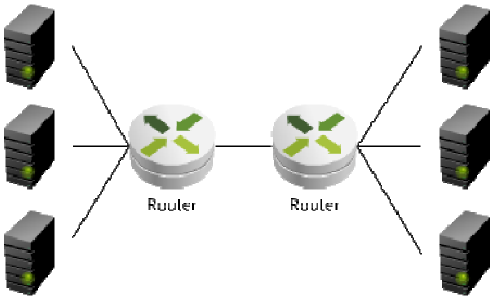

Figure 2: Given the link capacities, queue sizes, and a complete trace of all network traffic for this simple network, different congestion periods could be identified using different definitions of congestion.

Identifying when a network is congested depends upon the definition of congestion one employs. Loosely speaking, everyone would agree that "congestion" is the state of network overload. However, this is not a precise definition adequate for characterizing exactly when or for how long a network is congested.

More precise, but different, definitions are supplied by various authors and sub-disciplines in networking. Each uses the term "congestion" to describe different (but related) phenomena. Each of these meanings of congestion is useful in its own right. But regrettably the particular definition is not always clear in discussions. In the subsections that follow we explore different meanings of congestion.

3.1. Queuing theory definition

In queuing theory, traffic congestion is said to occur if the arrival rate into a system exceeds the service rate of the system at a point in time.[20] This is typically expressed as formula 1 in the sidebar. Note that traffic congestion is not just a binary proposition using this definition. Traffic

1:arrival rate > service rate

$$\lambda > \mu$$

2: congestion = arrival rate / e

$$p = \frac{\lambda}{\mu}$$

[20] See Gross (1998).

congestion, or traffic intensity, can be measured as the ratio of the arrival rate to service rate. Using this definition we could precisely account for the periods of time in which a given network resource is congested and even characterize the intensity of congestion (formula 2 in the sidebar). It is precisely the periods of time in which more network traffic has arrived then has been sent (i.e., all t for which p(t)>1).

If the arrival rate persistently exceeds the service rate of the system, the queue of traffic will grow without bound – there will be no steady state behavior of the system. The system will always be congested with service times getting longer and longer. On the other hand, if the arrival rate exceeds the service rate only occasionally, then congestion will be a transient phenomenon. Queues will build, but will eventually clear in the system.

Note that by this definition, if the input rate equals the output rate, the network is not considered congested. It says nothing about the steady state size of the network's queues. If during a previous period of congestion a backlog queue of traffic built up, and traffic then arrived and departed at precisely the same rate the backlog of packets would not grow but it would not shrink either. Admittedly such a perfectly balanced system is not likely to persist for long in the real world. But quickly draining the queues is considered to be a central requirement of "good" congestion control (since it makes the queues shorter to handle future transients when p(t)>1).This illustrates that there are other facets to congestion control than just keeping the average input rate below the average output rate.

3.2. Networking text book definition

In contrast, a popular academic textbook on networking[21] defines the buildup of packets in a queue not as "congestion" but rather as "contention." "Congestion," according to this textbook, is restricted to the situation in which a switch or router has so many packets queued for transmission that it runs out of buffer space and must start dropping packets if more arrive. A buffer must be filled to capacity for congestion to occur. (See Figure 3.)

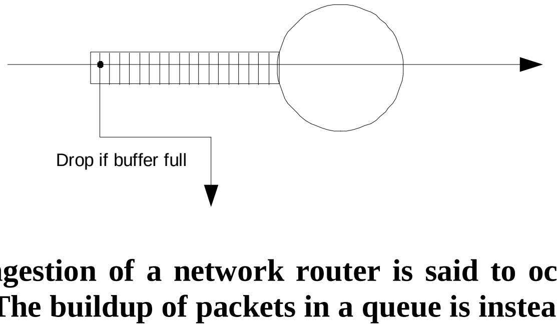

Figure 3: Congestion of a network router is said to occur if packets are dropped. The buildup of packets in a queue is instead described as "contention."

[21] See Peterson (2007).

The queuing theory community does not make a distinction between "congestion" and "contention." According to their definition, as soon as a queue starts to build traffic congestion is occurring. "Contention" would imply "congestion" to queuing theorists.

These different definitions imply a difference in how congestion is measured. According to the packet-dropped definition, congestion could be viewed as a binary phenomenon: either a switch or router is dropping packets or it isn't. If someone asked for a more quantitative measure of congestion one might reply with the number of dropped packets or the rate of dropped packets over some period of time. The measure of traffic congestion according to the queuing theory definition, on the other hand, is a unit-less ratio of arrival rate over service rate. It is a measure of intensity.

Note that a dropped packet is the only definition of congestion that is relevant to the TCP algorithm.TCP, and hence end users, do not react to any other definition of congestion. So we might also have called this the TCP definition of congestion.

3.3. A network operator's definition of congestion

Network operators tend to adopt another definition of congestion. They define "congestion" in terms of the load on a network over a particular period of time. (See Figure 4.) Ask a network operator how "congested" part of their network is and they will respond with the average utilization of a link over some period of time. This period of time is typically on the order of minutes or longer.

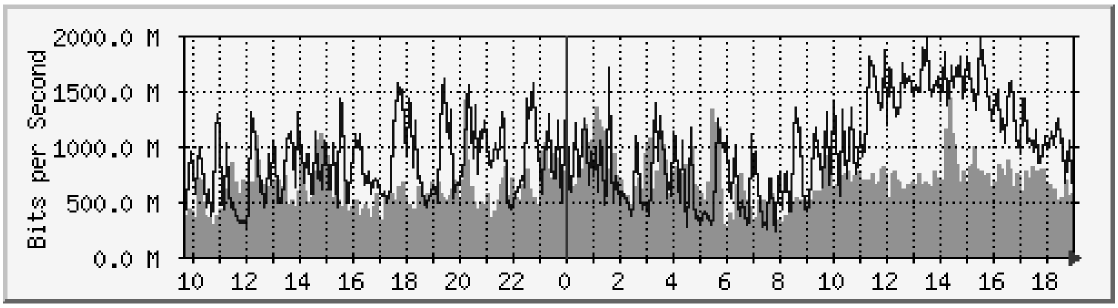

Figure 4: The figure above is a sample of a "Multi Router Traffic Graph" that is very common in the operational community. It shows the averaged utilization of a single link in both the inbound (black) and outbound (gray) direction over the course of multiple hours.

Comcast, in a filing with the FCC[22] describing their congestion management techniques, defined a "near congestion state" in the following manner:

> *For a CMTS port to enter the Near Congestion State, traffic flowing to or from that CMTS port must exceed a specified level (the "Port Utilization*

[22] See Comcast (2008).

> *Threshold") for a specific period of time (the "Port Utilization duration"). The Port Utilization Threshold on a CMTS port is measured as a percentage of the total aggregate upstream or downstream bandwidth for the particular port during the relevant timeframe. The Port Utilization Duration on the CMTS is measured in minutes.*

Comcast considers a port in their network to be "near congestion" if

> *"[O]ver any 15-minute period, if an average of more than 70 percent of a port's upstream bandwidth capacity or more than 80 percent of a port's downstream bandwidth capacity is utilized, that port will be determined to be in a Near Congestion State."*

A network might intermittently be "congested" by the queuing theory or networking textbook definitions of congestion and "not congested" or "not near congestion" from the network operator's perspective. This would occur for instance if there is a temporary spike in traffic that causes queues to build or overflow dropping and delaying a small number of packets while the longer term average remains below the congestion threshold.

Conversely it could also be the case that even though the utilization of a link exceeds a network provider's threshold, 70 or 80 percent in Comcast's case, the queues feeding that link could all be empty (as packets are sent on without delay upon arrival). In other words, the output link might be adequately servicing the demand from all the input links.

3.4. An economic definition of congestion

The field of economics provides yet another perspective on the definition of congestion. From the MIT Dictionary of Modern Economics[23]:

> *"When an increase in the use of a facility or service which is used by a number of people would impose a cost (not necessarily a monetary cost) on the existing users, that facility is said to be 'congested'.*

This definition is quite general,[24] and may be interpreted to apply more broadly than to the notion of TCP congestion control which is focused on whether packets are being queued (e.g., because the send-rate has been reduced in response to perceived congestion) or dropped. However, even if we apply this definition narrowly to the sort of situations anticipated by the queuing or network engineering text book definitions, it is clear that by

[23] See Pearce (1992)

[24] For, example, one's choice of what is identified as the "facility", the "people", or their "use" will vary with the problem under consideration (e.g., the relevant time scale for congestion as discussed previously). In the present context, the most natural interpretation of the "cost" is the cost of the added delay realized by end-users because of lengthening traffic queues or dropped packets. For example, one might include among the "costs" modifications in behavior or investments made in anticipation of congestion by end-users or network operators.

this economic definition a network switch or router could be "congested" even **before** a queue starts to build and before packets start to be dropped. This is in contrast to the queuing theory and networking textbook definitions of congestion that entail that a queue is building or packets are already being dropped. A switch or router would be considered "congested" by the economic definition if it is on the "knife's edge" where a marginal increase in the arrival rates of packets would cause the queues to start building.

A key difference between this economic definition and the other definitions is that the economic definition implies sharing among different economic entities so that cost is shifted from one user to another. The queuing theory definition just requires that offered load exceed capacity, and does not discriminate between one user and many users overloading a link.

Is this economic definition of congestion close to the network operators' definition based upon usage levels? The answer to this depends upon how one interprets the size of the increase in load needed to cause the delays or drops and the time period in which this increase has to occur. If one believes that traffic might increase by 25% in the next six months and the network is currently at 80% utilization then a network might be deemed congested now since the projected increase will assuredly cause delays or drops over that longer duration time period.[25]

3.5. Summary of congestion definitions

The preceding discussion highlights the danger in presuming that there is a single correct definition for what constitutes congestion, even in the technical language of researchers and network operators. When the economic complexities of resource allocation and the direct and indirect cost allocation implications of such decisions are added, it is not surprising that discussions about what the right way to manage congestion become contentious.

The preceding discussion should also make clear that whether one observes congestion as occurring will depend on the time-scale over which one observes traffic and the network elements (or potential bottlenecks) over which that observation is undertaken.

For example, the simple queuing and network definitions of congestion were applied to a general queue or network. Any particular network will consist of multiple links and routing nodes that may be congested at different times and under different configurations of traffic loadings. One could focus on a definition of the sort that said: "this flow is congested if it experiences packet losses" or "this network is congested if any of the

[25] There might also be an opportunity cost born by a network operator if the network is loaded to a level that precludes them from accepting or attracting new subscribers. Thus, a network might be perceived to be "congested" if it is near the point where additional traffic would trigger a business decision to deploy additional capacity. Historically, some network operators adopted a rule-of-thumb that additional capacity would be deployed if the average link utilization during the busy period of the day exceeded 50% of the link capacity.

flows experience packet losses," et cetera. Knowing that some flow somewhere at some time may have experienced congestion-related packet losses provides little useful information, by itself, regarding what the technical or economic consequences of the congestion event may have been. On other hand, more meaningful and detailed characterizations of congestion must perforce engage measurement details that can make generalized discussion more difficult. The appropriate measurement or characterization depends on the context in which congestion is to be managed. There is no "one-size" fits all solution.

4. Evolution of Internet Traffic and Congestion

Loosely speaking, the Internet has gone through three phases: first, the period before the Internet became a mass market data platform (pre-1990s); second, the 1990s when the Internet was growing rapidly via dial-up broadband access into a mass market platform; and third, since 2000, as the Internet transforms into the broadband Internet.

During the early history of the Internet, the long distance links were the primary congestion points in the network. The congested links were things like 56.6 kb/s leased lines. Computers in the local area network were comparatively uncongested with higher speed connections transferring data at 3 Mb/s over the early Ethernet networks.[26]At the time, most of the traffic was associated with applications like email, bulk file transfers, or low bit rate interactive sessions that were tolerant of the Internet's best effort service and the variable packet delays that sometimes resulted during periods of congestion. The users of the network were relatively sophisticated academic, government, and commercial researchers. Congestion was tolerated and seen as a shared problem to be jointly addressed. There were no expectations, for instance, that the early experiments to transfer live voice over the Internet would be a viable substitute for a regular telephone.

The commercialization of the Internet during the dialup era changed this markedly. For the first time customers were paying for a service and developed more demanding expectations about the quality of the network connections they were purchasing. While the Internet remained a "best effort" network, if that best effort was seen as not good enough, the majority of new users were not sufficiently sophisticated to try and debug the network problems themselves; rather they called their ISP's customer service line.

During the dial-up era of the Internet, most users were connected to the Internet via dial-up models with a maximum rate of 56.6 kb/s (really lower in practice). Moreover, these connections were intermittent (i.e., users were not continuously connected to the Internet). Although the Internet's rapid growth as a mass market service during the 1990s attests to the significant value consumers placed on Internet access even at these low data rates, many of the services that have emerged as dominant consumers of bandwidth would were simply not viable at such low data rates. This provided a natural suppression in per-user demand, and the maximum data rate imposed by the limits of dial-up

[26] See Crowcroft (2003).

modems, provided a hard cap on the amount of traffic any single mass-market subscriber could deliver to or receive from the network.

In this environment, when Internet congestion arose, it most commonly arose in the dial-up access link. The most common network resource facility that was constrained in this period was the data modem banks. When too many subscribers attempted to connect, excess subscribers received a busy signal. This is akin to the way that congestion was managed in traditional circuit-switched telephone networks. At least for a while, this scarce network resource carried a price. Popular dialup providers like AOL had service plans offering various numbers of hours per month for a flat fee, with additional hours for an incremental per hour charge. For example, subscribers could connect for $19.95 for 20 hours and $2.95 per each additional hour under AOL plan in 1996.[27] Eventually, unlimited dial-up access for a flat monthly fee proved a much more attractive service for consumers and most subscribers in the U.S. migrated to flat rate plans. Once such plans became common, per-user traffic (number of hours on-line) and the number of subscribers grew rapidly, resulting in rapid growth in aggregate traffic. However, the primary bottleneck remained the modem link speeds.[28]

Broadband Internet access offering 100s or even 1,000s of Kb/s data rates and "always on" connections started becoming available in the mid-1990s, but mass market demand for broadband Internet access did not really begin to take off until after 2000. The typical technologies for offering these first-generation broadband services were DSL (over telephone company lines) and cable modems (over cable television provider networks). Figure 5 presents the trends in home Internet access.

[27] Notice that this is an inverted price which penalized callers who stay connected for long periods of time (longer than 20 hours per month) by charging them a higher price per hour.

[28] For example see http://www.timewarner.com/corp/newsroom/pr/0,20812,669989,00.html which notes that since 1995 AOL implemented compression technologies to minimize the number of bytes transferred over the dialup lines. These compression techniques wouldn't have been effective at improving performance if the primary bottleneck was elsewhere in the network.

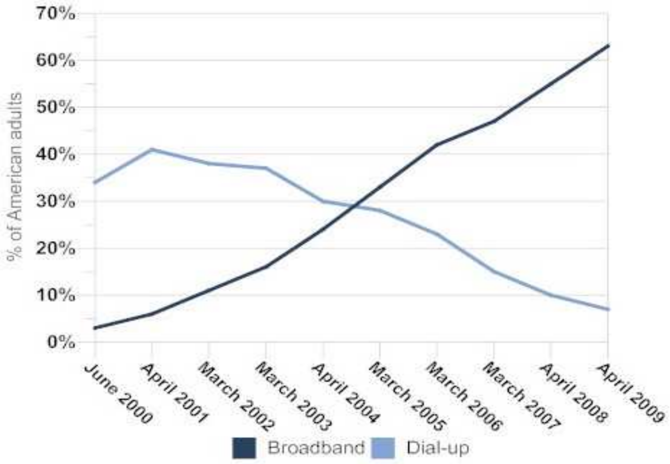

Figure 5: Trends in home Internet access: broadband vs. dialup: the percentage of adults who have broadband or dialup, 2000-2009. Source: Pew Internet & American Life Project Surveys

The transition to broadband represented a seismic shift in the nature of the congestion problem. First, broadband makes it attractive from the perspective of the user's experience to use much more rich media-intensive, bandwidth hungry applications like streaming video, peer-to-peer (p2p) file sharing for large files (mostly music and movies), and interactive gaming. Elimination of the dialup bottleneck was matched by complementary investments upstream (in more powerful multimedia-capable PCs in the home and home networks capable of supporting multiple users on a single broadband connection) and downstream (in rich media content and other on-line services that are worth going to).

Second, the larger access capacities makes it feasible most of the time to run even real-time applications like voice-over-IP (VoIP) over best-effort broadband services with a fair degree of reliability – even though this was not anticipated by the original designers. Many users today expect that voice over IP services (such as Vonage) running "over the top" on the best effort network will be a substitute for their dedicated phone lines. Moreover, Internet based tools like Speedtest (http://www.speedtest.net/) allow users to quickly gauge their current upstream and downstream data rates and then compare them either to their advertised "peak" rates (which they seldom realize) or averages across a number of markets. These elevated consumer expectations demonstrate the success of the Internet, but also raise the bar for the standard that access ISPs need to meet in terms of delivering a predictable and high-quality Internet experience to unsophisticated – but nevertheless demanding – consumers.

Third, in many cases, the primary bottleneck is now owned entirely by a broadband ISP that chooses service tiers and when capacity upgrades will occur. The time scales for upgrades are determined by providers' assessments of the collective market demand. In this environment, it is plausible to ask whether the congestion problem might be managed by the access ISP by simply always over-provisioning its access network ahead of current demand.

Holding aside the all important question of how access ISPs might be incentivized to make such investments, simply over-provisioning the links does not provide a guarantee against congestion in a world where individual broadband connections could potentially offer traffic at rates in excess of 10, 100, or even 1,000Mbps either continuously or for periods of time. TCP is an opportunistic protocol that will always speed up in an attempt to use all the available bandwidth. It will keep going faster until

(a) it runs out of data to send, or

(b) it backs off when it detects congestion, or

(c) it is going as fast as the end-machines can sustain.

Since PCs today can sustain TCP transfer rates of hundreds of Mb/s, outcome (c) is not common. For applications such as VoIP, which only generate data at a limited rate, option (a) applies. But to the extent that outcome (b) is the limiting factor, as it will be for bulk data transfer, there will always be instantaneous congestion events in the network; otherwise TCP would just continue to speed up.

Moreover, when one considers the capital intensity and high cost of continuously over-provisioning access networks, it seems improbable that we should simply rely on over-provisioning as the only viable congestion control mechanism (besides TCP). Furthermore, given the relative economics of scaling backhaul connections in the core of the network, for most networks we expect that, at least for the near term, the access networks will remain the dominant constraint on achievable throughput.[29] Large application and content providers have similar expectations. In 2007, the CTO of CBS Interactive noted that the rate at which they choose to encode video traffic is determined by their assessment of the common downlink capacities of end users. "If enough people can sustain a 2Mbps video feed we will up our bit rates."

That economic considerations should impose constraints on over-provisioning is certainly neither surprising nor new. During the early years of the Internet, John Nagle described the congestion problems facing the Ford Motor Company in operating its long haul TCP/IP network.[30]

> *"In general, we have not been able to afford the luxury of excess long-haul bandwidth that the ARPANET possesses, and our long-haul links are heavily loaded during peak periods. Transit times of several seconds are thus common in our network."*

[29] Congestion will occur in other locations in the network (peering points and other inter domain and intra domain links) as well of course.

[30] See RFC 896 (1984).

Fourth, the transition to broadband has changed traffic patterns on the Internet. For many years, the peak usage periods of access networks were during the business day when commercial users were at the office. However, for at least the last number of years the peak usage hours of many access networks are in the evening roughly between 8 PM and 10 PM. This is important to understanding the economics of congestion as the previously off-peak residential customers used to easily "fit" in the pipes that had been provisioned for the commercial users. Now, however, the usage patterns of the residential customers are driving the provisioning decisions of network providers.[31]

While the aggregate level of traffic continues to grow at double-digit rates, averaging 50-60 percent CAGR per year,[32] the mix of applications is changing as well. In 2008, Cho[33] noted:

> *The current traffic is heavily affected by an eruption of peer-to-peer applications but the crust underneath is also slowly rising with video and other rich media content. The crustal movement is slow at the macro level so that it is unlikely to cause a major quake in the near future.*

This is a good metaphor as the increasingly popular video traffic does not pose an imminent threat to the stability of the Internet, but the growth in video traffic will be significant, eventually fundamentally reshaping the traffic mix on broadband networks. The dynamic nature of Internet traffic however is not new. When broadband networks were initially being adopted the "symmetry ratio" of downstream to upstream bits (i.e., bits sent to the customer compared to the bits received from the customer) was relatively high (around 18:1 according to one participant).[34] In other words, customers predominantly downloaded content. With the advent of peer-to-peer traffic, this ratio changed dramatically to around 3:1 or 2:1. In other words, traffic became much more symmetrical as users uploaded music and other shared files. With increases in video traffic the symmetry ratios are reported moving back up (around 5:1 in at least some providers' networks.)[35]

Internet traffic has changed not only in volume but character. Aggregate traffic has grown because there are more subscribers and the average traffic per subscriber has grown. Subscriber traffic is also more symmetric (downstream/upstream) although this may change again if streaming video or some other (yet unknown) asymmetric traffic type becomes a larger share of traffic. Also, streaming traffic is typically less "bursty" than

[31] See Fukuda (2005).

[32] See, Minnesota Internet Traffic Studies (MINTS) at http://www.dtc.umn.edu/mints/home.php, for data on traffic growth rates.

[33] See Cho (2008).

[34] See Leslie Ellis's description from Nov 2008 available at http://www.translation-please.com/column.cfm?columnid=244

[35] Ibid. We emphasize these are anecdotal numbers.

traffic associated with large file transfers like peer-to-peer.[36] Furthermore, with the expansion of subscribership and the menu of potential applications, users and usage has become more diverse or fat-tailed. This raises interesting challenges in terms of predicting per user usage patterns over time (and at different time scales) and for categorizing users into usage-types.

Users are sometimes categorized as exhibiting "heavy" versus "regular" or "light" usage patterns. The relationship between the aggregate volumes of traffic a subscriber sends or receives and their contribution to congestion (in terms of causing packets to be dropped) is not always clear. It is possible that a "heavy user" does not disproportionately contribute to either packet dropping congestion or to usage during the aggregate peaks on a network. What is clear though is that there are very large differences in the volume of traffic sent and received by different subscribers. While most users may download less than 2 gigabytes of traffic in a month, the top users on a system can easily exceed 100 gigabytes. Figure 6 displays the average **daily** inbound and outbound traffic per user on a fiber network in Japan measured over a week in 2008.Each dot represents one user.

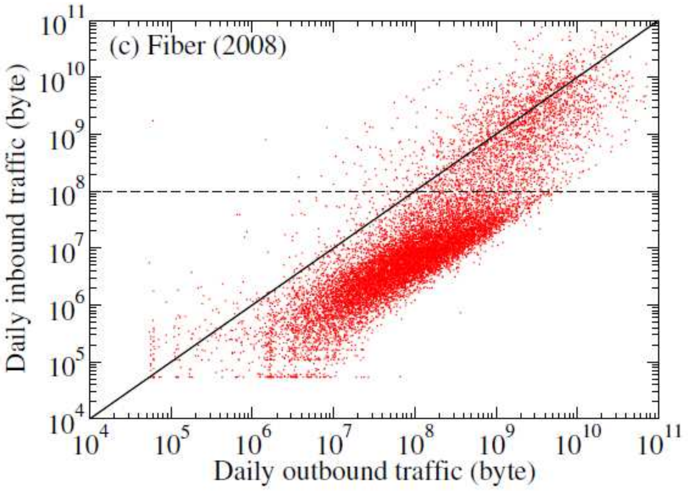

Figure 6: Correlation of daily inbound and outbound traffic volumes per user in one Japanese metropolitan prefecture for a fiber optic

[36] "Bursty" traffic has a high peak to average sending rate. The observed burstiness of a flow depends on the time over which it is observed (which defines the average rate) and so traffic that appears bursty over a short period may appear less bursty over a longer period. Streaming video of a particular quality has an optimal defined sending rate. Because it may be buffered it may be sent faster or slower than real-time, treating it more like a bulk file transfer that could be distributed via a peer-to-peer application. The ability to buffer video depends on the type of content (e.g., how sensitive is the viewing experience to needing real-time delivery?). Also, it is possible to vary the quality of the streaming video (send a lower resolution, lower bit rate image during periods of congestion and higher quality when congestion is low).

network in 2008.[37] Each dot above the dashed line represents users that sent more than 100 megabytes of traffic in a day.

As peak rates increase, and hence the possibility for sending and receiving ever larger amounts of traffic grows, there exists the potential for an increasing divergence between the volumes of traffic that different segments of the market send and receive.[38] This is not problematic in and of itself. A challenge will arise however if these very different usage patterns are associated with different underlying cost structures either in terms of the congestion they contribute to or in terms of the variable costs (such as usage sensitive charges from an upstream network provider).

In this paper, we do not offer opinions as to the desirability of being able to categorize users into distinct traffic types, as for example, "heavy" or "light" users. Certainly, a large measure of what providers seek to do in designing their service tiers (into services with different advertised peak rates, volume limits, or other features like on-line storage, static IP addresses, or multiple email addresses) is an attempt to induce users to self-select into categories based in part on their willingness-to-pay and in part on their expected traffic requirements. Identifying user types in a way that is expected to be correlated with their willingness-to-pay is a key element of what is needed to implement price discrimination. Whether such price discrimination is regarded as a good or bad thing, and whether traffic analysis is necessary to implement it, are debatable topics but not ones we wish to engage here. Let it suffice to say that while user typing may be used to implement price discrimination, this is certainly not its only (or obviously most useful) purpose. For network operators to adequately plan infrastructure upgrades, reconfigurations, and investments, they need to be able to forecast traffic. Their ability to do so appropriately is important for ensuring economically efficient congestion management over the long term. A better understanding of the components driving aggregate traffic growth in different parts of an operator's network may provide a better basis for planning such investments. Acting purely as analysts, we are always interested in more data. However, collecting more data has direct costs (e.g., measurement, computation, storage) as well as indirect effects (e.g., enabling better service customization, potentially enhancing security, or enabling better matching of costs and usage; *or,* potentially threatening privacy, adversely impacting competitive dynamics, or facilitating an abuse of any market power that may exist).

In light of the preceding, it is clear that the congestion environment is much more complex in the broadband Internet than in earlier periods, and that the nexus of congestion control is shifting into the network. Neither end-user devices, applications, nor inherent limits in link capacity provide an effective throttle on the volume of traffic that a potential end-node *may* deliver to the network. Moreover, prospects for continued

[37] See Cho (2008).

[38] In addition to distinct differences in the usage patterns of different types of users, there may be different numbers of each type; and they may be distributed differently across a network in ways that may be related to what they are doing (e.g., different on-net/off-net patterns) with resulting implications for aggregate traffic flows.

innovation and growth in rich-media services (e.g., from TV-over-Internet, sensor-enabled remote monitoring, wireless broadband expansion, machine-to-machine communications, et cetera) do not provide a confident basis for forecasting the saturation of per user demand growth.[39] The increased uncertainty and shorter time scales over which aggregate traffic may shift make it increasingly difficult to manage congestion relying solely on either the per-flow TCP mechanisms or the longer-term capacity expansion. Thus, it is not surprising that network operators confronted with this more complex environment have opted to explore a range of new traffic management techniques.

5. Evolving mechanisms for dealing with congestion

Congestion management is inherently contentious. The fundamental problem is that during periods of congestion, shared network resources must be allocated. Those who get more may be happy, while those who get less may be unhappy.[40]

There is no universally acceptable definition of what the right solution is, or even what a "fair" solution is. In economics, although in theory it is sometimes possible to separate efficiency and equity considerations, in practice, these are usually linked. Often, as a matter of simplicity, economists abstract from the distributional concerns (equity/fairness) and focus on efficiency by assuming a single economic optimizing agent (who presumptively maximizes total surplus and then allocates that surplus to participants). Indeed, few seem to object when a company actively manages the Internet usage of employees,[41] while decisions of a public ISP to manage the traffic of its customers is more likely to raise concern.

What is true more generally has also been true within the Internet technical community. What is interesting is that over the history of the Internet the answers to how congestion should be managed have varied. The dominant answer in operation at any time has almost always had detractors and competitors – no universal consensus has ever existed. This is unsurprising given that (in some respects) this is a debate about what is "fair" and about what is economically efficient to deploy and operate. We briefly trace some of the history of congestion management here -- it is a richer and more varied history than some might suspect.

[39] If we thought we could forecast long-term per-user traffic requirements (and ignored machine-to-machine traffic), we could forecast total aggregate traffic based on population and subscribership forecasts.

[40] Even if all parties agree *ex ante* (before congestion occurs) what an efficient or fair allocation might be, parties need not be happy *ex post* (when congestion-limited resources are actually allocated).

[41] Although, considerations of employee rights and privacy may arise in this context as well.

The problem of network congestion was anticipated in Paul Baran's pre-Internet papers at the Rand Corporation which detailed his proposal for packet switched networking.[42] His proposed architecture solved the congestion problem by investing network operators with the responsibility for arbitrating which packets were transmitted first. Communications officers operating each node of the network set the priority of traffic for different military units or different types of traffic. Quite literally congestion was to be solved with command and control decisions by the network communication officers. The hosts simply injected packets into the network. Once the priority levels were set, the network operated automatically forwarding packets according to the assigned priorities. Figure 7 is a copy of the "communications control console" from Baran's paper.

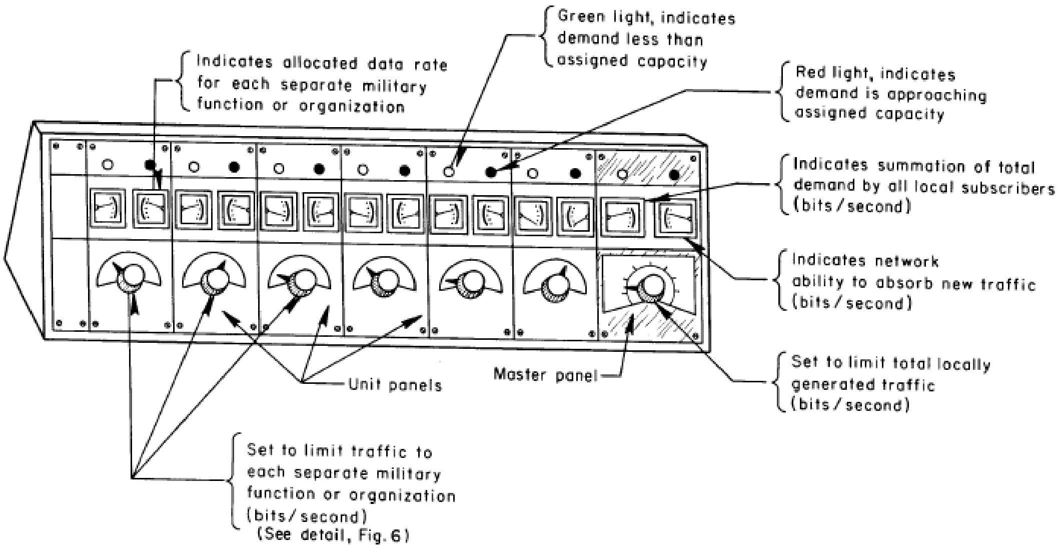
Green light, indicates demand less than assigned capacity
Indicates allocated data rate for each separate military function or organization
Red light, indicates demand is approaching assigned capacity
Indicates summation of total demand by all local subscribers (bits/second)
Indicates network ability to absorb new traffic (bits/second)
Unit panels
Master panel
Set to limit total locally generated traffic (bits/second)
Set to limit traffic to each separate military function or organization (bits/second)
(See detail, Fig. 6)

Figure 7: The communications control console depicted in Paul Baran's Rand paper proposing packet switched networks in 1964.

This is interesting as responsibility for resolving what was sent during overload was the responsibility of the network. The senders or receivers had to lobby the communication officer if they wanted priorities to be adjusted so they could get more traffic through.

Nothing like this was built for the actual Internet (though priority queuing would return with the later research on quality of service).[43] In the early days of TCP, hosts on the network would send packets into the Internet as fast as the advertised window of available buffer space on the receiver would allow. If a gateway dropped a packet due to overload, the host would timeout after not receiving an acknowledgement of the packet and resend it.

[42] See Baran (1964).

[43] See Campbell (1996).

Gateways (the precursors of routers in today's network) did have the ability to explicitly signal to the edges about congestion. They sent ICMP Source Quench messages to hosts if the gateway became overloaded. As noted in RFC 791:[44]

> *The source quench message is a request to the host to cut back the rate at which it is sending traffic to the internet destination. The gateway may send a source quench message for every message that it discards. On receipt of a source quench message, the source host should cut back the rate at which it is sending traffic to the specified destination until it no longer receives source quench messages from the gateway. The source host can then gradually increase the rate at which it sends traffic to the destination until it again receives source quench messages. The gateway or host may send the source quench message when it approaches its capacity limit rather than waiting until the capacity is exceeded.*

However this proved to be inadequate to control congestion and was later discontinued.[45] Parts of the Internet suffered from temporary periods of "congestion collapse" i.e. periods of time when the network was busy sending packets but most of the packets were duplicates of previous packets that had already been sent. In other words, little real work was being done. As described in one such incident:[46]

> *In October of '86, the Internet had the first of what became a series of 'congestion collapses'. During this period, the data throughput from LBL to UC Berkeley (sites separated by 400 yards and two IMP hops) dropped from 32 Kbps to 40bps. We were fascinated by this sudden factor-of-thousand drop in bandwidth and embarked on an investigation of why things had gotten so bad.*

The solution to this particular congestion problem was a series of new congestion controls invented by Van Jacobson based on a "congestion window".[47]

The cleverness of the TCP congestion controls involved how the congestion window expanded and contracted as a host continually discovered the available capacity in the network. This involved methods with names like slow start, additive increase/multiplicative decrease, and fast transmit and fast recovery.[48] For instance, slow start at the beginning of a connection doubled the size of the congestion window every

[44] See http://www.ietf.org/rfc/rfc791.txt.

[45] Hosts today ignore ICMP source quench messages.

[46] See Jacobson (1988).

[47] Van Jacobson introduced a "congestion window" parameter – cwnd – that was added to the per connection TCP state of hosts. The sending rate of a host would now be limited by either the receiver's advertised window or the congestion window tracked by the sender. For further discussion, see Peterson (2007) and Note 4 *supra*.

[48] See Peterson (2007) for details.

time a congestion window's worth of traffic was successfully sent until either a slow start threshold was crossed or a dropped packet occurred. This exponential rate of increase quickly grows the amount of traffic a sender is capable of putting into a network. While this is aggressive, it is actually slower (hence the name "slow start") than the previous behavior of TCP. Before the "slow start" modification, TCP would send a burst of traffic that was as large as the advertised window of the receiver.

TCP congestion control was a pragmatic and demonstrably working solution to the existing congestion problems. TCP congestion control could be deployed simply by updating the software on hosts without needing to modify any routers in the network. This change was followed by a long series of tweaks as new ideas like selective acknowledge were incorporated to improve performance.[49]

The research community was also active in exploring alternative means of detecting congestion and adjusting the sending rate. With the original TCP, the sending rate had to be increased past the available capacity, in effect deliberately creating loss as a means of finding the available bandwidth of the path. Alternatives were experimented with that included detecting the increase in round trip delay as a queue began to build in the network, cutting the sending rate before a packet loss was detected.[50]

However, while TCP was generally adopted as the right solution at the time and has proved to have enduring value over the following decades, Van Jacobson[51]in his original paper on the topic noted some limitations:

> *While algorithms at the transport endpoints can insure the network capacity isn't exceeded, they cannot insure fair sharing of that capacity. Only in gateways, at the convergence of flows, is there enough information to control sharing and fair allocation. Thus, we view the gateway 'congestion detection' algorithm as the next big step.*

This "big step" has been explored in research particularly in the form of quality of service research. But in practice little has changed about how shares of capacity are allocated to competing best-effort Internet traffic. The result is that most traffic on today's Internet will get the share of traffic that the TCP stack on a host calculates.

Under TCP, each flow will receive a roughly equal share of the bottleneck capacity. If there were 100 flows through a router with a capacity of 100 mb/s, each flow would receive approximately 1 mb/s under TCP fairness (assuming each of those flows wants

[49] See RFC 4614 "A Roadmap for Transmission Control Protocol (TCP) Specification Documents" available at http://www.ietf.org/rfc/rfc4614.txt.

[50] See http://www.ecse.rpi.edu/Homepages/shivkuma/research/cong-papers.html for an extensive bibliography of academic papers. For a more recent perspective on research challenges see "Open Research Issues in Internet Congestion Control" http://tools.ietf.org/html/draft-irtf-iccrg-welzl-congestion-control-open-research-04 (May 2009).

[51] Jacobson (1988), supra note 46.

more than 1 mb/s).[52] This may sound equitable or fair, but that clearly depends on one's opinion of what constitutes fairness. In the Internet today, a traffic flow is typically defined as the unique four-tuple of 1) source address, 2) source port, 3) destination address, 4) destination port. The key here is that a host can use multiple ports. Different hosts can therefore have different numbers of flows active in any bottleneck.

Consider four hosts that have each opened twenty flows and twenty hosts that have each opened one flow through the same 100mb/s pipe. In this case each of the twenty hosts would get 1 mb/s of throughput while the four hosts with twenty flows would each get 20mb/s (1mb/s over each of the 20 flows). Some might not view this as equitable. Now it could also be the case that ten hosts have each opened ten connections apiece. Each would then get approximately 10mb/s of capacity. Both of these scenarios would be considered TCP fair.

On the network today, popular peer-to-peer applications do open multiple simultaneous connections to different hosts (whether or not they open simultaneous connections to the same remote host is a different question). These multiple connections may or may not share common bottlenecks depending upon where in the network topology the congestion arises. So the distribution of capacity in a bottleneck link during periods of congestion on today's Internet isn't clear. Sitting at the edges, we can't comment on how equitably traffic is distributed during periods of congestion. Only network operators have such a view.

Alternative ways of defining and managing "flows" have been explored in the past. Rather than per-flow congestion control, a "congestion manager" was proposed by Balakrishnan[53] which could control the transmission of packets on a per-host or per-domain basis. The IETF has similarly explored other ways of defining flows.[54]

In more recent years, the topic has again been raised. In "Flow Rate Fairness: Dismantling a Religion," Briscoe argues that "it is actually just unsubstantiated dogma to say that equal flow rates are fair."[55] He notes that:

> *Fair allocation of rates between flows 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.*

[52] In practice, other factors are important as well such as the round trip time. Flows of traffic with smaller round trip times have an advantage under TCP and will thus get a larger share of the capacity.

[53] See Balakrishnan (1999).

[54] See for instance RFC 2309 (http://www.ietf.org/rfc/rfc2309.txt and RFC 2914 (http://www.ietf.org/rfc/rfc2914.txt).

[55] See Briscoe (2007).

While others like Mathis (2009) also question whether TCP fairness should remain the dominant paradigm, flow rate fairness has defenders. Floyd and Allman (2008) note the advantages of TCP fairness in terms of its simplicity and the minimal demands it places on technical or economic infrastructure. They also note the proven track record in the real world (which is indeed a good argument for any engineered system). They do, however, acknowledge limitations with the existing paradigm:

> *We do not, however, claim that flow-rate fairness is necessarily an *optimal* fairness goal or resource allocation mechanism for simple best-effort traffic. Simple best-effort traffic and flow-rate fairness are in general not about optimality, but instead are about a low-overhead service (best-effort traffic) along with a rough, simple fairness model (flow-rate fairness).*

In this paper, we are not taking a position on what is desirable. We highlight these differing views simply to demonstrate the existence of technical disagreement as to how congestion should be managed. This continues to be a healthy and productive debate that is engaging a wide range of technical stakeholders. We expect that interesting standards and new products will eventually be produced as a result of this work.

While the underlying principals of congestion management are being debated in the standards and research communities, network operators have been faced with what they perceive to be genuine problems during periods of temporary congestion on their network. They face demands from subscribers to improve (or maintain) application performance but have limited technical means that are popularly viewed as acceptable to arbitrate how a fixed amount of capacity is allocated among subscribers.

Access providers have deployed or experimented with network devices that implement provider selected congestion management policies based on traffic analysis that have met with opposition from a mix of stakeholders. These management decisions may be based on essentially any information included in a data packet– hence the common name for this technique is "deep packet inspection" or DPI. Use of DPI-enabled network management controls determine how traffic is handled by the network. The crucial point is that they often change the allocations that would result from the distributed actions of hosts' applications and TCP stacks. While it is certainly different than TCP fairness, it is not *de facto* unfair or welfare reducing, however, the wider Internet community *might* regard it as unfair or inefficient depending upon the policies that are implemented. Additionally, the prospect of ISPs looking inside users' packets has been compared to the postal clerks reading your mail, which raises privacy concerns.[56]

Another technique that some providers have adopted or experimented with is limiting the total volume of traffic that a subscriber can send or receive over some time span. Monthly limits on upload/download or the combined traffic volume are the most common types of

[56] See https://www.dpacket.org/blog/kyle/inaccurate-analogy-dpi-equivalent-postal-service-reading-your-mail (accessed 8-14-09) for an account of comments comparing DPI to the postal service reading your mail, reported by those who dispute the analogy.

limits.[57] Interestingly some Japanese broadband providers have instituted daily traffic limits (e.g. 15 or 30 GB of uploaded traffic).While volume limits are also employed to segment the subscriber market, they also may limit the congestion on a network (if aggregate volume of traffic is a predictor of expected contribution to periods of congestion).

Up to this point we have been focusing on congestion management in terms of the networking textbook definition of congestion (i.e., when packets are being dropped). However, as we noted, the network providers' common definition of congestion is based on the aggregate volume of traffic exceeding some threshold over some measurement interval. We imagine that the two are somewhat correlated, but there is no public research we are aware of to document that correlation.

Furthermore, even if aggregate link utilization over some period is linked to peak utilization during periods of congestion, this does not obviously imply how such aggregate utilization relates to the economic costs of providing service or managing congestion. If the traffic is off-peak, then some might argue that the incremental cost of carrying the traffic is low or close to zero. While that may be true, it may also be the case that there are significant usage-sensitive charges born by the ISP such as charges for transit services.[58]

On today's Internet, hosts have no information about any of these other ways of looking at congestion. Hosts or applications couldn't help reduce or avoid aggregate congestion even if they wanted to (beyond limiting their own demand). But this may be changing.

An example of these changes followed the public kerfuffle that erupted in the fall of 2007 involving the FCC, Comcast, and BitTorrent over Comcast's efforts to actively manage peer-to-peer traffic on its network.[59] Broadband customers (not just on Comcast's network) using over-the-top VoIP applications like Vonage (that are carried along with other best-effort Internet traffic) were complaining about the quality of their experience. Comcast adopted a technical approach by which they injected TCP reset packets into certain BitTorrent flows to cause those flows to be disconnected temporarily, with one effect being that more capacity was available for other over-the-top applications like Vonage. Of course, if you were a BitTorrent user whose session was interrupted, you might regard such treatment as unacceptable. Or, if you were a provider of over-the-top services that anticipated operating over Comcast's network, or especially, if you were

[57] See http://www.oecd.org/dataoecd/22/46/39575020.xls

[58] ISP intercarrier traffic charges may arise as a consequence of interconnection agreements that are impacted by regulation and business practices that may deviate significantly from underlying economic costs. ISPs may pay for transit on the basis of aggregate traffic or their inter-traffic flow may impact an ISPs ability to maintain a peering agreement. See Faratin *et al*. (2007) for further discussion.

[59] See, for example, "Comcast caught blocking BitTorrent traffic," October 22, 2007 (available at: http://www.vnunet.com/vnunet/news/2201667/comcast-caught-blocking).

anticipating offering services that might compete directly with offerings from Comcast, you might be concerned that such an approach boded ill for your ability to manage the quality of your service. While most folks in the technical community do not regard the approach of resetting TCP connections as a good long-term solution for congestion control, there is no general agreement as to what the appropriate response should have been. The FCC decided to investigate Comcast's traffic management practices and issued an order directing Comcast to end its "discriminatory" practices.[60] However, this hardly resolved the issue to everyone's satisfaction and the debate over what constitutes acceptable ISP traffic management continues.[61]

While the recent kerfuffle mentioned above is now well known, the subsequent work in the technical community provides an interesting example of cooperative efforts to improve traffic behaviors during periods of congestion.

The technical community, including representatives from Comcast and BitTorrent gathered at MIT in May 2008.[62] Both companies gave presentations explaining the challenges they face and their current approaches for managing congestion. Comcast detailed their trials of protocol agnostic congestion management (which has since become fully deployed). BitTorrent detailed their approach which interestingly does not rely upon TCP alone to determine its sending rate.[63]

> *[BitTorrent's client is] continuously sampling one-way delay (separating propagation from queuing delay) and targeting a small queuing delay value. This essentially approximates a scavenger service class in an end-to-end congestion-control mechanism by forcing bulk, elastic traffic to yield to competitors under congestion.*

Other proposals, research, and results of trials were also presented for improving congestion. These included cooperative efforts like the "Provider Portal for P2P Applications (P4P)" where providers could express the "virtual cost" of their intra-domain or inter-domain links. In this architecture, virtual costs could reflect any kind of provider preferences and could be based on the provider's choice of metrics, including utilization, transit costs, or geography. Applications would then cooperatively select peers, resulting in improved performance for all parties. During the workshop, presentations were offered by major ISPs, by academic researchers, and by third-party solution vendors (e.g., the results of successful trials by Pando networks).

[60] See http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-284286A1.pdf.

[61] See, for example, http://arstechnica.com/old/content/2008/08/reactions-to-fccs-comcast-spanking-come-fast-and-furious.ars.

[62] See http://downloads.comcast.net/docs/Comcast-IETF-P2Pi-20080528.pdf or http://trac.tools.ietf.org/area/rai/trac/wiki/PeerToPeerInfrastructure for additional materials from the workshop.

[63] See http://www.rfc-editor.org/rfc/rfc5594.txt.

We recommend reading RFC 5594[64] as a useful summary of the technical topics discussed. The workshop was well attended by a broad representation from the technical community and produced several ideas that are being pursued within the IETF community. Efforts are, for instance, under way to standardize algorithms that allow an application to have what amounts to a "scavenger" class of traffic that willingly backs off more than necessary (according to conventional TCP requirements) during periods in which Internet congestion is detected.

6. Policy implications, questions and discussion

We emphasize that we are not taking a position here on the desirability of any of these competing congestion management approaches. We however are very supportive of the continuing competition of ideas. Given this healthy competition, we expect technical approaches to congestion management to continue to evolve.

Given the stakes, though, it is a fair question how different approaches to managing congestion may affect the competiveness and health of the Internet ecosystem. How congestion is managed – how resources and costs are shared – has potentially far reaching implications. It is certainly possible that network operators, under the guise of managing congestion, may exploit their control over the network pipes in ways that are socially undesirable (either intentionally as a result of a desire to exploit any market power they may have or unintentionally because this is a complex system and it is difficult to anticipate all direct and indirect effects). It is also certainly possible that regulatory actions could prevent genuinely beneficial management practices, regardless how well intentioned the regulatory intervention may be.

We echo one of the recommendations that came out of last year's IETF meeting on this topic:[65]

> *Obtaining more data about Internet congestion may also be a helpful step before the IETF pursues solutions. This data collection could focus on where in the network congestion is occurring, its duration and frequency, its effects, and its root causes. Although individual service providers expressed interest in sharing congestion data, strategies for reliably and regularly obtaining and disseminating such data on a broad scale remain elusive.*

Just as this is a prudent course of action for the technical community, it is a prudent course of action for the policy and regulatory communities. We are currently working with a diverse group of network providers to systematically collect just this type of data.[66] Others are engaged in similar efforts to better understand the actual state of the

[64] See http://www.rfc-editor.org/rfc/rfc5594.txt.

[65] Ibid.

[66] At the time of writing, this project is still being launched. See http://mitas.csail.mit.edu for further details.

network.[67] Actual traffic data and appropriate analysis may assist both the technical community and wider-Internet community in assessing claims about the impacts of alternative control strategies and toward reaching consensus on what best practices might be (or at least, what bad practices might be).

The availability of data, however, will not ensure consensus. The same data may be viewed in different ways, and the range and diversity of data that may be collected is potentially immense. Thus we need further work toward defining appropriate metrics and ways of thinking about Internet traffic. There is an active research agenda here that challenges attempts by policymakers that might prefer a simple answer to what amounts to a difficult and complicated, multidisciplinary, set of questions. Earlier we presented several different definitions for congestion and recounted some of the intellectual history in the evolution of thinking about congestion to suggest the importance of this complexity. The policy community should recognize which definition of congestion is being appealed to with each type of data they are presented. However, we emphasize, there is no "right" definition of congestion. Each has implications and is important in particular contexts.

Another obvious question is whether network operators should be transparent about the congestion management policies they employ? It seems relatively uncontentious to regard voluntary transparency as desirable.[68] A number of providers publically document their policies.[69] This helps the outside technical community debug networking problems, promotes trust between participants, and helps set expectations about what types of applications and services will perform well in the future. Without disclosure, congestion management policies are often difficult to infer from the edges using probing techniques.

If voluntary transparency is desirable, should transparency of congestion management policies be required? Going even further, one might imagine mandatory disclosure of the resulting outcomes of applying any congestion management mechanisms. On these

[67] Some of the important initiatives in this area include MINTS (http://www.dtc.umn.edu/mints/); CAIDA (http://www.caida.org/research/traffic-analysis/), and M-Lab (http://www.measurementlab.net/).

[68] A possible counterargument is if the disclosure of traffic management techniques adversely impacted competitive dynamics or posed a collective security threat (e.g., enabling malicious users to better circumvent valid policies or launch distributed denial of service attacks). ISPs may feel compelled to "voluntarily" disclose policies because of market pressures, even if the equilibrium in which such disclosures are made is socially undesirable (e.g., there is a kind of Prisoners' Dilemma going on). While such a possibility cannot be ruled out, at present, we regard this as far-fetched and applaud voluntary disclosure of practices.

[69] See Comcast's Network Management Policy http://networkmanagement.comcast.net and “BT Total Broadband Fair Usage Policy” http://bt.custhelp.com/cgi-bin/bt.cfg/php/enduser/cci/bt_adp.php?p_faqid=10495&cat_lvl1=346&p_cv=1.346&p_cats=346

questions we are more skeptical. We would particularly not like to see such requirements interfere with network providers experimenting with alternative congestion management techniques.

There is no mandatory disclosure of routing or interconnection policies of network operators, however, these are as influential as congestion management in shaping the Internet ecosystem. Further it strikes us that it would be difficult to crisply define what separates congestion management techniques from other network management policies. Is the deployment of a cache for peer-to-peer traffic a congestion management technique or a way of reducing traffic load while improving application performance?

However, recognizing that the problem is difficult is not the same as assuming it does not potentially exist. There remains a valid concern that an ISP that may have market power may seek to use its ability to manage traffic to discriminate against or extract rents from an unaffiliated video service provider. Conversely, a municipality may over-invest in local broadband infrastructure in a misguided attempt to prevent all congestion through over-provisioning. In these and other cases, there will remain an important collective need to monitor the congestion health of the Internet ecosystem.

Another area where we lack adequate insight relates to user perceptions. End-user perceptions of the congestion state and the fairness of traffic allocations within that network will not depend on whether packets are queued or dropped, but on how their user-experience is impacted.[70] Today, there are only crude measures of user satisfaction available. These include things like the rate of complaints to a provider's help desk. Part of this is clearly subjective. The system may be "fair" in some way if users feel no worse off than their neighbors.[71]

How to reason about what could happen in the future is challenging in this discussion. There is the general perception that there are a series of pent-up applications that will emerge if network capacity is expanded, and that it is bandwidth limits (or imposed congestion) that gate the emergence of these applications. To the extent that this supposition is correct, one cannot hope to build a truly congestion-free network. New applications will always emerge to utilize that unused capacity. One must therefore assume that the users are in a perpetual state of mild frustration, anticipating the coming applications they cannot quite use today. In this case, planning for traffic and coping with congestion must be seen as a dynamic process.

While metaphorical reasoning always has weakness, determining what the right level of network investment is and how capacity should be shared is similar to the questions that

[70] On the other hand, some end users are likely to blame any failure of their experience on network congestion. For instance, the failure to or delay in resolving a DNS name is often ascribed to network congestion instead of the DNS server problem.

[71] Even the quality of telephone calls is measured subjectively based on user tests that rate call quality on a scale of 1 to 5, known as the Mean Opinion Score (MOS) (see http://technet.microsoft.com/en-us/library/bb894481.aspx for further information).

come up when considering the right student to teacher ratio in classrooms. There will always exists a general clamor for "lower" student to teacher ratios no matter what the current level is, yet there is also a recognition that a one to one ratio of teachers to students would be a waste of resources.

Similarly, once a student to teacher ratio is fixed, there is a question of how the teacher's time should be divided among students. Students, with differing demands, have to "share" the teacher's times. Should each student get an equal share of the teacher's time? If a student is out sick for a week, should the teacher spend more time with the student to catch him up, or should each student still get an equal share?

7. Conclusion

Internet congestion arises as a direct consequence of resource sharing. During congestion episodes, allocating more resources to some usually means allocating less to others. Economists argue that the allocation process ought to be efficient -- in an allocative (resources go to their highest value uses), productive (costs are minimized) and dynamic (investment is optimized) sense. Achieving this goal in a way that is also perceived to be "fair" is a difficult challenge that is rendered more difficult by the fact that the Internet has evolved vastly in size and importance in the global economy. The architecture and operation of the Internet have important economic and policy ramifications that engage the interests of a much wider community than the principally technical community of academics and network engineers focused on the design of data communications protocols or the operation of the public and private data networks that comprise the Internet.

In this paper, we have reviewed the technical history of congestion control management in the Internet, illustrating how thinking has evolved in response to new challenges that emerged as the size and traffic on the Internet grew. The very origins of the Internet lay in the observations that establishing a telephony circuit to transfer small amounts of data between two computers was inefficient. The telephony model was simply not a good fit for the type of traffic computers generated.[72] Similarly the provisioning of link capacity, routing policies, and other network management decisions are made based upon assumptions about the behavior of traffic, its growth rate, and the implications of congestion for application performance and end user satisfaction. In other words, how a network is managed crucially depends upon the traffic it is carrying. An enduring product of this process was TCP fairness and its associated tweaks that was ushered in by Van

[72] In 1965, Donald Davies noted that the connection setup times establishing a circuit to transfer small amounts of data was inefficient (Kirstein, 1998). Data traffic between the earlier computers did not have the well-understood characteristics of the Erlang distribution common in telephony networks. Contemporaneously Paul Baran at Rand made a similar observation that networks could be architected to avoid the costly connection setup times and the single points of failures of the tradition hub and spoke telephone networks. Both proposed that packet switches may be more efficient for variable length data traffic.

Jacobson in the late 1980s, and since then has provided the most important and widely used (short-term) mechanism for controlling congestion over the best-effort Internet.

Of course, the most important way to manage congestion is to provide adequate capacity. What is "adequate," however depends on one's perspective. It is ultimately a question of network architecture, investment, and user demands. From the perspective of network operators, it is a peak-load planning problem in the face of stochastic demands and costly capacity. Network operators must continuously weigh the costs of incremental investment in base-load capacity, the variable costs of peak-load management (e.g., using higher-cost facilities or workarounds during peak traffic episodes), and the potential costs imposed on users as a result of congestion (and potential for lost revenue). Furthermore, each network operator is trying to solve this problem in the face of complex market, industry, and regulatory dynamics that collectively comprise the Internet ecosystem.

Holding aside how the above market-mediated process decides what Internet infrastructure looks like at any particular point in time, there will be episodes of congestion (loads exceed available capacity) that will require management in the short-run. With the growing complexity of the Internet and the transition to a broadband Internet, network operators and ISPs see themselves as needing to more actively manage traffic over time horizons significantly shorter than the network investment horizon.

The use of techniques and technologies like volume capping, usage-based pricing, application prioritization, and Deep Packet Inspection all represent significant deviations from TCP fairness in terms of how resources are allocated during periods of congestion. The employment of some of these techniques also raises concerns about the potential for abuse of market power, for threats to privacy, or for threats to the openness of the Internet. Efforts by regulatory authorities in the U.S. and elsewhere to actively investigate ISP traffic management practices have demonstrated the importance of these issues for the health of the Internet ecosystem.

In this paper we do not offer a position on the merits of alternative traffic management practices. Our goal is instead to educate the wider community regarding some of the history of these issues within the technical community. Our assessment of this legacy and of more recent research efforts to characterize Internet traffic more carefully lead us to conclude that there is ample scope for useful innovation in ISP traffic management practices beyond TCP fairness. Consequently, we would caution against any regulatory policies that had the likely effect of enshrining TCP fairness and thereby limiting the scope of the Internet technical community's on-going experiments with how to best manage best-effort traffic over medium (month or less) to short time-scale (seconds to minutes).

8. References

Balakrishnan, H., Rahul, H. S., and Seshan, S. (1999). “An integrated congestion management architecture for Internet hosts,” SIGCOMM Comput. Commun. Rev. 29, 4.

Baran, P., (1964) *On distributed communication networks*. Rand Corporation Document Series, 1964.

Briscoe, B. (2007), “Flow Rate Fairness: Dismantling a Religion.” *SIGCOMM Comput. Commun. Rev*. 37, 2 (Mar. 2007), 63-74.

Campbell, A., Aurrecoechea, C.,and Hauw (1996), L., “A Review of QoS Architectures,” Multimedia Systems,vol. 6, 1996, pp. 138--151.

Cho, K., Fukuda, K., Esaki, H., and Kato, A. (2008),“Observing slow crustal movement in residential user traffic,” In Proceedings of the 2008 ACM CoNEXT Conference (Madrid, Spain, December 09 - 12, 2008). CONEXT '08. ACM, New York, NY, 1-12.

Cisco (2009), “Network-Based Application Recognition - Cisco Systems,” available at: http://www.cisco.com/en/US/docs/ios/12_1/12_1e11/feature/guide/dtnbarad.html (retrieved 7-1-2009).

Comcast (2008), “Comments of Comcast Coporation” Feb 2008. http://fjallfoss.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6519840991

Crowcroft, J., Hand, S., Mortier, R.,Roscoe, T., and Warfield, A., “QoS's Downfall: At the bottom, or not at all!,” Proceedings of the Workshop on Revisiting IP QoS (RIPQoS), at ACM SIGCOMM 2003, August 27, 2003.

Faratin, P., D. Clark, P. Gilmore, A. Berger, and W. Lehr (2007), "Complexity of Internet Interconnections: Technology, Incentives and Implications for Policy," paper prepared for 35th Annual Telecommunications Policy Research Conference, George Mason University, September 2007.

Feller, W. (1939), *Die GrundlagenderVolterraschenTheorie des KampfesumsDasein in wahrscheinlichkeitstheoretischerBehandlung*, ActaBioth. Ser. A 5 11–40. MR0690284, 1939.

Floyd, S. and M. Allman (2008), "RFC 5290: Comments on the Usefulness of Simple Best-Effort Traffic," Network Working Group, Internet Engineering Task Force, July 2008, available at: http://www.faqs.org/rfcs/rfc5290.html.

Floyd, S. (2004), "Thoughts on the Evolution of TCP in the Internet (version 2)," ICIR, March 17, 2004.

Fry, T. C. (1928), *Probability and its engineering uses*, Van Nostrand: New York, 1928.

Fukuda, K., Cho, K., and Esaki, H. (2005),“The impact of residential broadband traffic on Japanese ISP backbones,” Sigcomm Comput. Commun. Rev. 35, 1 (Jan. 2005), 15-22.

Gross, D. and C. M. Harris (1998), *Fundamentals of Queueing Theory*, Wiley Series in Probability and Statistics, Wiley-Interscience, February 1998.

IETF (2009), IETF Low Extra Delay Background Transport (LEBDAT) Working Group Charter, 2009 (available at: http://www.ietf.org/html.charters/ledbat-charter.html).

Jacobson, V. (1988),“Congestion avoidance and control,” Computer Communication Review, 18(4):31429, August 1988.

Kirstein, Peter (1998), "Early Experiences with the ARPANET and INTERNET in the UK," mimeo, available at: http://nrg.cs.ucl.ac.uk/mjh/kirstein-arpanet.pdf.

Kleinrock, Leonard (1961), "Information Flow in Large Communication Nets," Proposal for a Ph.D. Thesis, Massachusetts Institute of Technology, May 31, 1961.

Lehr, William, Jon Peha, and Simon Wilkie (eds) (2007), *Special Section on Network Neutrality: International Journal of Communication* (volume 1, 2007), August 2007 (http://ijoc.org/ojs/index.php/ijoc/issue/view/1).

Leiner, B.M., V.G. Cerf, D.D. Clark, R.E. Kahn, L. Kleinrock, D.C. Lynch, J. Postel, L.G. Roberts, and S. Wolff (1997), “A Brief History of the Internet,” *Communications of the ACM*, vol. 40, 1997.

Licklider (1963), "Topics for Discussion at the Forthcoming Meeting, Memorandum For: Members and Affiliates of the Intergalactic Computer Network". Washington, D.C., Advanced Research Projects Agency, 23 April 1963 (via KurzweilAI.net, retrieved 2009-06-15).

Lyon, M. (2004), *Where Wizards Stay Up Late*, Simon & Schuster: New York, 2004.

Mathis, M. (2009), “Rethinking TCP Friendly”, March 2009 at http://staff.psc.edu/mathis/papers/TSVAREA73.pdf (retrieved March 30th 2009).

Peterson, L. L. and B.S. Davie (2007), *Computer Networks: A Systems Approach*, Fourth Edition, Morgan Kaufmann: New York, 2007.

Stallings, W. (2008), *Operating Systems: Internals and Design Principles*, 6th Edition, Prentice Hall: New Jersey, 2008.

V. Jacobson and M. J. Karels (1988), "Congestion avoidance and control," ACM Computer Communication Review; Proceedings of the Sigcomm '88 Symposium in Stanford, CA, August, 1988, vol. 18, 4, pp. 314-329, 1988.