Section V:

Computer Applications, Data Communications, and Software Directions

Switch and Network Transitions and Ramifications

Gary L. Andresen

Vice President of Marketing

AltiGen Communications

Confronted by the rapid increase in demand for data-networking services, from business customers as well as Internet users, carriers are looking for a way to manage multiple services on a common infrastructure. In fact, the integration of voice, data, and video services over a single network has been a major carrier goal for many years.

Although data services clearly are driving industry growth, voice traffic still accounts for the majority of revenues—and bandwidth consumption. Currently, most carriers do not derive enough revenue from their data offerings to justify an investment in a standalone network; however, to handle the growing volume of data traffic, they rapidly are using up any surplus bandwidth they might have on the voice side of the network.

Although they cannot afford to maintain two separate infrastructures, prudent network planners know they must expand network capacity to accommodate new services. Any new equipment installations must support both voice and data applications. Although numerous technologies do so, only asynchronous transfer mode (ATM) can be implemented in both the switching and transport infrastructures.

ADSL Solves Problems But Creates Others

Looming on the horizon is the widespread deployment of asymmetric digital subscriber line (ADSL) technology, that will enable carriers to provide larger-bandwidth pipes, as well as move Internet calls, with their long hold times, off the local switch. Yet while ADSL promises to solve one set of issues, it also presents another significant challenge for carriers. Given that many ADSL systems use ATM to transport data traffic, how can carriers manage ATM facilities within their existing asynchronous and synchronous optical network (SONET) infrastructures?

The first aspect of this challenge is the need for broadband facilities. Most central offices today typically use a DS3 or STS-l electrical as their largest broadband facility. While this is sufficient for DS1-based services, it will not hold up under the demands of larger-bandwidth services. Emerging high-capacity services are prompting carriers to rely more and more on OC-3 and OC-12 optics for intra-office broadband facilities.

The second aspect of the challenge deals with the management of the broadband ATM facilities in the network. ATM, like SONET facilities, requires certain network-maintenance functions—grooming, private-line management, and network-testing capabilities.

A Modified Public Network Model

Given the magnitude and complexity of this challenge, carriers need a new network model, one that not only supports existing voice applications but also provides the infrastructure for growing data services. Currently, the network comprises voice switches interconnected via a transport infrastructure. It is possible to design a similar and complementary topology by adding ATM as an extension of the existing network.

This unified network can support all voice, data, and video services on a common transmission infrastructure, as well as interact with Internet protocol (IP) routers for Internet services. The new network model, shown in *Figure 1*, comprises three layers: Level 1, the transport layer; Level 2, the switching layer; and Level 3, the routing layer.

The foundation of the network is the transport infrastructure, which supplies the interconnectivity necessary to support the switching and routing infrastructures. The transport layer is the most efficient, cost-effective means for interconnecting the various switching platforms that provide services to end users. In turn, this new network model is the most efficient, cost-effective means by which carriers can manage ATM and IP traffic.

FIGURE 1

Telephony Switch Evolution

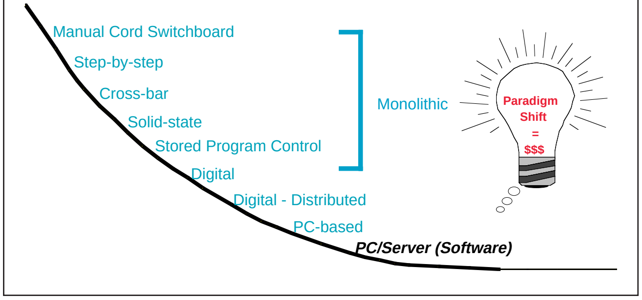
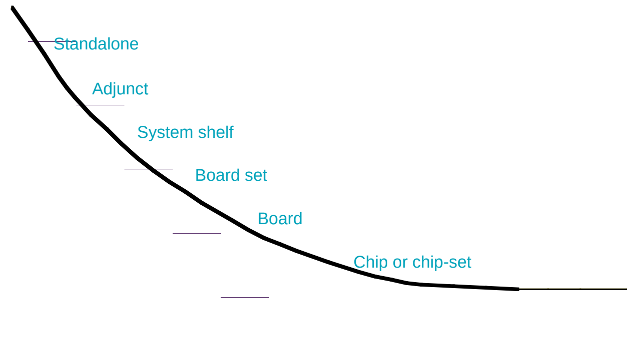


FIGURE 2

Adjunct Telephony System Evolution



The switch layer provides the call setup and teardown functions throughout the network that are required to deliver services on voice, frame relay, and ATM switching platforms. IP routing, for the delivery of data services, creates an additional layer above the network switching layer.

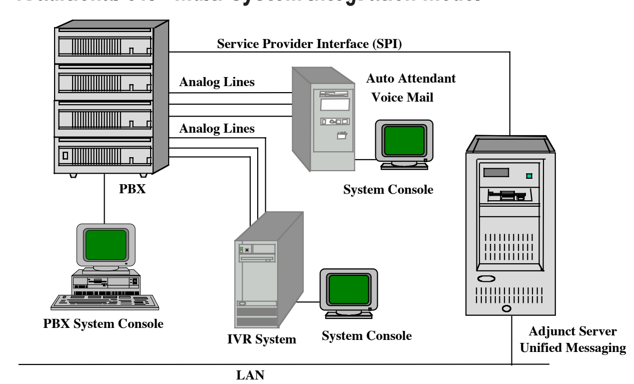
Connectionless and Connection-Oriented Transport

Using the transmission control protocol / Internet protocol (TCP / IP) protocol, carriers can provide Internet services over many different types of transport infrastructures: Ethernet, integrated services digital network (ISDN), frame relay (FR), ATM, or SONET. For end-to-end customer service, IP relies on a connectionless protocol that requires the administration of routing tables throughout the network. This connectionless routing is most efficient when the data to be transmitted contains a relatively small number of packets for each end-to-end flow.

When a large number of packets for a single end-to-end flow need to be transmitted, a connection-oriented protocol is more efficient. The carrier can set up an end-to-end path to transport all relevant packets, without the processing required at each switching platform to determine from the router header each packet's destination. The tradeoff between connectionless and connection-oriented data transmission necessitates an interaction between the IP layer and the switching layer, simply to ensure that network resources are used in the most efficient manner. Several methods for setting up the end-to-end connections for large IP traffic flows are currently under investigation; Ipsilon's IP switching and Cisco's Tag

FIGURE 3

Traditional CTI - Multi-System Integration Model



Switching are two examples that provide the interaction between the routing layer and the switching layer for the efficient support of IP traffic through large networks.

With today's network, voice and data customers are connected to the appropriate switching platform through the "network cloud," which provides a transparent interconnection for the switching platforms (see *Figure 2*). The transport network performs its job so efficiently that most people dealing with the switching infrastructure hardly know that it is there, yet it provides critical network functions.

ADMs and DSCs Add Up to More Efficiency

The transport network efficiently and reliably interconnects the customer and the switching systems that provide the customer-ordered service (see *Figure 3*). Add / drop multiplexers (ADMs) and digital cross-connects (DCSs) further enhance the transport network's efficiency. ADMs interconnect the customer's premise with the central office, as well as all the central offices in a carrier's network.

Ensuring network survivability for both linear and ring applications has long been a cornerstone ADM strength, enabling carriers to provide high-availability services to customers. Cross-connect systems primarily have served as vehicles for managing the interconnection of central office bandwidth. Such bandwidth interconnection is necessary, for example, to support service churn and to consolidate the payloads of partially-filled asynchronous and SONET facilities, thus making more efficient use of switch-port resources and interoffice transport bandwidth. The cross-connect has become the centerpiece of telecom networks, with nearly all the central-office bandwidth traversing it.

The Cross-Connect is the Workhorse of the Network

Because remote operations centers can prompt the cross-connect to redirect traffic, it has become the primary vehicle for managing non-switched (private-line services), as well as for maintaining the network. Economics has driven both cross-connect functions. A lower price per port on the cross-connect caused the migration of private-line services away from switching platforms, while carrier staff reductions created the need to provide remote test-access functions on the cross-connect for turn-up testing and troubleshooting of facility problems. The addition of ATM to the network by no means has eliminated the need for these critical transport-network functions; indeed, ATM actually has expanded the need to make the same functions available for asynchronous and SONET equipment on broadband ATM facilities.

Transport Layer Adds ATM Management Functions

With the addition of ATM to the network, equipment in the transport layer must add ATM facility-management functions to complement the management capabilities currently available for asynchronous and SONET facilities. Specifically, ADMs must incorporate virtual path (VP) ring functionality to provide VP trunks and survivability for ATM traffic, while DCSs must incorporate VP / virtual circuit (VC) connection management to groom traffic for more efficient use of network resources.

Carriers can add ATM functionality in two different ways:

- co-locate new ATM-based equipment with the existing equipment infrastructure
- upgrade the existing equipment by adding ATM functionality

Each method has its pros and cons. The major advantage of upgrading existing equipment is that it extends the economic life of the current investment. Such an approach also means the carrier does not have to create a new infrastructure to support new equipment. Instead, the carrier can enhance existing methods and support systems to accommodate the new functionality on the embedded equipment.

Short-Term Solution Poses Long-Term Challenges

Nevertheless, many carriers are choosing, at least in the near term, to deploy special-purpose ATM equipment, primarily because of its availability. Although this choice may be expedient in the short run, carriers will have to deal with several long-term issues: scalability, reliability, maintenance, and cost-effective management of private-line services.

Most special-purpose ATM equipment has been designed for the customer-premise environment, rather than the carrier environment, and thus cannot scale to the hundreds-of-gigabits-per-second speed that cross-connects currently provide for asynchronous and SONET facilities. In addition, special-purpose ATM equipment often does not satisfy carrier criteria for reliability and quality.

When carriers treat each service type independently of others in the transport layer, they end up with networks that are very inefficient and difficult to manage, simply because each facility has its own unique properties. Today, carriers use cross-connects or ADMs to multiplex facilities from the various switching platforms onto SONET facilities.

In a best-case scenario, carriers consolidate separate DSl and DS3 facilities onto a common higher-capacity SONET facility; the worst-case scenario reveals very little consolidation taking place, with each data network handled as dark fiber (see *Figure 4*).

FIGURE 4

Multi-Application CT Model

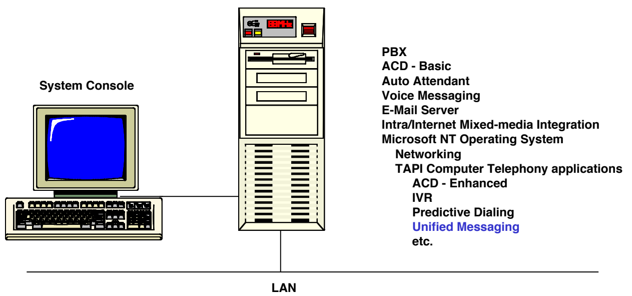
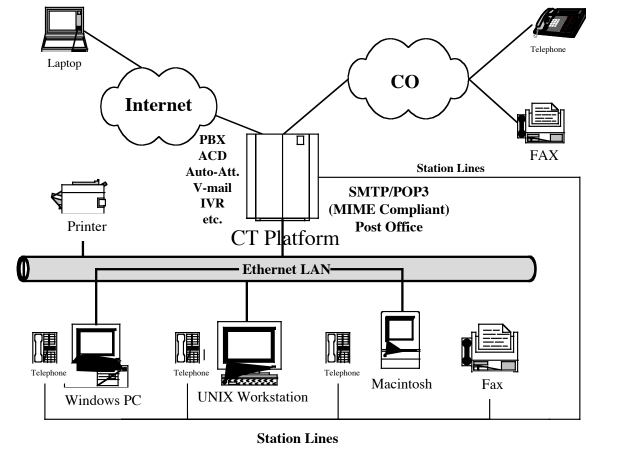


FIGURE 5

Computer Telephony — Multi-Application Model



Inefficient Use of Transport Facilities

Regardless of the approach, both significantly under-utilize network transport resources. Because peak loading for each type of traffic tends to occur at different times throughout the day, each dedicated facility has to be engineered for those separate peak times. On the other hand, a shared-transport medium means that a carrier has to support only aggregate traffic needs throughout the day and thus can size facilities more efficiently.

One of the ATM's central advantages always has been its ability to support voice, data, and video traffic—for switching as well as transport. This advantage first became evident in the campus environment, where most LAN and FR switching platforms have migrated to ATM at the switching layer.

ATM Entering the Carrier Market

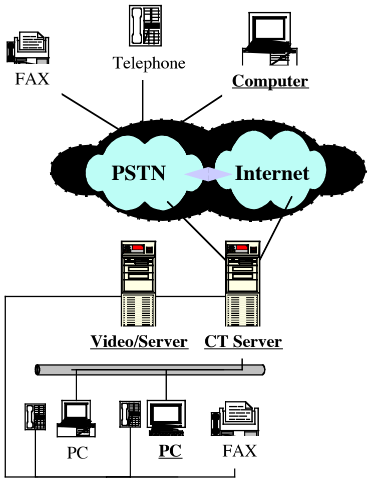
The network-transport layer in the carrier market has recently begun to provide ATM functionality to support dynamic service requirements. In addition to providing multiplexing functions for ATM switches and ADSL, ADMs now pick up LAN traffic from routers and convert it to ATM for transport on an STS-1 or STS-3c through the access network.

The presence of ATM in the transport infrastructure also is driving cross-connect systems to support ATM traffic, again, because carriers want to integrate all traffic types on a common infrastructure. With the addition of an ATM switch fabric, a cross-connect system, as shown in *Figure 5*, is capable of grooming ATM traffic from multiple sources onto a common backbone, thus boosting overall efficiency of the entire network.

AAL Translates into a Single ATM Backbone

Non-ATM services are converted to an ATM cell stream via the ATM adaptation layer (AAL) function, also referred to as "segmentation and reassembly" (SAR). For each service type, there is a distinct AAL function: AAL1 for DS1 adaptation, and AAL5 for LAN adaptation and frame-relay adaptation.

FIGURE 6
Multimedia Communications



With the use of AAL functions, a single ATM backbone can carry efficiently all types of service, including voice.

The ATM cell streams are managed through the network within VPs and VCs. The cross-connect will provide on ATM facilities the same functions it currently provides on asynchronous and SONET facilities—grooming, private-line management, and network maintenance. Adaptation-layer functions can be provided either within the cross-connect or at the edge of the network where the service originates.

The Hybrid Cross-Connect

With an ATM matrix, a cross-connect has the same capabilities as an ATM switch, but goes two steps farther: an ATM-equipped cross-connect is optimized for scalability (current SONET cross-connects host up to 200 Gbps of bandwidth), and it provides low-cost management of private-line services.

Figure 6 illustrates how an ATM-capable cross-connect can groom VPs for several partially-filled facilities and create a new highly-filled ATM facility. The VPs can be the trunks between ATM switches performing real-time call processing or VPs handed off to the network for end-to-end private-line services. Switching ATM cells within the cross-connect matrix means greater efficiency for the entire network and lower operating costs for the carrier.

Within this architecture, the ATM edge switch, which may include an IP router, now receives well-packed traffic on its interfaces. Only those ATM cells that need ATM-switched service are routed to the edge switch, thus minimizing the size and maximizing the efficiency of that switch. In addition, the only administrative tasks involve those circuits that use the switched services; the ATM VP cross-connect portions of the hybrid cross-connect, which is statistically provisioned as a private virtual circuit (PVC), handle all other ATM traffic.

Using CTI to Enhance Customer Service

Elizabeth A. Chambers
Product Manager
Versatility, Inc.

The integration of computing and telecommunications technologies has long been a topic of fascination. Since the vast increase in competition spurred by the Telecommunications Act of 1996, integration has become a virtual necessity. Computer telephony integration (CTI) can and should be used in many sectors of the business and personal arena, not the least important of which is customer service. This paper will review the available CTI technologies relative to their expression in a customer-service or customer-interaction environment. It will then describe how CTI provides solutions for many telecommunications problems.

Industry Challenges

Deregulation has created many challenges for the telecommunications industry. Acquiring new customers for new and existing services has become more of a challenge because there are more competitors. A second challenge involves accelerated product delivery: companies are trying to deliver products to customers in a matter of minutes rather than in hours or days. Minimizing subscriber churn is yet another important issue. Because there is so much new competition, many users will purchase a service and use it only until the end of the introductory period, afterward moving to the next competitor who offers a special incentive. This process is known as churn. For understandable reasons, telecommunications companies are trying to keep churn levels as low as possible, as well as recruit new customers from their competitors.

A fourth concern, associated with churn and customer interaction, is how companies can improve service quality indicators (SQI). This involves providing better-quality service to customers so that companies can differentiate themselves. A fifth challenge, which also lies in the area of customer interaction, concerns empowering sales representatives to entice the customer. Representatives must be able to offer potential buyers services and capabilities that they find exciting, new, personally tailored, etc. Finally, companies must provide enterprise-wide customer care.

Call Centers

Call centers are the arena for customer interaction. To many people, call centers might conjure unpleasant images-irritating phone calls from telemarketers selling credit cards or soliciting for insurance, to name just two of the most common types of calls associated with telemarketing. Because of this perception, established local exchange carriers (LECs) and interexchange carriers (IXCs) face the same acceptance challenges as new entrants. Indeed, call centers did start out as fundraising telesales. They are, however, evolving. Because it is so expensive to perform numerous professional visits and extensive customer and potential-customer mailings, call centers are becoming a cost-effective way for a company to interact with consumers for sales, as well as service. The term *call center* is even falling into disuse in favor of the phrase *customer interaction systems (CIS).* Customer interactions systems means that the call center becomes the primary focus for all customer interactions—from customer services, to billing inquiries and sales.

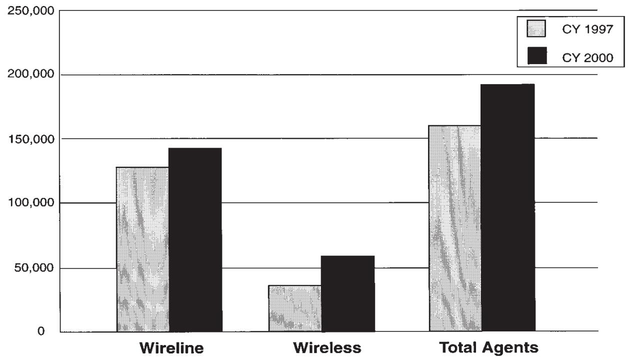
For telecommunications companies, the call center is the heart of the organization. It is the place through which customers can reach providers. As mentioned above, there is an evolution toward customer interaction systems that allow agents at a single site or multiple sites combined through multiple private branch exchanges (PBXs) and networks to communicate with customers. These agents, who face the challenges of many different systems, will have integrated desktops to help them deal with different billing, service order-processing, service-activation, service-acquisition, and service-provisioning systems. They will also have to utilize graphical information systems (GIFs). GIF will enable agents to pull up a map that displays the problem area when a customer calls with a complaint. They can then dispatch service personnel to that spot. Agents at the CIS will also handle customer requests for literature addressing their service.

There are far more than simple sales occurring at today's call centers. *Figure 1* illustrates current call-center trends. Research has shown that the average growth of telemarketing service centers or call centers and LECs has been just 10.9 percent each year since 1991, while the cellular industry has grown 472 percent each year, paging has grown 384 percent, other wireless has increased over 500 percent, and cable has shown a relatively low 45 percent annual growth. Call centers are becoming a very important part of the organization.

There are several ways in which CTI can help meet call-center challenges. First, CTI allows for reduced call time. This can be diminished in two areas: in outbound auto and predictive dialing, and in nearly every inbound instance. Saving time is

FIGURE 1

Current Call Center Trends



Sources: Corporate Annual Reports, SEC 10K Reports, Company Press Releases, Telephony, Interviews with 15 Carriers, Andersen Consulting, CTIA, PCIA, USTA

more than just dialing various numbers quickly. It also allows agents to dial automatically.

The inbound CTI applications are even more important. These include automatic number identification (ANI),Dialed Number Identification Service (DNIS),and interactive voice response (IVR). IVR is used by banks, for example, where customers call in for account balance information, etc. CTI allows data coming from an IVR to be linked to stored customer data and to the contact-management system. For example, there might be a customer who calls a bank with a question about whether a check has cleared. The customer places a call to the call center, enters the necessary account information, listens to the automated response, and is informed that the check has not yet cleared. That customer then wishes to speak to an agent. Unfortunately, when that customer is connected to an agent, all too often the agent requests the caller's name and account number-information that the caller has already supplied when interacting with the IVR. CTI technology is now available that will transfer the IVR information to the answering agent. This will not only save time for both agent and customer, but also present an image of efficiency and technical capability.

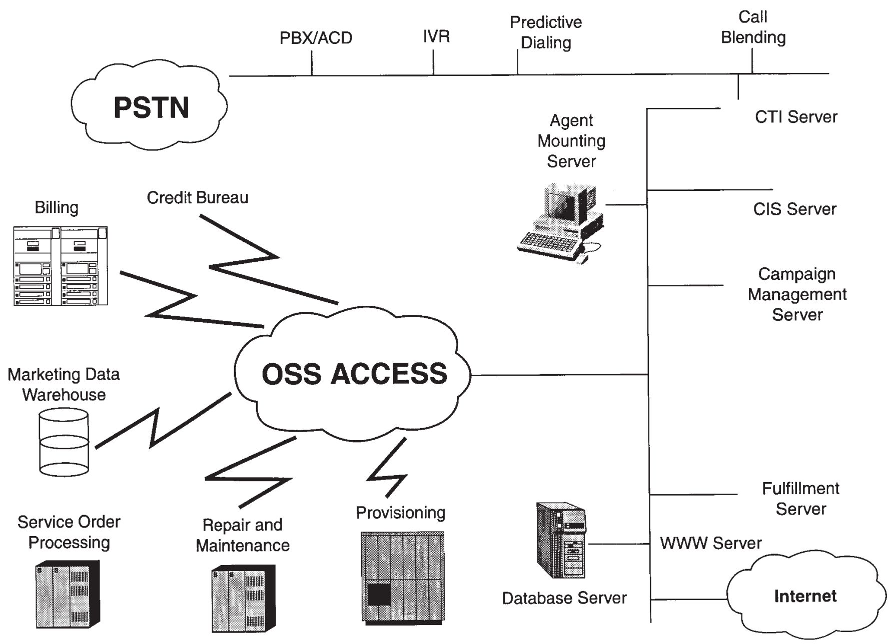
In addition, CTI provides better call tracking, which is very important for a call center. It is helpful for executives to know what customers call in about, what their issues are, how long it takes agents to deal with these issues, etc.

Another benefit of CTI is case tracking and escalation. A customer may call in one day to report a problem with a telephone, and then call again the next day to report that the problem remains unsolved. The agent should immediately see, based on the caller's telephone number, that he or she had called earlier and reported a problem. The agent should either know what solution was provided or be able to update the caller on the status of the solution. This can be done through CTI, again making the customer feel that the necessary information is being provided by the call center personnel as effectively as possible.

In addition, CTI enables calls to be delivered to the appropriate customer-service representatives (CSRs). When a call comes in through the PBX, it can be routed automatically to an appropriate agent, based not only on dialed number identification service code but also on information about the last agent to talk to the customer or agents assigned to work in particular geographic regions. All of these solutions contribute to improved operational efficiencies. Through CTI, a call center is able to deal with customers in a much more effective manner. There are additional ways in which CTI helps meet call-center challenges. Automated callbacks are one way of performing

FIGURE 2

CTI Technology Options



simplistic multimedia queueing in the sense that calling customers back and making commitments after sending literature or after they sign on to a new service can be automatic, or without any agent intervention.

Finally, CTI helps with World Wide Web integration. As an example of how CTI can help an organization, it may be helpful to imagine a company that works with telecommunications companies to provide activation, credit and fraud checking, up-selling, and cross selling. Through CTI, it is possible to activate a new phone in five minutes. When consumers purchase a telephone from a box, they call into an 800 number. This call triggers a screen pop, which in turn sets off an activation process, which checks for credit card fraud, checks credit card information, and within five minutes activates that telephone number and gives the consumer a new telephone number. This is very important, as most consumers want to be able to use a new purchase immediately.

Available CTI Technologies

There are a number of available CTI technologies, including the following: outbound dialing, inbound call delivery, IVR screen pop, call blending, integrated reporting, and Internet telephony. *Figure 2* shows how CTI can be used with legacy PBXs and / or automatic call distributors (ACDs). This technology will not last forever—it seems probable that within ten or fifteen years there will be true Internet telephony, and PBXs will have been replaced by software. Until that time, however, the large investment that organizations have made in legacy systems should not be ignored. To this end, there is CTI software that can interact with different switches to provide CTI through software and through software servers on the local area network (LAN).

Outbound CTI

The important terms related to outbound CTI are "preview," "progressive," and "predictive dialing." Progressive and predictive dialing refer to automatic dialing. Predictive dialing is the process of calling more consumers or businesses than there are agents available in the expectation that when one consumer hangs up, another consumer will be picking up his or her telephone and can automatically talk to a representative. Many consumers strongly dislike these types of calls—calls that are answered by the receiver and then, after a lag, by the agent—but they are particularly useful for call centers..Many organizations do this through CTI using legacy PBXs and ACDs.

Outbound CTI is faster than manual dialing, leading to significant performance improvements. When a call is made, it takes 10-14 seconds to pass through the network. Progressive dialing or predictive dialing eliminates that time, enabling agents to talk to customer after customer, increasing productivity and, hopefully, sales. A final positive aspect of outbound CTI is that agents or customer service representatives (CSRs) can be anywhere in the enterprise and still access the same software functionality.

Inbound CTI

As mentioned above, inbound CTI provides ANI screen pop, DNIS routing, and IVR screen pop. These types of capabilities minimize the amount of time that an agent must spend accessing customer information. The information is provided to the agent as soon as a call comes in, based on either the ANI, a DNIS code that will trigger the appropriate script, or the information that was entered into an IVR. It also displays scripting to lead an agent through a conversation based on previous caller contact. This is very important, because there is a great deal of turnover in call centers. Agents need enough information to tell them how to deal with various questions. People can be calling about anything, from a telephone problem, to wanting to add new services, to wanting to check billing, and agents must be empowered to handle all of these issues.

Finally, agents should be able to review information that comes from the IVR within context. This will inform them about the caller's history, for what reasons the calls were placed, whether the caller interacted with an agent or an IVR, etc. With this information, agents can more intelligently handle issues and questions according to each customer's personal history. This technology will allow telecommunications firms to distinguish themselves. If customers call into a cellular company where agents know nothing about them and cannot access their histories, then it is less likely that those customers will stay with the company. On the other hand, if customers call into a provider where CTI technologies allow agents to be more personal, then they are more likely to stay. This factor cannot be emphasized enough, because in the final analysis, services, in terms of cost and telephone equipment, may not be that different. It is the call center that will finally make the difference relative to the way the customer views the company.

Call Blending

Call blending allows many different technologies to be blended for a telecommunications company. This means that agents working at the call center can better utilize their time. They can take inbound calls and then, if there is a lag, switch easily to outbound calling. Through call blending, a universal call center can have different agent groups: inbound only, outbound only, inbound preferred, or outbound preferred. Call blending, managed by ACD queues, alleviates the problem of inbound peaks and valleys, allowing agents to be productive 100 percent of the time.

Integrated Reporting

Integrated reporting helps call center personnel remain aware of the different levels of customer interaction. This includes what types of interactions the caller is having—IVR or personal—as well as what has been done on the IVR. Other information provided by integrated reporting includes call outcomes and time spent on calls. A customer's interactions with the call center are stored in a single contact-history database, which greatly improves analysis of customer interaction. Again, the use of integrated reporting should enable agents to provide better customer service, thus encouraging the customer to stay with the provider.

The Internet

CTI and the Internet involve interesting and exciting new technologies. Despite media presentation, however, the necessary technology is still under development. Call centers have been talking about Internet integration for almost two years, but it is only recently that these technologies have started to emerge. Additionally, since many businesses are historically slow to adopt new technology, even that which is available is not yet prominent. It does seem probable that the near future will see video and telephone conferencing, fax over the Internet (FON), voice on the Internet (VON), and Internet ACDs.

Relative to today's technologies, companies such as Venturian Software, AT&T, and Teloquent have come out with solutions. The basic application is Internet "call me," although Forrester calls it "teleweb." *Figure 3* shows the Internet "call me" architecture, which dramatically extends the boundaries of the traditional call center. In this scenario, the user can be on a Web page for an enterprise, request a call about a particular service, and schedule a time that an agent should call. This changes the dynamic of the customer's interaction with the CSR. The CSR will call already knowing what the customer wants to discuss. Also, transaction time is shortened. Relative to economics, there is a high correlation between Internet callers, who are still a very small audience, and money spent. This should serve as a strong incentive for call centers to add this interaction method.

There are many other functionalities that will be available through CTI and the Internet. Some are already available, but are not being used widely. These capabilities are particularly exciting, as they mean that a single transport-Internet protocol will be able to manage voice, data, and video, effectively collapsing three networks into one.

Summary

CTI already brings many benefits to a call center, such as the following:

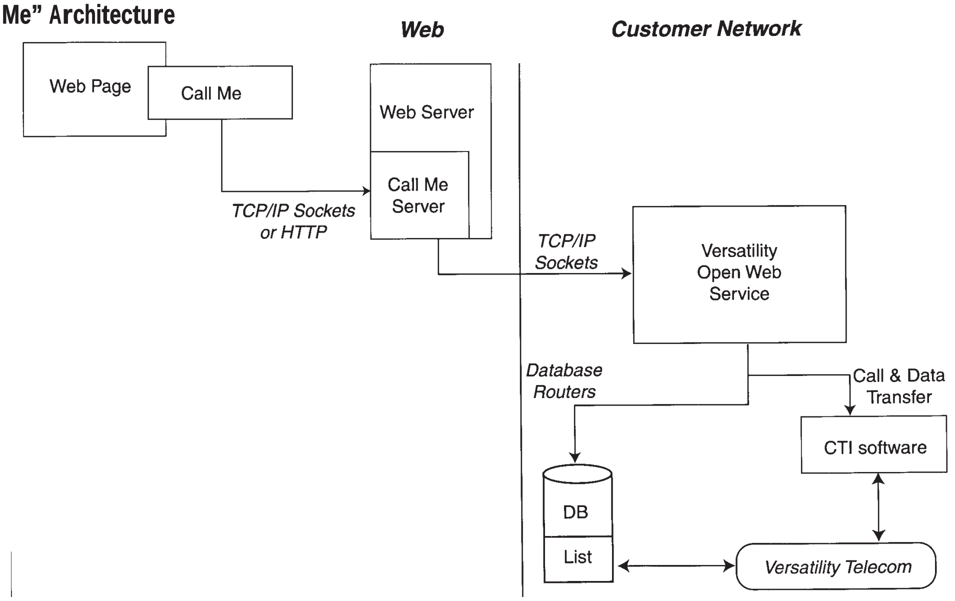
- improved acquisition and retention
- shorter transaction times
- blending calling opportunities
- integrated reporting

In addition to these, CTI and the Internet have already started and will almost certainly continue to provide a valuable new mechanism for customer interaction. It should be noted that this Internet involvement will probably not revolutionize the outward appearance of the call center. Rather, it will seem to the agents like just another IVR call.

Despite their benefits, many call centers have been slow to adopt any of the available CTI technologies. Surprisingly, even in the telecommunications area, there are few centers that have adopted CTI and use it on a day-by-day, minute-to-minute basis. It should certainly be expected, however, that the near future will see more widely adopted computer telephony integration solutions, since companies that use CTI are able to markedly differentiate themselves from the competition.

FIGURE 3

Internet "Call Me" Architecture



Current Status of National and International IN Standards and TINA-C

Jim Garrahan
Project Manager, Advanced Intelligent Networks
Bellcore

The public telephone network is continually evolving. From the days of the intelligent network (IN) to the advanced intelligent network (AIN) comes the next phase, which is telecommunications information networking architecture (TINA). This evolution addresses additional paradigms, such as object-oriented methodologies and distributed processing, which expand the narrowband arena and embrace broadband methodologies, concepts, and services.

Figure 1 is a matrix of information that compares the architectures of IN, AIN, and TINA. Imagine another dimension in this figure to address time, which would include where this work originated, current status, and where the technologies are headed. The portions of this matrix to fill in are the relationships among these initiatives: their similarities, differences, paradigm shifts, and how the technologies will evolve.

AIN's evolution involves national implementation and deployment of AIN 0.1 and AIN 0.2, which are described in the context of the framework architecture of AIN Release 1. From an International Telecommunication Union (ITU) perspective, the evolutionary process is described in terms of capability sets. IN CS-1 refers to IN Capability Set 1, the first set of ITU IN Recommendations. IN CS-2 and CS-3 also exist, and all three of these are described in the context of a long-term capability set (LTCS). Communication exists between the American National Standards Institute (ANSI), which governs North American IN standards, and the ITU. TINA has its own methodologies, and the evolutionary process is described in the terms of versions: versions 1 through 3 in the context of issue 4.

The Role of TINA-C

The Telecommunications Information Networking Architecture Consortium (TINA-C) was established in 1992 with about 40 manufacturers and network operators from the telecommunications and computing industries. The consortium's objective was to address the convergence of both industries. There have been several efforts to do that; TINA-C is yet another attempt involving convergence of telecommunications services, information services, and management services. References to the consortium's efforts are available at www.tinac.com.

One of the distinguishing characteristics of TINA-C is that it is addressing a software architecture. This is different from IN and AIN. For IN and AIN, although attention is paid to software paradigms and methodologies, a greater focus is on the physical entities and communication aspects among these en-

FIGURE 1
Status, Relationship, and Direction of Industry Initiatives

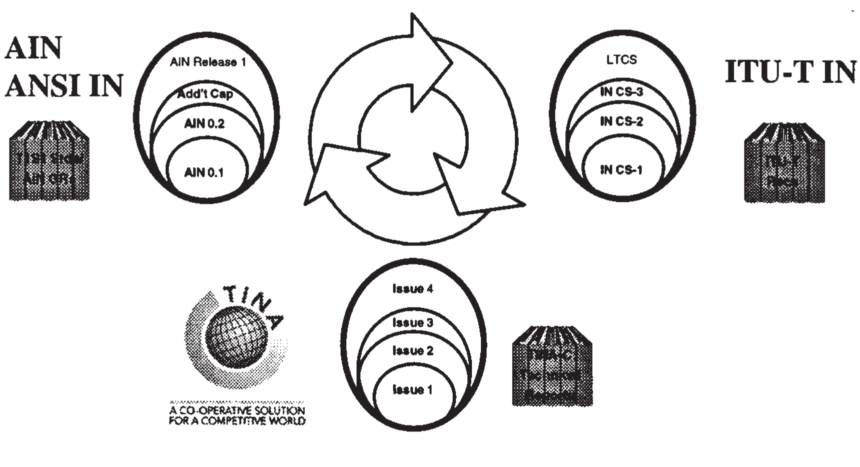


FIGURE 3
TINA-C Service Timeline

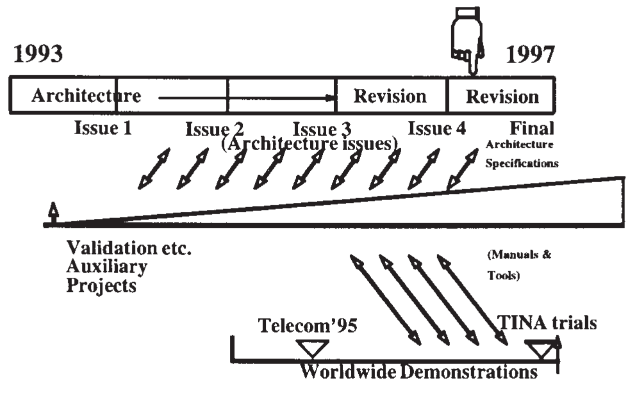
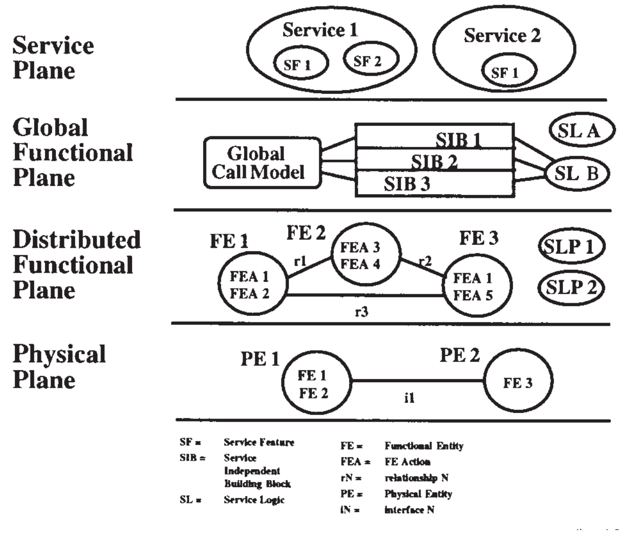


FIGURE 4
ITU-T IN Framework Architecture



tities: interfaces and protocols. TINA-C is addressing a distributing processing environment for telecommunications, information, and management services principally from a software architecture perspective.

AIN/IN Service Drivers

AIN evolved from the initial implementations of toll-free service and other alternate billing services (ABS). *Figure 2* illustrates this evolution. IN evolved into AIN 0.0, where introductions of call-models and functional modeling were made. From AIN 0.0, the call-models of AIN 0.1, AIN 0.2, and then additional AIN capabilities originated, which resulted in functional descriptions, capabilities, new reference points, and interfaces that have evolved over time.

The international activity associated with intelligent networking began in late 1989 with the specification of IN CS-1 Recommendations (the Q.121X series of Recommendations). Those Recommendations were initially released in 1992 and subsequently revised in 1995. Subsequent to that, the ITU worked on IN CS-2 with an approval phase completing in September 1997. Parallel efforts have been underway on the third capability set, IN CS-3.

The service drivers for IN CS-1 were predominately narrowband services addressing flexible routing, flexible user interaction, and flexible charging. These narrowband services provide a foundation of functionality, which are being applied to wireless access as well as broadband access.

TINA started with a blank slate. There was no attempt to work with an embedded telecommunications infrastructure. TINA started by addressing the service drivers from the perspectives of broadband, information services, and management services. TINA is using a three-phase approach (see *Figure 3*). There is a specification phase involving Versions 1 through 4 in a yearly release that started in 1993. Along with that effort, there was a validation phase as the specifications became more detailed. Prototyping was done to validate some of the architectural concepts which focused on open distributive processing (ODP). Use of object-oriented methodology instead of a strictly functional or interface view was adopted.

The third dimension introduced was to conduct demonstrations and trials. There was a trial at Telecom 95 in Geneva involving a TINA-C–compliant system to offer video-on-demand, messaging, and management services. This three-phase approach validated and demonstrated TINA systems.

Figure 4 shows the ITU perspective for architecture modeling in a four-plane approach: the service, global, distributed, and physical planes. These four planes represent different abstract views of an IN structured network and serve different needs for different users—there is no interaction among the planes. The services plane describes the service and features driving each capability set. The global functional plane addresses global functional logic and service-independent building blocks. The distributive functional plane (DFP) addresses the functional entities, functional entity actions, relationship among functional entities, and information flows; the DFP helps in defining the interfaces and protocols on the physical plane. The physical plane is what is traditionally associated with the service switching point (SSP), intelligent peripheral (IP), and service control point (SCP).

With respect to AIN, most generic requirements address the distributed functional plane and physical plane. There is a set of functionality and information flow descriptions that are used to define the protocol. This is also used in the ANSI IN standards efforts. The ANSI IN standards address a distributive functional plane and a physical plane in terms of interfaces and do not focus on the other two planes.

TINA-C, on the other hand, does address that upper-layer set of functionality. From a service architecture perspective, TINA addresses the equivalent of the service plane, global functional plane, and distributive functional plane.

FIGURE 5

Intelligent Network Architecture—Initial

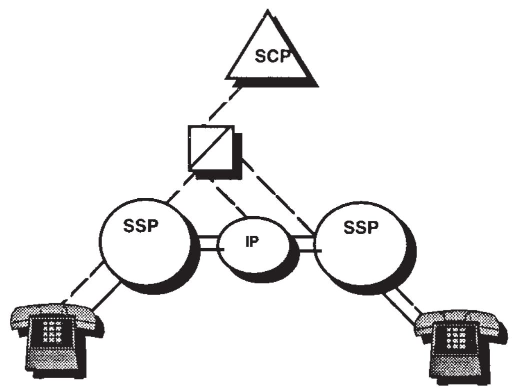
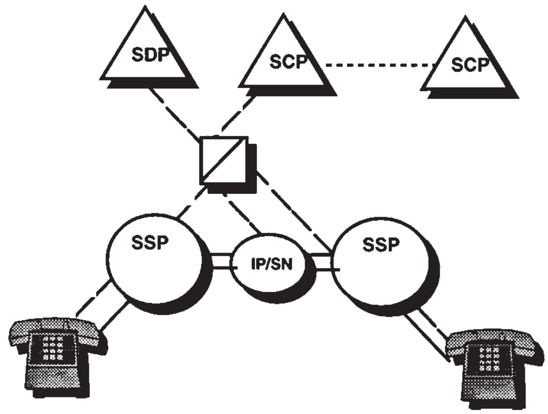


FIGURE 6

Intelligent Network Physical Architecture—Current

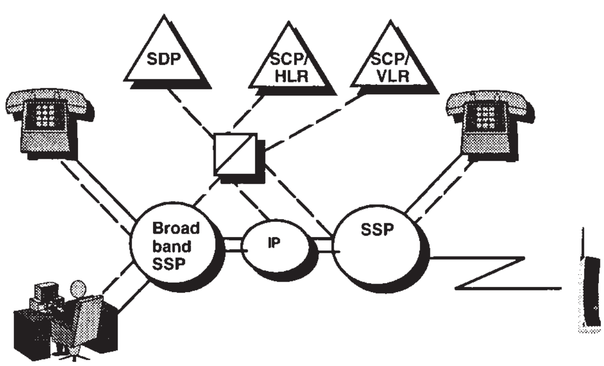


The TINA-C architecture focuses on five architectural views. There is an overall architecture describing methodologies and paradigms that are used by the underlying architectures—the service architecture, the management architecture, the network architecture, and the computing architecture. The overall architectures address the use of the distributive processing environment, ODP methodologies, object-oriented paradigms, layering, and separations of concerns. Therefore, the paradigms apply to all four underlying architectures. The service architecture addresses a services view of networks, which maps to a services plane view of an IN structure network.

The network architecture view maps more closely to the distributed functional plane of an IN structure network. There is an analogous set of functional entities, information flows, and relationships described. The management architecture maps to essentially all four planes of the ITU methodology. The ITU does not separate management per se, but there are management aspects included in services, the global functional plane, the distributed functional plane, and the physical plane. The management functionality addresses all four planes. TINA-C, however, has a set of specifications strictly associated with management.

FIGURE 7

Intelligent Network Physical Architecture— Evolving



The computing architecture does not map to any of the functionality described in the ITU IN conceptual model. The computing architecture addresses portions of the distributed processing environment, involving more of a software architecture perspective, which is not strictly addressed either in AIN or ITU-T IN.

The Evolution of the Intelligent Network Architecture

Figure 5 represents the starting point of the intelligent network architecture, and such an architecture should look very familiar. It is the initial physical plane of an IN structure network and is common to both AIN and ITU-T IN. The basic elements of an IN—SSPs, IPs, and SCPs—are included to provide VPN, toll-free, calling card, and alternate billing services.

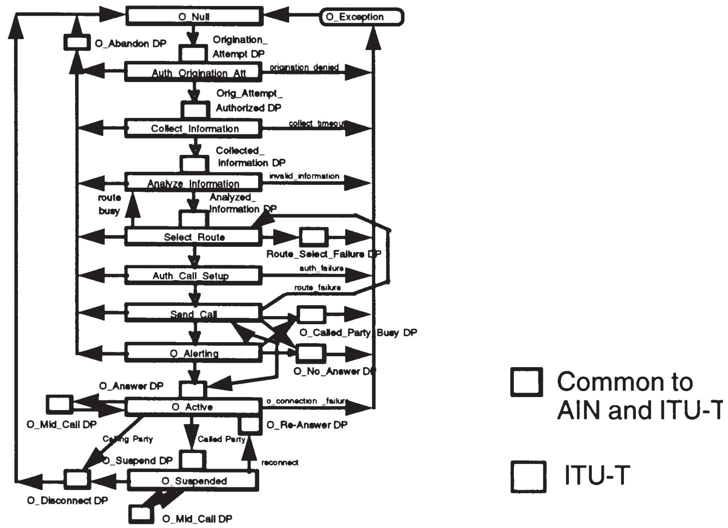
This architecture evolved from a variety of viewpoints, the outcome of which is depicted in *Figure 6*. One was to add functionality to intelligent peripherals in order to augment functionality from a service node or an adjunct. Additional nodes were described as service data points (SDPs). They address additional interfaces to facilitate SCP-to-SDP communications for internetworking mediated-access applications. As with the initial architecture, this scenario is common to both AIN and ITU-T IN, but it is not strictly part of TINA-C.

Figure 7 shows where the architecture is headed. First, different access arrangements for wireless and broadband access are supported. The functionality of home location registers (HLRs) and the context of an SCP are being addressed, allowing interworking with the mobility management application protocol—Interim Specification 41 (IS-41) in particular.

From a functional perspective, the call modeling concepts shown in *Figure 8* and *Figure 9* are common to both AIN and ITU-T IN. However, they are not strictly part of TINA-C. So when comparing these two figures, this discussion will simply refer to AIN and ITU-T IN. There are numerous similarities in terms of call modeling point-in-calls (PICs) and

FIGURE 8

Originating Basic Call Model



detection points (DPs). Differences exist in some of the terminology for PICs and DPs based on historical precedent. In addition, there is additional functionality described in ITU-T IN that supports mid-call or call-party handling functionality. AIN has an initial set of generic requirements associated with call-party handling, but most of the recent systems engineering was done in ITU-T. The notion is to use the ITU results for North American networks. In general, however, call model aspects are more specifically described in AIN for the originating basic call state model (O-BCSM) and terminating BCSM (T-BCSM).

Similarity between ITU-T IN and what is being described in AIN also exists in the T-BCSM (see *Figure 9*). The comparison in *Figure 8* and *Figure 9* is to the additional capabilities described in AIN, sometimes referred to as AIN 0.X. However, there are differences from a mid-call perspective. TINA does not have a notion of originating or terminating BCSMs.

Another architectural approach used for ITU IN was finite state machine (FSM) modeling (see *Figure 10*). The FSMs describe states and events associated with state transitions for each of the functional entities (FEs) that include a service switching function, a service control function, a specialized resource function, and a service data function. All of this appears on the distributed functional plane of the IN conceptual model.

ITU offers a detailed set of specifications associated with this set of functionalities. For instance, there are states described

FIGURE 9

TerminatingBasic Call Model

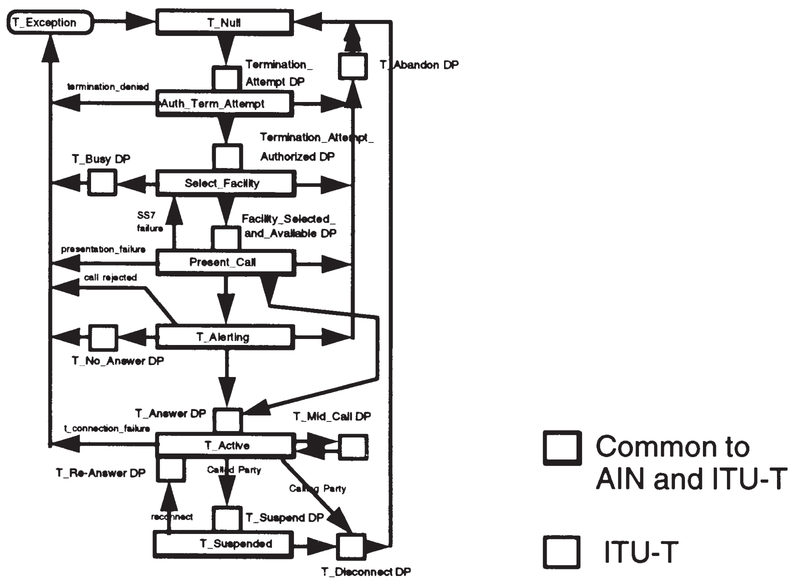
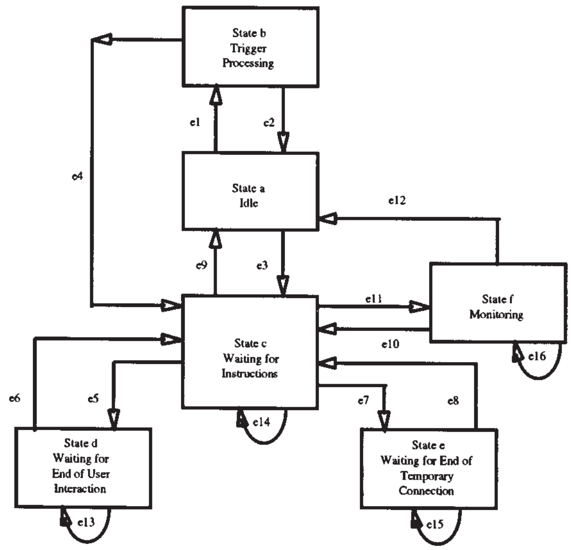


FIGURE 10

IN Finite State Machine



for monitoring or user interaction. However, this approach was not strictly used in AIN. There are other representations that were used for the specifications which were more appropriate; once again, this is a functional view that TINA-C does not have any direct mapping to.

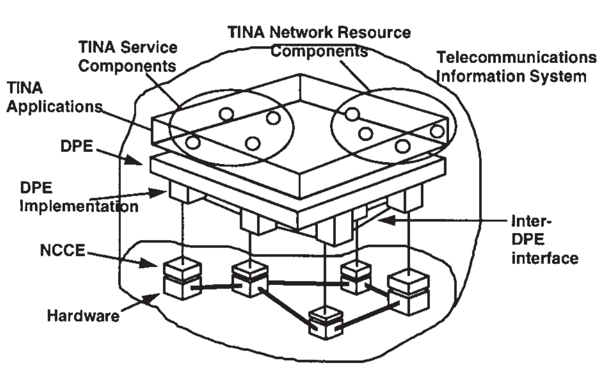
Functional Representations of TINA-C

As mentioned, the four architectures (service architecture, network architecture, management architecture, and computing architecture) are all clustered under the overall framework architecture. Some functionalities from AIN and ITU-T IN are noticed, but the depiction is done differently. TINA-C depicts functionality using an object-oriented approach from a distributed processing perspective. A service architecture view includes elements such as a session model in which call and connection separations are described. A subscription model is described showing the innovation of services and how they are accessed by customers.

With respect to the network architecture, a network resource model is described. This is similar to the set of FEs, FE relationships, and FE actions described in AIN and ITU-T IN. However, the TINA-C approach addresses a broader set of functionality in terms of multiple connections and multiple sessions for the multimedia services that are being addressed.

FIGURE 11

TINA-C—Business Model, Service/Resource Architecture



The management architecture is based on telecommunications management network (TMN) to address components such as configuration management and accounting management and uses TMN paradigms. ITU-T IN does not strictly address a management architecture, but IN CS-3 will address management aspects in more detail using TMN paradigms.

The computing architecture is where the distributed processing environment use of ODP, common object request broker architecture (COBRA), and object-oriented paradigms apply using the software architecture.

Figure 11 depicts the overall business model and service/resource architecture of TINA-C. Telecommunications services, information services, and management services are represented both as service and network components in a service architecture and a network architecture. These finally converge on both a computing architecture and a management architecture, in terms of how a service is offered. TINA is very broad in terms of the types of services that it addresses.

Figure 12 depicts a three-dimensional view of TINA. While AIN usually is shown in two dimensions to depict SCPs, SSPs, and IPs, TINA adds another dimension. This is both literally and figuratively true based on the approach used. Once again, TINA does not specify a physical architecture or set of physical entities. However, it does describe physical systems in terms of what would be a TINA-compliant system. For example, *Figure 12* shows that someone can deploy, define, or develop an SCP, which is using TINA-compliant paradigms or an IP or SSP. Those can be TINA-compliant physical entities, but TINA does not prescribe what those physical entities are; it is a function of the services to be provided. Nevertheless, there is a notion of having a set of objects that are available for reuse in whatever functions or different services are addressed. That is part of the specification process that TINA describes.

FIGURE 12

Overview of TINA-C—Architectural Separations

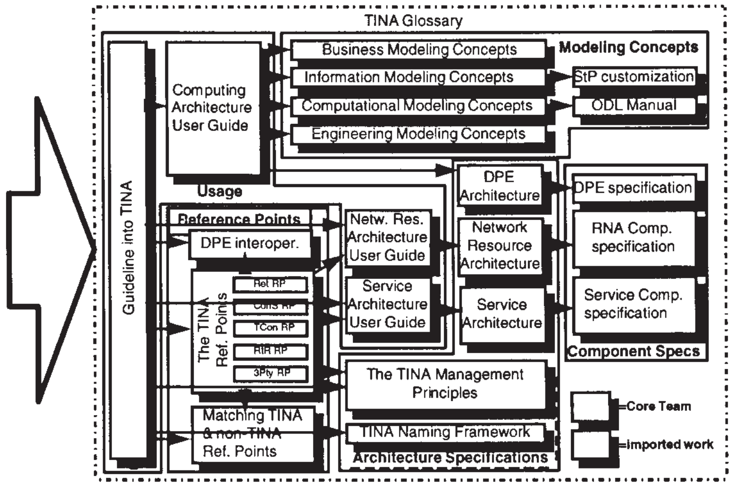


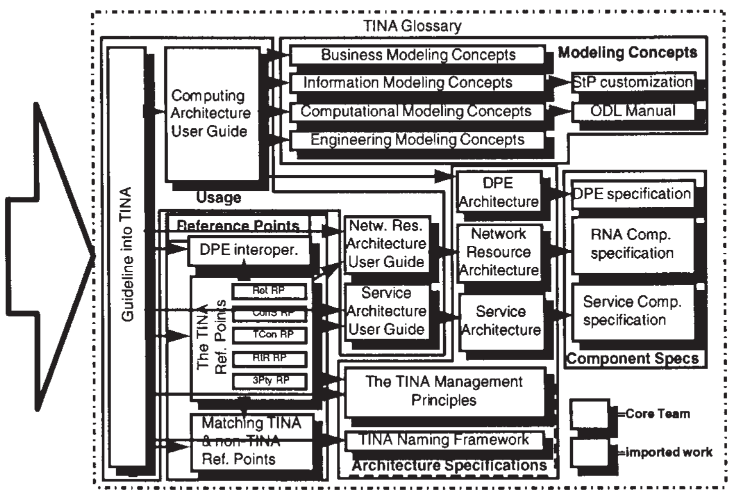
Figure 13 depicts the documentation that is available or being planned for TINA-C. Much of it is available from the Web site. The documentation describes the guidelines in TINA, which is an overall architectural view. There is a computing architecture view, a service architecture view, a management architecture view, and a computing architecture view. The current set of specifications that TINA is addressing is in Version 4, which is planned for completion by the end of 1997. It is very ambitious both in its depth and breadth.

Summary

AIN and IN are evolving in the context of supporting more services and access arrangements, both for wireless and broadband services. They address similar interfaces and similar physical entities. TINA is providing a set of paradigms and concepts that can be useful in the context of evolving both AIN and ITU-T IN. Some of these are being addressed in ITU-T IN CS-3 already where there are considerations and deliberations for the use of object-oriented methodologies. There are both formal and informal information exchanges between AIN and ITU-T IN, and TINA-C also has conduits into both the specification and standardization efforts of ITU-T IN and AIN.

FIGURE 13

TINA-C—Roadmap of 1997 TINA Documents



IP Voice Services

Bruce D. Jacobs
Counsel to the VON Coalition
Fisher Wayland Cooper Leader & Zaragoza, L.L.P.

David K. McGraw
Counsel to the VON Coalition
Fisher Wayland Cooper Leader & Zaragoza, L.L.P.

The development of the Internet and Internet protocol (IP)-based networks, including the development of IP voice services, is having a profound and beneficial impact on the United States and the world. Use of voice over IP networks is drastically reducing the cost of international communications, improving an imbalance in U.S. trade, and creating a foundation for broadband communications that have much greater capacity and functionality than is offered by the current public switched telephone network (PSTN). Gateways that permit users to access IP voice services with a telephone handset are providing the first Internet applications for people without computers.

The development of IP voice services promotes and is consistent with the goals of universal service. Not only is voice over IP making communications more affordable (particularly with respect to international service), but providers of IP voice services are willing to contribute to the support of universal service programs in a way that is competitively neutral among functionally equivalent services. Such contributions, however, should recognize that: (i) while other voice services may be functionally equivalent to IP voice services in some respects, IP voice services are properly characterized as "enhanced" or "information" services because of their intrinsic use of IP and computer processing; (ii) IP voice services, when separated from the provision of potential monopoly transmission facilities, should not be regulated; and (iii) IP voice services should not be required to participate in the current access-charge regime, which was developed for circuit-switched networks. Deployment of IP voice services on public networks for domestic use is proceeding slowly enough that any changes that must be made to establish competitive neutrality in universal service support can be accomplished without significant dislocations.

Background

Internet Protocol Generally

Transmission control protocol / Internet protocol (TCP / IP) is a set of rules that facilitates the communication of data among computers operating on a wide variety of networks with differing hardware configurations and operating systems, deploying different protocol-independent applications. This flexibility and the creativity of the millions of people who have made use of it, has made IP the common element in and the foundation for the phenomenal growth of the Internet, an interconnected group of thousands of computer-based networks.

On IP networks, all data, whether voice, text, video, computer programs, or numerous other forms of information, travel though the network in packets. Each destination in the system has a unique IP address, and packets are routed to their destination according to the address contained in a header. Data may be transmitted at the same time from one user to many users and data addressed to various users can share the same line.

IP networks trade increased use of computer processing for a decreased use of transmission facilities. As the cost of computer processing continues to decrease and the demand for communications bandwidth by consumers increases, IP systems will increasingly offer a more economical and substantially improved means for providing communication connections. By contrast, conventional telephone systems require more lines and a more complex topology. At the local level, the conventional telephone system works on the model that each customer's equipment must be connected by a continuous line to a telephone company switch, whether or not the line is actually in use. At the long-distance level, a continuous link must be established between each pair of users for the duration of a call, regardless of the actual information sent through that path.

IP networks also offer the potential of higher reliability than the circuit-switched PSTN. IP networks automatically reroute packets around problems such as malfunctioning routers or damaged lines. IP networks do not rely on a separate signaling network.

One of the key attributes of IP is its openness. Many networks, including the PSTN, operate as closed systems on which it is impossible for innovative developers to build new applications. (The failure of advanced intelligent networking (AIN) illustrates the problem of closed systems impeding the development of innovative products and services.) IP is a nonproprietary standard agreed on by a consortium of hardware and software developers, and is free to be used by anyone. As such, it permits entrepreneurial firms to develop new hardware and software applications that can seamlessly fit into the network.

Transmission and Reception of Speech Using IP

The first component in an IP voice system in which the user has a personal computer or is on a handset with an analog connection is the digitization of the speaker's voice. The next step (and the first step when the user is on a handset connected to a gateway using a digital PSTN connection) is typically the suppression of unwanted signals and compression of the voice signal. This has two stages: (i) the system examines the recently digitized information to determine whether it contains a voice signal or merely ambient noise and discards any packets that do not contain speech (this requires the system to be able to distinguish the properties of a human voice, which needs to be transmitted, from those of ambient noise, which can be ignored) and (ii) the use of complex algorithms to reduce the amount of information that must be sent to the other party. The development of increasingly sophisticated codecs is key to the success of suppression and compression.

Once the data has been compressed, it must be packetized and have protocols added. Some storage of data occurs during the process of collecting voice data, since the transmitter must wait for a certain amount of voice data to be collected before it is grouped together as a packet and sent through the network. There are also brief periods of data storage during the encoding and compression processes. Protocols are added to the packet to facilitate its transmission across the network. For example, each packet will need to contain the address of its destination, a sequencing number in case the packets do not arrive in the proper order, and additional data for error checking. Because IP is a protocol designed to interconnect networks of varying kinds, substantially more processing is required than in smaller networks, the network addressing system can often be very complex, requiring a process of encapsulating one packet inside another and, and as data moves along, repackaging, re-addressing, and reassembling the data.

When each packet arrives at the destination computer, its sequencing is checked to place the packets in the proper order. A decompression algorithm is then used to restore the data to an approximation of its original form, and clock-synchronization and delay-handling techniques are used to ensure proper spacing. Since packets of data travel through the network along various routes, they do not necessarily arrive at the destination in the proper order. Therefore, incoming packets are stored for a time in a jitter buffer to wait for late-arriving packets. The length of time in which data are held in the jitter buffer varies depending on the characteristics of the network.

On the open Internet, a large percentage of the packets can be lost or delayed, particularly during periods of congestion. In addition, some packets must be discarded because errors have occurred during transmission. These lost, delayed, and damaged packets cause substantial deterioration of sound quality. In conventional error-correction techniques used in other protocols, incoming blocks of data containing errors are discarded, and the receiving computer requests the retransmission of the packet, thus the message that is finally delivered to the user is exactly the same as the message that originated. Since IP voice systems are more time-sensitive and cannot wait for retransmission, more sophisticated error detection and correction systems are used to create sound to fill in the gaps. This process stores a portion of the incoming speaker's voice. Then, using a complex algorithm to "guess" the contents of the missing packets, new sound information is created to enhance the communication. Thus, the sound heard by the receiver is not exactly the sound transmitted, but rather portions of it have been created by the system to enhance the delivered sound.

IP Voice Configurations

The processing that was just described can be done either at the user's premises or at a service provider's premises. Initially, the typical configuration, shown in *Figure 1* has involved a PC located at the user's premises. This typically involves all the parties to the call getting on-line and connecting in a "chat room" established by a service provider. With the introduction of gateways (specialized computers that at a minimum provide an interface between IP network and other networks such as the circuit-switched PSTN), it is increasingly possible to use other equipment such as a common telephone handset to use IP voice.

The gateway can be located either at the user's premises, as shown in *Figure 2,* or at a remote site accessed by conventional phone lines, as shown in *Figure 3.* In this configuration, the gateway may accessed by dialing either a local telephone number or a toll-free number. In the former case, the end user's local exchange carrier (LEC) terminates the call at the gateway, and in the latter case the LEC carries the call to the interexchange carrier (IXC) point of presence (POP), and the IXC terminates the call at the gateway. In the future, there are likely to be hybrid appliances that have more functionality than a telephone handset but less than a PC. Such a device may have a computer screen and provide the user with text and video information simultaneously with the audio content of the voice conversation. In some cases, the use of a hybrid appliance may negate the need for a gateway; in others, a gateway may still be useful to provide additional computer power (see *Figure 4*).

In all of these cases, unless the user has a continual IP connection, it will be necessary for the party originating the communications to first establish the IP connection (either through the use of a dial-up modem or by using the keypad on a phone). Then, a separate process must occur to establish a connection with the terminating party or parties (such as dialing the additional phone numbers, in the case of a gateway connection.)

The ideal is for all users of IP voice to have a continual connection to an IP network. This permits optimal use of the flexibility provided by IP to manage broadband communications. For the foreseeable future, however, there will always be some users who can communicate with IP networks only by using legacy systems, including conventional telephone handsets. For these users, gateways will provide the only available access (see *Figure 5*).

IP Voice Services

The development of IP voice services is just beginning. The first two-way application was an attempt to reproduce speech as accurately as possible under the worst possible conditions, involving little bandwidth and high error rates. Applications have been developed that permit users with a personal computer to combine speech with other forms of data, including video, text, and graphics. Thus, for instance, two people can see each other while they talk and can trade text, jointly process a document, or draw and review diagrams on an electronic whiteboard. More advanced collaborative applications take advantage of the flexibility of packet-switching to

FIGURE 1

PC-to-IPN (Through Dial-up ISP)

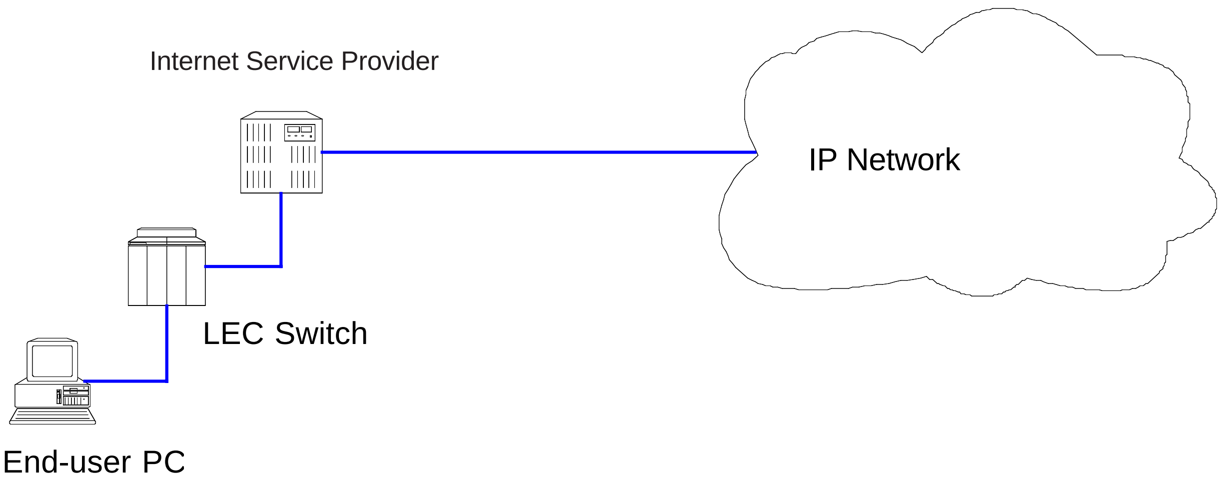


FIGURE 2

Handset-to-IPN (Via PBX/Gateway)

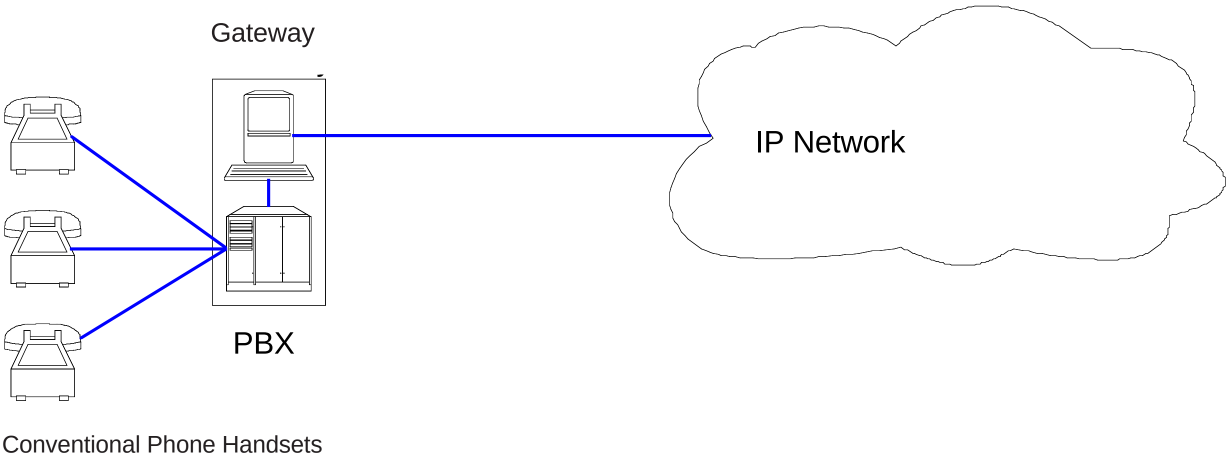


FIGURE 3

Phone-to-IPN

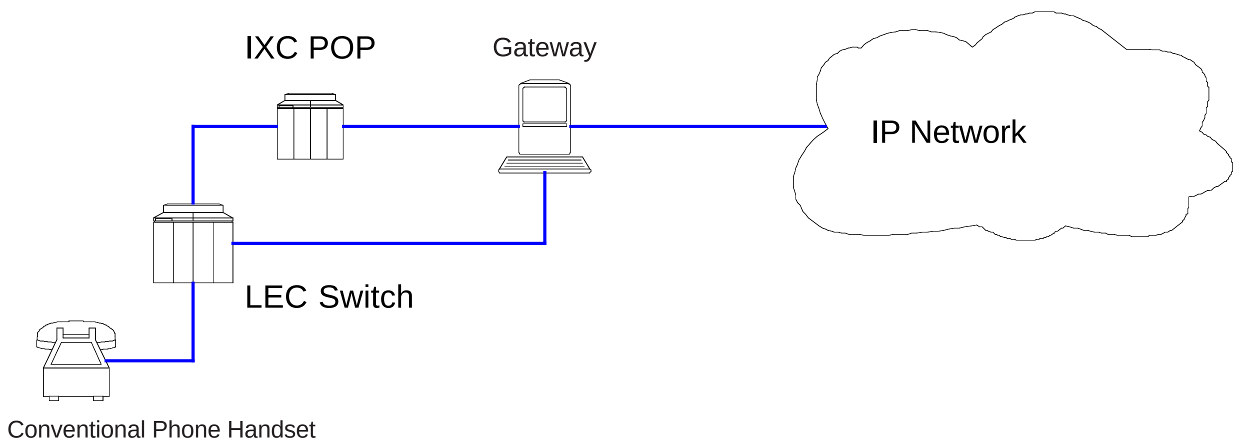


FIGURE 4

Hybrid Appliance-to-IPN (By Dial-up IPVSP)

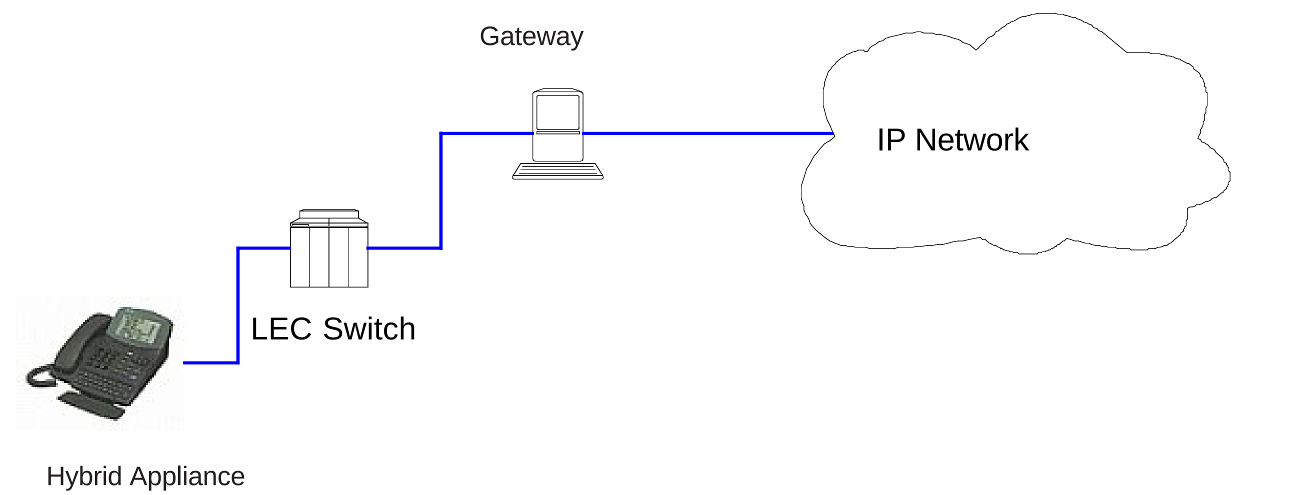
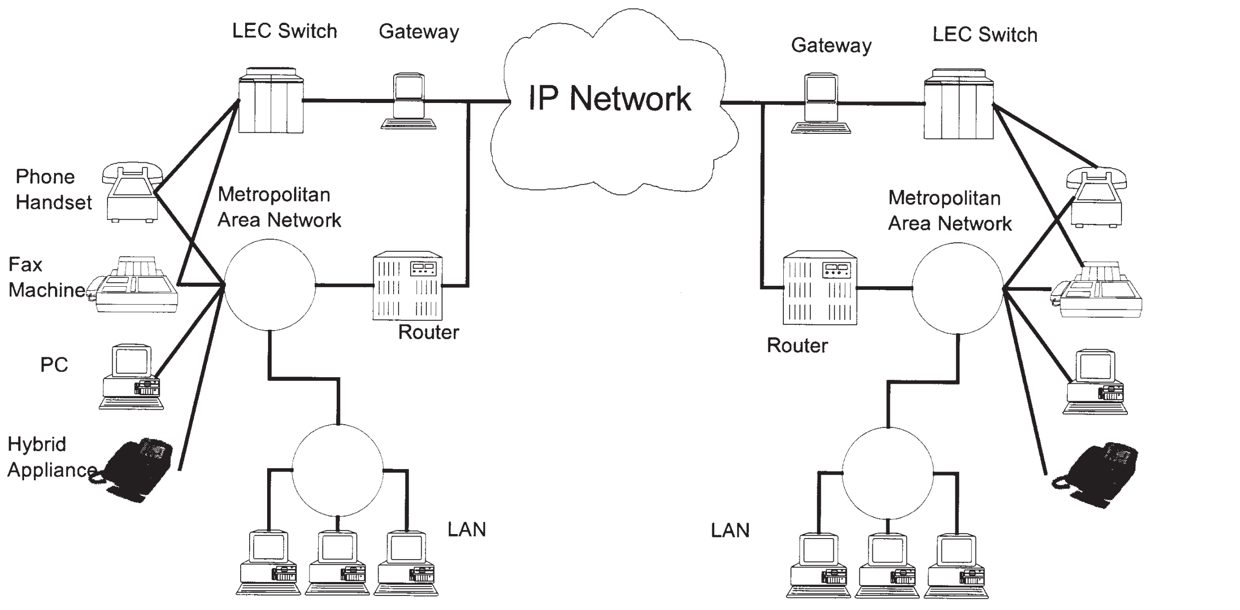


FIGURE 5

The Future (IP Dedicated Connection)



permit all of this information transfer to be shared simultaneously by many different users. The potential of these applications for education, business, and consumer use is extraordinary and goes far beyond what is practical in a circuit-switched environment using the PSTN alone.

These applications do not require the use of a personal computer. With gateways, users that do not have a PC nonetheless can take advantage of at least a subset of these services. As with PC-based IP voice services, the first gateway applications to develop have been simple ones that focus on providing the best quality with the lowest quality inputs, but more functional applications are being developed. By using the computer-processing capability of the gateway, the user of a simple handset can gain tremendous additional functionality, typical of IP networks. For instance, gateways may be used to provide inexpensive voice-conferencing that could be very useful for educational applications involving a few people or hundreds of people, something that would be impossible or far too expensive on the PSTN. The likely situation is that some users would access such a classroom using a PC, while others would do so using gateways and a simple telephone handset.

Gateways may be used to facilitate retrieval of information from databases on Web sites or elsewhere, using either handset-generated tones or voice recognition technology. Gateways that contain voice recognition software may be used by the hearing impaired with a TTY device to translate between text and speech, without the need, as is the case now, for a human relay. A voice recognition program could convert the hearing speaker's voice into text to be transmitted to the deaf user's TTY, and the deaf user's text could be converted to speech by voice synthesis software. (As voice recognition and language translation technology improve, it is not far-fetched to imagine the use of gateways to interpret for users that speak different languages).

Voice messaging systems and store-and-forward programs may be added to a gateway. All of the functions currently available to electronic mail users, such as the ability to store pre-programmed mailing lists and distribute a single message to numerous recipients, and the option to save, copy, and forward messages may be performed at the gateway level. Users may access a gateway when they contact directory assistance to permit them to create their own personalized directory, which might be retrieved and used with voice commands.

The IP Voice Services Business

The deployment of public IP voice services is being driven largely by: (i) the ability to bypass international accounting rates by using value-added data networks and (ii) building the foundation for a future in which IP networks are a ubiquitous way to efficiently manage information transfer, including voice, video, and data, between two people or among many people.

Almost all of the revenue being generated by the use of IP voice services today is from international services. The current international accounting rates regime for the PSTN supports rates that are in some cases far above costs, which causes a multibillion dollar annual U.S. trade imbalance. By using IP networks, which are widely-recognized by many foreign governments to be value-added networks that are not subject to accounting rate settlements, IP voice service providers can charge far less than those using the PSTN. The availability of this competition helps put pressure on foreign governments to reform their accounting rate structures and move them closer to true costs. In this respect, the FCC has championed the development of IP voice networks.

In its typical configuration, an IP voice service provider that provides international service to and from the United States does so from a single location, probably colocated with a PSTN switch. In the case of a call originating in the United States, the IP voice service provider routes the traffic to its gateway usually through an 800 number. In rare cases, there is sufficient traffic in the vicinity of the switch to justify the use of local business lines or integrated services digital network (ISDN) primary rate interface (PRI) lines to connect users to the service provider's gateway. Similarly, to terminate traffic, the service provider in most cases uses an interexchange carrier that uses feature group access to move traffic from its gateway to the end user, or again in relatively rare cases uses local business or PRI lines.

The development of public IP voice services is also taking place as a component of the deployment of IP networks more broadly as part of a general recognition of the value of broadband, IP-based networks. Networks that use IP offer an inherent efficiency and functionality for communications, particularly those that combine different kinds of data, including voice. In tomorrow's world, much of our electronic communications will take place over IP networks and such networks will replace even many local circuit-switched networks—local area networks (LANs) will be connected to metropolitan area networks that will be connected to wide area networks (WANs), using wired and wireless transmission paths. In that world, higher-speed bandwidth will be more abundant, connectivity will be continual, and users will be able to dynamically assign their bandwidth to a variety of simultaneous streams of information over a single connection, including voice, video, and other forms of data, all subject to the user's information management systems. Several carriers are deploying such networks nationally and on the local level in anticipation of a continued growth in demand for a variety of services, including voice. The deployment of gateways is a logical component of this deployment, since they make the network more accessible and increase network traffic.

There is a myth that the offering of these services is being driven by the benefits of avoiding interstate access charges through the characterization of these services as enhanced. In fact, the cost of deploying gateways and acquiring local PSTN connections through the purchase of local business lines is roughly three cents per minute, about the same as current access charges. The highest costs are for the gateways themselves, which are currently priced at approximately $500 per port, and for terminating access, which is billed per minute. Moreover, this estimate assumes that it is possible to originate and terminate sufficient traffic using local lines—an assumption that is very optimistic under current conditions. In many cases, it would be necessary to provide access using interexchange facilities. With access charges moving closer to cost, the deployment of gateways as a form of access charge arbitrage becomes even more unrealistic.

The other area in which IP voice services will grow rapidly is over private networks. Large corporations that operate intranets will add voice traffic to those networks, including the use of gateways for communications with traffic on their PBX. Some offices may replace their PBX entirely and put their voice traffic on a LAN that uses a gateway to connect to the PSTN.

There are a number of impediments to more rapid growth of public IP voice services. These include:

- the lack of availability of scalable products—at present, the largest gateway available offers 96 ports (modern switches have tens of thousands of ports) and there are no network management and billing systems that are capable of handling large volumes of traffic

- the lack of necessary standardization, to permit one user to communicate with another user—currently, there are several different vendors each offering their own nonstandard software and gateways

- frequent problems with congestion on the Internet that makes quality of service unreliable—a problem that will not be solved until either the Internet is substantially upgraded or new, managed networks are deployed at considerable expense, something which many existing carriers are unlikely to do given their enormous existing investment in legacy systems

- efforts of existing monopolists to ban or impose inequitable costs on IP voice services—several foreign governments and PTTs have tried to prohibit or restrict the deployment of IP voice services

Legal and Policy Issues

Public Benefits of IP Voice Services

As demonstrated above, the continued deployment of IP voice services is in the public interest. IP voice services are having a positive impact on international communications and U.S. balance of trade, facilitating the deployment of integrated wideband data services for which the public has shown tremendous demand, providing an opportunity for continued innovation through the use of an open architecture, and in the case of gateways, providing the first Internet application for people without computers.

Impact on Universal Service Support

In many ways, the provision of IP voice services in and of itself furthers many of the goals of universal service. For many people, such as immigrants communicating with family members back home, IP voice services make international voice communications affordable. For people without computers, IP voice gateways provide access to the Internet. As computer processing power increases even more, IP voice services will make communications even more affordable and universal.

Defining IP voice services as "information services" and not "telecommunications services" will not harm universal service goals. As discussed above, the deployment of IP voice in the near-term will not affect the revenue base for universal service contributions. Virtually all of the IP voice services being provided are either for international communications or on private networks that have no revenue. Under the current law, companies that provide only international services are not subject to universal service contributions. Moreover, the international services generally rely on interexchange facilities for originating or terminating access to or from the PSTN in the United States.

IP voice service providers are prepared to contribute in a competitively neutral way to the support of universal service programs. It is critical, though, to establish such a mechanism without defining IP voice services as "telecommunications services." As discussed immediately below, IP voice services meet the definition of both "enhanced services" and "information services." If, however, the FCC were to characterize IP voice services as "telecommunications services" in order to ensure that IP voice service revenue is included in the funding base for Universal Service, it would make it very difficult for U.S. companies to provide IP voice services in many foreign countries without being subject to the accounting rate regime that the United States is trying to reform.

IP Voice Services Are "Enhanced Services"

The origins of the Commission's "enhanced services" classification was its concern that facilities-based common carriers entering the data processing services market might either pass along the cost of their data processing investments to captive ratepayers or take advantage of their telecommunication facilities to unfairly disadvantage their data processing competitors. *Regulatory and Policy Problems Presented by the Interdependence of Computer and Communication Services and Facilities*, 21 RR 2d 1591, 28 FCC 2d 267 (1971) *(Computer I)*. By 1980, the differentiation between "data processing" and "telecommunications" had become unworkable, and in *Computer II* the Commission refined the distinction by developing the categories of "enhanced service" and "basic service." *Amendment of Section 64.702 of the Commission's Rules and Regulations (Second Computer Inquiry)*, 44 RR 2d 669, 77 FCC 2d 384, (1980) *(Computer II)*.

Against this backdrop, the Commission defined "basic service" as the provision of "pure transmission capability over a communications path that is virtually transparent in terms of its interaction with customer-supplied information." *Computer II* at 420. Enhanced service, on the other hand, refers to services, offered over common carrier transmission facilities used in interstate communications, which employ computer processing applications that act on the format, content, code, protocol or similar aspects of the subscriber's transmitted information; provide the subscriber additional, different, or restructured information; or involve subscriber interaction with stored information.

Id.; see also 47 C.F.R. § 64.702. Basic services were to be offered under tariff, according to *Computer II,* while enhanced services were unregulated. In the Communications Protocols decision, the Commission noted that there is a continuum between services that are clearly basic, such as "transmission with no changes in ... electrical signals," and clearly enhanced services, such as "creation, deletion, and alteration of information." *Communications Protocols under Section 64.702 of the Commission's Rules and Regulations,* 55 RR 2d 104, 95 FCC 2d 584, para. 3 (1983). In *Computer II,* the Commission emphasized that:

the carrier's basic transmission network is not to be used as an information storage system. Thus, in a basic service, once information is given to the communications facility, its progress towards the destination is subject only to those delays caused by congestion within the network or transmission priorities given by the originator.

Computer II at para 95.

The purpose of these categories was to ensure that the "licensed transmission facilities of a carrier are equally available to all providers of enhanced services," to minimize "the potential for a carrier to use its transmission facilities to improperly subsidize an enhanced data processing service without detection," and to benefit consumers by "enabling resale entities to custom-tailor services to individual user needs. *Id.* at para 87. Thus, the goal of the Commission in creating the categories was to enhance competition and foster increased technological development in the computer industry by keeping it free from regulation. Concern about regulation of enhanced services extended to regulation by state governments. *Computer II* at para 7. *See also Amendment to Sections 64.702 of the Commission's Rules and Regulations (Third Computer Inquiry),* 62 RR 2d 1662, 2 FCC Rcd 3072, at paras 18, 46 (1987) *(Computer III).*

The heart of the Commission's concern was that large carriers might abuse their control over potential bottleneck transmission facilities. *See e.g., Computer III,* 2 FCC Rcd at 3077, 3111 n. 25, 3112 n. 62; *see also, Frame Relay* at para 42. Under the "contamination theory," a service that bundled both basic and enhanced services remains unregulated as long as the basic services are not using transmission facilities over which the service provider might have market power and restrict supply.

The relevant terms for the purpose of the universal service discussion are "information" and "telecommunications" as defined by Congress in the Telecommunications Act of 1996 rather than "enhanced" and "basic" services. For the purpose of this paper, however, we are relying on the Commission's characterization of "information services" as a broader category than "enhanced services." *Implementation of the Non-Accounting Safeguards of Sections 271 and 272 of the Communications Act of 1934,*5 CR 696, 11 FCC Rcd 21905, at para. 103 (1996).

Since establishing the basic/enhanced dichotomy, the Commission generally has reviewed technology on a case-by-case basis to determine its classification. Since *Computer II,* protocol processing has generally been considered an enhanced service. While the Commission did permit AT&T to offer transmission of data over a packet-switched network following X.25 protocols as a "basic service," it emphasized that it was applying the definitions in *Computer II* "in a flexible manner so as to ensure that the transitional introduction of new technology in basic services is not inhibited." *Application of AT&T For Authority under Section 214 of the Communications Act of 1934, as amended, to Install and Operate Packet Switches at Specified Telephone Company Locations in the United States,* 94 FCC 2d 48, at para. 5 (1983). When protocol conversion services are offered, those services have generally been held to be enhanced. *Petitions for Waiver of Section 64.702 of the Commission's Rules by Pacific Bell et. al.,* 58 RR 2d 1664, 100 FCC 2d 1057 (1985). In 1985 the Commission noted that "packet switching is heading rapidly towards integration with facilities for conventional telephone service," and suggested that when the integration occurred, new regulatory approaches might be warranted. Petitions for Waiver of Section 64.702 of the Commission's Rules by Pacific Bell et. al., 58 RR 2d 1664, 100 FCC 2d 1057 (1985) at para 77. In *Computer III,* the Commission rejected a proposed shift in its definition of enhanced services that would have included a "change in content" test, but rather decided to continue to label protocol conversion as an enhanced service. *Computer III,* at para. 68.

In the *Frame Relay* case, the Commission was faced with the proposal of AT&T to offer a service that combined the transmission of data over a frame relay network, and the conversion of protocols. *Independent Data Communications Manufacturers Association and AT&T Petition,* 1 CR 409, 10 FCC Rcd 13717 (1995) *(Frame Relay).* The Commission concluded that AT&T's proposed service, which included accepting data with protocol information already attached by the customer, and merely transported that data across its frame relay network, the offering should be considered a basic service. *Id.* at para. 35. On the other hand, protocol conversion is, as it has always been, an enhanced service, and thus the Commission required AT&T to unbundle the two services. Id. at para 22. The basic frame relay service was to be offered on a common carrier basis under tariff, while the protocol conversion itself was an unregulated enhanced service. *Id.*

In the *Sections 271 and 272* case, the Commission soundly rejected the notion that the term "information service," a roughly equivalent term, only refers to a net conversion of content between one end of the transmission and the other. *Implementation of the Non-Accounting Safeguards of Sections 271 and 272 of the Communications Act of 1934, as Amended,* 5 CR 696, 11 FCC Rcd 21905, at para. 104 (1996). The Commission stated that "information services" do not merely refer to "services that transform or process the content of the information transmitted by the end-user," but rather that "the statutory definition makes no reference to the term 'content,' but requires only that an information service transform or process 'information.'" *Sections 271 and 272,* at para. 104. Therefore, even if the message received has the same meaning to the end user as was intended by the sender, the underlying process could still be considered an information service, so long as the data is processed in some fashion between the sender and receiver.

While the Commission might need to conduct a more thorough review of the technology before it could make any decisions on the issue, it is apparent that IP voice services are "enhanced." Internet service providers, including Internet voice service providers, process data, convert it from one form to another, add protocol information, process protocols, and perform a myriad of other tasks that constitute an enhanced service. As such, IP services, including IP voice services, fit within the definition of enhanced service established by the Commission and upheld on numerous occasions. It is clearly neither "pure transmission capability" nor "transparent in terms of its interaction with customer-supplied information." The processing performed on voice transmissions carried over the Internet is qualitatively different from that of conventional switched voice systems and from common carrier data transmission services that have been held to be "basic" services.

First, the suppression and compression used to enhance the efficiency of the system is sufficient to deem the system enhanced. *Computer II* included "bandwidth compression techniques" among the list of techniques that do not constitute an enhanced service (*Computer II* at para. 19), but there is a significant difference between the bandwidth compression techniques employed in 1980 and the more complex process currently used in voice over IP. The suppression and compression techniques commonly used in the voice over IP industry include the detection of whether or not the signal contains voice sound, and only transmitting only those portions it determines to be voice. It is important that these programs are not merely deleting silence, but are deleting nonvoice sounds that may have been transmitted across the channel in a conventional voice system. Thus, the system is actually interacting with the information in the incoming data stream, analyzing its content, and deleting portions that it determines to be unnecessary. This advanced process does "employ computer processing applications that act on the ... content ... of the subscriber's transmitted information." See *Computer II* definition at 420. Such a process also involves a "deletion" of information, which, according to the *Communication Protocols* case, makes it a clearly enhanced service. *Communications Protocols* at para 3.

Second, as with all of IP, the packetization and adding of protocols makes this an enhanced service. As noted above, the Commission has generally held such protocol processing to be an enhanced service. *See, e.g., Computer II* at para. 99; *Communications Protocols under Section 64.702 of the Commission's Rules and Regulations,* 55 RR 2d 104, 95 FCC 2d 584, at para 3 (1983). In cases in which data transmission systems have been held basic services, such as that in the *Frame Relay* case, the service was generally offered by a facilities-based common carrier that did not add or remove the control information, but merely used it to route data through the network. In those services, the customer generally created the protocol headers and trailers, gave them to the common carrier for transmission across its proprietary lines, and removed them at the end. In IP voice services, the service provider, generally not a facilities-based provider, is packetizing the data and adding protocol data, then releasing it for transmission. Thus, the IP voice service provider is involved in the addition, deletion, and processing of information in a manner not done by AT&T in the *Frame Relay* case.

Third, IP voice services employ storage of data. At the transmission end, data is stored during the transmitter recording process, and again briefly during the encoding and compression processes. Incoming data at the receiver's end is stored for a period of time in a jitter buffer. The purpose of this storage is to properly order the information and wait for late-arriving packets. This storage of data is not caused by the congestion of the network or transmission priorities of the originator.

Fourth, the voice reconstruction that occurs to compensate for lost packets and transmission errors in a voice-over-IP system makes it enhanced. While the *Computer II* definition of basic service included "error control techniques," there is a qualitative difference between the simple error control techniques used over the switched network and the systems used by Internet voice systems to enhance the delivery of voice over the IP network. In conventional error correction techniques, such as those envisioned by the *Computer II* definition, include checking incoming blocks of data for errors, and requesting the retransmission of any that contain errors. Thus, in conventional error control techniques, the message that is finally delivered to the user is exactly the same as the message that originated, even if some errors resulted during the original attempt to transmit the data. The typical IP voice error detection and correction system is quite different in that it uses retrieval of stored data and creation of new data to enhance the communication. As described above, this process of error correction not only processes and transforms the information, it also retrieves stored information and adds new content that did not exist in the original. The sound heard by the receiver is not exactly the sound transmitted, but rather portions of it have been created by the system to enhance the delivered sound. Thus, in the words of the *Protocols* decision, this is a "creation" of information.

The distinction between basic and enhanced services was originally whether the carrier offered a simple, pure, transparent communications path, or whether the provider used computer processes to add value. The focus of the distinction should be on whether the system adds value, not in a simplistic comparison of whether the format of the message is the same from the sender's perspective as the receiver's perspective. IP voice systems add value—by increasing efficiency and by providing a capability for integrating voice services with other forms of data—and therefore they should be considered enhanced, or "value added," services.

One of the key characteristics of enhanced services is that they combine basic transmission services and computer processing in a way that is often difficult or impossible to distinguish. That is the case for IP voice services, even if one were to regard simple handset-to-handset service through a gateway as a basic service. Particularly for the service provider operating the terminating gateway, it will often be impossible to determine if the packets it is receiving were transmitted by a gateway or a personal computer, and absent a regulatory requirement there would be no reason for the gateway operator to try to differentiate between the two.

The goals sought by the creation of the "enhanced service" category would be furthered by the inclusion of IP voice services in the category. The classification of service providers that offer access to the IP networks using transmission facilities that are generally available for lease poses no threat of improper cross-subsidization or market influence. On the contrary, as the Commission sought to promote when it established the category, the classification of IP networks as enhanced services would protect against unnecessary regulation, increase competition, and encourage the development of new technologies.

Access Charges

IP voice service providers should not be forced into the current regime for interstate access charges. The predominant use of IP voice—for international traffic—typically results in paying the same interstate access charges as is being paid by other entities that are providing international voice communications using the PSTN. To the extent that the current regime contains implicit subsidies to reduce the cost of local service, IP voice service providers will not carry sufficient traffic in the near future to undermine that aspect of universal service support, certainly before the regime is changed to make access cost-based.

As discussed above, the myth that IP voice is a form of access-charge bypass is just that—a myth. No rational service provider deploys gateways as a means of access-charge arbitrage and none appear to be doing so today. Perhaps more to the point, it would be unfair to require IP voice service providers to pay for a form of access (circuit-switched) that includes many elements that they do not need and does not provide the form of local connection (packet-switched) that in fact they do need.

Almost all IP voice traffic today that uses the PSTN and involves a per-minute charge to the subscriber is international traffic that predominantly uses toll-free 800 numbers for calls originating in the United States through a gateway and similar interexchange facilities for the termination of calls into the United States through a gateway. For this traffic, therefore, the service generates the same access charge revenue as would be generated if the traffic was handled more conventionally.

To the extent that the traffic is sufficient to aggregate it at and use local business or PRI lines, the service provider still pays for the use of the local PSTN, particularly for termination of traffic, which almost always involves a significant per-minute charge. It is generally understood that these business lines (particularly PRI lines) are priced above cost. (With access-charge reform, the cost of terminating a call on local business lines increasingly exceeds the cost of termination using feature group access, where much of the reform is to focus. It is completely conceivable then that IP voice service providers will locate their gateways in only a few key locations in a region and continue predominantly to use interexchange facilities to originate and terminate traffic.) Moreover, as end users, IP voice service providers pay multi-line subscriber line charges on their business lines and at least indirectly pay primary interexchange carrier charges (PICCs).

The current Part 69 access charge system unfairly requires IP voice service providers to pay for features and functions of the local exchange that they do not need. These unnecessary rate elements include: access tandem, tandem transport, local switching, and PICCs. IP voice service providers would not need switching and transport services if the local exchange network was capable of identifying IP traffic prior to the call being delivered to the switch. IP voice service providers also do not need equal access (1+ dialing) or trunk side signaling. Historically, the Commission has appreciated that the kind of tandem dialing that is required to use public IP voice services justifies paying for less expensive access. The current model also mandates that the access purchaser acquire services in a minimum 64-kbps channel without giving the purchaser the ability to maximize the use of that channel.

IP voice service providers would prefer, rather than leasing local PSTN lines, to be able to lease the type of packet-switched routing and transport that would permit them to provide an optimal service. In practice, this access could be provided by the local exchange carrier through newer, more efficient transport technologies such as a metropolitan area network (MAN). The MAN would not need a loop from every customer location to the central office, as a circuit-switched network requires. The customer's unique IP address can be used to direct voice, data, video, or any other manner of communication to the appropriate destination.

To encourage the development of MANs that facilitate the deployment of IP networks, incumbent local exchange carriers should be encouraging IP voice service providers to use trunk- side access arrangements (such as T-1 based ISDN primary rate interface service). At present, however, use of these services is deterred because digital T-1 lines and other trunk-side connections are only offered at premium rates. For example, a hunt group of 24 analog voice lines is priced as much as 50 percent less than equivalent trunk-side connection.

The Commission should also require incumbent local exchange carriers to offer Part 69 access elements on an unbundled basis. IP voice service providers could purchase loop sub-elements that would permit the local exchange carrier to identify data traffic, packetize it at the end office and transmit it to the IP voice service provider in a data-friendly packet environment. Such an arrangement would eliminate the need for current local switching or transport elements. If the incumbent local exchange carriers offered unbundled loop sub-elements, data-friendly technologies could be deployed by competing service providers. (Traditional Internet service providers have taken a similar position in response to complaints by local exchange carriers that web-browsing is causing congestion at the local PSTN. In the case of IP voice services, however, congestion is not even an issue, since users of IP voice services do not typically spend the hours at a time "on line" that is characteristic of web-browsing.)

To further promote the development of data-friendly local networks, the Commission also should require equal access and interconnection for competitive packet services. Packet service providers should have competitively neutral access to data traffic originating on the incumbent local exchange carrier's network. The Commission should also revise its collocation rules to eliminate unnecessary restrictions.

For further information contact the Voice on the Network Coalition at (510) 277-8110 or *www.von.com.*

Intelligent Contact Management: The Missing Link Between Information Technology and Productivity

Jeff King
Vice President, Business Development
COMPUS Services Corporation

Abstract

Expediting resolution of system problems and accelerating recovery from system downtime has become exceedingly expensive and complicated in the corporate computing arena. Effective corporate contact management infrastructures require a significant level of automation to handle the number of organizations and individuals demanding notification of system and application availability. Today's information technologists understand the value and impact of communication during computing interruptions.

This paper investigates the key criteria for increasing the volume, type, and quality of message delivery through intelligent contact management without disproportionate cost increase. A new approach to contact management has emerged that provides the entire corporation with the capability of effective and effortless message delivery and contact device management for all systems under a single platform.

Notification Issues

When system interruptions occur, a major challenge is getting the right people in front of the problem and giving management and users a timely and meaningful status. There are a number of reasons for this challenge. First, technical personnel are not professional communicators; they are individuals who resolve problems. Second, when a system has a problem, the error message is only meaningful to the technician. It has little or no informative value for management or users. Finally, the timeliness and depth of notification is insufficient to minimize the impact on productivity. As a result, the time and money lost from system outages is unnecessary.

Notification Strategies

Historically, notification strategies have been based on two delivery mechanisms, which have met with varying degrees of success depending on the tenacity of the organization responsible for notification. The two primary message delivery mechanisms are:

1. Pagers, which have been the backbone for notification because they are so widely used and do not require much organization beyond a list of pager numbers and a telephone.

2. E-mail, which has been popular for the desktop-computing professional, is a logical form of contact considering their environment.

These methods were often adequate for initial contact management strategies. However, as organizations move toward highly distributed enterprise-wide systems, these methods are not only highly inefficient but can greatly undermine the productivity with ineffective communication.

Recent advances in wireless communication technology have consolidated the pager, cellular phone, and e-mail into a single device. This technology is exciting, but purchasing new improved contact devices alone will not dramatically improve existing contact management strategies. As the corporation grows and business processes become more complex, new or more contact devices are not the answer. The situation requires a solution that automates the manual contact management process by tapping into network management systems and corporate applications to centralize message delivery and resolution support. A corporation must evaluate several important issues before implementing a contact management infrastructure.

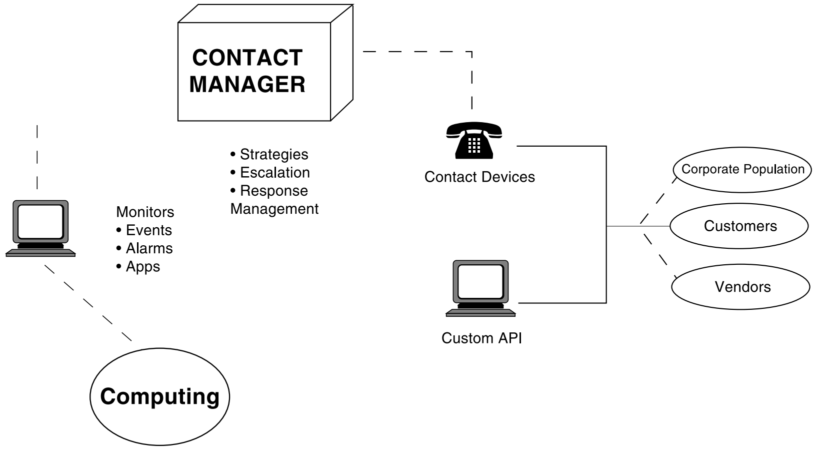
Maximum Contact Capability—Minimum Investment

The systems' platforms used by today's corporation can require unique levels and types of support. This is a result of business demands from the products or services provided by the corporation. Each system supports a segment of the business that must be supported based on its schedule and method of operations. IT managers must adapt system availability and recovery procedures to the needs of each business segment.

Beyond system technical support, the IT manager can expect diverse notification requirements from each organization to be as varied as their system requirements. While paging and e-mailing may seem to be the universal contact approach, it

FIGURE 1

Intelligent Contact Management



may not be up to the task in the future. Complex message types, time frames, and distribution agendas will be difficult to accommodate. System managers need contact management support to give them a common method of delivery with a wide variety of automated options to configure the message and contact procedure.

Keeping pace with business processes, reorganization, and technology upgrades is a staggering proposition. Building the confidence of the user community in such dynamic environments can only be accomplished through performance and communication. Contemporary contact management infrastructures must address the demands of new technology and keep the corporations' investments in contact devices fully utilized.

The corporate objective is to acquire a platform that will use all of the contact devices that a corporation has at its disposal. This application must provide APIs to address any new technology on the horizon. As a result of the implementation of a platform of this type, maximum contact capability with minimum investment would be realized.

Reliability

Contact management systems must be reliable. How can management have confidence in decisions they make when information is missing? Contact management strategies that are not reliable usually create problems bigger than the ones being solved, leaving managers to say, "If only we had known sooner." This statement means that productivity has been or will be lost.

When an IT organization is attempting to communicate with the user community or vice versa with passive devices such as pagers and e-mail, it can be frustrating. This is particularly true in time-critical situations when next-step decisions must be made quickly. The same frustration exists with mediocre contact management procedures. This predicament often makes the sender feel as though they are on a desert island and just wrote a note, put it in a bottle, and tossed it in the ocean. In both cases, it is unclear if anyone got the message. This inability of manual contact management procedures to automatically validate and escalate contact means making decisions without all the information that should be available.

The new intelligent contact management applications will solve this problem with response management and automatic escalation routines to ensure contact and message receipt is successful in a predetermined time frame. The responsibility of "hunting down" people via pagers and call list is eliminated.

Intelligent contact management provides control over six specific areas:

- who or what is to be contacted
- when the contact is to be initiated
- how the contact is to be initiated
- what information is to be delivered
- the bidirectional communication for each event
- alternatives when initial contract strategy is unsuccessful

Intelligent contact management applications will support all contact devices in the commercial marketplace and provide APIs to extend the interfaces to analog and digital equipment of the corporations choosing. This gives the enterprise the ability to take advantage of all technologies in the corporation and maximize the effectiveness of the entire management team. An intelligent contact management system is positioned to interface with all entities.

In *Figure 1*, instead of many applications each with their own unique method of providing contact technology, the corporation has a contact management platform to house strategies and interfaces with all corporate contact devices and applications. This delivers an element of control and reliability to a single point. The reliability in this case is based on deploying best-in-class contact management standards. The administrators now have the opportunity to fine tune contact management procedures that can be instantly implemented across all platforms. This architecture permits end users to access their own contact records without interfacing with administrators for any given application. The end result is reduction of maintenance expense and the reuse and redistribution of expertise throughout the corporation.

TABLE 1

Requirements for Intelligent Contact Management

Corporate Requirement	Application Must Provide
Archive	*A "printable" history of contact attempts and event activity.*
Database	***For contact strategies (Actions) and user profiles.***
Customization	*APIs for interfaces to various monitors and applications.*
Flexibility	***A script language embedded for ad hoc development.***
Escalation	*Nested escalation routine support.*
Persistence of Contact	***Contact device support of Wireless and Wire line interfaces***
Notification Types	*Formatting for various message types (Fax, Email, etc.).*
Analysis	***Reports providing various data views (user, devices, events, actions etc.).***

Finally, this architecture provides a reliable design, which allows persistence of contact across many devices for each user. Automatic escalation is executed in the event that there is failure to contact a targeted user. This ensures communication is completed to every department and organization for every event. This persistence of contact is the foundation for reliability.

Scalable and Heterogeneous

With the introduction of new technology, maintaining a dynamic environment is intimidating. At first glance, it appears almost impossible to plan for. However, when new technology can give the corporation a competitive advantage, it is incumbent upon the IT organization to embrace that technology. Corporate growth creates many new challenges for the IT organization, and technologists must adjust to this climate of constant change.

Technologists are quick to call for computing standards, but all too often ignore the people-side of that requirement. Imposing IT standards may be easier than convincing people to communicate thoroughly. The more new technology is introduced into the corporation, the more confusion is added to each support scenario. Contrary to popular belief, contact management is not optional if the corporation wants to protect its productivity. However, IT managers cannot force standards for the sake of having standards, without proving the overall impact is for the common good of the corporation. As formidable a task as it may seem, IT managers must endeavor to establish consistent standards for new changes whenever possible. Efficient standards in the computing arena have a positive impact on the economics of the corporation. To this end, scalable and heterogeneous environments allow managers to set standards in place that will evolve with the system changes.

Notification becomes more difficult because of geographical issues and time zones. The larger corporation can no longer depend on contact management procedures that address only a small amount of "need to know" technicians. The ripple effect of technology is so far-reaching that any interruption in systems availability can impact productivity on an enterprise level. This leaves the entire management team and, more expressly, the IT manager, in the dilemma of trying to create a more sophisticated contact management infrastructure. Some of the applications may have paging and e-mail capabilities, but that will not solve an enterprise-wide problem.

Intelligent contact management applications present a foundation to store and implement corporate contact management strategies. This can only be accomplished with a completely scalable and heterogeneous design. This gives an information technology manager the ability to absorb new technology and have it fit into corporate notification and support structures with minimal configuration requirements.

Centralized Control

Specifically, systems managers must be able to deal with the volatility of large-scale computing. As the corporation expands, its need for robust contact management infrastructure, the support to maintain that structure, is critical. There can be no exception of how and when communication will occur during system or process interruption. Roles must be clearly understood and defined, and confusion must be minimized. Messages from the corporate contact management system must be responded to with a sense of urgency. If an individual is contacted by the corporate intelligent contact management system, he/she knows they have a role to play in the recovery resolution. This centralized message distribution approach will enhance the corporation's ability to reduce the overall effect of system interruptions.

When attempting to establish centralized control with an intelligent contact management system, the several criteria must be met. Meeting the criteria outlined in *Table 1* will ensure the contact management system can support a corporate level contact infrastructure.

TABLE 2

Benefits from an Intelligent Contact Management System

Problem	*Solution*	*Benefit*
Long term technical compatibility	System architecture and custom APIs Support rapid development of Event Monitors, Application Interfaces and Communication Agent.	**Maximum Capability**
Embedded contact device inventory	System communication agents support all wire line and wireless devices. This will allow full utilization of all corporate devices.	**Minimal Investment**
Confidence in Contact Management	Persistence of contact through all user contact devices and routines to guarantee information delivery.	**Reliability**
Corporate Growth	Three tier architecture and supports diversity in deployment strategies.	**Scalable**
Legacy and Open Systems	Software architecture, APIs and third party interfaces maintain an open systems environment.	**Heterogeneous**
Control	Software is not application dependent allowing message distribution from a single point from all event sources (machines, applications and organizations)	**Centralized**
Access Flexibility	Web based user interface extends the reach of the system into the user community.	**Versatility**

Versatility

The technology to access the contact management platform can be an issue. The cost could be prohibitive if it is non-standard. System maintenance and the user-interface approach must be carefully weighed when selecting a system to improve the corporation's contact management infrastructure. An intelligent contact management system is required to help the corporation with the care and feeding of its primary systems; how much "tending" the contact management system needs is yet another concern.

The goal of an intelligent contact management is to allow the corporate systems and business processes to let the people know when they require assistance. This paradigm should also hold true for the intelligent contact management system itself. If the contact management system has a problem, it should contact someone. This concept addresses the problem with service interruption. However, there is still the issue of general maintenance of user profiles.

The idea of contact management deals with interfacing conveniently with the user community. In the manual environment, contact procedures require a great deal of the systems administrator's time. An intelligent contact management system will give the corporation the option of having the users manage their own profiles. This kind of versatility of access makes the system more accommodating for IT managers and users.

An intelligent contact management application eliminates all of these issues. It is designed to be set up and run with minimal attention. A Web-based interface will allow the system to move the user profile management away from the system administrator and back to the user. This removes the majority of the added load from the data center for application maintenance. When support is required, most routines can be executed remotely. This kind of versatility makes intelligent contact management a particularly attractive application to build and implement.

Intelligent Contact Management Business Issues

Table 2 summarizes the expected benefits from an intelligent contact management system. In the first column is the issue faced by a corporation deploying an intelligent contact management system. In the second column is the solution implemented by an intelligent contact management system to address that issue. The third column is the overall benefit derived from that solution.

Summary

Intelligent contact management is one of the most important challenges for all business entities today. Intelligent contact management has not yet come about as a discipline for the computing industry to aggressively embrace. However, after every service interruption, the sounds of lost productivity echo throughout the corporate hallways with management asking, "Why weren't we informed?" There are a variety of companies attempting to address the problem of contact management. Software and hardware developers are adding modules to their offerings to help alleviate the problem. The call is there for intelligent contact management, without a full understanding of what will fill the void.

System administrators, management, and end users need a contact management platform that provides a single, consistent, simple method of user device tracking and message delivery that accommodates the schedules and resources of people. Surely, it must meet the maze of technical requirements for multiple platform and legacy system support. However, paging, e-mailing, and new wireless technology is not enough to solve the real problem.

People are a corporation's most valuable resource. If people are left out of the communication loop, there is nowhere for productivity to go but down. Corporations must use intelligent contact management to take into account people and not technology. This is the only way we can truly integrate technology into our lives and make it work for us.

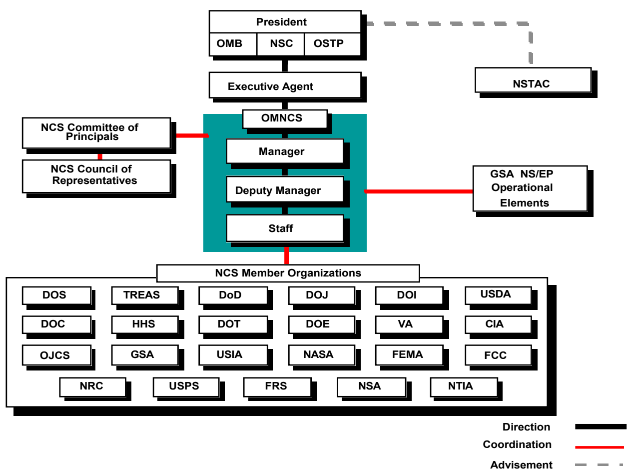
Security Concerns for IN Architectures and Services

Gabor A. Luka
AIN Program Manager
National Communications System

Intelligent networks have existed for several years but have drawn a great deal of attention since the passing of the Telecommunications Act of 1996. These networks will benefit most major industry participants: service providers, incumbent and competitive local exchange carriers, vendors, and even third-party solution consultants. There are other telecommunications users, however, that are frequently overlooked. One of the largest of these is the government. Terrorist attacks such as the bombing of the Alfred P. Murrah Federal Building in Oklahoma City and the World Trade Center in New York emphasize the need for diligence in securing vital assets for Federal Government national security and emergency preparedness (NS/EP) use. For public and private telecommunications networks that support NS/EP activities, information security is a paramount concern. This paper discusses issues and strategies identified by the Office of the Manager, National Communications System (OMNCS) to enhance advanced intelligent network (AIN) security in support of NS/EP telecommunications.

FIGURE 1

Introduction to the OMNCS



OMNCS

The Office of the Manager for the National Communication System (OMNCS) is tasked with ensuring the availability of a visible NS/EP telecommunications infrastructure for its 23 member organizations, including the Departments of Defense, State, Transportation, and other major federal agencies. While it is not possible to provide emergency communications to everyone in the government, the OMNCS serves those people responsible for providing emergency support to maintain continuity in government for those in a position of leadership. The OMNCS is shown in the center of *Figure 1*. To successfully fulfill its mission, the NCS is structured to foster interagency cooperation and serve as a focal point for joint industry and government telecommunications planning. In partnership with the NCS, the President's National Security Telecommunications Advisory Committee (NSTAC) provides industry advice and expertise to the President on matters related to national security telecommunications. NSTAC is comprised of about 30 industry CEOs from many of the biggest companies in the telecommunications industry, including carriers, vendors, and service providers.

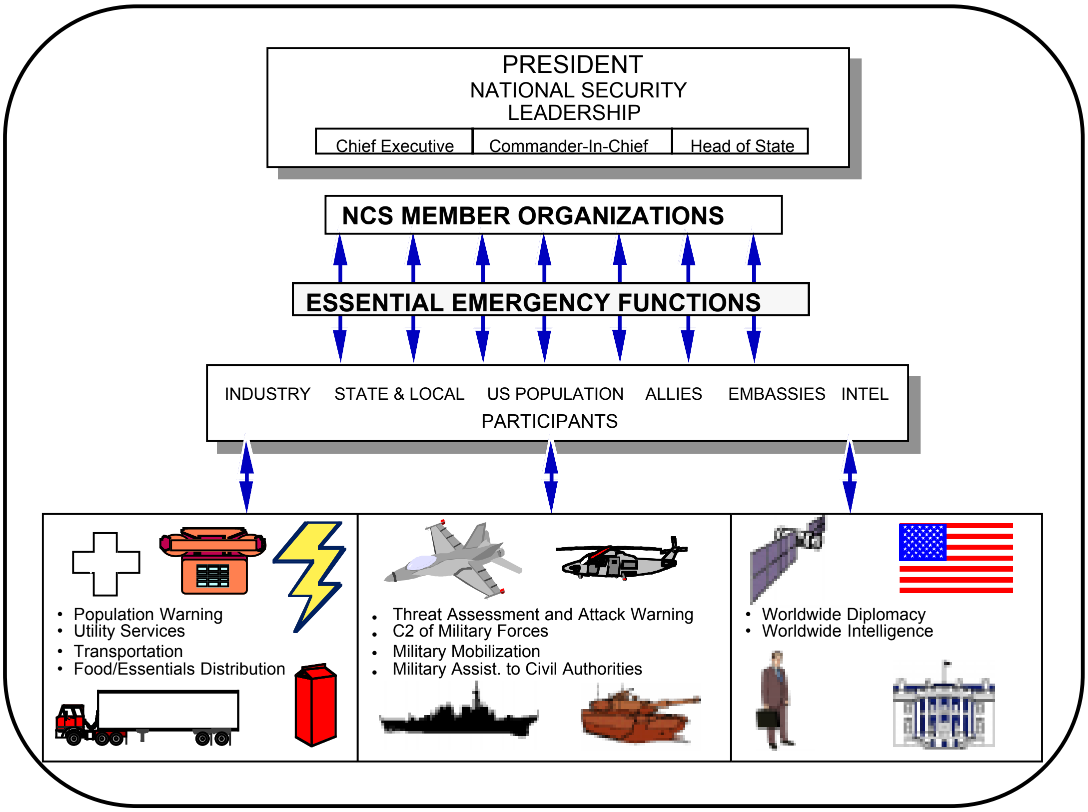
The OMNCS also administers the Government Emergency Telecommunications Service (GETS) program. GETS provides NS/EP users with priority switched voice and voice-band data service through the public switched network (PSN) that is available under a broad range of circumstances, including natural or human-made disasters. GETS implementation takes advantage of new and emerging technologies within the PSN, such as AIN capabilities. NS/EP users include personnel responsible for emergency functions or crisis management at the state, local, and even industry levels. NS/EP provides functions that support humanitarian assistance, whether it is Red Cross, food distribution, military operations, and even diplomatic missions. *Figure 2* illustrates the range of NS/EP telecommunications components. For all of these groups, security is a major concern.

Electronic Intrusion

The NCS has an Information Assurance Division responsible for identifying the security threats in the network today. *Figure 3* outlines the threat of electronic intrusion. Initially, hackers gather information, whether by stealing PIN information from trash cans or recognizing patterns of social engineering or behavior that are standard today. Hackers

FIGURE 2

NS/EP Telecommunications



frequently communicate via Internet bulletin boards and share information with others. Almost all information is accessible, given enough time and determination.

Once hackers find their desired information, there are several possible modes of attack. They can simply monitor the activity occurring in a network, look at the information that exists, determine what information is being passed, identify who the people are that are passing it, and see if they can use it to their own ends. Often, if hackers find something interesting, they will proceed to the next stage, which is penetrating a network. Frequently, they masquerade as an authentic user and utilize the assets of a network to their own intent. Finally, they can plant malicious agents in a network that may cause great damage.

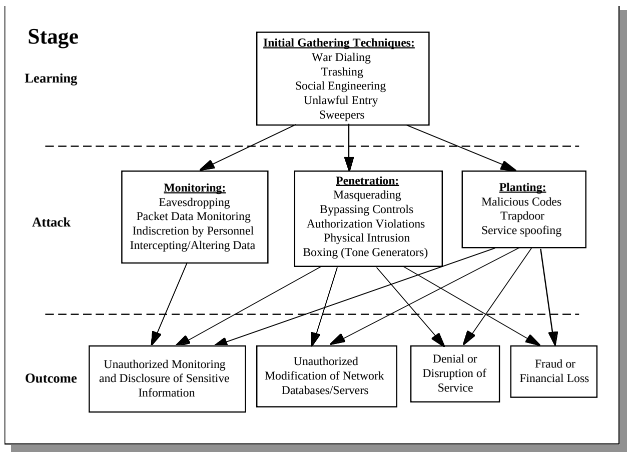
Outcomes of hacker incidents can vary greatly. They can range from the disclosure of sensitive information to the modification of parameters, including the database of authentic users, services, routing parameters, or billing instructions. Modification to any of these areas can result in an adverse effect on the network. These effects could range from simple disruption of service to one user to something as damaging as denial of service to a variety of users. The financial impact in the latter case could be devastating.

IN Risks

Many studies have been conducted by the OMNCS on security and reliability risks in the intelligent network area. Particular areas that have been explored are signaling networks, whether SS7 related or IS-41 related, service logic security, mediated access, and emerging technologies. In the

FIGURE 3

Electronic Intrusion Threat



Stages of the Electronic Intrusion Threat

SS7 and IS-41 arena, some of the vulnerabilities that have been identified include physical- and/or software-based disruption of common channel signaling (CCS) components, such as signal transfer point (STP) outages or the injection of false SS7 packets. Additionally, interoperability/interconnection concerns come to the forefront as new entrants are connected into a network and create potential points of failure.

The second major area of risk is service logic security. Currently, the network is intensively software-defined, and service logic programs (SLPs) are a definite area of vulnerability. Mis-programming, malicious alteration, or corruption of service parameters are major concerns. There is also the potential for financial or economic espionage that could cause a substantial loss of revenue. This could come in the form of unauthorized use of a service, such as accessing free long-distance or other chargeable services. Another concern is access to user databases and service platforms such as the SCP. Examples of SCP security breaches could include a hacker targeting NS/EP user databases, mobile identification numbers, electronic service numbers, or authentication keys. This access could come from an outside hacker or an internal user with inappropriate motives.

Another identified risk area is mediated access. This occurs when a new point of entry is introduced so that third parties can access that network, perhaps as a result of the unbundling mandated by the Telecommunications Act. Again, software can cause vulnerabilities here, as the mediation software necessary to interface points in a network can add a potential vulnerability to access by unauthorized parties. In addition to this projected danger, there is always the chance that mediated access will produce new sources of error that could endanger a network.

Finally, emerging technologies can also cause IN interface concerns. For instance, local number portability (LNP), personal communications services (PCS), and wholesale services can all bring their own security dangers. Since these are new services and technologies, there could be trouble detecting the vulnerabilities that may result from technology integration. Trouble-detection techniques often lag behind the implementation of the services, especially as companies drive for first advantage to market. Sometimes the drive for implementation of new technologies and capabilities comes at the price of minimizing or delaying the application of security measures—a potentially very high cost.

Potential Mitigation Strategies

While the risks associated with IN may be high, there have been solutions proposed to mitigate potential security problems. For example, SS7 and IS-41 internetworking best practices were well-documented in the FCC Network Reliability Council Report from 1996 and should prove helpful to all involved. It is also possible to implement SS7 enhancements, such as the high probability of completion (HPC) standard for critical services. The HPC is an American National Standards Institute (ANSI)–approved standard that identifies NS/EP users across the public switched network, establishes the appropriate initial address message parity level, and sets the calling party category to allow enhanced routing, exemptions from network management control, and alternate carrier routing. Additionally, working with suppliers and smaller service providers to ensure adherence to CCS security standards can cut the chances of risk through interconnection.

In terms of SLP protection, there are several actions that could mitigate security risks. One is to address security with the customer—identify what the customer's needs are, what data they think is important, and what level of security they would like to see implemented during service development. When the customer's security concerns are understood, it is easier to build them appropriately to guard the capabilities in their network. It is also advisable to use better security measures during software delivery, installation, and upgrading. This means that the supplier should start implementing security early on in the service development process and continue throughout the interaction process. Every upgrade should be viewed as an opportunity to evaluate security. Finally, it is important to ensure protection of critical data such as mobile identification numbers (MINs) and electronic serial numbers (ESNs).

Mitigation strategies for mediated access are similar to some of the solutions mentioned above. It may be helpful to issue a minimum set of requirements for interconnection, not allowing outside parties to interconnect to the SS7 network, service logic, the SCP, or databases until they have met certain criteria for security. If this does not happen, there could be a security breach in the network, which could result in a domino effect that could take out some of the service capabilities. The solution lies in the use of network partitions and firewalls and, once again, conducting thorough testing that involves customer participation.

There are fewer tried and true mitigation strategies for emerging technologies. Most people understand that standards are contribution-based. If security requirements are identified early on in the process, it will be possible to develop best practices for new technologies. The contribution to standards bodies ensures that security concerns are addressed. With that type of participation, it is possible to establish such best practices as encryption of over-the-air activation authentication, a process that greatly aids in minimizing fraud in the wireless network. Finally, it is important for all concerned to participate in ongoing security research and development. Since security is an industry-wide concern, there are often opportunities to conduct collaborative efforts for research and development.

The above are proposed solutions for the risks identified previously. There are, however, other general mitigation strategies. One involves personnel. Some say that investment in personnel is the most valuable investment a company can make. This investment should include the conducting of intensive security audits. It is also important to implement document control and improve the skills of security staff either through training or security awareness programs. Additionally, it is always a good idea to use the latest technical security solutions available.

The hacker world is a very dynamic one. As soon as a virus is introduced, there is a team identifying the solution to that virus or hacker threat. And as soon as that solution is identified, there is a new problem to solve. This makes it essential for all involved to stay on the leading edge of security solutions. It is important to keep informed about the latest threats as well as the latest solutions. These solutions include the use of network partitions and firewalls, ensuring access control and authentication, and deploying new security technologies and standards as they become available.

Summary

Standard large telecommunications networks such as those used by the government are not the only areas in which security is a concern. For instance, there are many wireless focus areas for security. Some of these include mobility management, local number portability, and personal communications systems. Educating new market entrants and partnering with them in security responsibilities should help in this area. Also, the convergence of IN and wireless IN (WIN) can prove challenging.

There is a trade-off between implementing security practices and procedures and the potential loss of revenue. Because of this, it is important to identify potential target markets when developing a business case for implementation—determine what kind of security must be offered and make that a factor in considering the type of security measures that must be implemented. Customer behavior and profiling, speaker verification, cellular authentication, and secure key distribution are the areas that will come into play most often.

The NS/EP NCS AIN program will continue to work with the industry to address security issues. While the governmental case may seem extreme, it can serve as a good example for the rest of the industry in security measures taken and level of participation. Through the GETS program, the NCS has an initiative to reach every state, contacting them to let them know about the programs and initiatives that the NCS has in place and the available opportunities. If this level of involvement is carried throughout the telecommunications industry, the major concerns about security should be handled to everyone's satisfaction.

Modeling AIN: A User's Perspective

Patrick McGregor
President
Nyquetek, Inc.

This essay presents a user's perspective of modeling advanced intelligent networks (AIN). It explains the Government Emergency Telecommunications Service (GETS) and the benefits it receives from AIN. Then, using GETS as the application example, it discusses how a user applies AIN modeling to identify potential service issues and guide service provider and vendor considerations.

By way of introduction, imagine that you do not work for an AIN company. Instead you work for a major brokerage, an automobile company, a bank, or a major airline. Late one night, you suddenly awake from a sound sleep with a powerful vision, a vision of an AIN application that can get your company one, maybe two more percentage points of market share. The vision is so strong that you can see a promotion, a raise, a bonus, and a big red Mercedes ahead. Life looks good.

The next day you go into the office and share your vision with the boss. He is excited and brings in marketing. Marketing is excited and suggests bringing in the telephone company to see if the vision can be realized. When you share your vision with the telephone company, its representatives say they can implement it in nine months—and do so affordably. Budgets are allocated, brochures are designed, the vice president puts it in her quarterly presentation to the president, and the president puts it in his semiannual presentation to the chairman. This looks really good.

Then one day your boss calls you in and asks if you know your idea will work the way the telephone company has promised, the way you want it to work. Wondering whether this is a trick question, you respond that you are dealing with the telephone company—this is their business and they can be trusted. The boss in turn answers that, first, this is business and no one is to be trusted. Second, whatever trust the telephone company deserves it earned twenty years ago, when it decided what was good for customers, invested all it could because regulation provided profit on whatever was invested, and tested new ideas for five years before deciding whether or not end users would see them. Now the telephone company wants to hear our ideas and do our development—for a cost. Is this different from other suppliers? Does this approach deserve more trust than we would place in any other supplier? The answer is no. We should have some trust in all suppliers, but today, in this competitive world where we must be responsive to the customer, we should not have any more trust in the telephone company than we would in any other major supplier. So, how are we to know if our service will work? Enter the use of modeling.

User Perspective

From a user perspective, then, the question is whether an idea for a new service will in fact be delivered in a timely fashion and will work the way it is supposed to work. Modeling is a tool that allows a user to determine if a telephone company will deliver the service with the performance and capacity promised.

"User" here means the value-adding retail service provider (see *Figure 1*), the middleman who represents the end user as part of the end user's corporate world. The user may be a separate retail provider, or part of the wholesale provider—the telephone company. The middleman receives the service from a wholesale service provider, or plant owner and operator, and is acting on behalf of the broader end-user, or subscriber, community.

From the user's perspective, then, the issue is assurance. What does the user need to know in order to feel certain that the owner-operator of a service will in fact deliver as needed, and how far back in the chain does the user have to go to get that good feeling? Modeling is the key to identifying concerns and creating the assurance dialogue for the user.

GETS and AIN Benefits

The user's perspective can be illustrated by the Government Emergency Telecommunications Service. A program of the National Communication System, GETS is designed for key people in government to use in a national security emer-

FIGURE 1
User Perspective

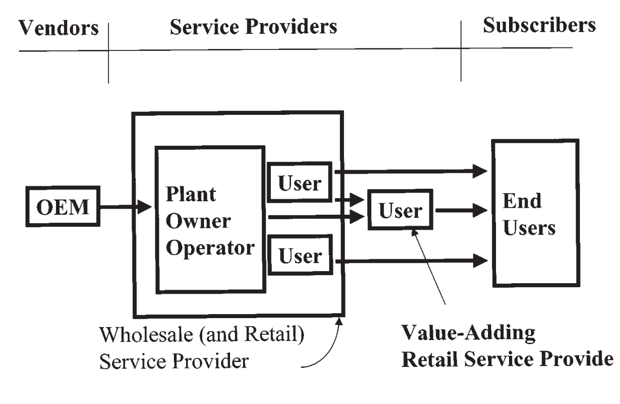


FIGURE 2

GETS Primary Strategy "Out from Zone" End-to-End Performance Congestion Only: No NMCs, No Damage (6ccs/line, IECs at P.01, Non-Stress Egress at P.01)

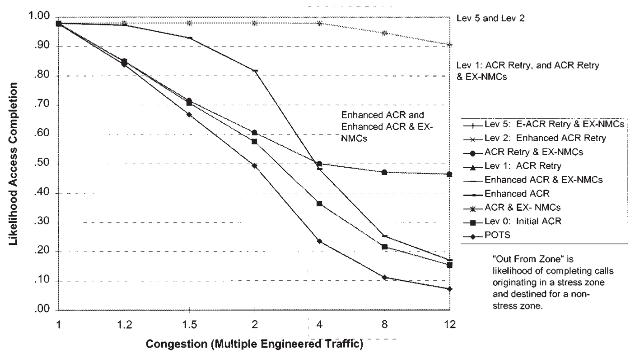
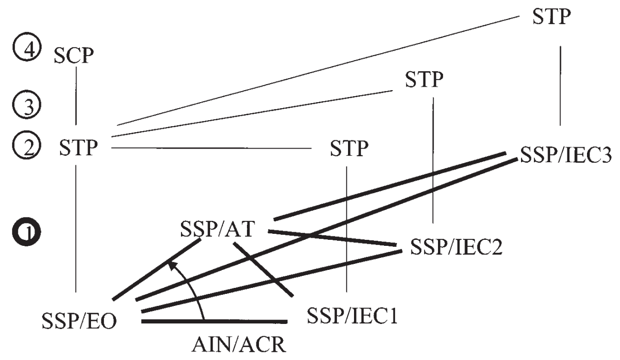


FIGURE 3

Simplified Topology



gency. The government must be able to continue to function and to function effectively when the communications network is overloaded and outages are prevalent because of a national emergency or major disaster. GETS gives key users certain features that enable them to use the public switched network (PSN) on a priority basis. The priority treatment will not preempt other types of communication, but it will ensure that the GETS call receives the best possible chance of being completed through the network.

GTE Government Systems, the GETS integration contractor, and Nyquetek, a subcontractor to GTE, provide GETS with AIN capabilities through subcontracts with the local exchange carriers (LECs) (i.e., GTE Government Systems is a user of the AIN services provided by the LECs). AIN enhances GETS' ability to complete calls in the event of an emergency coupled with severe PSN stress through its PSN priority treatment features.

Figure 2 illustrates the type of benefits AIN can provide in a broad regional disaster with some outages that lead to overload in different areas. On the left of the bottom axis, 1 indicates an overload of one. That is the average busy hour, the normally engineered rate of service for the telephone network. It represents a 98 percent likelihood end-to-end call completion. Moving to the right on the bottom axis, 1.2 means a 20 percent increase in call volume compared to volume in the average busy season, busy hour. Two times overload is becoming quite severe, and is well beyond Mothers' Day line traffic. Four times overload is four times the normal call volume, as occurred, for example, in the focused traffic of the Oklahoma City bombing disaster. Four times overload is extraordinary demand for telephone networks, which do not usually experience that level of demand for even their highest busy hour of the most busy day of the year. Eight times overload might occur in focused traffic situations with a major hurricane or an earthquake in California.

The bottommost curve is plain old telephone service (POTS). This curve shows that the probability of completing a call goes way down as overload increases, but the probability can be increased by using an AIN service capability feature called alternate carrier routing (ACR). This feature includes instructions to try multiple carriers. It routes the call first to one carrier and, if unable to get through, tries a second carrier. Alternate carrier routing brings the curve up a little bit. Enhancing ACR by adding the capability to retry carriers periodically brings the curve up to the top so the call eventually goes through the network. That is a nice AIN benefit that is good for the service.

The question is, will that benefit occur during conditions of severe network stress? Will it really work when the network is congested? When it has eight times overload? Nyquetek has modeled that situation. The company looked at a fairly intense network situation typical of a metropolitan business area with end-offices averaging six CCS traffic units per line (i.e. .17 Erlangs per line) and a three-minute holding time. If usage increases four times with a two times focus factor, the system will face an eight times overload with single points of failure that must be worked around.

To model these conditions, Nyquetek looked at the SS7 network as the first link in the AIN process (see *Figure 3*) and examined the likelihood of completing an AIN transaction for alternate carrier routing (ACR) when the network has eight times overload. For our hypothesis, we looked at a conventional SS7 network arrangement, beginning with the first link from the service switching point (SSP) to the signal transfer point (STP). The left-hand scale of *Figure 4* shows the likelihood that the end call will be completed using GETS (the likelihood is very high as seen by the top curve). The right-hand side of the scale shows the likelihood that an SS7 transaction will be blocked because of congestion on the SS7 network. The likelihood that an SS7 call will be blocked is in the range of one in 1 million to maybe two in 100,000, and is reflected in the AIN blocking line overlapping the bottom axis on the chart. The SS7 backbone is assumed set up to be 40 percent, utilized in its normal operating mode, with sufficient buffering allowed for that level of service. In this case a 10-message buffer has been applied.

Up to eight times overload will produce no stress on the SS7 network. Conceptually, that point is easy to understand. The only time a message is sent to the other end is when a trunk is available. Normally the trunks are about 90 percent utilized.

FIGURE 4

SS7/AIN Overload Performance (Baseline: Only GETS TCAP, No Outages)

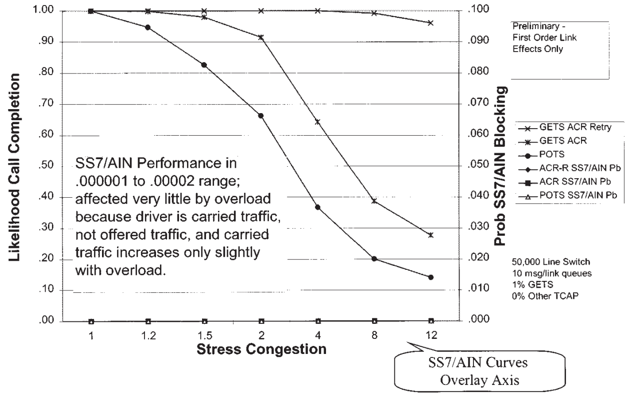
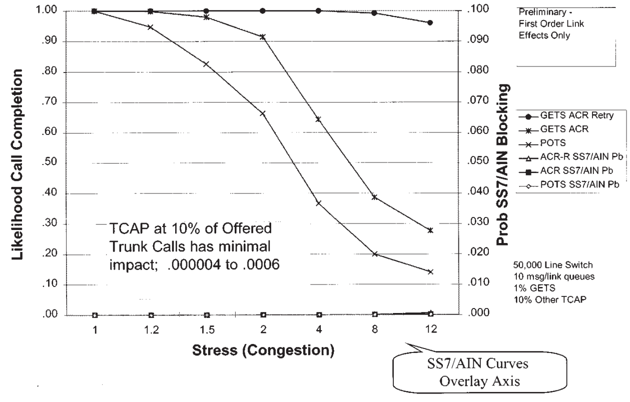


FIGURE 6

SS7/AIN OverloadPerformance (Case2: GETS TCAP + 10% Other TCAP, No Outage)



When eight times overload occurs, the most additional traffic that can be carried is that last 10 percent, which translates only into an 11 percent overload factor. As long as the SS7 network generates at most 11 percent more messages than it otherwise would, it does not matter how big the overload is. Although this explanation is somewhat simplistic, it explains the idea. The SS7 network is no problem in this scenario.

But what happens if one of the STP links is out? In this case, instead of the normal SS7 link utilization of 40 percent, SS7 link utilization would be 80 percent in the average busy season, busy hour. As the overload increases by 10 percent, utilization increases from 80 percent to 88 percent to 90 percent (see *Figure 5*). Some blockage becomes apparent—about a 1 percent traffic loss. The ACR retry has to go through many iterations of sending the transaction up to the SCP and regenerating the instruction to keep trying calls. Along with failure and eight times overload, the additional retry activity adds a detectable amount of blocking that can be seen on the SS7, but things still look pretty good. The top curve (GETS call likelihood of completion) in *Figure 5* is still very high and the AIN transaction blocking curves are still very low.

If, as in *Figure 6*, 10 percent of the other traffic also has TCAP transactions in the process, the system still is fairly healthy. Even 10 percent TCAP transactions and eight times overload is not a big problem.

However, if the system loses an STP and has 10 percent TCAP transactions and eight times overload, a potential problem begins to emerge (see *Figure 7*). The TCAP transactions are a function of the offered traffic, not the carried traffic. Any call that comes in has to go to the SCP before it can be routed. With eight times overload on the TCAPs, all the overload goes to the SS7 network, not only the 10 percent that the SS7 can carry, but the full eight times overload. This high mix of TCAP traffic places the SS7 network at much greater risk of performance degradation than is historically associated with its robustness for call signaling.

Once it is evident that the service benefits may become sensitive to the TCAP mix, the user can open the dialog with the telephone company to find the needed assurance that this facet of service design is being adequately addressed. In the case of GETS, such questions include with how much TCAP traffic will GETS AIN transactions have to compete.

Can the GETS traffic be separated from the other TCAP traffic and perhaps be given a different priority? Remember, the only time GETS has to work is when the network is severely congested.

FIGURE 5

SS7/AIN Overload Performance (Case 1: Only GETS TCAP, No Outages)

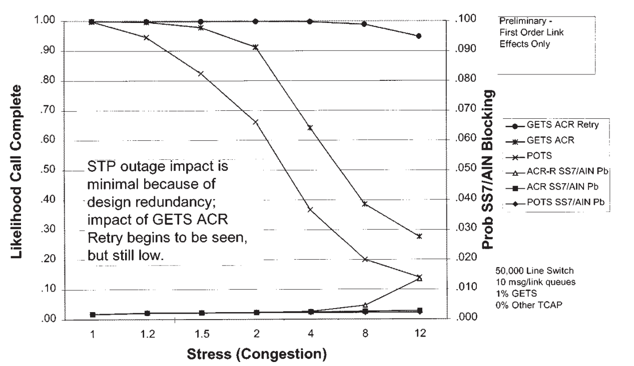
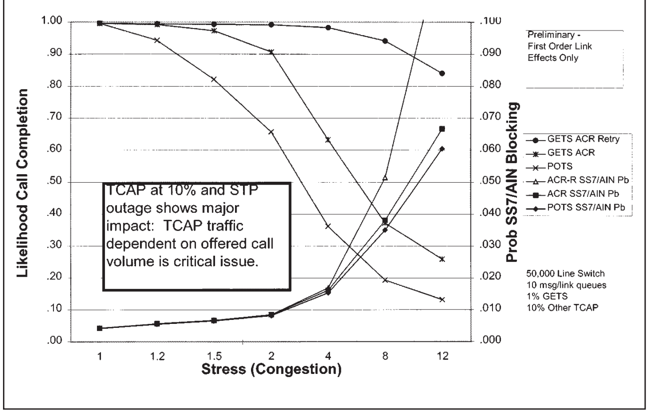


FIGURE 7

SS7/AIN Overload Performance (Case 3: Only GETS TCAP, + 10% Other TCAP, STP Outage



Owner/Operator and Vendor Considerations

From a more general standpoint than that of the models just discussed, the user can ask what processes the owner / operator uses to ensure service effectiveness. The user wants assurances that the telephone company will provide a service that performs as promised. The telephone company develops such services with sound engineering practices. But as with any industrial supplier / customer relationship, the customer needs some visibility into the supplier's engineering processes and considerations to get the needed assurance. Modeling can provide the customer with insights identifying possible concerns that help the supplier to better understand user needs and to be responsive to user requests for assurance.

In addition, as users become familiar with the supplier's dependence on certain systems for service delivery (e.g., the SCP), users may need vendor assurances that the systems used by the supplier provide the needed capabilities for the desired service. For GETS, this includes the capability of the SCPs to gracefully accommodate possible volume changes and to ensure that an equitable and effective share of processing resources will be available for the service. GETS must be sure its transactions will be processed under severe system stress, which is the only time it needs the system. Whether or not the supplier can provide such capability will in part depend on the capabilities of the systems he acquires from his vendors.

The user must be prepared to follow the supplier / vendor chain for as far as necessary to be confident the service will be delivered as promised. The process need not be onerous for either the user or provider. By using modeling to help identify areas of concern, the user and supplier (and vendors) can work effectively as a team to ensure the delivered service is delivered with justified confidence in its performance.

Conclusion

AIN offers potential for new, fast, economical, and flexible services from a user perspective. GETS is a good example of AIN benefits, but the flexibility also brings new complexities, both on the technical side and the business side. The owner / operator and the vendors must work as a team with the user—not just as the user's suppliers—to turn these complexities from problems into advantages.

Voice over Frame Relay Integration

Micom Communications Corporation

This paper originally appeared as a white paper from Micom, a Nortel (Northern Telecom) company, and is reprinted with permission.

Introduction

The rapid growth of frame relay (FR), driven initially by the need to connect remote office local area networks (LANs), is expected to continue as customers discover new ways to take advantage of its benefits and add new multimedia applications. Between 1994 and 1996, the number of worldwide companies using FR services grew ten times from 1,480 to 14,847, as shown in *Figure 1*. This trend is expected to continue.

Frame relay provides better performance, simplifies network management, and allows for consolidated data networks—all exciting benefits that, when combined, save money. This is great news for the network manager. The better news is that now it is possible to save even more by consolidating over the FR network, not only enterprise data traffic, but also voice traffic between offices.

This paper was written to:

- help the reader to understand the basics of frame relay
- explain why frame relay networks can now effectively carry voice
- describe the technologies to look for when deploying voice over frame relay
- show examples of how complete network integration has saved companies thousands on their wide area networking (WAN) costs

What Is Frame Relay Anyway?

The term Frame Relay can be used to describe the following:

- an interface specification—rules for connecting to the public network

FIGURE 1

Frame Relay Market—Worldwide Users

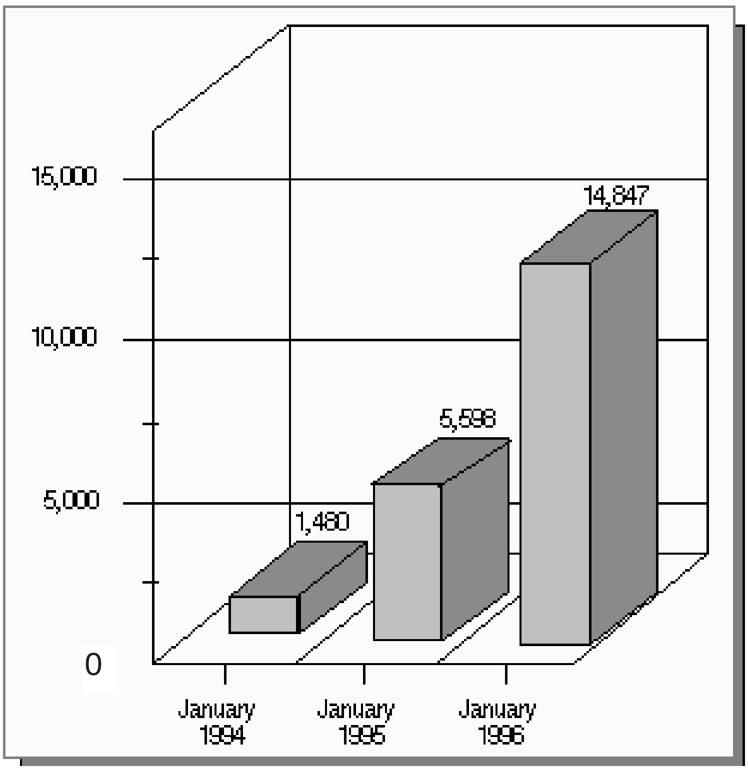
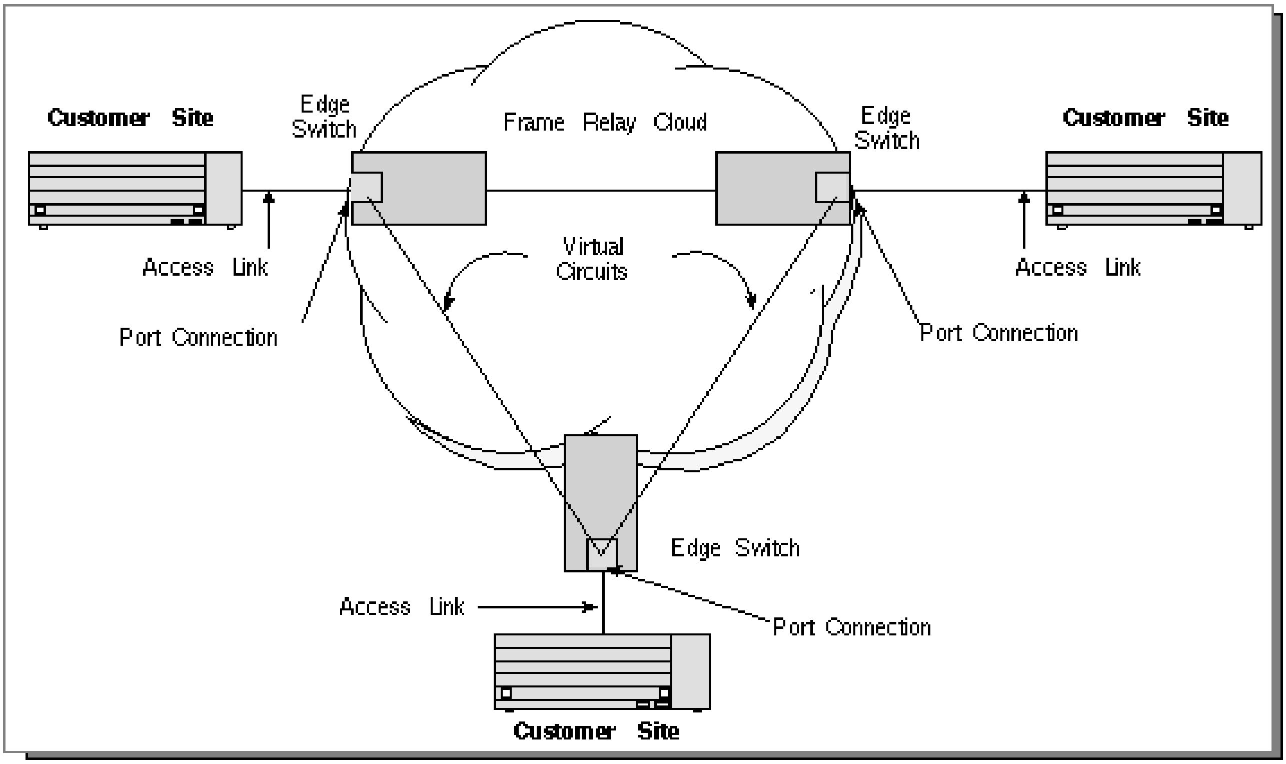


FIGURE 2

Frame Relay Network Elements



- a switching technology—a means of routing frames through the network

- a public service—offered by carriers for WAN connectivity

Frame relay defines a method for effectively routing frames of information across a WAN. Packet technologies are very suitable for data communications since they make very efficient use of communications channels. Frame relay is especially suited for data-intensive applications, including the connection of LANs over the wide area. It is also becoming popular as an access method for higher-speed asynchronous transfer mode (ATM) networks.

Frame relay is a packet-switching protocol based on X.25 and integrated services digital network (ISDN) standards. Unlike X.25 however, which assumed low-speed, error-prone lines and had to perform error correction, FR assumes error-free lines. By leaving the error correction and flow control functions to the end points (customer premise equipment [CPE]), FR has lower overhead and can move variable-sized data packets at much higher rates.

Like its predecessor X.25, FR is a shared service allowing multiple customers to use the WAN simultaneously. Several different corporations connecting their remote locations would be one example. Frame relay would not be gaining so much popularity if it did not offer some solid benefits, such as:

- saving money
- improving network performance and availability
- simplifying network management
- consolidating networks

To explain how FR provides these benefits, it is important to understand how frame relay technology works, the pertinent elements of a frame relay network, and how carriers charge for FR service.

How Does It Work?

Figure 2 shows FR network elements. Each location gains access to the FR network through a frame relay access device (FRAD). A router with frame relay capability is one example. The FRAD is connected to the nearest carrier point-of-presence (POP) through an access link, usually a leased line. A port on the edge switch provides entry into the FR network.

FRADs assemble the data to be sent between locations into variable-sized frame relay frames, like putting a letter in an envelope. Each frame contains the address of the target site, which is used to direct the frame through the network to its proper destination. Once the frame enters the shared network cloud or backbone, any number of networking technologies can be employed to carry it.

The path defined between the source and the destination sites is known as a virtual circuit. While a virtual circuit defines a path between two sites, no backbone bandwidth is actually allocated to that path until the devices need it. Frame relay supports both permanent and switched virtual circuits. Frame relay service providers typically charge for each of these paths.

A permanent virtual circuit (PVC) is a logical point-to-point circuit between sites through the public FR cloud. PVCs are permanent in that they are not set up and torn down with each session. They may exist for weeks, months, or years, and have assigned end points, which do not change. The PVC is available for transmitting and receiving all the time, and in that regard, is analogous to a leased line.

In contrast, a switched virtual circuit (SVC) is analogous to a dial-up connection. It is a duplex circuit, established on demand, between two points. Existing only for the duration of the session, it is set up and torn down like a telephone call. FRADs that support SVCs perform the call establishment procedures. Currently, all public frame relay service providers offer PVCs, while only a very small number offer SVCs. In the near future, there will likely be more SVC services available. SVCs and PVCs are compared in *Figure 3.*

By supporting several PVCs simultaneously, FR can directly connect multiple sites through a single physical connection. In contrast, a leased line network would require multiple physical connections, one for each site. A data link connection identifier (DLCI), assigned by the service provider, identifies each PVC. A header in each frame contains the DLCI, indicating which virtual circuit the frame should use.

The real benefit of frame relay comes from its ability to dynamically allocate bandwidth and handle bursts of peak traffic. When a particular PVC is not using backbone bandwidth, it is "up for grabs" by another.

When purchasing PVCs, the bandwidth or committed information rate (CIR) must be specified. The CIR is the average throughput the carrier guarantees to be always available for a particular PVC. Most carriers sell CIRs in increments of one kbps, from zero bits per second to T1 (1.544Mbps). Higher rates cost more.

For example, a remote sales office might have a physical link speed of 64 kbps supporting two PVCs, one for LAN data with a CIR of 48 kbps and another for asynchronous data at a CIR of 16 kbps. This means that, on an average basis, the LAN and asynchronous terminals should not transmit more than 48,000 or 16,000 bits, respectively, in a second. Like a leased line, as long as the amount of data the end device is trying to send does not exceed its CIR rate, it should get through the network.

Frame relay offers a unique advantage over leased lines. Unlike leased lines, FR can handle bursts of peak traffic. Since the leased line bandwidth is fixed, if the device attempts to send data at a rate higher than the line bandwidth—a large file transfer, for example—it will not get through, and perfor-

FIGURE 3

Network Connections

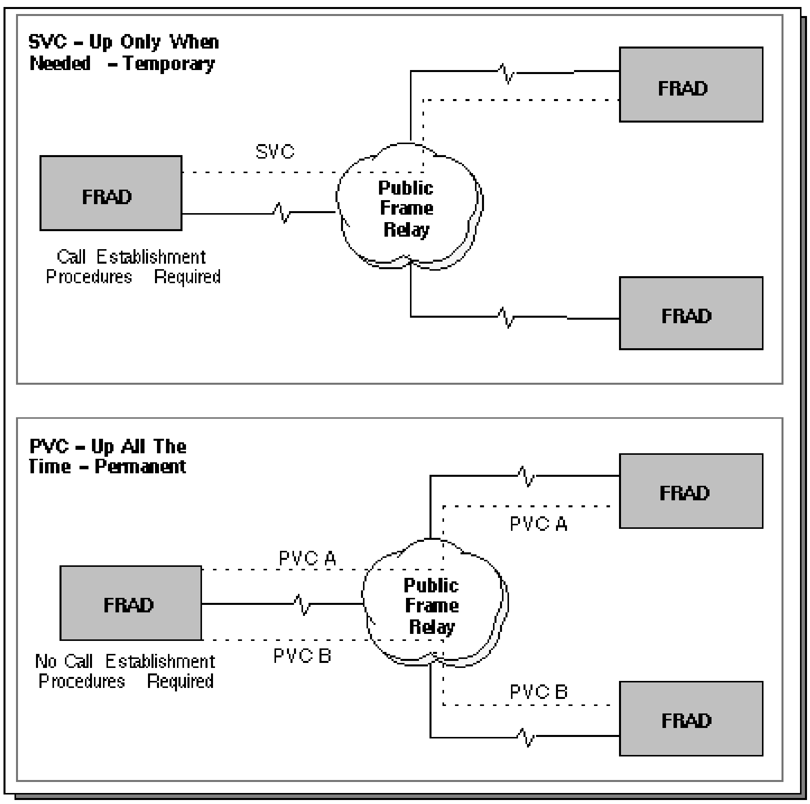
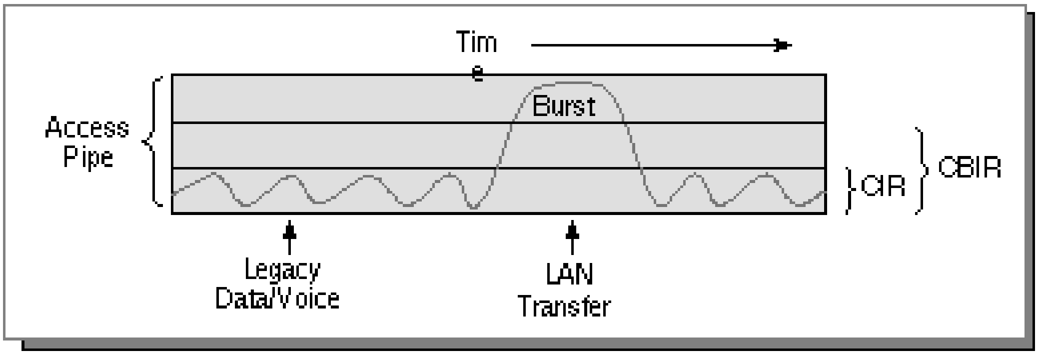


FIGURE 4

CBIR During A LAN Transfer



mance will be degraded. Frame relay, however, will allow a device to transmit data at a higher rate than the CIR for a few seconds at a time. This is called bursting. The idea is to let devices that temporarily need it, borrow, at no extra cost, network backbone bandwidth not being used by other devices. During heavy file transfers, for example, the LAN could use the 16 kbps allocated to the terminal, if the terminal was not in use. A device can burst up to the committed burst information rate (CBIR) and still expect the data to get through (see *Figure 4*). The duration of a burst transmission should be short, less than three or four seconds. If long bursts persist, then a higher CIR should be purchased.

Devices using the extra free bandwidth available do run a risk: any data beyond the CIR is eligible for discard, depending on network congestion. The greater the network congestion, the greater the risk that frames transmitted above the CIR will be lost. While the risk is typically very low up to the CBIR, if a frame is discarded it will have to be resent. Data can even be transmitted at rates higher than the CBIR, but doing this has the greatest risk of lost packets.

The FR network does try to police itself and keep congestion, and thus packet loss, down. It can do this in two ways. It can try to control the flow of packets with forward explicit congestion notification (FECN), which is a bit set in a packet to notify a receiving interface device that it should initiate congestion avoidance procedures. Backward explicit congestion notification (BECN) is a bit set to notify a sending device to stop sending frames because congestion avoidance procedures are being initiated. See *Figure 5.*

A second way to inform the end devices that there is congestion is through the local management interface (LMI). This specification describes special management frames sent to access devices.

A discard eligibility bit (DE Bit) is set by the public FR network in packets the device is attempting to transmit above the CIR or the CBIR for any length of time. It will also be set if there is high network congestion. This means that if data must be discarded, packets with the DE Bit set should be dropped before other packets. It should be noted that the network itself has no way to enforce congestion flow control. It is up to the end device to support and obey these codes. Finally, the frame travels to its destination where it is disassembled by the receiving FRAD, and data is passed to the user.

How Does This Benefit You?

Frame relay offers several benefits:

- reducing internetworking costs
- improving network performance and availability
- simplifying network management
- consolidating networks

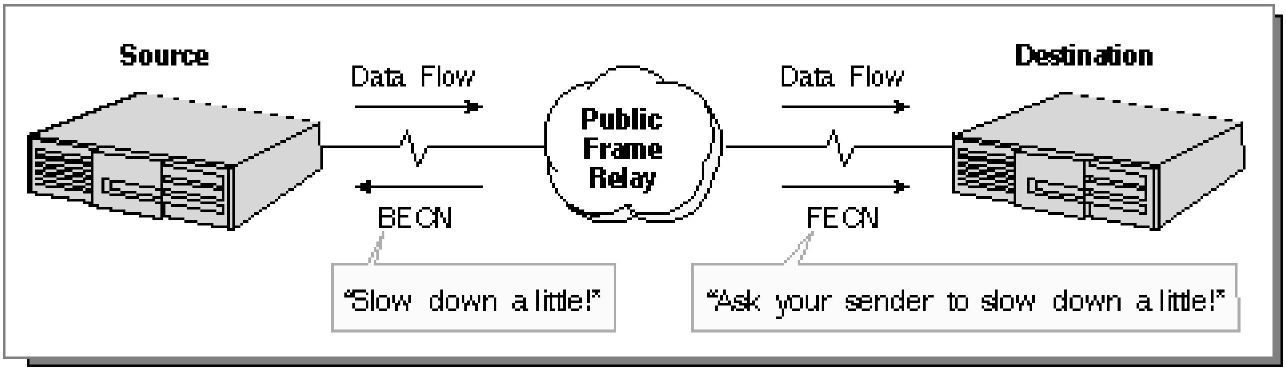
Frame relay improves a company's bottom line in several ways. Reduced access and connection costs show immediate savings. In most areas, the cost of an FR link is less expensive than a leased line of equivalent bandwidth. That is only part of the equation, though. Additional savings can be realized by taking advantage of FR's bandwidth flexibility. Since frame relay dynamically allocates bandwidth and allows for bursting over the purchased rate, a fundamental change can be made in the way networks are planned.

With leased lines, where the bandwidth is fixed, it is essential to plan and purchase bandwidth for the maximum traffic load. That means purchasing more than normally needed, just in case (like buying insurance). Therefore, the user ends up paying for unused bandwidth. Frame relay allows the user to purchase average bandwidth usage. Peak traffic bursts are still handled by the network "for free." In addition, each link can be tailored to its exact requirement because CIRs can be purchased in increments as small as 1 kbps. Purchasing only what is necessary can save as much as 40 percent over equivalent leased line service.

Frame relay also allows changes in network topology. In a typical network configuration requiring direct connections to multiple sites, multiple physical connections (i.e., leased lines) are needed at each location. Frame relay provides this same capability with a single physical connection. Multiple PVCs on the single link can be used to create direct logical connections between branch offices. There is no need to route all connections through headquarters. Direct connections can also improve performance. Since all data does not have to be

FIGURE 5

FECN and BECN



routed by a single location, as in a leased line star topology, there are no bottlenecks. Multiple links through a single physical connection means less customer premise equipment to maintain, including fewer data service unit / channel service units (DSU / CSUs). The mesh topology of the FR cloud also increases network availability and resiliency. If a network link fails, the switches in the cloud automatically reroute the data, and the end devices never even notice. See *Figure 6.*

As a side benefit, frame relay eases the pain of network management. The service provider is responsible for keeping the network up 24 hours a day.

Lastly, and probably most importantly, frame relay offers the ability to consolidate your networks. This is because it was designed to be "protocol transparent." Since frame relay can carry multiple data types, both LAN and SNA traffic travel simultaneously over a single network. That is a real money saver.

So the operation and benefits of frame relay, from cost savings to eliminating expensive and redundant networks have been discussed. Does this mean that the user is saving as much as possible? Not quite. Even though FR has helped to consolidate data networking, the fact is that the user has achieved only partial integration, as shown in *Figure 7.* There

FIGURE 6

Logical Mesh Network

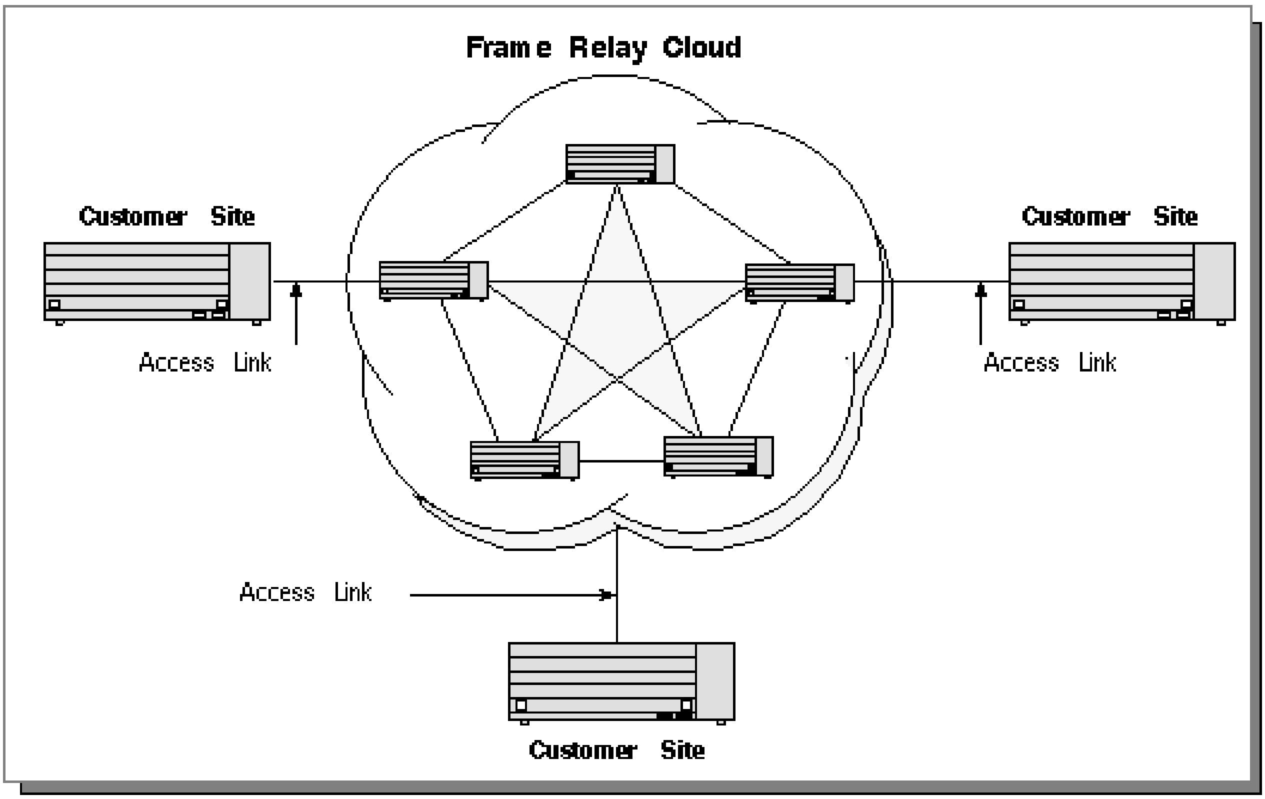
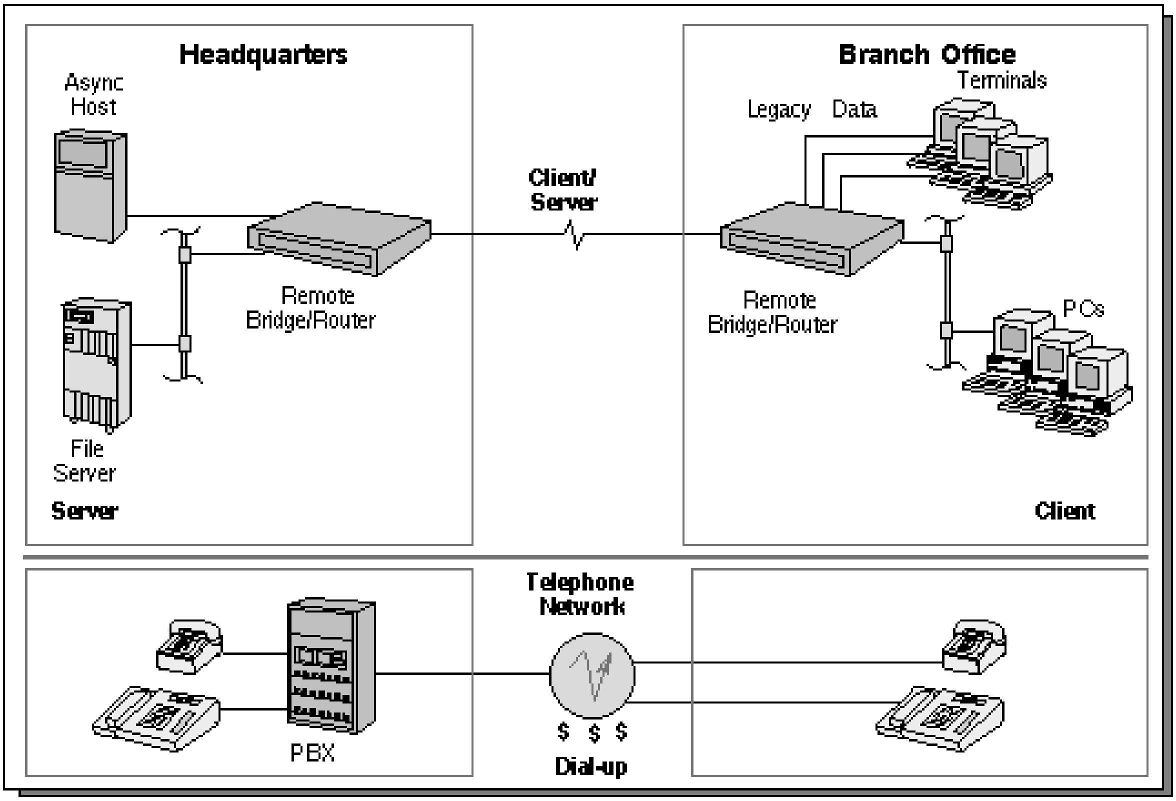


FIGURE 7
Partial Integration



is one network still costing unnecessary expense: the voice network between sites.

Integrating Voice over Frame Relay

By adding voice to the frame relay network, it is possible to fully integrate an enterprise network, as shown in *Figure 8*. Fully integrating an enterprise network means consolidating voice, facsimile, data, and LAN traffic over a single network. Frame relay allows full consolidation of a network by efficiently carrying voice and facsimile along with data.

Voice over frame relay is receiving growing attention. A number of new products have been delivered to help companies achieve this full integration, and a sub-committee within the Frame Relay Forum has been formed to develop standards. The benefits of this technology helped the enterprise integration market grow to $475 million in 1995 (see *Figure 9*).

Integrated networking is for anyone who has two or more sites sharing data and voice information. By fully integrating the enterprise network over frame relay, a user can receive all the benefits associated with frame relay plus the added savings in having voice calls ride free between remote locations. A move to frame relay, combined with voice and facsimile integration, can reduce remote communications costs up to 60 percent.

At this point, it is important to clear up one of the myths concerning voice over frame relay—the myth that it doesn't work. Today's FR public networks routinely transport a combination of synchronous and asynchronous data, voice, facsimile, LAN, and video traffic in thousands of installations world wide. The myth came about because early FR networks, based on medium-speed backbone links, were not particularly well suited for voice.

High-quality voice and video require very short, predictable delays. Private networks, which use leased line connections, could dedicate bandwidth to different data types and meet this requirement.

As discussed above, packet networks do not dedicate bandwidth. Instead, bandwidth is dynamically allocated, and the data is placed in packets for transmission. In cases such as X.25, which used relatively slow (up to 64-kbps), unreliable backbone lines, error correction was performed, which required protocol overhead and introduced delays. This environment often resulted in poor quality voice transmission.

When frame relay was first introduced, the transmission infrastructure was also in transition. Transmission facilities were being upgraded from an analog to a digital structure. These changes made the facilities more reliable so error detection and correction did not have to be performed by the network. The backbone lines also got faster (up to T1). While these factors greatly improved network throughput, the delays still presented a problem for time-sensitive voice, facsimile, or video transmission.

FIGURE 8

Full Integration

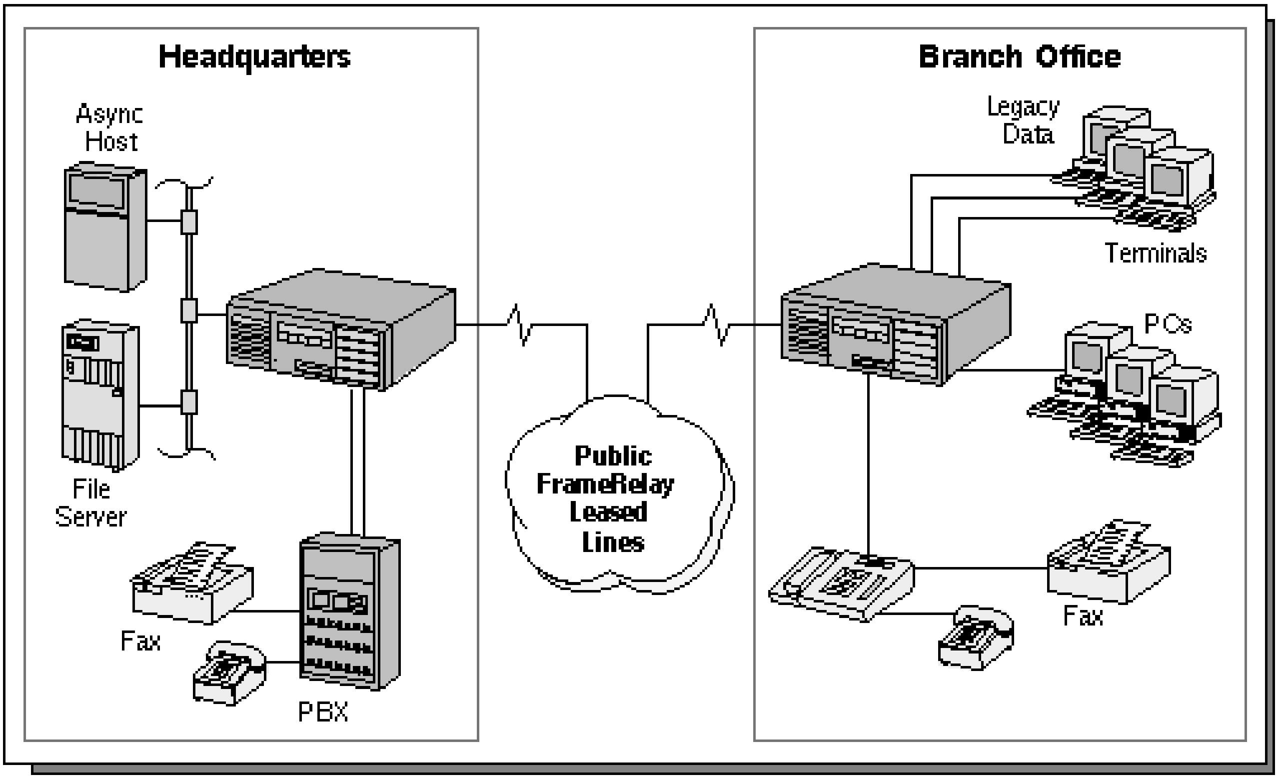


FIGURE 9

Integration Market

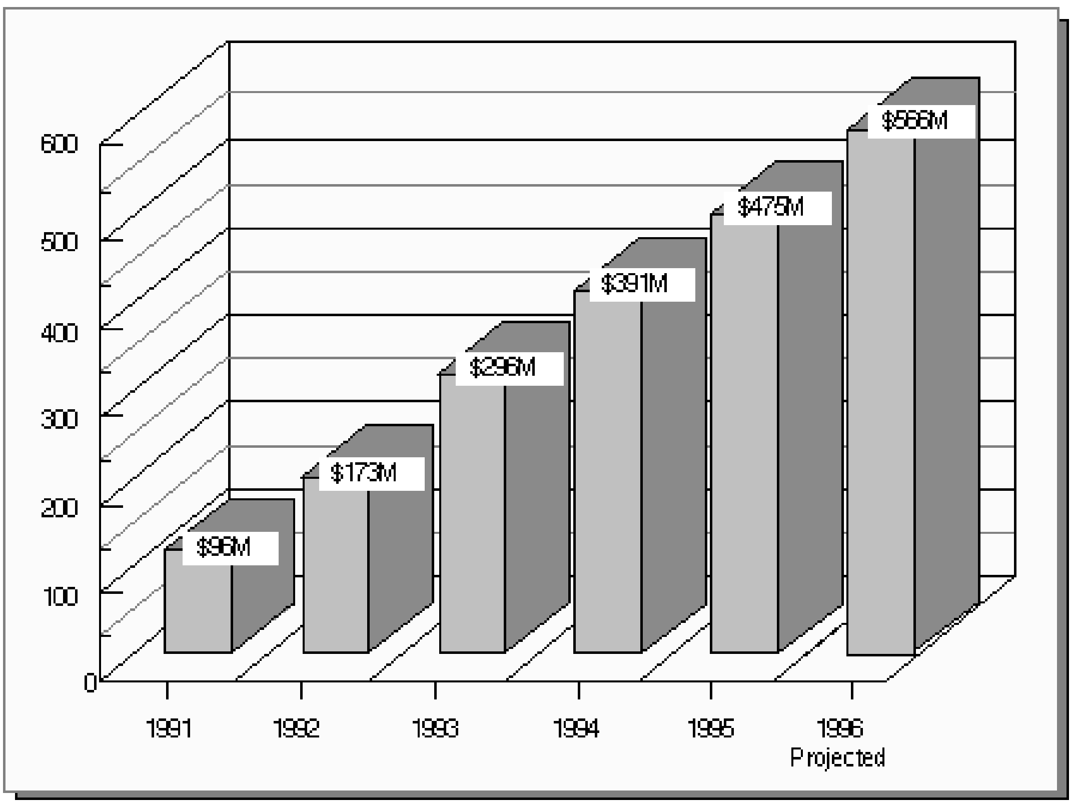
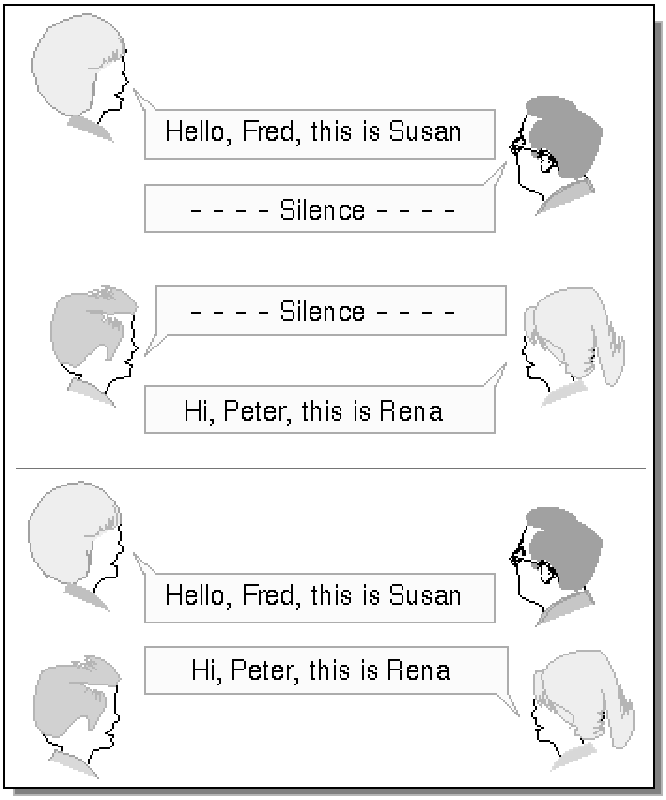


FIGURE 10

Silence Suppression



Today, the model for public networks is one with a frame relay interface and often a cell-relay backbone. These backbones are high speed—E3 (34 Mbps), T3 (45 Mbps), some even up to 600 Mbps—providing very fast transmission through the clouds. They are based upon technologies such as asynchronous transfer mode (ATM) or switched multimegabit data service (SMDS), which provide fast, predictable delivery of time-sensitive information. In addition, networking efficiency is very high, allowing transport of time-sensitive voice, video, and facsimile, along with data and LAN traffic, across the cloud.

The public frame relay networks are prepared to carry voice; now it is up to the individual organization. When deciding to integrate an enterprise network, it is important to make sure that the products chosen provide the best technologies available.

What to Look for When Integrating an Enterprise Network

Of primary importance is the quality of the voice and amount of bandwidth it needs. If it doesn't sound good, no one will want to use it. Toll-quality voice, the same as what is heard over the public telephone network, is required. This means the voice must have good audio and very little delay. Second, voice should use a minimum of bandwidth to allow other types of traffic. Digitized voice compression techniques make this possible.

Voice Compression

In today's public telephone network, the spoken words (analog) are converted to digital in order to travel through the network. Pulse code modulation (PCM), the name of this technique, is the benchmark for toll quality voice. It requires 64 kbps of bandwidth which, while optimized for speech quality, is not very efficient for integrated networking applications. Newer compression techniques, such as adaptive differential pulse code modulation (ADPCM) and low-delay code excited linear-prediction (LD-CELP), which use 32 kbps and 16 kbps respectively, as well as several vendor-proprietary algorithms, have been developed to provide good quality voice while requiring minimum bandwidth.

The current standard for voice compression is ITU G.729. This International Telecommunications Union algorithm's full name is conjugate-structure algebraic-code-excited linear-predictive (CS-ACELP) coder. It delivers an exceptionally high level of voice quality. The G.729 standard provides toll-quality voice and uses only 8 kbps of bandwidth. This is a definite requirement for voice over frame relay.

Speech, like data, is not contiguous; it comes in bursts as each person talks. Since it is bursty, voice can be further compressed and quality increased through silence suppression and background noise regeneration.

Studies show that conversations contain significant periods of silence. In fact, according to Bell Laboratories, these periods can make up as much as 60 percent of the conversation. When one person speaks and the other is listening, there is a 50 percent savings. The pauses between words account for another 10 percent. Through silence suppression, the silence is not digitized, freeing up bandwidth on the voice channel. Then, bandwidth can be used by speech or data from another channel. Silence suppression reduces the bandwidth requirement to an average of approximately 3.5 kbps during a conversation (see *Figure 10*).

While silence suppression saves significant bandwidth when used alone, it could result in "unnatural" silence between speech bursts for the listener. In order to save bandwidth and still offer maximum call quality, a technique called background noise regeneration is used. A measure of the speaker's background noise is sent across the link and stored at the receiving end of the conversation. From then on, whenever silence is detected, the local equipment regenerates the background noise from memory. This preserves the natural sound of the call without using any bandwidth for silence.

Prioritization

Ensuring reliable delivery and minimal delay for voice traffic could be a problem when trying to multiplex several traffic types over a link. In order to insure the delivery of these, the integration product must properly prioritize the traffic and minimize the transmit queue congestion. The optimal prioritization is as follows:

1. voice
2. protocol-sensitive sync (e.g., DLC)
3. async, LAN

When multiplexing several data types, queuing delay can affect voice quality. Hence, it is important to minimize the amount of time voice packets spend in the transmit queue. This is done by controlling the size of the lower priority packets, which insures that there is never too much low-priority data in the queue ahead of voice.

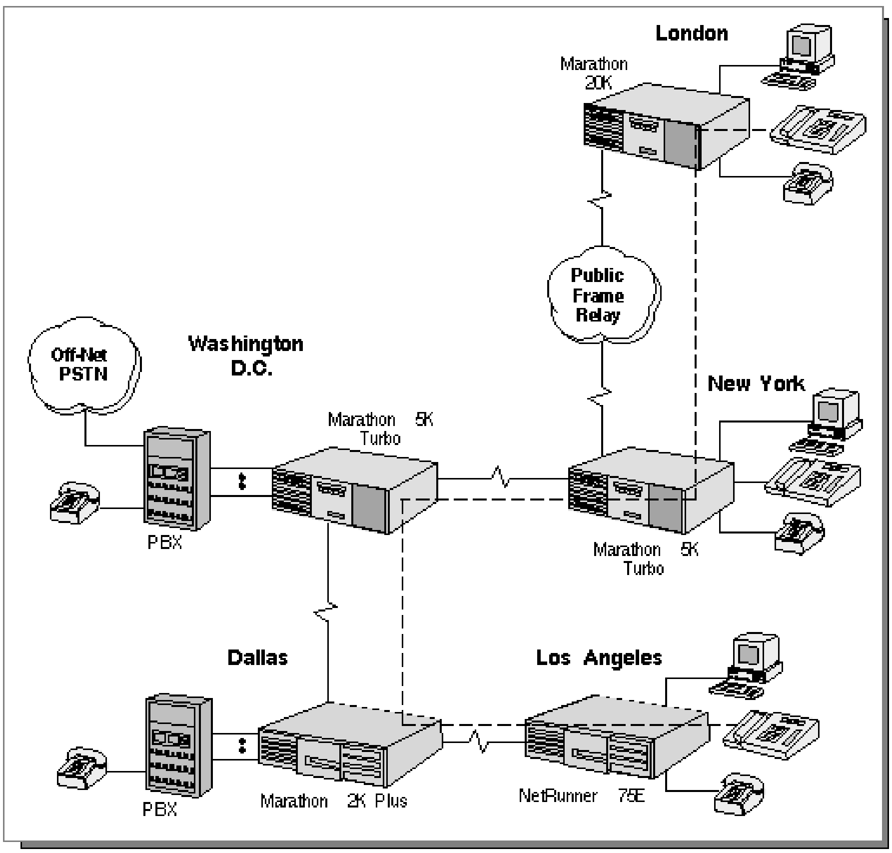
Voice Switching

Sometimes it is impractical to have direct connections (fully meshed) between all sites. In fact, networks are often designed with the minimum number of links required to connect all sites. Star or cascade topologies might be examples. Maintaining high-quality digitized voice in these networks introduces new challenges. With voice switching technology, the user can build a multi-site, integrated network with the same full connectivity of the logical mesh network, independent of the actual network structure. This means that even if there is no direct link between sites, the compressed digitized voice can be directed through multiple FRADs to reach its destination without leaving the network (see *Figure 11*). A network with voice switching performs like a simple distributed voice switching system or public branch exchange (PBX).

Without voice switching, a call between two sites not directly connected would have to be routed first to the headquarters, placed onto the company PBX, possibly converted to analog, and then redigitized and sent to the remote site. The result

FIGURE 11

Voice Switching



would be reduced quality, longer delay, and even higher network costs. Voice / facsimile can also be switched between FR and leased lines in hybrid networks.

Contention

Since phone lines are not used all the time, a one-for-one voice capacity is not necessary. This means that the number of voice ports at each location is dependent on the amount of time the ports are in use rather than the number of locations with which they can connect. Voice contention allows multiple originating voice connections to share a smaller number of receiving channels. This is especially true at major locations, like headquarters, where a small increase in the number of voice channels can greatly increase capacity. Voice switching and contention features go a long way to help keep the network costs down.

Facsimile Integration

To achieve full savings of integration, facsimile traffic between sites can also be carried over the FR network. Automatic facsimile demodulation will automatically detect a facsimile. Instead of trying to handle a modulated (analog) signal that would require PCM at 64 kbps to digitize, this technique determines what kind of facsimile modulation scheme is being used and converts it back to its digital format. The digital signal typically requires only 9.6-kbps or less bandwidth. The receiving channel then remodulates the facsimile signal for transmission to the receiving facsimile machine. Switching and contention features also apply to facsimile connections.

How Much Bandwidth Does Voice Use?

The best integration products use a combination of technologies to ensure the highest quality voice while not materially impacting data throughput. With state-of-the-art technologies, voice channels require very little WAN bandwidth. Following is an example using Micom's ClearVoice:

- The G.729 standard delivers toll-quality voice using only 8 kbps of bandwidth.

- Silence suppression technology takes advantage of the pauses in conversations.

- A phone is in use an average of 25 percent of the time (2 hours during an 8-hour day)

Putting this all together:

ClearVoice	‡	Silence Suppression	‡	25% Rule
8 kbps	‡	3.5 kbps	‡	1-2 kbps

An efficient voice channel, as shown in *Figure 12,* consumes an average of 1-2 kbps of WAN bandwidth, leaving the rest for other traffic.

Anything Else?

In addition to delivering efficient high quality voice, several other technologies are also required to insure the most successful implementation of voice over frame relay.

Multiplexed PVCs

Some FRAD products on the market require a different PVC for each traffic type. This means that for LAN, legacy data, and voice to travel between a remote site and headquarters, three PVCs would be required. If another remote site were added, six would be required, and so on. Since carriers charge for each PVC, having to use many of them will make a network expensive even with the cost savings of frame relay. Also, the guaranteed CIR bandwidth is assigned to each PVC, making administration more difficult and less efficient. With multiplexed PVCs, costs for extra PVCs are eliminated because all traffic types share a single PVC. The same bandwidth rules previously discussed apply. In *Figure 13,* which compares multiplexed and nonmultiplexed PVCs, instead of a PVC for each data type, only two would be required—one for each site—quite a saving on PVC costs.

Premise Frame Relay

Some FRADs provide an integral frame relay switching capability called premise frame relay. Premise frame relay allows the user to add voice to existing enterprise FR networks (see *Figure 14*). Data-only FRADs connected to premise frame relay can pass their traffic transparently, through an integration device, to the FR network. This allows complete integration by adding toll quality voice to the network without losing investment in existing equipment.

FIGURE 12

Voice Compression

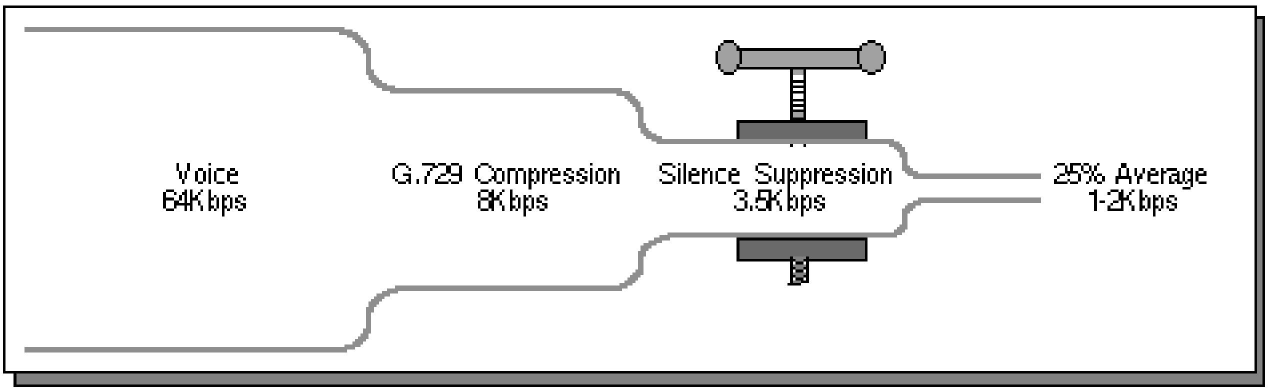


FIGURE 13

Non-multiplexed vs. Multiplexed PVCs

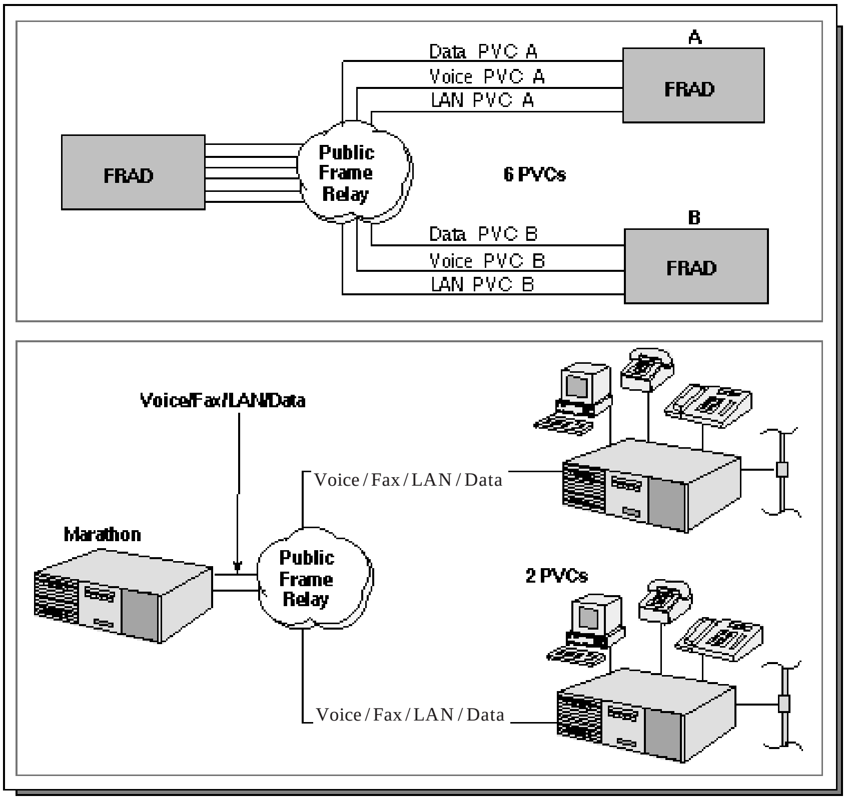


FIGURE 14

Adding Voice to Your Existing Frame Relay Network

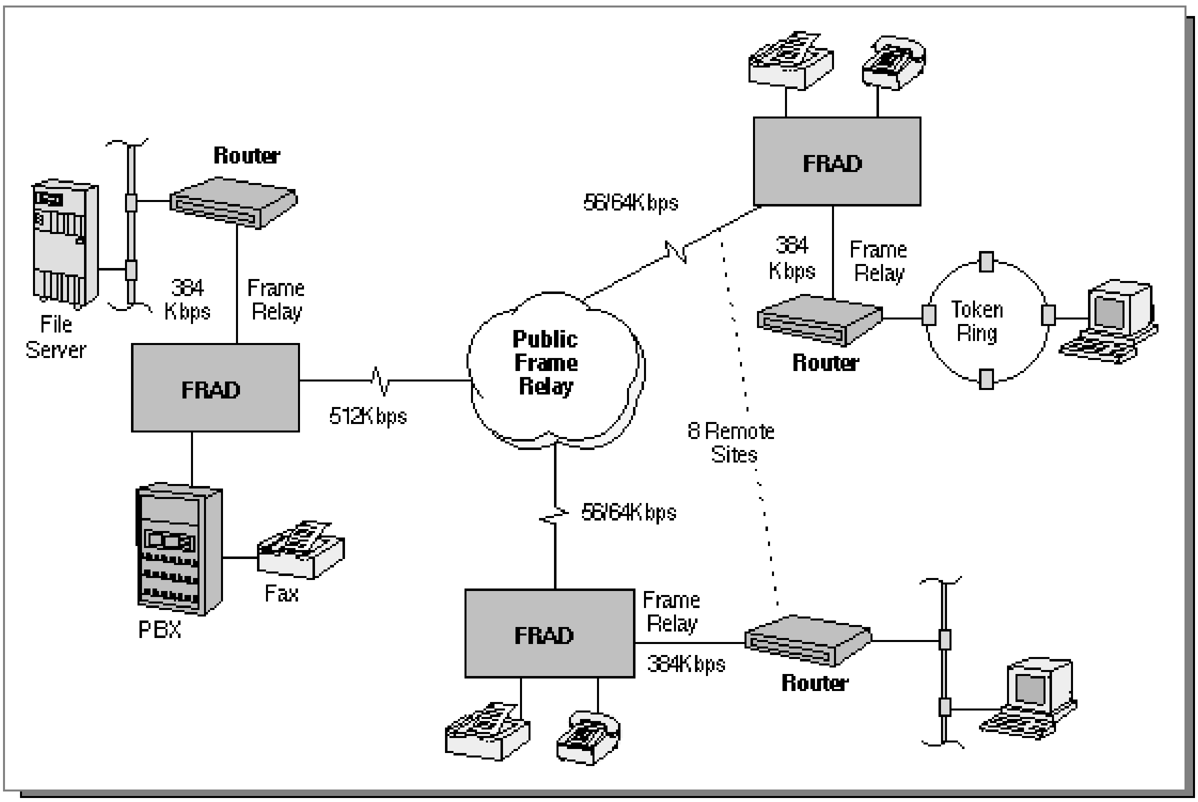
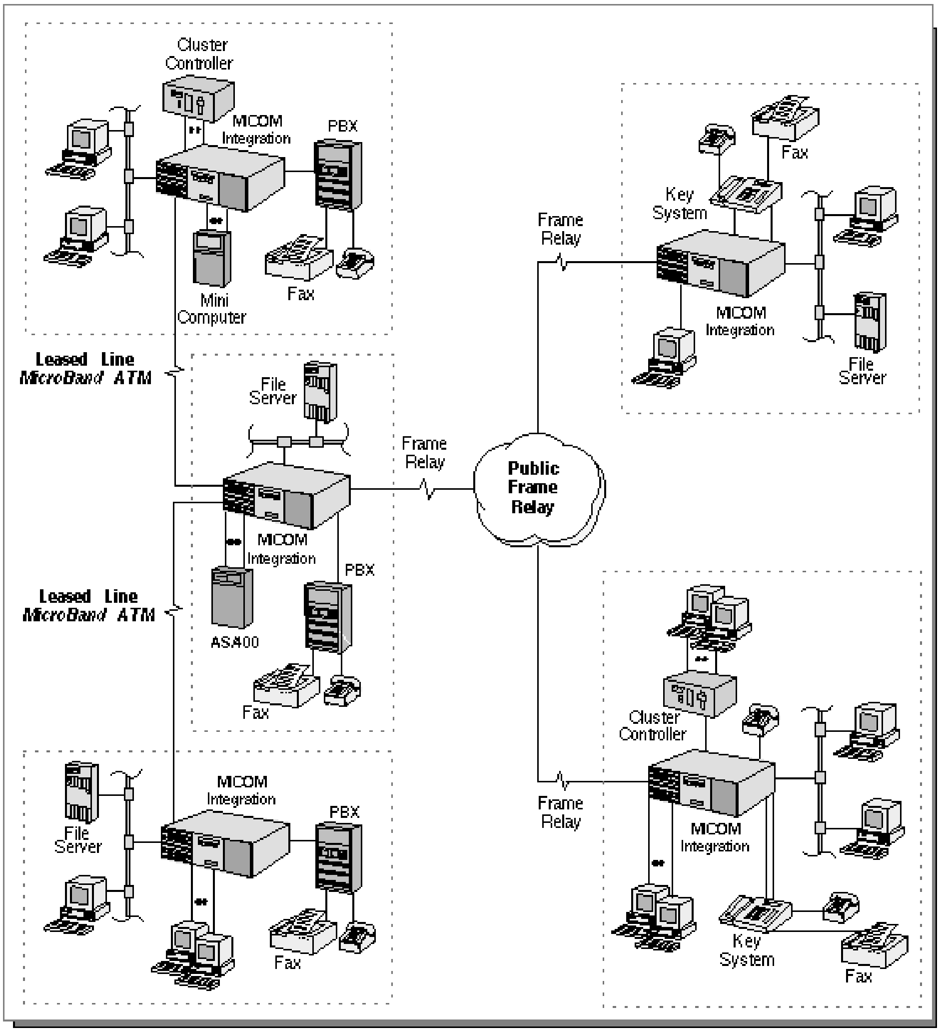


FIGURE 15
Hybrid Network



Flexible Network Design

Cost considerations aside, users should remember that frame relay is not a panacea. At times, leased lines will be more appropriate. As conditions change, the user needs products that can use one or the other service to best advantage. A key product feature is to allow such flexibility. A network that minimizes costs and maximizes performance sometimes requires a combination of public FR and private leased line networking—hybrid networking, shown in *Figure 15*.

Most important is the potential added cost savings and design flexibility that result from being able to choose the best available services for all segments of a network. With a seamless hybrid network, the user does not know if a phone conversation is being transported over the leased line or the frame relay environment.

When is hybrid networking the best to use? When FR connections are not available—which may still be the case in many geographical areas—or when public FR services are more costly than leased lines. Frame relay may be very cost-competitive over leased lines in one area, but not in another. This hybrid networking model could also be used for gradually migrating the network to frame relay from leased lines.

To build the most cost-effective hybrid network, the potential user should look for integration products that offer many of the frame relay benefits in leased line environments. The most important are:

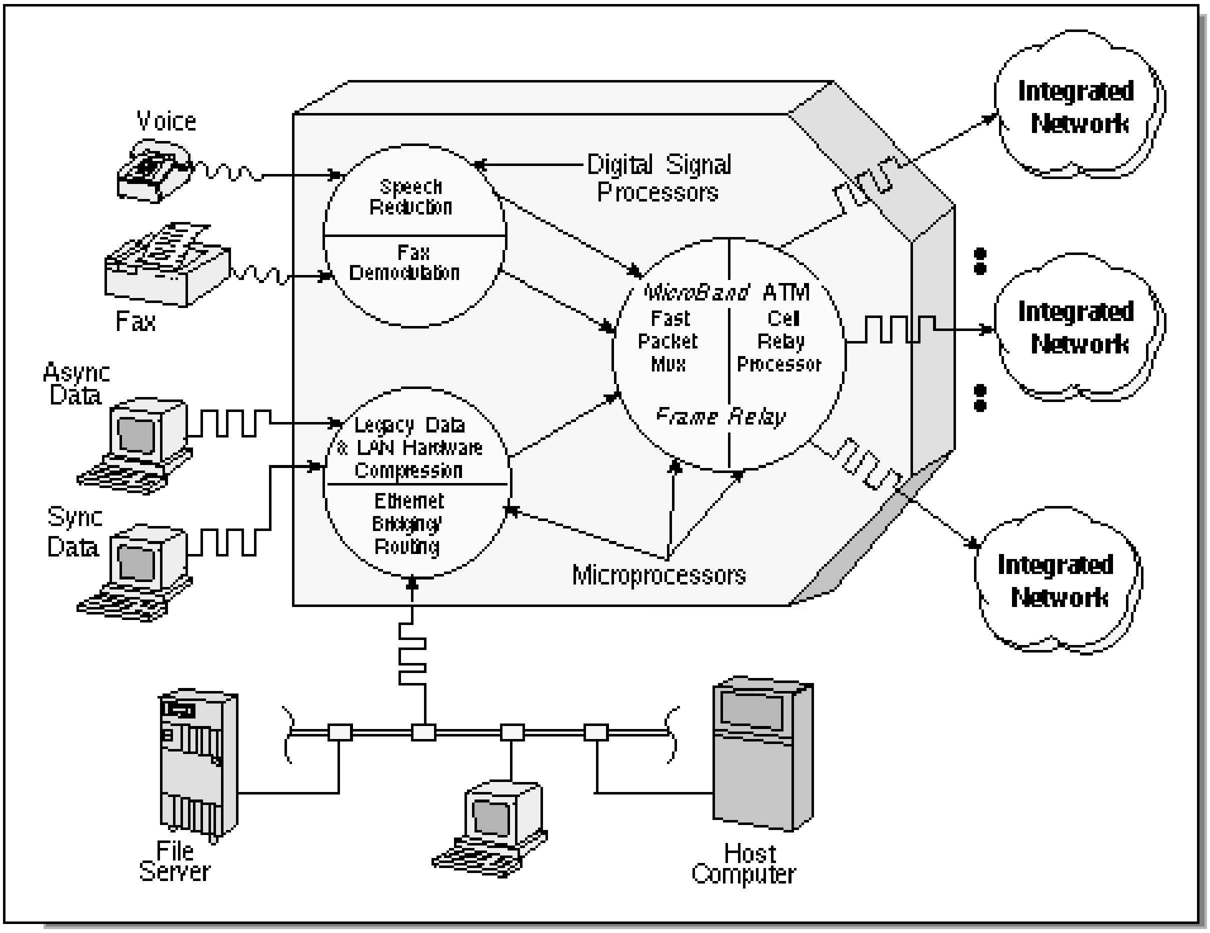
- high quality voice
- bandwidth management
- high availability
- seamless operation (including voice switching)

Typically, the voice technology used in frame relay environments is also used over leased lines. It is important to look for toll quality voice and efficient use of bandwidth.

Efficient use of bandwidth for voice is especially important. Less bandwidth required for voice means more bandwidth for data. It should be remembered that, unlike frame relay,

FIGURE 16

Combining Technologies



leased line bandwidth is fixed. If the peak traffic requirement is higher, then link speed performance is degraded. When integrating voice and facsimile over leased lines, especially slower (<T1) links, it is important to use the most efficient protocol.

Bandwidth-on-Demand

As with frame relay, a user should not have to pay for unused bandwidth on the leased line side either. Bandwidth-on-demand offers many bandwidth management benefits for leased line environments.

Bandwidth-on-demand is a technique for providing additional bandwidth when conditions require, such as during peak traffic periods. In this case, there would be a primary leased line link between two nodes. During average conditions this link is enough. When bandwidth requirements increase, a secondary link would be initiated, increasing traffic capability between the two nodes. The additional bandwidth can be added based on utilization or time of day.

Bandwidth-on-utilization causes the second link to be initiated automatically when the link reaches a predefined threshold. When the traffic decreases, the link is automatically terminated. Time of day can be used to bring up additional bandwidth based on a preconfigured schedule. For example, additional bandwidth might be brought up in the morning when everyone starts work and multiple file transfers are initiated. When the traffic levels off, the additional link is terminated.

Both of these functions save money by enabling you to purchase leased line bandwidth for average conditions but bring up additional bandwidth as necessary using switched digital services like ISDN, Switched 56, or modems, to maintain a high level of performance when required.

The bandwidth-on-demand feature can also be used to back up a failed link between sites. In the event of a lost link, a backup link is automatically initiated, insuring network reliability and availability in leased line environments.

Standards and Approvals

Of course, the chosen product should support all frame relay standards, for example, FECN and BECN flow control. It should also be approved for connection to all the most popular FR services and switches. It should also meet the safety and type approvals required in different countries around the world.

Summary

The best products provide a combination of all technologies, which allow the user to take full advantage of the attractive frame relay benefits including cost savings, network flexibility, and improved performance, plus achieve additional savings by integrating voice and facsimile along with LAN and data traffic. With seamless hybrid networking, users can combine private and public networking to get the best of all possible combinations (see *Figure 16*). The resulting savings pay for the network hardware, and from that point on, the savings go directly to the bottom line.

When evaluating products for integrating voice over frame relay, users should look for:

- complete connectivity of voice, facsimile, LAN, and legacy data

- voice technology that delivers toll quality voice and the most efficient use of bandwidth

- the flexibility to use the best networking method to meet the needs of a particular network

Intelligent Peripheral Realization

Thomas E. Newman

Enhanced Services Development Executive, Telecom and Media Solutions

IBM

This essay shares one vendor's perspective of building an intelligent peripheral (IP) and some experiences in doing so. It first will introduce characteristics of the IP that make it different from intelligent networks of the past. It will then explain IBM's design objectives in developing an IP system and the actual design that emerged. Some of the lessons IBM learned from the experience will be discussed, as well as some directions for future IP development that resulted from this experience.

IP Characteristics

The basic difference between the intelligent peripheral (IP) and other types of intelligent network elements is the actual network connection to the customer that IP offers (see Figure 1). The connection to the customer via the IP provides an opportunity to interact. Thus the service provider must deal with all the asynchronous customer events—everything that switch vendors have had the problem of dealing with for many decades. Opportunities to provide classic kinds of services such as voice response and voice recognition devices offer another area for IP development, as does integrated messaging, which can link voice mail to e-mail and similar services. Bringing messaging into the intelligent network is a very important function of service node implementations of IP. Another important impact of IP is in billing, because IP functions require longer holding times than transactional interfaces with service control points (SCPs).

Russell C. Sivey of GTE Telephone Operations captured the essence of what makes IP different when he said that IP is "one place where we can make a customer very happy or very sad." The IP captures a very different kind of interface to the customer than, for example, service control point (SCP) technology, which does not allow direct interaction with the customer. This makes all the difference.

Design Objectives

As IBM set out to enter the IP market, it developed several specific objectives. First of all, IBM wanted to develop capabilities that would allow it to compete in a broad market that included wireless as well as wireline. Another important consideration was that the system design solution would have to allow both intelligent peripheral and service node implementations—everything from voice mail through the classic IP services—within the same network elements. The design solution would have to be appropriate for a worldwide market as well as a national or North American market, because many customers need IP capabilities throughout the world. Finally, the platform would have to be open enough to interwork with other types of intelligent services that are currently available. Voice mail, as mentioned earlier, is one such service, but the IP system would also have to interact with external databases that contain valuable information which could be brought into call processing, and it would have to provide access to, and be accessible from, the Internet. The Internet opens up a new wave of opportunities because it, like the telephone network, is universally accessible. Interworking intelligent services combines universal access on the data side with universal access on the voice side, offering new opportunities to make services available to customers. Among those opportunities is letting customers control many of their own services from the multimedia-rich Internet.

Design Considerations

As at the start of any project, the company's developers felt they could achieve all these design objectives. Unfortunately, as any developer also recognizes, many things can get in the way of accomplishing the dream of extending capabilities to customers to allow them control over their own services. Issues such as cost and scalability emerge in service creation. From a cost perspective, how far down could the platform go? How far up from a performance perspective could the platform go to adequately handle companies that take millions of calls per day?

FIGURE 1
Classic IN Architecture

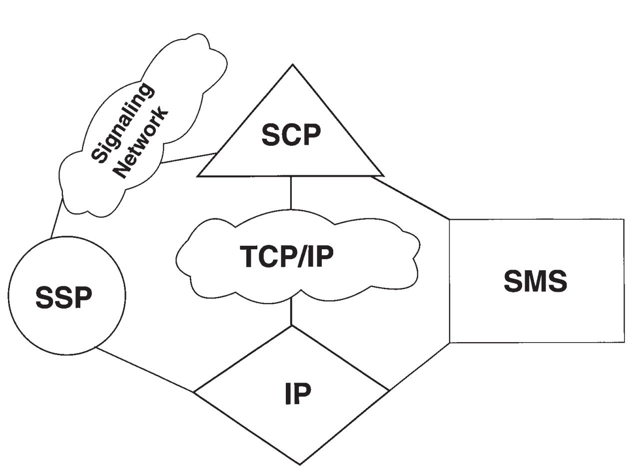
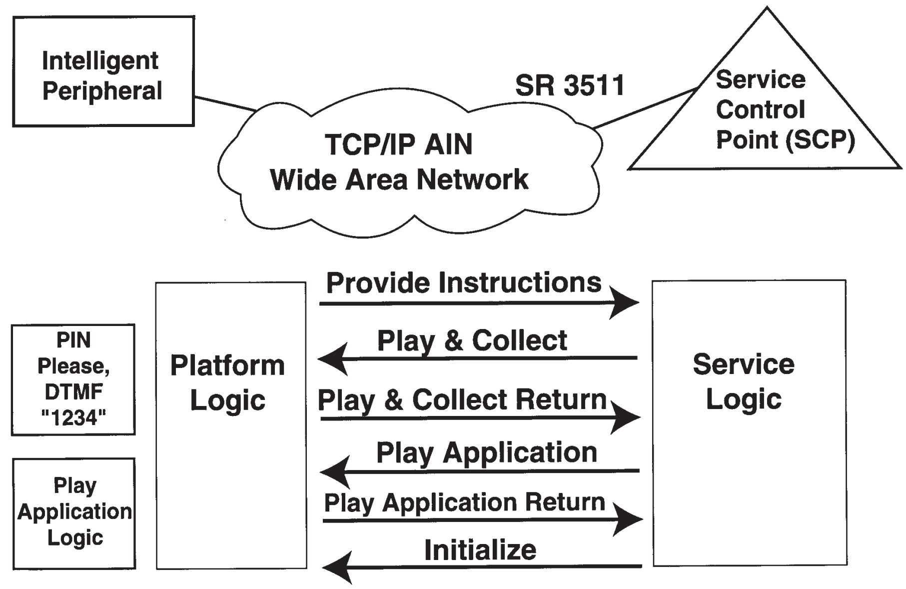


FIGURE 2
1129+ Protocol Flow



Another problem is that standard protocols do not exist. To compete in the marketplace, designers had to determine how to implement an open architecture that utilizes some of the proprietary protocols that switch vendors offer for tasks such as voice-activated dialing. System reliability was yet another issue to be addressed.

What strategy did IBM develop, given these design considerations? The company decided on a consistent hardware structure throughout the platform. It was important from an operations viewpoint to keep a common hardware platform throughout the product line. IBM chose one of the best server environments available, the RS/6000, as the basic server environment. The system is UNIX-based and offers high availability with automatic failover and recovery.

Another important requirement was for a relatively robust, high-function voice platform because voice services were critical for many of the initial customers. IBM selected the Direct Talk/6000 product line for its voice platform. Utilizing off-the-shelf hardware and software as much as possible can help manage costs, so IBM tried to use off-the-shelf products rather than building from scratch. IBM used the service creation environment in the Direct Talk/6000 product line. It utilized open network communication protocols (AIN and TCP/IP) to provide interfaces to other intelligent network elements, such as the service control point (see *Figure 2*).

A critical aspect of the design solution was that the platform needed to enable customers to build services wherever they wanted (see *Figure 3*). A classic view of the intelligent net-

FIGURE 3
Platform vs. Service Application

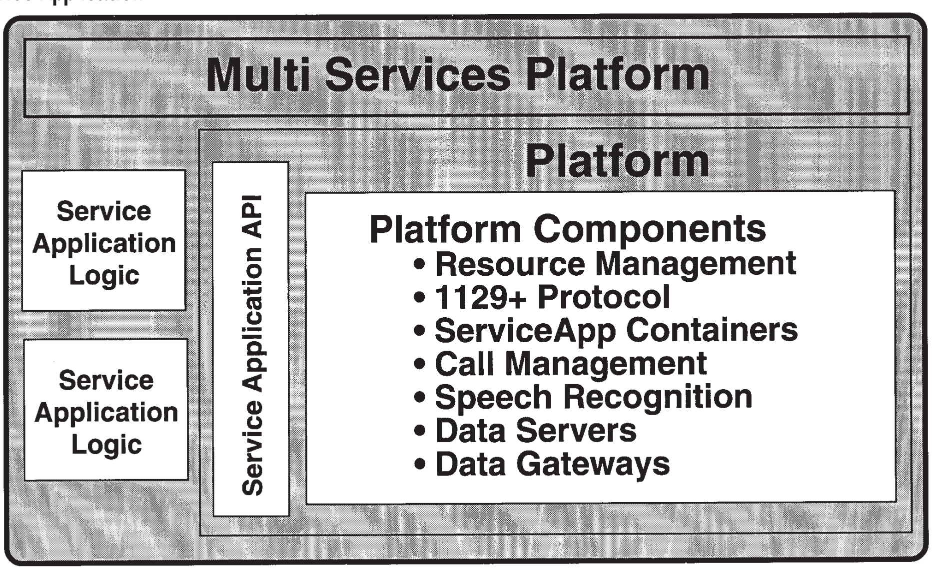
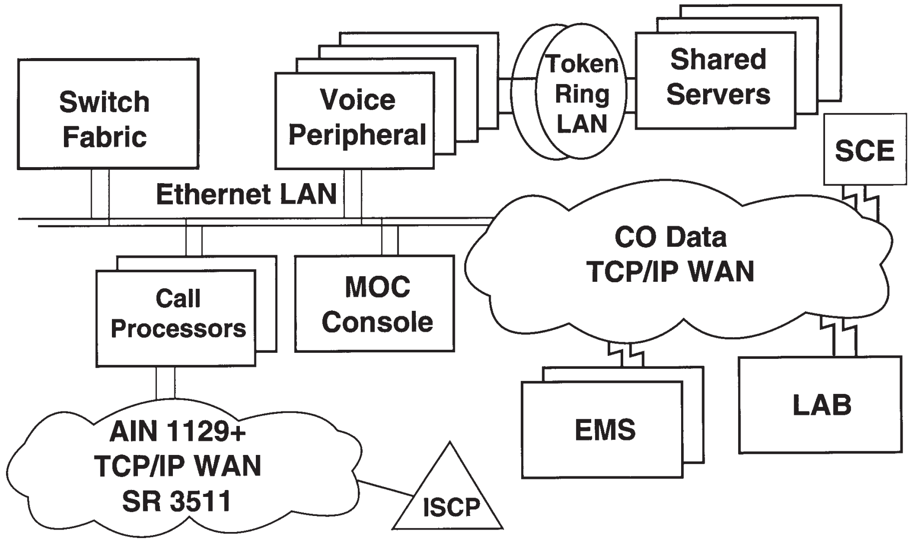


FIGURE 4

MSP/6000 R1.0 Hardware Connectivity



work houses the service logic in the SCP, and the intelligent peripheral remains a relatively dumb device that plays announcements, collects digits, and so forth. That was not IBM's strategy. Instead, IBM provided a robust service creation environment in the voice peripheral platform itself to enable customers to decide how much logic they want in the service node or IP and how much logic they want to keep in the SCP. The system supports an approach in which the SCP directs the IP to perform simple activities, as well as a much more robust service creation environment that allows customers to accept a play application command from the SCP and then house all the service logic for that application in the IP. That is the key strategy to IBM's solution. This approach gives customers a relatively open application programming interface (API) which allows any application developer to take advantage of the whole range of services that are available on the intelligent peripheral—all the way from speech recognition capabilities to switch-controlled applications such as conference bridges.

Many customers have high availability demands that require a robust dual-rail, dual-processor environment. IBM used an implementation that is very similar in structure to some of the successful solutions it has used in the past in the SCP environment (see *Figure 4*).

The first release offered a switch fabric that allows capabilities such as conference calls, outdials, etc., because many customers needed that service. Over time, these switching possibilities will become an optional feature in the platform.

The voice peripheral, which is the heart of the system, again is based upon the Direct Talk / 6000 product family. It provides the basic processing point for most of the service logic for handling calls and accessing a shared server base that offers speech recognition and similar devices, databases, etc.

Finally, the system links to the wide area network (see *Figure 5*), which helps manage the whole intelligent peripheral or service node environment. In addition to the IPs themselves, which are spread throughout the network, the system has the capability to provide an element management system that enables customers to manage those IPs more effectively. This capability allows anything from remotely monitoring the IPs to actually creating the services in a lab environment, testing them in the lab environment, and distributing them into the IPs. Most of the work can be done remotely from the IP, so the IP can be an unstaffed site that manages and monitors services as well as creates and distributes services and data remotely from the platforms themselves.

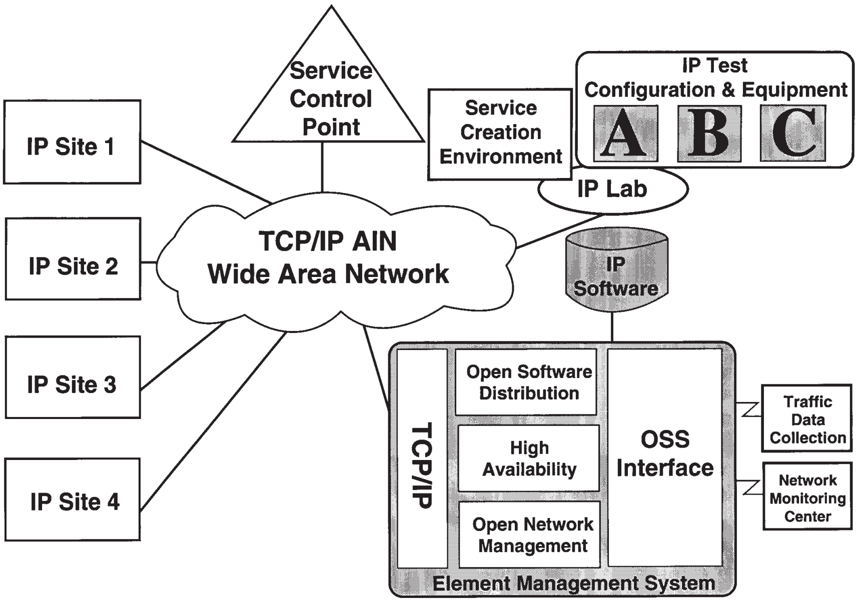
Lessons Learned

The first lesson IBM learned is that IP design is very difficult. If any company promises a provider an easily built and installed IP, the provider should look elsewhere, because it is likely to be an empty promise; it requires a lot of work. It also requires a very significant resource investment. IBM learned that the hard way. One of the initial mistakes IBM made was to underestimate the size of the job.

The IBM experience showed that the IP system could be made to quickly handle all the basic services, such as call management and speech recognition. The details were the really difficult part of the system design. The operations impact, service creation strategy, and reliability were the elements that made it difficult. Planning the entire process for implementation, including all those details, is the key to successful installation.

IBM used a strategy of buying off-the-shelf products that sometimes worked well and sometimes did not. I still think it was the right strategy, but in retrospect the company could have benefited from more diligence. Another problem area was integration of the diverse products chosen, which proved difficult. Tighter project management in the area of integration is very important. In addition, the project would have benefited from a longer interval for beta testing.

FIGURE 5
Wide Area Network



Critical aspects of the project that contributed to its success, on the other hand, included very good relationships with customers. The customers stuck with IBM and stumbled when we did. The dedication of the whole team—both at IBM and at partner companies such as Bell Atlantic, Southwestern Bell, Bell Canada, BC-Tel, and others—made success possible. Finally, successful IP implementation requires commitment. There will be hard times, and any participant must be committed for the long haul.

The Future

What is the future of IP implementation? As already mentioned, many opportunities come from Internet connectivity. Potential applications such as Internet control of telephone services, managing conference calls via the Internet, and a range of other service options open up by linking the Internet to the telephone network.

In terms of market expansion, the IP has not penetrated too deeply into the wireless environment; market growth will require more aggressive movement into that area. IP application also will expand worldwide. Some Core-INAP capability has been implemented in Europe, but it has not been deployed yet. Nevertheless, worldwide expansion will come.

Another area that will be developed is service creation. The service creation environment must evolve to take advantage of new technologies. Scalability and cost issues also must be addressed. The platform must accommodate small-scale implementations that make it easy to install and very cost-effective. Finally, IBM intends to be much more aggressive than it has been in the last year or so, in getting involved with setting standards to really make this particular market take off.

Exploiting the Opportunities for Intelligent Networks

Jo Piggot
Director
Schema

Introduction

The birth of the intelligent network (IN) in the mid-1980s was expected to herald a new era of improved efficiency and faster service creation. By freeing telephony services from some of the constraints of network infrastructure, especially the time needed to make network changes, innovative new services could be designed, tested, and launched in a fraction of the time taken for switch-based solutions.

Although today many operators can demonstrate success with IN–based services, there are still some important challenges for telecom operators to overcome before the market really takes off. For example:

- Each new service must add complexity that can extend the time taken for interoperability testing and avoiding feature interaction.
- Improvements in service creation are not always matched by improvements in billing systems or customer support processes.
- Although the use of toll-free and VPN services is growing strongly in many countries, some other IN services (for example, personal numbering) are taking longer than expected to become established.

Nevertheless, the European market for IN services is set to grow dramatically in the next five years, more than tripling in size to reach an annual value of more than $15 billion by 2002.

To gain an understanding of how the IN services market will evolve, Schema's research focused on a range of key questions and issues:

- Has the promise of IN been realized to the extent that operators are achieving significantly improved time-to-market for new IN–based services?
- What is the leading IN service in Europe, and how will this picture change by 2002?
- Is there any evidence to indicate that operators are using IN capabilities to launch innovative services and therefore differentiate their offerings in an increasingly competitive environment?
- Having started out as a technology for the fixed network environment, is IN being introduced in the mobile area to support new services and fixed-to-mobile integration?
- Which vendors are driving the market, and what new entrants are appearing?
- How important are IN standards in the market evolution process?
- What is the impact on IN deployment plans of the EC requirement for number portability?

The results of Schema's research reveal that significant investments in IN are being made by European operators, paving the way for the trial and launch of a wide range of new services. The findings also highlight an increased use of multiple IN services among large European businesses and a growing level of interest from smaller companies. However, they also suggest that telecom operators need to broaden their awareness and marketing campaigns in order to provide potential users with a clearer picture of the business benefits and opportunities presented by such services.

What Is Intelligent Networking?

An intelligent network is a communications network that is controlled by a software "services" layer, enabling new services to be developed without the need to modify network switches. In an IN environment, voice calls arriving at a public network switch are "suspended" while the switch asks an associated computer system for instructions on what to do with the call. For example, on receiving a call to an 800 number, the switch asks the computer where the call should be routed. The look-up reveals the "real" telephone number, and the call is allowed to proceed. Call-routing instructions may also vary according to a wide range of parameters, such as time of day or day of week.

This introduces an important IN concept: the service (the number translation process) has been removed from the infrastructure (the network). The 800 number is meaningful to the subscriber and independent of the network numbering.

Key elements of an IN include:

- *The service switching point (SSP).* This is a public network switch that has the ability to suspend calls while asking the computer what to do next.
- *The service control point (SCP).* The computer that "controls" the service.
- *The service management system (SMS).* This system allows the provisioning of IN services and provides management functions.
- *The service creation environment (SCE).* This is an important IN component for those operators that wish to develop their own services. It has capabilities for building services from telephony building blocks, usually via an easy-to-use graphical interface.
- *The intelligent peripheral (IP).* An optional IP provides the network with specialized "resources" such as announcements and tone listening. Typically, this will be directly connected to the switch.
- *The service node (SN).* A service node is a single system that combines the functionality of an SSP, SCP, and IP. This may be an attractive option for small operators looking for a lower-cost IN solution or for larger operators testing new services on a small scale.

In addition to these functional components, an important element of an IN solution is the service software that forms the basis for an IN service.

A key benefit of the IN approach is that changes to an IN service can be made more efficiently than would have otherwise been the case. This is because only a single alteration is necessary in the central computer, rather than having to make modifications to the software in every network switch (a time-consuming and expensive task).

Schema's study assesses the use of and prospects for seven major IN service types:

- *Toll free:* Also referred to as Freephone, toll-free services are those where the cost of the call is paid for by the called party—that is, they are free to the caller.
- *Toll shared:* Toll shared (sometimes known as split-charging services) are those where the cost of the call is shared between the caller and the called party. Typically, the caller pays the equivalent of a local-call charge, with the called party paying the remainder.
- *National rate:* This service type provides a single, location-independent number for all calls to an organization, with callers paying the full cost of the calls at a national rate.
- *Premium rate:* Premium-rate services (also known as shared-revenue services) are those where callers pay more than the standard charge for a call to reflect additional value provided in the form of entertainment or information. The revenue collected from callers is shared between the telecom operator and the organization providing the entertainment or information service.
- *Calling cards:* A calling card allows the user to call from any telephone and charge the cost of the call to an account other than that associated with the telephone—for example, an employer's account. A prepaid calling card allows calls to be made up to the value of the pre-payment.
- *Personal number:* These services offer a single number for callers to contact a party who may otherwise be reached via several other numbers (e.g., home, office, mobile). Typically, the call is delivered to the correct number through a sequence of call attempts, with the first based on time-of-day and then by redirection on no-answer. A range of features such as paging, call screening, fax or data traffic, and messaging services are normally available with such services.
- *A virtual private network (VPN):* A VPN is a public network service designed to connect PBXs in such a way that the customer has a high level of functionality (comparable to that which can be obtained with private networks).

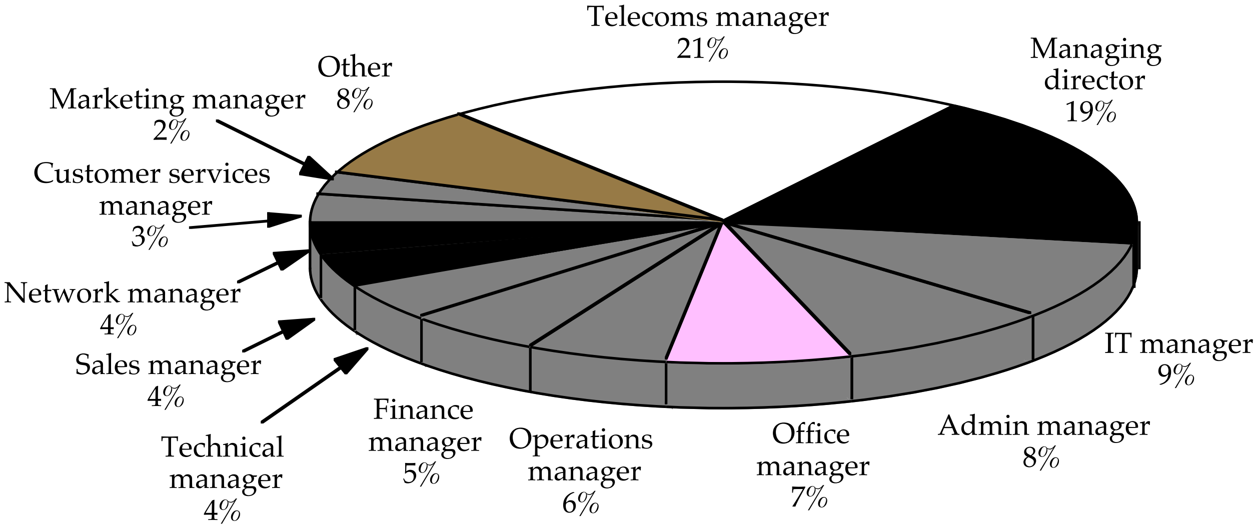
Toll-free, toll-shared, national-rate, and premium-rate services are often collectively referred to as number translation services. National-rate services are currently offered by relatively few operators and represent only a tiny proportion of the total IN services market. Readers should note that these types of services are referred to as "IN services," even where they have been implemented using a non–IN solution.

Study Objectives

The overall objective of the study is to provide telecom operators and suppliers with detailed insight into IN market developments, which will help them to identify opportunities and target resources more effectively. To provide this insight, the research program aimed to:

- investigate the awareness and use of IN services among corporate users across Europe
- understand the key technological developments that will affect future products and services
- review available products and supplier strategies
- assess the current size and value of the market for IN products and services and predict likely future growth

FIGURE 1
Distribution of Survey Respondents According to Job Title



- evaluate the driving forces influencing the decision to invest in IN and analyze the experiences of operators to date
- assess how operators approach the process of new service development
- assess the current status of IN standards

Methodology

The study research program took place between January and July 1997 and had four major strands:

- research with more than 30 leading suppliers to assess their IN product and marketing strategies
- investigation of IN deployment and service rollout plans through discussions with more than 30 metropolitan, regional, mobile, national, and international telecom operators
- a major telephone survey of more than 500 business organizations throughout Europe to assess user awareness of and requirements for IN services
- detailed case studies of 6 leading users of IN services

The telephone survey took place during April and May 1997 and involved structured telephone interviews with 505 companies in a range of industry sectors across Europe. An analysis of the telephone survey respondents is shown, together with a breakdown of the interviews by country, industry sector, and organization size. The results from the various strands of research formed a major input to the market forecasting process, which was based on a number of models developed by Schema (see *Figure 1* and *Tables 1-3*).

The Market

IN Services

In 1996, the major IN services in Europe generated revenues approaching $4.5 billion, over half of which was accounted for by number translation services. VPN services accounted for 26 percent of service revenues. By 2002, the European market will have an annual value of more than $15 billion in service revenues (*Figure 2*).

Number Translation Services

- The European market for toll-free / toll-shared services was worth more than $1 billion in 1996 and is expected to grow at a compound annual growth rate of about 30 percent over the study period to reach an annual value of approximately $5 billion in 2002.
- The United Kingdom is the most well-developed market for toll-free / toll-shared services, accounting for 40 percent of traffic and 35 percent of European revenues in 1996. This reflects the United Kingdom's position as Europe's leading market for call centers and telebusiness.

TABLE 1

Distribution of Telephone Interviews by Country

Country	Number of interviews	%
Austria	31	6%
Belgium	29	6%
Denmark	30	6%
Finland	31	6%
France	60	12%
Germany	59	12%
Italy	53	10%
Netherlands	42	8%
Norway	30	6%
Spain	51	10%
Sweden	29	6%
UK	60	12%
Total	**505**	**100%**

TABLE 2

Distribution of Telephone Interviews by Sector

Industry sector	Number of interviews	%
Finance / insurance	90	18%
Business services	83	16%
Retail / wholesale	80	16%
Transport	79	16%
Manufacturing	94	18%
Government / utilities	79	16%
Total	**505**	**100%**

- By 2002, the United Kingdom will still account for 25 percent of the revenue from toll-free / toll-shared services. Other leading markets will include France and Italy (20 percent and 18 percent of toll-free / toll-shared revenues in 2002, respectively.)

- A relatively under-developed call-center market—Germany—accounted for only 10 percent of European toll-free / toll-shared revenues in 1996 but will increase its share slightly over the study period to 12 percent of revenues by 2002.

- The market for premium-rate services has grown rapidly during the late 1980s and early 1990s in most European countries. However, it has also been subject to wild fluctuations and instability in a number of countries due to the impact of sudden regulatory changes and the poor image associated with "adult" services.

- Worth nearly $1.5 billion in 1996, the premium-rate services market is now maturing and even stagnating in a number of European countries. Some revenue growth is taking place in the market for "serious" business premium- rate services.

- Over the study period, the premium rate services market will grow slowly to reach a value of $2 billion in 2002. Leading country markets include France and the United Kingdom.

VPN Services

- Major corporate organizations are becoming enthusiastic adopters of VPN services, attracted by the opportunity to save costs and achieve a consistent level of functionality across all company sites.

- In 1996, European VPN revenues totaled $1.2 billion in which 50 percent was generated by U.K. VPN users. The United Kingdom has a mature VPN market, with both major national operators having launched VPN services during the early 1990s and other operators such as Energis following suit.

- Users of international VPN (IVPN) services such as those offered by AT&T—Unisource, Concert, and Global One—also accounted for a significant proportion (15 percent) of 1996 revenues.

- By 2002, annual VPN revenues will total just under $6 billion, with the United Kingdom still accounting for more than a quarter of these revenues.

Other IN Services

- Although calling-card services represent a huge market in the United States, they have only become significant in recent years in Europe. Their use has been stimulated by international players such as Global One and by a variety of new operators. However, in markets in the Nordic countries where the penetration of mobile communications is high, calling cards are still relatively insignificant.

- The adoption of personal number services has been slow to date in Europe, with subscribers in the low tens of thousands. Take-up has been restricted by a combination of high

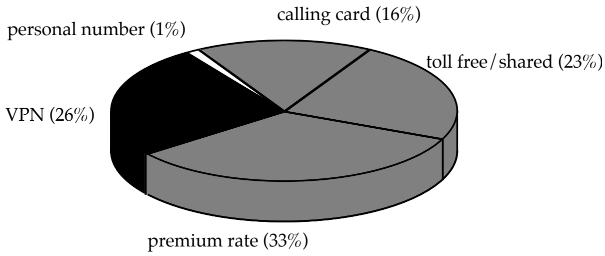
TABLE 3

Distribution of Telephone Interviews by Organization Size

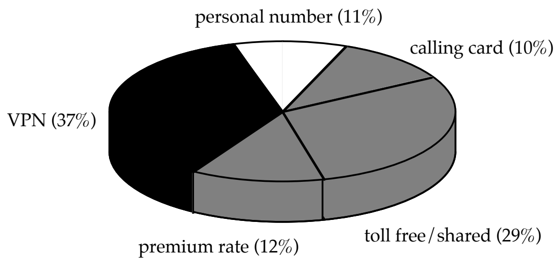
Organization size (employees)	Number of interviews	%
1-10	95	19%
11-100	88	17%
101-500	61	12%
501-1,000	87	17%
1,001-5,000	112	22%
5,001-10,000	23	5%
10,001+	35	7%
Not stated (501+)	4	1%
Total	**505**	**100%**

FIGURE 2

Share of European IN Service Revenues by Service Type, 1996 and 2002



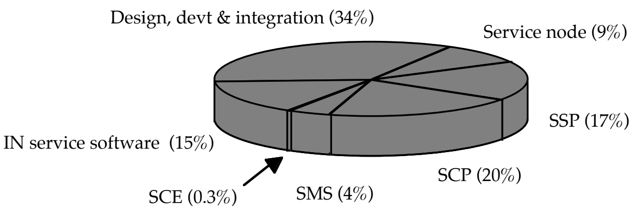
1996 revenue shares



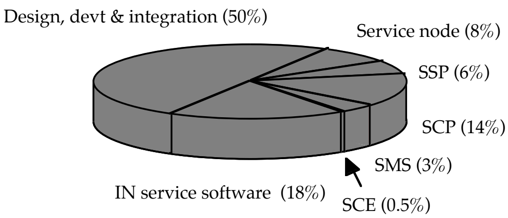
2002 revenue shares

FIGURE 3

Share of European IN Service Revenues by Product Type, 1996 and 2000



1996 revenue shares



2000 revenue shares

Note: SSP revenues do not include the public network switch, only the SSP software required to IN-enable the switch.

pricing, complex services and tariff structures, and channel resistance. Over the study period, personal number services are expected to become more popular, principally among mobile communications users. By 2002, revenues from personal number services will reach $1.7 billion.

IN Products

The European IN market has developed since the mid-1980s and witnessed an increasing amount of activity in the first half of the 1990s as a growing number of operators deploy IN infrastructure to support new services.

The move toward IN has typically been led by the incumbent PTO in each country. However, in competitive markets such as Finland and the United Kingdom, other operators have quickly followed suit. The United Kingdom is the extreme example with a number of national, regional, mobile, and metropolitan operators all having implemented IN in some form. Deployment continues to accelerate as emerging operators across Europe and the major IVPN operators introduce IN capabilities and services.

Schema's analysis indicates that:

- annual European IN product revenues will grow from $175 million in 1996 to a peak of just under $300 million in 1999.

- from 2000 to 2002, annual revenues will decline somewhat, representing the tail end of a major buying cycle across Europe.

FIGURE 4

Awareness of IN Services

Source: Schema survey; base = 505 European businesses.

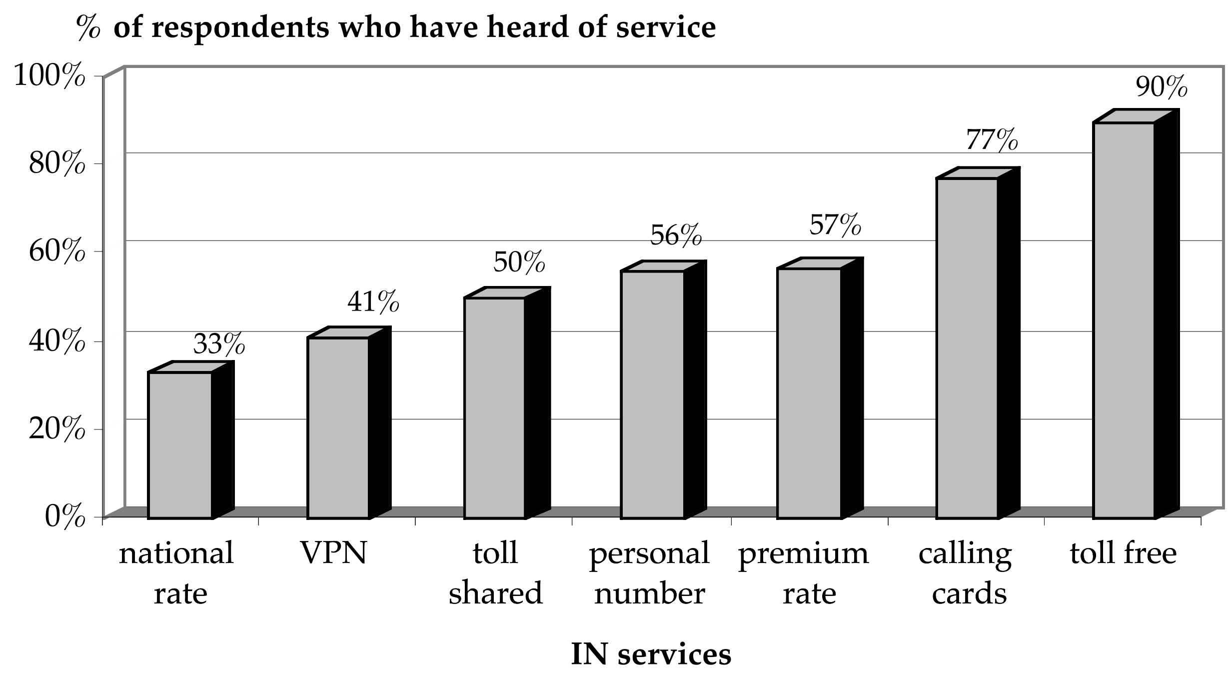


FIGURE 5

Current and Potential Use of IN Services

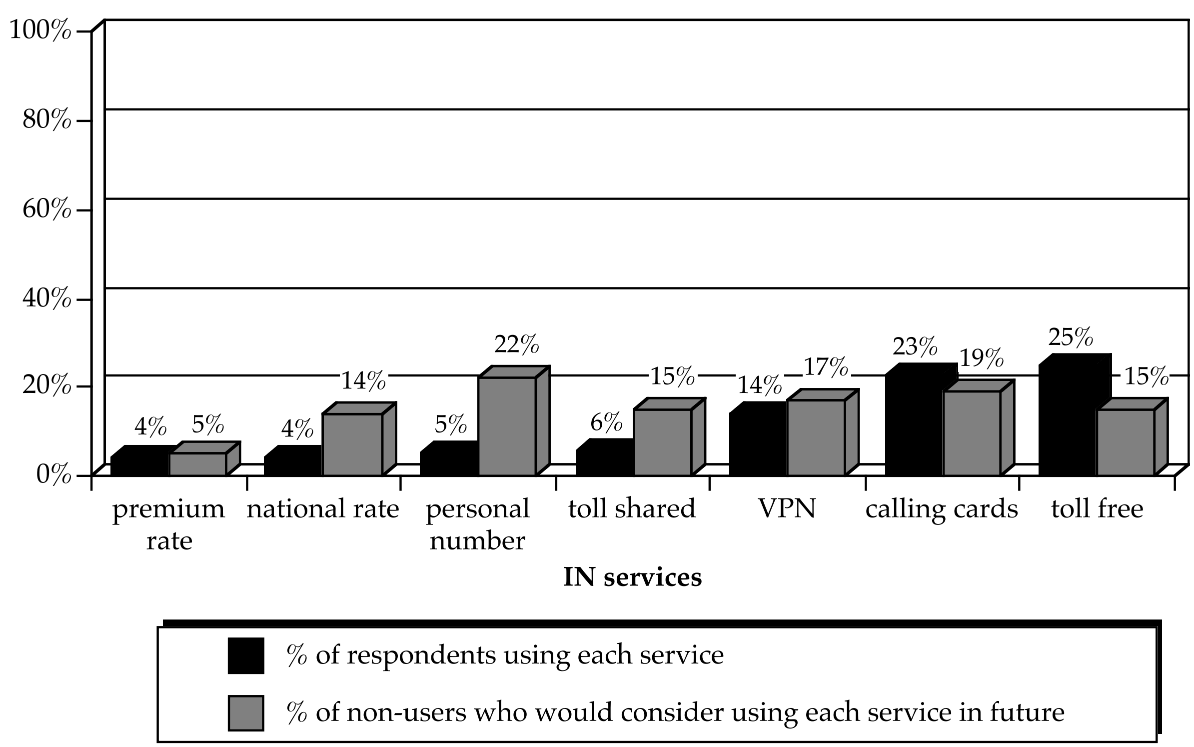
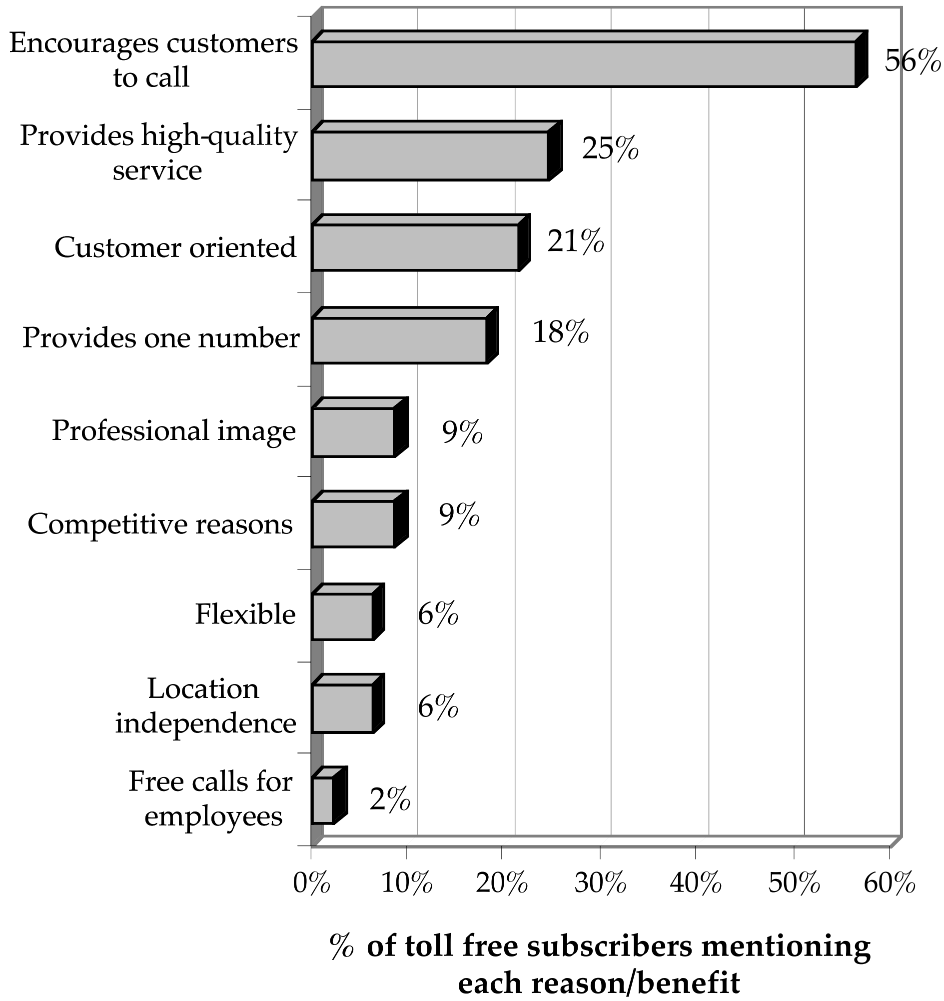


FIGURE 6
Benefits of Toll-Free Services

Source: Schema survey; base = 126 toll-free subscribers.



- over the seven-year period from 1996 to 2002, cumulative IN product revenues will total $1.5 billion, representing ongoing and new investments by over 120 operators.

- a large and increasing proportion of revenues will be generated by two main components—IN service software and the professional services associated with IN design, development, and integration. In 1996, these two elements accounted for 15 percent and 34 percent respectively of annual revenues, rising to 18 percent and 50 percent in 2000.

- of the remaining product elements, SCPs will generate the largest proportion of revenues—20 percent in 1996 and 14 percent in 2000.

- expenditure by U.K. operators will account for the largest proportion (15 percent) of IN product revenues over the period 1996–2002, followed by Germany with 11 percent and France with 9 percent. Germany, in particular, is already witnessing the emergence of new national operators such as o.tel.o and VIAG Interkom and is also expected to see the appearance of a number of metropolitan operators.

- in addition to expenditure by national operators in each country, IN investments by the IVPN operators in Europe will account for about 3 percent of total European product revenues (*Figure 3*).

User Perspectives

Schema's research with European businesses reveals that IN services are mainly in use by larger companies, and that there are few, if any, mass-market IN services at present *(Figure 4)*.

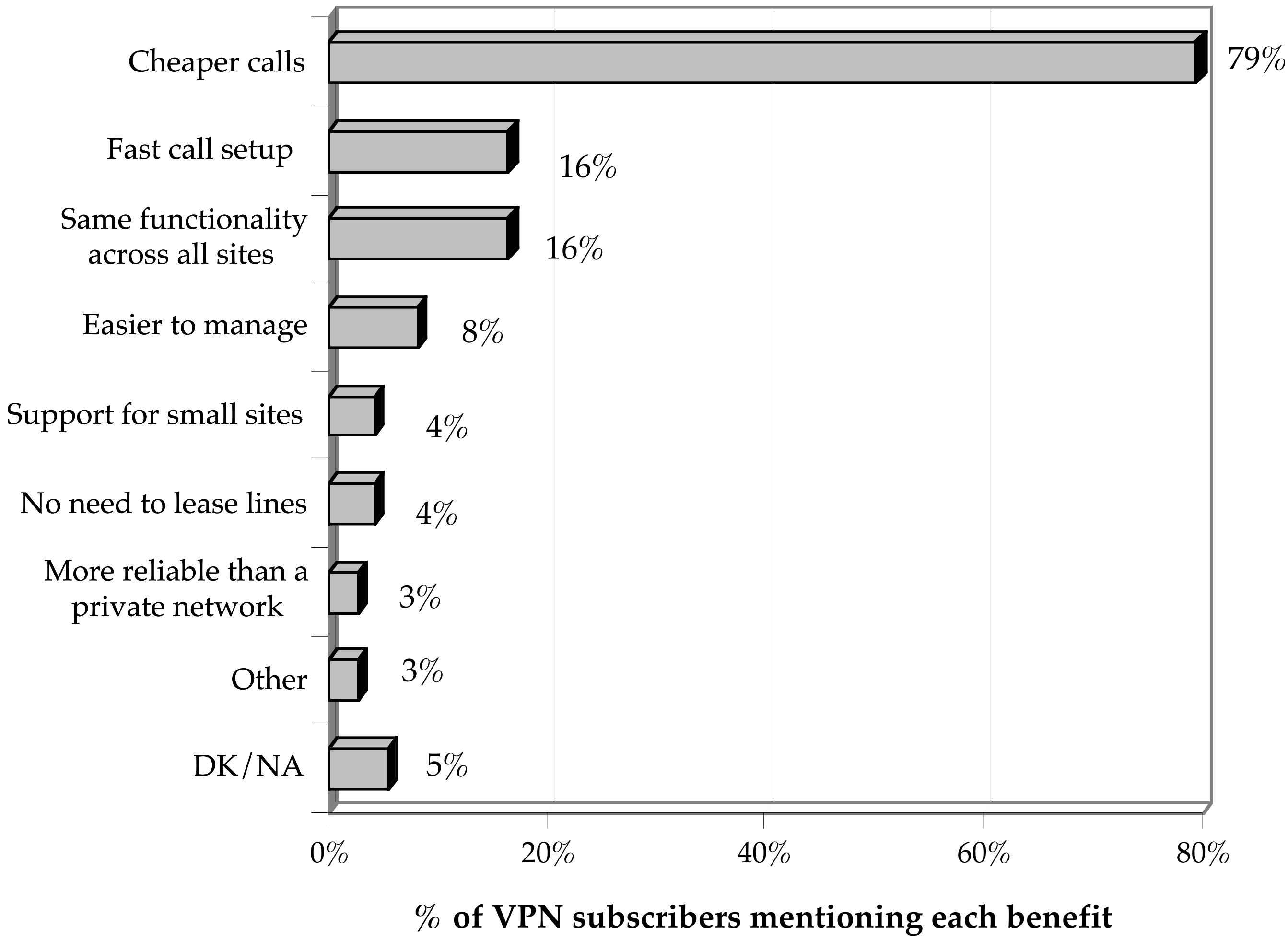
- Among European businesses, there is a high level of user awareness of established services such as toll free and calling cards.

- A wide range of services is in use among large organizations and, to a lesser extent, medium-sized companies.

FIGURE 7

Benefits of VPN Services

Source: Schema survey; base = 73 VPN subscribers
Source: Schema survey; base = 505 European businesses



- With the exception of the calling card, few IN services are in use by small organizations.

The research also suggests that there is considerable potential for the take-up of additional services (*Figure 5*):

- The service of most interest to non-users is personal number, with more than 20% indicating an interest in it.
- VPN and calling card services, already among the most used IN services, are of some interest to non-users.
- All the number translation services, with the exception of premium rate, are of interest to about 15% of non-users.

Corporate users perceive that IN services deliver a range of business benefits:

- Users of toll-free and toll-shared services believe that offering this form of access to customers encourages them to call and demonstrates a commitment to high-quality customer service.
- Premium-rate services are seen as a way to generate revenue from value-added services or to cover the cost of activities such as help desks.
- The potential for achieving cost savings is the overriding factor driving user take-up of VPN services.
- The main perceived benefit of calling-card services is that calls can be made from any location without the need for carrying cash.
- Improved contactability is seen as the primary advantage of personal number services.

Despite the encouraging level of interest in IN services, it is clear that many companies lack detailed information and would welcome the opportunity to evaluate these services more closely. This suggests that there are significant untapped opportunities that could be addressed through more extensive marketing and education programs by telecom operators (see *Figures 6* and *7*).

The Telecom Operators

Operators must invest in IN now if they are to survive in the 21st century. Today, operators are striving to provide value-added, customizable solutions that will give them a competitive edge and protect revenue streams from major corporate customers. This is a lucrative telecom market, and operators without IN will be unable to react quickly enough with sufficiently cost-effective solutions or indeed with any solution at all.

Intelligent networking is a telco initiative born from the desire for independence from suppliers. It was expected that this independence would allow operators the freedom to decide their own destiny and reduce the time to develop new services. In practice, dependence on suppliers is as strong as ever with most operators complaining about the cost and time for suppliers to develop new IN-based services. However, with experience and good advice, development times can be reduced by half.

Schema's findings show:

- there is an insufficient number of skilled IN resources to meet today's demand. Many operators are relying on contractors who previously worked for the longer established PTOs. Even so, there are still too few experts.

- operators changing to IN have seen service development times cut in half to around 6 to 9 months from idea to implementation. However, this has required process reengineering to take advantage of new ways of working.

- there are two schools of thought on the organizational approach to service development. Those operators with a formal customer-supplier relationship between the internal marketing and engineering departments and those that form multi-departmental teams. While the latter seems to be the most enlightened approach, it is the former that can often get the best results.

- service creation is much more complex than many operators first imagine. This means that many still rely on their suppliers to develop new services.

- service creation environments are being used as a rapid prototyping tool by many marketing departments. However, there can be several dangers: it may not be possible to transform the stand-alone prototype into a working network-based solution, and a "fascination with the fancy" can develop all too soon and prevent a real analysis of customer needs.

- few operators have a rigorous enough approach to requirement specification. This leads to ambiguities that result in redesign and ultimately delays.

- structured testing of new services is few and far between with too many operators leaving fault finding to "internal trials." This delays the service launch and does not decrease the probability of a fault being found when in operation.

- many new operators are asking suppliers for marketing consultancy to help them launch and position services. However, some have been disappointed at the quality of advice provided.

Technology and the Vendors

The Schema study looked at a wide range of suppliers of IN solutions and components. They fall broadly into a number of categories:

- "traditional" switch vendors—the companies that supply large-scale switches and associated equipment to the majority of the world's telecom operators. Examples include Ericsson, Nortel, Alcatel, Lucent, and Siemens

- "independent" vendors that produce all or most of the IN solution with the exception of the switches (e.g., Tandem, Stratus, Bellcore, IBM)

- "specialist" vendors that have evolved a solution from some specific product, such as an IVR, including the suppliers of service nodes (e.g., Comverse, Tecnomen, Vicorp)

- IT vendors that supply a platform or key technology to other suppliers in the chain or to system integrators for completion (e.g., Digital Equipment, Hewlett-Packard)

- software and systems integration companies that provide components or build complete systems from an IT platform (e.g., Logica, Sema, CMG, Sligos).

Issues for Suppliers

The emerging market in small IN systems—either service nodes, adjuncts, or scaled down "classical" solutions—is an important factor in providing IN capabilities to operators that would not, in the past, have been able to justify them.

Despite claims of ease of use for the service creation capabilities available from various vendors, many operators are choosing not to buy an SCE (or to buy one but use it only for prototyping and demonstration and evaluation purposes) but to rely on the chosen supplier to carry out the service creation tasks. This leads to three requirements: more off-the-shelf service packages, more investment in the service creation systems since this is still a complex task, and more investment in support for customers in the form of skilled staff.

In the past, the promise of IN to deliver rapid service development has been met to a large extent. This gain has often been largely nullified, however, by the difficulty and lengthy process of integrating the IN platform with the operator's OSSs, such as those that support provisioning, customer care, and, especially, billing. This is now the focus of investment by the major suppliers, both to provide complete off-the-shelf solutions in which the OSSs are included and to create well-defined interfaces that can be used to integrate the IN system with other equipment.

Further investment is coming from suppliers of advanced OSSs with the aim of providing better support for the operator's business processes, tightly coupled to the IN environment. Support for the whole customer life cycle is now becoming available, allowing the reduced time-to-market of the IN to be exploited.

Since capabilities of the IN to create services easily and quickly have led to increasing levels of service specialization and customization, moving toward the "market of one," the management capabilities of the IN platform have to be opened up to recognize this growing diversity. Customer access to the service management function is now an essential requirement. Simple access, such as that possible through a touch-tone telephone to modify simple parameters such as call-forward numbers or time-of-day settings, has been available for some time. What is now required is a means of providing more powerful and easy-to-use access by the customer's telecom manager, and this will be available through an Internet Web browser.

Many of the developments are placing considerable demands on IN platforms in terms of computing power and other factors such as storage. It is interesting to note that virtually all suppliers have been investing for some time now in ensuring that they were positioned to take early advantage of advances coming from the IT industry, both in hardware and software, through the use of as much standard technology as possible. Multi-processor configurations, particularly for SCPs to support high transaction rates, are the norm. IN suppliers also seem keen to take advantage of the ever-increasing processor power which is offered by their IT partners.

Standards

Virtually all of the IN suppliers have implemented key standards, such as the ETSI CS-1 capabilities, as they have become reasonably stable. This is usually modified, however, by a pragmatic attitude toward these standards, with some facilities being postponed in favor of specific feature developments for customers. Many suppliers expect to continue with this strategy, implementing CS-2, CS-3, and CAMEL phases 1 and 2 in the future, although there are many additional requirements which are to be met at the same time that lie outside the current scope of standards.

Fundamentals of IN and IN Capabilities

Barry Price
Principal Engineer, AIN/Wireless Products
DSC Communications

This paper will examine intelligent network (IN) capabilities in detail, focusing on IN functionalities such as call routing services, intelligent peripherals (IPs), and enhanced services. IN capabilities such as voice announcements and messages, text-to-speech, fax transmission and reception, and voice recognition are critical capabilities for putting intelligent services into the network. The paper will address service creation, particularly in regard to network architecture, tools, and users. Finally, it will present an overview of the significance of IN capabilities.

IN Services

Call Routing Services

One basic goal of IN is a call model that drives the switch to recognize a plain old telephone service (POTS) call or an enhanced call that uses AIN capabilities. It is important to understand how a switch processes a call. At specific points in a call (PICs), a mechanism called triggering, or trigger detection points, allows the network to make a decision about routing the call. Within the AIN model, the switch can suspend call processing for a particular telephone call, sending a message through the common channel signaling network using SS7 to a service control point (SCP). The SCP is an intelligent database that makes decisions about routing a call. The IN call models include a variety of trigger points for the origination and termination of telephone calls.

One class of IN services is call routing. *Table 1* provides an extensive list of IN–based call routing services. All of these services involve the AIN call model and SCP architecture to control the routing, screening, and monitoring of calls using AIN trigger points. Calls can be routed based on a variety of decisions: time of day, day of week, originating or destination phone number, or state of destination number (busy or not).

To understand the AIN model, it is also necessary to understand the network architecture. *Figure 1* shows an AIN architecture that builds on the IN architecture in place since the mid-1980s. The IN architecture added SCPs with service control of service switching points (SSPs) via SS7 signaling. The AIN architecture, then, has expanded to include additional trigger points (Bellcore AIN release 0.1 and 0.2 and ITU Command Sets 1 and 2) for more control, and equipment to process calls. The new AIN equipment includes intelligent peripherals (IPs), service nodes (SNs), and adjuncts. An adjunct is a database for information or a stand-alone service such as voice mail. IPs provide data and call processing capabilities, and SNs are IPs with service control. IPs and SNs are important because they are capable of multiple services on the same platform. An IP, SN, or adjunct equipment provides additional capabilities that enable additional enhanced services in the AIN. The general architecture shown in *Figure 1* applies to regional Bell operating companies (RBOCs), interexchange carriers (IXCs), and wireless networks alike.

TABLE 1

IN Call Routing Services

☎ Digit Extension Dialing	☎ 900 Number Screening
☎ Area Number Calling	☎ Call Blocking
☎ Disaster Rerouting	☎ Gateway to Information Providers
☎ Do Not Disturb	☎ Credit Card Calling
☎ One-Number Service	☎ Selective Routing
☎ Screening Lists	☎ Call Allocator
☎ Selected Call Acceptance	☎ Alternate Destination
☎ Preferred Call Forwarding	☎ Call Gate (PIN Screen)
☎ Routing by Day of Week	☎ Personal Access (Virtual Number)
☎ Routing by Time of Day	☎ Call Center Routing
☎ Number Portability	☎ Least Cost Routing

FIGURE 1

Advanced Intelligent Network (AIN)

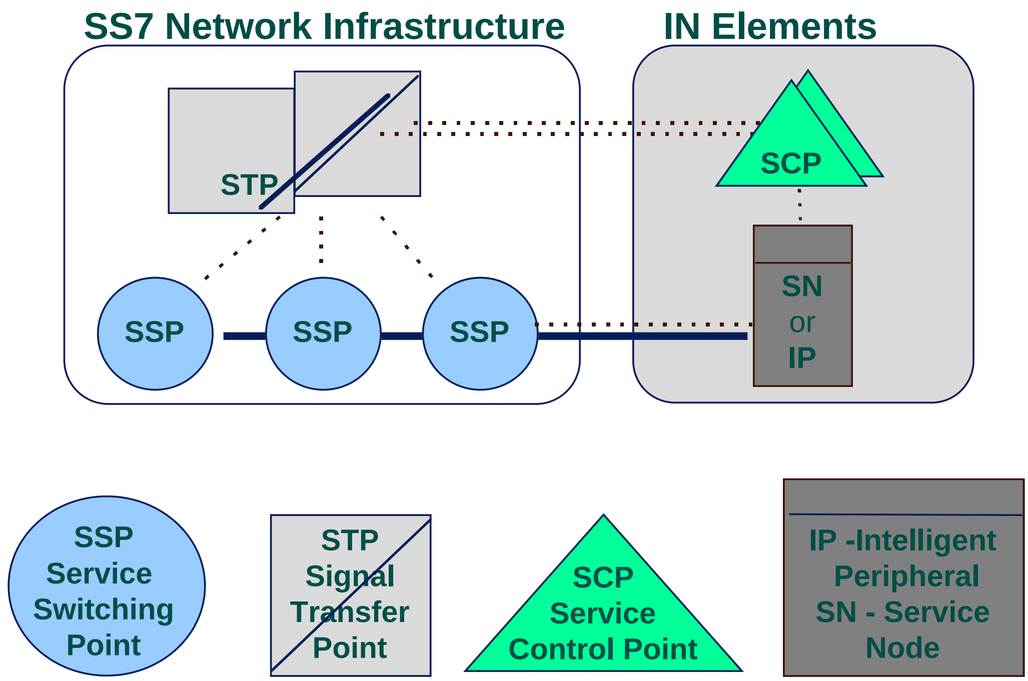


Figure 2 illustrates the wireless intelligent network (WIN). WIN extends the IS-41 standard to support IN capabilities to the wireless network. It also allows control to reside in entities other than home location registers (HLRs); such entities include the SCP and SN. WIN includes three call models—originating, terminating, and mobility management.

AIN provides a step forward in meeting the IN goals of standard open interfaces, vendor interoperability, and third-party developed services. These goals are achieved by standardizing the type of equipment, capabilities, and operational interfaces in the AIN network. IN architecture standardization creates a great advantage for simple service creation, service management, rapid service deployment, and efficient service switching, call processing, and call control. The following paragraphs describe the major components in the AIN architecture.

Service Switching Point

The circles in *Figure 1* represent the switches. A switch with the letters "SSP" has the capabilities to perform two of the functions in the network: call setup for trunking and SS7 interfaces for talking to intelligent databases. Any switch, regardless of its vendor, needs to be loaded with a package to make it an SSP. Other switches may be network access points, which are primarily switches that can only perform certain IN capabilities. This means that it is not a fully implemented AIN–type switch. Many older switches are of this type.

Signal Transfer Points

Primarily responsible for SS7 routing and network management, signal transfer points (STPs) can be seen as the traffic cops of the network. Whenever a problem occurs in the network, they are the nodes that must recognize whether there is a route failure, error failure, or congestion failure. The STPs are always in identical pairs; in fact, most of the SS7 network is engineered in a redundant mode, which means that there is always a second link from the switch. In case of failure, they are each engineered to carry 40 percent of the total expected load. If one dies, then the others should carry the full load.

Intelligent Peripherals

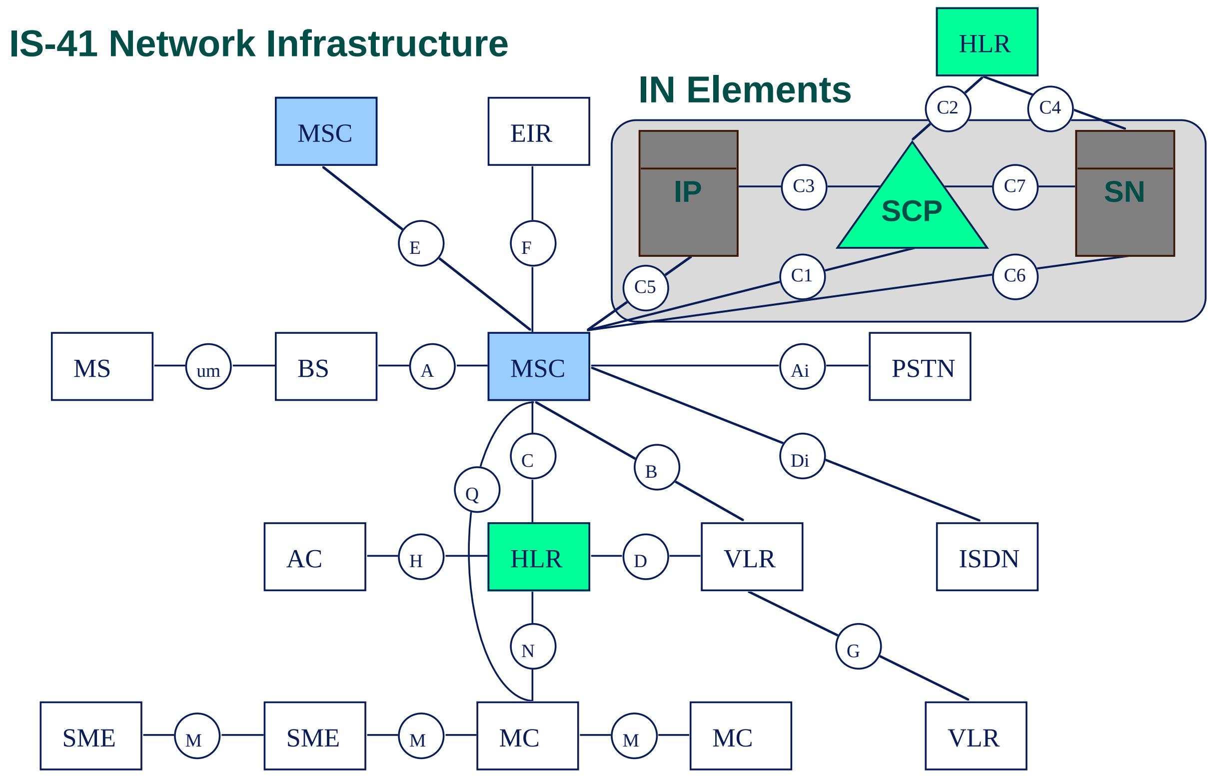
The SN or IP is the major new component in the AIN architecture. The IP adds resources for a variety of new AIN services. These resources include voice recognition circuits, voice synthesis, and voice announcements, among other new capabilities.

Figure 3 provides an example of IP implementation and operations. Its hardware components include media servers, a controller, communications elements, a data disk, and optional switches. All hardware must meet network equipment building standards (NEBS) and reliability requirements. It must also be fault-tolerant and have high availability, or N+1 redundancy. The IP must be scalable from a few ports (single T1) to thousands of ports. This is usually accomplished using a distributed local area network (LAN) architecture, where additional hardware is added at an IP site as needed.

The software components of an IP include media processing (e.g., voice recognition), resource management (connecting voice trunks to media processing), distributed communications, operations management subsystem, signaling interface subsystem, database subsystem, and service control. The two basic network interfaces used are the SSP interface (voice trunking) and the SCP interface (call control). The SSP interface can involve T1s; integrated services digital network primary rate interface (ISDN–PRI), or in-band NFA; direct

FIGURE 2

Wireless Intelligent Network (WIN)



inward dialing (DID); and so on. The SCP interface can be ISDN-PRI, GR-1129, GR-1129+, INAP, SS7, or IS-41, among other possibilities.

There are several typical GR-1129+ transactions, some of which are listed in *Table 2*. In addition to these transaction capabilities, the IP must provide for operations, administration, management, and provisioning (OAM&P) interfaces. *Figure 4* shows some example legacy interfaces that the IP and other AIN network elements must interface to.

Enhanced Services

IPs, as discussed above, are new network elements capable of multiple, programmable services on the same equipment. IPs, along with SNs and adjuncts, enable a new class of IN services known as enhanced services. Consisting of more than simple call routing, enhanced services use new IN capabilities to provide call processing. The SCP can route calls to IP equipment to provide a wide variety of new services.

Table 3 lists some of these IN enhanced services. Some of the services can be provided on a switch without an IP, but having an IP adds flexibility. A few of the most important enhanced services include unified (or multimedia) messaging, voice-activated dialing (VAD), voice-activated network commands (VANC), and personal assistant. Multimedia messaging allows a subscriber to receive and review voice, facsimile, and e-mail messages in a single mailbox. VAD allows a subscriber to dial any telephone number simply by saying a number, name, or word for the person or place being called. Voice-activated network commands allow subscribers to control the system with voice commands ("forward all calls"). Personal assistant enables customers to manage all communications (mail, schedule, VAD, etc.) through a speech-activated interface.

TABLE 2

Typical 1129+ Transactions

IP to SCP	Provide Instructions
SCP to IP	CallInfoToResource, PlayAnnouncement&CollectDigits, IPStayOnLine=True, Play Application
IP to SCP	CallInfoFromResource, InformationReturnBlock, PlayApplicationResult
SCP to IP	CallInfoToResource, Play Announcement, IPStayOnLine=False
IP to SCP	ResourceClear

FIGURE 3

Intelligent Peripheral: Implementation & Operations

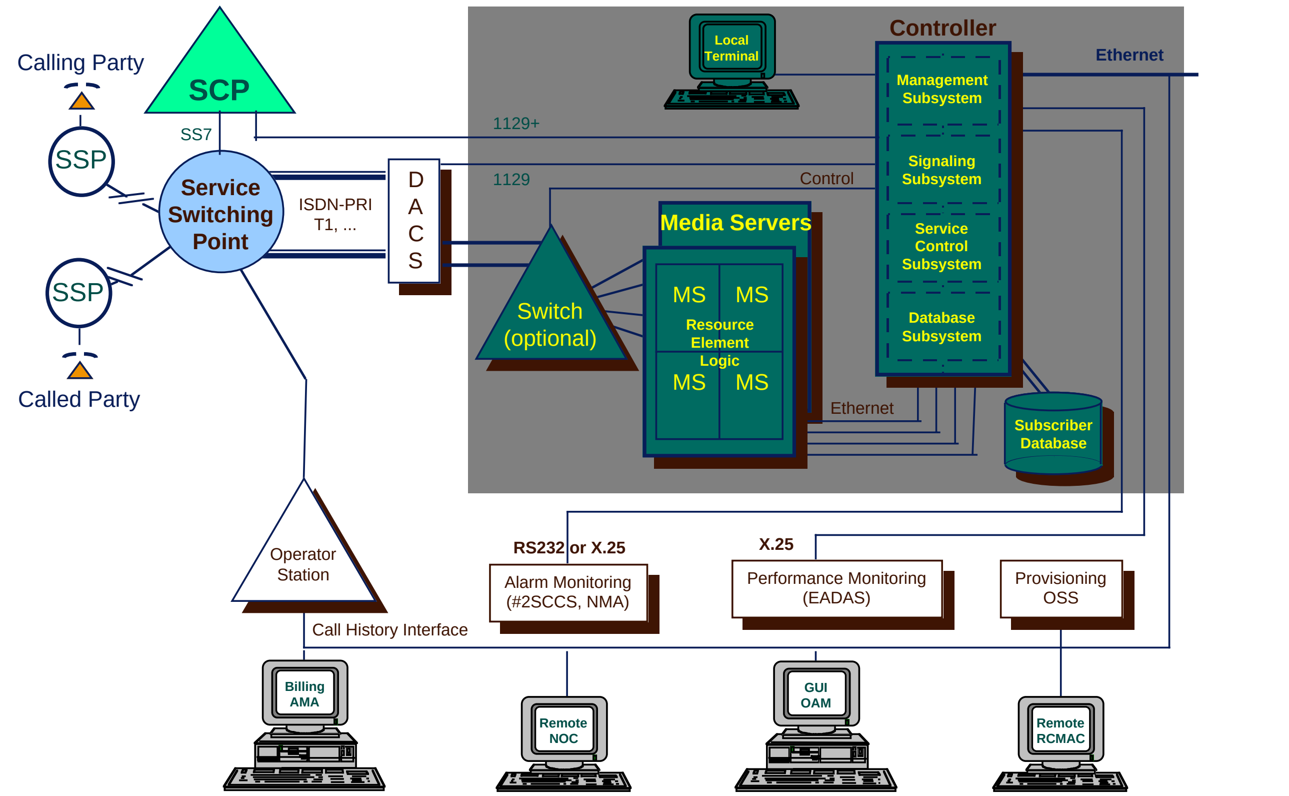


FIGURE 4

IN Operations, Provisioning, Service Management

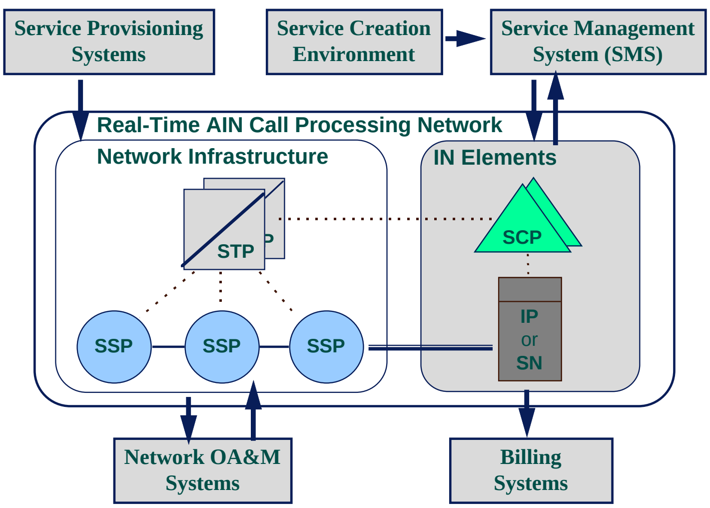


Table 3

IN Enhanced Services

Calling Services	Message Services	Network Control	Validation Services	Operator Services	Provisioning
Voice Activated Dialing (VAD)	Voice Mail	Custom CLASS	Credit Card Authorization	Automated Directory Assistance	Service Repair
Call Screening	Message Delivery	Voice Activated Network Commands (VANC)	900/976 Authorization	Intercept / Announcements	Subscriber Provisioning
One Number Service	Fax Mail	Personal Number Service	Calling Card Authorization	Personal Assistant	Order Taking
Caller ID Announcement	Fax Broadcast	Route with Voice	Long Distance authorization	Call Billing	Diagnostics
Interactive Voice Response	Never Busy Fax		IVR Gateway Authorization	Automated Operator Assistance	
Fax on Demand	Pager Notification			Customer Care and Assistance	
	Short Message Service				
	Universal Mailbox				

IN Capabilities

Important IN capabilities include voice announcements and messages, text-to-speech, facsimile transmission and reception, voice recognition, and e-mail messaging.

Voice Announcements and Messages

The capabilities of voice announcements and voice messages include the ability to play back a recorded voice, and dual tone multifrequency (DTMF) detection. IN voice services include operator intercept announcements, interactive voice response (IVR) services, voice paging, and voice mail.

Announcements, such as the common "the number you dialed is no longer in service," can be switch-based. SN–based or IP–based announcements, however, provide additional flexibility by letting the service providers use their own tools for recording and editing the voice messages.

DTMF detection is a particularly important service. It allows for prompt response and interaction with the caller. Basically, when a prompt is played, there will be a DTMF response, such as yes / no, menu selection, account number, etc. The ability should exist to interrupt prompts with DTMF, and there should also be a consistent user interface for DTMF usage (e.g., "*" means cancel, "#" means go ahead, "0" means help, etc.).

The term "voice messages" here refers to a voice mail system with the ability to record messages for later playback. It is also used for recording names, greeting messages, operator sessions, etc. This functionality should have playback controls, including play faster / slower, skip ahead / back, etc. Voice announcement and messaging capabilities are achieved through speech compression, the digital representation of voice for efficient transmission and storage.

Text-to-Speech

Text-to-speech refers to the conversion of text to spoken voice, or speech synthesis. There are four basic areas of importance in this conversion: The first, quality, is measured in terms of understandability and naturalness. The second, pronunciation, must take into account the expansion of abbreviations and recognize dates, time, currency, and message headers as well as proper name pronunciation. It must be sensitive to context and have a service-definable exception dictionary. Playback control, the third area of importance, includes DTMF interrupt, volume / speed / skip, and spell word. The fourth important capability involves being able to work in multiple languages.

IN text-to-speech services include an IVR system to get database information such as airline arrivals, a reverse directory with pronunciation of proper names, a spoken caller ID, directory assistance, and e-mail to voice. Other text-to-speech services include electronic messaging, financial information, message broadcasting, news and weather information, product information, and spoken short messaging.

Facsimile Transmission and Reception

Table 4 provides a list of facsimile capability features, data rates, and standards. Facsimiles make up a significant percentage of all phone calls (about 30 percent of all international calls). The T.30 protocol establishes a handshake between machines to negotiate data rate, page boundaries, error correction, and transmitting / called station identifiers.

TABLE 4

Facsimile Capabilities

Fax Features

- T.30 Protocol
- Error Correcting Mode (ECM)
- T.4 and T.6 Compression
- ASCII/Graphics Conversion
- Landscape/Portrait Rotation
- Graphic Inserts & Overlays
- Binary File Transfer
- OMR and OCR

Data Rates	Standard
300 bit/s	V.21
2400 & 4800 bit/s	V.27ter
9600 & 7200 bit/s	V.29
14.4 Kbit/s	V.17
28.8 Kbit/s	V.34

IN Fax Services

- Fax Broadcast
- Fax on Demand
- Never Busy Fax
- Fax Messaging

97-10.13

Fundamentals of IN 3/24/97 IN Capabilities Telecom Systems

T.4 and T.6 provide run-length encoding of multiple facsimile scan lines at a 20 percent to 70 percent compression with an average of 50 Kbytes per fax page. The fax capabilities also include ASCII conversion for cover pages and e-mail, graphics conversion for transmission of files from applications such as Microsoft Word and Microsoft PowerPoint as well as graphic overlays for letterheads, signatures, and advertisements.

In the future, facsimiles should have several additional capabilities. One of these, binary file transfer, will embed binary files in T.30 protocol. Other capabilities include optical mark recognition (OMR) and optical character recognition (OCR), which enables the translation of facsimile into e-mail format. The future will also see a facsimile/Internet relationship, whereby facsimile traffic could be rerouted to a data network for retransmission at the destination or retrieved via e-mail or a Web browser.

In this area again, IN provides enhanced services, including fax broadcast, fax-on-demand, never-busy fax, and fax messaging. The first of these provides for the scheduled delivery of the same facsimile to many machines. Fax-on-demand is an IVR system with data sent via facsimile (e.g., real estate property information). The third capability, never-busy fax, involves a busy or unanswered receiving facsimile machine that reroutes the call to a facsimile IP for later transmission to the receiving machine. This IP may be on either the receiving or transmitting side. Finally, fax messaging allows for the storage and forwarding of facsimiles very much like a voice mail service.

Voice Recognition

Voice recognition refers to machine recognition of spoken phrases. In this process, spoken phrases are compared to voice templates to indicate a match. Recognition can be either speaker-dependent, where each individual repeats their phrases several times to create templates, or speaker-independent, where predefined generic templates are used for all users. Voice recognition may be used for digit recognition, names and commands, and verification. Digit recognition includes spoken called numbers ("call 555-1234") and account number identification. The recognition of names and commands allows new enhanced services including voice-activated dialing ("call mom"), voice-activated network control ("forward all calls"), yes/no and menu selection, list/directory management, operator assistance requests, and information query ("what is the temperature in Dallas").

Verification is used for the purposes of authentication and security. Verification refers to the machine acceptance or rejection of a spoken phrase. The spoken phrase is matched to an individual user's voice template consisting of a password or digit string and requires an initial PIN or speaker-independent recognition for identification. Speaker verification systems are capable of accepting true speakers and rejecting impostors as a security mechanism for access to enhanced services.

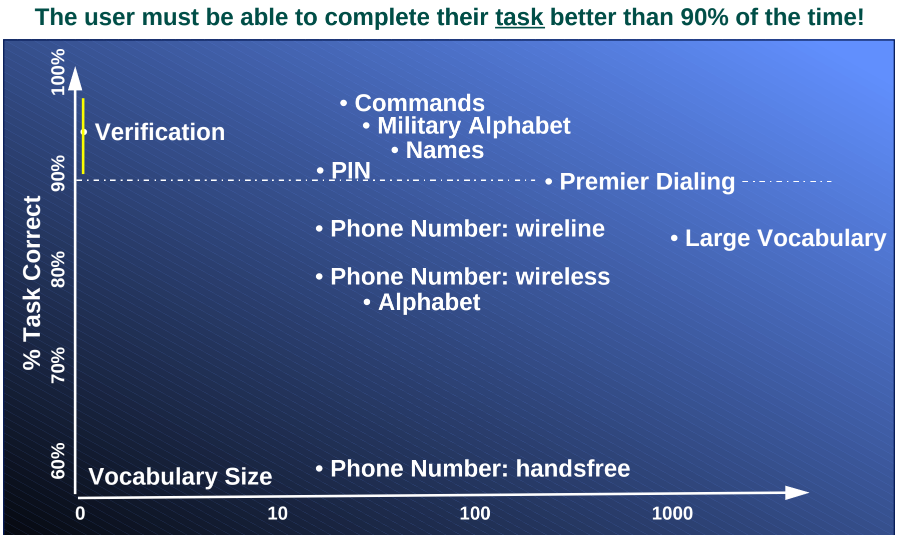
One of the newer capabilities of speech recognition is phonetic recognition. Phonetic recognition uses a predefined database of phonetic models to provide automatic-text-to-templates conversion for speaker-independent recognition. It can be used in both commands and proper names and allows rapid deployment of new vocabularies and services.

Voice recognition must be robust, which means being user-friendly. It must be continuous (allow naturally spoken phrases) versus discrete (having to pause between each word). A robust voice recognition system should include word spotting, which automatically rejects noise and words which are not in the vocabulary to be recognized; and voice barge-in, which is like DTMF prompt interrupt. Ideally the voice recognition capability should permit recognition of natural numbers and letters, such as 1-800, double 5, and alphanumeric characters. Finally, the recognition capabilities should be robust enough to handle handset independence, which means that the same vocabulary of templates is usable in the wireline and wireless networks across all handsets. Speaker-independent recognition must also work across multiple languages. The system must be able to recognize correctly regardless of age, gender, and dialect.

In order for a recognition system to be considered effective, the user must be able to complete a task more than 90 percent of the time—about the same rate as user errors in non-voice

FIGURE 5

Recognition Performance



recognition tasks. *Figure 5* plots approximate recognition performance across the recognition industry. These performance numbers are for continuous recognition—discrete performance numbers are higher. As can be seen in this figure, most numbers are approaching a satisfactory level: Verification is correct 90 percent to 99 percent of the time. Phrases like "call Grandma" are correct 95 percent of the time, while commands like "directory" and "add name" are correct more than 95 percent of the time. Premier dialing ("call business name") is correct 90 percent of the time, military alphabet recognition (alpha zulu) is correct more than 95 percent of the time, and PIN recognition is correct more than 90 percent of the time. If difficulties arise, it is often because of the use of a very large vocabulary, where correct completion is only at 85 percent. Correct completion with actual alphabet is at 80 percent. Recognition of a seven-digit to ten-digit phone number is correct about 85 percent of the time in the landline environment and is less for wireless use. The recognition industry's most challenging area is phone number recognition in the wireless, hands-free environment (with simultaneous road and car noise), where performance is currently only 60 percent correct.

It is the industry goal to have, eventually, transparent voice access of services. This should be available regardless of whether the unit is wireline or wireless. Correct completion of tasks on the first attempt should be at least 90 percent. There should also be secure access (using speaker verification) and the ability to recognize natural commands and information queries in multiple languages. Several possible AIN voice recognition enhanced services are shown in *Figure 6*. The ultimate goal is to have an effective human-machine dialogue that is as natural as talking to another person.

FIGURE 6

Voice Recognition Services

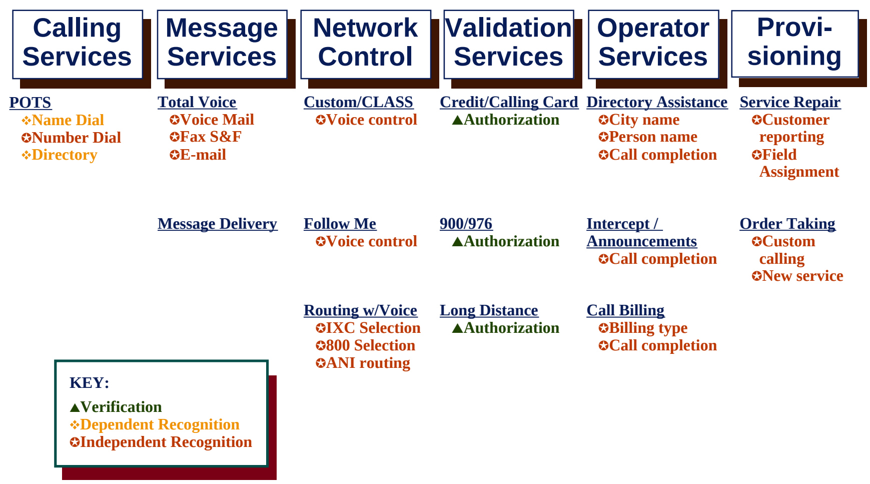
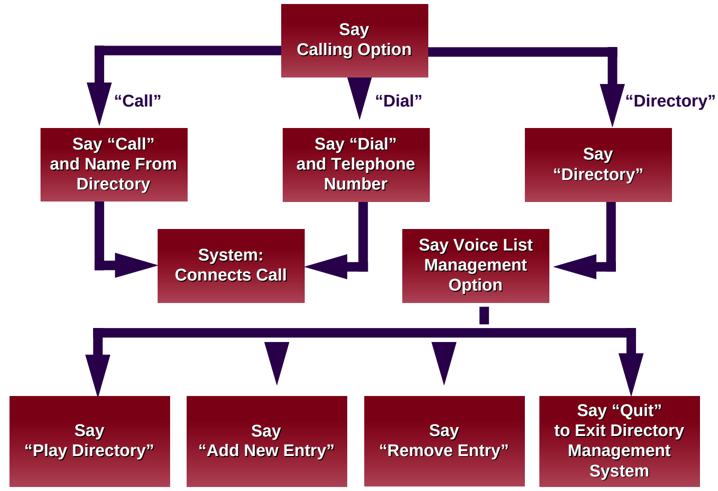


FIGURE 7

Voice-Activated Dialing Services



Today voice-activated dialing services are being deployed. VAD functions in an easy call flow, as shown in *Figure 7*. Voice-activated dialing requires not only voice recognition capabilities but also a good human factors design including a call/no call grammar ("call Grandma" and just "Grandma"), word spotting, out-of-vocabulary rejection, automatic enrollment coaching, name-too-long/short/similar announcements, voice-activated network commands, continuous spoken number dialing, voice list management commands with barge-in, and a good system for subscriber-utterance archiving for recording and troubleshooting the performance of the service.

IN Operations, Provisioning, and Service Management

IN functions, as discussed above, are made possible because of IPs. One of the greatest benefits of an IP platform is that as a single network element providing capabilities for a variety of new enhanced services, the IP provides a single administration and operations management interface to the service provider. These OAM&P functions have been organized into a telecommunications management network (TMN) model referred to as FCAPS, or fault management, configuration management, accounting management, performance management, and security. An example of IN operations, provisioning, and service management is given in *Figure 4*.

Fault management involves error and activity logging, interfaces to alarm subsystems, and diagnostics and fault isolation. Configuration management takes into account service-order processing, provisioning system resources, and conducting database management, including load balancing, archiving, backup, and disaster recovery. The service management system (SMS) is responsible for coordinating deployment of services and service data to the network. It is connected to the network infrastructure, IN elements, and all other operations support systems. The SCE is connected to the SMS for the development and maintenance of IN services. Accounting management deals with the billing record collection system. Performance management provides nonintrusive trace and line monitoring functions, performance analysis, statistics-collection logging and reporting, and traffic monitoring and control. The security area contains the system terminal screen interfaces and enables remote management.

Service Creation

One of the most integral factors of IN is service creation. Service creation distinguishes IN from central office (CO) or black box solutions by providing control to the service provider of the services offered. The building block approach is outlined in *Figure 8*. This approach allows for faster time-to-market, competitive differentiation and customization, economic efficiency, and vertical-to-horizontal integration. It also makes it possible to change a closed, single-vendor system into an open, multi-vendor, user-responsive network.

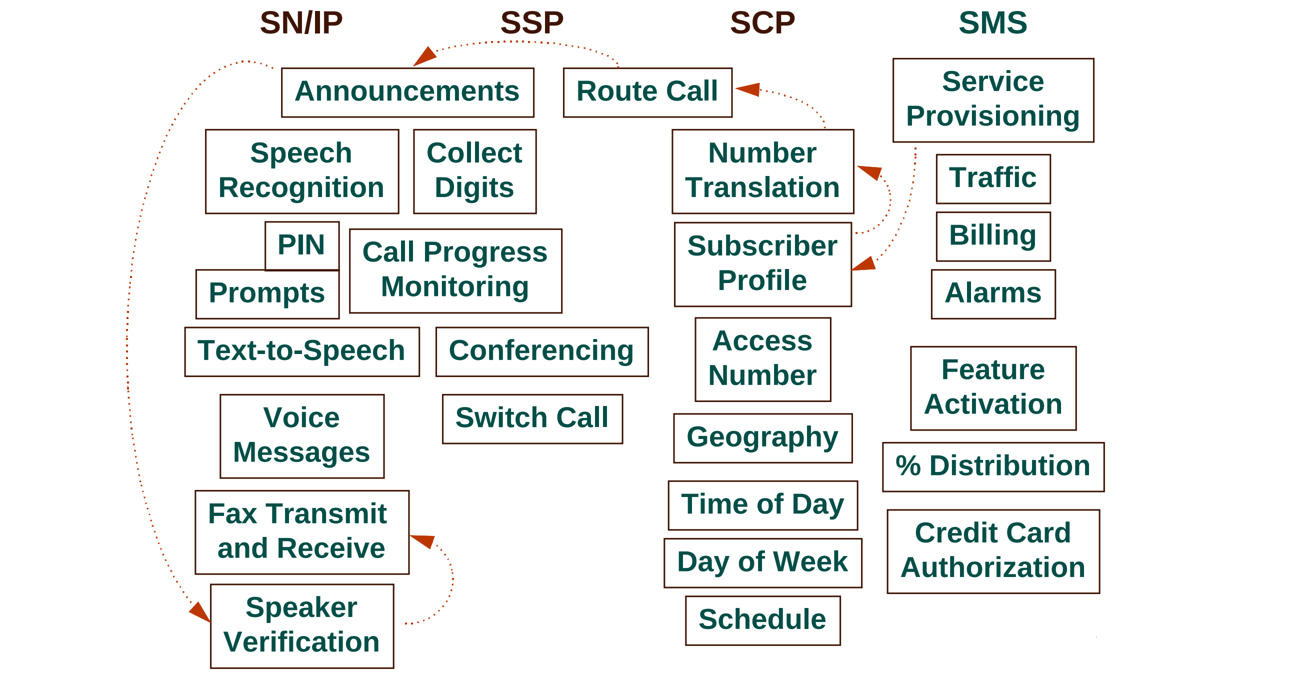
Many bodies have spent a great deal of time developing IN models to separate physical equipment, interfaces, and the logic of enhanced services. One such body is the International Telecommunications Union (ITU), whose four-plane IN model is shown in *Figure 9*. This model consists of the service plane, the global functional plane, the distributed functional plane, and the physical plane. The presence of the service plane allows for a solution to come from the user-level perspective. The global functional plane provides the solution in logically reusable parts, while the distributed functional plane focuses on physically reusable parts. The physical plane provides efficient, transparent physical distribution.

This model uses service-independent building blocks (SIBBs). Using this concept, it is possible to create service scripts by threading together precreated and tested capabilities. The SIBB concept uses object technology and allows for a flexible service logic format. Services are created by linking SIBBs together.

Figure 10 represents a similar service architecture that may be more familiar to those in the computer field. The software of an IP is organized into layers: The platform layer includes the base hardware, operating system, and middleware software.

FIGURE 8

Building Block Approach



The capabilities layer includes the media processing and resource and database management aspects of the IP. The applications layer provides the computer programmable layer of the equipment, including the reusable blocks, or SIBBs, that can be shared across services. Sequences of SIBBs implement applications. Finally, the administrative layer provides tools to create, modify, and maintain IN services. Service creation follows the evolution of computer systems in economic efficiency. Using this architectural model, there can be multiple vendors competing for each layer, which lowers costs.

Service creation involves everything that changes as a result of introducing new functionality. This includes service call flow, fulfillment (e.g., operator screens), traffic and billing records for the service, service assurance (e.g., alarming, trace logs), service activation, and business policies. It is important to remember that services can be created rapidly only when all components of the service can be created rapidly.

As a result, the SCE becomes very important. Available for use in the SCE are such tools as a graphical user interface (GUI); forms and call process parameters; database screen and report tools; flow chart specification languages; fourth-generation languages, UNIX, C/C++, and Java; voice prompt recording and editing; and recognition grammar definition utilities. If the service creation environment is used to its full potential, it should ensure network integrity as well as reduce

FIGURE 9

ITU Four Plane Intelligent Network Model

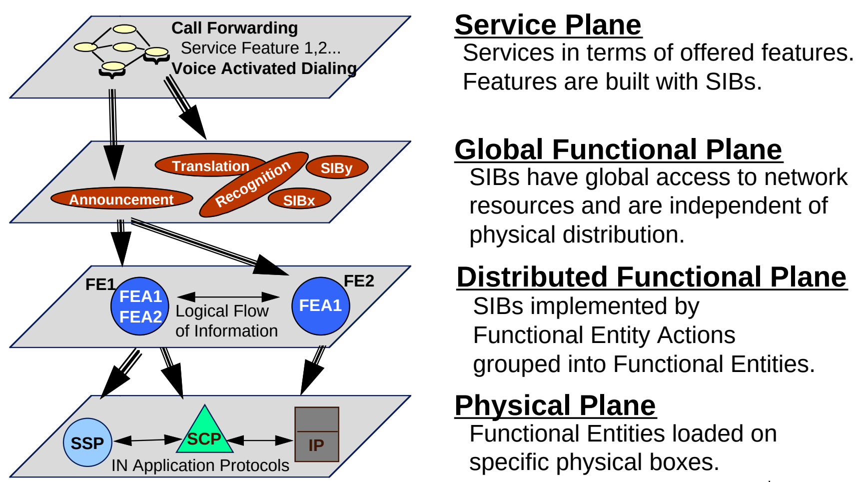
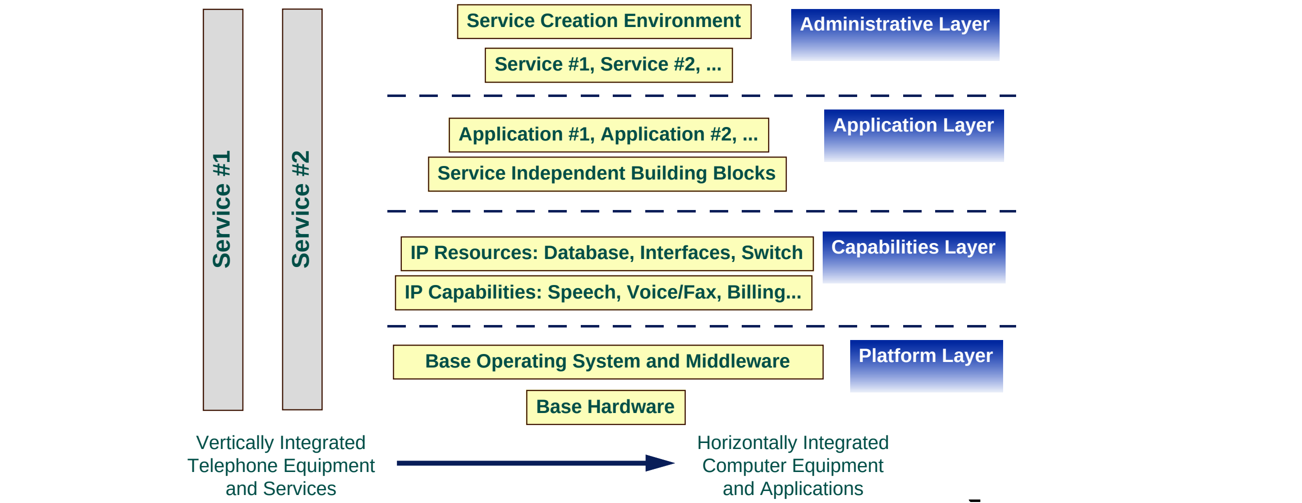


FIGURE 10

Service Creation Architecture



time-to-market by 70 percent and development and deployment costs by 50 percent.

Generally the following 10 basic steps are involved in service creation:

1. Service description (customer view)
2. Service specification (network design)
3. Service analysis
4. Service approval
5. Service development, environmental preparation, and operational support preparation
6. Test cases preparation
7. Off-line service testing
8. Service interaction testing (off-line)
9. On-line service deployment and testing
10. Service activation

IN empowers telecommunications companies and their customers by allowing them to create their own network services quickly, efficiently, and economically. Service creators may include network systems suppliers operating as systems integrators, telecommunications company staff in the form of service development teams, corporate telecommunications users who want company-specific services, third-party application developers, and even regular telephone users.

Significance of IN Capabilities

The advent of IN and AIN has made possible great leaps in telecommunications capabilities. New enhanced services based on new IN capabilities, architecture, and service creation are now possible. The underlying technologies for IN capabilities are already here and include announcements, voice messaging, text-to-speech, facsimile, and speech recognition. Speech recognition is particularly exciting because it is essential for providing intuitive user interfaces to these new services. Additional technologies that will facilitate IN services include e-mail, the Internet, and broadband data and video.

Network platform development and integration are currently in progress. While landline AIN has moved slowly toward deployment, the WIN has moved quickly, helping showcase AIN capabilities. Another major area of progress is computer telephony integration (CTI), which is driving open standards and lower equipment costs to provide competing services.

It is important to remember that IN should not be marketed just as a set of new capabilities. Rather, one must target specific markets, with specific enhanced services. Users must see the real value in using these enhanced services before usage—and revenue—increase substantially.

CTI and the Web-Enabled Call Center

Martin A. Prunty
Director of Consulting and Implementation
Alltel Call Center Solutions

Foundation technologies for combining the Internet and CTI are already in place. Call centers will never be the same.

For the past decade, computer-telephony integration (CTI) has been one of the hottest topics to hit the call center, promising reduced call volumes and handle times, as well as a higher level of customized service. However, in spite of this high interest, perhaps no more than 10% of all call centers have actually deployed CTI applications of any sort, and where such applications have been deployed, they are, in most cases, relatively simple screen-pops (see *Business Communications Review*, June 1997, "CTI beyond the Screen-Pop," pp. 57-60).

So we have been trapped in a vicious cycle: Vendors continue to aggressively introduce CTI technologies into the market, but call center managers remain reluctant to implement CTI until a "killer application" surfaces. While screen-pop offers many advantages, its benefits alone have not been enough to capture a large percentage of the market.

Today, however, that cycle may finally be broken. The global emphasis on electronic commerce and the use of the Internet as a delivery channel has sparked the development of new CTI applications that offer tremendous opportunities to call centers. The measurable advantages they offer are likely to surface as the "killer applications" that once and for all drive CTI into the mainstream of call centers.

Internet and Call Center Converging

Over the past two years, the Internet's explosive growth has awakened the corporate world to the potential for a new delivery channel for both electronic commerce and customer service. Many corporate Web sites have evolved from electronic versions of glossy brochures, offering little more than a profile of the corporation, into emulations of the call center that offers an alternative delivery channel from which transactions may be conducted in a self-service mode—without having to speak to a live agent.

Reprinted, with permission, from the February 1998 issue of Business Communications Review; copyright 1998 by BCR Enterprises; 950 York Rd., Hinsdale, IL 60521; 800/227-1234; www.bcr.com.

There have been some tremendous success stories already. Federal Express, for instance, recently indicated that it was conducting approximately 800,000 transactions per month on the Internet, a figure that the company claims is growing 10–12 percent per month. Since, in many cases, Internet transactions cost approximately one-tenth as much as live telephone transactions, the economic benefit of a functional, self-service Website is quite compelling.

However, it is also becoming clear that self-service will not address the needs of all Web users; there will always be users who prefer to speak with live agents. Indeed, there's a legitimate question as to whether it is in a company's best interest to fully automate its service delivery and risk losing the personal touch.

It is possible to draw a parallel between the evolution of functional Websites and the introduction of the voice response unit (VRU) several years ago. As call centers began to realize the benefits of the VRU, numerous self-service applications were deployed in a "standalone" mode—callers were unable to "bail out" and speak with a live agent. In the short history of self-service Web applications, this "bail-out" requirement is once again beginning to surface, creating a need to more fully integrate the Website with the call center, and CTI is the core technology to support this integration.

The Web-based "bail-out" transaction takes different forms, depending on the technology used and the specific requirements of the Web surfer. For the purpose of this discussion, three separate methods are addressed, along with their corresponding limitations and benefits: nonintegrated "call-me" transactions, integrated call-me transactions, and IP telephone transactions.

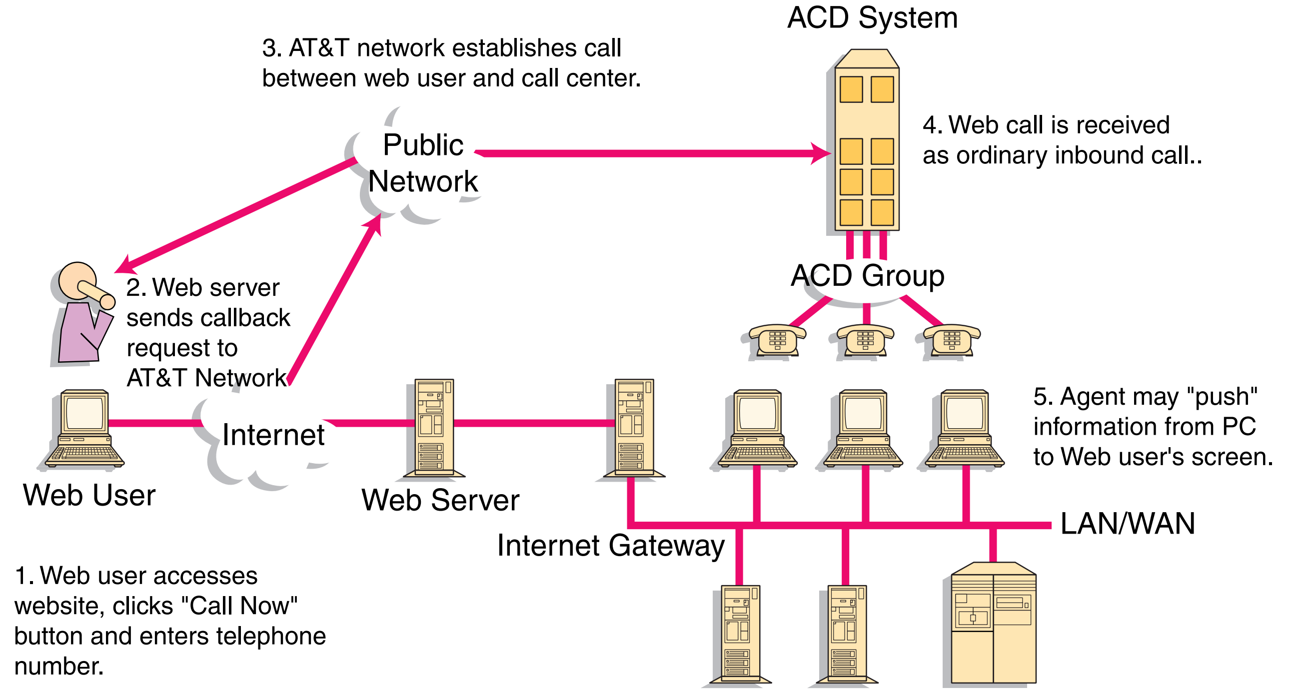
Nonintegrated Call-Me Transactions

The call-me transaction was the first form of Internet bailout to hit the call center. While this type of transaction takes several forms, AT&T's (www.att.com/callcenter) Project IA (Interactive Answering) has begun to take root in many corporate Websites. High-profile customers include Consumers Car Club, Personal Mortgage Corporation, and TSR, a third-party telemarketing, sales and reservations company.

With Project IA, Web users who wish to speak with live agents click on a "Call Now" button, then enter the telephone number where they can be reached. The AT&T network then sets up a call simultaneously between the Web user and the call center. The call center agent may then speak with the customer while the customer remains on line, connected to the

FIGURE 1

A Call-Me Transaction



corporate Website. Project IA also allows the call center agent to "push" information—such as graphics files or spreadsheets—to the customer's Web browser, further enhancing the communication.

Project IA is clearly a first-generation effort; it employs no CTI integration between the corporate Website and the call center (see *Figure 1*). Once the AT&T network has received information regarding the "Call Now" request, it sets up an inbound call to the call center, then connects the call to the Web user. This greatly simplifies the technology requirements necessary to establish a Web-initiated telephone call (see *Business Communications Review*, March 1997, "Distributing CTI across the Enterprise," pp. 63-66, and November 1997, "Call Routing Solutions for Virtual Call Centers," pp. 36-40).

Integrated Call-Me Transactions

There is another method of establishing a Web-initiated call-me transaction, which, from the customer's point of view, works in a similar fashion to the AT&T Project IA—users click on a "Call Now" button and then enter the telephone number where they can be reached. The difference lies in the means of establishing the call.

In an integrated environment, the Web user's telephone number is passed to the CTI server (sometimes a VRU platform) at the call center location, which in turn sets up a callback request through the call center automatic call distributor (ACD) system (see *Figure 2*). In some instances, the "Call Back" window offers the Web user a choice of callback times based on the current conditions of the call center. Once the call center agent is connected with the customer, the agent's screen receives the Web page used by the customer—this is called a "page pop"—ensuring that both agent and client are viewing the same information at the same time.

Myriad products now offer this CTI-based call-me transaction, including offerings from such companies as Edify (www.edify.com), Answersoft (www.answersoft.com), Versatility (www.versatility.com), and Spanlink (www. spanlink.com). The CTI software, often running on NT, Unix and other server platforms, also acts essentially as a VRU. Total up-front prices for such systems can range from about $350 to about $1,500 per agent seat.

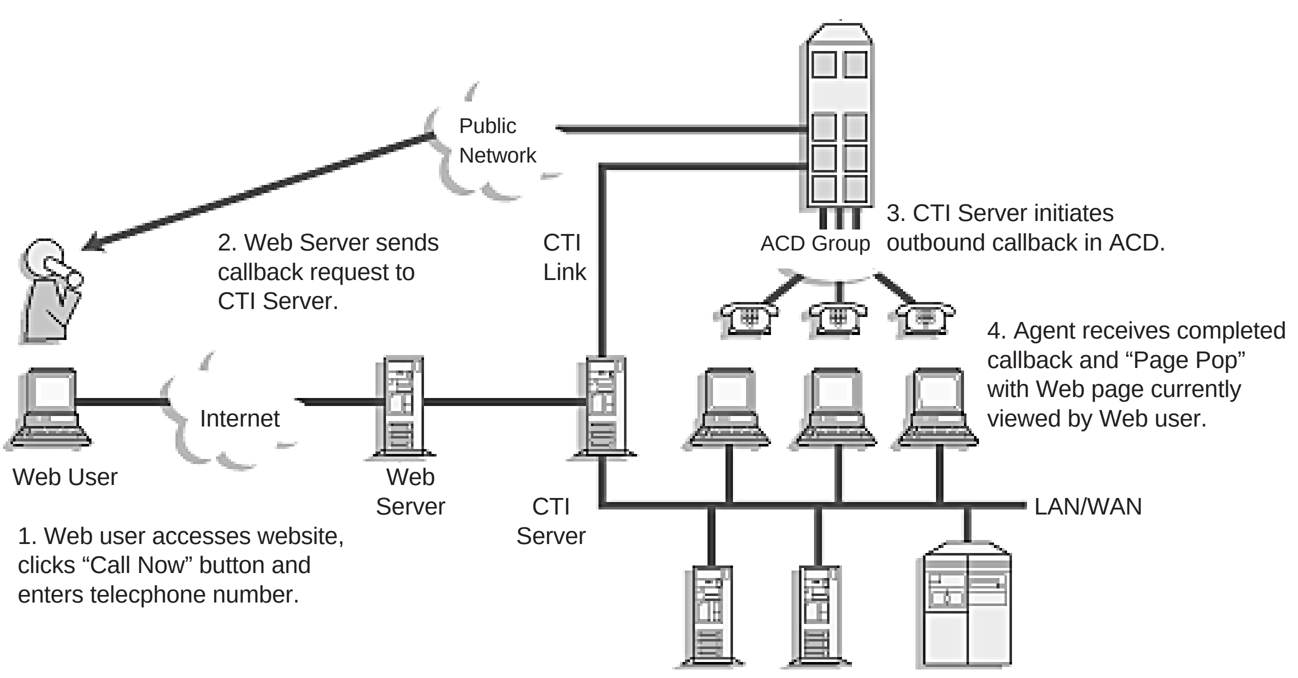
CTI-enabled integrated call-me applications offer some significant advantages over nonintegrated systems. AT&T's Project IA, for example, is billed on a cost-per-call basis, on top of 800-number charges. Furthermore, since 800 calls are received by an ACD as incoming calls in the nonintegrated system, there is no way to track their progress and agent activities, except by using a separate trunk group or dialed number identification service (DNIS) number.

By contrast, in the integrated call me transaction systems, there is only the cost of the outbound call, and tracking activities are far simpler. From a cost perspective, one key differentiator will be determined by calculating the unit cost per call of Project IA and comparing it with the capital outlay for the integrated application.

Integrated call-me systems are being provided by well-known PBX and ACD vendors, including Lucent Technologies (www.lucent.com/ callcenter), Nortel (www.nortel.com), Siemens (www.siemenscom.com), Aspect Telecommunications (www.aspect.com), Rockwell International (www.switch.rockwell), Mitel (www.mitel.com) Fujitsu (www.fujitsu.com/fbcs),

FIGURE 2

An Integrated Call-Me Transaction



and Intecom (www.intecom.com). GeoTel Communications (www.geotel.com) partners with other customer premises equipment (CPE) companies, such as those listed above, and is also active in comparable network-based call center CTI apps, partnering with the likes of AT&T, MCI (www.mci.com), and Sprint (www.sprint.com), plus overseas telecom carriers.

IP Telephone Transactions

Whether the call me transaction is integrated or not, its greatest limitation is the requirement that the Web user be equipped with a second phone line. Customers without a second line must drop off the Internet before receiving an agent's callback, which obviously can dilute the effectiveness of the transaction.

Some technical workarounds are being implemented to help in single-line situations. A product from call center software supplier Teloquent (www.teloquent.com) lets a Website user make an integrated services digital network (ISDN) or plain old telephone service (POTS) phone call to a live agent, dropping the Internet link for the duration of the phone call, then automatically reestablishing it when the voice call ends.

Recently, however, several companies have introduced technology that relies on Internet telephony to provide direct linkage to the call center. In this case, a user on a given Web page may click on a "Connect Now" button, but instead of entering a telephone number and waiting for a callback, the caller initiates an Internet protocol (IP) telephone call through the PC. The call is routed through the Internet to the call center on the same line used to connect to the Web. An IP telephony gateway then routes the IP phone call into the ACD system in a fashion similar to a normal telephone call. Once connected, the agent receives a page-pop similar to that in the prior example (see *Figure 3*).

An early leader in developing IP technology to integrate the Web and the call center is NetSpeak (www.netspeak.com), whose WebPhone software is being earmarked for call center-Internet gateway solutions from Rockwell International's Electronic Commerce Division (www.ecd.rockwell.com) and others. Rockwell's Internet gateway consists of agent WebPhones and server hardware / software at the call center. WebPhone software is also needed on the callers' PCs; NetSpeak's latest release, 3.1, retails at $20 to $50.

The Internet gateway server—costing from $40,000 to $50,000 for a 24-port system—is integrated into the ACD and is provided by either Rockwell or the customer. Rockwell says future Internet gateway server releases will empower agents to change callers' Web screens, jointly edit text and graphics, download data, software, and full-motion video to customers, compute collaboratively, and view the customer while conversing.

Lucent has a Web voice-enabled Internet call center capability that has been used for some time at its own Lucent Direct online catalog site (www.lucentdirect.com). In mid-1997, PC maker Micron Electronics (www.micronpc.com) became the first commercial customer to let potential buyers browse a Web page and talk to an agent over a single analog phone line.

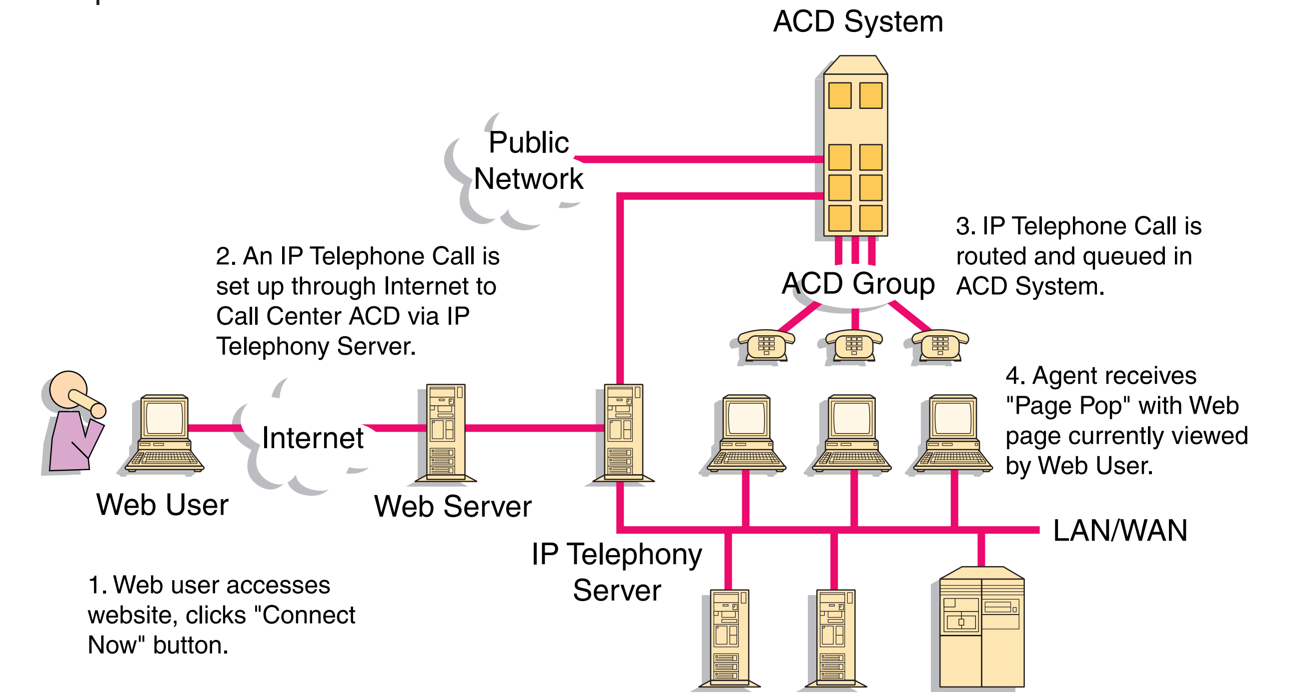
IP technology, however, still has some limitations, such as latency, for voice conversations, but ongoing research and development by IP telephone and gateway vendors should improve quality in the future. This would make such an approach to call center integration very promising.

Why Is Integration Important?

Integrating a Website with a call center operation offers users a self-service capability similar to that provided by most

FIGURE 3

An IP Telephone Transaction



VRUs, with the option of obtaining "live help" when necessary. Users can complete transactions in a single session, improving customer service levels and creating a viable alternative to the call center. In the short history of Web-based telephone calls, there is also evidence to suggest that the call resulting from the Web browser is significantly shorter than one that begins with the call center.

Newer strategies for integrating Websites with call centers offer even broader capabilities with greater benefits. For instance, if a Website has been designed to offer online purchasing, users may be reluctant to use credit cards on the Internet. Integration that supports simultaneous voice communication to the call center will allow a Web user to complete a high percentage of a transaction before it is necessary to speak with a live agent. Once the voice portion of the transaction is established, the user may finish the transaction by giving the credit card number over the phone, thus minimizing any concerns about security.

Consumers may see the voice link with a live agent as more secure than inputting a credit card number over the Internet. In the case of IP telephony, the entire voice communication is digitized, making it more difficult to intercept than in analog systems. Yet since solutions to such a challenge may be governed by perception, there is an inclination to favor the voice link.

In the near future, Web-call center integration will expand this concept well beyond on-line purchases. As more customers grow accustomed to using the Web as a customer service delivery channel, they will begin to access call centers directly from the Internet.

Soon, call centers will leverage this trend by encouraging customers to complete larger percentages of transactions on-line before linking to the call center. Given the projected growth of Web use in the future, channeling customers to the call center via the Web is likely to significantly reduce the expected handling time of a telephone call.

Conclusion

For the past 20 years, delivering customer service by telephone has remained largely unchanged. While voice response technology has improved call centers by reducing live agent calls, its use has been limited to relatively simple transactions. Longer and more complex transactions continue to require live agent assistance, and the demand for live agents has continued to increase.

The Internet clearly offers a more complete alternative by supporting a full range of transactions, almost regardless of their complexity. As its popularity continues to increase, its impact as a delivery channel will improve dramatically and may finally begin to stem the tide of demand for live agents.

To ensure that the needs of all users are met, Websites must be integrated with the call center, giving customers a full range of options without completely eliminating the valuable personal touch. Using CTI as a core technology, the Web will eventually evolve into another primary entry point into the call center, providing an advanced form of call setup far exceeding that of today's voice response technology.

CTI will continue to be one of the hottest call center technology subjects. The Web-enabled call center will be one of the "killer applications" that transform CTI from a talked-about technology to one that is actually being implemented.

Intelligent Peripherals and Service Nodes: Distributed Intelligence for Seamless Service

Susan E. Rudd
President
Acorn Communications

Today's telecommunications customers want access to services. They do not care if four carriers are involved or if a network is fragmented. A carrier that can conceal fragmentation, can interwork with other carriers, can deliver service to the end user, and can bill for service will be the victor in the marketplace. One provider could, in theory, offer everything, but new services will continue to be added, making it important to design services that are network and carrier independent.

One way to accomplish these goals is through the use of intelligent networks (IN). This paper will discuss traditional IN platforms and the capabilities that they promise, as well as the need to extend the IN model. IN with a single carrier model does not meet the requirements of the 1990s or the next century. A model is needed that takes IN architecture and extends it into a multi-carrier, multi-service world where service mobility is as easy as geographic mobility on a wireless network. This paper will examine the paradigms for solving this problem, while also discussing the requirements for platform and system architecture. Finally, it will explore alternatives to IN.

Traditional Platforms

To have a choice of competitive vendors, companies need to become hardware-independent. From the service and application development points of view, hardware independence requires open platforms. Historically, IN platforms have been closed, monolithic, and proprietary. In addition, the network elements that have actually been built often differ from the standards. As a result, IN solutions are more expensive, ranging from $2-to-$20 million. All of this leads to one important aspect—high cost of new IN platforms and services—and one important question: Why would a carrier move services off the switch and take the risk of multi-vendor integration only to be locked into another proprietary solution at the same or at a higher price? The carrier would be better served by a proprietary switch vendor who takes responsibility for the complete solution. Closed platforms have therefore led to a continuing preference for the non-IN switch-based model.

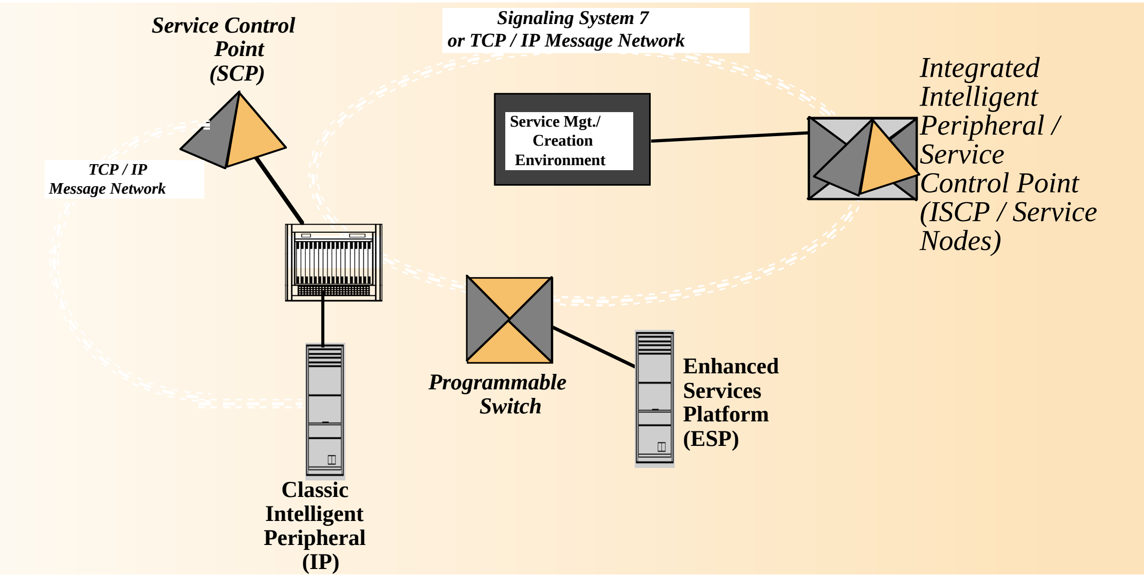
Figure 1 displays deployed platforms. There have been many platforms that do not fit the model. The most typical platform, from the end-user applications point of view, is the enhanced services platform (ESP). This is not an intelligent peripheral (IP). While many vendors place an SS7 protocol stack on the ESP and call it an IP, there is no service logic on it. If, however, the ESP sits behind a programmable switch, it can offer a very good solution for public network routing, etc. For this reason, many traditional voice-messaging vendors are moving resources onto such a platform.

Another example is the classic IP, which is designed as an element linked to the network through only a service switching point (SSP). The low performance this produces creates the "service node dilemma," which forces a trade-off between the correct IN architecture and application response time. There is a separate connectivity—1129+ or another interface—that allows direct communication with the service control point (SCP) database. If there is an IP application that requires high performance, there are good reasons to avoid routing through a telephony switch not designed for message traffic in order to obtain information from the SCP database. A backdoor channel to the SCP using 1129+ messaging over a transmission control protocol / internet protocol (TCP / IP) link, allows requests and responses to be sent rapidly between the IP and SCP. Alternatively, the subscriber service database can be integrated on the application platform. This configuration is called a "service node," or intelligent peripheral service control point (ISCP).

There are problems with either solution in a multi-vendor, multi-application environment, since the SCP database is still potentially a centralized bottleneck resource. In the 1129+ case, it is possible to have multiple vendors contending for a single common SCP. In the service node case, the database is "nailed" to the applications platform just as it used to be "nailed" to the switch. In both cases, there are potentially severe limitations on the information access. For IN to work well in a multi-vendor environment, the service profile databases need to be distributable, scalable resources. Appropriate subsets of information are available locally to the applications that need them, but are synchronized rapidly with the master profile.

FIGURE 1

De-Facto Standards That Have Emerged in IN



Promise of IN

IN potentially promises significant new services that operate independently over any carrier's particular transport mechanism. IN services should therefore operate seamlessly across wireless and wireline networks as well as interoperate between public and private networks. IN allows carriers to purchase these services from multiple third parties as well as from the switch vendors, so producing a much faster time-to-market for new services. IN theoretically allows for the mixing and matching of software from these different application vendors on the same platform, as well as making it possible to upgrade over time without a "forklift truck." If this promise is kept, IN should lead to lower costs for multiple new services from a plethora of new vendors.

Figure 2 provides one example of the interoperability potentially afforded by IN. Global systems for mobile (GSM) is chosen as the wireless system in this example because from the very beginning, the parties involved defined GSM wireless interoperability with wireline networks based on integrated services digital network (ISDN) standards. GSM was launched with specified service interoperability. As additional services are provided to a GSM wireless network, signaling and service logic can be linked to an intelligent server peripheral attached to the mobile switch. With the right signaling connections, such a peripheral could be attached to any switched telephone network anywhere.

With this architecture, wireless access to wireline service can be readily provided. The same services that are available on a wireline telephone can be available on wireless telephones. This is not an amazing concept. Rather, it is what users think they are buying with a personal communication service (PCS) phone. How is it possible to use the ISDN services that are on the other switch? In the GSM model, there can be a peripheral or a service node to link the two systems. The service profile for each user is linked and synchronized between the home location register (HLR), the service control point (SCP), and potentially multiple other HLRs and SCPs for roaming or number portability. Extensions such as these will allow IN architecture to be extended to a multi-carrier world.

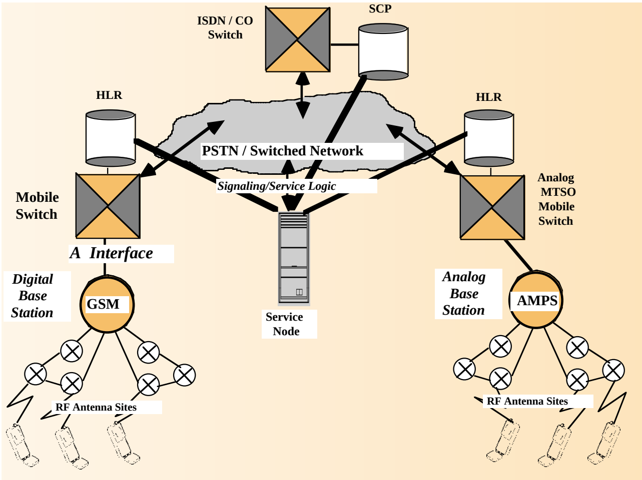
The distributed services model that will ultimately succeed is probably the GSM model. Providing interoperability without making one carrier's proprietary information (e.g., user's billing information) available to another carrier requires careful design. In the mobile world, this issue has been resolved by allowing one carrier to receive a subset of information about a subscriber—not enough to bill or "steal" the subscriber, but enough so that the application could run with high performance and create a record to bill back to the originating carrier. In mobile networks, this is called the visitor's location register (VLR), and is used primarily for roaming between service areas. This VLR approach can be extended to support roaming between services or between carriers.

Figure 3 depicts the ultimate network—the network that may be prevalent in the future. This example shows all narrowband services. At customer sites A and B on the left, there are private branch exchanges (PBX's). At site A there is also an associated telephony server. Computer telephony servers—that directly control calls on the PBX with standard message sets—are becoming widely available. The computer supported telephony application (CSTA) message set has already been developed by the European Computer Manufacturer's Association (ECMA) and virtually all major PBX vendors now offer open connectivity to the CT-Connect subset of CSTA available from Dialogic Corporation.

Site C has other local area network (LAN)-based services—such as the telephone application programming interface (TAPI) and telephony server applications programming interface (TSAPI)—that will eventually converge with true CSTA message capability. The rest of *Figure 3* shows:

FIGURE 2

GSM/Analog & ISDN Services



- traditional central-office switch, which has been provisioned with an SCP database
- cellular network with a mobile switching center (MSC) and its own HLR, linked to a programmable switching hub with a metro area HLR
- network servers—a "server farm"—attached to the hub and accessible from any network or carrier

If there is Web farming on the Internet, why not value-added service farming on the telephony network? (These servers are only logically centralized—they do not need to be physically centralized.) The important concept is that services are accessible to the server resource from any other network.

There are many questions associated with this architecture. Why is this a good idea? One reason is that an experimental service can be quickly set up and managed at a single physical geographic location and service offered to multiple remote cities. As a result, instead of performing service trials at each central office location, it is possible to run multiple trials at different locations from a single remote server. To make this work well, it is important to have the right signaling, a high-performance backbone, and a high-performance programmable switch matrix. As traffic grows for the live service, additional servers are distributed to support only the required level of demand in each service area. Capacity can scale linearly.

What is difficult about this approach? Assuming that the servers and the message data conform to appropriate telephony and networking standards, the key challenges are

- voice traffic control
- call control handling and
- retrieval and update of a user's service profile information

In *Figure 3*, the small VLR "pyramids" represent server- and application-specific databases. Use of these application-spe-

FIGURE 3

Seamless Campus and Metro Area Services

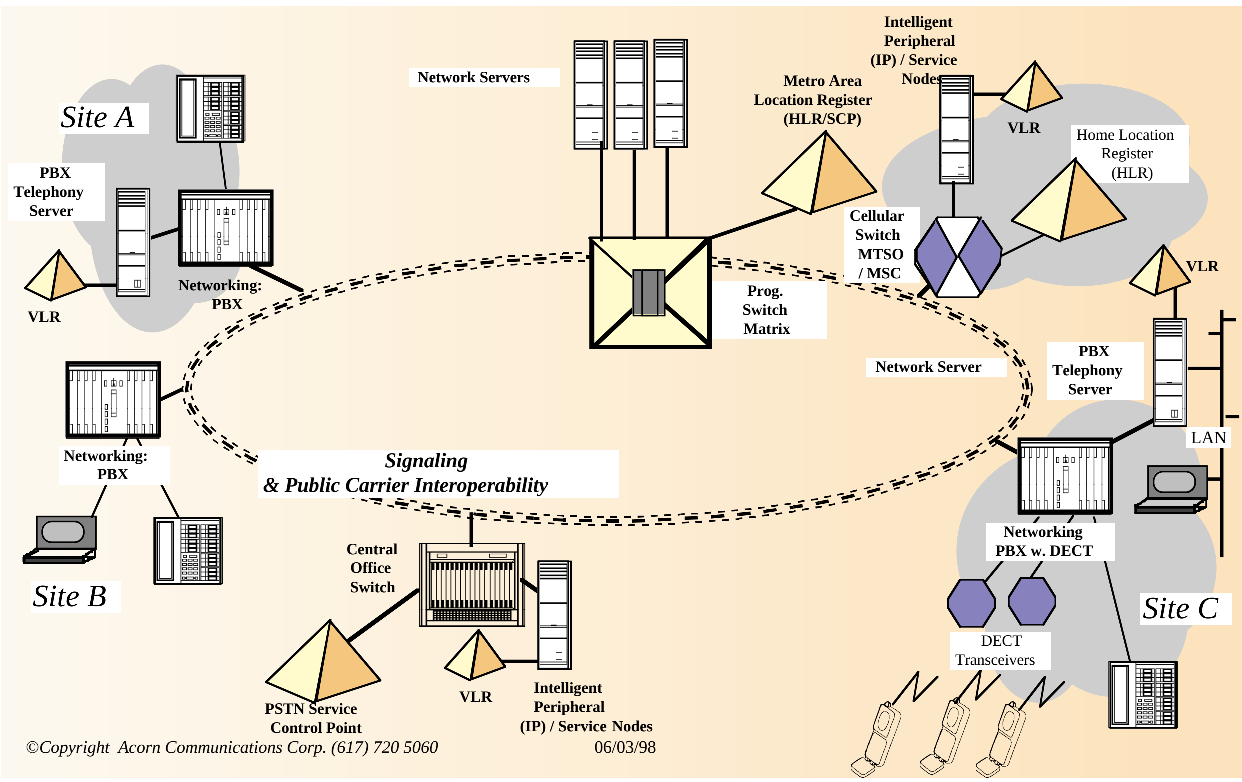
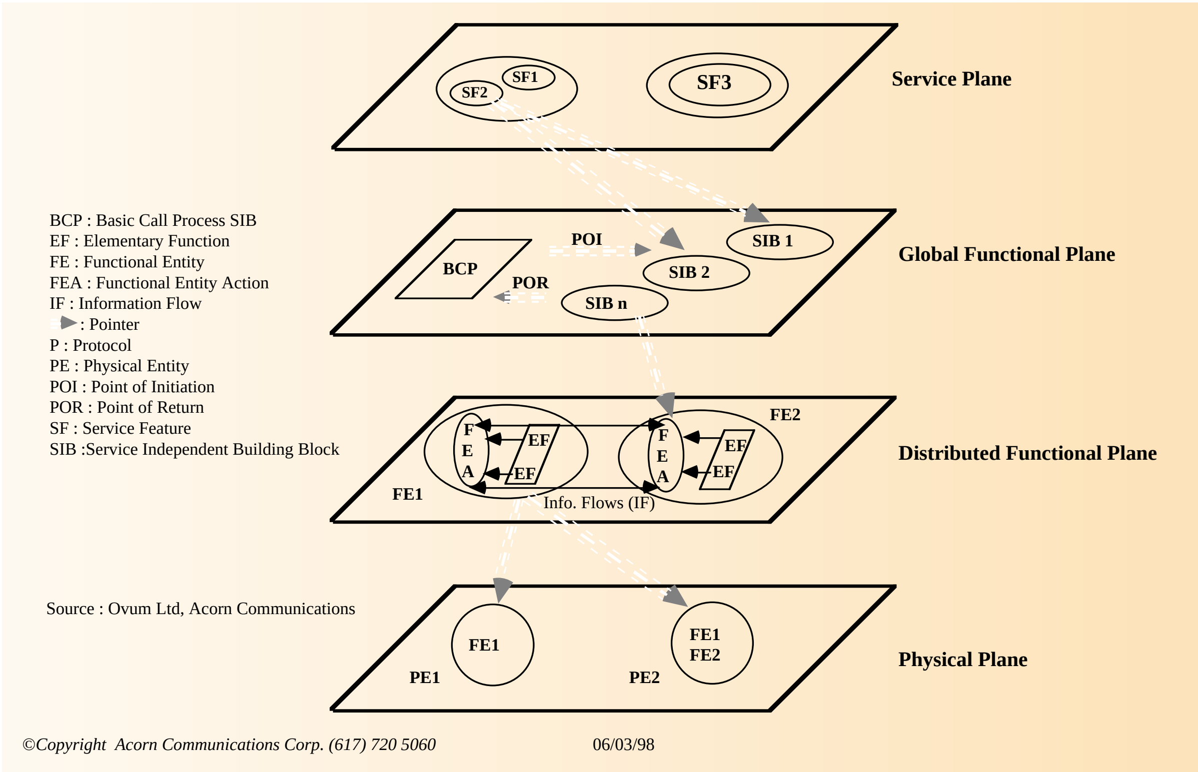


FIGURE 4

CCIT/IUTU-T IN Conceptual Model Planning



cific VLR's resolves the "service node dilemma." A subset of user information that is specific to each application or service exists at any of the different service nodes. Any of these applications or services may be offered by a different carrier, but the information in the respective databases is synchronized with the appropriate HLR or SCP for that subscriber's "master" record, using standard database interchange mechanisms and subscriber-authentication techniques.

Similar concepts are being developed for service portability and service mobility in the European Universal Portable Telephony Standards (UPTS). Some of these standards may eventually be adopted into the U.S. environment. Nonetheless, even with today's standards, it is possible to make application server databases and the carriers' master databases interoperable. This enables one subscriber to be provisioned by an HLR, another by an SCP, and yet another with a service node. With interoperability, each database request and update can be routed to the appropriate resource and the proper billing transactions can be created, even for a multi-leg, multi-service call across several carriers!

Figure 4 provides a conceptual model of IN based on the International Telecommunications Union (ITU) model, the worldwide standards model for IN. Service switching points (SSPs), signal transfer points (STPs), and other lower-layer elements that comprise network connectivity are not shown in this diagram.

IN, as shown here, is an information service architecture for any application. Application features are visible to the user from the service plane, which typically has an application programming interface (API) to the next layer via service independent building blocks (SIBs). At the SIB level—the global functional plane—there are telephony call-control commands based on telephony service logic and call flows. The distributed functional plane contains the underlying software elements for the service functionality and the database, and media resource modules. These, in turn, are mapped onto functional entities in the physical plane.

People with an information-systems background immediately recognize this as simply a services architecture that happens to apply to telephony services. They typically use different names for these layers, however.

In information systems terms, the top layer becomes the applications layer and has an API (or message set like CSTA) linking to the system logic environment, which has access to software modules for application functions, databases, and media resources at the third layer. These have access to different physical elements such as network elements and device drivers at the physical layer. (The Enterprise Computer Telephony Forum [ECTF] uses terms that are closer to these information systems terms.) Thinking of the IN services model in terms of information systems makes it easier to define voice services as a subset of multimedia-environment services with special class and quality of service parameters (e.g., low delay and moderately high throughput).

IN is a powerful generic service architecture for telephony that can and should be used to define complete services independent of the underlying network elements.

Requirements for Platform and System Architecture

The following are required to build platforms and systems with this architecture:

1. Multiple service profile databases. There are now computer standards for interoperability of standard query language (SQL) databases, many of which are now fast enough to be carrier-class service profile databases. Informix, Oracle, Sybase, and Tandem / Guardian have standard interfaces that allow them to interoperate in a high-performance telephony environment.

2. Multiple vendor's platform implementations. There are now IP platform vendors offering open, hardware-independent platforms that allow application developers to port software rapidly between their platforms. Digital Equipment Corporation (DEC) and Hewlett-Packard (HP) are two companies offering open platforms.

3. Multiple Applications. The creation of multiple diverse applications that can coexist, interoperate with one another, and share media and other platform resources is also critical to the success of IN. A number of independent suppliers are beginning to comply with common object request broker architecture (CORBA). It is likely that in four to five years it will be possible to exchange requests for telephony functionality (or telephony objects) between applications.

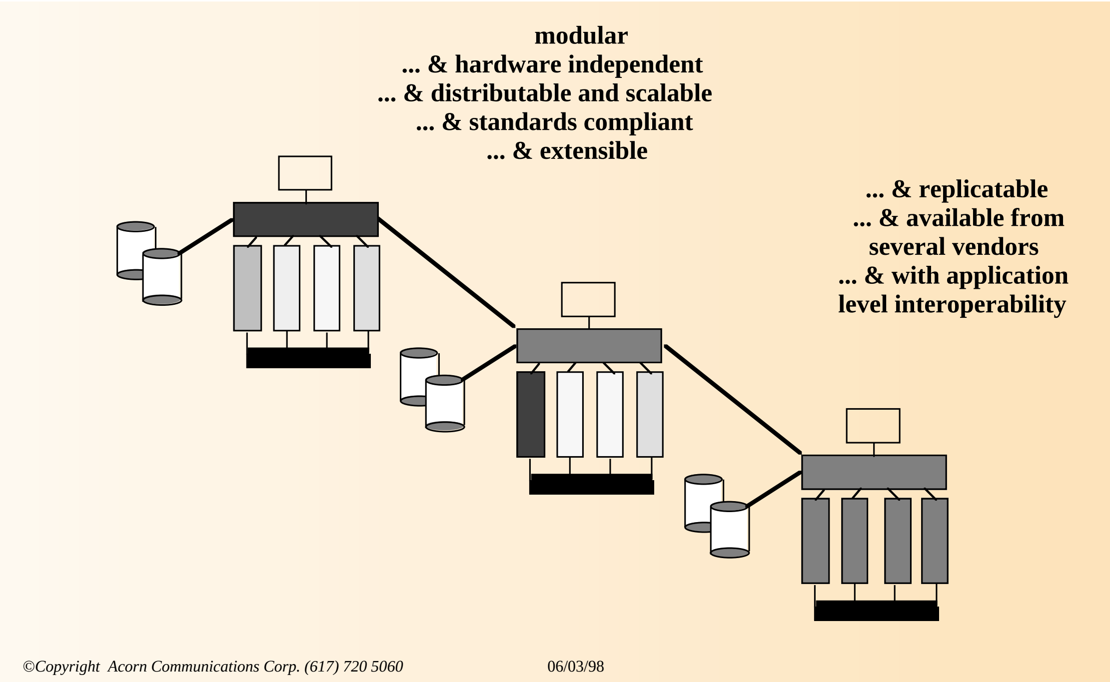
4. Multiple Media Resources. There are a variety of media resources and voice-processing boards that can be integrated into any vendor's computer platform. These include multiple digital signal processing (DSP) boards and interactive voice response (IVR), automatic speech recognition (ASR), and text-to-speech (TTS). Standards for common interfaces and interoperability of media and physical resources are being generated by the Enterprise Computer Telephony Forum (ECTF). Many of these standards (e.g., S.100—the standard that allows a single application server to have different media resources that are not co-located) will become de facto IN standards. Other standards from ECTF, including the telephony call control message set, will be forthcoming in the next year.

Figure 5 shows typical open-server platforms. For each server platform, there are likely to be one or more application-specific service profile database(s) and a set of call control functions—shown as linked in the diagram. There will be multiple application software elements—logical clients and servers—on any platform (four are shown for each physical server as an example). These access a variety of physical media resources shown at the bottom of each server. As mentioned above, the S.100 standard provides an open interface between media resources that allows different application vendors and physical servers to share resource capability across platforms, rather than the separate replicated resources as shown here.

ECTF and its private network constituency is racing ahead of IN in setting standards for all types of service interoperability for these platforms. Fortunately, some CTI designers are now looking at interoperability with SS7 and IN services architecture to link these two worlds.

FIGURE 5

Server Platform Criteria



Threatening Alternatives to IN

Because of the delay in implementing IN in today's narrowband public switched telephone network (PSTN), alternatives for providing new telephony services are now coming from the private network market and computer telephony integration (CTI).

1. New services will come from CTI, not IN. New services have, in the past, almost always originated as private-customer solutions and moved into public network services (e.g., voice messaging). One reason may be that carriers have typically focused on network deployment and quality of service rather than on the creation of new customer capabilities and value-added services. More importantly, innovation is associated with large numbers of creators experimenting simultaneously with different capabilities. Only the fittest applications survive, but many new ones are tested quickly. Historically, this experimentation has been far easier in the private-network environment. Microsoft and Novell almost instantly attracted large numbers of third-party developers to develop computer telephony applications by promoting open APIs on existing PCs and servers. IN needs to leverage the creative and innovative services of this burgeoning CTI world and offer open platforms to link these new capabilities into servers in the public network, at the same time making it easier for multiple competing third parties to develop applications for open-carrier platforms.

2. Volume is driving down the cost of CTI platforms. Already, low-cost voice-processing hardware and commercial databases are capable of supporting telephony call processing. Large volume from computer telephone integration (CTI) customers supports lower cost platform pricing. Both these factors—low cost and high customer volume—are driving the success of CTI technology.

3. Virtual public networks (VPN) represent a major new opportunity for service providers, but broadband / asynchronous transfer mode (ATM) networks or Internet service providers may be the winners. Traditionally, large companies have outsourced their transport requirements for on-demand bandwidth and virtual private networks of switched services to the telecommunications carriers, because it was more cost effective to outsource the specialized management of communications transport than to do it in-house.

There are many comparable reasons for creating a virtual public network. A virtual public network allows large companies to outsource services that primarily involve interactions with parties outside the corporation. The corporation can contract with a carrier for the operation of shared value-added telephony services for the benefit of their customers and suppliers and save significant amounts of money by using the switched network to concentrate traffic to secure telephony servers in the public network. For instance, many financial services for the consumer or small business market can be more cost-effectively delivered as a shared public network service (e.g., pay-by-phone for home banking).

VPN networks for shared access to corporate services require the following:

- distributed as well as centralized platforms
- parallel as well as conventional processors
- distributed as well as centralized databases
- openness and interoperability at multiple layers of the model
- extensibility and replicability of service elements
- performance in tens rather than thousands of milliseconds per event

Today, however, ATM networks and Internet approaches seem to be stealing a march on narrowband IN in delivering many of these capabilities. Throughput and delay remain issues, however.

Nonetheless, platforms are already changing to support VPNs with voice over IP or ATM. Focus is on applications and services rather than on network infrastructure platforms. The cost of the platforms is based on vendor-independent hardware shared by thousands of users. Platforms are becoming configurations of client / server functional elements instead of monolithic black boxes. Administrative functions are being standardized for easy integration with operations support systems, including service management, provisioning, profile update, network management, and bill-back applications. All that is missing is the following:

- enhanced throughput,
- manageable packet delay
- support for non-switch based network and PBX telephony call control over IP or ATM

An IN Response for Integrated Broadband/IN Networking

The larger interexchange carriers want to provide VPN applications with shared switched-network access to corporate databases. Their high-speed backbones—whether they are using broadband / ATM or sending IN narrowband messages over a broadband transport—allow for "server farms" to be located anywhere in the public network to offer remote functionality, as well as authenticated access through a corporate gateway to a corporate environment. There is no reason that a transport network of any type cannot be designed to use the IN services architecture and standards to hook carrier-based, network-class, and high-performance systems to corporate and eventually residential environments. One such network is shown in *Figure 6*.

Over time, narrowband and broadband networks will be linked. In *Figure 6*, the central office switches are attached to a high-speed backbone as specialty servers for the broadband network (e.g., for CLASS services), while continuing to be system controllers for their own narrowband copper networks. Conversely, a remote "server farm" on the broadband backbone can supply narrowband IN services to users in the narrowband local loop. This is a view of an ultimate and somewhat extreme scenario, but there is no reason why application servers for narrowband services cannot be located on broadband backbones provided that they support appropriate signaling and service functionality. From the backbone, applications can be routed to any of the narrowband switches, whether central offices or PBXs with minimal switch integration. The switch integration problems that are stalling IN deployment today are obviated.

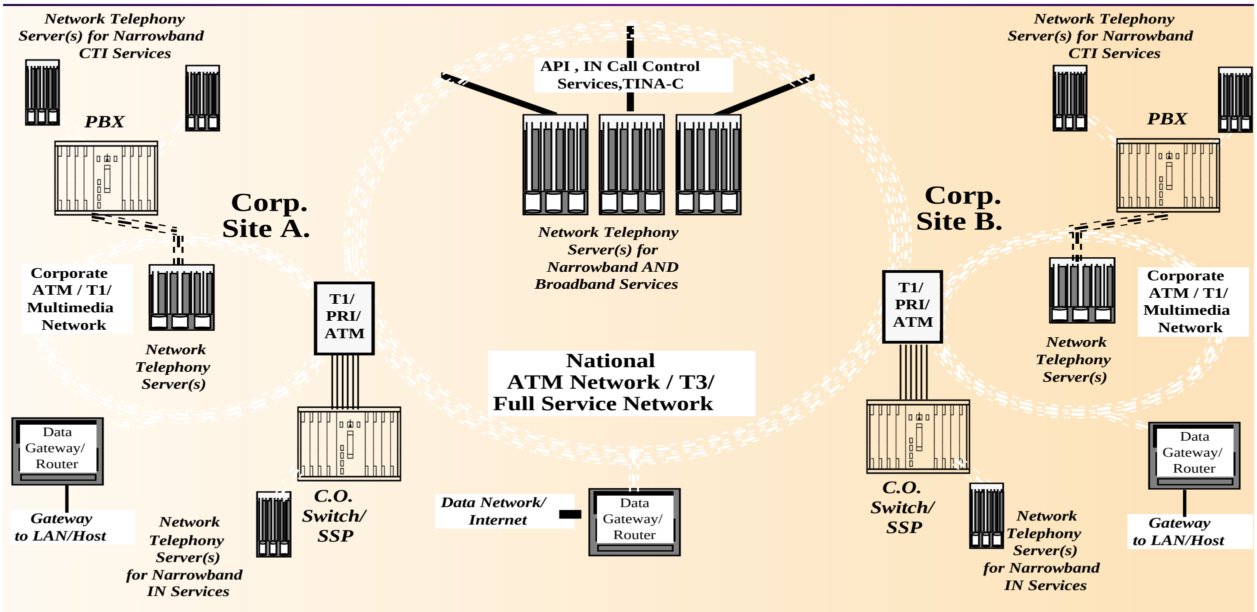
With an architecture like this, it is possible to start to launch at least some IN-like services, regardless of whether a particular switch has the right software upgrade or whether release AIN 0.2 (with the critical send-to-resource [STR] command) is available. This makes it possible to immediately promote new VPN service applications that link IN services to premise-based CTI services without disturbing the incumbent switch vendors.

Conclusion

While IN looks to be the best possible theoretical solution for new services, it is falling behind CTI and broadband services in development. By examining the architecture already in place from an applications point of view, not just a physical point of view, an application creator can today put together high-speed service access that uses a multimedia backbone to supply narrowband IN-like services. This system leverages the IN services architecture to achieve interoperability between public and private networks, wireless and wireline networks, and voice, data, and multimedia networks. Such an approach will stimulate tremendous service innovation from multiple third-party applications developers.

FIGURE 6

The Threat to IN: Servers on Broadband and Corporate Networks and a Few on Narrowband



The Cable Modem's Evolution and Revolution: A Historical View

Rouzbeh Yassini

Founder and Chairman, LANcity; *Inventor*, cable modem

Introduction

The cable modem is profoundly and positively changing society. This technology is enabling the cable communications industry to lead society's transition from its 20th century emphasis on paper communications to a 21st century emphasis on electronic media storage and transport. But what is a cable modem, and where did the technology originate? How do cable modems deliver high-speed data over cable services? How did the cable television infrastructure develop, and why is it so well suited for broadband data services? This paper examines the networking evolution that has positioned cable modem technology at the leading edge of the digital convergence revolution.

Birth of the Cable Television Industry

In 1948 in Astoria, Oregon, the cable television industry was born primarily as a means to provide improved television reception to viewers in remote areas. Regulation of this industry by the FCC began in 1960. In 1972, the first pay-per-view service, Home Box Office (HBO), was launched. Since that time, cable television has become a common source of video entertainment for more than 500 million viewers worldwide and is now found in approximately two thirds of American homes. While cable television became well known as a means for distributing high-quality one-way video services into homes, its marriage to television, a passive receiver of video broadcasts, obscured cable television's other ability to provide high-speed two-way data communications.

I-NETs

In the late 1980s, however, local and state governments, universities, and research institutions began employing the cable television infrastructure for two-way campus data networking. These institutional networks, or I-NETs, served the specific needs of the institutions they supported. Typically, I-Nets were used to provide connectivity between a campus's local area networks (LANs). The demand for local information on private institutional networks did not provide the impetus for cable modems as the next-generation technology, though. Rather, it was the demand for access to global information on the World Wide Web that drove the commercial success of the cable modem. Still, I-NETs did serve as the early proving ground for cable modem technology.

LAN Technology

The history of the cable modem is closely linked to the history of internetworking technologies and the explosive growth of the global Internet. By the late 1970s and early '80s, the networking industry had begun developing the LAN. Ethernet, which is the predominant LAN technology in use today, was first developed in 1973 by the Xerox Corporation. LAN technology, in fact, predated personal computer and engineering workstation development. Originally, LANs were used as an efficient means to link terminals to mainframes and minicomputers.

The LAN industry grew to its second phase with the explosive use of PCs in the commercial market place, along with the rapid adoption of engineering workstations for scientific and engineering computing. The need for PCs to share resources (e.g. printers and file servers), and the growth of network computing in the UNIX environment, led to widespread manufacture and use of network interface cards, bridges and hubs to interconnect PCs, workstations and servers across LANs.

At the same time that a vibrant data communications industry was building around LAN technologies, universities, research institutions, and government agencies were experimenting with and developing the building blocks for today's global Internet. These building blocks included the TCP/IP protocol stack, the algorithms and protocols for distributed management of a vast, decentralized network of networks, and switches—known as routers—for interconnecting these networks.

Internetworking

In the late 1980s and early '90s, LAN and Internet technology merged. Internetworking was then adopted as the principal means to interconnect isolated LANs. Wide area networks (WANs) evolved to link geographically dispersed LANs, while the TCP/IP family of protocols and routers provided the internetworking glue.

Although the Internet was widely used in research and academia prior to the mid-1990s, its explosive growth did not begin until the introduction of TCP/IP protocol software onto personal computers and Web browsers. This was the killer application that really launched the Internet's popularity. This phenomenal growth continues today and is responsible for catapulting the cable modem to its current position as the premier means to access the Internet and bandwidth-intensive data services.

HFC Network Topology

The cable television infrastructure is ideally suited to provide broadband, two-way data services. In fact, the HFC network topology can provide these broadband services far more efficiently than dedicated wire, telephone networks. This efficiency results from the cable television infrastructure design, which typically consists of a hybrid fiber-coaxial (HFC) network laid out in a tree-and-branch topology. A high degree of aggregation of users' traffic is made possible by this tree-and-branch topology, along with the cable network's shared-medium technology. What has resulted is an economy of scale unavailable to traditional telephone networks with their dedicated, twisted-pair local loops connecting each home to a central office.

Cable Modem Beginnings

By the 1980s, early interest in cable modem technology evolved, centered on the IEEE 802.4 token bus over CATV standard and its adoption by the automotive industry as a basis for the manufacturing automation protocol (MAP). Deficiencies in radio frequency (RF) technology at the time made early cable modem products costly to manufacture and maintain, and the resulting broadband data networks expensive to operate. Frustrated by these technical obstacles, by the late 1980s all but one manufacturer had abandoned efforts to provide commercial broadband networking solutions and instead migrated to other networking technologies and markets.

Meanwhile, the combined LAN and internetworking industries grew into a multibillion-dollar industry, providing data networking solutions to corporations, academic and research institutions, and governments. All these solutions employed baseband-networking technology, such as Ethernet, where digital signals are sent without frequency shifting. Although broadband networking's frequency division multiplexing offered the possibility to support multiple communications channels simultaneously, making digital convergence a reality, broadband's technical challenges still proved too daunting for all but one manufacturer determined to overcome cable TV aberrations.

LANcity's Invention

Early in 1993, it was LANcity Corporation of Andover, Massachusetts that overcame the previous 15-year technical, operational and management barriers to high-speed data over cable TV delivery. LANcity's cable modem technology provided the networking industry with the benefits of many other industry technologies, including cable TV, software, computers, and content providers. The cable modem was a critical piece needed to extend a network from a building to an entire city via the existing commercial cable TV infrastructure.

Obstacles Overcome

The newly invented cable modem technology was built from the ground up to overcome, at a minimum, the following key cable TV aberrations and requirements:

Cable TV Aberrations:

- excessive channel group delay
- limited distance due to delay through HFC plant
- upstream time division multiplexing
- microreflections
- varied operating levels
- varying plant losses
- time dependent plant level variations
- adjacent channel operation
- poor carrier to noise

Cable TV Requirements:

- long propagation delay, which can be greater than packet transmission time
- efficient support of a large number of potential users
- upstream channel noise and ingress
- plug-and-play installation
- interfaces standardized with existing networking solutions offered by the LAN industry
- use of any available frequency up to 750 MHz via software control
- efficient and manageable bandwidth
- significant group delay upstream (800 nanoseconds) and downstream (240 nanoseconds)
- frequency response tilt of up to 5 dB within 6 MHz micro reflections
- solid quaternary phase shift keying (QPSK) modulation technology for mass-volume production
- a scaleable solution adhering to internetworking principles
- Simple network management protocol- (SNMP) based management for end-to-end data services
- operation up to a 200-mile round-trip distance
- product installation, manufacturing and deployment occurring in less than one minute
- network privacy over shared media (data and manageability)
- a billing interface plus a centralized operations center
- price reduction from $10,000+ to less than $300.00 per unit
- successful implementation of a networking system on top of a two-way cable TV system
- robust-implemented physical access (PHY), and media access control (MAC)
- automatic adjustments for cable TV aberrations

Broadband Vendors

Broadband vendor evolution in the last eight years is shown in *Table 1*.

Cable Modem Mania

The successful launch and deployment of the cable modem offered the existing commercial cable television industry significant potential for new revenue streams. Early adopters included Continental Cablevision (now MediaOne), Tele-

TABLE 1

Broadband Vendor Evolution

1990	1993	1995	1998
LANcity Zenith Fairchild Ungermann-Bass Chipcom	LANcity Zenith Intel	LANcity Zenith Motorola Intel Hybrid Com21 Westend	3Com Bay Networks* CISCO Panasonic Sony Motorola Intel Com21 Teryon Hybrid Hayes Thomson Toshiba Libits Daewoo/Cadence NEC
Estimated Cost 2K to 20K	Estimated Cost 1K to 7K	Estimated Cost $500 to $995	Estimated Cost $150 ~ $350

*Bay Networks purchased LANcity
TABLE 1

Communications, Inc., Cox Communications, Time Warner, Cablevision, Adelphia, and Jones Intercable. Priced around $400 per unit, the cable modem broke the price barrier and created the cable modem industry—and cable modem mania.

Soon, multibillion-dollar industry leaders, including Zenith, Motorola, and Intel, realized the cable modem's potential, then repositioned their networking products for the broadband market. Today more than 20 vendors have entered the cable modem arena, which is now a multibillion-dollar industry itself. With the agreement of a data over cable service interface specifications (DOCSIS) standard for the next-generation retail version of the cable modem, mass marketing of this technology is expected by the end of 1998.

Cable System Installation

To deliver data over cable television services, two categories of equipment are required: headend equipment at the service provider's headend facility and a cable modem at each subscriber's location. Headend equipment interconnects the broadband CATV network to the service provider's own networking facilities. These facilities are home to various servers employed in the delivery of locally originated data services provided to subscribers, in addition to switches that provide access to the public Internet. Cable modems at subscriber sites are the means by which subscriber equipment, such as PCs, gain access to the broadband CATV network and the data services accessible through this network. A cable modem termination system (CMTS) at the headend controls to the broadband network, in addition to ultimate access to the data service provided across that network.

DOCSIS

A set of data over cable service interface specifications, better known as DOCSIS, has recently been completed by the Multimedia Cable Network System Consortium, also known as MCNS. Wide acceptance of DOCSIS has accelerated design and development of interoperable headend and cable modem equipment. As a result, more than 20 vendors are now manufacturing DOCSIS-compliant cable modems, in addition to cable modem termination systems and network management systems. By the close of 1998, DOCSIS cable modems will be available to the mass market through retail outlets. Additionally, all major multiple system operators (MSOs) will offer DOCSIS-based broadband data services in their respective service areas.

Now, with the addition of Sun Microsystems, Microsoft, and other software industry giants to the broadband arena, transparent, application-, and topology-independent networks capable of voice, data, and video communications will soon be enabled over the cable television infrastructure.

Cable Modems as "The Missing Link"

Since 1996, Wall Street has recognized cable modem technology to be a potential gold mine as it takes full advantage of the existing, underutilized cable television infrastructure that was laid in the ground 50 years ago. By the close of 1998, more than $15 billion will be invested in cable modem technology for deployment of broadband services through the acquisition of Continental Cablevision by U.S. WEST, the @Home Network public offering, the purchase of LANcity by Bay Networks, and the creation of Road Runner and Highway 1.

This "missing link" cable modem technology is having a profound effect on society. By providing last-mile connectivity, every household can now be transformed into the most powerful office possible. The cable modem revolution will assist in environmental conservation by encouraging home-office productivity over corporate commuting. Finally, as a facilitator for education, the cable modem will enable our children to rapidly and efficiently access a wealth of information from home.

Cable modem technology will contribute greatly to the 21st century digital convergence revolution. Because of this invention, the foundation for the information age has been firmly established, the "virtual information village" has finally been born, and the doorway to the future of computing has been opened.