NCS TIB 00-8

NATIONAL COMMUNICATIONS SYSTEM

TECHNICAL INFORMATION BULLETIN 00-8

The Convergence of Signaling System 7 and Voice-over-IP

September 2000

OFFICE OF THE MANAGER
NATIONAL COMMUNICATIONS SYSTEM
701 SOUTH COURT HOUSE ROAD
ARLINGTON, VA 22204-2198

This document was prepared under contract to the

Office of the Manager
National Communications System

Contract No. F41621-98-D5625
Work Order No. 0006

Prepared by:

Gary L. Ragsdale, Gerard P. Lynch, and Michael W. Raschke
Southwest Research Institute
P.O. Drawer 28510
San Antonio, TX 78228
210-522-3743 (VOICE)
210-522-5499 (FAX)

SwRI Project No. 10.03607

EXECUTIVE SUMMARY

Voice-over-IP describes the process of transporting voice audio conversations across Internet Protocol (IP) networks. Voice-over-IP must achieve low latency and low delay jitter to be accepted as a viable, toll quality service. The long-term incentives for Voice-over-IP are to be found in value-added, multimedia services; consolidation to a single, easily managed network; and reduced maintenance cost through the use of modern communications equipment.

Today, the Signaling System No. 7 (SS7) network is the backbone signaling method for the national telephone system. SS7 is the signaling network that manages and controls the separate voice network. New Voice-over-IP systems employ SIP/SDP or H.323 for signaling and Real-time Transport Protocol (RTP) for voice transport. Both signaling and voice occur on the same IP network.

Just like web servers and mail servers, Voice-over-IP servers stand out as targets for malicious individuals and groups. Voice-over-IP servers may soon suffer the same fate as e-commerce web sites. They may be targeted to obtain credit card or calling card numbers to commit extortion or fraud.

AT&T defined the toll quality service standards within the context of an all-analog network. Modern packet networks characterize performance in terms of packet loss, available bandwidth, and system latency. Fortunately for Voice-over-IP networks, voice packets have a high tolerance for truly random packet loss. Voice-over-IP networks must compete with data for network resources and contend with networks designed to provide the best data performance per dollar.

Fast path restoration within the high-speed core network is essential to reliable Voice-over-IP. Switching transitions of a few milliseconds will be noticed by the receiver codec and by the listener. Coordinated restoration within and across network domains is important when natural disasters, terrorism, or other NS/EP events disable many network components simultaneously.

New domain interconnections, such as Media Gateway Control (MEGACO), joining the Internet and public switched telephone network (PSTN) provide windows through which a network attack can originate. Therefore, the existing and new domain operators should take cooperative measures to certify that their domain interconnections can be trusted to protect the SS7 infrastructure. Unlike SS7, MEGACO signaling control occurs on the same network as all other traffic. This opens up MEGACO elements to common network security vulnerabilities. Intrusion in the Voice-over-IP network could be used to access end-user billing information contained in the PSTN, providing personal information such as home address or the credit card number used for automatic bill payment. These intrusions could also create fraudulent billing statements or hide toll conversations from the billing system. Voice carriers already have security measures in place to protect this information from other voice carriers; however, the strength of their security measures against the resourceful internet hacking community has not yet been determined. As more gateways are deployed, more avenues are available to intruders for finding chinks in the armor of the PSTN.

TABLE OF CONTENTS

Executive Summary ..... **i**

Table of Contents ..... **ii**

List Of Figures ..... **v**

List Of Tables ..... **vi**

List Of Acronyms ..... **vii**

1.0 Converging Networks **1**

1.1 100 Years of Telephony 1
1.2 The IP Revolution 1
1.3 Circuit-Switched vs. Packet-Switched 1
1.4 Voice-over-IP 3
1.5 The Move to IP Telephony 3
1.5.1 New Service Predictions 3
1.5.2 Market Predictions 3

2.0 Voice Carrier Networks **5**

2.1 Carrier Network Protocols 5
2.2 Bandwidth Efficiency 6
2.3 Voice-Data Integration 6

3.0 Data Carrier Networks **7**

3.1 Bypassing Incumbent Carriers 8
3.2 Voice Carriers Want to Play 8
3.3 Voice over DSL 9
3.4 The Cheap Long Distance Myth 9

4.0 The Signaling System No. 7 Network **11**

4.1 Architecture 11
4.2 Signaling 12
4.2.1 Service Switching Point 13
4.2.2 Signal Transfer Point 14
4.2.3 Service Control Point 16
4.3 Transport 17
4.3.1 Signaling Transport 17
4.3.2 Voice Transport 19
4.4 Value-Added Services 19
4.5 SS7 Protocol Stack 20
4.5.1 Level 1 - Physical Level (MTP 1) 21
4.5.2 Level 2 - Data Link Level (MTP 2) 21

4.5.3 Level 3 - Network Level (MTP 3) 22
4.5.4 Level 4 – ISDN User Parts (ISUP) and Application Parts (TCAP) 22
4.6 Security Features and Limitations 22
4.7 ITU-T SS7 Similarities and Differences to ANSI SS7 24

5.0 The Voice-over-IP Architecture 26

5.1 Signaling 26
5.1.1 IETF Voice-over-IP 26
5.1.2 ITU-T Voice-over-IP 29
5.2 Transport 32
5.2.1 RTP 32
5.2.2 RTCP 32
5.3 Value-Added Services 32
5.4 Similarities to SS7 33
5.5 Security 34
5.5.1 Security Risks 34
5.5.2 Security Solutions 35

6.0 Bridging the Gap 38

6.1 Benefits of Union 38
6.2 PSTN Gateways 38
6.2.1 IPDC and SGCP 38
6.2.2 MGCP 39
6.2.3 MEGACO 39

7.0 Reaching Toll Quality in a Converged Network 41

7.1 Latency and Jitter in the PSTN 42
7.2 Latency and Jitter in Voice-over-IP Networks 42
7.3 WANs: Panaceas and Plagues 43
7.3.1 Role of Traffic Variance 44
7.3.2 Impact of Link Failures 46
7.4 QoS in the WAN via Prioritization 47
7.4.1 IEEE 802.1p 48
7.4.2 Integrated Services (INT-SERV) 48
7.4.3 Differentiated Services (DIFF-SERV) 51
7.4.4 COPS 51
7.4.5 MPLS 52
7.4.6 ATM 53
7.4.7 Real-World Latency Mitigation 53
7.5 Packet Loss Across Multiple Network Domains 53

8.0 Reaching Reliability in a Converged Network 56

8.1 Multiple-Domain Network Fault Tolerance 56
8.2 Multiple-Domain Network Restoration 56

9.0 Achieving Security in a Converged Network 59
9.1 Changes in Signaling System 7 Security 59
9.1.1 IP Network Gateways 61
10.0 Conclusions 63
Appendix A SS7 Standards A-1
Appendix B List of References B-1

LIST OF FIGURES

Figure 1: Circuit-Switched Network vs. Packet-Switched Network 2
Figure 2: ATM Cell-Switched Network 5
Figure 3: SS7 Signaling Point Topology 12
Figure 4: Direct-Dial Voice Connection..... 13
Figure 5: Routed Voice Connection 14
Figure 6: STPs in the Worldwide SS7 Network 16
Figure 7: SS7 Interconnection Topology..... 18
Figure 8: SS7 Protocol Stack and the OSI Reference Model[12]..... 21
Figure 9: SS7 Service Management System Security..... 23
Figure 10: SIP Architecture[7]..... 28
Figure 11: H.323 Architecture 30
Figure 12: H.323 Signaling 31
Figure 13: H.323 to SS7 Call Signaling[20]..... 33
Figure 14: SIP to SS7 Call Signaling[19] 34
Figure 15: MEGACO Connection to the PSTN 40
Figure 16: Typical Voice-over-IP Latency..... 43
Figure 17: Multiple Domain Network Infrastructure..... 45
Figure 18: Core and Edge Networks in Multiple Device Environment 46
Figure 19: INT-SERV Architecture – RSVP Packet Flow..... 50
Figure 20: COPS System Architecture 52
Figure 21: Current Network Pyramid 57
Figure 22: Vulnerable ATM LANE Network..... 61

LIST OF TABLES

Table 1: Multi-Domain Call 54

LIST OF ACRONYMS

ACF	Admission Confirm
ACM	Address Complete Message
ANM	Answer Message
ANSI	American National Standards Institute
ARQ	Admission Request
ASCII	American Standard Code for Information Interchange
ATM	Asynchronous Transfer Mode
AT&T	American Telephone and Telegraph
CAP	Competitive Access Provider
CBR	Constant Bit Rate
CCIS6	Common Channel Interoffice Signaling System No. 6
CCITT	Consultative Committee on International Telegraphy and Telephony
CLEC	Competitive local exchange carriers
CMSDB	Call Management Services Database
CNAM	Calling Name (database)
COPS	Common Open Policy Service
DIFF-SERV	Differentiated Services
DNS	Domain Name Service
DoS	Denial-of-Service
DPC	Destination Point Code
DSL	Digital Subscriber Line
DWDM	Dense Wave Division Multiplexing
FCC	Federal Communications Commission
GCF	Gatekeeper Confirm
GK	Gatekeeper
GRQ	Gatekeeper Request
H.323	ITU-T Standard for Packet-Based Multimedia Communications
HTTP	Hyper Text Transfer Protocol
IAM	Initial Address Message
IEEE	Institute of Electrical & Electronics Engineers
IETF	Internet Engineering Task Force
IN	Intelligent Network
INT-SERV	Integrated Services
IP	Internet Protocol
IPDC	Internet Protocol Device Control
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISDN	Integrated Services Digital Network
ISP	Internet Service Provider
ISUP	ISDN User Part
ITU-T	International Telecommunications Union – Telecommunication Standardization Sector

IXC	Inter Exchange Carrier
LAN	Local Area Network
LIDB	Line Information Database
MC	Multipoint Controller
MEGACO	Media Gateway Control
MG	Media Gateway
MGC	Media Gateway Controller
MGCP	Media Gateway Control Protocol
MIN	Mobile Identification Number
MP	Multipoint Processor
MPLS	Multiprotocol Label Switching
MTP	Message Transfer Part
MTU	Maximum Transfer Unit
nrt-VBR	Non Real-Time Variable Bit Rate
NS/EP	National Security/Emergency Preparedness
OC-48	Optical Carrier 48
OPC	Origination Point Code
OSI	Open Systems Interconnection
PBX	Private Branch Exchange
PC	Personal Computer
PDA	Personal Data Assistant
PDP	Policy Decision Point
PEP	Policy Enforcement Point
POTS	Plain Old Telephone Service
PSTN	Public Switched Telephone Network
PTT	Public Telephone and Telegraph
PUC	Public Utility Commission
QoS	Quality of Service
RAS	Registration, Admission, and Status
RBOC	Regional Bell Operating Company
RCF	Registration Confirm
RFC	Request For Comments
RRQ	Registration Request
Rspec	Reserve Spec
RSVP	Resource Reservation Protocol
RTCP	Real-Time Transport Control Protocol
RTP	Real-Time Transport Protocol
rt-VBR	Real-Time Variable Bit Rate
SA	Source Address
SCCP	Signaling Connection Control Part
SCP	Service Control Point
SCTP	Simple Common Transport Protocol
SDP	Session Description Protocol
SG	Signaling Gateway
SGCP	Signal Gateway Control Protocol
SIP	Session Initiation Protocol

SLA	Service Level Agreement
SMS	Service Management System
SMTP	Simple Mail Transport Protocol
SONET	Synchronous Optical Network
SS7	Signaling System No. 7
SSL	Secure Sockets Layer
SSP	Service Switching Point
SSRC	Synchronization Source
STP	Signal Transfer Point
TCAP	Transaction Capabilities Application Part
TCP	Transport Control Protocol
TDM	Time Division Multiplexing
TOS	Type-of-Service
TPI	Tag Protocol Identifier
Tspec	Traffic Spec
UAC	User Agent Client
UAS	User Agent Server
UDP	User Datagram Protocol
VLAN	Virtual Local Area Network
VPN	Virtual Private Networking
WAN	Wide Area Network
WRED	Weighted Random Early Detection
WSP	Wireless Service Provider

1.0 CONVERGING NETWORKS

In the technology world, the phrase "more for less" is a universal truism. Advances in state-of-the-art technology continuously enable users to utilize fewer resources to accomplish the same tasks. Major drivers in this trend are cost and compatibility. Using the same equipment for multiple purposes reduces operational costs. Systems that are compatible can be readily intermixed in multipurpose environments. These factors have been fueling the fire of digital convergence that is currently knocking at the door of telephony.

1.1 100 Years of Telephony

There are over one hundred years of industrial experience in supplying reliable voice delivery over circuit-switched networks. A huge, multi-carrier network spans the globe for the purpose of delivering reliable, toll quality voice communications. Until recently, this circuit-switched infrastructure dominated all other networks, and still represents the preponderance of the revenue derived from communication services. It is the standard before which all other voice services must withstand scrutiny.

1.2 The IP Revolution

Since its inception, growth of the IP networks has proceeded like a landslide. The Internet is the most prevalent IP network. Most private IP networks connect to the Internet to enhance their reach. The Internet was originally developed as a means for researchers to exchange information and ideas.

The Internet was and is about communications. Researchers quickly took advantage of the Internet as a convenient and affordable means of exchanging electronic mail. Email quickly stood out as the primary motivation for some people to use the Internet. With the development of the World Wide Web, more users flocked to the Internet to take advantage of graphics in addition to text. Many view the Internet as the network of choice for interactive data communications. There are many initiatives to add streaming video and audio to the Internet as a way of enhancing its interactive services. The future is clear in that the Internet will carry interactive voice and video communications as a natural extension to its success in interactive data communications.

Now, with the significant financial investment in data network infrastructures, users consider IP networks to be the only networks they should need. This financial incentive, coupled with the simplicity of maintaining only one network, has created a desire to transmit voice over IP networks.

1.3 Circuit-Switched vs. Packet-Switched

Traditional telephone networks are based on circuit-switched technology. When a user makes a call, switches align circuit paths to provide a continuous circuit between the caller and the called party. The circuit exists for the duration of the call, regardless of activity on the circuit. Call participants are guaranteed a consistent level of performance

once the call has been established. For the duration of the call, the network resources supporting the call are unavailable to other users of the circuit-switched network. The top portion of Figure 1 shows a typical circuit-switched network. The green lines represent the physical circuit from one endpoint to the other.

The Internet is an example of a packet-switched network. Packet-switched networks deliver individual packets across a network from source to destination. In a route-diverse network, packets traveling from source to destination may traverse different paths. Some paths may offer better performance than other paths, leading to variability in the latency (quality) of the connections. Since packets traveling from source to destination can take any available path, network resources are not set aside for exclusive use by a single connection. Packet networks are also "best-effort" networks. No guarantees are made on the reliable delivery of a packet. The bottom portion of Figure 1 shows a typical packet-switched network. The maroon squares indicate packets traveling between the two computers. As can be seen from the figure, the maroon packets share the network with other packets and do not always take the same path.

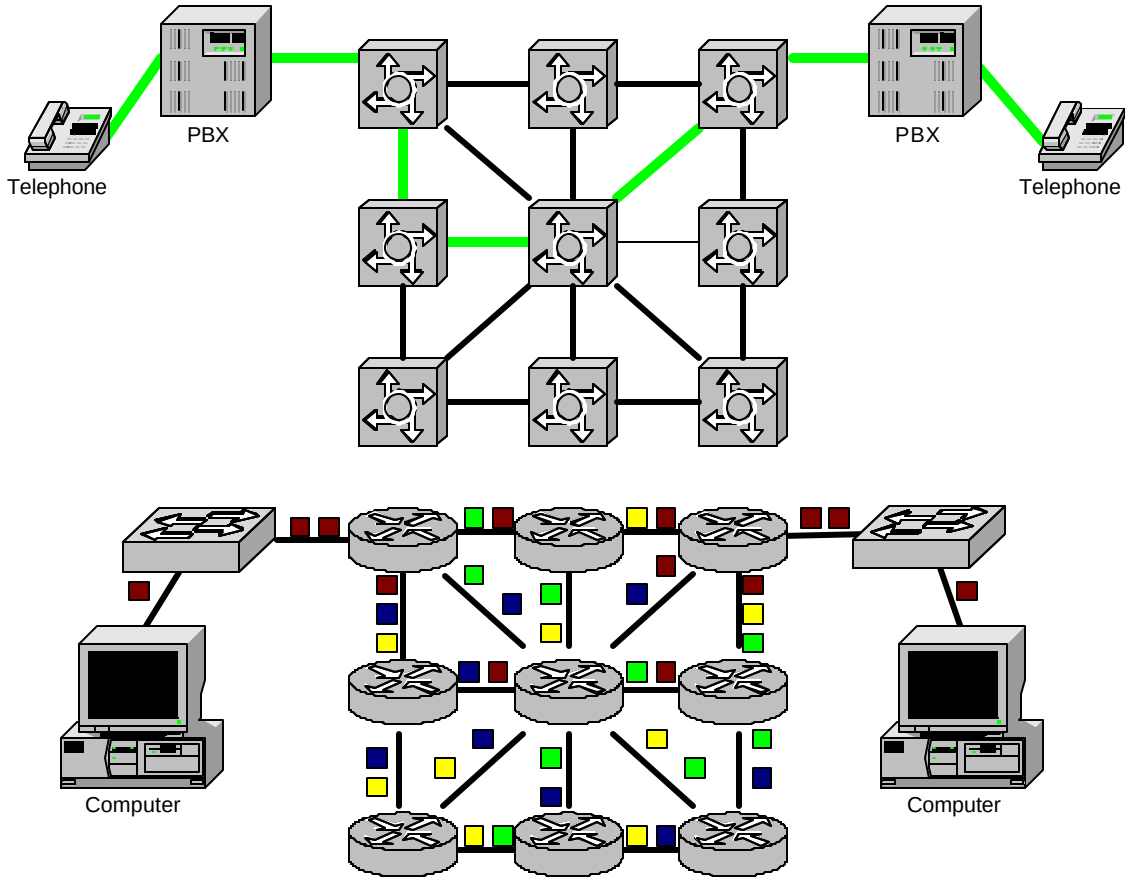

Figure 1: Circuit-Switched Network vs. Packet-Switched Network

1.4 Voice-over-IP

Voice-over-IP describes the process of transporting voice conversations across IP networks. Voice-over-IP must achieve low latency and low delay jitter to be accepted as a viable, toll quality service. Much less experience exists in providing reliable voice over IP networks as compared to providing reliable voice over circuit-switched networks.[1] Low cost telephone calls alone will not make Voice-over-IP a mainstream service. Rather, toll quality circuit-switched voice communications set the standard upon which Voice-over-IP will be judged, accepted, or rejected in the marketplace.

1.5 The Move to IP Telephony

Philosophically, Voice-over-IP is not about assigning IP addresses to telephones and turning personal computers into electronic telephones. It is about competition, market trends, and economy.

1.5.1 New Service Predictions

Voice-over-IP offers several advantages compared to the traditional public-switched telephone network (PSTN). Simplicity of services can be achieved through the integration of voice, email, and fax into a single coordinated service.[3] Some new services likely to arise from Voice-over-IP include web-based call centers, hi-fidelity audio exchange, unified messaging from multiple sources to a single inbox, virtual second line, voice cost reduction especially for on-hold conditions, real-time billing, remote teleworking via desktop conferencing, and enhanced teleconferencing with white board plus application sharing.

1.5.2 Market Predictions

The research firm IDC predicts that IP telephony will be the fastest growing network service of the 2000 decade.[4] There are market forces driving such predictions. The preceding sections describe forces leading to new value-added services facilitated by IP telephony. Financial incentives also lead to predictions of a major uptake in IP telephony within the commercial networks.

There is a huge investment in the PSTN network. The PSTN switching is hard to manage, difficult to maintain, and costly to operate. For example, a modification to central office switching software is a monumental task that is fraught with risk and encumbered by long lead times. The complexity of modifying telephone switching software drives the movement of value-added services such as caller ID, call waiting, and calling cards into the SS7 network where modification is easier and lower risk.

Aspiring PSTN service providers abandoned modifications of their telephone switches in favor of service control point (SCP) systems that create value-added services within the SS7 network. SCP processors direct basic switching functions resident within telephone switches, thereby orchestrating new service functions without modifying the telephone switching systems.

At the same time, developments in low-cost fiber optic and dense wave division multiplexers are changing voice network investment emphasis away from transmission

and onto switching. The total investment necessary to build and operate a fiber optic transmission system has fallen below that of building the voice switching systems that direct traffic onto the transmission systems. Consequently, service providers are giving greater consideration to the cost and lack of flexibility of telephone switches as it affects their bottom line.

Service providers see packet networks as providing higher value than the circuit-switched equivalent. First, the cost of creating and deploying value-added services within a packet network are lower. Second, packet networks support a wider variety of service opportunities than an SS7-equipped circuit-switched network. Finally, the investment cost and operational complexity of packet networks may be much less than currently experienced within the circuit-switched networks. Taken together, the cost, risk, and lack of flexibility inherent within current circuit-switched networks is driving service providers to consider packet networks as a new service platform.

2.0 VOICE CARRIER NETWORKS

Market factors continue to drive carrier networks from circuit-switched to packet-switched solutions. Cost analysis shows a steep increase in packet-switch device performance per cost compared with circuit-switch device performance per cost.[2] Telephone carriers recognize these trends and continue to increase their deployment of packet-switching products to maximize their switching ability per dollar.

2.1 Carrier Network Protocols

Asynchronous transfer mode (ATM) switching systems are making gains in the wide-area networks (WANs) despite limited uptake in the local area networks (LANs).[6] Service providers are relying on ATM within their core networks to provide the high speed backbone for an array of services involving combinations of voice, data, and eventually video.

The ATM protocol defines a cell-switched network that combines benefits from both circuit-switched and packet-switched networks. ATM uses virtual circuits to map a consistent path through a network without requiring that path to be exclusive to a specific connection. ATM can multiplex cells from different connections down any path, much the same way packets from many different connections can travel down the same path in a packet-switched network. Figure 2 shows an ATM cell-switched network. In this example, the maroon cells travel the same path through the network, but still share the network segments with other cells. ATM can also be tuned to efficiently carry packets across its network inside cells.

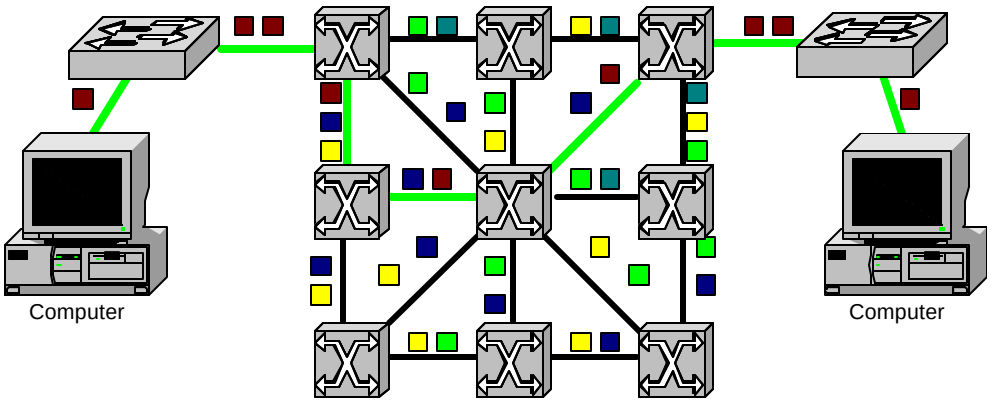

Figure 2: ATM Cell-Switched Network

Service providers adapt their ATM core networks to particular service applications along the network edge. These edge networks supply services such as Internet, circuit-switched voice, cellular mobile, and paging. Voice-over-IP is another edge network application that can be adapted to highly efficient ATM core networks.

2.2 Bandwidth Efficiency

Quality of Service (QoS) concerns form the primary motivator for use of ATM on the carrier networks. Carrier networks support users who pay for quality based on a service level agreement (SLA). QoS features in ATM enable carriers to handle high capacity circuits efficiently with the same performance guarantees available in the legacy circuit-switched networks.

2.3 Voice-Data Integration

Consumer demand for data network access continues to grow. Recently, the volume of data traffic carried across WANs has surpassed the volume of voice traffic carried across the PSTN. Carriers must respond to these demands by increasing their offerings for data services. Management and maintenance of two separate networks can be time consuming and costly. Sending voice over data networks offers carriers the best of both worlds. They can continue to offer their historic business services of voice while building their data networks that keep them on track with changing consumer demands.

ATM's characteristics serve as an enabling technology for organizations offering fee-based services. ATM's connection-oriented protocol lends itself to usage-based billing, which is a common way to market telephony. Users can also be readily billed for specific service levels. In an analogous manner to airline passengers purchasing first class tickets for better service, some users may purchase performance guarantees for their traffic. Internet retailers may pay extra to guarantee that their Voice-over-IP presale calls receive the minimum bandwidth that is necessary to ensure a quality audio conversation with a prospective buyer.

Many challenges and opportunities may face preparedness organizations as the move from circuit-switched telephony progresses to IP telephony. Emergency services resident within the circuit-switched network may cease to operate when a portion of the telephone connection traverses a Voice-over-IP network. These same services may not exist within the Voice-over-IP network. Therefore, the transition to Voice-over-IP networks may disrupt old arrangements for emergency telephone services during NS/EP events.

At the same time, Voice-over-IP networks rely heavily on prioritization of traffic as a means of ensuring voice quality. Additional consideration to traffic prioritization within Voice-over-IP networks with preference to NS/EP traffic could add valuable services to the emergency service suite to be used in times of national crisis or natural disaster. For example, if Voice-over-IP networks give preference to NS/EP voice traffic, then Voice-over-IP networks would automatically allocate available network resources to emergency preparedness agencies and away from less essential users.

3.0 DATA CARRIER NETWORKS

Paradigms created by the Internet differ radically from those defined by the PSTN. Internet and PSTN service providers differ in the way each generates revenue, how they define quality of service, and how they sell their services. Consequently, convergence of the commercial PSTN and data networks represents a major paradigm shift for PSTN and Internet service providers.

Consider for a moment the issue of revenue generation. Recent changes in Federal Communications Commission (FCC) regulations have created opportunities for businesses to create access networks that bridge the consumer into carrier networks. Some access providers seek to offer low cost services by bypassing traditional carrier networks and capitalizing on the Internet.

The Internet is an access network funded in part by access fees paid to access providers, also called Internet Service Providers (ISPs), to gain entry to the network. Fees are associated with access connection rates, not use of the network. Much more Internet revenue comes through services and advertising dispensed over the Internet. The vague, almost mystical Internet revenue relationships stand in stark contrast to the highly structured PSTN where most revenues come directly from telephone subscribers.

PSTN networks derive their revenues directly from customers for network usage. Yellow page advertising and flyers enclosed with telephone bills represent relatively small revenue sources for most PSTN service providers. Consequently, the customer-provider relationship for PSTN providers is very different from that of the ISP. Other factors make the PSTN paradigm substantially different from that of the data network provider.

Incumbent PSTN networks are heavily steeped in practices and regulations that run counter to innovation. Tariffs and settlements require lengthy reviews by public utility commissions (PUCs), the FCC, and even the courts. Often, political and special interest issues take precedent over new customer requirements. For example, PUC concerns regarding emergency telephone services called "life-line" services and low-cost residential services complicate the pricing and deployment of new services within the PSTN. For instance, California PUC concerns over inexpensive residential telephone services during the 1980s delayed the introduction of Pacific Bell data services for months and ultimately influenced data service pricing. The regulated, highly structured telephone marketplace stands in stark contrast to the free-market, innovative Internet marketplace.

Much of the innovation that we equate to the Internet comes from the ability of over 4,000 ISPs and "dot.com" value-added service providers to link their domains together into a cohesive network in a rapidly changing marketplace. Issues regarding the exchange of services and revenues are dynamic. Each new service creates the potential for new customer billing methods and a division of revenues along lines that are as new as the service itself.

The driving force behind Internet innovation has been the opportunity to provide new, value-added services. On-line retailing, news, weather, travel, chat rooms, email, Web

advertisements, and many more services represent new ways of interacting with customers. Even with all this rapid-fire innovation, there is still room for more.

For all of its prowess, the Internet lacks a full set of interactive multimedia services. The lack of interactive, toll quality voice is a key missing element in many services. Additions of toll quality voice communication to the Internet would fulfill the vision of Web-based "click-to-talk" customer service, virtual second line, Internet call waiting, remote teleworking, desktop conferencing, enhanced teleconferencing, and unified messaging. New services mean new revenues, new service providers, and more innovation. Innovation of this magnitude does not fit well within the PSTN paradigm.

3.1 Bypassing Incumbent Carriers

In the past, regulated carriers paid each other settlements for traffic exchanged at network interconnections according to well-defined settlement formulas. Structured settlement formulas do not provide for the new service pricing flexibility. This is especially true when the revenues for new services are derived indirectly through advertising, retail sales, or other third party relationships.

Innovation requires an open framework amenable to new pricing strategies, new communication methods, and new ways of dealing with suppliers. The open framework invites new services, many of which call for the integration of data with interactive voice. The next natural step will be the integration of interactive video. For now, Voice-over-IP promises to integrate interactive voice to an already interactive data network.

End users, competitive local exchange carriers (CLECs), and carriers stand to benefit from the integration of voice, data, and multimedia communication on an integrated IP network.[5] End users will see reduced voice cost largely due to the avoidance of per-minute carrier interconnection charges and more efficient use of network bandwidth, i.e., Voice-over-IP requires 6K-8Kbps as compared to the 64Kbps required by the PSTN.[5] New CLECs and ISPs will avoid the regulation-heavy PSTN and provide competitive voice services over the more versatile IP networks. CLECs and ISPs can offer new services not available in existing Regional Bell Operating Company (RBOC) networks such as hi-fidelity audio delivery, multicast conferencing, distance learning applications, and voice-based Web services such as voice recognition and Web-based call centers. Voice-over-IP therefore serves as an enabling technology to competitive business in the voice transmission market. The Internet paradigm provides a forum in which interactive multimedia can flourish without the encumbrances of the old PSTN paradigm.

3.2 Voice Carriers Want to Play

PSTN and wireless service providers (WSPs) find themselves limited to services that they can sell through a telephone handset. They realize that the dominance that once was the PSTN network is shifting to the Internet. With this shift, there will be a corresponding shift of new services and new revenues away from voice-only networks to multimedia data networks.[25]

PSTN networks provide at best 56Kbps of data transmission to new services. This is inadequate for multimedia applications such as desktop conferencing, enhanced call

centers, and consolidated message centers. PSTN and wireless service providers must update their infrastructure if they are to participate in new services having interactive voice content.

3.3 Voice over DSL

Local access providers, i.e., the RBOCs, CLECs, and the competitive access providers (CAPs), have an additional imperative of retaining control of the end subscriber. Access providers have the most direct relationship with the customer and want to preserve this highly valued relationship. Their strong desire to retain their access connection to their customers is driving Southwestern Bell and other access providers to update their local access infrastructure to DSL.

Initially, access providers marketed DSL services as a data-only access alternative to telephone and cable TV modems. The prospect that telephone and cable TV modems might capture their circuit-switched voice revenues is leading them to expand the role of DSL to voice as well.

New DSL products provide up to 14 simultaneous voice conversations and hundreds of kilobits of data transmission over a copper phone line.[22][23] In the central office, these DSL access lines connect with the PSTN for conventional voice services or an ATM-based data network for Voice-over-IP and other multimedia services. It is clear that DSL access with voice and data services gives the access provider a vehicle to market Voice-over-IP and other interactive multimedia services directly to the end customer.

3.4 The Cheap Long Distance Myth

Convergence of voice and data onto single packet networks does not hinge on cheaper long distance service. Today, PSTN services generate approximately 80% of the global communication revenues.[24] Analysts predict that the rapid uptake of the Internet and the growth in new services will reduce the PSTN market share by 12.5% in only four years. Analysts and providers agree that the prime motivator for Voice-over-IP and network convergence is new services.[2]

Deregulation of the long distance carriers initiated a free-for-all in long distance pricing that continues today. Interstate rates fell from $0.34 per minute in 1980 to recently advertised rates as low as $0.02 per minute as offered by Sprint. Deregulation and competition produced a 100-fold reduction in long distance rates as seen by consumers over a twenty-year time frame. This is an unprecedented bonus to consumers that places a typical long distance call in the same discretionary spending category as a roll of breath mints. Therefore, aspirations of cheaper long distance rates lack substance, are misleading, and, if fulfilled, will have a minor impact upon the consumer.

In truth, communication services providers need new value-added services and new revenue streams to maintain their vitality. ISPs need interactive voice and data as enabling technologies for new value-added services. PSTN providers need to grow beyond the telephone handset and sell their voice-related services through multimedia personal computers (PCs). PC makers want to replace every telephone, cell phone, and TV set with a multimedia PC or personal data assistant (PDA). Therefore, the

convergence of voice and data networks into a data-centric, Voice-over-IP network is the first step in a new multimedia services paradigm.

It may be a mistake to think of the Internet and the PC as a one-for-one replacement for the telephone. In the consumer's mind, toll quality voice added to a PC-based desktop conferencing system may be much more than a replacement for a telephone. Consolidated paging, voice mail, and email through a wireless PDA may be much more than a replacement for a cell phone. The consumer will probably view voice integrated with data as a completely new service. Plain old telephone service (POTS) may become a part of that basic services group shared by taxicabs and newspapers in which there is fundamental usefulness, basic features, and a lack of modern sophistication.

Consumers may find the concept of PSTN and Internet service providers vanish behind a wave of integrated voice and data services. The fact that their telephone calls use the same data network as their Web pages will be completely transparent to the consumer. The lack of consumer awareness may perpetuate a service distinction between Internet and plain old telephone service. Only the ISP and PSTN service providers will realize the convergence manifest in the consolidation of their infrastructures and the ability to diversify their revenue streams to new multimedia services. The old and the new will continue to exist within the consumer's mind. Long distance telephone services will continue to compete as a commodity service. Pricing for new Voice-over-IP services will have little regard for the idea of a cheaper long distance call.

4.0 THE SIGNALING SYSTEM NO. 7 NETWORK

In the 1960s, the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T), formerly the Consultative Committee on International Telegraphy and Telephony (CCITT), developed a digital signaling standard called the Common Channel Interoffice Signaling System No. 6 (CCIS6) for the Public Switched Telephone Network (PSTN). CCIS6 provided a mechanism for out-of-band signaling between remotely located telephone offices. Almost as soon as the deployment of the CCIS6 network began, the ITU-T initiated work on the Signaling System No. 7 (SS7) standard that is now prevalent throughout the world. CCIS6 was still being deployed in the US when SS7 was being introduced in 1983. In the US, the first uses for SS7 consisted of providing access to remote databases rather than call setup, management, and teardown. In international communities, including Europe, the opposite was true. There the concept of widely using remote databases is still rather new, but growing. Finally, it wasn't until 1996 that the widespread deployment of SS7 was accomplished. These deployment timeframes depict the PSTN network as a network that is reluctant to change. A converged voice and data network should prove to be a more flexible network.

Today, the SS7 network is the backbone signaling method for the circuit-switched PSTN and is paving the way for the "Intelligent Network" that includes a plethora of new value-added services. The most important reason for deploying the SS7 network throughout the world is to enable telephone companies to share subscriber information and perform efficient call signaling procedures.

4.1 Architecture

The architecture of the SS7 network contains many worldwide-interconnected nodes that communicate signaling messages for connecting telephone calls across the PSTN. The nodes are identified as signaling points, and the transport connections between them are identified as data links. The SS7 network in the international plane conforms to the ITU-T Q.7nn series of recommendations. The SS7 network in the national planes may conform to either the ITU-T recommendations or a customized version thereof. In the US the ANSI T1.nnn standards are used for the SS7 protocol. See Appendix A for a list of the individual standards.

Exchanges of SS7 messages between the international and national planes require the use of SS7 gateway converters to ensure that a given nation's SS7 messages conform to the ITU-T recommendations when traversing beyond the scope of the national plane. The key to the success of the SS7 network is the tremendous level of built-in redundancy. This redundancy provides the SS7 network with the option to route signaling messages along alternate paths if it encounters a failed node, a failed link, or network congestion.

SS7 network redundancy provides for a highly reliable PSTN network. High reliability and the ability to route signaling around failed network components are especially valuable during times of national crisis and natural disaster. It is during NS/EP events that portions of the PSTN and SS7 networks are likely to be damaged. SS7's design

lends itself to rapid restoration and graceful operation even during outages and crisis situations. Similar consideration should be given to Voice-over-IP service providers, especially when their Voice-over-IP networks exchange traffic with the PSTN network.

4.2 Signaling

The SS7 network nodes, or signaling points, originate signaling messages, receive signaling messages, and transfer signaling messages to and from other signaling points in the network. Three basic types of signaling points exist within an SS7 network: (1) Service Switching Points (SSPs), (2) Signal Transfer Points (STPs), and (3) Service Control Points (SCPs). The following sections briefly describe each type with respect to their functions within the SS7 network. Figure 3 illustrates a simple connection topology of the SS7 signaling points.

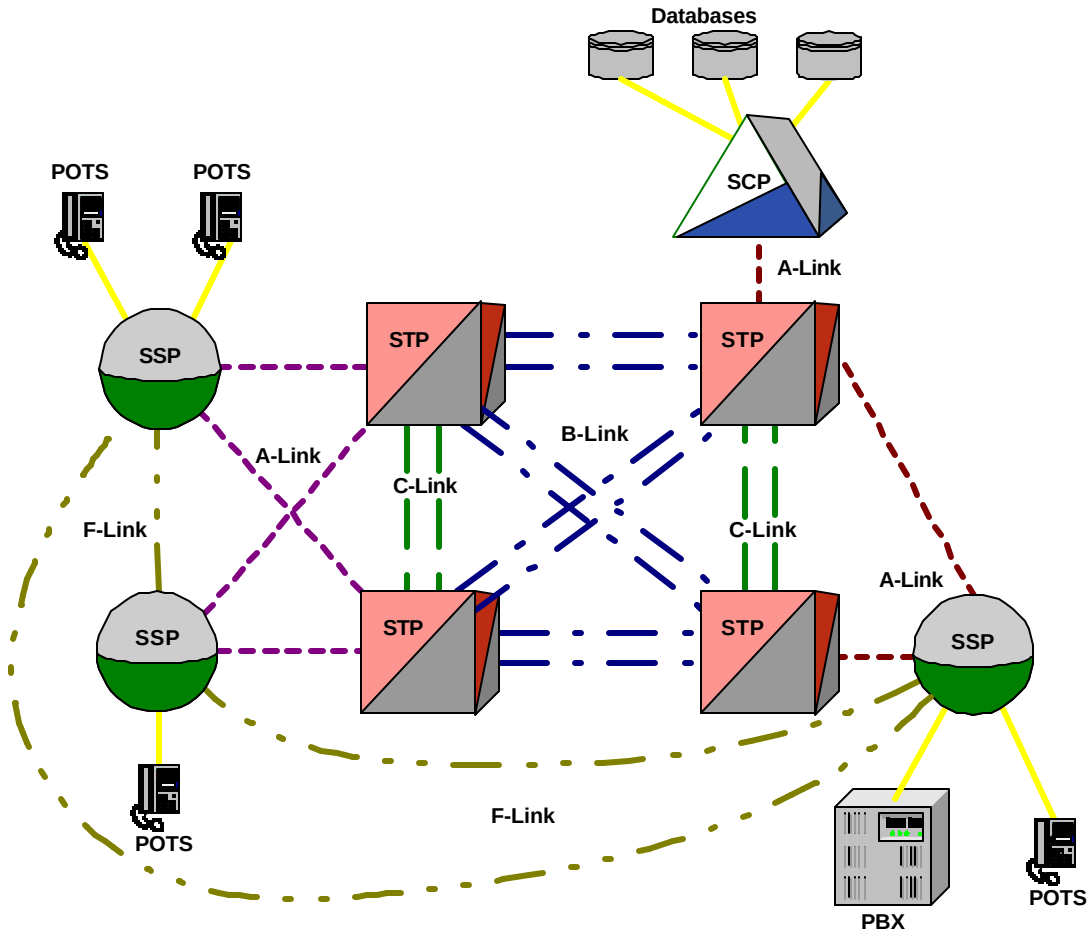

Figure 3: SS7 Signaling Point Topology

4.2.1 Service Switching Point

The Service Switching Point (SSP), usually a part of a PSTN end office or tandem switch, routes and connects calls frequently under the direction of a Service Control Point (SCP). (The SCP is described in Section 4.2.3.) As such, the SSPs are the source of and destination for SS7 signaling messages. The SSP uses these messages to obtain setup, management, and teardown instructions for interconnecting voice circuits and circuit-switches, thereby constructing the connections required to carry the telephone calls. To establish a direct-dial voice connection as shown in Figure 4, the SSP first translates the called number and other signals from the originating switch into SS7 messages and sends the messages to a nearby STP. An STP associated with the originating SSP expands the SSP messages into a set of SS7 signaling messages that address each of the tandem offices and the destination office comprising a complete voice connection. Having established all of the circuits and switches comprising the connection, the STP transmits corresponding SS7 messages to the designated tandem SSPs, terminating SSP, and their associated circuit-switching offices.

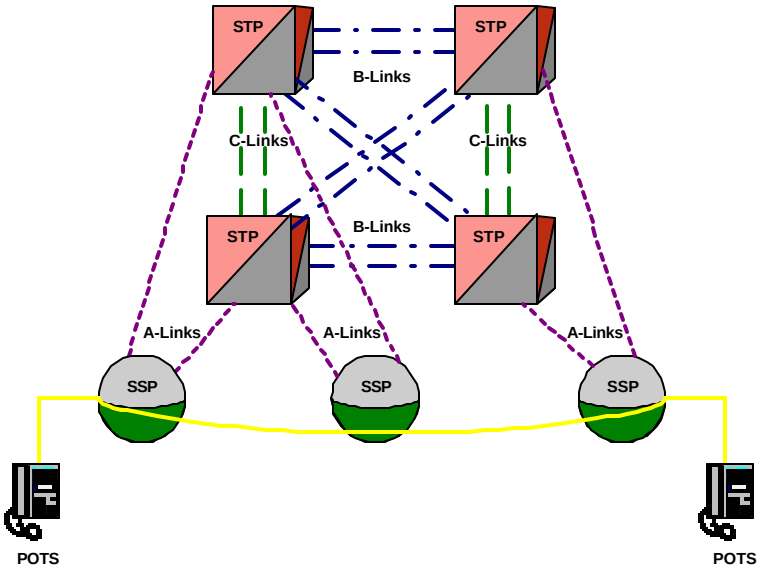

Figure 4: Direct-Dial Voice Connection

If an STP has insufficient information in order to determine which interoffice circuit to use to route a call, an SCP must be used to complete the call. For example, consider the case of a toll-free call (e.g., calls beginning with area code 800, 877, or 888) based on Figure 5. Toll-free numbers do not specify the location of the called party. STPs resolve a toll-free destination ambiguity by sending a database lookup message via intermediate STPs to an SCP equipped with a toll-free number translation database. The SCP responds to the SSP with a translation of a dialed toll-free number into a subscriber telephone number that specifies an unambiguous area code, end office code, and

extension. The inquiring STP is given sufficient information and generates the appropriate SS7 signaling messages to route the call to the end office switch and the called party.

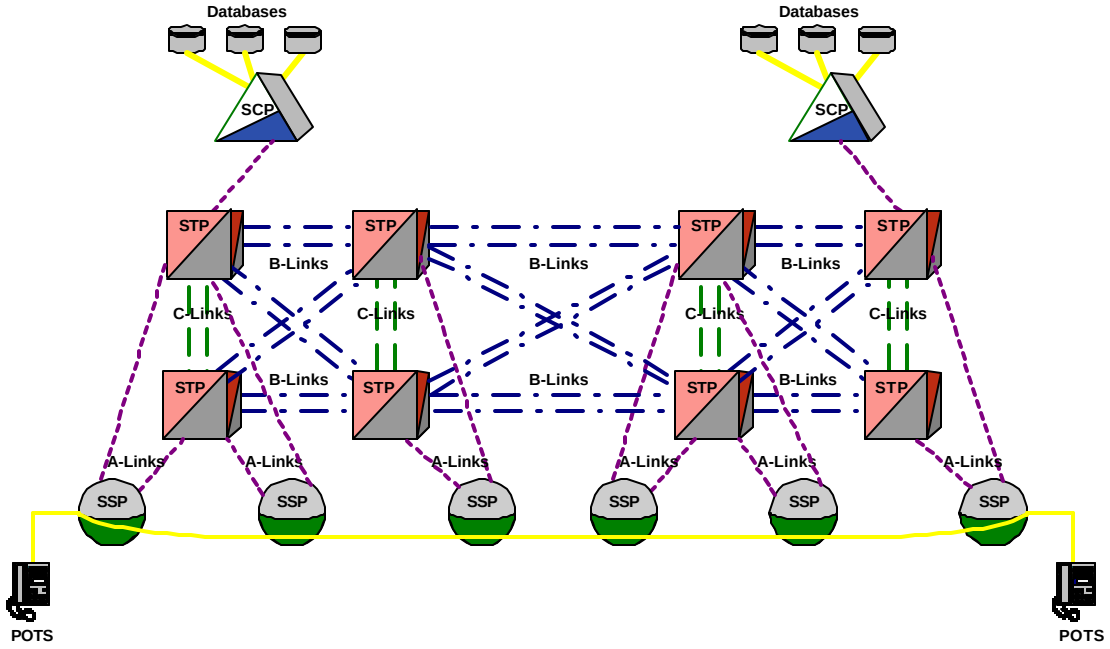

Figure 5: Routed Voice Connection

4.2.2 Signal Transfer Point

The Signal Transfer Point (STP) routes SS7 signaling message traffic through the SS7 network and provides traffic measurements. The STP does not originate traffic other than network management traffic and is never the final destination of SS7 signaling messages. The STP is the packet-switch of the SS7 network and routes each SS7 signaling message from an incoming signaling link (see Section 4.3) to an outgoing signaling link based upon the routing information contained in the SS7 signaling message. STPs improve the utilization of the SS7 network by eliminating the need for direct, or associated, signaling links between all signaling points.

STPs process both circuit-related traffic and database-lookup traffic. Circuit-related traffic is referred to as ISDN User Part (ISUP) traffic and consists of signaling messages originated by an SSP in an attempt to request a connection on a particular dedicated voice-circuit (typically a 64Kbps channel). STPs interpret the message to determine which digits were dialed, and routes the message based on a global title translation of the digits (typically the first six digits consisting of the area code and the office code). In a mobile network, a popular beneficiary of SS7 value-added services, the global title digits consist of the mobile identification number (MIN). The MIN identifies a cellular terminal, or end device.

Database-lookup SS7 traffic is referred to as Transaction Capabilities Application Part (TCAP) and consists of signaling messages that require processing by the STP in conjunction with the Service Control Point (SCP) database. For example, again consider the case of a toll-free call. The originating SSP invokes the services of the STP to resolve the destination office of the toll-free call. The SS7 message provides the toll-free dialed digits to the STP, which in turn uses this information to determine how to route the database lookup request to the appropriate SCP. The SCP replies to the SSP, via the STP, with the requested routing number. The STP may then generate the appropriate SS7 signaling messages to route the call to the proper terminating PSTN switch.

Figure 6 illustrates the three levels of STPs: (1) a national STP, (2) an international STP, and (3) a gateway STP.[10] A national STP performs the STP functions within the scope of a national SS7 network. Commonly a national SS7 network may use a customized SS7 protocol to meet specific requirements. In the United States, the ANSI T1.nnn standards are used for the SS7 protocol. An international STP performs the STP functions within the scope of the international SS7 network using the ITU-T Q.7nn series of recommendations for the SS7 protocol. A national STP may not connect directly to the international SS7 network without enlisting the help of a gateway SS7 STP. The gateway STP translates SS7 signaling messages between the national SS7 network and the international SS7 network, or another national SS7 network. Thus, the gateway STP and the message translation operations it performs is key to providing worldwide connectivity through the SS7 network and may be a prime point of attack either in the US or abroad.

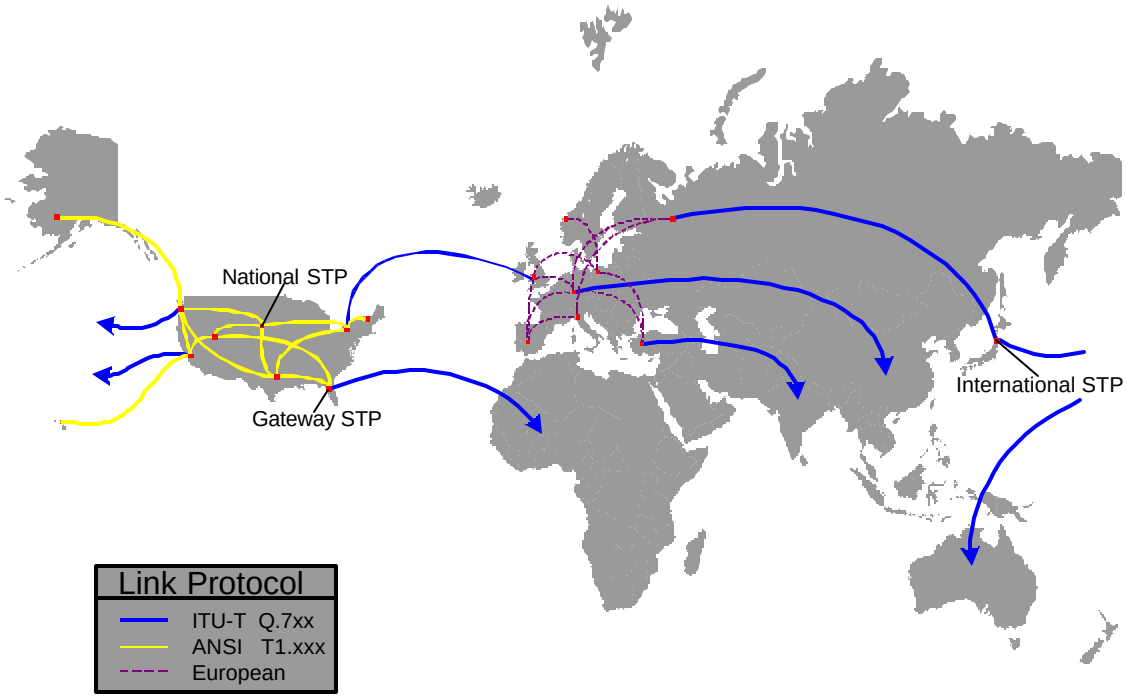

Figure 6: STPs in the Worldwide SS7 Network

4.2.3 Service Control Point

The primary role of the Service Control Point (SCP) is to provide the SS7 network with many value-added services as detailed in Section 4.4 via the vast intelligent network databases of information. Essentially, the SCP serves as an interface to these databases and the software applications that manage them. It makes the Intelligent Network intelligent.

The SCPs obtain the database addressing information from the SS7 messages and provides access to the appropriate database. Requests are routed to the database application based on the database number. Replies are generated for the originating SSP or STP. If a database fails, the STP and SCP reroute the request to other databases. Some of the common databases, also known as subsystems, are the Call Management Services Database (CMSDB) and the Calling Name (CNAM) database.[10] The CMSDB provides toll-free number translation services, network management services, and call sampling services for traffic studies. The CNAM database provides the called party with the name of the party associated with the calling number. The calling number is readily available via the SS7 signaling messages while the calling name comes from a CNAM database lookup.

4.3 Transport

Two different transport mechanisms exist within the traditional telephone network, which contains two distinct planes of operation. The signaling plane utilizes the packet-switched SS7 network as a mechanism for transporting signaling messages that control operation of the voice plane. The voice plane utilizes the circuit-switched central office network as a mechanism to transport voice conversations and data exchanges under the directions of the signaling plane.

4.3.1 Signaling Transport

Signaling transport is performed in the SS7 network over redundantly interconnected bi-directional data links between network nodes, or signaling points. Originally, the CCIS6 network contained data links that operated at 2.4Kbps or 4.8Kbps. Today, the common SS7 data links operate at 56Kbps with an additional 8Kbps of control information. Recently, high-speed data links (e.g., 1.544Mbps) have been deployed, and TCP/IP links (e.g., 100Mbps) have been introduced to substantially increase the speed of the SS7 network and to support many new value-added services.

The interconnection topology (shown in Figure 7) characterizes the SS7 data links as one of six different types. **A-Links**, or access links, directly connect a SSP to a STP, or a STP to a SCP. Each SSP has two A-Links, one to each paired home STPs. **B-Links**, or bridge links, interconnect STP pairs to other STP pairs at identical hierarchical levels and carry signaling messages beyond their initial point of entry into the SS7 network. **C-Links**, or cross-links, interconnect mated STP pairs and are deployed in link pairs. During times of traffic congestion, C-Links carry traffic from one STP to a mated pair and do not normally carry non-congested traffic. **D-Links**, or diagonal links, are similar to B-Links and interconnect STP pairs to other STP pairs at different hierarchical levels. D-Links carry traffic beyond the initial point of entry into the SS7 network. **E-Links**, or extended links, interconnect a SSP to a remote STP for enhanced reliability purposes and provide an alternate path to a STP for use during times of congestion. **F-Links**, or fully associated links, directly connect local SSPs without utilizing a STP. SS7 signaling messages for call traffic between the two SSPs remains isolated to the F-Link. Therefore, F links bypass the security precautions provided by the STPs and may be of NS/EP concern. Finally, not all networks deploy D-Links, E-Links, or F-Links due to the added cost or network topology.

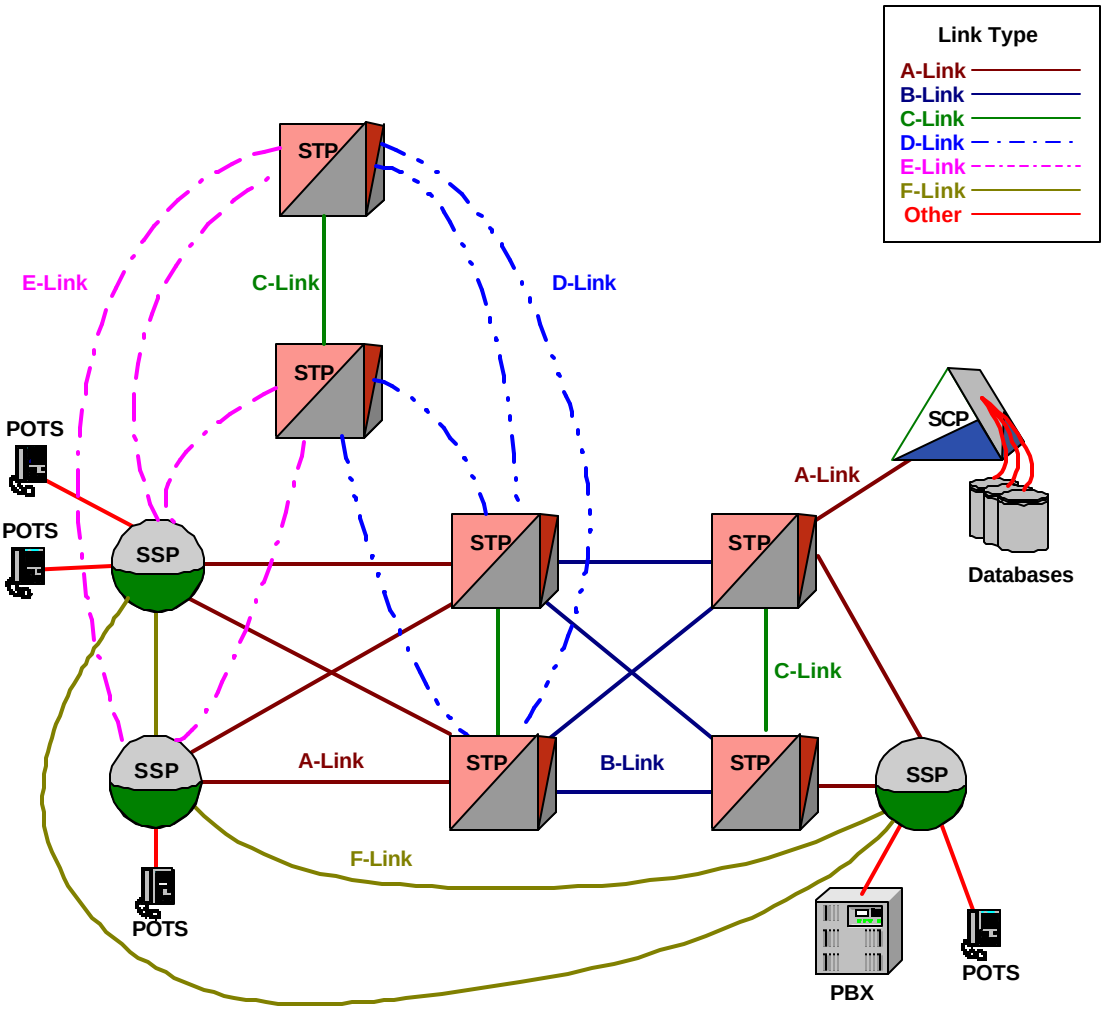

Figure 7: SS7 Interconnection Topology

Additionally, SS7 supports three modes of data link signaling: (1) associated signaling, (2) non-associated signaling, and (3) quasi-associated signaling. Associated signaling (least desired) provides signaling points with a single direct SS7 link. This method is very inefficient, since all signaling points require an associated signaling link to each of the other signaling points in the SS7 network, i.e., a fully associated SS7 network is a fully meshed network. However, associated signaling is simple to implement, such as between two SSPs that constantly share traffic in a heavily populated area. On the other hand, non-associated signaling consists of linking signaling points through multiple STPs. Quasi-associated signaling consists of using a minimal number of STPs to link signaling points, which reduces the overall delay in the transmission of the SS7 signaling messages. Therefore, quasi-associated signaling is the most desired in the SS7 network.

4.3.2 Voice Transport

Voice transport is performed in the voice plane under the direction of the SS7 network in the signaling plane. The voice transport plane consists of central office class 4 and class 5 digital switches that create dedicated circuits between endpoints. These dedicated circuits operate at 64Kbps near the edge of the network and at up to 9.6 Gbps, or faster, in the middle of the network. Dedicated circuits offer very predictable connections since they carry only one type of signal and are not shared with any other transmissions. These characteristics result in a high quality (toll quality) transmission of voice conversations between distant endpoints. Over the years, the public has become accustomed to this quality and will continue to demand the same quality of service in a world that contains a converged voice and data network such as a Voice-Over-IP network.

4.4 Value-Added Services

The SS7 network forms the backbone for the emerging intelligent network (IN), and as such, provides a multitude of new services in addition to call setup, management, and teardown operations. These new services provide benefits to the telephone companies in the form of increased revenue and reliability, and to the end-user in the form of increased efficiency and functionality. The SCP databases store the required information that makes these services possible. A list of value-added services offered via SS7 is presented here to illustrate the types of services demanded from the PSTN.[10]

Reduced Central Office Dependency on Dedicated Circuits – If the STP determines via the SS7 network that the called party currently has a busy line, then the ingress switch will not attempt to create a dedicated switched circuit to the egress switch. Before the SS7 network was widely deployed, the ingress switch was required to make the dedicated switched circuit to allow the egress switch to signal the calling party using audible in-band tones (i.e., busy and ringback). This in-band signaling would essentially waste a circuit if the called party had a busy line. This new SS7 network scenario saves the dedicated switched circuit for other revenue generating calls if the called extension is busy.

Call Screening (including anonymous call rejection) - consists of blocking certain calling parties from connecting to a called number through the SS7 network. An SCP database lookup provides the information on whether to block the call or not. If the call is blocked, it may be routed to a voice message prompt stating that it was blocked.

Find-me Service - consists of forwarding screened calls to another number in an attempt to allow privileged calls to find the subscriber.

Follow-me Service - is similar to the Find-me service in that screened or unscreened calls will be forwarded to another number during subscriber defined time periods.

Local Number Portability - allows the subscriber to move to a new address or to change their telephone company without changing their telephone number.

Computer Security Service - screens unauthorized users from accessing a computer via a modem connection. Either an access code or an authorized calling number is required for permission to connect.

Call Pick-up Service - enables a subscriber to pick up an incoming call from another telephone when notified via a paging service. Additionally, incoming calls may be screened before this service is invoked.

Store Locator Service - provides a business with the opportunity to advertise a single telephone number and request that incoming calls be routed to the nearest store.

Multi-location Extension Dialing - allows subscribers to be reached at any location using a single number and without using a PBX.

Calling Number Delivery - provides, to the called party, the name of the calling party that accompanies the calling number. The calling number is readily available via the SS7 signaling messages, but the calling name comes from a CNAM database lookup.

Outgoing Call Restriction - allows the subscriber to place restrictions on the destination of outgoing calls. These restrictions may include area codes, office codes, or complete numbers. A very common use of this service includes blocking 900 toll calls.

Busy Number Ring Again - gives a caller an option to request a ringback when a called number is busy, or unavailable.

Automatic Callback - provides the caller access to a database and an opportunity to enter an access code that triggers the system to automatically call back the user at a predefined number. Automatic callback is a popular security measure for ensuring that databases are accessed from known parties operating at known locations.

Three-way Calling - creates a conference call between three parties at three different numbers.

Although this list of value-added services is quite extensive, there remain many more value-added services that the SS7 network cannot provide and that a converged voice and data network may well provide. For example, the capabilities to exchange hi-fidelity audio and to combine paging messages, voice messages, and video messages into a single inbox would be possible in a converged network. Currently, it is extremely difficult for the SS7 network to provide value-added services when new services require changes to traditional circuit-switching equipment.

4.5 SS7 Protocol Stack

The SS7 network is a packet-switched network that exchanges signaling messages in support of telecommunications operations including value-added services. As such, the protocol for the SS7 network is a layered protocol with some similarities to the OSI seven-layer reference mode. Figure 8 illustrates the SS7 four-level protocol next to the OSI seven-layer reference model.

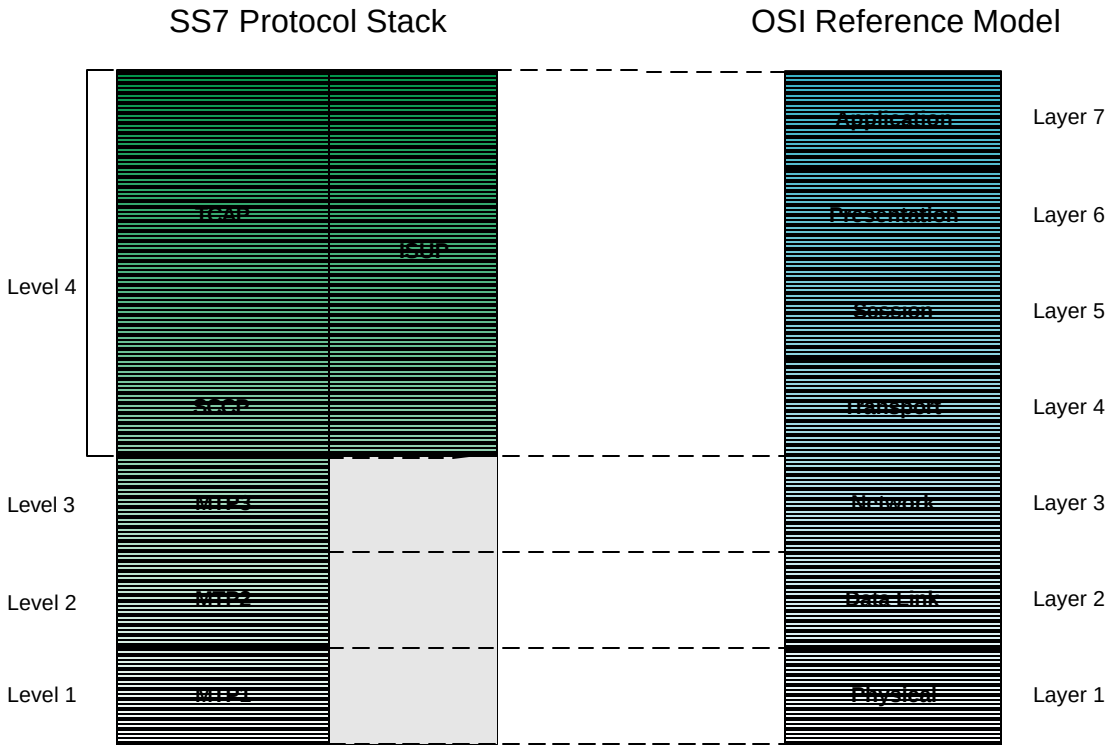

Figure 8: SS7 Protocol Stack and the OSI Reference Model[12]

The SS7 protocol stack is split into two parts: the message transfer part (MTP) and the user/application part (ISUP, TCAP, and SCCP). The MTP is further split into three levels: MTP 1, MTP 2, and MTP 3.

4.5.1 Level 1 - Physical Level (MTP 1)

The physical level (MTP 1) of the SS7 network is very similar to the physical layer of the OSI model. As any protocol should stipulate, the physical level, or layer, should be open for the use of any type of transmission medium, including transmission rates. Typically, the MTP 1 links in the SS7 network between SSPs and STPs and between STPs and SCPs are DS0A (56Kbps) and V.35 (64Kbps). The MTP 1 links in the SS7 network between STPs are often DS1 (1.544Mbps), or faster.

4.5.2 Level 2 - Data Link Level (MTP 2)

The data link level (MTP 2) of the SS7 network mirrors the data link layer of the OSI model. MTP 2 is predominantly concerned with the accurate transmission of the SS7 messages. It includes error detection, error correction, and message packet sequencing. If the SS7 message is not transmitted or contains errors, the level 2 protocol requests a retransmission of the SS7 message. The level 2 protocol receives the necessary routing

information from the level 3 protocol and provides the required functions to route the message packet to the next destination.

4.5.3 Level 3 - Network Level (MTP 3)

The network level (MTP 3) of the SS7 network performs four important functions: (1) message discrimination, (2) message distribution, (3) message routing, and (4) SS7 node and link status determination. First, MTP 3 determines, via the message point code, if the destination address is one of the local addresses. If so, then it passes the message to the message distribution function. If not, then it passes the message to the message routing function. Also, the MTP 3 determines the operational state of SS7 nodes or links. When problems are detected, the MTP 3 attempts to mitigate the problem by rerouting the message and informing the adjacent node to send traffic over alternate links. The combination of these functions allows the SS7 messages to be effectively routed through the SS7 network and around congested or failed components to the appropriate destination.

4.5.4 Level 4 – ISDN User Parts (ISUP) and Application Parts (TCAP)

The ISDN User Parts (ISUP) and the application parts (TCAP) of the SS7 network perform several different functions. The ISUP part supports the call setup, maintenance, and teardown operations of the PSTN network. The TCAP part supports the access of the SCP applications and databases.

4.6 Security Features and Limitations

Service provider domains often consist of many SSPs, STPs, and SCPs. They manage SS7 networks from a central SS7 network management system called the Service Management System (SMS). The provider's SMS connects with each SCP and provides an interface for manual or automatic updates of the SCP databases. Through an automatic mechanism, the SMS becomes a central authority for updating multiple SCP databases. Therefore, it is important to secure access to the SMS to prevent unauthorized, and possibly destructive, changes to the valued SCP databases as shown in Figure 9.

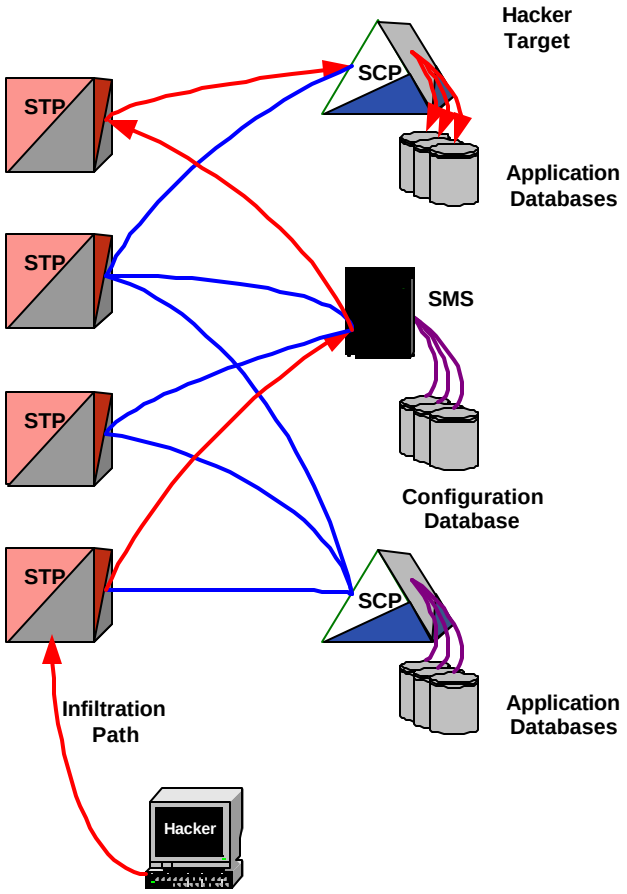

Figure 9: SS7 Service Management System Security

Domain operators interconnect their PSTN networks to exchange traffic that traverses more than one network in route to the called party. As a part of their voice network interconnection, domain operators interconnect their SS7 networks for the purpose of exchanging signaling information associated with the voice traffic traversing their domains.

At the network interconnection points, SS7 gateways route, translate, and screen messages, including network status messages, that pass between two interconnected SS7 networks.[9] The SS7 gateways filter these SS7 messages by comparing the fields in the SS7 messages against predefined criteria that define allowable message exchanges. As a result, these gateway functions help to ensure network security by analyzing the SS7 message traffic and blocking corrupt or unauthorized traffic.

A primary application for interconnecting SS7 networks is for the exchange of customer credit and billing information. For example, SCPs in one domain may access the Line

Information Database (LIDB) controlled by a SCP in another domain. Access to the LIDB provides the means for handling services such as calling card verification, call forwarding, and speed dialing across the interconnected networks.

In the past, service providers limited SS7 component access to their trusted employees. Regulatory changes and the introduction of new services like Voice-over-IP are opening the service provider network to greater access by third parties and external systems. The impact and risks associated with changing the SS7 environment is discussed in later sections.

4.7 ITU-T SS7 Similarities and Differences to ANSI SS7

Several important differences exist between the ITU-T recommendations and the ANSI standards for the SS7 network that are of concern for interoperability.[11] For example, the ANSI standards (United States) states that the ANSI SS7 network is *generally* compatible with the international ITU-T SS7 network. In fact, the ITU-T recommendations permit, as an option, the customization of a national SS7 network that meets the specific needs and desires of that nation's telecommunications network. However, traffic that emanates from a national SS7 network into the international SS7 network MUST be 100% compliant with the ITU-T recommendations. Section 0 identifies three types of networked STPs in the SS7 network. The gateway STPs are required to perform conversions between a national SS7 network and the international SS7 network. This process simply cannot maintain a transparent conversion of all value-added services between different networks.

A majority of the differences between the ANSI SS7 network and the ITU-T SS7 network exist for several reasons. First, some portions of the ITU-T recommendations are for use at the discretion of the individual nations. For example, the ITU-T recommendations permit the use of four different formats for routing labels, which identify the destination, while the ANSI standards only permit the use of one routing label format. Second, the ANSI standards make extensions to the SS7 protocol available. For example, the lengths of several parameters in the Message Transfer Part and the ISDN User Part of the SS7 protocol are not consistent between the ANSI network and the ITU-T network. Finally, the ANSI documents and the ITU-T documents are not reviewed on identical schedules. This differentiating review schedule tends to allow enhancements in the ITU-T documents that are not in the current version of the ANSI standards until the next review data.

More notable differences between the ANSI standards and the ITU-T recommendations include provisions for network congestion control and additional security procedures for preventing access to the STPs by an unauthorized user or entity. The network congestion control is accomplished by providing different priority levels for messages that identify messages that may be discarded during times of network congestion. Incoming international traffic needs to be evaluated at the gateway STP and assigned a priority level before transmitting the message into the protected US SS7 network. This assignment procedure at the gateway STP seems to be a prime target for intrusion. Additionally, authorized access to the protected US SS7 network is controlled at the STP via a process of comparing the destination point code (DPC) and the origination point

code (OPC) of each SS7 message to an authorization list, which is setup by the STP administrator. The STP may then either allow or disallow entry of each SS7 message into the network based upon this list. This list forms the basis for entry into the protected US SS7 network. If an unauthorized user compromises this list, then unintended SS7 messages might be allowed entry into the network. Traditionally, physical access to the computers controlling the list served as an important element in the security of the list. As data networks extend the convenience of many tasks to the employee's desktop, security concerns increase. Once list maintainers allow list modifications from alternate physical locations, the need for enhanced list security becomes imperative. Network data security such as virtual private networking (VPN) tunnels and strong authentication will be necessary to protect the lists from other network users.

5.0 THE VOICE-OVER-IP ARCHITECTURE

The Voice-over-IP technology offers the opportunity to capitalize on the prevalence of IP networks to both increase availability of voice communications and reduce the infrastructure costs associated with voice-only networks. The following sections explain how Voice-over-IP works from a data perspective focusing on usage wholly within an IP network.

5.1 Signaling

As in the world of circuit-switched telephony, Voice-over-IP signaling provides the mechanism for joining the call participants in real-time. Signaling in a Voice-over-IP call can be generalized into five operations: User Location, Session Establishment, Session Negotiation, Call Participant Management, and Feature Invocation.[7] User Location defines the mechanism for the caller to locate the called party. Session Establishment focuses on how the called party responds to the call request. Responses to a Session Establishment request include accepting the call, rejecting the call, or transferring the call to another number or service (such as voicemail, or a pager service). Session Negotiation occurs when a call is accepted. The calling and called parties exchange information about their respective capabilities (such as encoding standards supported by their codecs), and once they agree to a compatible set of capabilities, the call completes. Call Participant Management service allows additional parties to be added or removed from an active call. Feature Invocation allows either party to invoke features available in traditional phones such as hold, transfer, and mute. These five operations are predominately handled by either SIP/SDP or H.323 with signaling occurring on the same network as transport. SIP/SDP and H.323 offer competing methods to handling Voice-over-IP signaling.

SIP/SDP touts simplicity while H.323 emphasizes completeness and traditional telephony. Most companies are developing and emphasizing H.323 compatibility, yet are implementing SIP/SDP today. This is mostly due to the relative complexity of H.323 and implementation compatibility problems.

5.1.1 IETF Voice-over-IP

The Internet Engineering Task Force (IETF) created an architecture for Voice-over-IP that focuses on passing voice across an IP network. This architecture mirrors other common Internet services such as HTTP, SMTP, and DNS. SIP/SDP messages are sent in plain text and therefore are relatively easy to read without additional software. The following sections describe the core protocols that provide the signaling discussed above.

5.1.1.1 Session Initiation Protocol

The Session Initiation Protocol (SIP) defined in RFC-2543 provides four of the five basic signaling operations described in Section 5.1. It does not provide Session Negotiation. However, Session Description Protocol (SDP), which is described in Section 5.1.1.2, does provide Session Negotiation. Under SIP, the participants in a call use software on

their computers, or Internet telephone appliances, called user agents. Typically, participation is bi-directional, so each user agent has both a client (UAC) to initiate a call and a server (UAS) to answer or reject a call. SIP servers form the signaling backbone of a SIP Voice-over-IP call. SIP servers provide one to three functions: registration, proxy, and redirection. A UAC updates its location with a SIP registration server. SIP proxy servers forward SIP messages between the UAC, UAS, and other proxy servers. SIP redirection servers return forwarding instructions to the UAC or proxy, but do not forward messages.

SIP offers User Location, Session Establishment, Call Participant Management, and some degree of Feature Invocation. Figure 10 shows the SIP architecture. SIP uses SIP servers to determine location of the called party. A caller configures the user agent software with the email address of the caller. The email address then becomes akin to a phone number as the identifier of a UAS. The UAS then registers the email address with a registration server to indicate the caller's current location. The caller instructs the UAC to send an INVITE message to the local SIP proxy server. The local SIP proxy server uses DNS to locate the proxy server corresponding to the domain of the called email address. The local SIP then forwards the INVITE message to the email domain proxy server. The email domain proxy server queries the email domain's registration server to determine the exact location of the user within the email domain. If the exact location of the user is within the local domain of the email domain proxy, the INVITE message is forwarded to the called party's UAS. If the user is within the email domain, but not local to the email domain proxy, the INVITE message is forwarded on to the proxy local user, which in turn forwards the message to the called party's UAS. If the called party is not within the email domain, the proxy will act as a redirection server. It will return the message to the calling party's local proxy server with the new email domain of the called party. The process will then repeat until the called party's UAS is reached.

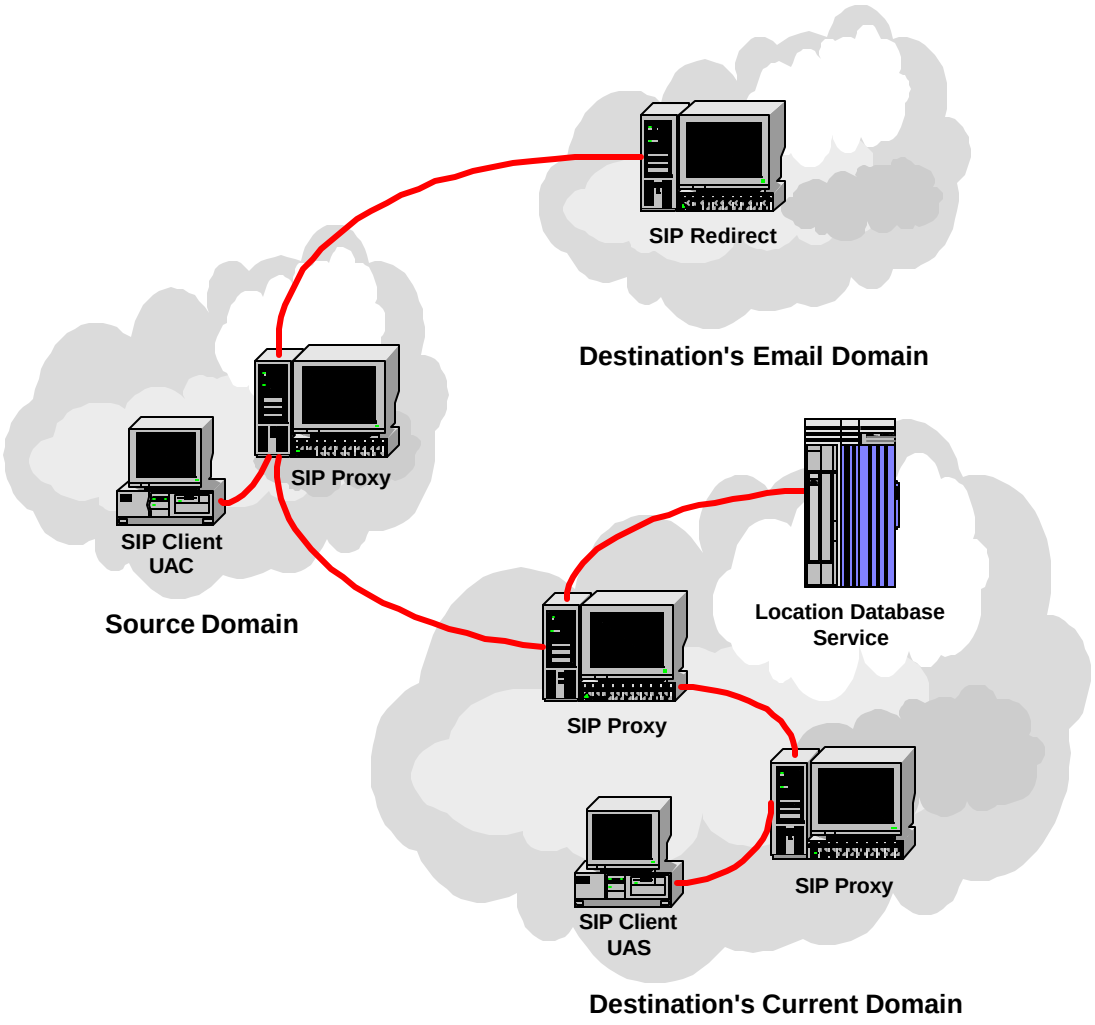

Figure 10: SIP Architecture[7]

Once the UAS is found, Session Establishment begins. The UAS receives the INVITE message. The message is accepted, rejected, or redirected. The INVITE message is an ASCII text message containing *TO, FROM, SUBJECT, CALL-ID, CSEQ, CONTACT, REQUIRE, CONTENT-LENGTH*, and *CONTENT-TYPE* fields. *CONTENT-LENGTH* and *CONTENT-TYPE* define the characteristics of the message body. SIP also allows for extensions to be defined with new header fields. This flexibility enables SIP to provide feature invocation of services not specifically defined in SIP such as call transfer. Extensions can also be used to provide call participant management such as multiparty conferencing.

5.1.1.2 Session Description Protocol

The Session Description Protocol (SDP) defined in RFC-2327 provides the remaining signaling operation of Session Negotiation. SDP establishes the media stream types, the

addresses and ports for the media streams, the payload types, the start and stop times for broadcast media streams, and the originator identifier. The SIP INVITE messages contain an SDP body that allows the involved parties to negotiate session parameters such as the desired type of codec during the Session Establishment operation.

5.1.2 ITU-T Voice-over-IP

The ITU-T created an architecture for Voice-over-IP similar to the PSTN. This architecture mirrors other common telephony services such as ISDN. The following sections describe the core protocols that provide the signaling discussed in Section 5.1.

5.1.2.1 H.323 Protocol Suite

The ITU-T H.323 recommendation defines a protocol suite to provide multimedia communications.[8] The H.323 architecture consists of four main components as illustrated in Figure 11. The terminal, as the communications endpoint, is generally a PC or an Internet telephone appliance. The gateway provides bridging services to other networks and the gatekeeper (GK) provides control and routing. The multipoint control unit (MCU) serves three roles: (1) a multipoint controller (MC), (2) a multipoint processor (MP), and (3) a T.120 server. As a MC, the MCU coordinates control functions for a multipoint conference. Conference participants perform their call setup signaling with the MC. The MC ensures that all participates negotiate to a common protocol (such as requiring all participates to use G.729 for audio). The MC also tracks the availability of additional services such as video or data. If mixing is desired, such as allowing more than one participant to speak simultaneously, the MC will control the MP to perform the mixing. In this way, only one combined audio stream goes to each participant, instead of each participant receiving an audio stream from each of the other participants. This method reduces the requirements on each participant's terminal and typically reduces network bandwidth requirements. If the MC negotiates data sharing, the MCU will also function as a T.120 server. A T.120 server enables collaborative development by sharing workspaces.[20]

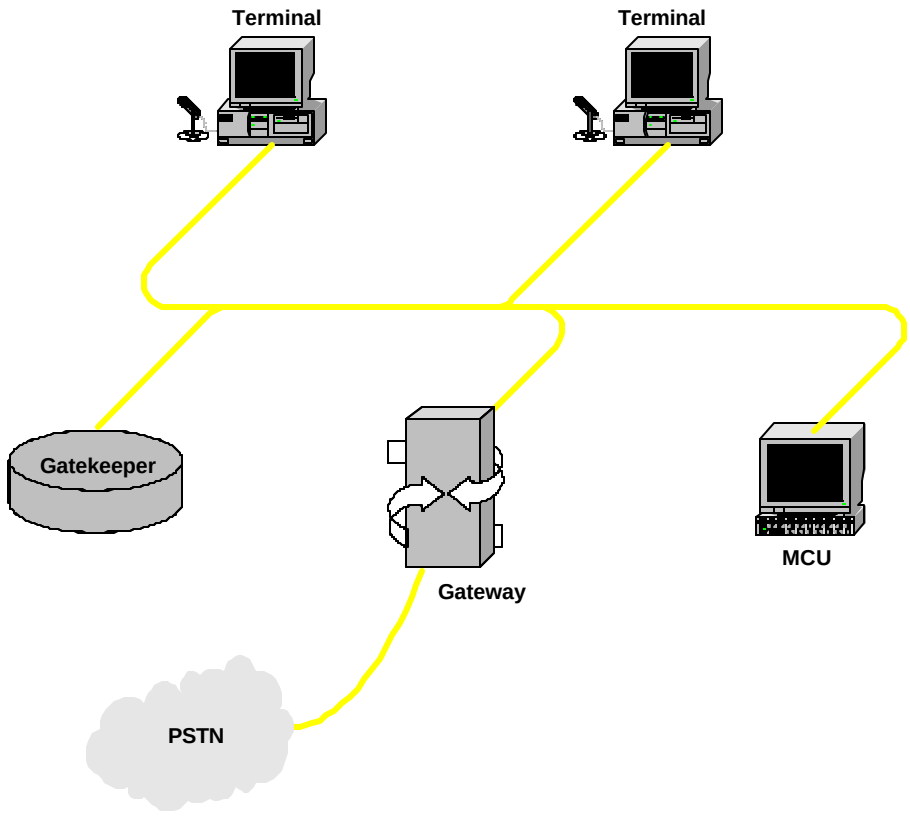

Figure 11: H.323 Architecture

Within this suite, H.225.0 controls the signaling and includes two sub-protocols: RAS (registration, admission, and status) and Q.931.

The RAS channel provides the User Location function of signaling. Call endpoints use RAS to register with a Gatekeeper that is logically close to the endpoint within the network (a.k.a. in the "network zone" serviced by the Gatekeeper). Gatekeepers serve as the administrators of a network zone. RAS also provides the mechanism for a Gatekeeper to perform address resolution, call prioritization, and call status.

The Q.931 protocol provides the Session Establishment function of signaling, which includes handling the setup, call proceeding, alerting, and connect messages necessary to complete a call. The Q.931 signaling protocol of H.323 is an adaptation of the Q.931 signaling protocol developed originally by the ITU-T for use in ISDN.

Figure 12 describes a direct endpoint signaling exchange using H.323. Terminal A attempts to discover its Gatekeeper by sending a Gatekeeper Request (GRQ) multicast to 224.0.1.41. The Gatekeeper responds with a Gatekeeper Confirm (GCF) unicast to Terminal A. Terminal A registers its location and related information by sending a Gatekeeper Registration Request (RRQ) to the Gatekeeper. The Gatekeeper responds with a Gatekeeper Registration Confirm (RCF). Terminal A initiates a call to Terminal B by sending a Gatekeeper Admission Request (ARQ) to the Gatekeeper. The Gatekeeper

responds with a Gatekeeper Admission Confirm (ACF). Terminal A sends a Q.931 setup message to Terminal B. Terminal B sends a Q.931 call proceeding message to Terminal A. Terminal B initiates the return audio path by sending an ARQ to the Gatekeeper. The Gatekeeper responds with an ACF. Terminal B sends a Q.931 alerting message and a Q.931 Connect message to Terminal A.

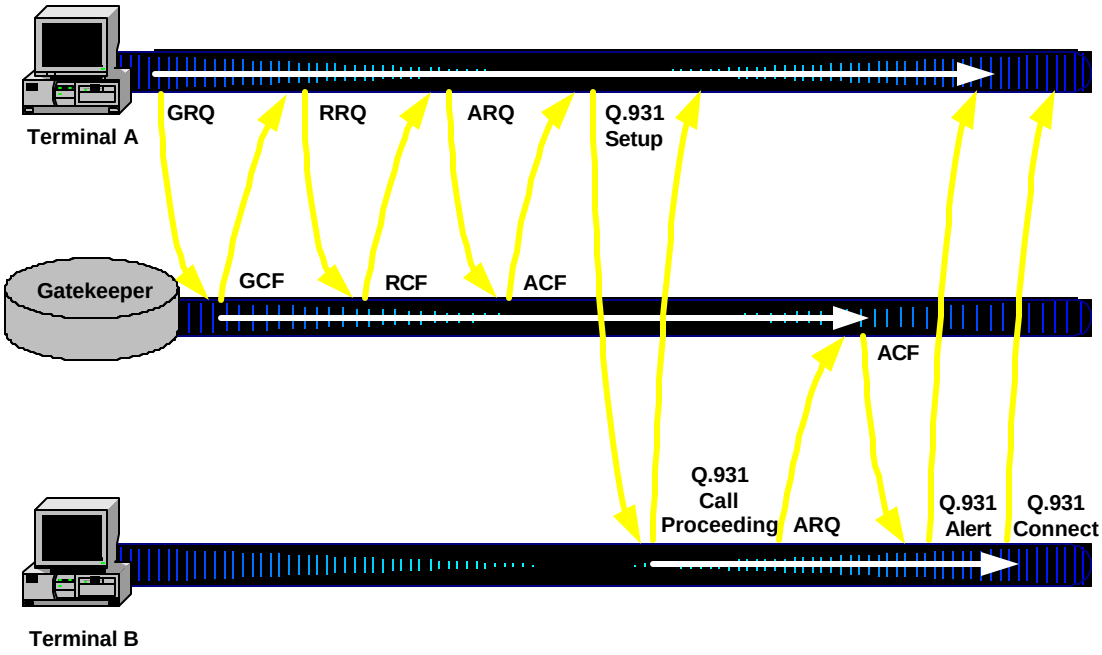

Figure 12: H.323 Signaling

H.245 handles the remaining signaling functions of Session Negotiation, Call Participant Management, and Feature Invocation. H.245 performs a master-slave determination to establish which endpoint terminal takes precedence in a tie-breaking condition. For example, if each endpoint terminal supports the same set of audio codecs, the master will choose the audio codec for the session. The master-slave determination process also identifies which Multipoint Controller (MC) will control conferencing if multiple MCs are available. The capability exchange negotiates the media types, channels, bit-rates, and other session parameters. H.245's media channel control establishes and tears down

logical media channels. Logical media channels include audio, video, and data transport streams between conference endpoint terminals. If during an audio conversation, a video stream is made available, H.245 will create a video channel to transport the video stream. The conference control function performs Call Participant Management.

5.2 Transport

Once signaling completes the call, exchange of audio begins between the parties. The IETF defines the Real-time Transport Protocol (RTP) in RFC-1889 for transmission of audio using Voice-over-IP. Its companion protocol, the Real-time Transport Control Protocol (RTCP), monitors the receipt of the packets to provide feedback about the performance of the connection. Both H.323 and SIP rely on RTP and RTCP to transport audio packets and monitor the performance of the transport of the packets.

5.2.1 RTP

The RTP header describes five important characteristics about the RTP payload: (1) the packet order, (2) the timestamp, (3) the payload type, (4) the frame mark, and (5) the synchronization source. A sequencing field is used to define packet order, enabling detection of lost packets or reordering. The RTP timestamp enables intramedia synchronization to manage jitter. The payload type identifier lists the encoding of the payload. The encoding information was also exchanged during signaling using SDP or H.245, but is repeated here to facilitate dynamically changing the encoding type to adapt to changing demands or network conditions. The frame marker bit identifies the beginning or end of an audio frame to assist in intelligent discard by the network or synchronization by the higher-level applications. The synchronization source (SSRC) identifier identifies the source of audio payload for coordination in a multiparty conversation.

5.2.2 RTCP

The RTCP header provides feedback about the performance of the connection to each participant. For example, RTCP header fields provide quality of service feedback on packet loss, jitter, and round-trip delay. RTCP headers contain information that enables intermedia synchronization such as to provide lip-sync between an audio and a video stream. Identification header elements provide data such as email address, phone number, and name of the participant. RTCP headers also provide some session control, enabling users to leave a conference session or exchange small text notes.

5.3 Value-Added Services

The Voice-over-IP technology offers several advantages to the traditional public switched telephone network (PSTN) as stated earlier. Simplicity of services can be achieved through the integration of voice, email, and fax into a single coordinated service.[3] Some new services likely to arise from Voice-over-IP include web-based call centers, hi-fidelity audio exchange, unified messaging from multiple sources to a single inbox, virtual second line, voice cost reduction especially for on-hold conditions, real-time

billing, remote teleworking via desktop conferencing, and enhanced teleconferencing with white board plus application sharing.

5.4 Similarities to SS7

The H.323 standard was designed by members of the telephone industry and therefore has many similarities to SS7. The signaling similarities are best explained when considering a call between a POTS telephone unit and an H.323 terminal as shown in Figure 13.

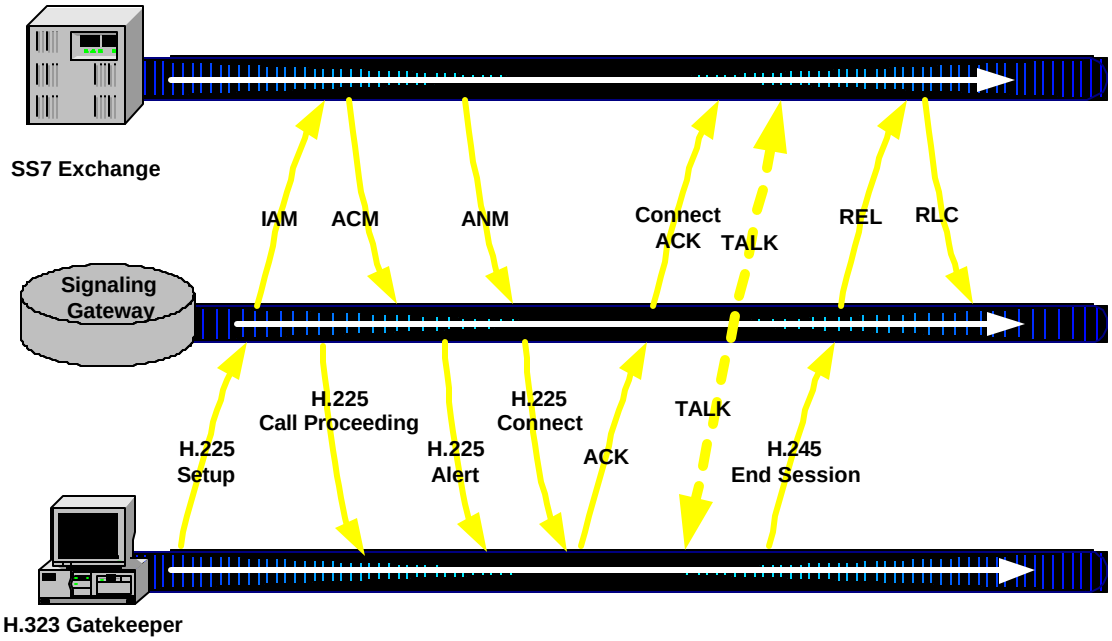

Figure 13: H.323 to SS7 Call Signaling[20]

The H.323 gatekeeper sends a H.225 Setup to the H.323 gateway. The gateway translates the H.225 Setup into an ISUP IAM that is sent to the STP and on to the SSP. Once the ISUP IAM is sent, an H.225 Call Proceeding is sent from the gateway to the gatekeeper. The ISUP ACM connecting response from the SS7 network is translated by the gateway into an H.225 Alert. The ISUP ANM answer response from the SS7 network is translated by the gateway into an H.225 Connect. A similar comparison of SS7 call setup initiated by a SIP user is shown in Figure 14.

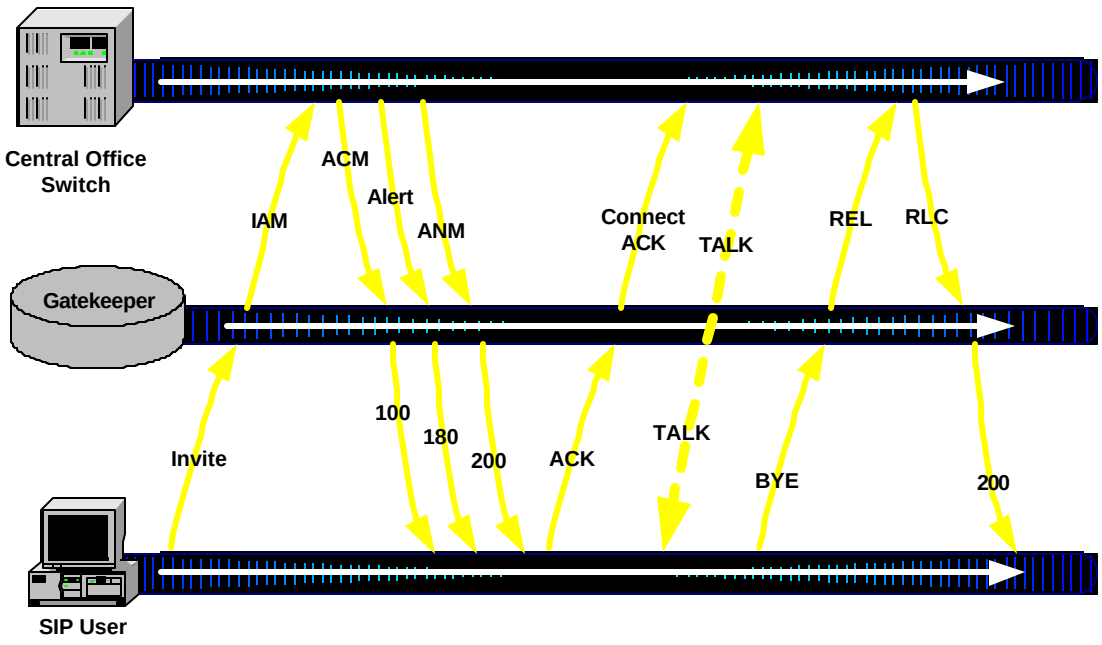

Figure 14: SIP to SS7 Call Signaling[19]

Similarities between Voice-over-IP and SS7 end with signaling. Voice transport and signaling occur over separate networks in the SS7 architecture. In Voice-over-IP, both signaling and transport occur over the same network. The SS7 voice network is based on circuit-switching compared to packet-switching in Voice-over-IP networks.

These differences between circuit-switched voice and Voice over IP are key factors in areas such as reliability, fault tolerance, restoration, and security. Reliability, fault tolerance, and restoration are major new issues for Voice-over-IP networks that are not adequately addressed within the Internet data-only paradigm. Consideration should be given to Voice-over-IP reliability, fault tolerance, and restoration, especially during times of crisis or natural disaster. Voice-over-IP reliability, fault tolerance, and restoration are discussed later in section 8.0. The remainder of this section considers security issues raised by Voice-over-IP systems.

5.5 Security

A Voice-over-IP system (ignoring gateways to the PSTN) suffers similar security concerns faced by other devices connected to a public network such as the Internet.

5.5.1 Security Risks

SIP servers and H.323 Gatekeepers, Gateways, and MCUs act as servers allowing Voice-over-IP applications to work. Just like web servers and mail servers, these Voice-over-IP servers stand out as targets for malicious individuals and groups. Denial-of-Service (DoS) attacks on Voice-over-IP servers can render a group of Voice-over-IP terminals

useless. This can create a particularly frustrating experience to Voice-over-IP users who will compare service outages to the reliable PSTN.

Compromise of a Voice-over-IP system can have serious ramifications. Although most Voice-over-IP servers today do not contain billing information, this looms on the future as a means of paying for enhanced quality of service. When that happens, Voice-over-IP servers may suffer the same fate as e-commerce web sites. They may be targeted to obtain credit card or calling card numbers to commit extortion or fraud. Even today, some risk exists. Voice-over-IP servers may contain information users desire to keep private such as their current location, physical mailing address, or traditional PSTN phone number.

Compromised Voice-over-IP servers can create a nuisance to users by having the server continuously "ring" all Voice-over-IP users registered with the server. This scenario can spiral into a DoS condition, since IP stacks can handle only a finite number of active connections (sockets).

Today, misconfigured email servers can act as unwitting open relay hosts. Open relay hosts forward email messages to their destination while obscuring their true source. In a parallel example, an open relay Voice-over-IP server could forward calls either anonymously or under the pretense of the compromised organization.

Finally, Voice-over-IP data exchange today typically occurs in clear-text messages. This means that anyone with a network sniffer can listen in on a Voice-over-IP conversation. This poses a significant privacy concern considering the prevalence of network sniffers, and the current lack of "wire-tapping" laws pertaining to Voice-over-IP.

Therefore, it is easy to imagine network outages brought about unwittingly through misconfiguration of a Voice-over-IP server or deliberately through deliberate exploitation. In either case, Voice-over-IP servers represent a growing vulnerability within the national infrastructure as interactive voice communications migrate away from the public circuit-switched network and onto packet-switched networks. A growing component of the national infrastructure will come to rely on Voice-over-IP components for their day-to-day communications and depend on it when disaster strikes. Careful consideration should be given to the protection of Voice-over-IP network components and to the appropriate allocation of Voice-over-IP resources during times of national crisis and natural disasters.

5.5.2 Security Solutions

Solving Voice-over-IP security vulnerabilities requires several steps. The first step involves improving the integrity of Voice-over-IP servers. This is accomplished by both hardening the operating system of the server, and protecting the server behind a firewall. Firewalls can provide a significant deterrent to server compromise. Modern firewalls also provide limited protection against DoS attacks.

Unfortunately, placing a Voice-over-IP server behind a firewall is not a simple task. The rules of the firewall must be modified to permit traffic to the Voice-over-IP server. For H.323, this requires allowing H.225 (RAS and Q.931) and H.245 to pass through the firewall. If H.323 terminals also exist behind the firewall, and are permitted to call users

outside the firewall, RTP must be allowed through the firewall. A separate RTP session is used for audio and video. As can be seen, the number of connections (sockets) that a firewall must pass to enable Voice-over-IP is significant. This poses a challenge to many existing firewalls to add support for these connection types. Further complicating the problem is the need for complete inspection of these packets. Since many of these connections use dynamic port assignments, the only mechanism for tracking each connection as a part of a permitted call involves extracting information about related connections from existing connection headers and payloads.

Conference calls using multicast also present security problems. Older equipment that do not support multicast treat such packets as broadcast data. Packets are then forwarded to network sections without call participants. Sniffer programs can exploit this to maximize its eavesdropping potential by deploying near one of these older devices.

Borrowing from security methods from other Internet applications can solve some of the other vulnerabilities with Voice-over-IP. Nonrepudiation can be achieved through the exchange of digital certificates and the use of digital signatures attached to each message. This method is currently employed for e-commerce web sites to established encrypted communications sessions (typically using the secure sockets layer, SSL). To deter eavesdropping, the RTP session can be encrypted between terminals. This can readily be accomplished by using Virtual Private Networking (VPN) solutions including IPSec. Encryption algorithms must be made with care for the VPN. RTP relies on UDP, since generally there is not sufficient time to retransmit a lost or corrupt packet. Encryption algorithms that span multiple packets may place a requirement that all packets be successfully received to enable decoding. Methods such as this may significantly undermine performance designs for Voice-over-IP systems.

All of these techniques can mitigate vulnerability, but extract a price on the performance of the Voice-over-IP session. Packet inspection, digital signature, and data encryption add latency to a connection. As previously discussed, latency margins are already tight for Voice-over-IP sessions using the Internet. Use of these solutions can push the system latency over the critical 150 ms level. The price of security therefore takes a particularly costly toll on Voice-over-IP.

5.5.2.1 H.235

Recognizing the need for security, the ITU-T established the H.235 specification. H.235 provides mechanisms for authentication, encryption, and data integrity. H.235 relies on extensions to the H.225 and H.245 standards, which implicitly require recent versions of the protocol (at least version 2 and version 3, respectively) to be employed in the H.323 session. For call admission, H.235 uses extensions to H.225/RAS to exchange support, either private-key or public-key encryption, for communications between the endpoint and gatekeeper. An integrity check value in H.225/RAS assures data integrity during call admission. During call establishment and control, H.225/Q.931 utilizes the validation performed during call admission to authenticate the connection. Data integrity of the H.225/Q.931 message may be achieved via H.235 tokens, since an integrity check value was not added to the H.225/Q.931 message. Call establishment can also be encapsulated in IPSec, minimizing or eliminating the need for H.235 enhancements. H.235 defines the use of H.245 Object Identifier tags to negotiate encryption schemes among participants

and distributes session keys. H.235 also supports dynamic key changes to enable conference participants to be individually removed from a session.

6.0 BRIDGING THE GAP

The momentum behind implementing Voice-over-IP continues to build. Vendors of Enterprise network equipment like Cisco continue to enhance their product lines to facilitate integrating Voice-over-IP technology into the existing Enterprise class networks. Large corporations are deploying this new technology, and carriers will have no choice but to provide hooks into their networks to allow some level of billing to continue as they adapt their business models to the data-centric era.

6.1 Benefits of Union

ISPs, Competitive Local Exchange Carriers (CLECs), RBOCs, and Inter Exchange Carriers (IXCs) see cheap telephone calls, avoidance of access charges, and avoidance of settlements as a temporary incentive to building Voice-over-IP networks.[2] The long-term incentives for Voice-over-IP are to be found in value-added, multimedia services (e.g., push-to-talk Web pages, Internet call waiting, enhanced call center functions); consolidation to a single, easily managed network that is flexible and easily expanded; and reduced maintenance cost through the use of modern communications equipment. The carriers currently spend approximately 75% of their revenues on network maintenance.[2] Modern all-data networks are easier to maintain and operate, which results in lower costs than their conventional circuit-switched counterparts.[15]

6.2 PSTN Gateways

The only network that rivals the PSTN in size is the Internet. Although the exponential growth of the Internet indicates its eventual ubiquitous status, today the PSTN still reaches more individuals than the Internet. This reality is especially important when focusing on interactive voice communications. Most Internet users do not currently have the ability to easily establish voice communications over data networks, yet they have ready access to a telephone to establish data communications over voice networks. Therefore, to increase the value of new Voice-over-IP solutions, interconnections with the PSTN must occur. This provides Voice-over-IP users with access to the same user base as the PSTN. Security concerns resulting from this interaction are discussed in section 9.0.

6.2.1 IPDC and SGCP

Protocols to join Voice-over-IP with the PSTN have been in development for some time. The first noteworthy methods were cooperative ventures by vendors to generate momentum. The Internet Protocol Device Control (IPDC) protocol was a venture by Level 3 Communications, 3COM, Alcatel, Ascend, Cisco, Ericsson, Lucent, Nortel, Selsius, Stratus, Tekelec, and Vertical Networks. This method mirrored the SS7 control by creating a centralized signaling scheme. Around the same time, Cisco also partnered with Telcordia (formerly Bellcore) to propose a different approach known as the Signal Gateway Control Protocol (SGCP).

6.2.2 MGCP

In April of 1999, the IETF merged IPDC and SGCP into the Media Gateway Control Protocol (MGCP) as defined in RFC-2705. MGCP is an attempt to merge the services of Voice-over-IP with the PSTN. Elements of MGCP focus on bridging the two networks. As a result, MGCP elements communicate directly with both Voice-over-IP elements as well as traditional PSTN elements.

MGCP relies on a Call Agent to perform the majority of the work integrating Voice-over-IP with the PSTN. The MGCP Call Agent controls signaling. It interacts with the H.323 Gatekeeper and the SIP Proxy Server in the Voice-over-IP realm. The Call Agent interacts directly with PSTN elements such as ground-start/loop-start in the local loop, Q.2931 in ATM, SS7 ISUP messages in the PSTN, and the ISDN D channel (Q.931). The MGCP Gateway controls voice traffic transport. It interacts with the H.323 Gateway in the Voice-over-IP realm. The Gateway interacts directly with PSTN elements (such as the conventional analog in the local loop), user cells in ATM, a DS0 in the PSTN, and the ISDN B channel.

The MGCP exchanges eight messages with a Gateway: (1) the NotificationRequest, (2) the Notify, (3) the CreateConnection, (4) the ModifyConnection, (5) the DeleteConnection, (6) the AuditEndpoint, (7) the AuditConnection, and (8) the RestartinProgress. The NotificationRequest message instructs the Gateway to await an event, and reply to the Call Agent with a Notify when the event occurs. The Call Agent sends CreateConnection, ModifyConnection, and DeleteConnection messages to manage the state of a connection. The Call Agent monitors the status of a connection with the AuditEndpoint and AuditConnection messages. The Gateway uses the RestartinProgress command to alert the Call Agent of a change in operational status of the Gateway.

Despite these features of MGCP, it lacks the wide range of support from vendors necessary to stand out as the clear choice for integration of Voice-over-IP with the PSTN. Politics has also played a role in MGCP's acceptance. As an IETF standard, many companies who support the ITU-T simply shunned the protocol.

6.2.3 MEGACO

The Media Gateway Control (MEGACO) protocol is the evolution of MGCP with coordination from both the IETF and the ITU. MEGAGO is also known by the ITU-T designation H.248. MEGACO is the latest evolution in gateway protocols.[26] Despite its youth, MEGACO holds the most promise for the immediate future due in large part to the backing of the two leading standards bodies in this technical area.

The MEGACO architecture relies on three major elements: (1) the signaling gateway (SG), (2) the Media Gateway Controller (MGC), and (3) the Media Gateway (MG). SGs interface with the SS7 STP and the MGC. To the SS7 network, an SG appears to be another STP (see Figure 15). The SG handles the MTP stack locally and passes messages to the MGC using the Simple Common Transport Protocol (SCTP). The MGC handles the control of a call. The MGC sits between SGs and MGs serving as the central intelligence point of MEGACO. MGs convert voice between telephone traffic circuits and IP network data packets. MGs connect directly to SS7 SSPs and appear as an SSP to SS7.

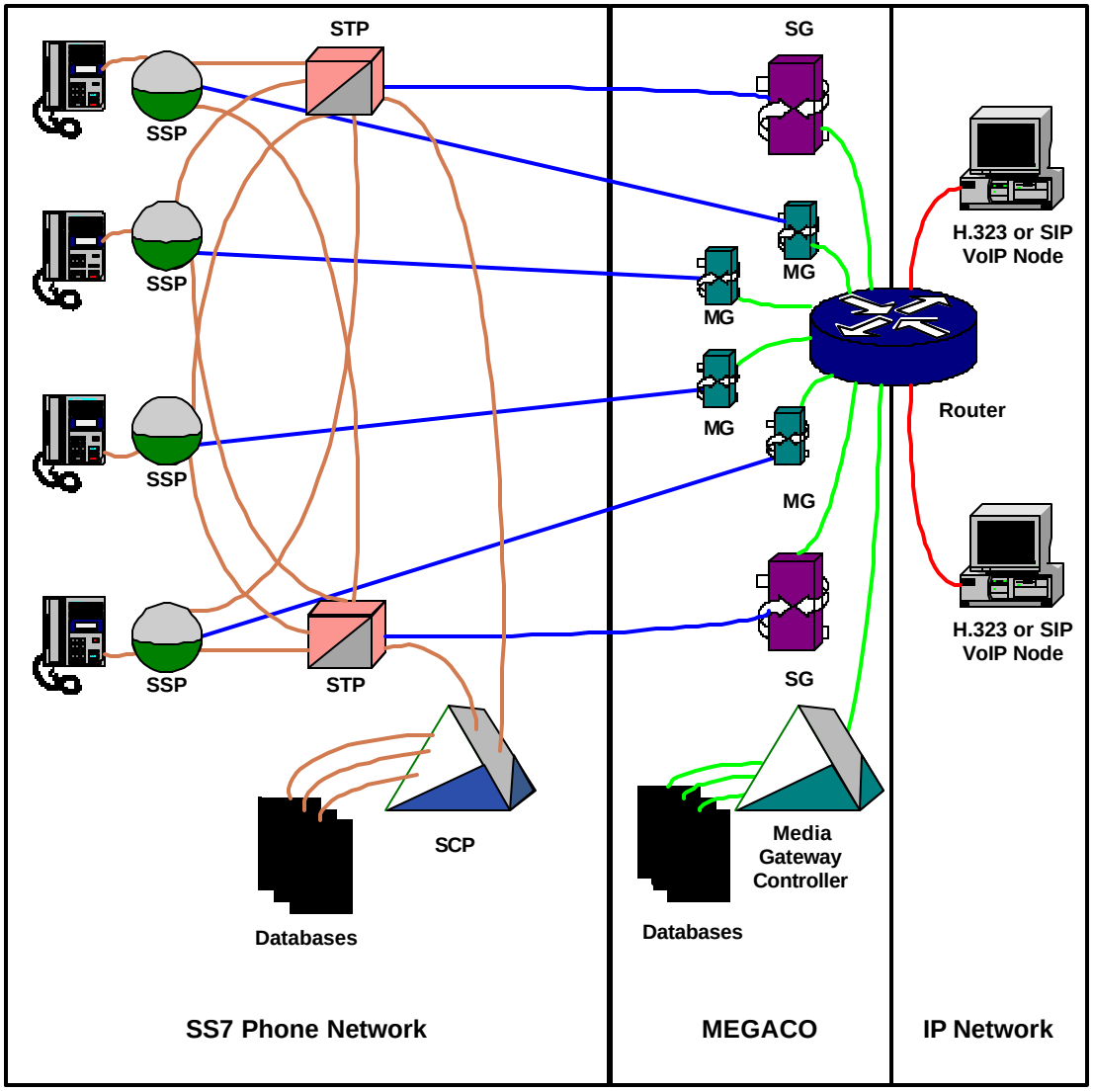

Figure 15: MEGACO Connection to the PSTN

It is important to realize that the evolution of Voice-over-IP builds upon experiences gained from earlier protocols. This is true of MEGACO as well. As was stated, MEGACO is an evolutionary step forward from MGCP, which has its roots in IPDC and SGCP. Moreover, some of the MEGACO functionality resident within the MGC can be traced back to similar functionality in the Proxy Server for SIP or the Gatekeeper for H.323. The MGC also performs functions similar to the SS7 SCP.

The major benefit of MEGACO comes from the placement of the MGs. MGs are spread throughout the network to minimize the need for SS7 links, provide fault tolerance, and convert analog voice into IP sooner.

7.0 REACHING TOLL QUALITY IN A CONVERGED NETWORK

American Telephone and Telegraph (AT&T) first defined "toll quality" telephone communications during studies conducted in the 1950's and 1960's. These studies lead to the development of the "via net loss" plan, call completion rates, audio levels, and other parameters that define "toll quality" service objectives within today's circuit-switched telephone network. AT&T documented its toll quality service objectives in the "Notes on Distance Dialing (1975)".[16] These objectives set the standards for the orderly development of direct distance dialing and global telephone communications for the next two decades.

AT&T defined the toll quality service standards within the context of an all-analog network. Bandwidth constraints, trans-circuit power loss, additive noise, and circuit crosstalk were among the serious design problems inherent to all-analog networks. Technical realities bounded toll quality service standards of the day.

Modern packet networks characterize performance in terms of packet loss, available bandwidth, and system latency. Packet loss represents data dropped or irrecoverably damaged by the network. Packet loss typically stems from data collision and buffer overflows. Fortunately for Voice-over-IP networks, voice packets have a high tolerance for truly random packet loss.

Bandwidth defines the available data speed within the network. Bandwidth is limited by physical media connection rates and network congestion. Voice traffic consumes as little as 8kHz in a data network, which is typically a small faction of the available network bandwidth. Bandwidth plays a larger role in voice performance indirectly via system latency.

System latency is the sum of the delays experience by a packet as it travels from source to destination. Although individuals may disagree on how much delay is tolerable in an audio conversation, it is generally accepted that 150 ms is the break point where the delay becomes annoying for most people.[3] Latency becomes intolerable to most people at 250 ms. Queuing introduces the most latency into a data network. Each network element performs some degree of queuing to allow the device's processor to make a routing or switching decision. Latency is closely tied to bandwidth. Queuing delays increase during periods of congestion, since the processors in the network elements must work harder to handle the increased volume. Bandwidth also impacts latency by limiting the maximum transfer unit (MTU) size for a given link. Lower speed links typically have smaller MTU sizes, since the cost of retransmitting a large packet is significant on a low speed link. Smaller MTU sizes translate to more packets per payload byte, which in turn translates into more overhead consuming the link's bandwidth and increased latency.

A special case of latency is jitter. Jitter quantifies the average variation in delay. Voice traffic cannot tolerate much jitter. Without compensation, jitter has the effect of speeding-up or slowing-down audio with annoying effects upon the audio codec and the perceived quality of the conversation.

7.1 Latency and Jitter in the PSTN

Latency and delay jitter have seldom been significant quality factors in the PSTN even during the days of analog systems. Circuit-switching injects little delay in the transmission of voice information. Delay jitter is virtually nonexistent. Delay in Circuit-switching networks has been less than twenty-five milliseconds for most conceivable voice connections. Notable exceptions occur on international calls that traverse geostationary satellites. Until the last decade, the cost of such international telephone calls and alternative communication means, such as fax and telex, relegated concerns for international call latency to a minor issue affecting a few callers. Undersea fiber optic cables eliminated the need for satellites in all but the most remote regions and eliminated the need for long delays on most international calls.

Modern digital technology and fiber optic transmission systems eliminate many analog constraints within a circuit-switched telephone network. It is much easier to achieve the toll quality in networks consisting largely of digital switching and transmission. A side benefit has been the general improvement in data transmission over circuit-switched telephone network as evidenced by the success of V.34 and V.90 modems. The transition to digital technology raised the quality of voice communications to the point that callers exceed the original toll quality standards on the majority of calls. Thus, AT&T set a standard for toll quality voice communications that today's Circuit-switching technology regularly exceeds.

7.2 Latency and Jitter in Voice-over-IP Networks

Although digital switching has improved performance in the PSTN, it does not mean Voice-over-IP networks relying on digital switching will have the same performance characteristics. Voice-over-IP networks face additional obstacles not present in the PSTN. Voice-over-IP networks must compete with data for network resources and must contend with networks designed to provide the best data performance per dollar. The PSTN network was designed to provide a specific quality of service for every voice connection.

Figure 16 shows a realistic Voice-over-IP audio path and the delays associated with each element.

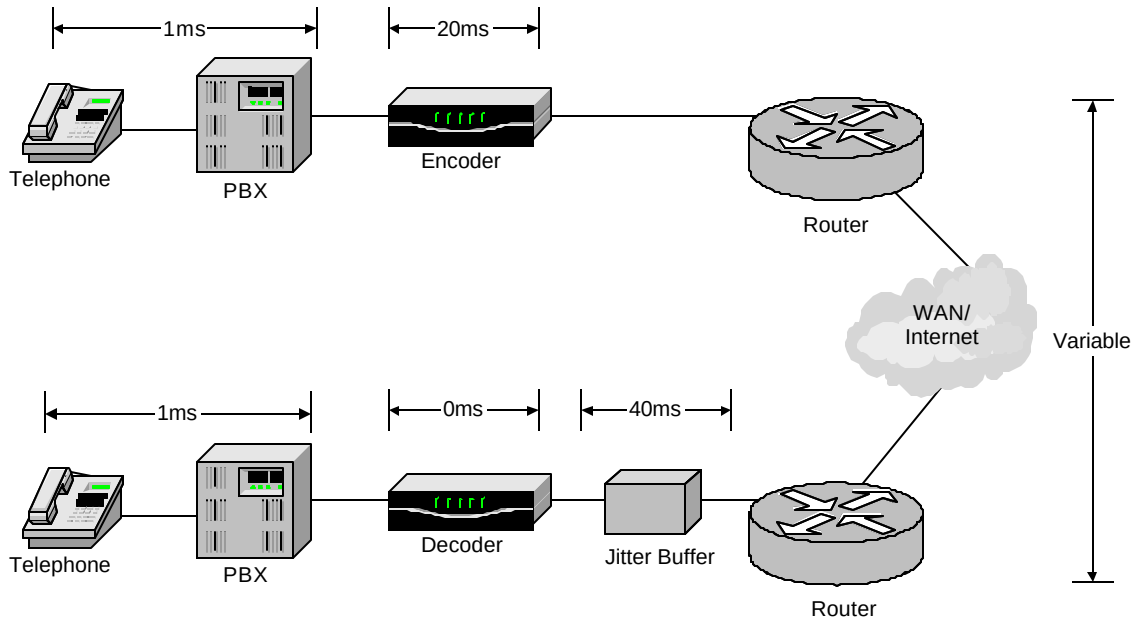

Figure 16: Typical Voice-over-IP Latency

In the figure, several elements stand out as the larger contributors to latency. At the transmitting end, the encoder introduces the most delay. G.729 and G.723 codecs take about 20 ms to generate an RTP packet.

At the receiving end, the decoding occurs quickly, but the jitter buffer introduces a large delay. Delay jitter is mitigated by adding a jitter buffer that holds the first voice packet for a fixed period of time representing the average value of the delay jitter. In so doing, the jitter buffer adds a fixed delay to the voice delivery. Since RTP packets typically are 20 ms in size, the jitter buffer in the figure allows only two packets to be buffered within 40 ms. For systems striving towards "toll quality" while competing with data traffic, this is probably a minimum jitter buffer size.

The remaining and typically the most significant delay is the WAN connection. Public WANs, such as the Internet, tend to introduce the most variability into delay (jitter). In the figure above, if the Internet delay exceeds 90 ms, the total system latency will be greater than 150 ms.

7.3 WANs: Panaceas and Plagues

In one respect, WANs are the great panacea to achieve all of the benefits of Voice-over-IP: additional features, reduced operational costs, convenience, and flexibility. Unfortunately, WANs also bring with them the plagues of excessive delay and delay variation (a.k.a. jitter). Studies show that an end-to-end delay of greater than 150 ms is excessive, and greater than 250 ms is unacceptable. For example, experience taught satellite network operators that the 250 ms delay inherent to geostationary satellites was unacceptable to satellite telephone network users. User resistance to voice latency explains why satellites were never a mainstay of commercial telephone networks.

Latency is the critical issue in the acceptance or the rejection of Voice-over-IP as a replacement for conventional circuit-switched telephony.

7.3.1 Role of Traffic Variance

The primary QoS challenges for Voice-over-IP is latency and delay jitter. Studies show that latency is a fixed delay introduced primarily by packet processing in routers and gateways.[5] Delays through ATM, SONET, DWDM, and other switching devices are negligible by comparison.

Congestion adds to the latency of the network by holding packets in queues. Variations in traffic intensity cause variations in congestion and in queue latency. The industry refers to these latency variations as "delay jitter." Jitter appears to the user as awkward pauses in the middle of a word. In the extreme, the jitter may either overrun or starve the audio decoder, leading to a loss of synchronization with further disruption to an otherwise orderly conversation.

A jitter buffer added to the Voice-over-IP connection is a common solution to delay jitter. Jitter buffers store voice packets in memory for a time equal to the average delay variation. Large delay variations equate to large jitter buffers and large built-in delays. Jitter buffers mitigate delay variation, but they compound latency problems. Consequently, it is necessary to keep delay jitter to a minimum for reasons of quality and cost.

US network infrastructure consists largely of interconnected service provider domains operating in tandem to provide the end-to-end connection. Likewise, the international network infrastructure is a set of interconnected national domains operating in tandem to provide international telephone, Internet, or other services. Figure 17 illustrates the global infrastructure composed of interconnected enterprise, US carrier, and international carrier domains.

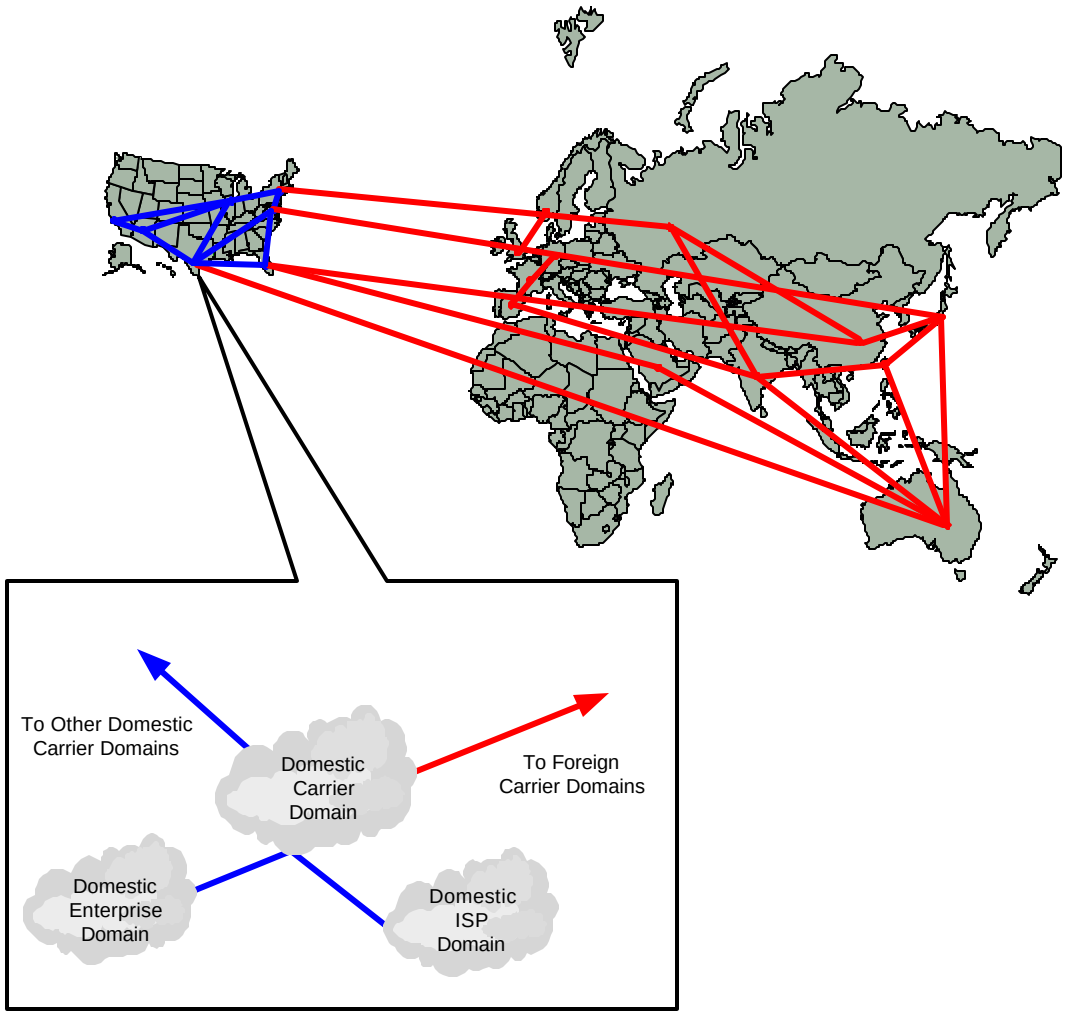

Figure 17: Multiple Domain Network Infrastructure

Each domain consists of interconnected switches, routers, and gateways operating under the supervision of a service provider (also called a "carrier" or "common carrier") or corporate enterprise. Each domain and its interconnection with other domains introduces latency and jitter depending upon each domain's design and prevailing traffic conditions. Total trans-network latency and jitter are the sums of the latencies and jitters across all domains comprising the end-to-end Voice-over-IP connection.

The additive effects of latency and jitter lead to the conclusion that close coordination must exist between the carrier and enterprise domains for quality of service objectives to be met. It is apparent from available H.323 gateway performance that only two gateways should be involved in a Voice-over-IP connection. Furthermore, the number of supporting routers cannot be excessive.

Coordination between network domains would likely avoid H.323 gateways and IP routers at the domain interfaces except where a given domain interacts directly with a

circuit-switched telephone network or a routed IP domain. Figure 18 illustrates the concept of core and edge networks within a multiple service domain.

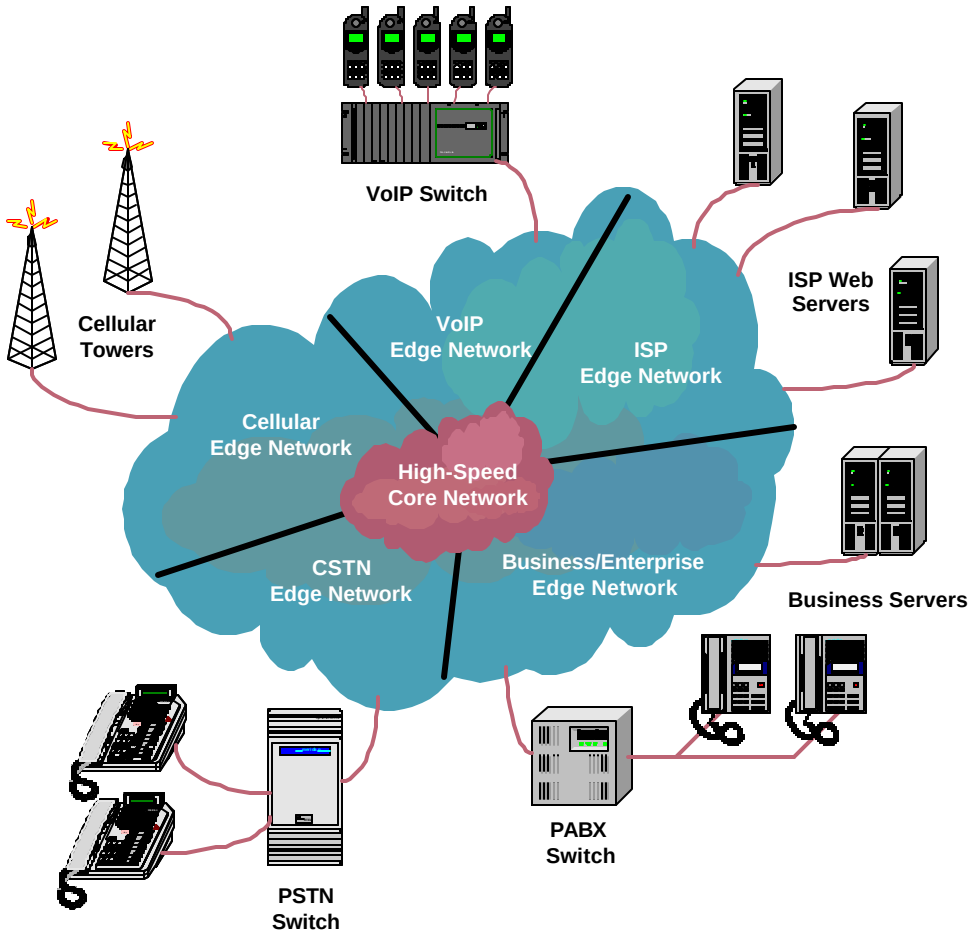

Figure 18: Core and Edge Networks in Multiple Device Environment

7.3.2 Impact of Link Failures

Network link failures alter the network topology and lower the bandwidth available for connections through portions of the network. Consequently, congestion, delay, and jitter increase following loss of a portion of the network unless proactive countermeasures are taken.

Restoration of failed components and recovery of disrupted voice connections involve considerations that are not common to IP networks. Data networks recover from failures by retransmitting lost data packets and routing data around the failed component. Most data applications are insensitive to occasional delays. Therefore, the time required to

recover data traffic can be relatively long. Redirecting traffic around failures may add delay to the data network session depending upon the design of the data network.

Client-server session connections and other types of data transfer sessions depend upon IP sockets as a part of the session control. During network recovery, software managing a given IP socket may detect excessive delay, time out, and terminate the data session. Recovery from failed IP sockets requires a time consuming recovery process. Retransmission, rerouting, and IP socket/session reestablishment are time-proven recovery measures in data networks. Delays associated with session recovery are tolerable within the Internet and other data-only networks.

Delays associated with conventional data network restoration are not acceptable within an IP network supporting Voice-over-IP. For example, retransmission of lost Voice-over-IP packets is an ineffective recovery technique. Retransmission introduces excessive delay and adds to the loss of voice quality created by the failure. Voice-over-IP networks must ignore missing packets. The receiving Voice-over-IP decoder must detect the information loss and substitute "quiet" information, "white noise" information, or other measures designed to reduce listener confusion.

Decisions regarding redirection of traffic along alternate paths must pay careful attention to latency and delay jitter. Redirected traffic may increase congestion above design limits along alternate paths leading to quality degradation on a broader scale. Redirection should not increase the number of gateways or add large numbers of routers to the restored Voice-over-IP connections. Otherwise, the redirection will reduce the Voice-over-IP QoS below toll quality.

Fast path restoration within the high-speed core network is essential. Switching transitions of a few milliseconds will be noticed by the receiver codec and by the listener. Again, the concept of fast core networks like that proposed by the DIFF-SERV model is appropriate for a national or international Voice-over-IP network.

7.4 QoS in the WAN via Prioritization

Prioritization of traffic is a compromise between infinite bandwidth, packet processing, and cost and is performed primarily for the purpose of eliminating delay jitter.

Ideally, unlimited application of bandwidth would eliminate jitter by ensuring that all network connections, routers, and switches have sufficient bandwidth to support the most intense traffic loads. Bandwidth is more available and less expensive than ever before due to the many innovations in fiber optics, DWDM, and high speed electronics. Even so, infinite bandwidth equates to infinite cost, neither of which is practical within commercial networks. Furthermore, limited foresight and uneven application of network management techniques by domain operators reduces traffic predictions and applications of network bandwidth to an error prone endeavor. Consequently, congestion will occur from time to time within portions of the network with attendant increases in delay jitter.

Prioritization of IP packets is a computationally intensive task. Processing must be fast and packet queuing kept to a minimum if low latency and jitter are to be achieved. Otherwise, the prioritization process adds to the latency rather than mitigating it. Again, the DiffServ QoS model makes the greatest sense from a performance standpoint by

relegating the prioritization process to the network periphery where traffic volumes are relatively low and away from the network core where traffic aggregates into dense, high speed flows.

Priorities established at the network edge are useful throughout the network for switching and restoration decisions. Therefore, priorities computed in one domain should travel across the domains traversed by a given connection. Each domain should act upon the priorities according to standards set by domain operators. Conveying priority information across the network and acting upon the information in a priori manner creates a reliable framework for providing toll quality service and reliable service restoration. The framework could show preference for value-added services, premium services, and especially for services supporting NS/EP efforts during disasters and national incidents.

Several methods are available to mitigate WAN delays. Unfortunately, most require a significant portion of the network to support the selected method for the method to have any real impact on latency. Therefore, the widespread use of recognized protocol standards and standard operating procedures amongst the domains comprising the Voice-over-IP infrastructure is an essential requisite for achieving reliable toll quality and rapid service restoration. Lessons learned from years of successful telephony teach the importance of cooperation amongst the domains along lines defined in standard committees, bilateral operating agreements, and international treaties. It is within these forums that agreements supporting prioritization of traffic, restoration priorities, and other practices supportive of NS/EP agencies are likely to develop. NS/EP agencies should foster forums for Voice-over-IP standards and practices. It is within these forums that the NS/EP agencies will most likely see their special needs met and ensure the integrity of the Voice-over-IP infrastructure during times of crisis. The following sections point to some of the standards that will play a primary role in achieving toll quality service within the Voice-over-IP infrastructure.

7.4.1 IEEE 802.1p

IEEE 802.1p provides a mechanism for prioritizing Ethernet frames. IEEE 802.1p relies on the IEEE 802.1Q frame structure. IEEE 802.1Q defines an extension to the standard Ethernet frame by inserting a Tag Protocol Identifier (TPI) and TAG field into the Ethernet header between the Source Address (SA) field and the Type/Length field. The TPI field is used to signify the presence of the IEEE 802.1Q extension by containing a value outside the possible range of Ethernet type or length entries. In IEEE 802.1Q, the TAG field is used to establish VLANs. IEEE 802.1p takes the TAG field, and uses it instead to provide three bits of user priority. To take advantage of this prioritizaiton, network devices must use Ethernet and be capable of processing IEEE 802.1Q Ethernet frames.

7.4.2 Integrated Services (INT-SERV)

Integrated Services (INT-SERV) provides a method to reduce variability in delay and reserving bandwidth in a network to support real-time protocols. The philosophy behind INT-SERV is simple: guarantee a service level for delay sensitive applications, and

manage loads on network components to provide "enhanced best-effort" performance. INT-SERV is defined by the IETF in RFC-1633.

INT-SERV has four components: flow specification, the signaling protocol, admission control, and packet management (classification and scheduling).[5] INT-SERV signaling relies on RSVP to make service class reservations. RFC-2211 defines two classes of service in addition to the standard "best effort" service offered by packet networks. The *Guaranteed Service* class of service attempts to create a virtual circuit. The *Controlled Load* class of service attempts to control the delay experience by packets.

RSVP packets travel through the network from source to destination (see Figure 19). At each router along the path, RSVP requests a service class. When the destination receives the packet, it returns the packet to the source. This allows the source to know the minimum service parameters accepted by all routers in the path. Routers use admission control to determine whether to fully, partially, or not reserve the requested bandwidth. Admission control is largely based on existing commitments and available bandwidth.

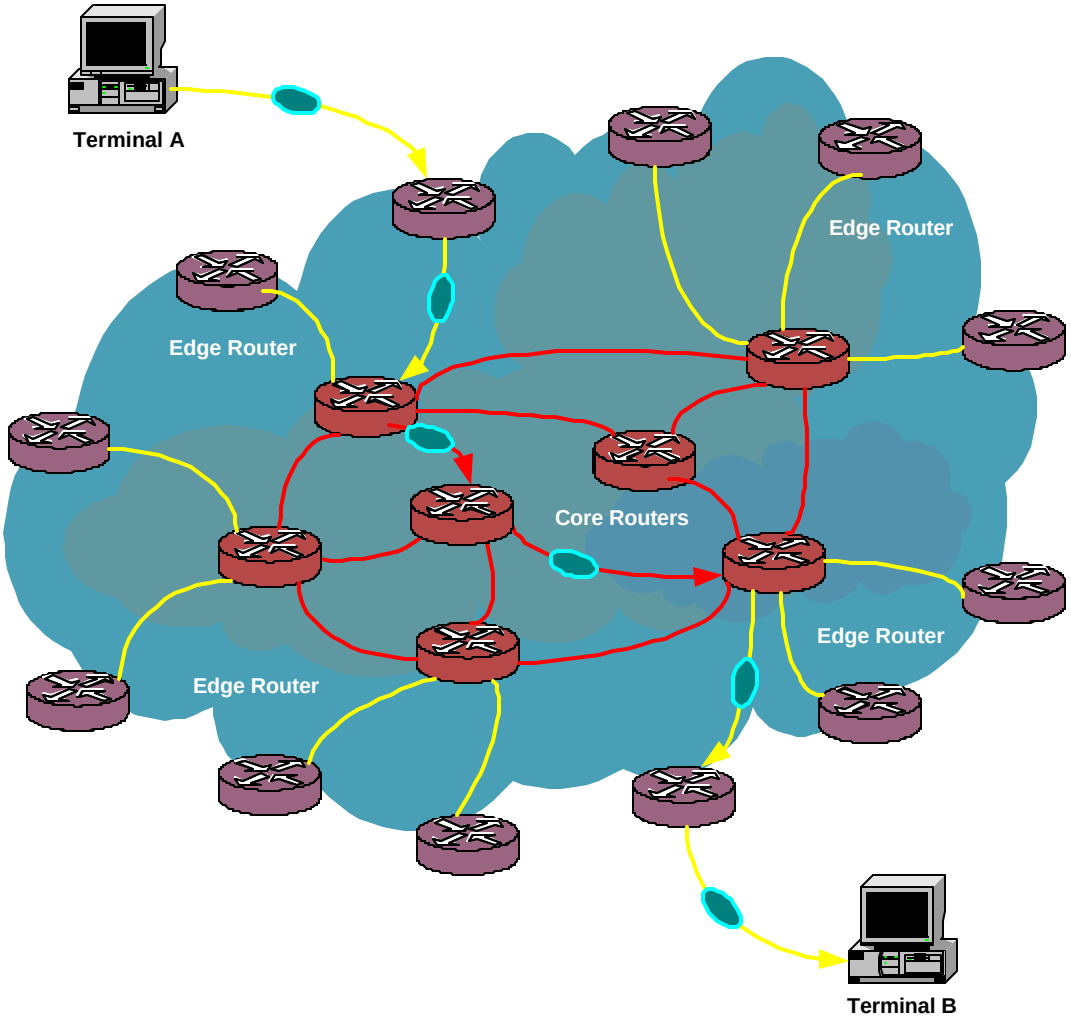

Figure 19: INT-SERV Architecture – RSVP Packet Flow

The path defined by RSVP is known as the flow specification. The flow specification is detailed in the *Flowspec*. The *Flowspec* consists of two parameter sets: the Reserve Spec (Rspec) and the Traffic Spec (Tspec). For a *Guaranteed Service* class, the Rspec defines a maximum allowed packet loss. For a *Controlled Load* class, the Rspec defines a maximum allowed packet delay. Both classes of service specify the Rspec parameters of *sustained rate*, *smallest packet size*, and *largest packet size.*

All data conforming to the RSVP *Flowspec* travels the same flow specification from source to destination. This is accomplished by routers performing packet management. Routers classify received packets to determine if the packet matches a flow specification. Packets matching an existing classification are then scheduled to continue along the flow specification with the associated queue prioritization.

INT-SERV suffers several real-world deficiencies. INT-SERV requires all devices in the path to support RSVP. This is possible to achieve in a LAN, but difficult to achieve in a public WAN. RSVP is a half-duplex protocol. This requires voice conversations, which are typically bi-directional, to create a flow specification for each direction of traffic. RSVP is initiated by the endpoint, which enables the endpoint devices and their users to exploit the admission control process. This has the potential to create an admission control, since without a policing measure, all endpoints could create flow specifications. INT-SERV also requires a significant processing effort at each network element in the flow specification to perform the classification and scheduling. INT-SERV deficiencies lead to the generally held view that INT-SERV is not scalable into large Voice-over-IP networks.[21]

7.4.3 Differentiated Services (DIFF-SERV)

Differentiated Services (DIFF-SERV) uses the IPv4 Type-of-Service (TOS) field (one byte wide) to provide expedited packet delivery. DIFF-SERV is defined in the IETF's RFC-2475. In IPv6, the traffic class byte is used. DIFF-SERV requires edge routers to classify traffic, mark the traffic for rapid routing decisions by core routers, and police connections to ensure they do not violate the user's service level agreement (SLA).

DIFF-SERV achieves scalability within large networks by forcing all complexity out of the network edge devices where there are fewer flows and smaller traffic volumes. In so doing, DIFF-SERV streamlines the core network for maximum traffic-carrying capacity and avoids changes in the core network stemming from new service offerings. Core routers are only responsible for checking packet precedence, enabling their resources to be directed at high-speed switching.[18] Service differentiation, traffic admission, traffic shaping, and other computationally intense functions occur at the network edge. This leads naturally to the use of high-speed protocols within the core network including ATM, SONET, and DWDM. It also facilitates the eventual implementation of photonic switching.

DIFF-SERV marks the TOS field based one of two schemes: quantitative DIFF-SERV or priority-based DIFF-SERV.[5] In quantitative DIFF-SERV, a deterministic or statistical measure of throughput, delay, jitter, and loss are calculated to provide a quality of service to the data flow. In priority-based DIFF-SERV, data flows are ranked on a relative priority scale.

7.4.4 COPS

The Common Open Policy Service (COPS) provides a means for regulating traffic at the network edge to control entry into the core (see Figure 20). COPS defines a Policy Enforcement Point (PEP) that communicates with the Policy Decision Point (PDP). The PDP resides on a Policy Server with persistent connections to the PEPs. PEPs cohabitate edge routers. When an INT-SERV or DIFF-SERV packet attempts to enter the core network through an edge router, the PEP consults the PDP for a policy decision. The PDP can accept, modify, or reject the SLA of the packet. Since the PDP is in constant communication with all of the PEPs at the network edge, the PDP maintains the network state and can adjust individual PEP policies based on its knowledge of the current

condition of the network core. COPS essentially provides a centralized means of intelligently enforcing SLAs for packets entering the core.

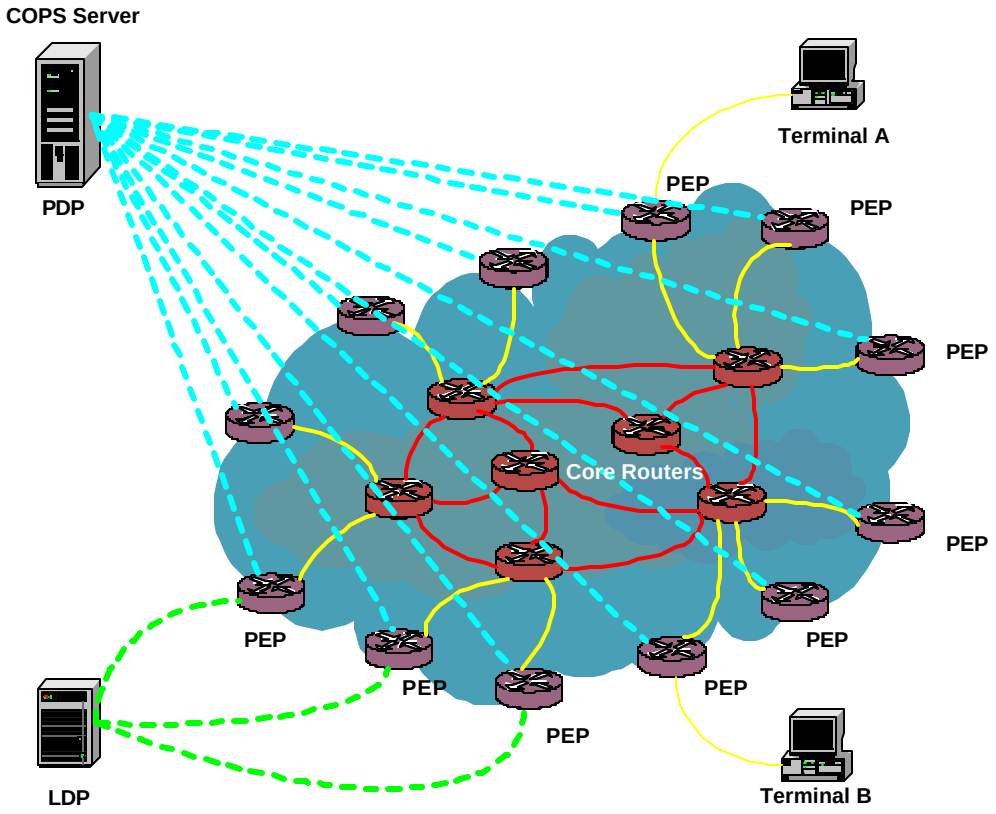

Figure 20: COPS System Architecture

7.4.5 MPLS

Multiprotocol Label Switching (MPLS) provides a generic method for increasing switching speeds while offering a class of service. MPLS works at the edge of a network to increase switching speeds at the core in a manner similar to DIFF-SERV. MPLS labels packets by prepending an MPLS header. The label is protocol-independent and could be prepended to packets, frames, or cells for communications between devices that support MPLS. The major driver behind MPLS is to enable the switch to perform "cut-through" switching. "Cut-through" switching occurs when the device begins switching the packet before the complete packet has been received. Typically, switches have to process most of the header before a switching decision can be made. With MPLS, the

information needed to perform a switch is prepended at the front of the packet to allow switching to occur after only a few bits have been received.

7.4.6 ATM

Asynchronous Transfer Mode (ATM) offers native support for QoS provisioning. By establishing CBR, rt-VBR, or nrt-VBR virtual circuits, latency and jitter can be reduced. ATM is well suited for the transmission of real-time traffic such as voice.

7.4.7 Real-World Latency Mitigation

As vendors attempt to create solutions to provide SLAs to voice, hybrid solutions involving INT-SERV and DIFF-SERV are developed focusing on either carrier class or enterprise class networks.[5]

Cisco's IP Telephony targets enterprise class networks. It manages individual flows along the edge of the network by packet classification. As with DIFF-SERV, the Cisco solution uses the IPv4 ToS field to create traffic classes with priority given to voice over data. Core routers use INT-SERV (RSVP) and DIFF-SERV (priority queuing) to provide high-speed switching and congestion avoidance. Weighted random early detection (WRED) is also used in the core for congestion avoidance, dropping data packets if necessary to maintain the SLA for voice.

Lucent's Gateway approach targets service provider networks using high capacity switches or routers within the network core. Lucent employs traffic aggregation along the network edge to reduce computational requirements within the carrier's high speed core network. This solution relies on the IP network to properly set the DIFF-SERV IPv4 ToS field. Edge routers then regulate flow into the core network based on the ToS field.

7.5 Packet Loss Across Multiple Network Domains

Operating packetized data and voice services over a common IP network is full of contradictions and contrast. Earlier discussion highlighted differences between data and voice regarding delay with the later service being quite sensitive to delay and delay variation. On the other hand, data services are insensitive to both impairments. Now compare the delay relationships with the relationship packetized data and voice have regarding packet loss.

For its part, most data applications are extremely sensitive to lost packets. Great lengths are taken within data networks to detect, correct, or retransmit corrupted or lost information. After all, a financial statement, report, or electronic funds transfer is of little value if data is missing or corrupted.

Packetized voice, on the other hand, is insensitive to packet loss.[3] Studies show that up to 5% of all voice packets may be lost without a significant loss in voice quality. Voice-over-IP codecs and the ever-adaptable human ear easily mask the occasional loss of information. The high tolerance to packet loss within Voice-over-IP services does not imply room for complacency amongst Voice-over-IP service providers. Packet loss can be overcome with silence substitution, noise substitution, packet repetition, packet interpolation, and frame interleaving.[3]

Packet loss is fundamentally an independent random process occurring within each network domain. The probability of successful packet delivery is a time varying function of domain congestion. For instance, a domain experiencing a 3% packet loss has a probability of packet delivery equal to 97% so long as traffic conditions remain constant. Probability of packet delivery will differ between domains due to differences in traffic intensity and traffic engineering decisions implemented within each domain.

Voice-over-IP packets may traverse several domains on their way from the transmitter to the receiver. It is necessary to think in terms of the probability of end-to-end packet delivery within a Voice-over-IP connection. Each domain has its own, statistically independent packet loss process and a time varying probability of packet delivery. The probability of end-to-end or "connection" packet delivery is the product of the packet delivery probabilities for the domains through which the packets must pass. For example, consider a Voice-over-IP conversation that traverses four network domains with independent packet delivery probabilities of 97%, 98%, 98.5%, and 99%. The probability of a Voice-over-IP packet arriving at the receiver is 92.6%, which is well below the acceptable threshold. Clearly, it is not sufficient for the domains in the example to operate their networks at a 95% probability of packet delivery.

The number of domains traversed in a complex, multiple domain network (like the US network) is also a random variable. A Voice-over-IP connection may traverse relatively few domains during low traffic periods while traversing several additional domains during heavy traffic periods. The number of traversed domains may increase significantly following a failure or NS/EP event within the national network.

Suppose that a community of international domain operators decide that they wish to achieve toll quality Voice-over-IP connections with as many as eight domains comprising the connection. Eight domains may seem like a large number, but consider an international Voice-over-IP call originating in San Antonio, Texas, and terminating at a cellular phone in Kyoto, Japan. The hypothetical call might originate within a San Antonio business domain and traverse the following list of domains on its way to the cellular telephone receiver in Kyoto:

Domain number	From Domain ?	? To Domain
1.	San Antonio enterprise domain	ISP domain
2.	ISP domain	US domestic carrier domain
3.	US domestic carrier domain	US international carrier domain
4.	US international carrier domain	Japanese international carrier domain
5.	Japanese international carrier domain	Japanese domestic carrier domain
6.	Japanese domestic carrier domain	Japanese cellular carrier domain
7.	Kyoto cellular carrier domain	Kyoto cellular telephone

Table 1: Multi-Domain Call

Our hypothetical call traversed seven domains without transiting intermediate countries or intermediate carriers. It is easy to see that the number of domains constituting a Voice-over-IP connection could easily reach eight or more within a complex set of interconnected domains. If 95% packet delivery across eight domains is the hypothetical

design goal, then each domain must maintain a packet delivery probability of at least 99.75%.

Voice-over-IP domains should cooperatively consider traffic engineering questions with a common goal of achieving better than 95% probability of end-to-end packet delivery. The allowable probability of packet delivery within each domain and the number of domains allowable within a connection must be considered in combination. Having agreed to traffic engineering standards, each domain should then apply capacity and switching within their respective domains in a manner supporting the common goal of toll quality voice.

Traffic engineering is not a precise process. The number of paths and combinations of a switching system are hard to conceptualize, as the domain network becomes large. It is even harder to imagine the proper combinations of capacity, switching, and switching instructions when a portion of the network fails or is destroyed by a disaster, terrorism, or other NS/EP event.

Domain operators may choose to employ priority schemes when traffic engineering measures are not sufficient to maintain toll quality connections under all circumstances. Prioritization of traffic according to the importance of the Voice-over-IP connection would augment traffic engineering with a degree of network intelligence. Domain operators could agree to priorities for different classes of Voice-over-IP connections. Priorities might show preference to different classes of subscribers and their calls based upon premium service options versus economy calling. Certain value-added services might be willing to pay extra for the surety that their calls were of high quality even during times of heavy congestion. Calls from disaster relief, public safety, military, and national security organizations could be given preference above all other types of calls during times of national crisis or natural disaster. Priority information shared between domains could improve connection quality and, at the same time, provide the necessary information for sophisticated fault restoration.

8.0 REACHING RELIABILITY IN A CONVERGED NETWORK

8.1 Multiple-Domain Network Fault Tolerance

Domain operators must exercise good network planning and management to achieve toll quality QoS, especially after a network component failure. The operators should have proactive fault detection and restoration measures in place prior to a crisis. Restoration measures should include alternative path switching, fault tolerant systems (e.g., SONET rings, cable operating along physically separate right-of-ways), and restoration bandwidth held in reserve for failures.

Domain operators should coordinate their failure countermeasures for maximum effectiveness. Specifically, domain interconnections must have the same considerations for fault tolerance, fault restoration, and restoration capacity as found within each of the domain. Otherwise, failures at the domain interfaces will lead to substandard voice quality, disconnected calls, and failed call attempts. Critical domain interfaces should be duplicated and physically separate with provisions for automatic restoration through the alternate interface.

8.2 Multiple-Domain Network Restoration

Trends in network architectures and Voice-over-IP development are heavily influencing the evolution of competitive, multiple-domain network architectures. Network architectures are evolving toward schemes where a general-purpose core network acts as a high-speed, highly reliable switching fabric supporting many applications resident along the periphery of the network.

Core networks currently consist of ATM, SONET, and DWDM layers arranged in a service pyramid as shown in Figure 21.[6] Each layer aggregates traffic from the higher layers and adapts the aggregated traffic to the transmission facilities provided by the lower layer.

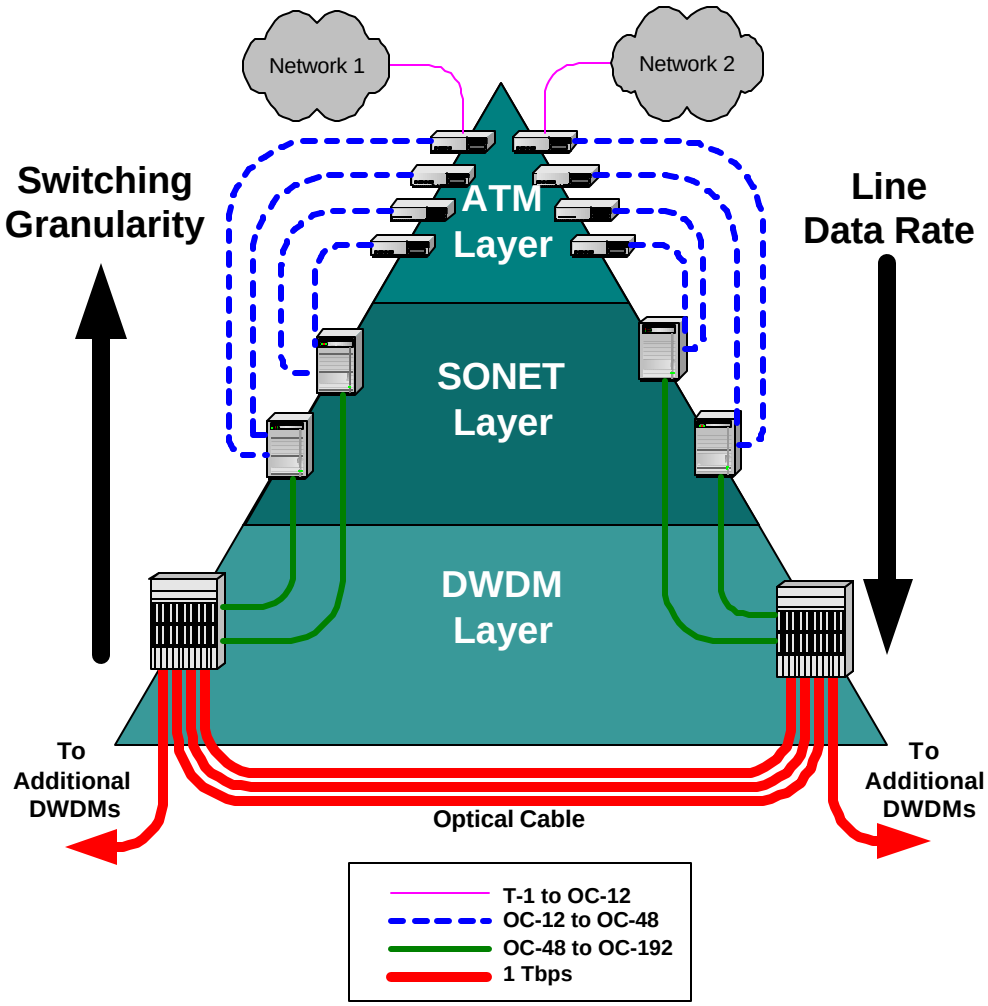

Figure 21: Current Network Pyramid

The exact configuration of the service varies from carrier domain to carrier domain. In some domains, the ATM layer may be omitted in deference to SONET TDM muliplexing. Other domains do not require DWDM. Instead, these domains rely entirely on ATM switching and SONET multiplexing services. However, SONET is inadequate for aggregation and restoration in the highest density core networks where ATM switches regularly use OC-48 interswitch trunks. Thus, the innermost core of the network may consist exclusively of ATM and DWDM layers. The combinations comprising the core network will vary from domain to domain, but it is clear that the core network is and will be fast, general purpose, and employ as little computation as absolutely necessary to perform switching, multiplexing, traffic prioritization, and restoration.

In contrast, future domain edge networks will be application-specific with computationally intensive algorithms providing value-added services. Traffic in the network periphery has low relative intensity. Edge devices can perform complex functions like speech compression, traffic prioritization, and security enforcement. Domains frequently involve several edge networks, each offering a different service (e.g., Internet, cellular, PSTN, Voice-over-IP).

The number of interconnections within and between domains will increase as the number and types of services increase. Domain operators interconnect their edge networks to their core networks for high speed transmission between regions. Domains will exchange traffic either by interconnecting their edge networks or by interconnecting their core networks.

Figure 18 illustrates a fully meshed interconnection of domain core networks and edge networks. Edge networks will exchange traffic either through direct connection or indirectly through connections between domain core networks. Connections between domain cores will exchange high volumes of traffic without knowledge of the traffic's application. Instead, priorities and QoS requirements specified by the originating edge network will propagate into the core network as switching and restoration instructions. The core networks will act on the priorities set by the edge networks during switching decisions and when core network faults lead to restoration of core transmission paths.

Decisions regarding restoration and prioritization will exist in different forms and for different purposes, both within the core and the edge networks. The core networks will restore high speed, general purpose transmission paths without direct knowledge of the application traffic carried on the restored path. In contrast, the edge networks will make detailed decisions regarding restoration and prioritization of a specific set of services. For example, voice prioritization, restoration, and latency reduction algorithms will reside in the Voice-over-IP and Internet edge networks. Special considerations for NS/EP traffic will primarily reside within the edge networks.

The need for restoration methods exists in each of the domain network layers.[17] The recommendations for restoration in each of the layers apply equally well to core networks and edge networks. In other words, domain operators should implement restoration methods within their application specific edge networks and within their general purpose core networks.

Restoration between domains is equally important. As mentioned earlier, failure of domain interconnections may isolate regions or block traffic bound from one domain to another. Coordination of restoration information within and across the domains is necessary for the restoration process to work effectively during equipment failures. Coordinated restoration is even more important when natural disasters, terrorism, or other NS/EP events disable many network components simultaneously.

Coordinated restoration requires that domain operators adopt recognized priority assignment, fault reporting, and restoration methods. These methods will generate switching information used for alternate path assignment, connection establishment, and other traffic management functions. The methods should consist of standard restoration protocols that convey necessary switching information throughout the nation's infrastructure.

9.0 ACHIEVING SECURITY IN A CONVERGED NETWORK

Each PSTN carrier operates an SS7 network within its domain. Interconnections between SS7 domains provide end-to-end signaling for voice traffic crossing a multi-carrier network. PSTN carriers closely guard the internal operation of their respective SS7 networks with access limited to authorized personnel. Security measures are relatively easy to enforce when compared to the wide-open access often found in the Internet. Until recently, there has been little concern that an attacker might infiltrate a carrier's SS7 domain given the level of most carriers' physical security measures.

9.1 Changes in Signaling System 7 Security

Changes brought about by deregulation of the telecommunications industry are opening the otherwise closed nature of the SS7 network. Under FCC cohabitation rules, PSTN carriers must lease space to other service providers within central offices, toll offices, international telephone gateways, undersea cable heads, and other circuit-switched network facilities. ISPs, competitive access providers (CAPs), competing interexchange carriers (IXCs), and other third parties may install and service equipment within PSTN facilities containing STPs, SSPs, and SS7 links. Cohabitation creates an opening for a would-be hacker to enter a SS7 facility and gain access to the network by masquerading as one of the facility's tenants.

PSTN operators and the tenants of their facilities should work together to screen personnel and enforce security within cohabited SS7 facilities. Otherwise, the physical defense of the SS7 network may be compromised. If so, then an attacker may disable or commandeer SS7 equipment with potentially serious impact to a critical part of the nation's communication infrastructure.

The original designers of the SS7 architecture may not have foreseen a rapidly changing, highly competitive telecommunications marketplace. The early SS7 network operators were RBOCs, AT&T Long Lines, foreign public telephone and telegraphs (PTTs), and other trusted members of a telecommunications fraternity that span the over 150 years of telegraph and telephony. Prospects of terrorism, hacking, and network exploitation were remote within this highly cooperative and trusted group.

Deregulation, competition, the rapid growth of wireless service providers (WSPs), the explosion of ISPs, and the advent of Voice-over-IP service providers create a whole new set of SS7 operators. Many of these new operators, their SS7 systems, and their staff lack the track records and experience of the long-standing service providers. There is reason to question whether the original SS7 security measures are sufficient within this rapidly changing communication environment.

For example, much thought was originally given to fault tolerance within the SS7 architecture. Redundant STPs and fully meshed linkages between STPs and SSPs make a properly configured SS7 network impervious to individual link or STP failure. Only the destruction of several directly related SS7 devices would lead to major degradation in

network signaling performance. Natural disasters and other NS/EP events could render a portion of the signaling network inoperative. Even so, the balance of the SS7 network will operate normally. Deliberate destruction of an SS7 network would require a coordinated attack at several points for significant impact. Otherwise, the SS7 network will route around destroyed elements, keeping the signaling functions in tact. However, the SS7 network is not impervious to attack or failure.

Consider for a moment a denial of service attack aimed at one or more STP. Two possible attack scenarios are worthy of concern: (1) flooding the network with valid SS7 messages or (2) directing malformed or misleading SS7 messages at STPs or SSPs. Note that the links between the STPs and SSPs are low-speed, 56Kbps connections. In the first scenario, a relatively small number of signaling messages can saturate the signaling links leading to SS7 network congestion and failure. A flooding attack would direct a steady stream of SS7 messages at several, strategically chosen STPs, thereby subdividing the network.

The second attack scenario requires the attacker to have greater knowledge of the SS7 network. Under the new competitive paradigm, there are many more people involved in SS7 operation. As knowledge of SS7 equipment becomes more widespread, then it becomes more accessible to a would-be attacker. Hacker web sites may soon describe known SCP, STP, and SSP vulnerabilities much as they now publish known vulnerabilities within Internet routers and servers.

A well-positioned attacker armed with prior knowledge of SS7 software vulnerabilities could direct malformed messages at chosen SS7 devices. An attacker might craft SS7 messages to induce improper switching states within the SSPs. Maliciously crafted messages might force the STP or SSP software into known but uncorrected malfunctions.[13] Exploitation of known SS7 software vulnerabilities is analogous to the many successful attacks leveled against known Internet router and server vulnerabilities.

Either scenario renders a portion of the SS7 network inoperative. An associated portion of the national telephone network would be isolated or even useless. Hacking within the circuit-switched telephone infrastructure seems increasingly likely given the state of SS7 security and the rapid evolution of SS7-dependent services.

It is clear that new domain interconnections such as MEGACO provide windows through which an attack can originate. Therefore, the existing and new domain operators should take cooperative measures to certify that their domain interconnections can be trusted to protect the SS7 infrastructure. Lacking sufficient industrial cooperation, the FCC or other government agencies should mandate certification of domain interconnections and require security processes within each of the domains. These cooperative or regulated measures should be designed with emphasis on SS7 network protection. However, preventive measures cannot totally eliminate the possibility of a denial of service event.

For example, PSTN carriers currently employ firewalls within their SS7 network to avoid denial of service attacks. The firewalls appear at the SS7 domain interfaces and filter malformed or improperly directed messages before they enter the domain's SS7 network. As such, current SS7 firewalls are largely a defensive measure.

It may be sufficient for an attacker to flood a given firewall and its domain interface. A flooding attack against the domain firewall would deny access from the rest of the network to the domain. Current firewall technology has only limited defenses against a flooding attack directed at the firewall.

New firewall protocol innovation could increase the effectiveness of SS7 security by mounting countermeasures against the source of the attack. For example, SS7 firewalls could detect extraordinarily high volumes of SS7 traffic and signal other firewalls of the impending attack. Strategically located firewalls armed with the appropriate information could block undesirable traffic at its source. Protocol concepts proposed by Smith and Bhattacharya for Internet firewalls could apply equally well to the SS7 firewalls.[14] See Figure 22 for an illustration of coordinated firewall countermeasures. Therefore, carriers should consider creating SS7 firewall protocols that signal other firewalls of detected attacks. Through active detection, coordination, and countermeasures, cooperating firewalls could block undesirable SS7 traffic before it renders portions of the telephone network inoperative.

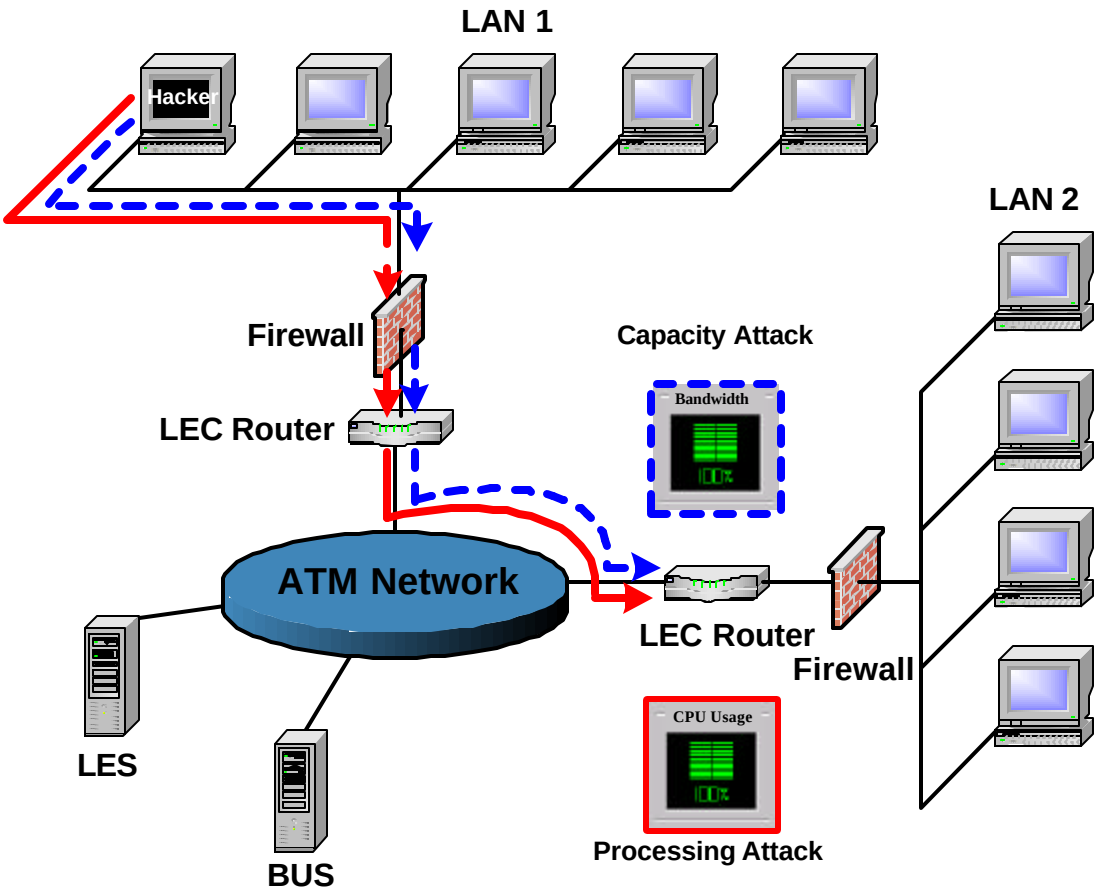

Figure 22: Vulnerable ATM LANE Network

9.1.1 IP Network Gateways

IP Network Gateways further complicate security concerns by merging the security vulnerabilities of IP Networks and the PSTN. Gateways enable a security compromise of

one system to lead to compromise of the other. The architecture for MEGACO calls MEGAGO elements (such as the SG) to masquerade as SS7 elements (such as the STP). This provides simplicity in implementation, since the existing and well-established SS7 system does not require modifications to support MEGACO. Unfortunately, this also poses a significant security risk.

Unlike SS7, MEGACO signaling control occurs on the same network as all other traffic. This opens up MEGACO elements to common network security vulnerabilities as discussed in section 5.5.2. MEGACO elements must be sufficiently hardened against intrusion. Otherwise, MEGACO elements could be used as a back door into the SS7 network. Since SS7 treats the SG as an STP, a compromised SG could send messages into the SS7 network causing switches to incorrectly create circuit connections. Without restrictive measures in the SS7 network, the MEGACO SG could therefore create a DoS condition in the circuit-switched voice network by opening as many connections as possible. Problems of this nature can be remedied by limiting the number of simultaneous connections allowed by a MEGACO SG.

Gateways also increase the privacy concerns mentioned in section 5.5.2. A compromised MEGACO SG could query or modify information controlled by the SCP. This intrusion could be used to access end-user billing information, providing personal information such as home address or the credit card number used for automatic bill payment. These intrusions could also create fraudulent billing statements or hide toll conversations from the billing system. Voice carriers already have security measures in place to protect this information from other voice carriers; however, the strength of their security measures against the resourceful internet hacking community has not yet been determined. As more gateways are deployed, more avenues are available to intruders for finding chinks in the armor of the PSTN.

Interactions between networks that create security risks should be the special concern of NS/EP agencies. The networks are a critical part of the nation's emergency response infrastructure during times of crisis. Hackers, terrorists, and saboteurs may use the changes in the national voice communication infrastructure as an avenue for carrying out their destructive agendas. Steps should be taken to simultaneously shore up network security, prepare for disaster through restoration measures, and dispense critical resources to emergency response agencies in times of crisis. Interactions between circuit-switched and packet-switched networks are creating new vulnerabilities within the national infrastructure. At the same time, new innovations within the nation's communication infrastructure offer opportunities to implement needed NS/EP capabilities that were not achievable within the inflexible PSTN architecture.

10.0 CONCLUSIONS

Voice-over-IP presents many new opportunities to reduce network infrastructure costs while increasing the availability of value-added voice services. Key among the differences between SS7 and Voice-over-IP is the networks used for signaling and voice transport. The separation of the signaling and transport networks in SS7 provides some natural avenues to provide quality, reliability, and security. All three of these areas pose a challenge to Voice-over-IP networks that has not yet been fully resolved. Many mechanisms exist that are designed to impose packet prioritization in a traditionally "best-effort" network such as the Internet. These prioritization schemes all suffer from limited reach within the network, only offering prioritization among like-configured devices. NS/EP needs for prioritized packet delivery suffer the same problem as Voice-over-IP and will benefit from solutions implemented to bring prioritization to Voice-over-IP.

Network reliability plays a central role in the acceptance of Voice-over-IP as an alternative to the PSTN. Again, NS/EP needs for reliable packet delivery during times of national crisis will benefit from the improvements in the national data networks necessary to support Voice-over-IP.

Network security will soon emerge as a major concern for Voice-over-IP networks. Current Voice-over-IP solutions choose performance over security, opening the door to electronic eavesdroppers. Interconnections between the Internet and PSTN enable each system to be impacted by security vulnerabilities of the other. This will surely accelerate the need to harden Voice-over-IP systems to prevent their use as backdoors into the PSTN network.

APPENDIX A
SS7 STANDARDS

ITU-TSS SS7 Standards [10],[11]

SS7 Introduction

Q.700 – Specifications of Signaling System No.7

Message Transfer Part (MTP)

Q.701 – Functional Description of the Message Transfer Part

Q.702 – Signaling Data Link

Q.703 – Signaling Link

Q.704 – Signaling Network Functions and Messages

Q.705 – Signaling Network Structure

Q.706 – MTP Signaling Performance

Q.707 – Testing and Maintenance

Q.708 – Numbering of International Signaling Point Codes

Q.709 – Hypothetical Signaling Reference Connection

Q.710 – Simplified MTP Version for Small Systems

Signaling Connection Control Part (SCCP)

Q.711 – Functional Description of SCCP

Q.712 – Definition and Functions of SCCP Messages

Q.713 – SCCP Formats and Codes

Q.714 – SCCP Procedures

Q.716 – SCCP Performance

Telephone User Part (TUP)

Q.721 – Functional Description of TUP

Q.722 – General Function of Telephone Messages and Signals

Q.723 – Formats and Codes

Q.724 – Signaling Procedures

Q.725 – Signaling Performance in the Telephone Applications

ISDN Supplementary Services

Q.730 – ISDN Supplementary Services

Q.731 Series – Number Identification Supplementary Services

Q.732 Series – Call Offering Supplementary Services

Q.733 Series – Call Completion Supplementary Services

Q.734 Series – Multiparty Supplementary Services

Q.735 Series – Community of Interest Supplementary Services

Q.736 Series – Charging Supplementary Services

Q.737 Series – Additional Information Supplementary Services

Data User Part (DUP)

Q.741 – SS7 Data User Part

Management

Q.750 – Overview of SS7 Management

Q.751 – Network Element Management Information Model for the Message Transfer Part

Q.752 – Monitoring and Measurements for SS7 Networks

Q.753 – Management Functions MRVT, SRVT, and CVT and Definition of the OMASE-User

Q.754 – Management Application Service Element (ASE) Definitions

Q.755 – Protocol Tests

ISDN User Part (ISUP)

Q.761 – Functional Description of ISUP

Q.762 – General Function of Messages and Signals

Q.763 – Formats and Codes

Q.764 – Signaling Procedures

Q.766 – Performance Objectives in the ISDN Application

Q.767 – Application of the ISUP for International ISDN Interconnections

Transaction Capabilities Application Part (TCAP)

Q.771 – Functional Description of TCAP

Q.772 – Transaction Capabilities Information Element Definitions

Q.773 – TCAP Formats and Encoding

Q.774 – TCAP Procedures

Q.775 – SS7 Test Specification General Description

Test Specification

Q.780 – SS7 Test Specification-General Description

Q.781 – MTP Level 2 Test Specification

Q.782 – MTP Level 3 Test Specification

Q.783 – TUP Test Specification

Q.784 – ISUP Basic Call Test Specification

Q.785 – ISUP Protocol Test Specification for Supplementary Services

Q.786 – SCCP Test Specification

Q.787 – TCAP Test Specification

Monitoring and Measurements

Q.791 – Monitoring and Measurements for SS7 Networks

Operations, Administration, and Maintenance (OAM)

Q.795 – Operations, Maintenance, and Administration Part

ANSI SS7 Standards [10]

T1.110 – Telecommunications SS7 – General Information

T1.111 – Telecommunications SS7 – Functional Description of the Signaling System Message Transfer Part (MTP)

T1.112 – Telecommunications SS7 – Signaling Connection Control Part (SCCP)

T1.113-1992 – Telecommunications SS7 – Integrated Services Digital Network (ISDN) User Part (ISUP)

T1.113-1995 – Telecommunications SS7 – Integrated Services Digital Network (ISDN) User Part (NxDSO Multirate Connection) (supplement to T1.113-1992)

T1.114 – Telecommunications SS7 – Transaction Capabilities Application Part (TCAP)

T1.115 – Telecommunications – Monitoring and Measurements for SS7 Networks

T1.116 – Telecommunications SS7 – Operations, Maintenance, and Administration Part (OMAP)

T1.118 – Telecommunications SS7 – Intermediate Signaling Network Identification (ISNI)

T1.226 – Telecommunications – Operations, Administration, Maintenance, and Provisioning (OAMP) – Management of Functions for SS7 Network Interconnections

T1.609 – Telecommunications – Interworking between the ISDN User-to-Network Interwork Interface Protocol and the SS7 ISDN User Part (ISUP)

T1.611 – Telecommunications SS7 – Supplementary Services for Non-ISDN Subscribers

T1.631 – Telecommunications SS7 – High Probability of Completion (HPC) Network Capability

APPENDIX B
LIST OF REFERENCES

[1] Mohsen Guizani, Ammar Rays, and Mohammed Atiquzzaman, "IP Telephony – Guest Editorial," *IEEE Communications*, April 2000.

[2] S. Schmelling and V. Vittorer, "Evolution or revolution?" Telephony, Vol. 235, No. 20, November 1998, pp. 28-40.

[3] Mahbub Hassan, Alfandika Nayandoro, and Mohammed Atiquzzaman, "Internet Telephony: Services, Technical Challenges, and Products," IEEE Communications Magazine, April 2000.

[4] Donna Bergmark and S. Keshav, "Building Blocks for IP Telephony," IEEE Communications Magazine, April 2000.

[5] Bo Li, Mounir Hamdi, Dongyl Jiang, Xi-Ren Cao, and Y. Thomas Hou, "QoS-Enabled Voice Support in the Next-Generation Internet: Issues, Existing Approaches and Challenges," IEEE Communications Magazine, April 2000.

[6] Elizabeth Clark, "ATM: Alive and Well on the WAN," http://www.networkmagzine.com, May 2, 2000.

[7] H. Schulzrinne and J. Rosenberg, "The IETF Internet Telephony Architecture and Protocols," IEEE Network, May/June 1999, pp. 18-23.

[8] J. Toga and J. Ott, "ITU-T standardization activities for interactive multimedia communications on packet-based networks: H.323 and related recommendations," Computer Networks, Vol. 31, 1999, pp. 205-223.

[9] Cheryl Buckles, "SS7 gateways serve and protect," Telephony, November 20, 1989.

[10] Travis Russell, Signaling System #7, 3rd Edition, 2000.

[11] National Communications System, "Signaling System Number 7 Standardization," Technical Information Bulletin 98-5, June 1998.

[12] Intel Corporation, "The OSI Model Revisited," Intel, 1988, http://www.intel.nl/training/olc/course/cert/fn2/mod_11/lesson_1/fn21101.htm.

[13] Yet Chang, Simon Ng, and Francois Stradler, "Inter-Networking of Wireless Communication Carriers and Local Exchange Carriers: An Engineering Perspective," *XVI World Telecom Congress Proceedings*, World Telecommunications Congress, 1997, Toronto, Ontario, Canada.

[14] Robert N. Smith and Sourav Bhattacharya, "Operating Firewalls Outside the LAN Perimeter," *IEEE International Performance, Computing, and Communications*, Phoenix and Scottsdale, Arizona, February 10-12, 1999, IEEE 0-7803-5258-0/99, 1999.

[15] Wanjiun Liao and Jen-Chi Liu, "VoIP Mobility in IP/Cellular Network Internetworking," *IEEE Communications*, April 2000.

[16] "Notes on Distance Dialing," American Telephone and Telegraph Company, Engineering and Network Services Department, Systems Planning Section, 1975.

[17] Gary L. Ragsdale and Gerard P. Lynch, "Mechanisms and Solutions for Denial of ATM Services," NCS TIB 99-7, November 1999.

[18] William Stallings, "Differentiated Services," *Communications Systems Design*, February 2000, pp. 46-49. http://www.csdmag.com.

[19] Global Knowledge, "Converging Voice and Data Networks."

[20] Uyless Black, Voice over IP, Prentice Hall, New Jersey, 2000.

[21] Carlos M. Pazos, Marek R. Kotelba, and Andrew G. Malis, "Real-Time Multimedia over ATM: RMOA," *IEEE Communications*, April 2000.

[22] Tim Greene, "Voice over DSL talk of the town," Network World, April 10, 2000. http://www.nwfusion.com/news/2000/0410voicedsl.html.

[23] Tim Green, "Voice-over-DSL turns heads at ComNet," Network World, January 31, 2000. http://www.nwfusion.com/archive/2000/86139_01-31-2000.html.

[24] George H. Dobrowski and Ajay Sharma, "Emerging Technology Series #2: VoDSL: The Facts Behind AAL2," *Communication Systems Design*, May 2000.

[25] Kevin K. Whang, William Fink, Bala Krihnamurthy, I-Far Lin, David M. Rouse, and Henrik V. Sorensen, "Voice over PacketStar Gateway Solution for Service Provider Networks," *Bell Labs Technical Journal*, October-December 1998.

[26] IETF, "MEGACO Protocol," http://www.ietf.org/ids.by.wg/megaco.html.