International Journal of Communication 1 (2007), 548-566 1932-8036/20070548

net neutrality issue

Measuring the Network - Service Level Agreements, Service Level Monitoring, Network Architecture and Network Neutrality

DIRK GRUNWALD
DOUGLAS SICKER
University of Colorado at Boulder

In this paper, we argue that the network neutrality debate hinges on issues that go beyond whether a carrier is unfairly discriminating in their treatment of packets across their network. We consider four issues central to this bigger discussion concerning network neutrality. First, we discuss the implications of *oversubscription* on the access networks. Next, we describe the lack of *service level agreements* for end-users and the coincident inability for users to *monitor* the quality of the network. Lastly, we consider the evolving architecture of the access networks and what the current design, dominated by *asymmetric uplink and downlink* characteristics means for future services on the Internet.

Index Terms—SLA, Network Neutrality, Oversubscription, Network Architecture

I. Introduction

In this paper, we argue that the debate surrounding *network neutrality* has suffered from failing to properly address what ensuring (or failing to ensure) network neutrality may mean technically and within the practical context of evolving Internet and last-mile broadband architectures and services in the United States. We begin in the introduction by summarizing issues to consider in addressing network neutrality concerns. The paper then considers in detail these issues that are currently missing in this debate, particularly from a technology perspective. Section III discusses Oversubscription, Section IV discusses Service Level Agreements and Monitoring and Section V examines Network Design and the importance of Symmetric Capacity. Finally, in Section VI we consider how these technical issues might impact the future of Internet development.

Network neutrality is the non-discriminatory carrying of varying types of Internet traffic by Internet Access providers. While we view the goal of network neutrality as a desirable condition, we argue that many of the proposed remedies to ensure "network neutrality" may, in fact, hurt the deployment of

---

Dirk Grunwald: **dirk.grunwald@colorado.edu**
Douglas Sicker: **douglas.sicker@colorado.edu** (primary author)
Date submitted: 2007-06-07

innovative services. For example, some services, such as interactive voice, are most effective when their data streams receive special treatments. Prohibiting such treatments may render the Internet less useful for such services. Further, although regulators can mandate equal service, such mandates will be difficult to enforce because contracts in existing access networks don't precisely define the service that is being provided. Equally important, there are no standards or methodologies for assessing when the network "is not neutral" and the organization of the network makes such determinations difficult. Since the Internet is a network of networks, each owned and operated by different organizations, assigning responsibility for service problems is difficult, and there are limited incentives for network operators to cooperate in resolving such problems.

A central point in this discussion is that Internet access providers do not guarantee the performance of the access service offered to users, and the performance of these links can vary substantially, depending on the design and utilization of the access network. To illustrate, when a consumer buys a 10 Mbps service from their cable access provider, that data rate is not a guaranteed rate, rather it represents a "best potential" data rate. This data rate is impacted by the design of the access network, the decision by the access network operator on how to allocate resources within the constraints of the network design, the number of customers sharing that network (i.e., the number of neighbors that are also using their cable service at that time), the type and amount of traffic on the network and the delay on the entire network path. Actual data rates can drop to a fraction of that 10 Mbps and this decrease can negatively impact the users' applications. An important point to consider is that during times of congestion, what matters is not the best potential rate, but rather *the lowest maintained (or guaranteed) rate*.

However, the technology exists to provide service guarantees. Commercial Internet services use *service level agreements*, or SLAs, to specify the quality and quantity of Internet connections. We argue that Internet access providers have traditionally not been concerned about SLAs, but that in the future the rise of bandwidth and delay sensitive applications might motivate carriers to provide some service level information. In fact, we argue that by providing this information, the carriers may be able to diminish the threat of having regulatory obligations placed on them by demonstrating that they are providing a ”neutral” access platform.

The debate about prioritizing traffic obscures an equally or more important issue - the asymmetric bandwidth of access networks will both limit solutions to scalable content distribution and increase the likelihood of media consolidation. The Internet has increasingly moved from a "network of peers" to a "client server" model, where large computer servers (which require extensive capital) provide content to home clients. This evolution was mandated by the poor performance of early access networks and a lack of mature software to manage distributed content. The Internet has seen considerable innovation in "distributed computation" for challenging computational problems, such as a cure for cancer[1], but the lack of sufficient "upload bandwidth" limits the solutions to distributed solutions for data-intensive problems in the Internet and future services that may have significant impact on economic

---

[1] See, for example, the "Cure Cancer With Your Computer" project at http://www.nfcr.org/Default.aspx?tabid=274

competitiveness and the scalability of Internet services for different access network technologies.

Rather than offer specific policy recommendations, this paper seeks to refocus the net neutrality debate by discussing these four issues that have heretofore not been adequately addressed:

1) *The need to consider oversubscription policies in assessing how network neutrality problems are likely to change over time and for different types of access networks.* Oversubscription or multiplexing of multiple user traffic onto shared facilities is a staple of network design, demanding that when congestion occurs, it must be managed – either all pending traffic yields to the congestion, or traffic is selectively prioritized. With the growth of rich multimedia and real-time traffic, the needs of different applications must be reconciled, as such applications demand that bandwidth requirements and time sensitivity parameters be met in order to operate successfully and acceptably. We conclude that changing traffic patterns will either force a lowering of current multiplexing rates or will induce greater reliance on explicit QoS to address increasing congestion issues.

2) *The need for Service Level Agreements (SLAs) and generally accepted standards to provide an informed technical basis for identifying and diagnosing network neutrality problems.* We highlight the need to focus not only on peak data rates, but also on average and minimum assured rates for both uplink and downlink connections.

3) *The need for appropriate monitoring metrics and infrastructure to collect, aggregate and report those metrics to complement and make useful the SLAs and standards noted above.* Metrics are needed to allow end-users and carriers to monitor the SLAs and QoS actually delivered. We show that end-user implemented P2P monitoring applications may prove useful in dealing with the challenges of monitoring e2e service quality.

4) *The need to focus on symmetric capacity in network design, which means paying attention to the uplink as well as downlink capacity as part of network neutrality policy.* We note that as the Internet evolves into the platform for our mass-market communications infrastructure, it is shedding its legacy focus on symmetric link capacities (where nodes are able to send and receive at approximately equal data rates) and becoming increasingly asymmetric (emphasizing higher downlink speeds to support broadcast delivery of rich media content, with much lower uplink speeds). This is consistent with the evolving duopoly competition between cable television and telephone carriers, each seeking to deliver a triple play of television, telephone, and broadband Internet access services to homes. Lastly, we note that this may harm innovation in Internet services and technology because it is biased against such important technologies as P2P and other end-user/edge-based approaches that have played such an important role in the evolution of the Internet to date.

II. What is the Net Neutrality Debate About?

Net neutrality, the principle that Internet access providers should not discriminate among different instances of Internet traffic is at the core of an important policy debate about how the Internet will evolve. Incumbent access network providers view the threat of regulation as premature and an unnecessary interference in their businesses without justification or demonstration of harm. At the same time, proponents of net neutrality claim it is indispensable to the innovation and openness of the Internet that society has come to expect and cherish, holding that the Internet will become the primary vehicle for social expression, eclipsing traditional means of expression including print, radio, broadcast television and high definition cable. For example, at the National Conference for Media Reform in January 2007, Bill Moyers stated [1]:

> *I can tell you reading about those days that educators, union officials, religious leaders and parents were galvanized by the promise of radio as a classroom for the air, serving the life of the country and the life of the mind until the government cut a deal with the industry to make sure nothing would threaten the already vested interests of powerful radio networks and the advertising industry. And soon, the public largely forgot about radio's promise, as we accepted the entertainment produced and controlled by Jell-O, Maxwell House and Camel cigarettes.*

Such statements highlight that one faction views the network neutrality debate in the context of preventing media consolidation and preserving access to media outlets for improved public discourse involving far greater numbers of people, including minority opinions; while others argue that service providers either have every right to discriminate among traffic types since they own the networks (an argument that prompts serious questions of what society wishes to permit in trade-offs between property rights and social policy goals), or that there are simple and compelling economic reasons for discrimination among traffic types based on certain applications' consumption of network resources.

The argument of compelling economic rationales for discrimination is strengthened by the current shift from traditional Internet access applications, such as web browsing, toward increasing multimedia traffic such as voice, music, and video that require quality of service treatments that exceed the requirements of traditional applications. Typically, emerging applications require "delay sensitive" treatment or "high bandwidth." Delay sensitive traffic (VoIP or video conferencing) must be delivered with very little delay and very little variance in delay. High bandwidth traffic involves transferring large masses of data (such as pre-recorded audio or video files) quickly. Complications arise when both types of traffic are combined on a single network. Incumbent access network operators argue that current network implementations are under-provisioned to support both delay sensitive and high bandwidth traffic and propose shifting part of the cost of upgrading those access networks to the providers of demanding content. Companies selling services that emphasize delay sensitive (VoIP) or high bandwidth content (IPTV, video distribution, etc.) and those concerned about media consolidation believe the access providers' argument is an attempt to extract excess rents in an environment lacking effective competition.

In simple terms, network neutrality advocates contend that consumers should be able to access the content and applications of their choice without network carriers "discriminating" among these choices. For example, under network neutrality a consumer could use Vonage phone service without the carrier blocking or otherwise degrading that service. However, there are several situations where a carrier might justify discriminating against certain content and applications, such as illegal content, or due to network security or network management requirements. This might include prevention of malicious reconnaissance such as vulnerability scanning, or blocking packet delivery of potential malware (such as computer viruses and worms), or the carrier may deem it necessary to limit its network traffic to maintain certain performance expectations. A more contentious justification is the economic claim that the content or application provider should compensate the carrier for the use of its network. Indeed, ATT CEO Ed Whitacre was recently quoted as saying, ”What [Google, Vonage, and others] would like to do is to use my pipes free. But I ain’t going to let them do that.” [2]. Similarly, the CEO of Verizon, Ivan Seidenberg, stated that broadband application and content providers should “share the cost” of operating broadband networks [3]. In November 2005, a Verizon Wireless executive stated that network neutrality should not apply to wireless carriers and indicated that carriers should be allowed to block traffic as they think appropriate [4].

The Federal Communications Commission articulated four principles with respect to network neutrality, directed at ensuring that “*broadband networks are widely deployed, open, affordable, and accessible to all consumers*” [5]. According to the FCC's principles, in order to encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet, consumers are entitled to:

- Access the lawful Internet content of their choice.
- Run applications and use services of their choice, subject to the needs of law enforcement.
- Connect their choice of legal devices that do not harm the network.
- Competition among network providers, application and service providers, and content providers.

They also stated that they have “jurisdiction necessary to ensure that providers of telecommunications for Internet access or Internet Protocol-enabled (IP-enabled) services are operated in a neutral manner” [5]. In other words, if they want to enforce network neutrality they have the authority. However, a footnote in the statement indicates that the principles are subject to “reasonable network management.” This of course raises the question as to what constitutes reasonable network management; the concern here being that network management becomes a justification for selective content blocking and other similar measures.

Clarification was brought in a March of 2005 action where the FCC fined Madison River Communications (a rural phone company) for blocking connections to Internet phone providers (such as Vonage). As a result of FCC pressure, Madison River also agreed not to block such calls in the future. In this case, while several of the aforementioned justifications for blocking content were provided, the FCC found none of these acceptable [2]. This action, together with the policy statement, suggests that the FCC does view network neutrality (or at least to the extent it involves competition) as a socially desirable goal.

As this example illustrates, the reasonable desire of the carriers to recover their network costs is countered with the concern that the carriers will exhibit anti-competitive behavior and thereby erode much of what has made the Internet successful. Telecommunications regulators have decided to move away from the heavily regulated model to embrace a model based on market discipline. However, with (at best) only two broadband players in most markets, many question whether sufficient competition exists to ensure the desired market discipline.

The mixing of high bandwidth and delay sensitive data means that the Internet may need to shift away from the best-effort model where all bits are essentially treated equally to a differentiated service model where certain bits can be given priority on the network. As network carriers move to this differentiated service model, they will have the ability to decide whose bits are provided with priority treatment. This decision will likely favor their own bits or bits from carriers who pay for priority treatment. The concern is that those who do not pay could be relegated to a service level that does not support their application demands. This is a weighty concern as the majority of the innovation on the Internet has been done by the application providers, not the network providers.

The underlying issues about network access, provisioning, discrimination and quality of service are at the core of most network neutrality debates. However, we argue that although these points are important, they can also serve as distractions from a broader range of issues that may have an equally or more profound impact on the future of Internet development and national economic competitiveness.

III. Understanding What You’re Buying -- Oversubscription

Internet access services typically advertise and compete on the basis of data rate. For example, one provider advertised DSL service capable of “up to 1.5 Mbps download/896 kbps upload” [7] while a competing cable company advertised service supporting “up to an unbelievable 6 Megs” but later caveats this with “Actual speeds may vary and are not guaranteed. Many factors affect speed.” [2] [8] One of the most important factors that affect speed is oversubscription.[3]

Oversubscription, also known as statistical multiplexing, is really a statement of the relationship between a marketing or advertising contention about the rate that a customer could enjoy (e.g., 1.5 Mbps) and the amount of capacity available on the network if all users are actively attempting to secure the advertised rate. For example, if a network has a total capacity of 15 Mbps and is shared among 100

---

[2] Qwest advertises several flavors of broadband. For example, Qwest offers its "Qwest Connect Silver with Windows Live" service that offers "Up to 1.5Mbps download/Up to 896Kbps upload" for $44.99 per month (standard), but at the time this service was viewed, they were offering a promotion of $13.00 per month discount for a month, or for life, if you agree to a two year contract (see http://www.qwest.com/residential/internet/pricing.html, viewed July 25, 2007). Note: This URL may change as technology and offers change; they may no longer be available or described as this.

[3] The text is available if you go to http://www.comcast.com/# and then click on "Learn" and then "High Speed Internet" it loads a flash player that provides you with interactive description that includes both claims about Comcast being much faster than DSL and also the disclaimer. Note: This URL may change as technology and offers change; they may no longer be available or described as this.

users, all of who are told that they are purchasing a 1.5 Mbps service, the oversubscription ratio is 10-fold: the ratio between the 150 Mbps that would be required to serve all users simultaneously and the 15 Mbps that is actually available. Oversubscription ratios are a design choice made in any part of the network that is shared. All networks share capacity in Internet switching, routing, backhaul, and transport. At first glance, oversubscription sounds like service provider fraud, but it is not, for both legal and practical reasons. Legally, service providers use the terminology "up to" in describing speeds as well as other qualifications as we saw earlier in this section. Practically, web surfing performance is largely determined by being able to support short high-speed bursts. Traditional Internet traffic such as web browsing is characterized by multiple short bursts of data (as successive transactions between a web browser and a server generate and fulfill requests for data) separated by periods of idleness. Performance is largely dictated by achieving low latency and high burst data rates while guaranteeing that data is delivered error free. Since sustained high data rates are not necessary, it is feasible to share a channel with high capacity among multiple users. So long as requested bursts are not often coincident, it may well be true that a user on a ten times (or more) oversubscribed network sees the advertised rate (or even a better rate) almost anytime that it is important to the user experience. Oversubscription does not typically impact traditional Internet traffic profoundly.

In addition to web surfing, file transfer is another common Internet traffic type; the emphasis here is on high average rate and error free transmission, but with allowable significant variation in instantaneous rate. Non real-time multimedia traffic has similar needs to those of large file transfers, albeit likely with some higher tolerance for errors than, say, a file of financial information. Of course, error correction can be handled at higher levels of the protocol stack. Real-time multimedia traffic makes two fundamentally different tradeoffs: a need for sustained data rates (at a rate that depends on the application, such as voice, video, music, etc.) and some tolerance for errors [6]. Latency is important if the multimedia is part of an interactive communication and less relevant if not (for example, one way streaming of a television program). Quality of Service (QoS) is a term used to denote an application's requirements, or the state of a network in terms of meeting those requirements. As we've seen, for multimedia these tend to focus on attributes of data rate (or "bandwidth"), rate variation, latency, and error rate. Oversubscription would tend to have the most significant and unacceptable impact in a real-time interactive multimedia application.

Service providers usually consider oversubscription ratio choices proprietary information. This leaves customers to discipline service providers through a qualitative assessment of service quality, expressing their preferences through purchase decisions. Of course, when there are only one or two broadband options in the market, it may be difficult to express a preference. Because of the lock-in effects around service decisions, as well as variability and noisiness in qualitative assessment, we can expect this effect to be present, but slow and fickle. In spite of a lack of public information, industry trade publications suggest that service providers select oversubscription ratios ranging from 10 to 100 [9], [10], [2].

IV. Understanding What's Needed – Service Level Agreements and Network Monitoring

Although regulators could mandate a network neutrality obligation, such mandates will be difficult to enforce because contracts in existing access networks don't precisely define the service that is being provided. As we have mentioned, given that broadband access networks are highly oversubscribed, end-users often receive far lower throughput performance (and longer latencies) than might be suggested by broadband service advertisements. Given this situation, it is more worthwhile to have information, or commitments, concerning the lowest expected throughput. With such information, it is possible to begin to assess the ability of a broadband connection to support a set of applications, particularly during times of network congestion. Such information might be communicated with subscribers through the use of service level agreements (SLAs).

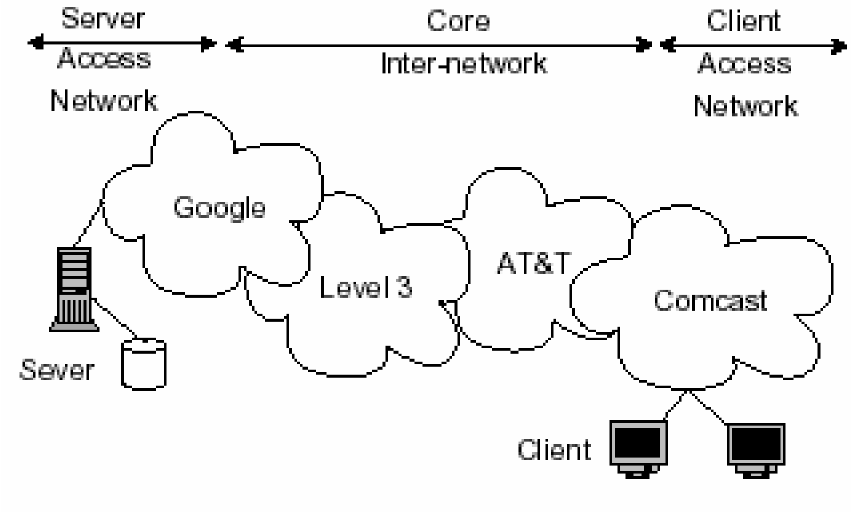

Figure 1. Example Autonomous Systems Between Author's House and Google

However, an SLA is meaningless without the tools to monitor the performance of the network. Such tools could take many forms. For example, the network provider, Comcast as depicted in Figure 1, could offer a network monitoring service to the subscriber. This tool would allow the subscriber to gain a view of Comcast's network performance. However, delay includes network performance beyond the local access network, namely the server access network and the core network. In our example this includes the performance of Google, Level-3 and AT&T. This would require monitoring and participation by entities outside of the access provider's network. Such monitoring tools could be provided by each carrier or could take the form of a distributed peer-to-peer application that would run on a large number of end-users'

PCs. The advantage of these systems is that they could provide information about other networks, whereas the Comcast tool would offer only limited views into other networks. It might be that all of the tools could be beneficial in a complementary manner by providing detailed pictures of location network conditions as well as more general information about the performance of other networks and the Internet.

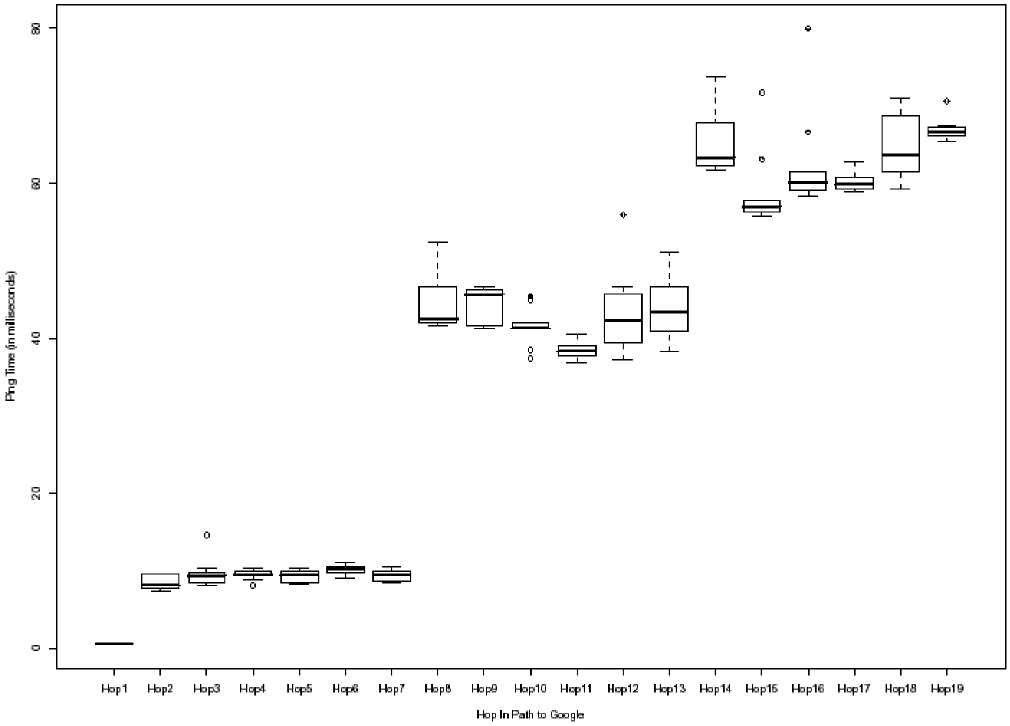

Figure 2. The Round Trip Time For A Data To Traverse The Network Depicted in Figure 1

It should be noted that even if the tools exist, there are no standards or methods for assessing when the network "is not neutral." Just defining "neutral" could be a significant point of contention. From a technical perspective, legislation could prescribe "neutral" access, latency, bandwidth or availability/reliability. However, "reasonable" network architectures can dramatically influence the value and variance of these metrics. Many of the metrics are greatly influenced by geography more than network design.

To provide a simple examination of the performance of the networks depicted in Figure 1, we ran a "ping trace" as shown in Figure 2. This test involves sending data through the network and having each router or "hop" report the round-trip delay for that packet. In this experiment, we have measured the delay 10 times and reported the mean and variation using boxplots; the middle line in each box indicates the mean while the extent of the box indicates the inter-quartile range of the round trip time. While not necessarily a statistically significant representation of the delay in networks, the delays that we observed across these networks were dominated by large transit times; these transit times correspond to the traversal of large distances. In this particular ping trace, there are four "regions" of network delay. The first, and least delay, is that of the home network. The next six "hops" are in the Comcast network with the last reported time being the interface between Comcast and AT&T. The large jump in packet delays is due to the large distance between Colorado and San Francisco, which is the next hop on this particular network path. The next sequence of six "hops" involve the data moving between different routers in San Francisco prior to being relayed to Seattle, Washington; again, the increased delay is due to limitations of the speed of light over long distances. The last sequences of hops are transitions between the AT&T network, the Level3 network and the internal Google network. The measurements for some later hops are lower than that for earlier hops because of differences in how routers prioritize the processing of the ping packets.

There are two things to take away from this diagram. The first is that the variance in ping times is less for the access network than for any of the other networks. Although it would be dangerous to extrapolate results from measurements using a data type specifically indicated for Internet measurements, this set of measurements indicate that it may be difficult for operators of access networks to guarantee network performance precisely because the most variable aspects of service are beyond their control. And, as mentioned, the significant delays in the network are due to geographic distance and the distributed nature of the Internet. Each datum retrieved from Google, YouTube, iTunes or other sources make similar traversals through the Internet. This also influences availability – If the AT&T network were to fail, the Comcast users would be affected, but would have no direct understanding of who is to blame.

A reasonable approach to improve the performance of the network for moving large non real-time content might best be addressed by placing the content closer to the user, using either peer-to-peer distribution networks (such as BitTorrent or the Joost video distribution systems) or commercial content distributions systems such as Akamai. Another improvement could be gained by having direct interconnection between the client and server networks. In these approaches, Comcast could directly connect to Google or archive Google content closer to the client. In effect, this organization mimics the existing mechanisms used by cable operators to efficiently distribute video data.

Obviously the content caching approach would improve the metrics for real-time services; however, direct interconnection could improve the performance of real-time traffic (as could quality-of-service optimized connections among the carriers). And, if much of the data in the Internet is pre-generated content, efficient management of that data makes the existing resources more available for time-sensitive data. The ping trace also suggests that the network under examination was fairly stable and had relatively low latency. In fact, the performance of this network would readily support a high-quality VoIP conversation.

Each of the points above demonstrate the value in allowing the network to evolve to support the demands of the users and suggest that network neutrality legislation should consider the potential for such network evolution. There are many “reasonable” network organizations that could circumvent the prescriptions of network neutrality legislation while still increasing media consolidation or two-sided payments for use of the access network. If an access network provider collaborated with a content distribution company such as Akamai to distribute high-bandwidth content, that content would have an advantage over the content of other providers. However, adding such content caching is both reasonable and desirable for both the access network operator and the subscribers of that network since it improves performance. Are such changes in network design “neutral”? Many involved in the network neutrality debate would feel that such a network design would work against the goals of fair and equal media access, but have not articulated alternatives to such “reasonable” network organizations.

V. Network Design –Asymmetric v. Symmetric Capacity

The previous section discussed the relationship between the organization of access networks and the efficient delivery of rich media. This section considers how network design has changed from symmetric uplink and downlink speeds to an asymmetric design favoring faster downlinks; and examines the impact to innovation and potential net neutrality solutions.

The original Internet was designed as an overlay network operating over the existing phone system and specialized high speed links; over time, that network evolved to include a multitude of interconnects, including ATM networks, frame relay and so on. One aspect of the early Internet was that most of these links were symmetrical – except for rare cases, such as satellite links, each of the (few) nodes in the Internet could send data at about the same rate at which it could receive data. This symmetrical organization, coupled with the relatively slow link speeds available at the time, drove the evolution of many early Internet applications. For example, modern “peer-to-peer” applications distribute content among a number of servers within the network; early predecessors to such applications were the common ‘Usenet’ news protocol that physically copied “news articles” from one machine to another – the content was distributed across systems because the costs (in time and money) of retrieving it on demand were too high for the technology of the time. This early application dealt with slow link speeds by storing contents with end-users, at the edge of the network, and utilizing the uplink for distribution rather than waiting for content to be delivered from a remote point, over a long haul network. One can see how providing symmetric uplink-downlink speeds would foster innovative applications that utilize such uplink capacity.

This assumption of symmetrical networks is still true for most institutional, commercial and educational uses of the Internet, but stands in contrast to most consumer access networks. For example, in the United States, most broadband consumers access the Internet using either DSL or DOCSIS cable modems. DSL connections typically offer a 1.5Mb/s downlink and 250kb/s uplink; cable modems offer a 5-6Mb/s downlink and 300-500kb/s uplink. Both of these stated bandwidths suffer from the problem of imprecise specification, as discussed in Section III. By comparison, a commercial T1 connection offers a regulated symmetrical 1.54Mb/s connection (typically at a much higher price than comparable DSL or DOCSIS connections).

This shift from a "symmetrical Internet" to one biased for download speeds is a by-product of the commercial uses of the Internet and the "client-server" model of many evolving Internet services. In the United States, as the Internet was commercialized, greater emphasis was placed on being able to access rich content (graphics and so on), and many of the technologies used to access the Internet were adopted before the technology to provide higher speed symmetrical access was firmly developed. By comparison, countries that adopted Internet technologies later were able to offer high speed symmetrical connections; for example, purchasing 100Mb/s download and 100Mb/s upload using VDSL technology in Japan costs $40 and similar prices and services are available in Korea, Italy, Portugal, Sweden and other countries. By comparison, consumer broadband services available in the United States offer much lower data rates. For example, Verizon's FIOS (Fiber to the Home) program only promises downlink speeds of 5, 15 or 30Mb/s and uplink speeds of 2 or 5Mb/s.

One might argue that these differences are related to technology or topography, but it seems more likely due to the nature of the duopoly market for mass market broadband evolving in the United States. For example, the Verizon FIOS deployment is using the majority of their available bandwidth to provide broadcast television services to compete with cable television. Under current regulations, fiber installations are not obligated to be shared by other carriers, as are DSL connections. Presumably there is more money to be made selling broadcast TV services using the available bandwidth. In particular, if the full available bandwidth were available to consumers, they might purchase television services from other parties using Internet protocols such as IPTV. This would reduce Verizon to a simple carrier rather than a vertical service provider. Similar decisions are made by cable companies deploying the DOCSIS services; it is possible for these networks to offer a broader range of download and upload speeds, but the bandwidth occupied by those services must be traded against broadcast television and proprietary services.

However, trading downlink and uplink bandwidth for broadcast television services has a hidden cost, because it retards further innovation in the Internet and potential services that may have strategic impact on the security and prosperity of the United States. We outline three such examples; our examples span past, present and future Internet applications and are illustrated in a variant of our pedagogical examples in Figure 1. There are three changes to underlying issues enabled by a robust Internet design.

The first is pricing pressure through competition – in the current DSL market, customers can choose multiple Internet Service Providers (ISPs) – this is indicated in the diagram by a sea of ISPs within the access network. This competition is an artifact of the regulations governing DSL connections, but the separation of the client access network into a physical transport and distinct ISP's allows customers to pay for different values in peering and transit relationships; in the absence of such "virtual competition," non-incumbent providers must overcome the fixed costs of deploying their own infrastructure, effectively limiting competition. The reason limited bandwidth affects competition is that when any resource is limited, the argument can be made that the resource would be better managed through means other than competition or that the resource is too limited to support the additional fixed costs needed to enable competition. Despite statements that networks providing richer competition are difficult to design, some efforts, such as Utah's Utopia metropolitan networks [13] appear to offer practical architectures that allow competition among multiple retail service providers, including a choice of ISP, voice and television services.

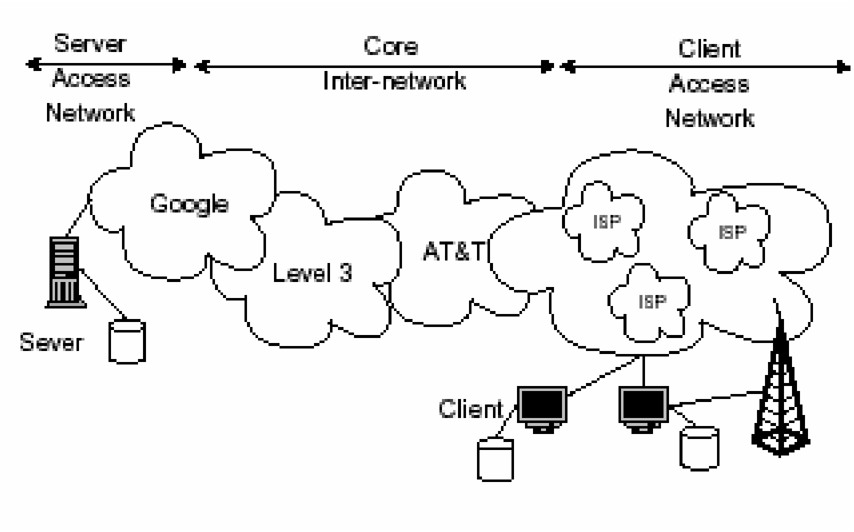

Figure 3. Potential Internet Architectures Precluded By Asymmetrical Bandwidth

The second issue facing U.S. markets is inefficient solutions to existing distribution problems that tend to favor incumbents or companies that control the client access network. Consider an example of conducting a search for a video stream using Google and downloading that stream using YouTube. The interaction with Google requires very little resources from the network, both in terms of total bandwidth and priorities; the video download requires more bandwidth but is (relatively) insensitive to variation in latency. If the content of YouTube were distributed through clients in the client access network, the video content could be sourced from local systems, decreasing bandwidth demands on common (longhaul) networks, therefore increasing speeds. This is the essence of “peer-to-peer” networks, and was the basis of earlier Internet applications such as Usenet. In effect, the lesson is “keep big stuff local.” In fact, Joost is a new peer-to-peer streaming video service that attempts to build just such a peer-to-peer network for distributing legal video content. From our prior examples, we saw the pure “client-server” model of Internet communication will require significant growth and expansion in the Internet core; adopting content distribution networks such as Akamai allows the bandwidth to be reduced, albeit at the cost of a “corporate structure” that regulates which large content sources will be efficiently distributed. By comparison, appropriate peer-to-peer systems allow the *end-users* of the Internet to decide whether to participate in the distribution of large content, as illustrated in Figure 3. In such systems, the content is provided by nearby systems. As our earlier discussion of the delays in accessing remote data, illustrated by the measurements in Figure 2, we can see the advantages of a peer-to-peer architecture for content movement.

As our simple measurement of network delay indicates, the only way to beat the speed of light is to retrieve data from local computers. If technologists are interested in reducing the “crushing demand” of video downloads, then technology that allows retrieval from local copies is essential. This solution gives broader voice to media sources and fosters the “media fairness” goals of advocates for network neutrality – for example, using a peer-to-peer system, the availability and accessibility of a video could be determined by the predilections of the clients subscribing to the network. Recent research indicates that adopting these network architectures would directly benefit access network providers [15]. Alternatively, the distribution companies could franchise their content provisioning across multiple clients, as is currently happening with commercial BitTorrent distribution. In either case, it is the network participants (the end in the end-to-end argument) that define the network function. All of these methods depend on the availability of significant "upload" bandwidth to end-user client computers; if the network design precludes such capabilities, then the incumbent content distribution (*e.g.,* television or television-like services) will remain the dominant content distribution service.

The third and last issue facing U.S. markets is the lost opportunity for innovation in other communications technology enabled by the Internet. Arguably, if we, as a country, were to deploy a cellular telephone system after the development of the Internet, we might examine variants of existing designs. Existing cellular systems use large towers to cover large areas in part because the acquiring tower locations is difficult, provisioning the tower is expensive and such central locations allows the rapid coverage of large areas. In the presence of a universal Internet, an alternative strategy would be to enable a multitude of smaller wireless cells that face less objection to deployment due to their size and coverage; such system doesn’t obviate the need for “wide area” coverage, but it allows mobile networks to “fill in” high-density metropolitan zones much more efficiently while preventing “dead-zones” in wireless deployment. It also allows the deployment of emerging 3G standards such as UMTS, which require much closer tower placements. However, the vision of any such “distributed cellular” infrastructure is limited by the availability, quality and regulations on Internet uplinks. Most acceptable use policies for current client access networks preclude the “sharing” of Internet resources; moreover, the limited upload bandwidth available under existing DSL and cable modem connections would practically limit the ability of a distributed cellular installation to offer reliable connections since the cellular network ties into and would be limited by these networks.

All of these challenges will be easier to overcome in countries that either encourage or mandate symmetrical access networks or at least regulate the uplink along with the downlink. If the network neutrality debate in the United States ignores these issues, it may be that long term solutions to “fair” media access or innovative applications will be precluded by incumbent network designs. Many incumbent networks would argue that such a network organization is optimized for (commonly unrealized) use of uplink bandwidth; however, metropolitan networks such as Utopia [14], which are not beholden to incumbent markets or regulations, demonstrate that there are innovative solutions that enable broadcast TV services, competitive voice services and data services that enable novel network applications by competition among users of a shared physical infrastructure.

Thus, we encourage policymakers to focus on *uplink policies* as well as downlink policies. Such an emphasis will likely be rewarded by a richer, more reliable and scalable Internet design that better serves the national interest. This issue is largely ignored in the current debate on network neutrality.

VI. Access Network Technology and the Impacts of Neutrality

The prior sections described three issues that face network neutrality legislation: knowing what you've bought (oversubscription), being able to know if you're getting what you bought (SLAs and monitoring) and asymmetric network design (higher uplinks speeds and lower downlink speeds). These issues will affect different network types in different ways. We now contrast how these issues impact cellular, DSL and cable networks differently, focusing on the impact on cellular networks.

Different physical layer networks face different challenges in meeting particular QoS expectations. Cellular wireless networks provide examples of the unique challenges in meeting QoS expectations presented by this particular type of physical layer network. Wireless networks have two particular characteristics of interest: a shared physical channel and a propensity for variations in error performance (caused, for example, by fading during relative motion of transmitter and receiver) [11].

Traditional cellular telephone networks solve the QoS problem for interactive voice by emulating the traditional circuit switched infrastructure of the public switched telephone network. Calls are blocked unless capacity can be dedicated to the call and the channel can be expected to have a low enough error rate to sustain intelligibility; what errors do occur are mitigated through error concealment. Recent cellular networks, though, are much more likely to select a packet switched infrastructure rather than circuit switched, with the goal of efficiently handling a wide variety of applications. These modern wireless networks are frequently designed first via natural extensions of computer local area network principles rather than as extensions of PSTN principles. Indeed, this trend is evident in almost all networks. QoS for voice calls – and for other multimedia traffic – then becomes a configuration and management choice for the network rather than an intrinsic characteristic of the network design. Thus, network neutrality of new cellular networks is in the hands of the carriers. The carrier may choose to prioritize certain traffic to maintain an expected QoS level for particular applications. It is likely that traffic for lucrative premium services would be prioritized. The QoS needs of various media are either presumed solved adequately by abundant over provisioning or through explicit QoS semantic overlays on networks. In either case, though, we tend to exacerbate a fundamental characteristic of the economics of cellular wireless Internet access: the expectation and reliance on shared use. Part of the cellular infrastructure is spectrum, which is a natural and limited resource that is increasingly consumed by bandwidth-consuming applications, such as streaming video.

The extent to which a particular physical network type is cost sensitive to oversubscription depends on the relative share of network cost that is devoted to shared infrastructure versus dedicated (per customer) infrastructure. The component of the network that is devoted to transport and backhaul of Internet traffic tends both to be more common among different physical network types and a smaller fraction of total cost than the portion dedicated to last mile link access. In other words, whether you are using a cellular network, cable network or DSL network to connect, each of these networks will connect to and use a common (shared) backhaul network to send signals over physically long distances. This portion of the network is relatively inexpensive in comparison to the expense of the “last mile.” This is because the long haul network spans one long segment with relatively few connections, while the “last mile” is a tremendous amount of physical infrastructure consisting of all the connections and physical infrastructure linking each individual end-user to the inner network. Consequently, most variation in sensitivity to oversubscription will arise from the last mile link. Thus, it is much less costly to build or upgrade a long haul portion of the network that is capable providing higher traffic rates.

DSL links do not share the last mile link;[4] a dedicated copper twisted pair is used to carry traffic to each end-user, although there is sharing in the DSL access multiplexer (DSLAM) at the telephone company's central office. Thus, the limiting agent for oversubscription in the case of DSL is the DSLAM. Both cable systems and wireless Internet access systems are more thoroughly shared. Cable systems share the bandwidth dedicated to cable modem use among a number of end-users over a portion of the electromagnetic bandwidth on coax cables that reach all the end-users. Similarly, wireless systems share capacity in the electromagnetic spectrum, either licensed or unlicensed among users. In the case of cable modem and wireless Internet, users experience oversubscription deficits due to the shared last mile link. Consequently, we can expect oversubscription to be a more important economic driver for cable and wireless networks than for DSL networks.

Further, there are subtle but important differences between cable and wireless networks. Both cable and wireless networks increase per user bandwidth (with increasing demand or increasing numbers of subscribers in a location) by geographic splitting: the subdivision of a shared area into smaller units that are independently shared. In cable plant this is called node splitting; in wireless plant it is called cell splitting.

Cable node splitting involves adding additional electronics to a node or placing aggregation nodes deeper in the cable plant (closer to end-users); importantly, though, these nodes are still located within the plant that the cable company already owns and maintains. Wireless cell splitting may involve the more expensive task of finding, acquiring, and equipping suitable additional cell sites in real property not owned (or not yet owned) by the operator (new or existing towers, new building placements) as well as adding backhaul capacity to these locations. Moreover, a key piece of shared infrastructure, the wireless base station, is the subject of intentional cost shifting using advanced radio technologies adding cost in the RF and signal processing portions of the base station in return for cheaper subscriber terminals. This is an eminently sensible choice in terms of overall system economics, but has the side effect of increasing the sensitivity of wireless system economics to oversubscription choices relative to those of other technologies.

The overall impact of these physical differences is that wireless network economics are likely to be more sensitive to oversubscription choices than either DSL or cable networks. If wireless Internet access network economics tend to be more sensitive to oversubscription ratios than other network types, it follows that they will be disproportionately impacted by changes in usage patterns that affect the relationship between oversubscription ratios and user perceptions of service. Earlier, we noted that

---

[4] Note: This is not true in the case of DLC.

multimedia traffic has differing QoS requirements. In particular, real-time multimedia traffic is (usually) not bursty but involves a sustained and steady demand for capacity. Non real-time multimedia traffic is similar to file transfers in that it benefits from high average capacity. However, multimedia files tend to be large relative to many other file types, so we can consider non-real time multimedia to represent the special case of large file transfers.

Unfortunately, either steady demand for capacity or multiple large file transfers work directly against the acceptability of high oversubscription ratios for end-users. In our example using a 10x oversubscription ratio, if any ten of the 100 users attempt to access their “up to 1.5 Mbps” by streaming 1.5 Mbps video streams – rather than doing bursty web surfing – the remaining 90 customers are out of luck (or, more typically, all subscribers will see tangible service degradation). Similarly, if many users decide to simultaneously download large audio-video content files to the hard drives on their computers or portable devices, all users will quickly see much slower performance than the advertised “up to 1.5 Mbps” rates. Hence, a shift towards increasing multimedia traffic as a fraction of total traffic on Internet access networks will increase pressure on service providers to lower oversubscription ratios, and wireless network providers will be disproportionately affected relative to cable or DSL providers. To some extent the impact can be managed by explicit QoS configuration choices, but the overall impact is unavoidable.

In thinking about network neutrality within the wireless context, there exists an interesting history that might influence the thinking of both policymakers and consumers. Our common view of wireless is in terms of cellular telephony and WiFi (and likely more the former than the later). The cellular systems in the U.S. have a long history of using proprietary equipment to maintain control of equipment used on their networks. They also have a long history of embracing closed architectural choices and implementing differentiated pricing models for service selection (for example, paying extra to access the Internet on your cellular phone). In this sense, it should not be surprising to see these carriers taking steps that would violate some aspects of network neutrality. However, users have recently come to think of wireless to include WiFi and most implementations of WiFi (keeping with the traditional Internet model) do not violate network neutrality. As more handsets begin supporting both cellular and WiFi services, it will be interesting to see how wireless carriers respond to users that might make use of their handsets in ways that undermine potential profits for the wireless providers.

The issue of wireless network neutrality has arisen in a separate policy debate, namely municipal WiFi. In a Request For Proposal (RFP) concerning the development of a citywide WiFi network put out by the city of San Francisco, reply comments contained language directly targeting issues of network neutrality. Quoting from the list of “demands,” replies included (but were not limited to the following) [12]:

- Open Access and Network Neutrality
- Assurances that free services don’t lag substantially behind premium services, possibly requiring at least one-third the bandwidth of the premium service.
- Free service that is robust enough for general web use, email and messaging, and VoIP (in the initial rollout)

What makes this interesting is the explicit public call for network neutrality within a wireless network. Also of interest are the explicit demands for measurable bandwidth characteristics and support for real-time services such as VoIP. This shifts the network neutrality debate into defined and measurable expectations.

VII. CONCLUSION

In this paper, we argued that the network neutrality debate hinges on issues that go beyond whether a carrier is unfairly discriminating in their treatment of packets across their network. We considered four issues central to network neutrality. First, we discussed the implications of *oversubscription* on the access networks. Next, we described the lack of *service level agreements* for end-users and the coincident inability for users to *monitor* the quality of the network. Lastly, we considered the evolving architecture of the access networks and what the current design, dominated by *asymmetric uplink and downlink* characteristics means for future services on the Internet.

For network neutrality legislation to be meaningful, it should address the intended goals of a neutral network rather than focus on simply legislating the treatment of "bits in flight." The Internet is a combination of content and distribution, and the limitations of different technologies emphasize one aspect or the other.

References

[1] B. Moyers, "Bill moyers' speech at the national conference for media reform." Transcript available at http://www.freepress.net/news/20357, January 2007.

[2] "At stake: The net as we know it, business week online." http://www.businessweek.com/technology/content/dec2005/tc20051215_141991.htm, December 2005.

[3] "Verizon says google, microsoft should pay for internet apps, informationweek." http://www.informationweek.com/story/showArticle.jhtml?articleID=175801854, January 2006.

[4] "Verizon wireless: Scrap network neutrality." http://techdirt.com/articles/20051110/1313223_F.shtml, November 2005.

[5] "FCC 05-151, policy statement, federal communications commission," September 2005.

[6] Z. Wang, *Internet QoS: Architectures and Mechanisms for Quality of Service*. Morgan Kauffman, 2001.

[7] Qwest, http://www.qwest.com/residential/internet/pricing.html (viewed July 25, 2007). Note: This URL may change as technology and offers change; they may no longer be available or described as this.

[8] Comcast, http://www.comcast.com/# (viewed July 25, 2007). Note: This URL may change as technology and offers change; they may no longer be available or described as this.

[9] S. Newman, “Broadband access platforms, Mckinsey and company.” http://www.fcc.gov/oet/tac/april26-02-docs/BB-Access-Tech.pdf, April 2005.

[10] “Earthlink enterprise t1/t3 frequently accessed questions.” http://www.earthlink.net/biz/highspeed/enterprise/pdfs/Enterprise_FAQ.pdf, 2005.

[11] T. Rappaport, *Wireless Communications: Principles and Practice, 2nd ed.* Prentice-Hall, 2001.

[12] “Webpage: The tech of techconnect.” http://leftinsf.com/blog/index.php/archives/525, December 2005.

[13] K. Moerman, J. Fishburn, M. Lasserre, and D. Ginsburg, “Utah’s utopia: an ethernet-based mpls/vpls triple play deployment,” *Communications Magazine, IEEE*, vol. 43, no. 11, pp. 142–150, 2005.

[14] S. Cherry, “A broadband utopia.” http://www.spectrum.ieee.org/may06/3434, May 2006.

[15] Thomas Karagiannis Pablo Rodriguez, Konstantina Papagiannaki. “*Should Internet Service Providers Fear Peer-Assisted Content Distribution?*” Proceedings of The Internet Measurement Conference 2005, Berkeley, CA, Oct., 205.