PRE-CONFERENCE DRAFT

April 29, 2000

Analyzing the Internet as a Common Pool Resource: The Problem of Network Congestion

Gerald Bernbom
Director, Research and Academic Computing
Indiana University

International Association for the Study of Common Property
IASCP 2000
Constituting the Commons: Crafting Sustainable Commons in the New Millennium
Bloomington, Indiana USA
May 31- June 4, 2000

Analyzing the Internet as a Common Pool Resource:
The Problem of Network Congestion

Abstract

The term "Internet" is used broadly to describe a global collection of multiple, inter-related resource facilities, each of which may be analyzed as a common pool resource (CPR). The Internet is comprised of a physical network infrastructure (network commons), a vast and distributed collection of information resources (information commons) that are accessible using this infrastructure, and a global communications forum (social commons) that is created and supported by the Internet.

This paper focuses on the physical network infrastructure and ways in which this aspect of the Internet functions and is managed as a commons, with particular attention to the CPR problem of overuse and network congestion. The paper discusses the major variables that influence problems of Internet appropriation and provision: the **physical world**, including the design principles and technical facilities of the Internet; the **individuals involved,** specifically the actors and the roles they play in provisioning, appropriation, and use of network resources; and the **rules in use** that govern the operation of the Internet, with primary attention given to network protocols as universal, mutually agreed to mechanisms that govern (among other things) the appropriation and use of network resources. The paper discusses the **CPR problem of network congestion**, with particular attention given to the technical facilities used to measure network use and determine conditions of congestion, and presents various ways of responding to overuse and network congestion, in three broad categories: (i) **increased provisioning**, (ii) **restricting access**, and (iii) innovation, by changing the **rules of appropriation**.

Analyzing the Internet as a Common Pool Resource:
The Problem of Network Congestion

> *"The Internet is the world's largest distributed system; it was designed and engineered for redundancy (it has an abundance of routes and connections) and resilience (it easily recovers from a mishap). The Internet is not a single company or a group of companies, nor even a single network. It is a worldwide mesh or matrix of hundreds of thousands of networks, owned and operated by hundreds of thousands of people in hundreds of countries, all interconnected by about 8,000 ISPs (Internet Service Providers). No single organization controls the Internet; not the U.N.; not the biggest ISPs; and the Internet has long since outgrown control by the U.S. government."* **[1]**

Introduction

The term "Internet" is used broadly to describe a global collection of multiple, inter-related resource facilities, each of which may be analyzed as a common pool resource (CPR). **[2]**

The most fundamental of these Internet resources is the **physical network infrastructure** (network commons) -- the optical fiber, copper wire, switches, routers, host computers, and end-user workstations, as well as the protocols and standards that allow these diverse and distributed facilities to be interconnected and communicate with one another.

The Internet refers as well to the **information resources** (information commons) -- the web-pages, text files, documents, images, databases, audio and video files, indexes, catalogs, and digital libraries that are accessible using this physical network infrastructure.

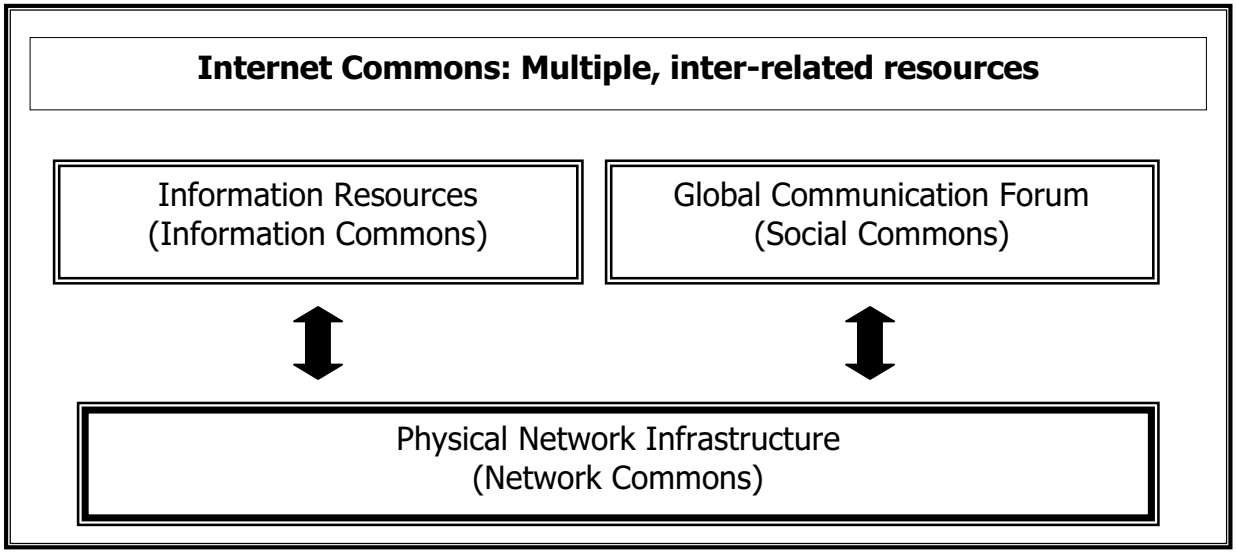
Finally, the Internet also describes a **global communication forum** (social commons) -- the e-mail messages, list-servers, news groups, discussion groups, chat rooms, and other facilities to enable communication between individuals and among groups, using the physical network infrastructure to send, receive, and store messages.

These resources are separate but interdependent. They work together but are not identical to one another. The information commons and the social commons depend upon the facilities of the physical network infrastructure to transmit messages, and to store and retrieve data. The utility

of the network commons is typically realized by users through their use of information resources or their participation in a communication forum.

Each of these resources has the characteristics of a CPR, namely it is a resource, without regard to ownership or property rights, for which exclusion of beneficiaries is difficult, costly, or technically infeasible, and in which exploitation by one user limits availability of the resource or reduces its value to others. **[3]** Each resource is also subject to CPR dilemmas and free-riding: either through overuse, or as a result of inadequate investment and maintenance necessary to sustain and enhance the resource.

Despite interdependence among the resources and their similarities as CPRs, each resource has different characteristics -- different attributes, rules, patterns of interaction, and outcomes -- and needs to be analyzed on its own. **[4]** For example, overuse in the social commons might display itself in lengthy, irrelevant or redundant communications that waste time and attention of all recipients or members of the affected discussion group. Overuse in the information commons might be characterized as pollution: inaccurate and unreliable data intermixed with higher-quality data, abundant but useless data overwhelming relevant and useful data, or misrepresentation of data for individual gain. Overuse of the physical network infrastructure manifests itself as congestion. Just as roads and highways become crowded or even impassible when too many cars and trucks try to use the same stretch of road at the same time, the Internet becomes congested when too much data is being sent at the same time on the same physical network link.



This paper discusses the physical network infrastructure and ways in which this aspect of the Internet functions and is managed as a commons. The paper focuses on the CPR problem of overuse and network congestion, which may be viewed as a problem of appropriation and as a problem of provision. Ostrom **[5]** observes that, "Both types of problems [appropriation and provision] are involved in every CPR to a greater or lesser extent." And that, "The structure of an appropriation problem or a provision problem will depend on the particular configuration of variables related to the *physical world*, the *rules in use*, and the attributes of the *individuals involved* in a specific setting." [emphasis added] The first five sections of this paper discuss these three variables as they define the CPR problem of network congestion.

- The first two sections discuss the **physical world** by looking at design principles (section 1) and technical facilities (section 2) of the Internet.
- The next two sections discuss the **individuals involved** by looking at the actors (section 3) and the roles they play (section 4) in provisioning, appropriation, and use of network resources.
- The paper then discusses the **rules in use** that govern the operation of the Internet (section 5), with primary attention given to network protocols as universal, mutually agreed to mechanisms that govern (among other things) the appropriation and use of network resources.

Section 6 of this paper discusses the **CPR problem of network congestion**, with particular attention given to the technical facilities used to measure network use and determine conditions of congestion. Section 7 discusses ways of responding to overuse and network congestion, in three broad categories: (i) **increased provisioning**, (ii) **restricting access**, and (iii) innovation, by changing the **rules of appropriation**.

1. The physical world: Internet design principles.

A number of fundamental principles are embodied in this design for the Internet that will have relevance to consideration of the physical network infrastructure as a common pool resource:

Distributed system. The Internet has no single owner and no central management. It was designed as a distributed system of cooperating but autonomous entities that can interconnect and share data by following some minimal set of rules. These rules, specified as network protocols, describe in detail the methods for sending data or instructions between independently-operating computers and computer networks. As discussed further in section 5, protocols function as self-enforcing agreements that are embedded in the technological infrastructure of the Internet itself to control and monitor network communications.

Network of networks. The Internet is not one network, but is comprised of many networks, each operating independently and autonomously. An autonomous network may be very small, belonging to a single college campus, research laboratory, government agency, or small business. Or it may be very large, operating on a national or international scale. Each such network participates in the Internet, and the Internet itself is the aggregate sum of these hundreds of thousands of networks.

Peer-to-peer. The Internet has no explicit hierarchy. Communication between any two networks is on a peer-to-peer basis, with each an equal participant in the inter-exchange of data and instructions. In practice, the Internet does have an implicit hierarchy, with some networks (usually large regional, national or international networks) serving as aggregation points for traffic from many smaller, more local networks, much as highway systems have secondary roads that carry local traffic and are feeders to larger, higher-capacity primary roads. But in the Internet all such arrangements are temporary, and the hierarchy changes frequently as local networks establish new peering relationships with one another and with different larger networks, and either continue or dissolve prior peering relationships. (Note: A peering relationship is an agreement between two networks to exchange traffic.)

Open standards. The network protocols that describe how computers exchange data and how networks interconnect are freely published. Software developers, computer manufacturers, network equipment manufacturers, and network service providers have an incentive to use the same set of common standards -- in order to assure that their software, hardware and services will interconnect and inter-operate with the rest of the Internet.

Best effort. The basic design of the Internet depends on taking any data to be transmitted (an e-mail message, a data file, a web-page, etc.), dividing it into smaller packets of data, and sending each packet independently toward its destination. There is no permanent path established from source to destination as there is for example when a telephone call is made and a single, enduring circuit is established for the duration of the call, and thus no guarantee of uninterrupted transfer of data. Each packet travels independently, and the network makes a best effort to deliver it to its intended destination. For example, if the network determines that one path to the destination is blocked or non-responsive, it may begin sending packets by an alternate path. The receiving system acknowledges packets as they are received and reassembles them into their original order. Some packets may not reach the destination and, depending on the application and the network protocol in use, the sending system may be asked to re-transmit them.

The autonomy of network participants, the rule-based requirements for interconnection, and the peer-to-peer nature of interconnection agreements establish some of the basic rules and decision-making arrangements for the management of the Internet as a commons. As will be discussed further below, the 'best effort' design principle is a significant physical attribute of the Internet in the analysis of overuse and congestion.

2. The physical world: technical facilities of the Internet.

The basic elements of the physical network infrastructure are:

- Data packets.
- Optical fiber links.
- Routers and network switches.
- End-nodes (e.g., host/server computers, personal computers).

In the Internet (i.e., any network using the Internet protocols) data to be transferred from one location to another across the network is broken down into smaller *data packets*. Each packet consists of a header, which specifies such information as the source (network address where the

packet originated) and destination (network address to which the packet is being sent), followed by the data itself.

Data packets are transmitted over *optical fiber links* that may span a few meters in a single building, or several hundred kilometers between cities or across national boundaries. (Less frequently, and usually over shorter distances, copper wire is in place of optical fiber.) Each link is characterized by the speed at which it can transport data, measured in megabits per second. This speed is used as a measure of the network's capacity, and the telecommunications industry has established several standard capacities in which network capacity is provisioned. Some examples:

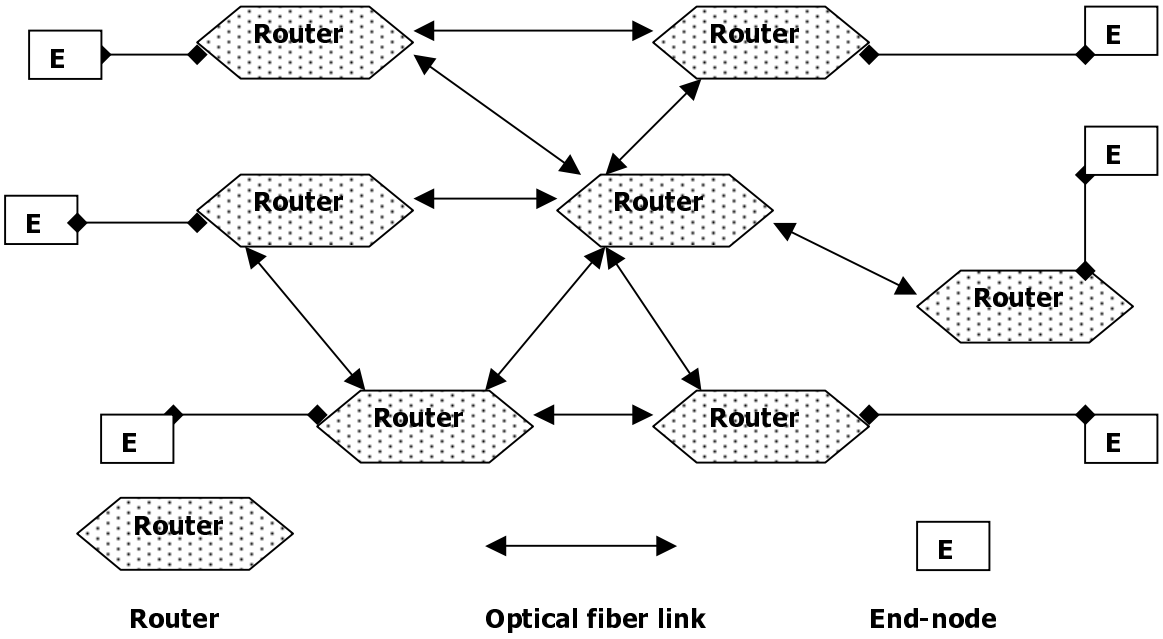
DS-1	1.544 megabits/sec.
DS-3	45 megabits/sec.
OC-3	155.52 megabits/sec.
OC-12	622.08 megabits/sec.
OC-48	2.488 gigabits/sec.

Important to note is that the capacity of a link is defined by the telecommunication signaling equipment at each end of the optical fiber, and is not primarily a characteristic of the fiber itself. As faster signaling equipment is designed and implemented, the capacity of a given physical fiber may be vastly increased (though the manufacturing and engineering quality of any specific fiber may set an upper limit).

Routers are the switch-devices that interconnect networks, locally or over wide areas. A router receives an incoming packet, examines its header to determine the destination, and transmits the packet along the path toward that destination -- either directly, if there is a direct connection to the destination, or else to the next router on the path toward that destination. A single packet may pass through many routers on its way from source to destination. Routers provide traffic direction (selecting one path from among many on which to send a packet), traffic control (responding to excess traffic by queuing packets or selectively discarding them), and filtering (selectively blocking packets or permitting them to pass, based on some characteristic of the packet itself).

End-nodes are the individual host/server computers and personal computers that are the sources and destinations of network traffic. Each end-node has a network interface that allows the computer to send and receive data. This network interface has a maximum rate at which it can transfer data, and thus is the terminal point at which the rate of network traffic flow is controlled. As a practical matter, when two or more end-nodes connect to the same physical network, the combined speeds of these network interfaces may be far in excess of the physical network's capacity to carry data.

Ownership of these facilities is a combination of public and private ownership. Optical fiber links are primarily owned by for-profit corporations but may be owned by national governments. The routers than manage network traffic are owned by for-profit corporations, governments and government agencies, universities and research centers, and non-governmental organizations. End-nodes are owned by all of these, and in addition are owned by private individuals. The next section of this paper describes some of the chief roles played by the owners and users of these facilities. Worth noting here is that the transit of a given packet from one end-node to another typically goes across the facilities of many owners, both public and private, and there is no direct relationship -- commercial or otherwise -- between the users who are using the facilities and all of the facility owners on whose network equipment this data is being transferred.



3. Actors in the Internet action arena.

There are five chief categories of actors involved in the management and use of the physical network infrastructure of the Internet:

- Internet service providers.
- Internet exchange points.
- "Local" network managers.
- Internet application software developers/providers.
- Internet users.

There exists a complex, nested set of relationships among these actors, characterized as:

- Multiple, inter-related **market arrangements** for the provision and use of network capacity and network services.
- Multiple, inter-related **bilateral and multi-lateral agreements** among Internet service providers for the exchange of network traffic.
- Multiple and sometimes overlapping **collective action arrangements** for the development of Internet standards and for research and development of new Internet services and protocols.

Internet service providers (ISPs) are companies or other agencies that provide one (or more) of these categories of Internet service: access, transit, and content. **[6]**

An *access provider* is an ISP that provides dialup or dedicated (e.g., leased-line, cable) Internet connectivity to business or residential users. Many access providers operate networks on a local or regional scale, but some are national or international. Local, regional or national telephone companies may also be Internet access providers.

A *transit provider*, sometimes called a backbone service provider, carries network traffic between other Internet service providers. Transit providers typically operate large-scale networks that are national or international. In the U.S., the major long-distance telephone companies (MCI, Sprint, AT&T, etc.) are also Internet transit providers.

A *content provider* focuses on specific types of Internet service, such as media services (e.g., Broadcast.com), portal sites (e.g., Yahoo), search engines (AltaVista), or others. Most

content providers do not operate any network facilities outside their own premises, instead buying services from access and transit providers, but some ISPs (America OnLine, for example) are both an access and a content provider. Even if they offer no physical network services whatever, content providers are important actors in this arena because the kinds of use they enable for end-users can have a tremendous effect on congestion and overuse.

An Internet Exchange Point (IXP) is an Internet service provider that allows autonomous and independently-administered networks to connect with one another in order exchange data in a managed environment. An IXP is a separate entity from the networks that connect to it and is usually administered independently of these networks. Autonomous networks that use an IXP are not connected directly to each other. Instead, each is attached to the IXP, which then interconnects the networks. This arrangement provides a neutral third-party broker between networks which may operate in competition with each other, or which may use different technology or operate by different rules internally. **[7]**

Internet Exchange Points are characterized as either public or private, but this refers not to ownership, but to their basic rules of operation. A public IXP places no restrictions on which networks are permitted to connect. There typically are some practical requirements such as a minimum connection speed for participating networks, or agreement by connecting networks to provide access to certain technical information for network management and performance. Such "public" Internet Exchanges are often operated by private, for-profit enterprises that charge connection fees. While these IXPs provide the technical facilities for interconnection in a neutral site, individual networks connecting to an exchange point also need to negotiate their own peering relationships with one another in order to establish the pathways for data exchange. Examples of major Internet Exchange Points in the U.S. are the Chicago NAP (Network Access Point), operated by Ameritech Advanced Data Services, and the MAE-East and MAE-West IXPs, operated by MCI WorldCom.

A private IXP operates under some policy criteria that select which networks may interconnect. Criteria might be the type of network, the size of the network in terms of traffic-volume, or the typical uses of the network (e.g., commercial, educational). One example of "private" IXPs are

the Federal Exchange Points (FIX-East and FIX-West), where only U.S government agency networks are allowed to interconnect. Another is the STAR-TAP (Science, Technology, and Research Transit Access Point), funded by the National Science Foundation to provide Internet interconnect services for international high-performance research and education networks. (For a discussion of technology and policy issues involving national and international Internet traffic exchange, see **[8]**.)

"Local" network providers/managers. Many large organizations and institutions -- universities, research laboratories, corporations -- manage and operate their own private networks, which are also connected to the global Internet. The network mangers in these organizations are effectively Internet Service Providers (ISPs), providing Internet access service to their own members (employees and customers, faculty and staff, etc.). Unlike commercial ISPs, which have a diverse clientele of residential and commercial customers, these local networks have a more homogeneous population of users who all share some institutional affiliation. The operation of these local networks may involve establishment and enforcement of institution-specific rules, sometimes codified in the form of an Acceptable Use Policy (AUP). An institutional AUP may define allowable uses of the network, prohibit some specific uses, establish expectations about privacy, or set limits on the amount of network resource that any user may consume. Either pursuant to, or separate from, the AUP a local network may also establish a "gateway" or "firewall" between itself and the rest of the Internet. Gateways and firewalls restrict access to internal resources from the outside or to block transit of specified data-types or transactions either into or out of the local, private network. They are often used in corporate networks to protect commercial data and proprietary information, and are increasingly used throughout the Internet as security mechanisms to detect and filter out potentially harmful network traffic. Gateways, firewalls and AUPs all provide potentially effective means of articulating and enforcing "local" rules of behavior on network use and misuse. Local network managers also make ongoing decisions about provisioning network capacity to the users of their own institutional network.

Internet application software developers/providers. Individuals do not use the Internet directly. Rather, they use various software applications -- e-mail programs, file transfer programs, web-

browsers, audio players, etc. -- which in turn make use of the physical network infrastructure of the Internet. Different programs have different characteristics in terms of how much network capacity they use, how large their data files are, how rapidly they transmit data across the network, etc. These characteristics are specified by application software developers and a user's only option in most cases is either to use, or not use, any of the available programs.

Internet users. Use of the Internet is measured in terms of the number of Internet hosts (computers connected to and accessible via the global Internet), or the number of individuals who use these computers to transmit or receive information via the Internet. Estimates of the number of Internet users vary widely. By some recent estimates there are more than 78 million Internet hosts and 288 million individuals with Internet access world-wide. **[9]**, **[10]**

4. Actors and their role in provisioning, appropriation, and use.

Transit providers and *access providers* are the chief actors in making decisions about provisioning network capacity -- for example, the number and speed of network links they install and operate (the equivalent of whether they build slow, secondary roads or high-speed, interstate highways), and the peering relationships they establish with other ISPs (whether they make direct connections to high-speed facilities or allow their traffic to take a roundabout route on slower links to its destination).

The information resources offered by *content providers* and the software programs used to access and retrieve this information directly affect the appropriation of network capacity by defining the volume of data available for access, and the amount and rate of data transferred in response to any specific access request.

Internet exchange points make decisions about the provisioning of network capacity -- for example, in the speed of network switches and routers with which they equip the IXP. They also influence the appropriation of network capacity, either directly, as when a "private" IXP establishes and enforces restrictions on network traffic, or indirectly, as when the IXP builds a framework to encourage multi-lateral peering agreements among ISPs, which can improve

network flow rates by offering a wider range of potential paths by which to reach its destination. (See, for example, **[11]**.

Internet application software developers make decisions about the appropriation of network capacity -- most significantly, their choice of network protocols and application design that affects the volume of data moved across the Internet and the rate at which data is transmitted.

Users make the ultimate decision about how they use the Internet -- for example, what application programs they run, how many messages and of what size they send, what information resources they retrieve, and what time of day they use these facilities. Given the very large number of Internet users the behavior of any individual user is almost imperceptible, but the aggregate behavior of thousands, or hundreds of thousands, of users is significant in terms of consuming network capacity.

5. Rules that govern the Internet.

The ability of the Internet to function at all, given the highly distributed nature of its ownership, operation, and management, is a result of the many sets of rules which users and providers adhere to. The most influential of these rules are the standards and protocols that define the Internet, which are discussed in detail following.

The original rules of the Internet, and the most fundamental rules that govern how networks and network-connected computers interact with one another, are expressed in the form of protocols, which are technical specifications for the exchange of data and instructions. The Internet is defined as the collection of inter-connected networks that use the Internet Protocol (IP) or, more broadly, the Transmission Control Protocol/Internet Protocol (TCP/IP) suite of protocols. Protocols define how messages are exchanged and tasks are performed **[12]**, such as:

- Making and receiving requests.
- Sending and receiving data.
- Rejecting requests or data.
- Acknowledging receipt of requests or data.

- Pausing and restarting transmissions.
- Setting transmission priorities.
- Handling error detection, correction, and retransmission.
- Numbering and sequencing packets in a transmission.
- Handling addressing and routing of packets.

Some protocols are closely tied to specific Internet services and are familiar to users. For example: the File Transfer Protocol (FTP) is used to transfer files from one computer to another across the Internet, and the HyperText Transfer Protocol (HTTP) is used by the World Wide Web to transfer web-pages from a server to a web-browser. Other protocols are largely hidden to users, and function to manage the operation and performance of the Internet. For example: the Routing Information Protocol (RIP) defines how routing information is exchange among nodes on the Internet, and the Open Shortest Path First (OSPF) protocol provides a method for determining the best path by which to route traffic based on its intended destination and the condition of the available network links (e.g., number of intervening routers, speed of each link, congestion on each link, etc.).

There is no single governing body that directs the development of protocols or enforces their use. A protocol achieves widespread use through a process of consensus among Internet users and providers, influenced substantially by the availability and utility of software programs that use the protocol. There are international forums through which new protocols are proposed and various, quasi-official rule-making bodies that approve or endorse protocols as "standards".

The Internet Engineering Task Force (IETF) is a "large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual." **[13]** The IETF meets three times a year and conducts activities through working groups in various areas (routing, security, etc.). A standards-setting process involves electronic publication and broad distribution of proposed standards via Requests for Comments (RFCs). There are three levels of standards defined by the IETF:

- Proposed standard - which entails a complete, credible specification of the standards and a demonstrated utility.
- Draft standard - which entails multiple, independent, interoperable implementations of the standard and at least limited operational capability to demonstrate that it works well.
- Standard - which entails demonstrated operational stability.

The World Wide Web Consortium (W3C) has a similar mission to the IETF, but with a more narrow focus (the Web rather than the entire Internet): " to lead the World Wide Web to its full potential by developing common protocols that promote its evolution and ensure its interoperability." **[14]** The W3C defines its role as a standards-making body as follows: " W3C contributes to efforts to standardize Web technologies by producing specifications (called "Recommendations") that describe the building blocks of the Web. W3C makes these Recommendations (and other technical reports freely available to all." **[15]**

Both the IETF and the W3C operate as self-governing bodies with broad participation of the Internet technical and engineering community from both public and private sector organizations. In the early days of the IETF, participants were almost entirely from the research and education sector in which the Internet had been developed. Even corporate participants were frequently from research laboratories or other R&D branches of large information technology firms.

With the rapid rise in popularity and commercial promise of the Internet, participation from the corporate sector has increased. But the influence of the information technology industry on Internet protocols and standards is less through their participation in standards-making bodies, and more through the product-oriented decisions they make about what protocols to use and what standards to follow in the design and development of network hardware and network-based application software. There is a potential conflict for these corporations between cooperation (using standard protocols to interoperate with the rest of the Internet) and competition (employing proprietary technology to gain market advantage).

The size and extent of the Internet, its large installed base of standards-compliant technology, and the network externalities that accompany global interoperability are all influences on

corporations to cooperate. Standards bodies have no enforcement authority. It is the nature of the Internet itself as an interoperable network of networks that enforces compliance with standard protocols.

While protocols and standards are the most basic rules of the Internet, and the only set of rules that operate universally, other rule-making occurs in more local settings. Internet exchange (IX) agreements, especially at "private" IXPs, and peering agreements set rules on how traffic is passed between networks, which can have an effect on availability or congestion. Acceptable use policies may also be in force in selected networks within the Internet, setting limits on types or amounts of use.

6. The CPR problem: overuse and congestion.

The history of the Internet has been a history of increasing use, periodic congestion, relief (through increased provisioning of network capacity), followed by further increases in use. In order to stay on top of use and congestion, and to be alert to possible network problems and failures, measurement of network performance is of increasing importance to a number of organizations, both public and private.

The National Laboratory for Applied Network Research (NLANR), with funding from the U.S. National Science Foundation (NSF), provides technical, engineering, and traffic analysis support for NSF High Performance Connections sites and other high-performance network service providers. Through a monitoring program of approximately 100 network locations in the U.S., NLANR measures network performance and congestion and assesses network topology. **[16]**

Matrix Information and Directory Services (MIDS) is a private corporation that provides Internet performance benchmarking and diagnosis services, including the Matrix Internet Quotient (Matrix IQ) which measures performance and congestion of links to several thousand destinations in the networks of Internet Service Providers world-wide. MIDS freely publishes a number of its performance analysis reports, while offering others on a for-fee basis. **[17]**

Individual network operations centers (NOCs) routinely measure performance and congestion of their own networks. Some of these NOCs freely publish the results of their analyses. Monthly reports of network performance and utilization are published by MCI WorldCom, which manages and coordinates the NSF-funded very high-speed Backbone Network Service (vBNS) high-performance research and education network. **[18]** Indiana University, which manages the network operations center for the Abilene/Internet2 backbone network, provides online, real-time measurements of network utilization along each link in the Abilene network via the "Abilene weather map" **[19]**. (Abilene is a high-performance research and education network developed by the University Corporation for Advanced Internet Development (UCAID), a not-for-profit consortium of over 170 US universities.)

Some typical measures of network performance include:

- *Latency*: the round-trip delay time between the time a packet is sent toward its destination and the time the source receives acknowledgement that the packet was delivered.
- *Packet Loss*: a measure of "lost" packets, usually expressed as a percentage of total packets sent, that do not result in an acknowledgement of receipt from the destination.
- *Reachability*: a yes/no determination of whether or not any packets at all are delivered and are acknowledged. Packets that arrive but are unacknowledged are for practical purposes the same as undelivered, because the sending system will usually consider them lost if the receiving system does not respond.

All of these are end-to-end performance measures, in that they may involve packets traversing multiple physical network links and crossing boundaries from one autonomous network to another. High latency (slow round-trip delay time) will be experienced by Internet users as slow response time or long waits between system outputs. For audio and video delivered over the network, high latency may result in pauses, distortion, video image 'jitter' or other interruptions of continuous service. Packet-loss may result in any of these same symptoms, as a typical response is for the sending system to retransmit packets that are not acknowledged thus introducing the same artifacts as when packet delivery is delayed by latency. Evidence of reachability problems is more straightforward; non-responding sites simply cannot be connected to.

Another common measure is *Percent Utilization*: a measure of the average traffic on a specific network link, typically measured continuously over fairly short time intervals, and expressed as the ratio of volume of traffic carried over total capacity of the link. Percent utilization is a valuable measure in locating specific points of congestion, something that end-to-end measures cannot focus in on.

Internet congestion may be ephemeral, the result of a brief burst of data traffic on the network. Congestion may be sustained, as when tens or hundreds of thousands of users simultaneously access the web-site of a single popular event (World Cup or Super Bowl news) or retrieve a single source of information (the report of U.S. Special Prosecutor Kenneth Starr). Or it may become chronic, when the total use of a network consistently exceeds the capacity available.

When congestion occurs, it may be seen as primarily a provisioning problem, that network capacity is not increasing at the same rate as network use. Internet congestion can also be viewed in a number of ways as a problem of appropriation and use.

Unequal or shifting demand for resources. Different Internet applications make use of different amounts of network capacity. A network designed to function very well when carrying primarily e-mail messages, interspersed with occasional file transfers, will not function as well when most users are accessing web-pages that contain graphics, images, and perhaps audio or video clips. The problem is both that these new applications demand more capacity from the network, and that the new applications overwhelm and crowd-out the earlier applications. So that even though individual e-mail messages may be small, they will still wait and compete with other larger demands on the network. This might be described as a free-riding problem, with newer applications taking more than their fair share of the bandwidth at the expense of an earlier category of applications. (See: innovation, following)

Overwhelming demand for resource. In some circumstances, a single use of the Internet (or a single category of use), may consume virtually all of a network's resources. This is most common as a relatively local phenomena, for example when the connection from a corporate or

university network to the rest of the Internet is entirely taken up by a single user transferring a number of very large data files (this happens periodically when very large scientific data sets are exchanged between institutions), or when the connection is taken up by a number of users transferring a very large number of data sets (this happened recently at a number of U.S. college campuses when students began using a popular Internet-based program to exchange MP3 digital music files). This might be described as a free-riding problem, with specific users taking more than their fair share of capacity to the detriment of all others.
(See: AUP and interdiction, following)

"Unfair" demand for resource. The design of the Internet anticipated occurrences of congestion, and the TCP/IP protocols have the ability for a sending system to detect congestion on the network and slow down its rate of transmission. But it has been pointed out **[20]** that some applications do not respond this way to network congestion. Instead they are designed to use as much capacity as is available at any time. This is a reasonable strategy in periods of excess capacity, as it shortens the total transmission time for the application or improves the quality of the transmission. (This feature is increasingly characteristic of streaming audio and video applications, in which sound and picture quality benefit from higher capacity and low latency.) But in periods of congestion, the "well-behaved" applications slow down their rate of transmission while these free-riding applications speed up their rate of transmission, thus continuing the period of network congestion. This can be viewed as a free-riding problem in which the free-rider is not the Internet user, but the software designer who chose to forego congestion control in favor of obtaining maximum network throughput.
(See: innovation, following)

7. Provisioning and appropriation: responses to overuse and congestion.

Increased provisioning.

Increasing network capacity has been the chief response to overuse and congestion throughout the 30-year history of the Internet. Continued improvements in the underlying technology have made major increases in capacity possible at comparatively modest costs. Moreover, the ease

with which an optical fiber link may be upgraded, by changing the signaling equipment at each end and without digging up old fiber and replacing it, has made continued upgrade and capacity increases feasible.

Ostrom, et al **[21]** point out that one element of solving CPR problems involves "creating incentives…for users to invest in the resource". When Internet services are sold commercially there are incentives for Internet service providers to expand their facilities to keep pace with demand. Access providers will lose customers if they cannot offer reliable and high-quality access. Content providers will not attract users if their site is unreachable or slow to respond. Transport providers will attract few access and content providers if they cannot offer fast network service. Government-funded networks have incentives of their own to offer adequate capacity to avoid congestion -- economic development and competitiveness of national industry, support for research and education, service to health-care and other social good endeavors, etc. -- although there will also be competition for financial resources with other government priorities.

Acceptable use policies and interdiction: restricting access to the CPR.

Ostrom, et al **[22]** also point out that "restricting access" is the other element involved in solving CPR problems. One of the earliest networks that made up the Internet was the U.S. government-funded NSFnet, which operated under an Acceptable Use Policy (AUP) that restricted access primarily to research and education uses and excluded commercial activity. With the vast expansion and commercialization of much of the Internet, AUPs can no longer be enforced over the entire global Internet, but they are still effective mechanisms of restricting access within the boundaries of individual networks that participate in the Internet, or restricting access to "private" Internet Exchange Points. Government-funded high-performance networks in the U.S. and elsewhere continue to restrict access to research and education uses.

The very high speed Backbone Network Service (vBNS) operates under an AUP that restricts use to support of "research and education in and among US research and education institutions and for private or personal communication incidental to such activities." **[23]**

The Abilene/Internet2 backbone network operates under a Conditions of Use policy, which states that, "Abilene traffic primarily and clearly serves the teaching, learning, research, and clinical missions of US higher education, plus the related support infrastructure and services. Abilene does not seek to compete with commodity Internet or other telecommunications services, and is not intended to carry commercial traffic…" **[24]**

Such policies restrict access and thus address problems of congestion on a "local" level, although in both these cases locality is broadly defined by a national or international community of users whose affiliation is based on their mission and use of the network, not geography.

Also on a local level, AUPs may also be employed to restrict access within the narrower confines of a single institution's network, by specifying what network traffic may enter or leave the local network. For example, Indiana University operates under the guidance of a policy for facilitative and fair usage of information technology resources. **[25]**

This policy states the general principle that:

> "technology resources at Indiana University…will be used equitably and only in support of the University's missions of research, instruction and learning, and community service."

Regarding equitable use of resources, it goes on to observe that:

> "individual users or processes may be identified as using what appears to be, in comparison with other users and processes on the same system or network, an inordinate amount of technology resource."

The policy further defines "inordinate amount of technology resource" as follows:

> "a user or process is consuming a resource to a level such that service to other users is degraded, or where the actions of a user could cause degradation if the user is permitted to continue their practice or activity."

On the basis of this policy Indiana University may take action to limit or restrict certain uses of its own network or its connections to the world-wide Internet, in order to conserve resources and protect the use of the network by others.

Innovation: changing the rules of appropriation.

A final category of response to congestion and overuse is innovation -- the development of new protocols, practices, or methods to improve network performance or diminish the impact of certain high-demand uses of the network. In other words, to change the rules in use that govern appropriation of network resources.

One of these innovations is *differential service* (or DIFFSERV), described by the IETF Differential Service Working Group as "mechanisms [that] allow providers to allocate different levels of service to different users of the Internet." **[26]**

As noted above, different network applications have different characteristics in terms of the volume and rate of data they send or receive, or different tolerances for latency. The current practice is to send all traffic on the same network links, each competing for the same capacity and each receiving the same "best effort" service from the network. Differential service is intended to allow such traffic to be separated, with each able to request and (if available) receive the level of service needed for these different applications. DIFFSERV neither restricts access nor increases the capacity that is provisioned, but it attempts to introduce a method for segregating traffic so that, for example, high-capacity and low-capacity network applications can operate concurrently on the same network without interfering with one another. In practice, DIFFSERV will need to be accompanied by a decision-making process that determines how much capacity to allocate to different classes of service and how to assign applications to these classes. These are among the areas being explored and developed in the evolution of differential service.

Another area of innovation directly addressed to the problem of network congestion is *end-to-end congestion control.* Many applications use a network protocol that detects congestion on any given physical link and reduces its consumption of resource on that link until congestion eases. This protocol is use by applications (e.g., e-mail, file transfer) that require every data packet to be delivered, and will request a packet be re-sent if discarded due to congestion. It is a matter of self-interest for these applications to slow down the rate of transmission, to reduce the

number of discarded packets and the reduce amount of resource consumed by multiple attempts to send the same packet. But other applications use a protocol that does not ask that every packet be acknowledged and do not re-transmit lost packets. Audio and video applications, for example, are highly sensitive to latency and cannot afford the overhead and delay involved in acknowledging and re-sending packets. They also have mechanisms to interpolate or otherwise fill in for lost packets. As a result these applications send packets in a continuous stream and do not slow down in conditions of congestion. Worse, as the "well behaved" applications reduce their rate of consumption, these "unfair" applications take advantage of the available capacity that is freed-up. The Internet research community has analyzed this dilemma and proposed a number of potential technical solutions. One of them, for example, is:

> "to support the continued use of end-to-end congestion control as the primary mechanism for best-effort traffic to share scarce bandwidth, and to deploy *incentives* for its continued use. These incentives would be in the form of router mechanisms to restrict the bandwidth of best-effort flows using a disproportionate share of the bandwidth during times of congestion. These mechanisms would give a concrete incentive to end users, application developers, and protocol designers to use end-to-end congestion control for best-effort traffic." **[27]**

The New York Times reported that this approach, which they refer to as a "virtual penalty box" is now being tested in network routers, and they describe its functioning as follows:

> "When a router experiences congestion, it takes a random sample of its traffic. If a certain host computer is over-represented in that sample, its packets are placed at the end of the line." **[28]**

While this remains an area of research and development, not yet widely deployed, it is noteworthy as an attempt to implement a policy of fairness directly in the technical infrastructure of the Internet itself.

Conclusion

As all measures of Internet use increase -- number of users, number of host computers, volume of information available, and volume and rate of data transfer required by Internet applications --

congestion will be a continuing problem. Continued increases in provisioning, supported by advances in networking technology that increase the capacity that can be delivered on current optical fiber media, will play an important role in responding to congestion. Acceptable Use Policies will continue to be effective in selective domains or for selected exchange points. But innovation in network protocol design in order to change the rules of appropriation is also a very promising avenue of response to the CPR problem of network congestion.

Ostrom's discussion of institutional options for solving commons dilemmas, specifically arrangements in which users "make a binding contract to commit themselves to a cooperative strategy" **[29]**, strongly parallels the way that network protocols are developed and enforced. Underlying all protocols is a cooperative strategy -- namely, to assure interoperability of all networks that participate in the Internet by agreement on use of common protocols. And protocols do establish a binding contract -- if any individual, software application, or network service provider ignores or discards a standard network or application protocol, interoperability with the rest of the Internet is diminished or eliminated.

Analysis of network congestion as an appropriation problem leads as well to consideration of the following observation:

> "At the most general level, the problem facing CPR appropriators is one of organizing: how to change the situation from one in which appropriators act independently to one in which they adopt coordinated strategies to obtain higher joint benefits or reduce their joint harm." **[30]**

The function of protocols in the Internet is to organize actions and engage all actors -- providers, appropriators, and users -- in a coordinated strategy to obtain the joint benefit of interoperability -- the ability to communicate through the exchange of information and instructions. In the same discussion, Ostrom also observes that:

> "Almost all organization is accomplished by specifying a sequence of activities that must be carried out in a particular order." **[31]**

The latter description of organization is precisely the definition of a protocol: instructions for carrying out activities in a specific order in order to accomplish a specific outcome. Protocols are mechanisms of organization *par excellence*.

There are incentives for designing protocols and applications that are more "fair" -- as a commercial offering they would protect users against "free-riding" applications -- and there are incentives for router and other hardware manufacturers to implement standards that improve overall utility of the Internet. Similarly, there are incentives for developing differential service protocols, again because the entire Internet provider community benefits from standards that improve the utility of the Internet, and also because differential services may be the basis for creating a differential pricing mechanism for Internet use. This motivation and the necessity for innovation is observed by Ostrom in discussing responses to CPR problems:

> "Appropriators in many settings are strongly motivated to find better solutions to their problems if they can. The economic livelihood of the appropriators depends on their ingenuity in solving individual and joint problems." **[32]**

Ultimately, there is no single lever that can control the provision and use of Internet capacity and thus regulate network congestion. Market forces, government policies, and the independent decisions of individual users and network managers all influence the provisioning of network resources and the rate of consumption. Institutions of collective choice play a significant role, specifically to promote or direct technological innovations that can address existing congestion problems or anticipate and avoid future ones. This takes place in the standards development activities of such organizations as the IETF, W3C, etc., where public and private interests -- corporations, governments, universities and research labs, and individual technicians and engineers -- work together to develop standards for Internet architecture and design. These groups meet, often face-to-face, in locations around the world; they communicate intensively via e-mail; and they freely share their works-in-progress, posting them on the Internet and requesting comments. To take one of these groups as an example, the decision-making of the IETF standards-development process has been characterized as "rough consensus and running code." These set the minimum criteria for innovation in the Internet: something works ("running code"), and it is judged useful by other Internet providers and users -- useful enough, in fact, that it is taken up and becomes widely used (the most practical evidence of "rough consensus").

References

[1] John S. Quarterman, Peter H. Salus. How the Internet Works. <http://www.mids.org/works.html>

[2] Charlotte Hess. "Untangling the Web: The Internet as a Commons," Workshop in Political Theory and Policy Analysis, Indiana University, March 1996.

[3] Ostrom, Elinor; Burger, Joanna; Field, Christopher B.; Norgaard, Richard B.; Policansky, David. "Revisiting the Commons: Local Lessons, Global Challenges", Science, 9 April 1999, Vol 284, pp.278-282.

[4] Hess. 1996.

[5] Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press, 1990. p.47.

[6] MIQ Ratings Methodology - How We Compare Performance of ISPs. <http://ratings.miq.net/method.html>

[7] Bilal Chinoy, Tim Salo. Internet Exchanges: Policy Driven Evolution. <http://ksgwww.harvard.edu/iip/cai/chinsal.html>

[8] Directorate for Science, Technology and Industry; Committee for Information, Computer and Communications Policy; Working Party on Telecommunication and Information Services Policies. *Internet Traffic Exchange: Developments and Policy*. Paris: OECD, 1998.

[9] Matrix Information and Directory Services, Inc. <http://www.mids.org/>

[10] Global Internet Statistics. <http://www.glreach.com/globstats/index.php3>

[11] Multi-Lateral Peering Agreement proposed by the Chicago NAP <http://nap.aads.net/MLPA.html>

[12] Sheldon, Tom. LAN TIMES Encyclopedia of Networking. McGraw-Hill, Berkeley, California, 1994.

[13] Internet Engineering Task Force Overview. <http://www.ietf.org/overview.html>

[14] About the World Wide Web Consortium. <http://www.w3.org/Consortium/>

[15] ibid.

[16] <www.nlanr.net>

[17] <www.mids.org>

[18] <www.vbns.net>

[19] <www.abilene.iu.edu>

[20] Sally Floyd. "Promoting the Use of End-to-End Congestion Control in the Internet. IEE/ACM Transactions on Networking, Vol. 7, No. 4, August 1999.

[21] Ostrom, et al, 1999.

[22] Ostrom, et al, 1999.

[23] Acceptable Use Policies for NSFnet Program Backbone Network Services. January 10, 1997. <http://www.vbns.net/aup/aupfaq.html>

[24] Conditions of Use for the Abilene Network. November 29, 1999. <http://www.ucaid.edu/abilene/html/cou.html>

[25] Information Technology Facilitative/Fair Usage Policy. September 2, 1999. <http://www.itpo.iu.edu/IT11.html>

[26] Differential Service for the Internet. March 30, 1998. <http://diffserv.lcs.mit.edu/>

[27] Sally Floyd, 1999.

[28] Sara Robinson. "Multimedia Transmissions Are Driving Internet Toward Gridlock". The New York Times, August 23, 1999. pp. C1,C4.

[29] Ostrom, 1990, p.15.

[30] Ostrom, 1990, p.39.

[31] ibid.

[32] Ostrom, 1990, p.34.