December 2002

CONVERGED NETWORKS AND THE GRID: 21 ST CENTURY INFRASTRUCTURE CHALLENGES

Samir Chatterjee
Claremont Graduate University

CONVERGED NETWORKS AND THE GRID: 21ST CENTURY INFRASTRUCTURE CHALLENGES

Samir Chatterjee
School of Information Science
Claremont Graduate University
samir.chatterjee@cgu.edu

Abstract

This tutorial paper discusses issues in two fundamental technology paradigms, namely, "network convergence" and "Grid" that will become the foundation of the 21st century information infrastructure. These emerging networking and computing technologies are poised to impact education, research, and business. We present packet voice and video architectures and discuss fundamental changes taking place within the telecommunications industry due to convergence to the Internet Protocol. The impacts of network convergence on carriers, enterprises and service providers are presented. We then discuss the impact on end-users and the importance of public policy to promote competition in the convergence space. Finally, we present the "Grid" computing architectures which is an emerging field distinguished from conventional computing by its focus on large-scale resource sharing, innovative applications, and, in some cases, high-performance orientation. Both the core technological components and the value proposition for businesses are discussed.

Keywords: Convergent networks, The Grid, packet voice and video technologies, infrastructure, voice over IP

Introduction

At the infrastructure level of information technology, two fundamental paradigm shifts are underway:

- "Network Convergence" refers to the global Internet (and its associated TCP/IP protocols) becoming the basis for transmitting all media types including voice, video and data (Dunstan, 2001). This infrastructure convergence (also called voice/video and data network) is already having a deep impact on the telecommunications industry networks, business model, and strategy.

- The "Grid Problem" which is a top priority research program funded by National Science Foundation, is defined as flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources (Stevens, 1997). "Grid" computing differs from conventional computing by its focus on large-scale resource sharing, innovative applications, and in some cases, high-performance orientation. Grid will enable truly virtual organizations that not only communicate over IP networks but also will enable collaboration across boundaries.

Network convergence and grid together present a solid foundation for an IT infrastructure. This tutorial explains the basic issues involved and discusses the challenges that remain.

In Section 2, we present a brief history of the development of packet voice and video technologies. The implications of convergence on industry are outlined in Section 3. Section 4 highlights the impact of convergence on service providers and enterprises. We discuss how enterprise infrastructure in the 21st century will gradually change to enable a single integrated network. Examples with Voice over IP are presented. Section 5 lists some of the technical and business challenges of convergence. Section 6 presents the basic issues of how "Grid" will work and the challenges of building such virtual organizations.

Packet Video and Voice Architectures

In 1996, the U.S. Federal Communications Commission paved the way for open competition in the telecommunications marketplace, promising new opportunities for business and better value for customers. Today, thanks to Internet Telephony, the promise of an open world of communications is starting to come true. New innovations and cost reductions are taking place that will fuel higher productivity, lower costs, and increase profitability on a global scale. Both businesses and end-users will benefit. This open communications also hastened competition and squeezed the margins of all participants.

To understand the industry and various firms involved, it is necessary to take a quick look at how IP Telephony or Voice over IP (VoIP) technology works (Schulzrinne and Rosenberg, 1999).

The two scenarios for using VoIP are (1) PC-to-PC and (2) PC-to-phone. When a PC-to-PC set up is involved, the telephone network (PSTN) does not play any role. The VoIP soft client on a user's PC encodes the human voice, places it into IP packets and sends it from the source LAN. These packets travel through the Internet across many routers and networks until the message arrives on a user's PC, where the digitized voice information is decoded and played through the PC's sound card and speaker system. A multimedia PC is assumed in this case. Since the entire call path stays on the Internet, the cost to the user is zero. In the second case, however, when a PC needs to call a regular PSTN phone, the two worlds of Internet and PSTN must be connected. Today the connection is through hardware called IP Telephony gateway (DeSerres and Hegerty, 2001). The sender encodes voice into IP packets and sends it across the Internet. However, this time the packets reach an IP telephony gateway where information is decoded and terminates onto a circuit-line to the destination phone through the "last mile" copper wires. Some termination charges are incurred. For interoperability and signaling issues to be resolved for such connections, two standards (an ITU-T standard called H.323 and IETF standard called SIP) are competing for dominance within the VoIP industry (Schulzrinne and Rosenberg, 1999). The additional components include the use of gatekeeper, servers, signaling set up (and interfacing to SS7 network), accounting, access, and billing.

The evolution of IP Telephony started with hobbyists who first developed and experimented with voice over the Internet for curiosity sakes. Commercial viability of VoIP took shape in 1996 when a small Israel-based company called VocalTec manufactured IP telephony gateway that would connect the two disparate worlds of IP and PSTN. This product led some Telcos to jump onto the bandwagon to leverage "rate arbitrage", i.e., enabling long-distance and international phone calls at a fraction of today's cost due to toll-bypass. These providers gave birth to a new industry called ITSPs (Internet Telephony Service Providers) that simply connect two (legacy) phones, but a large portion of the call would be placed over the Internet. ITSPs are placing gateway POPs at various cities to originate and terminate calls. The third generation of service providers started building networks that integrated voice and data. These are companies that are building IP-based PBXs, and voice-enabled routers to put voice and data traffic on a common network. The fourth generation or next-generation Telcos include CLECs, Packet CLECs, and others that offer far more advanced and enhanced application services to the end-user.

Video coding technology also progressed enough to reach a stage where it is feasible today to send high quality digital video over the Internet (http://middleware.internet2.edu/video). Both H.323 and SIP are being used as signaling protocol for establishing packet video calls between end users and also for multi-party conferencing. Service providers and carriers thus started upgrading their networks to carry voice and video along with data on their IP networks. The belief is that a connectionless IP network is cheaper to build and provision than a circuit-switched network. Moreover, there is industry-wide optimism that a converged voice and data network can be used to provide next-generation services, thereby increasing revenue opportunities.

Evolution of Telecommunication Services

The telecommunications industry, its revenue and business model and its technology is changing quickly. The continuing process of deregulation combined with remarkable network technology advancement and the advent of new service offerings via the Internet provides a powerful push for evolution. Digitization of information is at the heart of this change.

The Past "Old" Structure

In the early nineties, one could present a simplified illustration of the three major communication services that reached end-users at home or business (Clark 1998) (Figure 1). Plain old telephone service was carried over twisted-pair copper wires and a telephone device was used to talk and hear. Coaxial cables carried analog television signals that were one-way distributed through TV sets. On-the-air TV signals were broadcasted by television stations. Radio programming which was also carried over the air was accessed through transistor sets. The three important points about this simple old structure are:

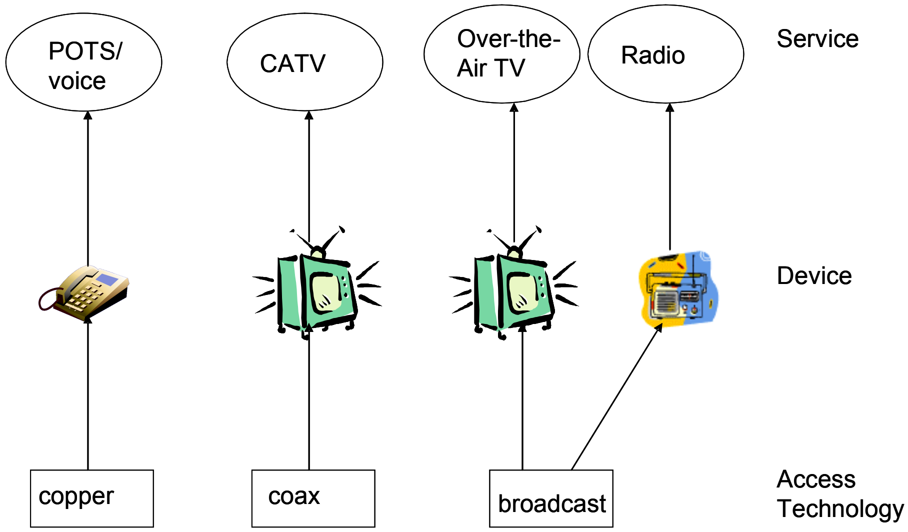

Figure 1. Past Communications Industry Structure

(modified from D. Clark (1998))

- A tight coupling of the delivery technology and the service. The service providers who installed and maintained the copper pairs and central office switches understood the requirements of telephone service.
- This clarity about service and what specific technology was needed to optimize that service so that end-users received maximum value led to clearer revenue and business models. Investment in infrastructure meant better and enhanced services in that business which led to greater profitability.
- In almost all cases, service and technology was provided by one and same company leading to a vertically integrated industry structure.

The Present "Emerging" Structure

After the mid 1990's, this picture began to change significantly. With Internet service becoming a huge commercial success, several "old" industry players felt that perhaps it was better to be in multiple line of business, which would increase revenue and hence yield profitability. Along with the three common services, Internet service suddenly dominated the picture (Figure 2) and two new access technologies came into commercial viability: (1) wireless (or cellular) and (2) satellite. The solid lines in Figure 2 indicate existing service while dotted lines show reasonable possibilities. The several important points about this rather complex picture are:

- The vertical industry relations were replaced by a complex matrix structure.
- The Internet is uniquely positioned in that several services can be delivered over it, and the Internet can be delivered over all the access technologies.
- Besides the well-known end-user devices, several new devices (cell phones, satellite handsets, web-appliances, PDAs, smart wrist-devices) came into the picture.
- Although not depicted clearly in the picture, the underlying technology for service delivery shifted to packet-switching and heavy use of web-based technologies.
- Various service providers are in multiple line of business and their networks are being used to deliver services for which their performance was not optimized.

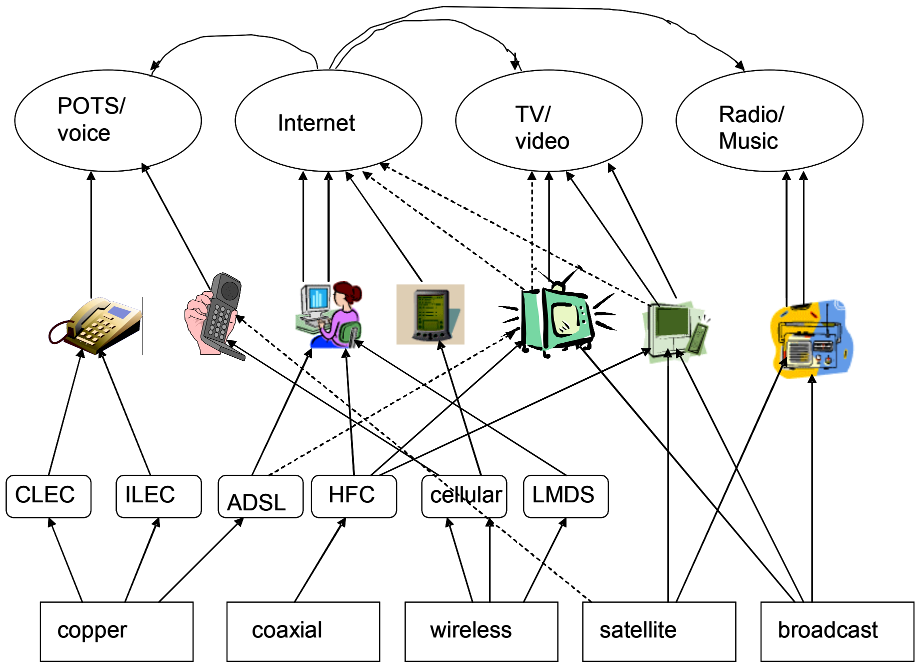

Figure 2. Present Complex Matrix Type of Industry Structure

The Future "Possible" Structure

Although it is extremely difficult to predict the future of the Internet, it is reasonable to believe that it will become the common "converged" infrastructure that will provide a host of existing services as well as a number of new enhanced communication services (Figure 3). What will be the type and nature of these services? What will be the implications on industry structure as these converged services evolve? To answer these questions, one needs to understand two salient characteristics about services:

- The open structure of the Internet led to a flurry of innovations in a very short span of time compared to any other service evolution in the past. Hence it is imperative that the future course of the Internet preserves this open standardization process. Open standards would further fuel unprecedented innovation rather than inhibit them as some of the present forces seem to indicate.

- The type and nature of services that will be successful depends largely on what perceived "value" it brings to enterprises and individuals.

New services and business paradigms are being created by the development of several technologies at once, technologies that were developing more or less independently but that now don't seem independent at all (Bartholomew, 1997). Businesses are spilling into each other's areas. Competition is opening up old protected economies. Political and national barriers are less of a factor than before. Access technology and access device will become seamless and we will begin to see the notion of service ubiquity. Technological convergence (Bartholomew, 1997) enables ultimately, "any client, any network, any bit, any format". With it only one thing is certain: Enterprises will see increased global competition, their profit margins will shrink and only the nimble adaptive enterprise will still remain in business. Understanding convergence and its power is a key first step towards a firm's survivability.

Impact of Convergence

Service Providers

In an environment where convergence is eradicating boundaries between telecommunications, computing, and entertainment, what can service providers do to succeed or even survive? The successful service provider will:

- Deploy an efficient, robust, cost-effective multi-service network infrastructure
- Offer a vast array of value-added, customizable enhanced services
- Must make money delivering

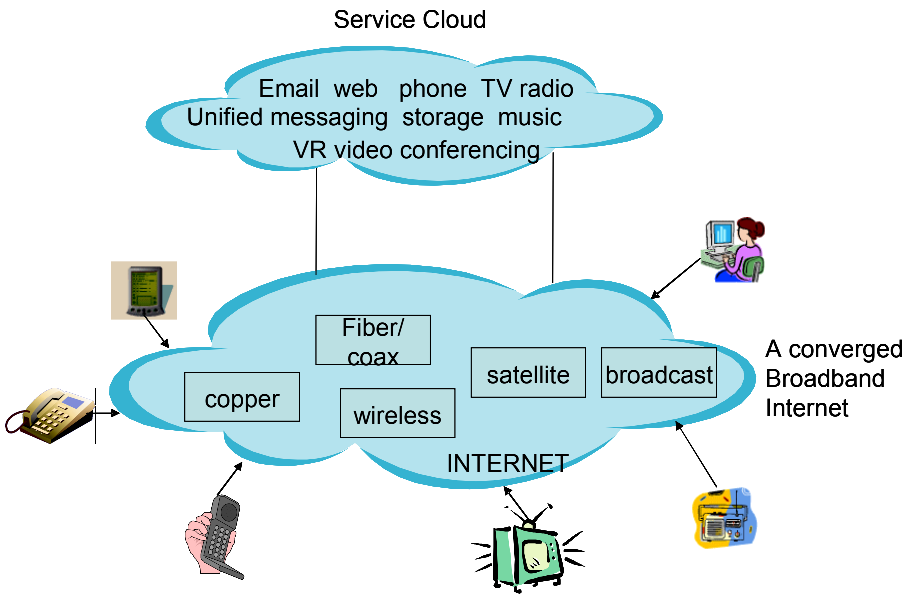

Figure 3. A Future View of "Converged" Network and Service Cloud

These objectives are not really new. What grants them a new flavor is the increasing rate of technological advancement that is currently feeding the advent of an all-IP network (Serras and Hegerty 2001). The Internet has the potential to unify all media types (voice, data and video) into a global infrastructure that will deliver enhanced services. Service providers also face extreme competition because of deregulation. Competition brought service rates down, a trend that is likely to continue. Advances in optical networking made bandwidth abundant and a commodity. Hence, it is becoming more and more difficult for service providers to compete on price.

Service providers must embrace convergence since they know that service differentiation will be the only weapon against the threat of competition. The many ways to differentiate within a converged infrastructure are simply not possible with today's separate vertical networks. The new converged network will span multiple domains (wireline and wireless, national and global, public and private). Perhaps for the first time service providers will be able to customize telecom services for their clients at a price they are willing to pay.

Enterprise and Corporations

The business case for enterprise adoption can be seen easily from the growth of Voice over IP (VoIP) technology. In a Business Week survey (Baker 2000) on the outlook for VoIP, which accounted for less than one percent (1%) of global telecommunications ("Telecom") traffic in 1999, was expected to increase to seventeen percent (17%) by 2003, and more than thirty percent (30%) by 2006. Unfortunately, later data than 1999 was not found.

Within the enterprise, although convergence was talked about previously in the form of Computer Telephony Integration (CTI), VoIP provides far richer set of services in a much more cost economic way. Some of the value proposition and benefits of VoIP include:

- *Lower cost* – Less expensive call rates by bypassing the PSTN. Because an IP phone call takes place across a data network, long distance calls becomes local. This change can lower telecommunications expenses significantly, particularly when placing calls or sending facsimiles internationally.

- *Maximize network bandwidth utilization* –Again, a significant cost reduction. A typical two-way IP telephony call occupies 25% or less of the bandwidth of a traditional voice call. And, because the voice is delivered as IP data packets, it can travel over the same data channels as any other data. The result? More network traffic over fewer leased lines.

- *Convergent and enhanced applications* – From opening up new business opportunities to supporting interactive Web pages, IP telephony provides value-added applications to businesses. For example, a telemarketing company based in New York City can install IP telephony gateways in its headquarters and its Houston office. By calling through the IP telephony gateway to access Houston prospects, the company may be able to service a new market without incurring traditional long-distance charges for each call. Or, a company may promote its products on the Internet by placing a "Talk-to-Agent" web application that will help businesses turn web browsers into buyers with the click of a web icon. Speech recognition technology will enable new convergent applications especially in wireless services.

- *Tie-line replacement* – Businesses today maintain expensive PBX and tie lines that interconnect branch offices with headquarters. By completely replacing these less flexible items, corporations migrating over to VoIP will save significantly.

Today, most enterprises maintain two separate infrastructures. For voice, they use PBX and leased circuits that connect to the PSTN. For data communication, they use local area networks and routers. However, with convergence, early adopters are connecting their PBX to the LAN via enterprise IP telephony gateways. They are carrying some of the voice traffic over IP data networks. Over time, the PBX can be completely decommissioned and IP phones and desktop video software can communicate directly over IP networks.

Business Users

For the employee within the corporation, gaining access to any media information from anywhere at the right time seems to be the Mantra. A converged network can reduce the number of devices that people usually carry these days to access information. Instead of a cell phone, a PDA, and a laptop, one single access device would be sufficient for accessing information and making fast decisions. IP provides the universal connectivity. By working with the service provider, each division within a company or even each employee would be able to customize the information that he or she wants. Receiving any information in any format using any access device would prepare the sales and marketing people to respond better to customers than what they can do now.

Challenges – Technical and Business

To enable real-time voice and video communications over IP networks, two critical technical challenges must be met:

1. *Quality-of-Service (QoS) issue*: Before we explain what types of QoS architectures need to be in place, it is useful to consider how quality is measured for voice communications. Over the years, AT&T and Bell researchers developed a metric known as Mean Opinion Score (MOS), which is a rating on a scale of 1 to 5. A MOS score of 4.0 signifies *toll* quality calls. ITSPs have been measuring Voice over IP calls over regular Internet and found MOS scores are about 3.73. So, quality is not that bad.

   Today's Internet is based on a "best-effort" service design which does not guarantee delivery. Moreover, at times of congestion, applications may suffer burst loss, which is detrimental to playback of voice and video. Hence QoS mechanisms are needed. Of the several approaches to QoS(Wood and Chaterjee, 2002), Integrated Services (IntServe) and Differentiated Services (DiffServ) are two of the popular mechanisms. IntServ uses the signaling approach of phone networks in which you reserve resources before you transmit. The reservation today is done by a protocol called RSVP. In the Diffserv approach, reservation is not used. The packets are tagged by appropriate priority and the network is engineered to provide a per hop behavior (PHB) using sophisticated queuing and buffer management techniques. Diffserv seems to be the more scalable approach. Quality is not just a network issue. Even within the application, one can improve encoding mechanisms as well as network protocol stack processing time. Better voice and video codecs can provide much improved quality and performance.

2. *Service creation challenges*: If IP Telephony simply replicates what our current phone systems can do, it will not find wide acceptance among users. However, it needs to do more and can. IP telephony can provide a whole new experience in

communication by integrating web, voice and video into a seamless whole. Proponents of this technology believe that the entire communication experience can change. This can be done through "service creation" possibility. One reason the web was a success is because end-users could easily create innovative applications and make that available to the others. Today's phone network is a closed system. But IP telephony can lead to new services that are created by end-users by simply integrating the web with telephony and video. This opens up huge potential business models for providers.

Given the current situation of the Telecom industry, which is in the doldrums, effective strategy and business models are needed. Convergence and differentiated services may be mechanisms for service providers and carriers to distinguish themselves from the competition. But the catch is to design effective pricing schemes. Usage based pricing or per-packet pricing is not popular with customers. The flat-fee per month with unlimited usage is popular. However, for effective engineering of the network, this scheme is not favorable. The optimal pricing strategy for voice and video over IP is still an open research problem.

The "Grid" Architecture and Issues

The National Science Foundation Partnerships for Advanced Computational Infrastructure (PACI) program is helping create a national "Grid" architecture (Stevens 1997). The National Technology Grid is a name that is derived from the notion of the electrical power grid that transformed the U.S., and indeed the world, during the past century. A similar transformation is envisioned with the Web and the Internet as they diffuse into our everyday lives – at home, school and the workplace. The Grid could evolve into a national-scale metacomputer for everyone. However, the first challenge is to know how to integrate the variety of networking, distributed computing, scheduling, security and organizational aspects of the Grid.

Such a high-performance platform to support high-end science and engineering requires bandwidth. This bandwidth is being put in place by Federal agency networks through very high-speed backbones, such as NSF's VBNS, the Department of Defense's Defense Research and Education Network (DREN), and the Department of Energy's ESNet; grass-roots programs, such as Internet2 (www.internet2.edu) and the President's Next-Generation Internet (www.ngi.gov). However, software tools to provide middleware functions need to be developed, and a set of applications utilizing the broadband networks in a variety of disciplines needs to be demonstrated.

The I-WAY experiment that was performed in 1995 (Stevens, 1997) brought together a large number of scientists whose research needed multiple supercomputers, visualization devices, and databases inter-connected via high-speed networks (ATM at 155 Mb/s). Multiple immersive virtual reality systems were tested across the continent. The main issues that came out of that experiment were:

- Besides high-speed, the network should be robust, stable and persistent.
- A critical mass of users is required to enable true grid environment. A large number of people, devices and data must be in place to justify such a platform
- End-users should be involved with computer scientists in development of experimental middleware
- Full-scale deadline-driven demonstrations are essential for motivating integration and deployment of new technologies.

Considerable progress has been made in the construction of such a national infrastructure since the I-WAY experiment. However, the term "Grid" has been inflated to mean almost anything from advanced networking to artificial intelligence. What is of interest is to understand if there is a real Grid problem and what are the issues.

The real and specific problem that underlies the Grid concept is coordinated resource sharing and problem solving in dynamic multi-institutional virtual organizations (Foster 2001). The sharing is not primarily file exchange but rather direct access to computers, software, data, and other resources. This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly just what is being shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form what we call a virtual organization (VO).

The following are examples of VO: storage service providers, consultants engaged by a car manufacturer to perform scenario evaluation; members of an industrial consortium building a new aircraft; a crisis management team and the databases and simulations systems that they use to plan a response to an emergency situation. The basic set of requirements for such VOs include:

- Highly flexible sharing relationships ranging from client-server to peer-to-peer
- Precise level of control (fine-grained and multi-stakeholder access)
- Delegation of local and global policies
- Support diverse usage modes, from single user to multi-user
- Service scheduling at a reasonable cost
- Performance related issues.

Today's distributed computing architectures do not support such requirements. The Grid project is currently exploring these and many other fascinating projects (http://www.globus.org).

Summary

Internet enabled rapid convergence of voice, video, and data applications resulting in widespread impact on the telecommunications industry. In this tutorial we explain the infrastructure-related issues of such a convergence phenomenon. We also discuss the Grid problem and associated set of issues. The Grid has the potential to transform the Web and the Internet into a national-scale metacomputer that everyone can use as a valuable resource. Several challenges remain towards the 21st century infrastructure.

List of Acronyms

ATM	Asynchronous Transfer Mode
CATV	Cable TV
CLECs	Competitive Local Exchange Carriers
CTI –	Computer Telephony Integration
Diffserv	Differentiated Services
H.323	Signaling standard from International Telecommunications Union
IETF	Internet Engineering Task Force
Intserv	Integrated Services
ITSPs	Internet Telephony Service Providers
PBX –	Private Branch Exchange
PDA	Personal Digital Assistant
PHB	Per Hop Behavior
POP	Point of Presence
PSTN	Public Switched Telephone Network
QOS	Quality of Service
SIP	Session Initiation Protocol
SS7	Signaling System No 7
TCP/IP	Transmission Control Protocol/Internet Protocol
VBNS	Very High Speed Backbone Network Service
VO	Virtual Organization
VoIP	Voice over Internet Protocol

References

Baker, S. "The Talking Internet", *Business Week*, May 2000, pp. 174-192.

Bartholomew, M. F. *Successful Business Strategies Using Telecommunications Services,* Boston, MA: Artech House, 1997.

Clark, David D. "Implications of Local Loop Technology for Future Industry Structure", *1998 Telecommunications Policy Research Conference*, Alexandria, VA, 3-5 October 1998.

De Serres, Y., and L. Hegarty, "Value-Added Services in the Converged Network", *IEEE Communications*, (39:9), September 2001, pp. 146-154.

Dunstan, S. "Converged Voice and Data Services in the new Network", *Telecom J. of Australia*, (51: 2), 2001, pp. 17-31.

Edell, R. , and P. Varaiya, "Providing Internet Access: What We Learn from INDEX", *IEEE Network* (13: 5), September/October 1999, pp. 18-22.

Foster, I., C. Kesselman, and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", Intl., J. Supercomputer Applications, 15(3), pp. 200-222, Fall 2001.

Keil, T., E. Eloranta, J. Holmstrom, E. Jarvenpaa, M. Takala, E. Autio, and D. Hawk, "Information and Communication Technology Driven Business Transformation – A Call for Research", *Computers in Industry*, 44, 2001, 263-282.

Schulzrinnem, H., and J. Rosenberg, "The IETF Internet Telephony Architecture and Protocols", *IEEE Network*, (13:3), May/June 1999, pp. 18-23.

Stevens, R., P. Woodward, T. DeFanti, and C. Catlett, "From the I-WAY to the National Technology Grid", Communications of the ACM, 40, 11, (Nov 1997).

Wood, S., and Chatterjee, S., "Network Quality-of-Service for the Enterprise: A Broad Overview", *Information Systems Frontiers* (4:1), 2002.