IEEE★USA

NETWORK TRAFFIC MANAGEMENT AND THE EVOLVING INTERNET

A WHITE PAPER

By

The Committee on Communications Policy

Institute of Electrical and Electronics Engineers - United States of America

2 November 2010

PURPOSE

Network Traffic Management (NTM) is a collection of techniques that may be used by Internet service providers to attain optimum performance for diverse classes of users. These techniques include the use of performance measures to define optional service levels tailored to different user needs, and to assure quality of service.

This paper discusses the nature of NTM, its effect on the orderly delivery of existing and future services, and its potential value in developing effective telecommunications policy. It has been prepared by technologists, for the use of stakeholders in the ongoing debates that will shape the further evolution of the Internet. It does not take positions in those debates, but attempts to increase the reader's understanding of the potential opportunities that NTM provides and steps, such as the standardization of performance measures, that could enhance those opportunities.

This white paper was prepared by the **Committee on Communications Policy** of **The Institute of Electrical and Electronics Engineers-United States of America (IEEE-USA),** with special assistance from CCP members *G. Matthew Ezovski, Emily Sopensky, John Richardson, Dan Lubar, Earl Turner,* and *Eric Burger*. It represents the considered judgment of a group of U.S. IEEE members with expertise in the subject field. A roster of committee members is provided at the end of this document.

White papers are designed to provide balanced information on public policy issues in technology-related areas that may affect the interests of technical professionals. This document does not constitute a formal position statement of the IEEE-USA, and its contents do not necessarily reflect the views of IEEE-USA, IEEE or other IEEE organizational units. IEEE-USA has issued this whitepaper to enhance knowledge and promote discussion of the issues addressed. IEEE-USA advances the public good and promotes the careers and public policy interests of more than 215,000 engineers, scientists and allied professionals who are U.S. members of IEEE. IEEE-USA is part of IEEE, the world's largest technical professional society with 375,000 members in 160 countries.

TABLE OF CONTENTS

Purpose 2
Table of Contents 3
Executive Summary 4
Introduction 6
Digital Convergence Drives the Internet 7
Fundamental Questions of Network Traffic Management 7
Stakeholders and Their Interests 8
Service Provider–Driven Network Management 10
Establishing Principles of Network Operation and Traffic Management 12
Performance Measures and Quality of Service 12
User-Selected Service Levels 16
Models for Managing Traffic 19
Managing Traffic in the Core 20
Managing Traffic at the Edge 20
Relevance to Net Neutrality 23
Supplemental References 25
2010 IEEE-USA CCP Membership Roster 27

EXECUTIVE SUMMARY

Network Traffic Management (NTM) is a collection of techniques that may be used by Internet service providers to attain optimum performance for diverse classes of users. These techniques include the use of performance measures to define optional service levels tailored to different user needs, and to assure quality of service. Traffic management is already common in the portion of a network where it is possible to smooth traffic flow without affecting performance.

NTM is critically important to the proper functioning of the Internet, yet, NTM can also be misused to create unfair access or use of the Internet. The paper presumes that an objective exposition of NTM's technical issues will help policy makers, regulators, and the industry develop fair and informed regulations and policies.

The Internet, as an international interworking of independently operated, autonomous networks, has, by definition, neither a central governing body nor a policy enforcer. However, the United States has historically been a leader in Internet governance. Any action taken by the U.S. government in Internet governance has a far-reaching impact on how other governments look to the Internet. Moreover, because of this impact, the U. S. government's positions in Internet governance, even if in an area that is wholly within the purview of the United States and not the global Internet, come under intense scrutiny. Thus, it is of utmost importance that the debate on Internet availability and accessibility is based on technological facts, capabilities, and projected growth.

Equally important, the framework for network operation and traffic management needs to rest on clear and simple principles, whether established by the market, by law, by regulation, or jointly by all three. These include the following:

- **Competition**, to assure user choice, to compel efficient pricing, and to stimulate innovation
- **Minimal regulation**, to remedy market failure and to encourage maximum investment
- **Nondiscrimination**, as to originator, consumer, content, applications, or services
- **Service levels**, to accommodate different user needs for bandwidth, latency (see below) and availability
- **Transparency**, to specify service levels and disclose methods of traffic management
- **Performance measures**, to provide quantitative metrics for evaluating service at multiple levels.

The stakeholders in the Internet are varied. End-users, enterprises and application providers access the Internet and provide content. Infrastructure providers offer wireless, wire-line and cable access. Some providers interconnect access networks. Other providers interconnect core networks. Content distribution networks accelerate delivery of Internet applications and content. And local, state and federal legislators and regulators strive to serve the public interest, as well.

A perfect market with perfect information keeps actors from extracting unfair rents or impairing service. However, as this paper indicates, there have been times when Internet service providers, particularly in a monopolistic situation, have unfairly blocked applications that compete with their own offering. Conversely, there have been legitimate situations where a service provider had to manage bandwidth usage to continue offering service. One way of both managing a network

and allowing a user to decide what service to use is to have a robust set of performance metrics. Performance metrics are a crucial first step in ensuring a predictable market, while mitigating the contentious issue of network neutrality.

This paper offers the following conclusions:

- Fair network traffic management practices are urgently needed in today's Internet.
- Performance measures constitute the basis of network traffic management.
- Technological metrics offer advantages to all stakeholders in Internet service and use.

INTRODUCTION

By its very design, central control of the Internet is untenable. Although the Internet originated in the United States, that system now connects world commerce, governments and individuals, melding many separate networks that follow many different models and policies. Service providers and enterprises control their own networks, with no central authority or operator governing the Internet. Nevertheless, U. S. policies and practices strongly affect the practices of other countries, in part, because many of the Internet's key nodes have remained in the United States, as the Internet has continued to grow dramatically.

One of the pressing Internet questions facing policy-makers today is whether service providers may differentiate their services to achieve optimum performance.

This question implies a host of sub-questions: What performance do various users need and expect? How should that performance be measured? Who should pay for any such performance differentiation? And, in the absence of performance differentiation, what incentives should apply to discourage congestion, or practices otherwise causing poor performance for other users?

The technical term for providing performance differentiation is *network traffic management (NTM)*. The paper argues that if non-discriminatory metrics and competitive pricing structures are applied, network traffic management can prove beneficial to all parties. If regulators and law-makers require transparent traffic management practices, their use will promote the development of the next generation of communication and collaboration technology. Parenthetically, traffic management, or traffic engineering, is not new. It has long been used in the central portions of networks to smooth traffic flow, thus allowing providers to build to average, rather than peak, demand.

The twin issues of how Internet traffic is managed, and what performance should be expected, are currently resolved using vague rules and interpretations by both service providers and regulators. This localized way of managing can make the implications of network traffic management difficult to understand. Wharton Professor Kevin Werbach said: *"...we don't have a regulatory structure for that new, converged, broadband Internet infrastructure."*[1] Clearly, the Internet is very different from highly regulated legacy telephone networks, and regulations appropriate for the Internet must evolve with the technology.

Further illustration of this point can be found in the stark architecture differences between the Internet and telephone networks. The Internet's architecture is based on a decentralized infrastructure — a network of many networks. Each of these networks has one major feature in common — the ability to dynamically receive or send (i.e., *"route"*) data packets to and from specified networks. The telephone network is very monolithic by comparison, as it is based on hardwired interconnections and switches, with few dynamic properties apart from its signaling system, which is separate from its voice channels.

This subtle but important distinction offers some explanation as to why regulation from the telephone era has struggled to keep up with fast-changing Internet-era networks and technology.

---

1 *New Rules for a New Age: Creating an 'Economic Stimulus Agency' out of the FCC* Knowledge@Wharton, April 01, 2009; Available http://knowledge.wharton.upenn.edu/article.cfm?articleid=2197

DIGITAL CONVERGENCE DRIVES THE INTERNET

Over the past decade, the Internet has grown to supplement and supplant numerous legacy platforms for communication: the traditional voice network, networks for small packets of low-latency data, and even television networks. The IP-based, packet-switched Internet of today is becoming the world's primary platform for multimedia point-to-point and broadcast communications. With digital convergence in all usage sectors, digital data makes the Internet an attractive tool for all.

As the Internet replaces specialized networks, there are fundamental new considerations that must go into managing it. Given the *always-on* and *24-by-7* nature of the Internet, its management must work well technically and without discriminating as to the user or the message.[2] Further, management cannot be static or lack flexibility, either of which risks inhibiting growth. The management of the Internet must tolerate new applications that demand bandwidth, such as voice over internet protocol (VoIP), cloud computing, video uploads and downloads, and the inevitable advent of entertainment high definition television in both 2D and 3D. Clear procedures for managing traffic must be pervasive and accepted by all providers and users.

FUNDAMENTAL QUESTIONS OF NETWORK TRAFFIC MANAGEMENT

Network management concerns are driven by issues that affect business, especially economic, regulatory, legal and operational issues. Regardless, the fundamental question of network traffic management is simple, as noted above: At what reliability and speed must service providers deliver a particular item of content at a particular time to a particular user? While there are some accepted practices, the current state of network traffic management is quite young, inconsistent, and in need of oversight.

The question posed provokes only more questions:

- Should all types of transmissions (data, audio, images, and video) be treated equally?
- Must an Internet service provider support applications that directly compete with its own product offerings?
- Should user prices reflect the cost of service or the value?
- How should a user's quality of service (QoS) be measured?
- Can one user's QoS be improved without diminishing another's?
- Are there ways to minimize impaired availability of service?

An artificial differentiation has also been introduced: a distinction, with regard to NTM, between wireless and wire-line communications. In fact, network traffic management is applicable to both, especially for the Internet. How a network is built and how content is delivered does not alter the need for network management, though it may alter the particular techniques.

2 Network traffic management and next-generation Internet issues should not be confused with the polarizing issue known as *"net neutrality,"* that connotes the inadmissibility of preferential treatment of data packets. The distinction is further described in the section: *Relevance to Net Neutrality*.

STAKEHOLDERS AND THEIR INTERESTS

Stakeholders in the network traffic management debate are numerous but can be grouped by their functions and aims.

Network Access and Transport Providers, who get packets onto the Internet and across the Internet:

- Internet Service Providers (ISPs) deliver Internet service to end-users. End-users can be consumers, enterprises or other networks. They need to provide superior quality of service to preserve and expand their subscriber base.
- Network Providers operate networks that connect ISPs together. Many network providers are also ISPs. Examples include incumbent local exchange carriers, multi-system cable operators and backbone providers. They face needed capital expansion of their networks to meet increasing demand.

Content and Application Providers, who connect data servers to the Internet, or deploy software for peer-to-peer servers across the Internet. All the following companies require prompt, reliable delivery suited to the nature of their material:

- Companies providing traditional content, such as Internet broadcast of radio, movie and television distribution, news and information
- Companies and individuals providing end-user applications such as shopping, Internet search, e-mail, and calendar management
- Companies and individuals providing enterprise applications, also known as Software-as-a-Service or SaaS, such as enterprise applications running outside the enterprise network in a hosted computing facility
- Companies providing application infrastructure applications, also known as Platform-as-a-Service or PaaS, such as computational systems for developing applications; and Infrastructure-as-a-Service or IaaS, such as computer and storage facilities companies providing content and application caching and acceleration also known as Content Distribution Networks or CDNs.

Content and Applications Users, who request content or application processing across the Internet:

- Individual and enterprise end-users. Individuals wish to download and upload material conveniently and affordably. Enterprises rely on the Internet for effective communication with customers, suppliers, and employees.
- Application consolidators, aggregating information from other Internet-connected information and application sources, often at the user device, not necessarily in the network.

Legislators, who propose legislation that promotes the economic and social benefits of the Internet.

Regulators, who administer such legislation, establishing regulations and guidance, when necessary and within their authorities.

In the current climate of consolidation, the lines between these categories of stakeholders are not always clear. For example, some companies can be both content providers and Internet service providers.[3] Similarly, different categories can share similar interests: facilities providers and content providers, for example, benefit from the expansion of CDNs, as they both decrease the backbone infrastructure required and improve user experience.

Historically, ISPs and facilities providers have opposed regulations limiting their ability to manage their own networks. Conversely, corporate content providers have sought to limit network management practices to allow for the flexible development of services that compete with offerings of the ISPs themselves. As Bauer, Clark and Lehr note[4], however, corporate content providers are among the primary drivers of both the expansion of Internet infrastructure and the need for network traffic management to ensure more robust, reliable broadband service. A sound network traffic management policy must address and support these complex interests.

3 J. Kosman, K. Whitehouse, and C. Atkinson. *"FCC readies Comcast net neutrality trap." New York Post.* August 9, 2010.

4 S. Bauer, D. Clark, W. Lehr. *"The Evolution of Internet Congestion."* 37th Research Conference on Communication, Information and Internet Policy (www.tprcweb.com). Arlington, VA. September 2009. Available http://mitas.csail.mit.edu/papers/Bauer_Clark_Lehr_2009.pdf.

SERVICE PROVIDER–DRIVEN NETWORK MANAGEMENT

As more bandwidth becomes available to the Internet, developers and customers will always find new ways to use it. Higher definition video, broader use of remote application hosting and increased peer-to-peer traffic will find a way, in the end, of consuming all the bandwidth that is available — especially during peak periods[5].

To respond to this bandwidth utilization reality, service providers rely on a range of network management tools and techniques, as well as other approaches, to service these demands on a dynamic and on-going basis.

For example, service providers must be able to assure that they can maintain the quality of their services over short time periods (bursting or increasing packet size, for example), and also over the longer term (setting up a larger connection to the Internet to handle peak demand periods). Generally speaking, assuring quality takes the form of infrastructure improvements that scale to meet demand, either by using new facilities or by creating more efficiencies within existing facilities. The service provider industry faces these operational challenges daily.

To address such demands, providers can lay more fiber or acquire more spectrum — but doing so can sometimes be too costly, depending on circumstances. In a backbone or core network, adding capacity or routes can be relatively cost effective. But in the access network, especially wireless, it is often presumed that the available resources will eventually be used at capacity. Additionally, deployment costs are often substantially higher in the access network, due to its distinctive goal: overlaying entire geographic regions with connectivity to individuals. Because providers must use available bandwidth more efficiently to face the growing challenges of delivery and economy, using NTM can become crucial to the interests of providers large and small.

At a high level, service provider network management is also driven by the type of ISP involved — network core or backbone provider, or user-facing access providers. Both can have very different network management needs. Additionally, the type of customers an ISP has often governs specific needs of that ISP's network. For example, sophisticated commercial users may have specific, well-defined requirements, while home users may have widely varying ideas of what is, or is not, an important service.

Across this diverse set of consumers and circumstances, without building new infrastructure, cost effective solutions are usually found through NTM. Assuring network "quality of delivery," and operation that supports the underlying broadband products and services offered, often depend on basic network management practices or technologies, such as load balancing, spam filtering and bandwidth monitoring.

In a wider context, network management used by ISPs falls into one or more of the following three general NTM "categories."

5 Jon M. Peha, *"The Benefits and Risks of Mandating Network Neutrality, and the Quest for a Balanced Policy,"* Proceedings of 34th Telecommunications Policy Research Conference (TPRC), 2006.

- **Economics Driven** (increasing revenue and lowering expenses). The main goal in an economics-driven NTM context is to generate greater cash flow, one based on implementing NTM practices that balance the cost of operations against ability to provide quality services to the clients. NTM practices such as bandwidth engineering, for example, assure a good subscriber ratio.

- **Regulatory or Statutorily Driven** (assuring legal compliance). The Communications Assistance for Law Enforcement Act (CALEA) of 1994 is an example of a legally motivated network management requirement. Under CALEA, e-mails, VoIP phone calls, and other forms of electronic communications may be subject to CALEA's "lawful intercept" requirements. Using the right network management tools and procedures, ISPs can usually easily record and supply the electronic correspondence required by law.

- **Operationally Driven** (maintaining operational quality). Networks, whether wired or wireless, are dynamic data transmission media. Over time, they can have widely varying data volume, type and duty cycle fluctuations that can be both random and significant, as far as network operations are concerned. As such, the network operator must be able to take steps to assure a level of quality. Such steps may be to bring a redundant Internet link online dynamically during peak demand periods — or even to close a port on a user's service automatically, because the user has an e-mail virus that is impacting the e-mail gateway the ISP maintains.

In today's Internet, efficiency may sometimes be gained through relatively small investment in time and infrastructure. Because unnecessary expenses adversely impact the cost to the consumer and a provider's competitiveness, network traffic management is clearly needed to avoid such inefficiencies.

ESTABLISHING PRINCIPLES OF NETWORK OPERATION AND TRAFFIC MANAGEMENT

The framework for network operation and traffic management needs to rest on clear and simple principles, whether established by the market, by law, by regulation, or jointly by all three. These include the following:

- **Competition**, to assure user choice, to compel efficient pricing, and to stimulate innovation
- **Minimal regulation**, to remedy market abuse and to encourage maximum investment
- **Nondiscrimination**, as to originator, consumer, content, applications, or services
- **Service levels**, to accommodate different user needs for bandwidth, latency, and availability (see below)
- **Transparency**, to specify service levels and disclose methods of traffic management
- **Performance measures**, to provide quantitative metrics for evaluating service at multiple levels.

The remainder of this paper discusses the metrics necessary to provide network services based on these principles, as well as current and future approaches for utilizing those metrics.

PERFORMANCE MEASURES AND QUALITY OF SERVICE[6]

The International Telecommunication Union (ITU), the part of the United Nations that addresses international telecommunications policy and standards issues, defines QoS as "*the collective effect of service performance which determines the degree of satisfaction of a user of the service*" in its ITU-T Recommendation E.800[7]. ITU-T Recommendation E.802 for users and ISPs[8] includes quality of service (QoS) criteria.

6 " Internet QoS: Pieces of the Puzzle" Aref Meddeb, ISITCom; *IEEE Communications Magazine*; January 2010; pp 86-94.

7 ITU-T Rec. E.800, *"Terms and Definitions Related to Quality of Service and Network Performance Including Dependability,"* 1994 (revised in 2008).

8 "Quality of Service Management for ISPs: A Model and Implementation Methodology Based on the ITU-T Recommendation E.802 Framework," Eva Ibarrola, Fidel Liberal, Armando Ferro and Jin Xiao, *IEEE Communications Magazine*, vol. 48, no. 2, *February 2010*, pp.146-153.

To guarantee QoS , the network provider reserves network resources for the requesting customer, according to a service level agreement (SLA)[9], a contract between the user and the provider. These resources are reserved and allocated based on several parameters: bandwidth or throughput, packet loss, latency, jitter, and availability (uptime). These five performance measures are used most frequently in network traffic management as quality of service measures.

1. Bandwidth or Throughput

   The easiest metric to understand in network QoS is bandwidth or throughput. Simply put, in a QoS environment, the service provider guarantees certain send and receive rate for data. Above the send rate, the sender can experience any number of agreement enforcement behaviors to prevent overuse of network resources. Below this rate, traffic is allowed to pass through the network at the same rate at which it is sent.

   Service providers generally enforce bandwidth restrictions at the edge of the network; that is, as close to the sender or receiver as possible. The practice of *throttling,* which has gained substantial attention in the consumer ISP setting, refers to a service provider's practice of restricting the rate at which a user can upload or download information, despite the availability of network resources.

   If throttling and other bandwidth restrictions fail to limit appropriately the volume and demands of traffic flowing through the core of the service provider network, the guaranteed level of service performance for other users may not be attainable.

   Note that these mechanisms assume that all traffic stays within a single service provider's network. However, if a server is outside of the service provider's network, then congestion at the interconnection point, the Internet backbone, or in the server's access network will reduce the actual speed of the data transfer from the server to the user. This congestion, and resulting poor QoS, is entirely outside the control of the access network provider.

2. Packet Loss

   The Internet does not inherently guarantee that when a packet is sent it will be received by its intended recipient. There are many reasons why packets can be lost, including *physical interference* and *overfilled queues*.

   Physical Interference. Electrical or magnetic interference from other devices can cause transmissions to be unintentionally distorted along their paths, possibly rendering them impossible to route to their final destinations. Wireless transmissions are particularly susceptible to such interference. Additionally, physical breakage of links can introduce physical interference, causing packets to be lost.

9 According to Meddeb in *IEEE Communications Magazine,* Jan. 2010, Service Level Agreement, or SLA, is generally business oriented and does not deal with technical aspects. Its technical specifications are commonly described in the service level specification (SLS) and service level objective (SLO). An SLS is defined as an operational guideline for the implementation of the service. An SLO is a subset of an SLS, which specifies, among other things, the goals to be achieved by the SLS. As implemented by ISPs so far, SLAs for residential users are usually referred to as terms of service. In general, there is no mention of traffic prioritization (CoS), and little assurance on service quality. In fact, the residential SLA serves more to limit the ISPs' responsibility, rather than to protect residential users. On the other hand, corporate SLAs are usually more stringent than residential ones. In general, there are objectives on network availability, latency, jitter, and packet loss.

Overfilled queues. When a network device, such as a router or switch, receives a packet, it may not be immediately able to send that packet along its next link. This delay frequently occurs if a large number of packets have come from a variety of sources but are headed to the same destination. In this case, a device may need to temporarily store the received packet in its memory, with subsequent packets stored in the order in which they were received. This memory, referred to as a queue, is a finite resource, since routers and switches have limited memory. If, when a packet is received, it cannot be immediately forwarded, and if there is no queue space available for it, the router may drop the packet. This non-delivery will occur most frequently in highly congested sections of networks.

Many protocols, such as Transmission Control Protocol (TCP), implement mechanisms to guard against packet loss. By using such mechanisms, large files can still be transferred reliably across the Internet in the face of packet loss. Other protocols, such as User Datagram Protocol (UDP), are not designed to guard against packet loss. As such they are utilized in situations where packet transfer integrity can be sacrificed for the sake of performance, or where recovery from packet loss is the responsibility of the application rather than the network protocol. Examples where performance is more important than reliability of transfer include live streaming applications, video, and VoIP. The electric power grid is an example where applications can capitalize on repetitive telemetry of system measurements to avoid the need for packet loss safeguards in the network protocols.

3. Latency

Latency is a measure of the time between sending and receiving information. It consists primarily of propagation and processing delays.

Propagation Delay. Regardless of what technology is used for transmission, some amount of time is necessary for a bit to travel over a network link. Known as propagation delay, this link traversal time is dependent on the length of the connection and the medium, whether wireless, optical, or electronic.

Processing Delay. In addition to propagation delay, processing delay at each device or node that a packet traverses can introduce additional latency**.** This latency can impact the performance of applications that depend on near-simultaneous transmission and reception of data, such as real-time voice or video applications. For example, if high latency links are used, perceptible delay could be observed in a video conference between one user's statement and another user's reaction.

For non-real-time applications, such as e-mail, latency has minimal importance relative to bandwidth constraints.

4. Jitter

Jitter is the variation in the end-to-end latency observed by packets following a particular path in a network. One can think of latency itself as the average of end-to-end delay and jitter as the variance in this average. Jitter is an important QoS metric because of its implications for real-time web applications, For example, if a receiving computer doesn't know exactly when it will

finish receiving an entire video screen because of jitter, it has to account for that uncertainty by correspondingly increasing the time before it displays that screen to allow for smooth presentation of the entire video.

While some factors that produce latency do not change over time, such as link propagation delay, other less-dominant factors can change. For example, in discussing packet loss, we referred to the impact that network congestion can have on device queue sizes, causing forwarding devices to not have enough memory to store incoming packets. Similarly, if a packet arrives at a forwarding node and there are many packets to be forwarded ahead of it, additional delay can be incurred. This delay varies based on network conditions and is the primary source of jitter.

Network applications with strict reliability constraints must account for this variation in latency to ensure smooth performance. High quality of service in the network can limit the application's degradation.

Table 1. Sample Network Requirements by Application

Application	**Bandwidth** (Mb/s)	**Acceptable Packet Loss**	**Target Latency** (milliseconds)	**Target Jitter** (milliseconds)
VoIP	1 – 5	Up to 1%	150	50
Telepresence	8-10	Up to 0.05%	150	30
Ordinary Power System Control	Negligible	Generally much greater than 1%	2000 -6000	Not applicable

Sources:

T. J. Kostas, M. S. Borella, I. Sidhu, G. M. Schuster, J. Grabiec, and J. Mahler, "Real-time voice over packet-switched networks ", *IEEE Network*, vol. 12, no. 1, pp. 18 - 27, Jan/Feb 1998.

T. Szigeti, C. Hattingh. "Quality of Service Design Overview". *End-to-End QoS Network Design*. Cisco Press: Nov. 2004.

Sample network requirements for availability are not generally published, but should be established. For example, extremely high requirements are observed for the public switched telephone network. Cable television networks also strive for high availability.

5. Availability (Uptime)

Availability is the percentage of time a connection is available from the user to the Internet. Three major categories affect availability. The first includes natural disasters, such as flooding of switching centers, cable cuts, power failures and kinetic attacks on infrastructure. The second has to do with the network routing infrastructure, such as network congestion, equipment failure, downtime for maintenance and upgrades, or misconfiguration of network equipment. The last has to do with business issues, such as tiered service, capped service and throttled service. This last availability constraint is often negotiated with the service provider, who may charge additional fees for priority packet forwarding or guaranteed uptime.

A JITTER METRIC EXAMPLE

In some cases it is possible to improve the jitter metric for jitter-sensitive traffic, while having no noticeable impact on other users.

For example, consider a user downloading a large file using the file transfer protocol (FTP). During the file transfer, the user also starts a VoIP call. Packets from the file transfer will be in the user's router, ahead of the packets from the VoIP session. This *"head-of-line queuing"* is not an issue for bulk data transfer. However, having the VoIP packets wait for the FTP packets can introduce significant jitter. Jitter is the issue here, because the number of FTP packets ahead of the VoIP packets is random.

Most users would not mind having their file transfer slightly delayed to be able to use the Internet for real-time, interactive streaming media, such as VoIP. While research shows that QoS mechanisms in the backbone rarely make a difference to delivered QoS, QoS in the access network can have a significant impact on user QoS, because access networks tend to have considerably less bandwidth and higher latency than core networks.

Additionally, the ITU recommendations acknowledge that these network metrics must be viewed within context, since the importance of each metric may vary with application or data type. New QoS terms that have been defined include quality of experience (QoE), quality of business (QoB), and quality of perception (QoP).[10]

USER-SELECTED SERVICE LEVELS

Several models adopted in other industries for standardizing requirements are worthy of consideration for transplantation to Internet traffic management. For example:

- The *food industry* has standardized requirements and labeling for shelf life.
- The *shipping industry* allows different rates for different delivery times.

The core metrics of quality of service — bandwidth, latency, jitter, packet loss and availability — are fundamental and clearly defined. As such, they are open to very little interpretation. However, today's service providers present consumers with substantially fewer concrete metrics, often couched in vague terms:

- Speeds of "*up to*" 50 megabits/second
- Qualitative marketing-driven terms, such as *"Powerboost,"* or plan classifications, such as *"Fast/Faster/Fastest"*
- Unspecified upload and download speeds.

10 Op.cit., "Quality of Service Management for ISPs," p146.

The transience of Internet congestion makes it difficult for a service provider to identify exactly what QoS it will be able to deliver to a particular customer at any point in time[11]. Uniform QoS is especially difficult to maintain when subscribers are sharing available bandwidth, such as cable broadband channels or Wi-Fi. The prevailing practice for addressing this issue is to provide upper bounds on how much performance can be expected; rarely is a corresponding lower bound presented. A lack of range in the choice of service classes forces a one-size-fits-all approach; a lack of transparency and consistency regarding quality of service and performance causes frustration for subscribers. Consumers and providers alike need better metrics for buying and selling Internet service to improve user experience and enhance network utilization. NTM needs standards and well-accepted guidelines that can be used to define service classes that support major use cases with both critical and non-critical delivery needs, such as VoIP, telepresence, streaming video, general web browsing, and so on.

Subscribers benefit from clearer metrics by gaining better understanding of the services they purchase, assuming the metrics are presented to the consumer in a transparent way that is standardized across the industry. Rather than only knowing best case metrics, and therefore being unable to depend on any minimum service guarantee for essential services like VoIP, subscribers might choose higher (or lower) classes of service to meet their own particular needs. Providers benefit from clearer metrics by gaining the ability to prioritize traffic according to a user's explicit service request, avoiding complex issues of privacy, censorship and favoritism.

User-selected, tiered-service levels accommodating widely different user needs can help ISPs manage traffic. ISP offerings can be optional, although competition among providers would likely bring QoS improvement, especially if failure to offer the desired performance results in subscriber loss. Transparency and minimal regulation could assure that the tiers applied to transmission only and did not discriminate as to content, applications, or users. If offered, tiered services would avoid the need for brute-force blocking, filtering, and throttling.

Tiered approaches have been proposed for the purpose of grouping ranges of quality of service together into simple, easily-understood options. Tiered service levels can create opportunities for a variety of innovative value and economic models but need standards and safeguards to create an open and fair market.

Along the lines of the examples of the food and shipping industries, where logistics are so important, here are a few keys to the tiered approach for ISPs.

- ISPs provide consistent labeling for key metrics by tier — similar to food nutrition labels.
- Traffic data includes subscriber-selected tiers so that data can be routed and prioritized appropriately.
- ISPs set bandwidth and QoS tiers, pricing, and metrics tailored for different subscriber needs, allowing subscribers to choose their desired service plan. For example, a subscriber can choose to make a VoIP call using real time QoS (an important call), or choose to use a lower QoS level to conserve QoS quota or cost, and the decision can be facilitated by the application itself.
- Market-oriented models that exist or have been tried include tiered pricing tied with availability.

11 S. Bauer, D. Clark, W. Lehr. "The Evolution of Internet Congestion." 37th Research Conference on Communication, Information and Internet Policy (www.tprcweb.com), Arlington, VA, September 2009. Available http://mitas.csail.mit.edu/papers/Bauer_Clark_Lehr_2009.pdf.

Effectively communicating these models to the average consumer will require study in the areas of marketing and user experience. It will produce challenges for both the application provider and the service provider, but effective communication of service levels will reap substantial benefits for all.

MODELS FOR MANAGING TRAFFIC

Given the need for network traffic management noted above, two main approaches have evolved — managing traffic in the core and managing traffic at the edge. Though techniques in these areas are not mutually exclusive, service providers' goals in applying these techniques differ between the core and the edge.

Comprehensive end-to-end traffic management across the Internet is not currently feasible. Large consumers of service provider resources often employ service level agreements, however, to guarantee QoS within a single provider's network, illustrated in Figure 1. Given that much of today's Internet traffic does not stay within a single service provider's network, as illustrated in Figure 2, techniques may need to be developed to support end-to-end traffic management.[12]

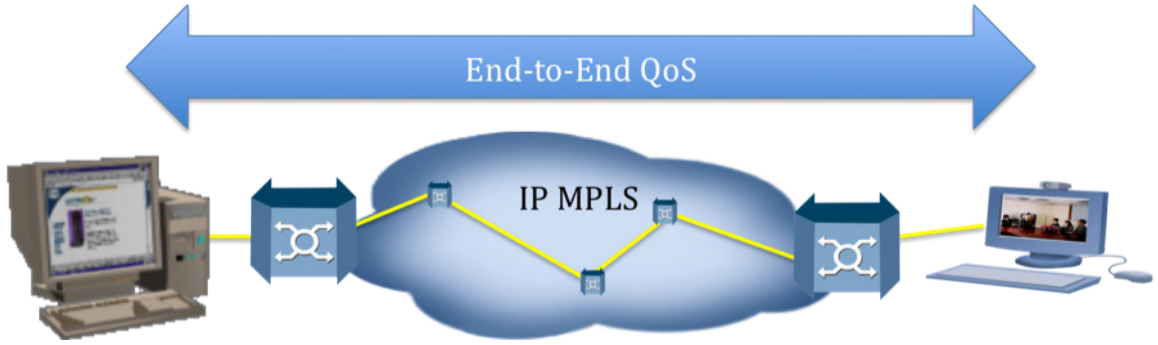


Figure 1. Managing Traffic: Single Carrier or ISP Backbone. (MPLS defined below)

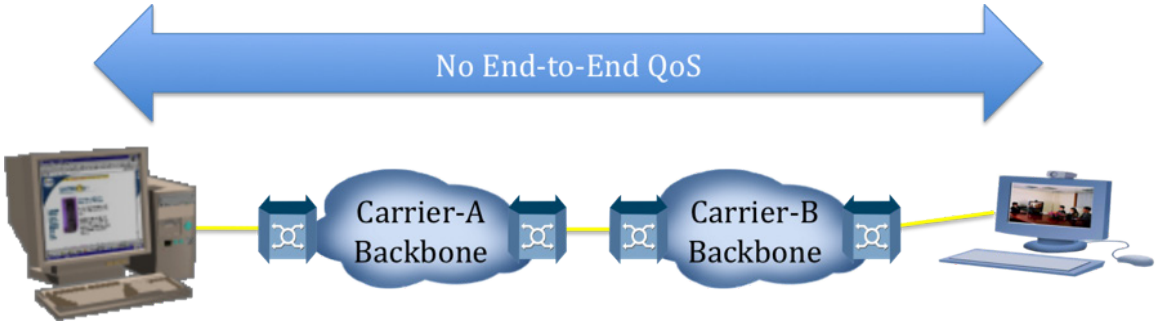


Figure 2. Multiple Backbones Traversed[13]

12 Testimony of David Clark, FCC Open Internet Workshop, January 13, 2010. Available http://www.openinternet.gov/workshops/innovation-investment-and-the-open-internet.html

13 [13] Figures 1 and 2 courtesy of IEEE-USA CCP member N. E. (Earl) Turner.

MANAGING TRAFFIC IN THE CORE

Service and facilities providers employ a variety of strategies in the core network to enhance network efficiency. Tactics include tagged priority routing (for example, Multiprotocol Label Switching (MPLS) where packets are labeled to facilitate network routing decisions), bundled QoS management, and latency management through priority queuing. When excessive network congestion surfaces, selective discarding of packets also becomes an important tool. Monitoring of the efficacy of these strategies is usually accomplished by analyzing the performance metrics discussed earlier.

These techniques allow service providers to optimize flow through the core network and to ensure that groups of users, perhaps united by geography, do not unnecessarily receive poor service. Studies have shown, however, that the current bottlenecks in service provider networks are in the access or edge, networks, not the core[14]. Similarly, the network traffic management debate has focused mainly on the access network; as such, the discussion of core network traffic management is limited in this paper.

MANAGING TRAFFIC AT THE EDGE

Contrary to its appellation as the access network, the edge of the service provider network, which connects individual users to the core, is actually the primary bottleneck in today's Internet. In the core network, Internet providers are able to add capacity along a single route to improve the experience for large numbers of users. At the edge, one possibility is to distribute servers more widely to handle such high-bandwidth traffic as movie downloads or entertainment television. However, the cost of expanding edge facilities for aggregating user communications to transfer traffic to the core is high because of widely dispersed wireless cell sites and subscribers to digital subscriber line, coaxial cable, and fiber. Accordingly, a high return on traffic management efforts is likely to occur at the edge. If such results prove largely satisfactory, they alleviate the problem of dealing with end-to end protocols and their complications across diverse core and ISP networks.

ISPs manage and monitor their networks for unusual patterns, misuse and fraud. It is in their best interest to keep the networks running smoothly. Service providers can save significant inter-carrier charges by blocking spam and botnets from within their networks, and likewise, can save on internal bandwidth needs by filtering spam and other attacks at the edge of their networks. Through their usage charges, the user pays for this service. The improved transport protocol, Internet Protocol version 6 (known as IPv6), offers expanded capabilities for dealing with traffic management, QoS, and priorities. Other attempts to manage the edge of the Internet, where usage demands are high, have resulted in the development of several techniques — blocking, filtering, and, most notably, throttling — some of which have been deployed in questionable ways. Clear, standardized metrics would help both providers and subscribers understand their usage patterns, delivered network performance, and how different use cases require different service levels. VoIP and movie downloads are two very different applications with different measures of quality and performance. Metrics can help determine the variance in costs among different service levels.

14 Jon M. Peha, *"The Benefits and Risks of Mandating Network Neutrality, and the Quest for a Balanced Policy,"* Proceedings of 34th Telecommunications Policy Research Conference (TPRC), 2006.

Finally, standardizing and updating metrics among service providers would ultimately provide choices to ISPs and subscribers, where none has existed before. Two examples follow.

> In 2008, the U.S. Broadband Coalition, (a broadbased coalition that includes IEEE-USA[15]) called for a national broadband strategy, which included setting up a Metrics Working Group to identify the best criteria to measure availability, adoption, cost, and speed for Internet services.[16]
>
> More recently, Google banded together with the New America Foundation's Open Technology Institute and the PlanetLab Consortium to support an open platform for academic researchers to enable examining overall network performance. This group's metrics platform is called the Measurement Lab (M-Lab). It intends to deploy Internet tools for users to determine the source of network connection slowdowns — including whether connections are being throttled by their ISP.[17] Actual attempts to extend metrics to the individual subscriber are sporadic, and not a requirement for service. A promising start is the FCC's Consumer Broadband Test tool, which gives subscribers a measure of their instantaneous bandwidth. The FCC also plans to deploy hardware-based testing tools throughout the national network.

None of these activities has netted metrics that the industry employs globally. The nature of the Internet and its origins still dictate voluntary adherence to any industry-wide metric, making it extremely important that suggested metrics fully support the principles of network traffic management to gain acceptance.

15 Coalition members listed at http://bb4us.net/id8.html.
16 The Coalition working group charter is available at http://bb4us.net/id19.html.
17 M-Lab2 Web100 based Network Diagnostic Tool (NDT) is found at: http://ndt.iupui.donar.measurement-lab.org:7123/.

COMCAST CASE SHOWS NEED FOR STANDARDIZED METRICS

The need for industry-wide accepted metrics is related to another component of network-edge traffic management — direct intervention or filtering based on network congestion, or other metrics. Metrics are often used to trigger proactive network management, presumably in the interest of users.

An example of this was doubtless some congestion metric used by Comcast to flag excessive peer-to-peer (P2P) traffic on its network for the well-known case (Comcast v. Federal Communications Commission, 2010}. The case involved Comcast's direct filtering of P2P file sharing traffic on its network. A minority of its users created a significant amount of P2P file sharing traffic across its network, creating very high network utilization. Such high utilization rates are not healthy for a functioning network, so Comcast took action to minimize the impact of the minority.

When the incident was cited by the Federal Communications Commission (FCC) as a violation, Comcast — as a network manager — stated that it took action because the P2P file sharing users were less than five percent of users that were significantly impacting the performance of the other 95 percent of users. As far as a majority of Comcast users were aware, the Comcast network was under a denial-of-service attack. One technology issue is there is no defined metric of what level of service is guaranteed to users.

The FCC lost its case in April 2010. The U.S. Court of Appeals for the District of Columbia Circuit said that the FCC had overstepped its authority in condemning an ISP for limiting P2P file sharing, even if it the traffic was legitimate. The Court's ruling on Comcast's appeal of the FCC sanction was not unexpected. The case underlined the fact that while the FCC retains the authority to protect consumers, the FCC has not established it has regulatory authority over Internet service providers.

MADISON RIVER DECREE SHOWS NEED FOR SERVICE PROVIDER RESPONSIBILITY

In contrast to the Comcast case, another incident illustrates blocking in the interest of the provider. Madison River Communications is a North Carolina local exchange carrier. They offered an unregulated ISP service running over their regulated DSL service. In early 2005, the Madison River ISP blocked the IP ports used by Vonage, a VoIP provider. At the time, 200 of Vonage's subscribers were Madison River ISP subscribers. The presumption was that since Madison River was primarily a telephone company, and Vonage directly competed with that service. Madison River blocked their competitor Vonage's subscribers.

The FCC sent Madison River a letter of inquiry on their network operations. In the end, Madison River and the FCC agreed to a Consent Decree. Madison River said they would not block VoIP ports, or otherwise impair VoIP services. Also part of the decree was an acknowledgement that the FCC was not asserting any particular rule, regulation, or law against Madison River. Madison River made a $15,000 payment, not a fine, to the Treasury to settle the investigation.

The Madison River event indicated the FCC's willingness to intervene on behalf of consumers, where there was a clear case of abuse of market position. However, since the case was never adjudicated, the FCC never established its authority to regulate ISP services.

RELEVANCE TO NET NEUTRALITY

Network traffic management has generated much discussion in both technical and political fora, and that discussion has included such notable participants as Tim Berners-Lee, World Wide Web inventor, and Vint Cerf, co-developer of the Internet Protocol.[18] Much of the debate has focused on the issue of *network neutrality*. Common Cause defines net neutrality as *"the principle that Internet users should be able to access any web content they choose and use any applications they choose, without restrictions or limitations imposed by their Internet service provider."*[19] Google CEO Eric Schmidt has said that net neutrality implies that *"if you have one data type, like video, you don't discriminate against one person's video in favor of another. It's OK to discriminate across different types."*[20]

Clearly, a universally accepted definition of net neutrality does not exist. Instead, the term serves as an umbrella term for a range of views on regulation of the Internet.

In its strictest interpretation, net neutrality precludes network traffic management through the use of QoS mechanisms, as they inherently prioritize some packets over others. In its broadest interpretation, the central issue of net neutrality is whether special alliances between application or content providers and ISPs are acceptable for Internet traffic.

By focusing on the parameters that define quality of service instead of particular applications or providers, it is possible to separate the content of data from service and network management of that data. Network traffic management techniques can be effective, while also being content- and application-agnostic. The Internet itself allows the separation of transport from the application and content transaction. This ability lets the broad, though still contentious, interpretation of network neutrality to be separated from the practice of network traffic management.

18 A Note to Google Users on Net Neutrality sidebar. http://www.google.com/help/netneutrality_letter.html

19 Common Cause on Media and Democracy addresses Net Neutrality here: http://www.commoncause.org/site/pp.asp?c=dkLNK1MQIwG&b=4773657

20 "Google's Schmidt on Verizon and Net neutrality," Ina Fried, cnet News, August 4, 2010 5:15pm PDT. http://news.cnet.com/8301-13860_3-20012723-56.html

CONCLUSIONS

The following conclusions emerge from the foregoing discussion of network traffic management and its potential for improvement of Internet operations.

Fair network traffic management practices have become urgently needed in today's Internet.

Competing interests have big stakes in the efficient, economical and fair management of the Internet. Users' accelerating demands on the Internet for ever-increasing performance create a need for network traffic management. So a sound network traffic management policy must address and support these complex interests to reach an acceptable resolution of the issues. The need for transparency is foremost in the use of any network traffic management mechanism. Transparency is achieved through standards and regulation in support of fair management practices bounded by clear and simple principles of competition, nondiscrimination, user choice, and performance measures.

Performance measures constitute the basis of network traffic management.

Network management allows more efficient use of limited transmission resources at modest investment than does uncritical expansion of facilities. Performance measures — in the form of metrics for bandwidth, packet loss, latency, jitter and availability — form the foundation for fair and transparent network traffic management. Given such metrics, user-selected service levels can be established to accommodate widely different user needs. Such service levels can facilitate consumer choice, orderly markets, and any necessary regulation.

Technological metrics offer advantages to all stakeholders in Internet service and use.

By focusing on purely technical performance measures, network traffic management stands apart from the contentious discrimination issues surrounding network neutrality. Network traffic management offers a technically sound compromise serving both network neutrality proponents and opponents. If standard, technologically neutral metrics and user-selected service levels are introduced and updated, network traffic management can prove beneficial to both the service provider and the user, supporting controlled, reliable user experience in the evolving Internet.

SUPPLEMENTAL REFERENCES

1. M. J. Copps. "Remarks at the State of the Net Preconference of the Congressional Internet Caucus." Net Preconference of the Congressional Internet Caucus. Washington, DC. 26 Jan. 2010. http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-295974A1.pdf

2. Federal Communications Commission. *Fiscal Year 2011 Budget* Estimates. Washington, DC: FCC, Feb. 2010. http://hraunfoss.fcc.gov/edocs_public/attachmatch/ DOC-296111A1.pdf

3. J. Gozdecki, A. Jajszczyk, and R. Stankiewicz, "Quality of Service Terminology in IP Networks," IEEE Communications Magazine, Mar. 2003.

4. D. Kravets. "Analysis: FCC Comcast Order is Open Invitation to Internet Filtering." *Wired.* 20 Aug 2008: http://www.wired.com/threatlevel/2008/08/analysis-fcc-co/ #ixzz0phPxhpUZ

5. A. Schatz. "FCC to Rule Comcast Can't Block Web Videos." *Wall Street Journal*. 28 Jul 2008: http://online.wsj.com/article/SB121720316961088595

6. C. Kang. "Court rules for Comcast over FCC in 'net neutrality' case." *Washington Post*. 7 Apr 2010: p. A01.

7. Federal Communications Commission. "File No. EB-08-IH-1518: Memorandum Opinion and Order, Formal Complaint of Free Press and Public Knowledge Against Comcast Corporation for Secretly Degrading Peer-to-Peer Applications." 20 Aug 2008: Available http://www.wired.com/images_blogs/threatlevel/ files/comcastdecision.pdf

8. Federal Communications Commission. *Connecting America: The National Broadband Plan*. Washington, DC: FCC, 16 Mar. 2010. http://www.broadband.gov/download-plan/

9. Federal Communications Commission. "Broadband Availability." Accessed 12 Jun 2010. http://www.broadband.gov/maps/ availability.htm

10. Google Inc. and Verizon Communications Inc. "Verizon-Google Legislative Framework Proposal." August 9, 2010. http://www.scribd.com/doc/35599242/Verizon-Google-Legislative-Framework-Proposal.

11. Federal Communications Commission. "Broadband Network Management." 7 Jan. 2009. http://www.fcc.gov/broadband_network_management/

APPENDIX: INTERNET SELF-REGULATION ORGANIZATIONS

- North American Network Operators' Group (NANOG) www.nanog.org
- American Registry for Internet Numbers (ARIN) www.arin.net
- Internet Engineering Task Force (IETF) www.ietf.org
- Messaging Anti-Abuse Working Group (MAAWG) www.maawg.org
- Internet Corporation for Assigned Names and Numbers (ICANN) www.icann.org

Together, these organizations can make sanctions against misbehaving network operators and network users; and thwart global cybercrime, identity theft, cyberextortion, botnets and spam.

Source: ITIF Brief before FCC GN Docket No. 09-191

2010 IEEE-USA CCP MEMBERSHIP ROSTER

Officers: Eric Burger, Chair; Dan Lubar, Vice Chair. **IEEE-USA Staff:** Deborah Rudolph

IEEE Society Members:

Charles Einolf, Broadcast Technology Society
Madeleine Glick, Photonics Society
Weibo Gong, Control Systems Society
John Healy, Reliability Society
James Isaak, Computer Society
Ferdo Ivanek, Microwave Theory & Techniques Society
Stanley Klein, Power & Energy Society
David Kunkee, Geoscience & Remote Sensing Society
Wayne C. Luplow, Consumer Electronics Society
Luke Maki, Technology Management Council
Dhawal Moghe, IEEE Region 5
John Newbury, Power & Energy Society
Kambiz Rahimi, IEEE Region 6
Tirumale Ramesh, IEEE Region 2
Curtis Siller, Communications Society
Christopher Stiller, Intelligent Transportation Systems Society
Wesley Snyder, Robotics & Automation Society
Erdem Topsakal, Engineering in Medicine & Biology Society
Gary Yen, Computational Intelligence Society

Members:

Michael Andrews
Jack Cole
Ann Ferriter
William Hayes
Richard Lamb
Stuart Lipoff
Michael Marcus
George Mattathil
Mike Nelson
Robert Powers
John M. Richardson
Paul L. Rinaldo
Emily Sopensky
Carl R. Stevenson

Corresponding Members:

Scott Atkinson
Stacey Banks
Judy Boggess
Thomas Cylkowski
Terry Davis
Upkar Dhaliwal
Hillary Elmore
Mark Enstrom
Matthew Ezovski
Rich Fruchterman
Jonathan Garruba
Norman Lerner
Philip Olamigoke
Wayne Pauley
Brennan Price
Al Reinhart
Raj Subbu
Norman Earl Turner

IEEE-USA
2001 L Street, NW, Suite 700
Washington, D.C. 20036
+1 202 530 8332
+1 202 785 0835 fax
Web: www.ieeeusa.org

IEEE-USA Staff: *Deborah Rudolph*
E-mail: *d.rudolph@ieee.org*