



# **Stranded By the Disconnect**

**Bandwidth Supply and the Innovation Economy**

by Professor Michael Kleeman

University of California, San Diego

Senior Fellow, USC Annenberg Center for Communication

## We've Reached A Point of Disconnect

In an earlier paper, I discussed the point of disconnect we have reached between today's customers needs and the traditional voice, Internet and broadcast video network architectures. This disconnect is driven largely by emerging multimedia and multi-modal Internet-based traffic and because of this disconnect our infrastructures and their underlying economics are struggling to catch up.

Three major factors have contributed to this disconnect:

- **Users want to do more with their connection independent of location.** Consumers now expect high quality video, music sharing, real time chat and voice and video conferencing, often all at the same time both at home and on the road.
- **Big files are now increasingly move from computer-to-computer instead of originating on central servers.** The technology commonly used for audio and video file transfers (applications like Limewire and BitTorrent, for example) use participating users' computers to deliver traffic, as if cargo trucks were let loose on surface streets.
- **Latency, jitters, and low quality-of-service are now unacceptable.** The Internet has gone from a best efforts complement to everyday activities to a principal platform for business and personal activities requiring quality of service.

In short, the demands of bandwidth consumers are outpacing the ability of bandwidth suppliers to keep up. There are several supply-side and demand-side management techniques for addressing this growing disconnect, including building additional capacity (increased supply), improving compression technology (in essence increased supply for some services) and triaging network traffic (again, more supply by making do with less).

Increasingly, though, networking companies are considering or implementing options to address the disconnect by *limiting* bandwidth supply and affordability (demand) — including, brief Internet brownouts for high-volume users during peak periods, blocking or limiting some peer-to-peer traffic (famously), bandwidth caps, usage-based pricing, and discontinuation of service for high-volume users.

However, limiting bandwidth supply or increasing consumer prices in ways that decrease demand could have long lasting effects on both the Internet economy and, given the pervasiveness of data communications, our overall economic health.

**Unless we ensure an adequate supply of quality bandwidth at reasonable prices, many current and future business models will be stranded, which will have serious implications for economic growth, sector innovation, and national competitiveness in the Internet sector.**

In this paper, I look at the impact this disconnect has on emerging applications and discuss our options for dealing with the problem, addressing the following topics:

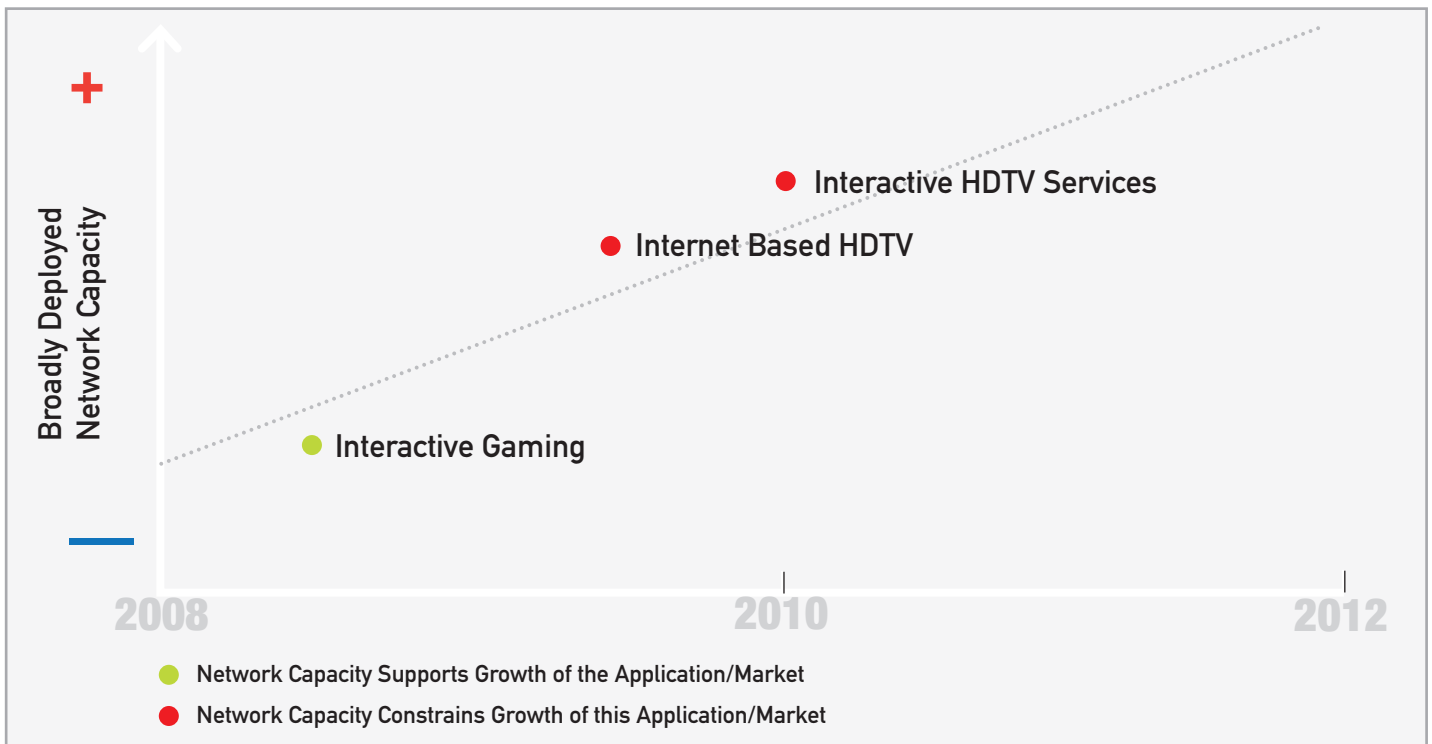
- **What the disconnect means for innovative business models.** In this paper, I profile five key emerging consumer and business applications — video entertainment, gaming, health-care, financial services and video-conferencing — assessing their vulnerability to this disconnect.
- **What we need from the U.S. networks.** Next, I review what network attributes are required to meet the new demands being placed on today's Internet infrastructure.
- **What innovation is required to close the gap between bandwidth supply and demand.** Finally, I discuss the innovation that will be required to develop the networks needed to meet the needs of U.S. consumers and businesses.

There is, or should be, little debate about the importance to all sectors of the economy of our communications infrastructure keeping pace. Accordingly, seeking a better understanding — and even a rough consensus — about the sources and impacts of this disconnect is important in identifying solutions.

## What the Disconnect Means for Entertainment, Healthcare and Business Applications

This paper examines the point of disconnect for applications in industries, which face significant challenges if networks are not sufficiently. In each we briefly discuss which emerging applications require more than the network is currently able to deliver, illustrating the current limits on innovation and growth.

### Entertainment



### Media Entertainment

When we think of radio or television we typically think of entertainment delivered via over the air television, satellite or cable delivered — sent and received in real time. Increasingly, however, these audio and video media are being distributed over the broadband Internet in both linear and time-shifted manners.

Customers utilize directory programs, such as Apple's iTunes or the producer or distributor's site (such as CNN or NBC), to locate the content they want and if necessary pay for the download. Once authorized by the distributor the audio files, music, TV programs, audio books and other specially produced content, are distributed to the user device, an iPod, computer or other media player. Or users can access real time radio and TV programs over

the Internet and access the programming directly from the station or producer, such as NPR, at the same time they are aired, or if desired at a time more convenient for the listener.

As the speed of networks increased the number of devices available to the user for accessing content has increased as well. Video content, originally only available over the air on an analog television when the broadcaster scheduled airing, now can be seen on a TV, computer or handheld device (such as video enabled mobile phones).

This content is controlled by the computer, the handheld, a digital set-top box with digital video recorder or a new class of device, the Internet Media Appliance or Home Media Center. A specialized computer based device designed for simple consumer control, these devices first were based on a traditional computer

platform with specialized software (from HP or Microsoft) but have now evolved to resemble high end high design devices such as the AppleTV. These innovations are made possible by high speed broadband, but their wider adoption is constrained by the limited availability of consumer broadband connections exceeding 20 megabits per second.

Behind these new devices are new services and business models. Media search and download sites, such as iTunes, NBox are the new “zero inventory” record and video stores. Customers

“Content licensing, marketing and sales, customer hardware and software, were all challenges, but without access to customers there was no chance of a successful business. The huge diffusion of broadband today is wonderful. Today we could have been successful, but there just wasn’t sufficient broadband deployment to support the model.”

—Jonathan Taplin, co-founder of internet software company that failed because of broadband capacity.

purchase their content over the Internet at prices comparable to physical products but receive no disk, no box and no packaging. Some sites distribute content for free or at a discount, but insert advertising into the stream or on the web site. And when the customer has registered at the site, and the systems there keep track of who the customer is and what they have examined searched and purchased they can position highly targeted advertising.

These business models were delayed for years by the lack of wide-spread broadband. For example, start up firms such as

!ntertainer, were unsuccessful because networks were not yet ready to support a network-based video distribution business.

Jonathan Taplin, co-founder of !ntertainer, reflects that their challenge was to get the sufficient network connectivity to support a sufficient number of customers to make the rest of the challenges worth while.

Broadband, according to Taplin, is critical, but innovative economic arrangements are equally important. “We also need microtransaction support, the ability to charge for 5, 10 and 50 cent transactions without the costs of doing so undermining the profitability.” And he argues that access to customer knowledge, with the explicit approval of the customer, can improve the user experience. “To really succeed we need to know how to create an ad-insertion infrastructure that leverages the power of individual customer knowledge in the way that Google does, plus with customer permission, add information about customer interests.”

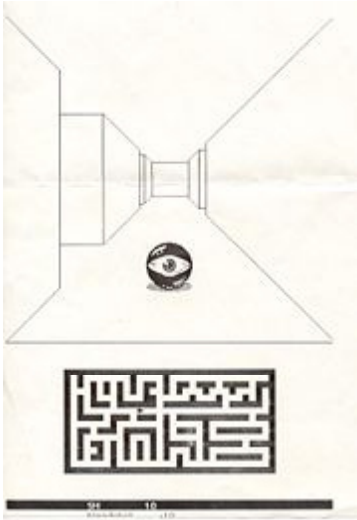
Taplin’s comments remind us that access is only part of the solution to providing services that customers want. But today customers have access to hundreds of programs channels, video and music on demand, plus the ability to store programs for later viewing, and even the choice of different devices to view the programs, at home or on the road. All of this is made possible with technology from numerous vendors and investments by wireless firms, cable and telephone companies, along with their IT suppliers.

Making this possible today are numerous technology elements all based on access to and enabling a high-speed broadband network. The network needs to be capable of carrying more than 20 million bits per second with high reliability and low delay to hundreds of thousands of customers.

The expanded broadband network has driven new device technologies like media centers, new means of compressing information to enable multiple high definition television channels, new handheld devices and even new user-generated content. These technologies, made economic by the mass entertainment market, are capable of supporting more services than traditional one-way entertainment.

The demand for these services, and the enabling value added services infrastructure, are growing rapidly, and these markets are still limited by the lack of high speed access and quality end to end network management. Until these are in place carrier-specific services such as FiOS or U-verse will be able to provide quality video, but other providers will not have access to a uniform and quality network for delivery.

## Gaming



Maze Wars on XEROX Star Computer ca. 1980

Almost thirty years ago interactive gaming between researchers on different continents represented a significant amount of late night traffic on TCP/IP networks. These were visual multi-player such as Maze Wars, originally invented by Steve Colley at NASA in the early 1970s.



Halo3

Today games are much more sophisticated. Halo3, for example, runs as a networked multiplayer game on the Xbox from Microsoft. They are also a significant market area. Before the Halo3 launch in September 2007 retailers took more than 1.7 million preorders for the game. More than 1 million gamers played the game online over Microsoft's Xbox Live system on launch day. In the first week Microsoft reported over \$170 million in sales of the game.

Halo 3 represents the cutting edge of publicly available games today. It is highly interactive and a multimedia experience with high quality graphics, multi channel audio, interactive voice and text chat and players navigating around a three dimensional virtual world. Halo3 is a fast paced game where a delay will cause the player's character to lose territory or even its virtual life. It is a complex program with stunning graphics and player experience, and that is part of the reason it has sold so well.

"We worry a lot about the player experience" said Christopher Novak a Design Director at the Microsoft Gaming, the company that created Halo3. High-speed networks are important to gamers, but what is more important is speed of response or 'ping times,' the delay between when your console sends out a player move and when the other players consoles receive it. "Imagine a ping time of over a second, it would destroy the experience for the player."

"We worry a lot about the player experience. Imagine a ping time of over a second, it would destroy the experience for the player."

—Christopher Novak,  
design director at the  
Microsoft Gaming

Microsoft and other game developers have to deal with the variable quality of the Internet and accommodate it as best they can, given they have no control over the networks. "Players see the game, they don't often think about the underlying network when play gets slow or less intense," said Novak.

Part of the startup of each interactive multi-player game session involves searching the game network for other players and games. "We actively examine each available game to see which would provide acceptable game play performance" said Novak, "we want the player to have a good experience." According to Novak the current generation of games could not be played on past networks and the future generation of games cannot be played on current networks.

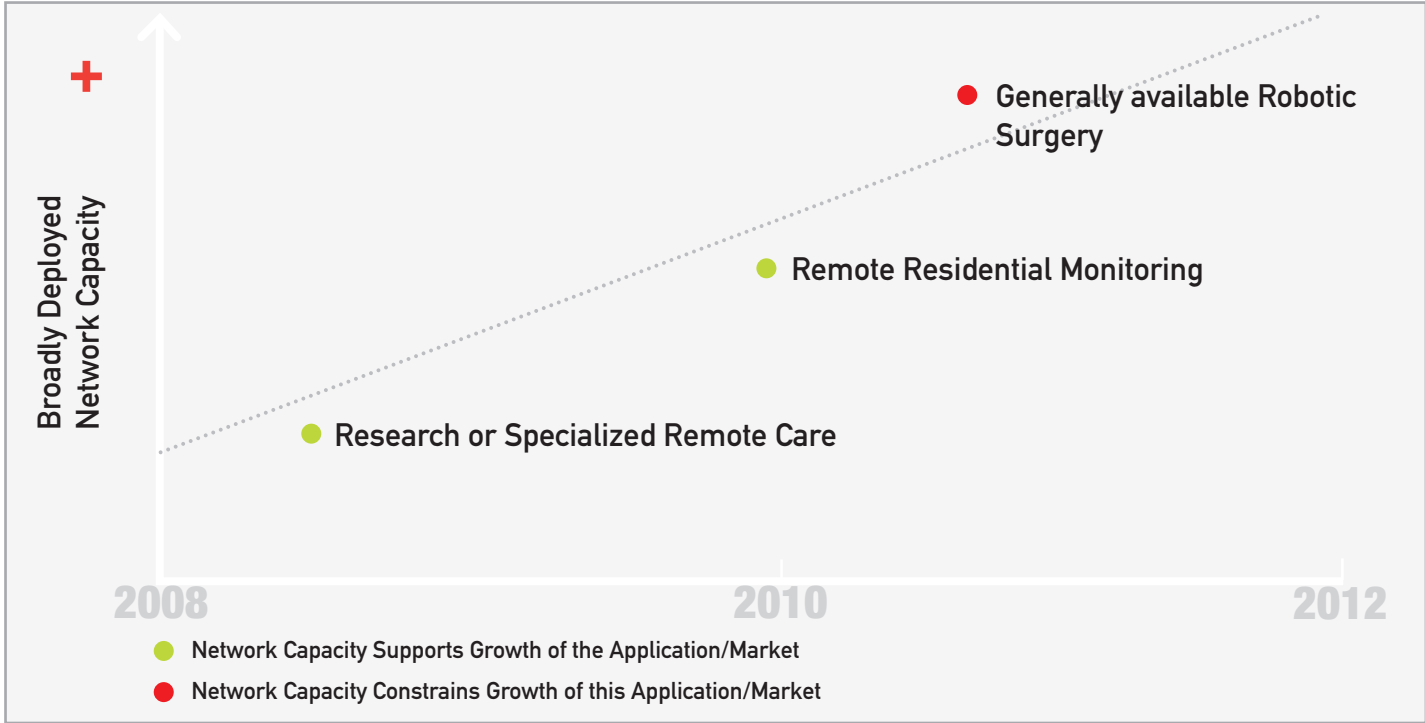
High-speed, low latency, networks supported the adoption of the next generation of interactive computer games. These games are played on both high end personal computers (the video processing is very resource intensive) and increasingly on dedicated consoles like the Microsoft XBOX, Sony PSP and Nintendo Wii. The consoles are optimized for the high end graphics of gaming but also support other entertainment options such as DVD playback and quality audio.

Making this experience possible is a combination of cutting edge computer technology and quality networks. If the network is not reliable then the multi-player on-line game will not be enjoyable and low latency, reliable networks are essential. People can play across the continent or across the ocean with an experience of being in the same room. In many markets this experience is avail-

able now, but as cited above the game developers have to work hard to get around the limitations of today's network that are delaying the next generation of gaming.

Gaming is constrained in terms of the network effect as without good quality fewer players can interact, especially as high bandwidth and low latency are needed.

### Healthcare



In healthcare video has numerous applications, especially in the case of rural healthcare where specialist medical resources are remote or where continuing education facilities are distant. The Federal Communications Commission recently granted \$417 million for the construction of 69 statewide or regional broadband

telehealth networks in 42 states and three U.S. territories to connect more than 6,000 public and non profit health care providers. The network will connect health care sites at hospitals, clinics, universities and research centers, behavioral health sites, correctional facility clinics and community health centers to broadband telehealth networks. The network will provide distance learning and remote consultation and diagnosis, both enhanced by video.



Remote Monitoring

However, broadband speed is not as critical as reliable and managed network access and controlled quality of service for many healthcare applications. "Few applications need very high-speed data, even high quality image studies can be received in a few minutes and still be viable. Sometimes in urgent care it helps, but often these are not distance applications," said Dr. Brent Lowenshon, former director of IT Advanced Technologies Kaiser Foundation, the not for profit arm of the Kaiser-Permanete HMO and current head of medical research and development for the Product Care Institute. According to Lowenshon, particularly for monitor-

ing patients at their homes, “the network has to be reliable or it risks the health of the patient.”

“The network has to be reliable or it risks the health of the patient.”

—Dr. Brent Lowenshon,  
former director of IT Advanced Technologies Kaiser Foundation

Wide scale reliable and predictable networks are critical for Remote patient monitoring, remote or robotic surgery and distant care require networks that stay up, have few errors, and in the case of robotic surgery, have a predictably low delay. Further

adoption of these applications will require quality managed networks, something that is difficult across multiple carriers.

Robotic surgery is being trialed by the military as a way of providing front line care in more locations of higher risk. But civilian applications exist for performing delicate procedures on a small scale, where human hands cannot maneuver with the same delicacy and care. Currently, such devices are nearly always used on local networks but in the future, with quality and reliable networks, they can be remotely deployed and monitored.

Remote monitoring – critical in the weeks following hospital discharge for some major procedures and valuable in the chronically ill – would allow health care providers to better monitor patients at home. Widespread use of remote monitoring would be particularly applicable to situations where the patient still needs monitoring but not access to urgent care, and would lower costs for patients making quality care more affordable.

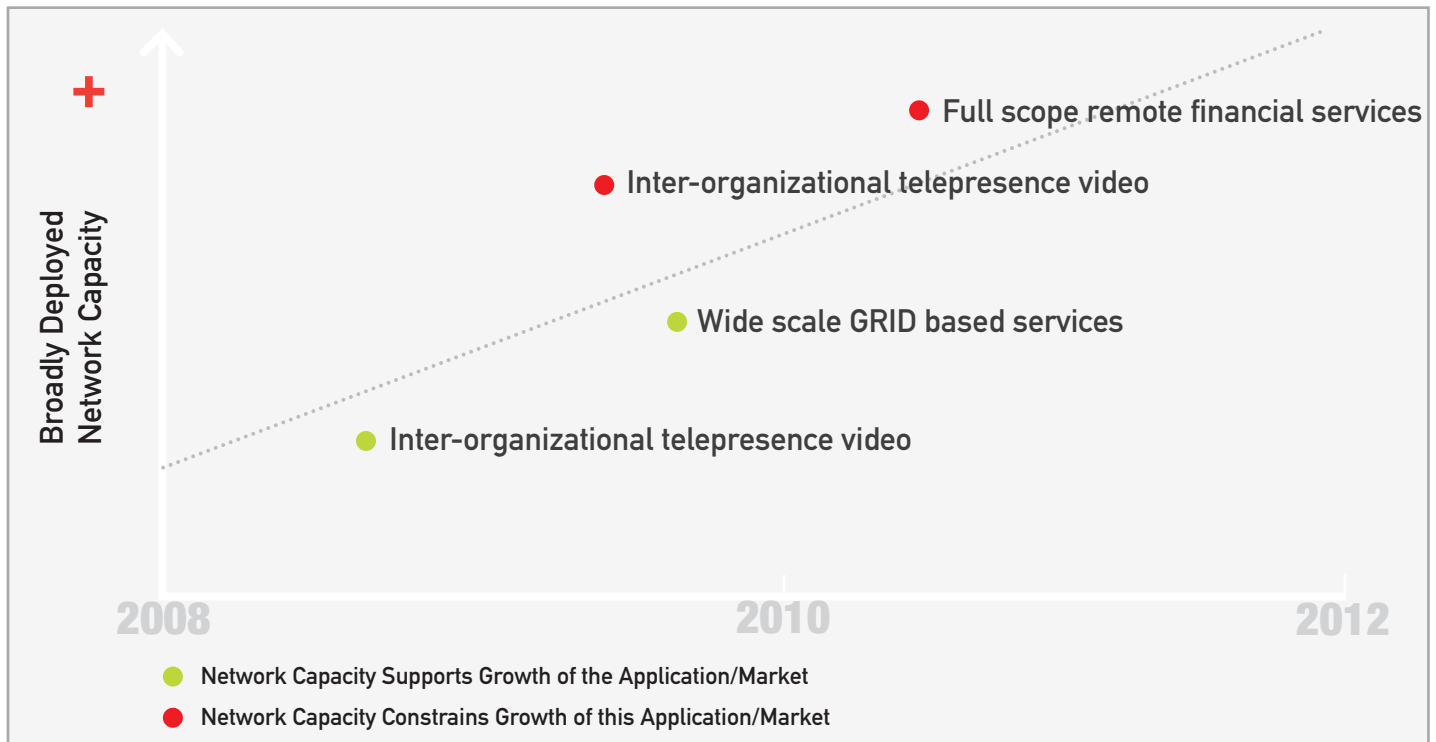
Remote monitoring requires reliable, predictable, networks between patients and caregivers. Additionally, it requires “smart” devices in the home that can communicate the patient’s health status and the status of the device itself to the central monitoring location. With a high-speed and reliable networks, and electronic patient records, doctors will have a more complete picture of your diagnoses, treatments and medications and be able to conduct appointments with patients in their homes. Although they may not be as personal as in-person visits, house calls will return

Healthcare applications highlight the demanding job of maintaining and operating complex networks. In healthcare applications the impacts of failing to provide reliable communications can be more serious. Quality of Service (“QOS”) refers to speed, reliability, packet loss, and the quality of the delivery of packets. In order to support QOS required for emerging healthcare applications, network operators need to accomplish many things, including overall network monitoring (looking at overall traffic

and monitoring for faults) and making decisions on which traffic should get priority.

Currently, networks are unable to meet the requirements of these healthcare applications in any large scale manner. Remote care is in use in research settings, but its potential for wide scale deployment is limited. However, there is strong pent-up demand for reliable, cost-effective residential monitoring capability, which would be a driver of lower health care expenditures. Unless significant improvements are made in quality and reliability the networks, deployment of this cost and life saving technology will be limited.

## Business Applications



### Grid Computing and Financial Services

Financial services firms' networks have evolved from providing community-based services to supporting highly complex, computer models, often requiring massive computational capabilities. The end-to-end speed and reliability of networks are critical, especially in applications such as international currency trading, where thousandths of a second can impact profitability.

This has led to the use of "Grid" based computing networks, which can support the massive computational demand. The term Grid was coined by Larry Smarr former director of the Illinois Supercomputer Center and current Director of the California Institute of Telecommunications and Internet Technology (CALIT<sup>2</sup>), based at UC San Diego. Smarr likened Grid to an electric power grid, where the user could plug into vast resources when needed. According to Dr. Ian Foster of the University of Chicago, an expert on Grid computing, "In financial services, the question isn't so much who uses grid, but who doesn't... Where you see grid applied most often is in areas... where even an eight-hour batch cycle wasn't enough to run reevaluations on a whole portfolio."

Just about all financial service firms use Grid computing – including firms such as Wachovia, Merrill Lynch, Morgan Stanley -- to do some part of their derivatives pricing and revaluations and for risk management, according to Foster.

In addition to Grid computing, which when performed outside the boundaries of a building requires high speed and reliable networks, financial service firms involved in trading depend on the networks for their business. This need for network quality, security and reliability will delay wider deployment of Grid computing on public networks, but the need is growing and is likely to become a problem for financial services companies in the near-term.

### Conferencing

Being in the same room is what meetings are all about, and the Internet can make this possible without traveling. Making this experience as real as possible has been the long-term goal for video conferencing. Early systems utilized a consumer TV and black and white camera, but quickly systems moved to color and multiple flat screens. But the consumer experience was less than perfect and the network was a large part of the problem.

"The connections were based on ISDN and just didn't work reliably," said Bill Gewirtz, who held senior customer support positions at AT&T during the early days of video conferencing. You needed high-speed to make the video quality work and it had to be digital and ISDN was all we had that could be dialed or connected to different customers, and ISDN just didn't work well." Part of the problem was unless the video systems were

directly connected on a private corporate network there was no other digital service available. And ISDN was a 64 kilobit per second speed service, so to get decent quality video you needed multiple ISDN lines connected together. If one line was lost then the signal might drop or degrade.

The Internet solved this problem and high-speed networks make a whole new experience possible. High definition video conferencing with instant, reliable and easy to use connections is now a reality. With the Internet and a high-speed connection there is no need to dial multiple phone numbers and hope the lines stay up. Anywhere you have compatible equipment and a high-speed line you can install and use the system. Long-term video conferencing supplier, Polycom, has seen this transition from tube TV and ISDN to high definition flat screen Internet based video conferencing. Their new systems enhance the experience with positional audio and stunning video. Polycom systems can work for small groups or large meetings and are widely used in government, industry and education.



*Widespread Video Conferencing Requires Upgraded Public Networks*

A relatively new supplier in the video conferencing marketplace, building on its vast experience with Internet and broadband services, is Cisco with its TelePresence product. The Cisco project was lead by Michael Dhuey, an Apple Computer veteran and co-developer of the iPod. He brought camera technology, displays and visual effects capabilities to the project, working with experts from the motion picture industry to create the proper lighting for the TelePresence product. One user found the experience so realistic he almost went to move items out of the way in the room only to find that they were in the remote city. In a form of technological quid-pro-quo the TelePresence suite was featured in a TV series “24”, where technology and Internet communications are a feature of the story.

The key to this technology’s rapid acceptance is the ease of use and cost savings. In fact products like these can pay for them-

selves in multiple ways, both economically and environmentally. By reducing travel to remote locations personnel save time and money, and gain productivity. Executives can meet with clients from six cities and three continents in one day without leaving the office. The return on investment on even an expensive system can occur in one or two years. Key to greater economic and social benefits is expanding the use of the systems beyond the company walls. The larger the network of high-speed connections available for video conferencing connections the greater benefit for all users. Ultimately, this argues for the use of the public networks for these services.

The environmental impacts of travel substitution are considerable, especially with repeated meetings (such as project reviews) or large group meetings. But these technologies use less energy and pollute less than even a single long distance airplane trip to attend the same meeting. Work is underway to further quantify the detailed savings but one goal is to utilize video teleconferencing systems to offset more than the operating impacts of the IT and network industries, effectively making the technology sector at the least carbon neutral. With a network connection of the right reliability and QOS one could even connect in from home, permitting people to work from home occasionally, and still be fully engaged in the work at the office, or across the globe.

Video-teleconferencing is enabled by the Internet but is a demanding application, especially when multiple high definition screens are used to enhance the experience, which requires hundreds more times the capacity than a phone call. In the same way that peering and network congestion can cause problems with gaming quality, video conferencing quality can suffer if the network is not running properly. Having all of your video systems on a managed corporate network can address some of this, but it to maximize the benefit use of the public Internet is required. As these systems increase in number and spread out in geography and networks utilized we can expect to see QOS problems increase. This will create further pressure to increase network speed, quality and reliability and delay its wide adoption for several years.

Major corporate customers avoid some of this risk by getting a managed network service all from one carrier and this approach supports most corporate video services today. This avoids the peering risks and in many cases traffic from one company location to another never go over the public Internet, but rather a special parallel network running the same Internet protocols of TCP/IP and interconnected to the public Internet to let corporate

users access non-company sites. Some firms utilize an entirely separate TCP/IP network which is not interconnected with the public Internet for performance and privacy reasons, as does the US military.

However this does not address the needs of inter-company communications, so the public Internet must be improved to reliably support growth of these applications.

## What We Need From the Network

What we will explore here are some of the end user services that the Internet enables and core functions necessary to create these services. We will also examine how they work, and what needs to be done to insure that these networks grow to meet the expanding demand for end user services and continue to fuel the economic and social engine that is now such an integral part of our daily lives.

Here are some of the different dimensions of service and how they impact end user applications capabilities.

### Speed of Access

This is what most consumers look for when they order Internet services, how fast it will operate. Usually measured in megabits, it is a key determinant of the user experience in web browsing and file downloads. As we will see later it is not the sole determinant of experience, as many other factors come into play.

### Speed Across the Network

This is what gamers and financial traders look for, how fast their traffic gets to the servers (computers that provide the content or services) and other resources located in the ‘cloud’ of the Internet. Measured in ping times (how fast the signal get to the remote location) it is gaited by access speeds but many other factors influence this speed, especially the number of network connections the signal must transit to get to its destination.

### Reliability

We take plain old telephone service “POTS” reliability for granted. Until recently most POTS lines were copper and were powered from the phone company central office. They seldom failed and when they did were back up quickly providing the basic service of POTS. This kind of reliability is essential if we are to use networks to help in home monitoring patients who have been discharged from a hospital or have chronic illnesses.

Internet service (and services such as voice over Internet protocol or “VoIP”) are more complex and require power locally, a complex of hardware and software to operate and when they work are amazing. When these services fail they are harder to diagnose and repair. Than traditional phone services. We expect our cell calls to drop, on occasion, and we expect our Internet connection to fail at times. However our experience and expectation is that land line calls, emergency calls and health related services have high reliability. This goal can be realized but will require focus and investment.

### Always/Anywhere

Mobility is somewhat new in communications, it is only since 1984 that we have had cellular phones and only in the last 10 years that we have had mobile data. The ability to have service always and anywhere changes the way we work, how we care for our families and even how we play. But to make this work requires sophisticated products, remarkable power systems and power control (that small battery that would not run a flashlight for 10 minutes runs your cell phone for days) and expensive and complex networks.

### Security

The Internet is a shared network, all of our traffic is intermixed with that of the other users. And wireless data is carried over radio signals that anyone can receive. How we guarantee that our communications are protected from unlawful receipt is much more complex today than it was in the POTS era. When important information, from our private conversations to million dollar financial trades are carried on networks secure communications matters.

### Management

In order for the video your are watching on your PC to be of any use it must download quickly and reliably. Not only are there myriad dimensions to manage but there are multiple companies providing carrier services and even more providing end user applications. Coordinating and managing these services to provide a quality of service (called QOS by carriers) is important to users and demanding for carriers.

## Network Innovations Required to Address Bandwidth Demand

Moore's Law famously states that computing power doubles every 24 months. Practically speaking bandwidth-consuming applications are constantly in a race with the ability of the network. If network capacity pulls to far ahead of demand, network investments become stranded. Today, we're seeing industry applications that are stranded by a lack of network capacity. Innovation at the network level is required to address this disconnect.

### Peer to Peer – Sharing the Wealth and the Work

In the very early period of the Internet content, text, web pages and even music, was requested from an site and delivered from computers, called servers, at that site. As the Internet grew in connected users and geographically this model became difficult to scale, especially to meet peak demand and demand from overseas. So a new architecture was born called network-based caching. Content was copied to multiple servers in data centers around the world and when a user wanted to access, for instance, a web page, they would be directed to a nearby server. These were shared amongst multiple content providers and offered better performance and economics than most providers could make available on their own.

These caching networks evolved to dedicated content distribution networks (CDNs) optimized for the high capacity and bandwidth of audio and video content. But network capacity and storage can be expensive if your customers are asking for thousands of copies at once from around the globe. And this cost limited the distribution of large files such as video streams to those larger players who could afford these high costs.

Innovators in Europe and the United States created a new model called Peer to Peer (P2P). In this model a central director computer would distribute parts of different content streams to different end user computers running the necessary software and then when someone wanted, for instance, a video, it would direct the requesting computer to ask each of these computers to send it their parts. The new recipient would then become another node in this distribution network. P2P networks, such as BitTorrent and Limewire, grew to share files for free between users (often violating copyright in the process).<sup>1</sup> Recently these networks became a point of contention between carriers, such as Comcast, and some

users, since one P2P user can overwhelm the local network connection denying others sharing that connection a decent quality of service.

While the P2P approach is a convenient means for audio distribution that bypasses centralized control and (sometimes copyright), when it comes to video distribution it becomes quite economically attractive. The BBC and numerous corporations use a P2P approach to video distribution called Kontiki from the firm Verisign. Kontiki, like other P2P programs, stores and distributes video content not from a centralized set of servers but from each computer that has obtained a copy. P2P spreads the cost of distribution amongst the users who are paying for their own hardware and bandwidth costs.

But P2P demands that all network connections being used have to be robust enough to support the P2P traffic without compromising the effective use by others sharing these access resources. Easier and lower cost distribution of large media files means more people will share them, and all of us benefit because of a broader selection of content from more and often less known content creators. So while peer to peer enables new services (such as certain forms of VoIP) and services providers it does place new demands on the network. Where the network is not properly engineered for symmetrical two way traffic it can create bottlenecks and cause degradation of service for some users. In order to address this need and take advantage of P2P carriers need to add new capacity, especially in the access segment of the network.

### Adding capacity – Building Bigger (and More) Highways

The amount of information that is moved over the Internet is increasing daily. Compression technologies minimize the amount (and cost) but that only makes it more attractive for people to load more on the network. Two recent papers explore this constraint, one by this author<sup>2</sup> and more recently one by Nemertes Research, suggesting that unless more capacity is added to the networks innovation and economic growth will be stifled. Adding capacity in the core of the network is a scaleable and comparatively (in engineering terms) straight forward process, since core capacity is shared by so many users and easy to monitor for growth. Capacity in the access network, that is getting from the user to the core, is where the biggest constraints, and risks, are.

---

<sup>1</sup>The same P2P logical architecture underlies VoIP (Voice over IP services) such as Skype where the central server or directory provides the connection details but the call traffic does not route though the central server, rather it goes directly between the two (or more) users.

<sup>2</sup> Michael Kleeman, *Point of Disconnect* --August 2007.

Nemertes Research, *The Internet Singularity, Delayed: Why Limits in Internet Capacity Will Stifle Innovation on the Web*--November 2007.

With the advent of mobile Internet access we now have two basic ways of getting to network resources, wirelessly and over some form of physical link, like a phone or cable line. And users expect that no matter how they access the network they can get the content they want, be it music, their email or even TV programs. But each of these network access approaches has its limits.

In the wired network service providers use a combination of fiber optic cable (plastic or glass fibers that carry signals in the form of light) and copper to reach the customer. There are differences in how they use each and how close the fiber is extended to the customer's premise. Verizon has an architecture they call FiOS that takes fiber-based signals to the customer's home and then uses wireless (WiFi, which is comparatively short range wireless Internet access) and copper in the house to distribute the Internet access and copper for the TV signal. AT&T calls its approach U-verse and it takes fiber to a device near the home and then uses copper to the home and, like Verizon, WiFi and copper in the home. Cable companies, like Time Warner, Cox and Comcast, use fiber to the street and then coaxial copper to and in the home. So despite the differences promoted by the carriers, they all use a mix of fiber and copper to get voice, data and video to and in the home. Each of these access (also called last mile) technologies has advantages and disadvantages, but as they are more broadly deployed they will let customers access the Internet at speeds between 20mbs and 100mbs. This is fast enough for a compressed high definition video over the Internet, very fast web browsing and voice services.

The wired world is moving quickly to serve customers' needs for high-speed service at the same time that many users are increasingly going mobile with their data and video. While it is easy (albeit expensive) to add capacity in the wired world it is much more complex to do so in the wireless domain.

Radio is a shared medium, everyone in a town can receive the local AM radio station and hear the same program. If I want to add more programs I need to add more radio channels. A fiber to your home is a dedicated medium, and if I want to give your neighbor more capacity I can put a separate fiber into their home. Also fibers can carry billions of bits per second and copper can carry tens of millions, effectively making wired access network capacity fast enough for any user. Radio signals have different characteristics by frequency or the radio channel, and it is far less than fiber, especially if it has to be shared by thousands of users.

The early mobile telephones were wide area phones with few users and expensive. Even in what AT&T then called Improved Mobile Telephone Services (IMTS) were only 11 channels, *in an entire city*, and the unit was hardly portable. This was the state of the art of mobile telephony for 20 years until AT&T Bell Labs invented and then the phone industry deployed cellular telephony, leading to the mobile networks of today.

While the IMTS service used one set of frequencies to cover a whole city cellular technology divided the city into smaller areas or cells. In doing so the system could reuse the frequencies many times so that a channel could carry numerous calls at the same time, in different cells. The phones, together with a central switch, were able to change from one cell to another, and one frequency to another, without dropping the call.

The cells were connected to the central switch by cables, fiber or radio links. The cells can be larger (called macro cells, originally 8-10 miles across) or smaller (microcells). Each time the carrier splits the cells and makes more, they add new masts, antennas and radios, and increase the capacity of the system. This concept of spatial reuse is what makes cellular so valuable and permits it to continue to grow.

If the trend toward smaller cells and the trend toward mobile devices being used as a substitute for wired devices then the next logical step is a cell in a home. This concept is becoming a reality in the femto cell.

“The subscriber gets seamless and higher speed mobile service, the carrier effects an increase in their network capacity without building hundreds of new towers and radio systems and users in their cars or in the street get better service. It is a win all around.”

—Chris Rice, Chief Technology Officer at AT&T

The high speed line (using fiber or copper) connecting the Internet to your home connects to the Internet and the cell network, and your home becomes a very small (or femto) cell. Cellular coverage in the home is excellent, and the signal is carried over the wire or fiber to the home, thus taking the load off of the public cells and effectively increasing their capacity. Chris Rice, the Chief Technology Officer at AT&T

is a big supporter of femto cell technology, “the subscriber gets seamless and higher speed mobile service, the carrier effects an increase in their network capacity without building hundreds of new towers and radio systems and users in their cars or in the street get better service. It is a win all around.”

## Network Management and Control

In 1980 when mobile meant IMTS there was only one radio tower in the city for telephone service. Now, with femto cells, there will be, in effect, thousands. There are millions of high speed access lines, computers and wireless phones and data devices. The volume of data being managed is massive and the threats to the network numerous. In order to insure that this complex of equipment, traffic, signals and applications works for the user the networks need to be well engineered and constructed. But they also need to be actively monitored and managed. This is the job of the network management and control organizations at carriers, and the network operating center or NOC.

In addition to monitoring the health of the network and providing technical support to keep the network working, the function of network management is to protect the network from attack and insure that customers receive the service they expect and pay for. The idea of QOS was discussed earlier and monitoring congestion of the network and removing the cause of it is part of the role of the personnel in the NOC. The NOC is also responsible to provide data to traffic engineering departments where network upgrades are planned and peering points designed, sized, and upgraded.

Examining traffic flows, and the specific sources and types of traffic is important in maintaining network quality. In order to do that it is important to have the switches, routers and special purpose computers, like systems from Narus<sup>3</sup> examine the contents of packets. If this process discloses millions of media packets from one user that are slowing service for others the NOC can limit the speed of that user or if that user has paid for a higher quality of service the NOC can insure their packets get priority.

Part of the value of what is called deep packet inspection is that it can disclose attacks on the network that might cause slowdowns or failures. The simplest one logically is a denial of service (or DOS) attack. The basic idea is to overload a network asset with traffic so it cannot provide normal service to all customers. DOS attacks can be targeted at content servers or routers, and usually come from one source through which viruses generate attack messages from thousands of computers all over the network. Without visibility of the entire network (to see where attacks are coming from and blocking them before they enter the core) and packet inspection the NOC might not be able to divert a DOS attack and if it did it would be hard to determine what traffic was legitimate and let it pass thru. Attacks can also contain computer viruses targeting servers or network routers and packet inspection is critical to determine and protect against these attacks. Attacks are common in the Internet, they occur thousands of times a day and originate from the U.S. and dozens of foreign locations. The fact that most of us never experience the impacts of these attacks is a reflection of how well the NOC personnel at the different carriers do their job.

## Power – Making Things Run While Using Less Energy

High-speed networks, compression, mobile devices, caching servers, network operations, even femto cells, all have one thing in common, they need electricity to run. Power is the fuel of communications, nothing can work without it and the industry is in a constant battle to increase capabilities while reducing the power requirements of those capabilities. Consumers want more speed, more features on their phones, larger screens and better quality pictures and all of this requires processing and power.

Fortunately there are three trends that help attack this issue and keep our phones and networks working late into the night. The first is the steady increase of the price and energy performance of the brains and heart of most of the devices in telecommunications today, microprocessors. The price and unit power for the devices that run a cell phone have dropped steadily, and designers have taken advantage of this decline in costs to add functions and fea-

<sup>3</sup>For an interesting illustration of this kind of protection approach see [www.narus.com/NSSdemo/NSS\\_Demo.html](http://www.narus.com/NSSdemo/NSS_Demo.html)

tures, like video and Internet access, to the mobile device. Likewise home devices are faster and more functional, less expensive than their predecessors and often smaller.

The second trend is the improvement in battery technology. Today most mobiles (and portable computers and even the Tesla electric car) use lithium ion (Li-ion) batteries. Li-ion batteries hold more power per unit of mass, can be recharged more and lose less capacity over time (but eventually like all chemical batteries do need to be replaced). They are certainly better than those before them, but they are not ideal.

All batteries take time to charge, degrade and use chemicals most of us would rather not touch (additionally Li-ion can catch fire or explode if not properly managed, so Li-ion batteries include power management controllers). Over time many designers and scientists see specialized capacitors being used to power our mobiles.

These can be charged very fast, do not degrade, are not toxic and high power densities.

The third trend in power is a concern to reduce the power requirements of network devices and the power consumption of all devices when not in use. New technology can reduce the amount of energy used by the central radio equipment in mobile networks by almost 40%, while improving signal quality. Phone companies are actively working with their equipment providers to reduce energy demands, effectively giving us more digital bang for our kilowatt (and carbon) buck.

While network power consumption is important, it is only part of the picture. Over 10% of the energy used in some homes is consumed by devices not in use, but in standby (the TV that comes on when you press the remote button is one example) The Energy Star goal for products when not in use of under one watt is significant. Computers, printers, set top boxes and Internet access devices all consume power and most are not in use the majority of the time. Selectively powering down or significantly reducing the power they use can have major impacts, literally reducing the number of new power plants that are built in the US in the coming years. Power is important and the less we use for each task the more we can do.

## **Investment – The changing demands on carriers**

The communications services industry in the United States drives big investments. Investments in major programs, such as FiOS and U-verse will run into the billions of dollars and generate thousands of jobs. These investments, and those of the cable, wireless, satellite and Internet content and services industry are major drivers of the economy. When new radio spectrum is auctioned by the FCC the costs just to acquire the right to use the frequencies can run into the multiple billions and then the winning bidder has to invest more to build out the network. The communications business is big business, and it has big risks. Unlike the days of POTS there are no guarantees that carriers will earn a return on their investments, yet much of the industry is still regulated in ways that were created when profits were guaranteed.

The investments in core Internet services is well understood and demand for new services, while not predictable in specific volumes is predictable in terms of its direction, up. But the investments in access services, like wireless, fiber or cable to the home, are risky, since there is a large central investment plus a cost to connect each home. If a certain number of customers in an area do not subscribe much of the investment is not in productive use.

These scale effects, where a high fixed cost plus lower incremental costs, drive the economics, can lead to profits when there are a lot of subscribers and losses when the customers go elsewhere. This is part of the reason that carriers market promotion prices to get customers connected. New services, new spectrum and new carriers all bring added risks to the new entrants and to the incumbents. In the communications marketplace carriers have always faced risk, from the market and technology. However the uncertainty of regulatory changes can cause unanticipated risks that all players would prefer to avoid. Regulations need to be streamlined to insure that a few key social needs and rights are addressed, but not in ways that are overly burdensome to carriers. These needs include universal service, nondiscriminatory access, emergency calling, legal intercept and interconnection of networks. None of these demand micro level management or regulation.

The Internet market will impact billions of dollars in business as new markets for content, advertising and services emerge. All of this will drive new technology investments and innovation. It is a virtuous cycle where the players bet big and hope for big returns, and in the process the customers benefit.

## Summary

The communications sector has always been a leader in technology innovation and the Internet has and will continue to change the way we all live. Because of the Internet and the government and corporate organizations that helped create, build and run it we have new firms like Amazon, Google and Yahoo, computers are true global communications tools, and independent media creators have a new voice and market. The cycle of innovation will continue and spur new technologies, firms and breakthroughs.

In order for this to occur, however, we need to be mindful of what it takes to make it all work. Monitoring and managing the network is hard work and requires the right tools so that the benefits we all realize can continue. Especially in the broadband media area the need for business and technical arrangements for QOS and cross network peering are critical. And we as a society need to recognize that a reliable, robust, communications network is critical to our well being and the incentives need to be there for its construction, operations and maintenance. Delays in realizing this infrastructure will impact business, entertainment and health-care and risks stalling the cycle of innovation that has fueled much of the technology growth in the past 20 years.

In this paper we have discussed innovations in applications and technologies and their benefits to society. What we have not discussed is that much of the innovation is generated not by individuals or organizations seeking profit but researchers and their sponsors who work to create new tools and capabilities to better the way we live. It is to these innovators who dream, create and share that this document is dedicated.