I.S.F.
Improved Systems Factors

VOICE OVER IP (VoIP) AND ITS USE IN CORPORATE NETWORKS

BY MALCOLM HAMER

Introduction

This paper aims to describe the current state of Voice over IP (VoIP) technology and to highlight the advantages and drawbacks of the use of VoIP in various parts of a corporate voice infrastructure. Before tackling the main topic, it starts by presenting a brief history of data communication from the 1960s onwards and then describes the emergence of VoIP in the last few years.

The origins of packet switching

From the earliest days of telecommunication, switched networks (like the telephone network, in which any user can call any other user) have been based on circuit switching. In circuit switching the end-to-end communication path for each individual call is set up at the start of the call and remains exclusively dedicated to that call – even if the users remain silent for the whole call. There was no practical alternative to circuit switching, given the state of technology between 1892 (when the first automatic telephone switch went into service) and the early 1960s.

The first switched network that resembled a "data network" was the telex network, which was deployed starting in the 1940s. The telex network, like the telephone network, was based on circuit switching.

The first departure from circuit switching was proposed around 1965, in the form of a proposal for a packet switched data network. In the USA, packet switching was proposed by Paul Baran at the RAND Corporation. At roughly the same time packet switching was independently proposed by Donald Davies at the National Physical Laboratory (NPL) in the UK. Baran proposed packet switching as a solution to the problem of building a national computer-to-computer network that could survive a first-wave nuclear attack and thus preserve command-and-control communication. Davies conceived packet switching as a way of building a national data network that could efficiently handle "bursty" data traffic.

Specifications for a workable set of protocols to implement packet switching were worked out in the US between 1966 and 1969, under the DARPA-funded Arpanet project. The project aimed to build a proof-of-concept pilot network linking a number of universities. The first part of Arpanet went live in 1969. The packet switching protocol definitions were then refined as

Arpanet grew. These definitions became stable around 1975. They form the foundation of the Internet today.

Over the next 15 years Arpanet evolved into the Internet. But even up until the early 1990s it remained a much smaller network than the present Internet (in terms of the number of users). The user population remained a closed community of about 60,000 researchers in universities and research laboratories, and it carried a tiny fraction of the traffic that the Internet carries today. Also, there was no World Wide Web "layer" in the Internet. The commonly used protocols were telnet, FTP, and various email protocols. The switching nodes were mainly Unix minicomputers, running software that performed the packet switching function, rather than the purpose-built routers that are used in the Internet today.

Blended packet/circuit approach for commercial use

Arpanet was not set up to operate as a commercial network. In fact, it was too costly, for the amount of traffic that it could carry, to have become a commercial network. However, it demonstrated that a large-scale switched data network was a practical proposition and it sparked interest in developing the technology for a commercial network.

While the DARPA-funded work was continuing to refine the packet switching protocols used in Arpanet, several separate, parallel efforts to develop the first commercial data network were started during the early 1970s. In designing commercial data networks, pioneers such as the company Telenet (not to be confused with the telnet protocol) abandoned pure packet switching in favor of a blended packet/circuit approach. In this blended approach, "Switched Virtual Circuits" (SVCs) are established across the network for the duration of a "data call".

The change from the pure packet switching approach to the blended approach was made in order to overcome the problem of the high cost of the large amounts of computing power that are needed in each switch in a pure packet switching network. This processing power is needed in a pure-packet-switched network to individually route each data packet on each leg of its journey through the network, based on a point-in-time assessment of the conditions in the network. By contrast, the blended approach makes routing decisions only once per call, at the start of a call. After that, all the packets follow the same path marked out by the SVC. The blended approach, as used by Telenet, became the basis of a series of international standards, particularly the CCITT X.25 standard. These standards were enthusiastically supported by several European PTTs (government-controlled Postal, Telephone, and Telegraph authorities).

A further development of the SVC concept was the Permanent Virtual Circuit (PVC). This is a permanent mapping between two end points (creating a perpetual data call). From the user's point of view, a PVC resembles a leased circuit.

From the late 1970s through to the late 1990s, so-called "X.25 networks" became very popular as private corporate networks. (Strictly speaking, X.25 is a standard for a connection between a terminal or computer and the network, not a description of the protocols used within the network. However, the term "X.25 network" is widely used.) X.25 networks were also the foundation of most of the pre-Internet public-information networks. These networks were collapsed into the Internet by their owners when their owners entered the Internet Service Provider (ISP) business. Many of these X.25 networks are still very much alive and well, as explained later.

In addition to its long-running success as the basis of X.25-based networks, the blended packet/circuit approach was further developed in the late 1980s and early 1990s. The first evolution of the approach was to strip out a lot of the protocol features that were concerned with error detection and correction. This was done because, in a world of decreasing bit error rates on circuits, these features represented an unnecessary processing burden. This simplification led to Frame Relay networks. These were not deployed as extensively by corporations as X.25 networks had been. However, frame-relay technology was quite successful in the public network arena. This was mainly because users found that the performance of PVCs on most frame relay networks approached the performance of leased circuits, but at a much lower price. Like public X.25 networks, public frame relay networks are very much alive and well today. They are the most important part of most ISPs' networks, as explained later.

Yet another evolution of the blended packet/circuit approach was Asynchronous Transfer Mode (ATM), in which the variable packet size was replaced by a fixed-length packet or "cell". ATM has not been as widely adopted as X.25 or frame relay, for a number of reasons. First, many corporations and public network providers were confused by the way that the standard was initially presented by its designers. The designers of ATM had gone to great lengths to make the design seem new, choosing the confusing name "ATM" (which can be confused with Automated Teller Machine in banking circles), and introducing the term "cell". Second, vendors rushed to bring out products before the standard was stabilized. Third, the launch of ATM was quickly followed by the Internet boom, with the associated re-branding of IP as the "ultimate protocol". And fourth, Cisco acquired one of the leading ATM vendors (Stratacom) in April 1996, causing potential customers to worry about what changes Cisco might make to Stratacom's product line.

Circuit switching for data

Although the original Arpanet team looked at packet switching as a way of building a resilient network, they regarded the efficient use of bandwidth as an important secondary goal. This was mainly because the bandwidth available for data was expensive and quite limited at that point in time. There were no digital leased circuits: the circuits used to interconnect the Arpanet switches used modems operating over analog voice circuits.

When the effort to develop a commercial data network was started, the goal of the designers at companies like Telenet was to minimize the cost of all parts of the network, including both the switches (which represented a large, capital investment for the data network operator) and the bandwidth to be acquired from AT&T to interconnect the switches (which represented the largest ongoing expense for the data network operator). Because they saw the bandwidth as "expensive", they wanted to make efficient use of it. This meant retaining the packet-switching element of the blended packet/circuit approach. At the same time they wanted to minimize the cost of the switches. This was achieved by means of the circuit-switching element of the blended approach, which avoided the heavy processing load that pure packet switches have to handle. It is important to understand, however, that the "bandwidth is expensive" view was a customer's view of leased circuits, based on the prices that AT&T (at that time a monopoly) was charging. It was because of their perception of bandwidth as expensive that the designers at Telenet ruled out pure circuit switching. (Had AT&T, at that time, tried to build its own data network it would have taken into account the true cost of data circuits, rather than the prices that it was charging its customers.)

AT&T, and most PTTs and telephone companies around the word, were fully aware of the fact that transmission costs were much lower than then-current leased-circuit prices and that transmission costs were falling rapidly – and likely to continue to fall. In fact, the billions of dollars that started to be poured into the development of new telephone switches in the 1970s (leading to the No.1 ESS and its successors in the USA, and various other switches around the world such as the Ericsson AXE10 and the British System X) were invested by telephone companies, PTTs, and governments specifically because the total cost of telephone networks, once dominated by transmission costs, had become dominated by switching costs. (In the 1950s and 1960s, R&D funding was directly mainly at transmission systems. This was because, in that period, transmission costs dominated the costs of telephone networks. This R&D led to transmission systems that achieved much lower transmission costs, at which point attention was turned to the cost of switching.)

Because telephone companies and PTTs were aware of how cheap bandwidth really was, and how much cheaper it was likely to become in the future, the few efforts to develop switches for data networks that were sponsored by telephone companies and PTTs resulted in the selection of circuit switching as the best solution. Circuit switching made less efficient use of bandwidth, but resulted in significantly less expensive switches. Overall, this led to the most cost-efficient networks for network operators who face true transmission costs, rather than inflated transmission prices.

One of the few public circuit-switched data networks to go into commercial operation was the NORDIC data network. Developed jointly by the Scandinavian PTTs in the late 1960s, NORDIC became a big commercial success. It remained in service for many years, used heavily by Scandinavian companies, particularly the Scandinavian banks. From a switching point of view it was a completely stand-alone network, sharing only transmission facilities with the telephone network.

Several other PTTs, while recognizing that circuit switching was the best approach for data, wanted to avoid the cost of building a completely separate data network like the NORDIC network. They saw an opportunity to further reduce the costs of data switching by building combined voice/data switches. This approach came to be known as ISDN (Integrated Services Digital Network). These PTTs reasoned (correctly, as it turned out) that, although data would be important one day, the total bandwidth consumed by data would remain smaller than that consumed by telephone calls. If data switching were to piggyback on voice switching, the data switching function would enjoy economies of scale, and thus much lower switching costs than in data-only networks. A combined data/voice network would also be easier and cheaper to operate, because support processes (such as numbering plan administration, call record collection, and customer billing) could be common to the voice and data services.

The costs of transmission continued to fall in the 1970s and 1980s, as transmission technology that was developed in the 1960s (particularly digital transmission) was deployed everywhere. This should have made ISDN a big success. Unfortunately, the PTTs' plans for implementing ISDN as a public service ran about 15 years behind schedule for a variety of reasons. ISDN did not become a dependable and widely available service until the late 1990s. By this time the Internet was already firmly entrenched as *the* global public data network. As a result, ISDN found only a few niche applications, such as videoconferencing, leased circuit backup, and dial-up access to the Internet. Had ISDN been widely deployed in the early 1980s, as originally intended, things might have turned out very differently.

Comparison of circuit switching, packet switching, and blended circuit/packet switching

Before continuing with the story of IP, it is useful at this point to summarize the differences between the three different approaches to building a data network that emerged in the 1960s and 1970s:

- In a *circuit-switched* data network, the path that data will follow during a call is determined at the start of the call. As a result, the switches need to make a call-routing decision only once per call (rather than once per packet). In a circuit-switched data network, a fixed part of the capacity of each switch-to-switch transmission link is reserved for the duration of each call and is exclusively assigned to that particular call. This part of the capacity (or "bandwidth"), expressed in bits per second, is available for the call whether or not data is flowing at any given time. Circuit switching thus makes less efficient use of transmission capacity compared with the other two approaches, but requires the least amount of processing power in the switches (to handle the route-selection function).

- In a *packet-switched* data network, the path that each individual packet takes through the network is individually determined by the switches (or "routers") at the time that the packets pass through the network. Different packets may take different paths, depending on second-by-second changes in conditions in the network. Also, packets passing between one switch and another share a common transmission link between the two switches. No part of the capacity of the link is in any way reserved in anticipation of the need to carry packets between any two network end-points. Because of this, it is possible for the capacity of individual transmission links to be exceeded by the instantaneous demands of packets arriving at a switch and requiring transmission over a particular link. The switch handles this situation by sending packets that it cannot find room for on the link in question over another link, heading in the general direction that the packet needs to go, but not as directly as the first-choice link. Because transmission links can be filled to capacity with packets, packet switching makes the most efficient use of the transmission capacity; but it requires the greatest amount of processing power in the switches to handle the packet-routing function.

- In a *blended packet/circuit-switched* network, the path that data will follow during a call is determined at the start of the call. This path is the SVC. In contrast to the situation in circuit-switched networks, packets passing between one switch and another share a common transmission link: there is no fixed part of the capacity of the transmission link reserved for the call. However, there is generally some arrangement to keep track of how much of the capacity of each link has been committed to SVCs, based on the predicted *average* bit-rates of the SVCs that are assigned to the link. (Once all the capacity is committed, no further SVCs are allowed on a link: an alternative path through the network has to be found for the next SVC.) This makes it unlikely that the capacity of the links will be exceeded by the instantaneous aggregate demands of SVCs assigned to the link. However, from time to time these predictions will turn out to be wrong and a link will become congested. In this situation the switches will hold packets in a queue until the congestion subsides. This results in packets being delayed in delivery. In a well-designed blended network, the software that handles the tracking of capacity commitments will make continuous revisions to the rules used in predicting average bit-rates, based on actual measured averages, thus making congestion less likely. Blended packet/circuit-switching makes efficient use of transmission links because the capacity is pooled for all SVCs going over the link, rather than broken down into many parts (one for each call) as in a circuit-switched network. It requires much less processing power in the switches than in packet-switched networks. It is therefore an good compromise, taking the best features of circuit switching and packet switching. Another feature of many blended packet/circuit-switched networks is a mechanism for prioritizing the packets from different sources or within different SVCs, so that, in the event of congestion occurring on a link, the packets of the higher-priority SVCs are given priority. This allows "quality of service" (QOS) targets for packet delay to be set for each end-point or SVC.

The evolution of purpose-built IP routers

Although the first version of Arpanet used purpose-built hardware for the packet switches, once the definition of IP was stabilized, off-the-shelf minicomputers, running Unix, started to form the basis of Arpanet. The cost of these minicomputers made Arpanet too expensive to be the basis of a commercial data network. (Its costs were covered by government and university sponsorship.)

From the time that integrated-circuit microprocessors (CPU chips) and memory chips were first manufactured in large quantities (around 1970), to the early 1980s, the price/performance characteristics of these chips improved every year. This progressive cost reduction was driven by demand from the computer industry. Use of chips started with minicomputers, like the Honeywell 516 and the DEC PDP11, and accelerated with the growth of the personal computer industry from 1981 onwards. By the early 1990s, the power and price of chips had reached levels that nobody had predicted in the late 1960s when the first packet switches were being built.

Looking at the price/performance trends in chip production, the founders of Cisco (and a number others) recognized, in the early 1990s, that the point had been reached where standard CPU and memory chips could be used to build a packet-switching device that would be able to compete with the X.25 switches of manufacturers such as Telenet (by then part of Sprint), Hughes, Northern Telecom, and others. Such switching devices (or "routers" as they became known) would also be of interest to government laboratories and universities as replacements for the Unix minicomputers used as packet switches is the Internet. Cisco therefore designed and started to manufacture routers. These used powerful processor chips, together with a stripped-down version of the Unix packet switching software used in the Internet, and an input/output design optimized for data communication. Another important feature of routers was that they generally came with ethernet interfaces so that they could be connected to a Local-Area Network (LAN) on the user-facing side of the router and a leased circuit on the network-facing side.

Routers were an immediate success. Cisco's timing was lucky. Many large companies had reached the point where the traffic on their X.25 data networks was reaching the limits of the X.25 switches that they had installed in the late 1980s or early 1990s. The manufacturers of these switches had not upgraded their designs to take advantage of the most recent chips; and so the cost of upgrading existing corporate networks by buying bigger switches was enormous. At the same time, the use of LANs was becoming widespread. Many companies already used LANs to connect office PCs to the corporate X.25 networks via ethernet/X.25 gateways, and many data centers used ethernet to interconnect systems within the data center. This provided an ideal situation for the introduction of router-based networks, where the end-point routers would have ethernet interfaces to interconnect with office LANs and data center LANs.

In parallel with the move towards routers in corporate networks, the Internet was starting to change and grow. Three developments between 1986 and 1994 had sparked the transformation of the Internet from a semi-private niche network to a “people’s network”. These developments were:

- The adoption, in 1986, of the Domain Naming standard, developed by Jon Postel, Paul Mockapetris, and Craig Partridge. This allowed end-user devices, like PCs, to “hide” the hard-to-remember 16-bit IP addresses from users and represent them, instead, in a more friendly format like “www.amazon.com”.

- The development of the Hypertext Mark-up Language (HTML) and Hypertext Transfer Protocol (HTTP) by Tim Berners-Lee in 1989, incorporating practical implementations of the concept of hypertext (originally proposed by Ted Nelson in the mid-1960s), together with the concept of text and graphics pages built from multiple sources.

- The development, in 1993, of the first browsers, like Mosaic, which turned the concepts and standards defined by Berners-Lee into usable PC programs.

These developments made possible (a) the evolution of the World Wide Web, and (b) a much more usable email service, based on the “name@domain-name” address format. (These are the only two aspects of the Internet that most users today are aware of.)

As a result of the transformation of the Internet, routers were purchased by the thousand by ISPs, and by other stakeholders in the evolving Internet, as well as by companies that were building private IP networks to replace the corporate data networks that they had built in the 1980s.

The rate of adoption of IP took many people by surprise, including vendors such as IBM and DEC. Throughout the 1970s, 1980s, and early 1990s, both IBM and DEC had stuck to proprietary network protocol stacks (SNA and DECnet), rather than embrace X.25 and various public-domain, upper-layer protocols that were emerging for use on top of ethernet and X.25. When IP was re-born in the mid-1990s, they initially ignored it, and then begrudgingly started to provide limited support for it in their products. But it was too late. SNA and DECnet were swept aside by IP. The position of IP was further consolidated when Microsoft and Novell redesigned their PC-to-server networking software to use IP.

The evolution of the Internet into a commercial enterprise

As the developments described above made the Internet easier to use, and of interest to a much larger user base, many Internet Service Providers (ISPs) came into existence. Some of the ISPs were newly formed companies. However, many (including AOL) evolved from pre-web information-service companies that delivered information services via X.25 networks, accessed via dial-up connections. On the whole, the pre-web information-service companies did better financially than the new companies. They were able to use their own networks as “access” networks to inexpensively bring users’ connections to a small number of central points, where the connections entered the “IP core” of the ISP’s network.

The ISPs sold access to the Internet to the general public and to companies who wanted to establish websites. This led to rapidly increasing traffic levels. As a result, the ISPs had to quickly add huge amounts of capacity to the Internet itself. The ISPs rapidly transformed the Internet from a non-profit utility, serving a closed community of academic users, to a commercial data network serving the whole world. The ISPs were big customers of Cisco and other router manufacturers, as well as big consumers of bandwidth that they had to acquire from domestic and international carriers.

Although the ISPs bought a huge number of routers, it would be a mistake to think that their networks (which collectively form what we now regard as the Internet) consist entirely of circuits and routers (although most corporate are exactly like this – just circuits and routers). The ISPs, or at least the ones that survived, are very conscious of their costs. Although routers have fallen in price since they first came on the market, and although they are much cheaper than the Arpanet/Internet Unix minicomputers of equivalent power used up until the early 1990s, in general routers are still not competitive with switches that use a blended packet/circuit approach. As a result, most ISPs use X.25 networks and frame-relay networks as "access networks", between their customers' premises and their main operating centers. A typical customer connection thus consists of a PVC through the X.25 network or frame-relay network to a main operating center, where the PVC finally plugs into the router-based backbone. Thus, *a substantial part of what we regard as "the Internet" is based on X.25 and frame-relay switches*. These switches avoid the high costs of processing power inherent in handling data on a pure-packet-switching basis.

The change of voice networks from analog to digital

Between 1964 and 1993, in parallel with the evolution of data communication, telephone networks around the world were being converted from analog transmission and analog switching to digital transmission and digital switching. This started with the first commercial deployment of PCM transmission in the UK in 1964 and eventually led, by the early 1990s, to the complete conversion of the core of most national telephone networks to digital switching and transmission. This "core" included everything except the local loop.

The same digital technology that was used in the public switched telephone network (PSTN) was applied to PBXs. In this case it proved practical and economic to convert the extension-to-PBX "loops" to digital transmission, because these cable lengths are much shorter than public network local loops.

Several solutions to the problem of creating digital local loops in the public network were developed in the late 1980s and early 1990s, none of them completely satisfactory. These have been used to implement 2B+D ISDN connections and, more recently, DSL services (combining switched voice with point-to-point data paths, which are used to connect customers' equipment to an ISP's access network at the central office).

The origins of Voice-over-IP (VoIP)

Many potential uses of packet switching were discussed and tested between the formulation of the packet switching concept in 1965 and the emergence of the Internet in its present form in the mid-1990s. However, one idea – that of carrying voice on a packet switched network – was rarely discussed, and only then as a theoretical possibility rather than a practical proposition.

Carrying voice over a pure-packet-switching network was not considered practical, for two reasons:

- First, it was clear that the variable delays and disordering of packets that occur in packet networks are completely incompatible with the requirements of voice communication, namely, the delivery of digitized samples of the analog signal at regular intervals and in the correct order. It was known, from many experiments with voice communication, that users are very sensitive to variable delays in the voice path. Users will tolerate a small fixed delay, such as occurs with satellite transmission. However, if the delay varies, users find it very disturbing.

- Second, it seemed obvious, even as late as the mid-1990s, that the cost of a router powerful enough to handle any reasonable number of concurrent telephone conversations would be very high compared with the cost of a traditional digital circuit switch (such as a public network switch or a PBX).

In the latter half of the 1990s the ultimate economic structure of the Internet (in terms of who pays for what and on what basis) was still unclear. The Internet was initially regarded as more or less a "free" resource, because the cost of its infrastructure was borne by universities and government-funded research organizations. Even as the ISPs emerged, providing Internet access to the general public, their pricing plans were very simple. The pricing plans tended to present the customer with a zero marginal cost for bandwidth consumption, once access was paid for.

These early pricing schemes resulted in a lot of talk about "bandwidth is free". Not only was this a rallying call for the Internet-based start-up economy of 1997-2000, but it also provided the motivation for the development of PC hardware and software to implement VoIP, so that Internet-connected PC users could make VoIP telephone calls at zero cost. Further developments along these lines added video to voice for PC-to-PC videoconferencing.

The quality of these early implementations of VoIP (and videoconferencing) was very low for two reasons:

- First, the Internet, at that time, had a rather patchy infrastructure. There were large parts of the Internet that were under-capacitized; and, even where capacity was adequate, failures of various parts of the network were common, leading to periods of poor performance.

- Second, the designers of the VoIP solutions were aiming to make them work with fairly low-bandwidth connections, such as 8 kbps or 12 kbps, because most uses had only 56 kbps or less via dial-up connections to the Internet and wanted to be able to "surf" at the same time as talk using VoIP.

Both these constraints had more or less disappeared by 2003: the performance of the Internet had become more predictable; and users started to acquire high-speed connections via DSL and cable. However, it is important to bear these initial constraints in mind when thinking about the first seven years of VoIP development.

Cisco identifies VoIP as an opportunity to sell more hardware

The fact that users seemed prepared to tolerate the shortcomings of VoIP in its early experimental forms prompted vendors to start to take the technology more seriously. Cisco, in particular, recognized the opportunity that VoIP represented: if people could be persuaded to tolerate the quality of VoIP, and if voice network operators could be persuaded that routers were cheaper than traditional telephone network switches (purchased from Lucent, Nortel, and others), then Cisco would stand to sell large numbers of very powerful routers for telephony. There were similar opportunities within private corporate voice networks.

In 1998 John Chambers started making speeches about the upcoming "convergence" of voice and data. He predicted that data traffic would soon outstrip telephone traffic. The implication of this prediction was that voice would start to be viewed as a *secondary* use of bandwidth and therefore one that could enjoy economies of scale by piggybacking on the data infrastructure (i.e. the Internet and corporate private IP networks). Such a "convergent" approach could then carry voice at near-zero marginal costs.

So appealing was this story that several new telephone companies were founded that tried to build IP telephone networks to compete with the established telephone companies, using Cisco routers in place of traditional voice switches. (Note that these companies were trying to build their own networks. This was a much more costly proposition than recent IP telephone services, like Vonage, which are based on VoIP over the public Internet, rather than a purpose-built network.) Building VoIP networks using routers quickly turned out to be a recipe for losing a great deal of money. Such networks only appeared viable under highly unrealistic assumptions, where the costs of interconnecting with the existing telephone networks, and the lack of a workable solution for the local loop, were ignored. It was also quickly recognized that a Cisco router does not come with many of the functions that are a vital part of commercial telephone network operation, such as numbering plan administration, call record collection, and customer billing.

A number of established telephone companies, reading about VoIP, thought that perhaps they should be doing something with it. It was clear to them that there was no immediate cost-saving opportunity domestically. However, a number of pilots were set up internationally using VoIP. Almost all of these have now been abandoned – party because of problems with service quality and reliability, and partly because there was no easy way to establish a tariff arrangement that was acceptable to all stakeholders.

These commercial failures of VoIP in the public network arena have caused the vendors, led by Cisco, to focus on private networks, specifically PBXs, call centers, and corporate site-to-site private networks. This move has, in turn, spurred the established vendors of PBXs and ACDs to start incorporating VoIP in their products, fearing that Cisco will make inroads into their markets.

Recent VoIP products and applications of VoIP

The focus of Cisco and others on the private-network VoIP market from 1999 onwards has produced a number of products and VoIP applications that are important in considering where VoIP will go in the next few years:

- <u>IP Telsets</u>: PBX telsets are now available that perform the digitization and packetization of the voice signals in the telset, presenting the signal as a twisted-pair ethernet interface. Assuming that you want to convert voice signals to IP, the cheapest place to do it is probably in a digital telset. There is already a need for an analog/digital conversion chipset in the telset. This chipset can be upgraded, at very little incremental cost, to perform packetization along with digitization. Thus, the cost of building IP telsets *should* be only slightly higher than the cost of comparable digital PBX telsets. (This does not necessarily mean that vendors, like Cisco, will price them the same as traditional digital PBX telsets. But in the long run, any price differences will tend to disappear.)

- IP-Ready Conventional PBXs: PBX vendors such as Avaya and Nortel have added ethernet interfaces to their switches, to allow the aggregated packet streams from, typically, 96 telsets to be fed into the PBX. The twisted-pair ethernet cables from each telset must be aggregated into a single ethernet connection outside the PBX. This is done by connecting the cables from the telsets into a device in the wiring closets on each floor, where the “horizontal” wiring connects to the “vertical” network service the building. This device is typically an ethernet switch, which in turn has a single ethernet connection to the PBX. The per-port cost of the ethernet switch is roughly the same as the per-port cost of conventional digital ports on a PBX; so the move to IP does not really save any money. However, enthusiastic purchasers of VoIP technology can make their purchases look attractive by booking the cost of the ethernet switches against their general IT/desktop-services budget, instead of against the voice budget. More will be said about this later.

- Pure-IP PBXs and ACDs: Cisco, and other vendors outside the voice telecomms industry, have developed switching products based on VoIP, particularly for applications such as call centers. The established PBX vendors are now trying to emulate these products. The most well-known product of this type is the Cisco ICS7750 switch, and its associated Call Manager software suite. (“ICS” is short for “integrated communication system”, a term which is also sometimes used by other vendors to refer to pure-IP PBXs in general.) These products bring added value in call center applications in that they simplify implementation of CTI (Computer/Telephony Integration) compared with trying to do CTI with traditional ACDs. Also, they make it possible to create call centers that handle a mixture of workflows, such as voice, real-time Internet “chat” calls, and email response – although very few call centers have so far successfully implemented the handling of mixed traffic workflows by individual agents.

- Private VoIP networks: If a company has several pure-IP PBXs, like the Cisco ICS7750, or several IP-ready PBXs, these can be interconnected via a corporate IP network, such as a corporate IP network originally built to handle only data traffic. This provides a new way of creating a private voice network, without using the traditional approach of leased T1/E1 circuits, or T1/E1 channels derived from higher-capacity leased circuits using time-division multiplexers (TDMs) or ATM switches. These networks allow intracompany calls to be placed without incurring local or long-distance call charges from local and long-distance telephone service providers.

Residential VoIP services

Before looking at how the above products and applications can be used in a corporate voice network, it is useful to briefly look at recent developments in residential VoIP services.

A number of new, non-experimental VoIP products and services have become available for residential use in the last few years. In particular, a number of companies have created a residential VoIP market, most notably Vonage (although they were not the first). Larger players, like AT&T, have now started to enter this market. What Vonage did was to launch a service based on a VoIP adaptor box that is inserted between a standard analog telset (or a fax machine) and a high-speed Internet connection (such as a cable modem or a DSL connection). The box performs analog/VoIP conversion, just like that performed within an IP telset used in a corporate environment. The box also interacts with servers, located at Vonage's operations centers, to control the setting up and receiving of calls, without the need for the box to be connected to a switching system like a PBX.

Vonage provides a set of interfaces between the Internet and the telephone network so that calls can be made between a Vonage subscriber and a normal fixed-line or mobile telephone number, and vice versa. By taking advantage of number-portability regulations, Vonage is able to allow its customer to move their existing telephone numbers over to their Vonage service.

What distinguishes Vonage from earlier VoIP-via-the-Internet services is that the design of the box is based on a no-compromise VoIP encoding standard that uses the same bit-rate as a conventional voice network (64 kbps, based on the G.711 voice encoding algorithm). Because the datastream has to be broken down and enclosed in packets, the total bit-rate consumed by a Vonage box is about 80 kbps. This would clearly have been useless in the early days of the Internet, when almost everyone was using dial-up connections at speeds of 56 kbps or less. However, given that a cable connection or DSL is a prerequisite for using the service, adoption of G.711 is no problem. Residential use of G.711 sets a new standard for VoIP. (The box can be optionally reconfigured to use G.729 voice compression, in order to reduce the bandwidth used.)

Some of the early VoIP telset products (and some of the special-purpose telsets for use with dealer turrets) support only low-bit-rate voice, using G.729 voice compression. Their designers had mistakenly assumed that there was a need to minimize the amount of bandwidth used for voice communication – so that VoIP could be carried economically over corporate IP Wide-Area Networks (IP WANs). This was a big mistake. G.729, although comparable to voice quality on a mobile telephone, is noticeably inferior to normal telephone communication. It includes optional silence suppression (often enabled), which means that the outgoing speech path is completely muted when the user stops speaking. This often causes the other party to keep asking, "Are you still there?". Based on interviews with users in various pilots, G.729 is described by call center staff as "tiring to listen to all day" and "often the cause of mishearing important information"; and by trading floor staff as "completely unacceptable for any form of trading when multi-million-dollar transactions are being done using voice".

More recent telsets support G.711, but many corporate VoIP networks have already been set up to operate using G.729 in order to minimize bandwidth consumption.

It seems likely that, as people start to notice the difference between G.729-configured and G.711-configured telsets and networks, and use services like Vonage at home, the use of G.729 will be abandoned in favor of G.711.

The economics of residential VoIP

Given that Vonage and its competitors use a technical standard that makes less efficient use of bandwidth than conventional telephone networks (80 kbps versus 64 kbps), it is reasonable to ask how the service can be made economically attractive to customers. The answer is that Vonage takes advantage of the fact that Internet service pricing is based on the marginal costs of network capacity (such as that of cable television networks and telephone local-loop cabling). By contrast, conventional telephone service pricing is based on fully-loaded average costs. So, the cost of the 80 kbps of capacity used for a Vonage call, together with the monthly subscription the user pays to Vonage to cover it operating costs (including the cost of its interfaces to the conventional telephone network), is less than the cost of subscribing to, and using, a normal telephone line.

Another factor that favors Vonage and its competitors is the fact that the many different taxes charged on conventional telephone services have not, so far, been paid by the VoIP providers. Over time it seems likely that these services will attract the same taxes as conventional services.

The long-term impact of services like Vonage is hard to predict. Residential VoIP users are in the minority. If everyone were to use this type of VoIP service then every part of the infrastructure would need to be expanded significantly – the cable television networks, the telephone companies' local loops, and the IP core of the Internet. If this had to be done then it would be unlikely that VoIP could continue to be priced based on marginal costs, because it would become the dominant use of bandwidth. However, such a situation is a long way off. Only a small proportion of the population in the US has high-speed Internet connections.

Corporate use of VoIP

We now come to the main topic of this paper: given these various developments in VoIP in the last seven years or so, and given the products that are now available for corporate use, where does it make sense to utilize VoIP technology in a corporate environment?

An important point which I will make now, and return to later, is that anyone considering an implementation of VoIP must first think very clearly about what the objective is, and whether a VoIP solution will meet that objective. It is not a good idea to rush into buying VoIP technology just because it is there and it sounds interesting. (A lot of companies discovered in 2000 that redesigning their business model to fit the Internet was good for some companies but not for all – in fact, not even for the majority.)

In the rest of this paper I will examine five possible uses of VoIP in a business setting:

- VoIP via the public Internet as a way of minimizing telephone call charges
- VoIP over a corporate IP network as an alternative to a dedicated private voice network
- VoIP as a way of serving remote sites using a central-site PBX or ACD
- VoIP telsets as an alternative to conventional telsets, without a change of PBX
- A pure-IP PBX as an alternative to making an existing PBX "IP-ready"

VoIP via the public Internet

While services like Vonage demonstrate that VoIP via the Internet can be turned into a fairly reliable and economic service, VoIP via the Internet much less suitable for use in a corporate environment. There are two ways in which VoIP via the Internet could *theoretically* be introduced in a corporate setting, as follows.

The first way is to allow calls to be established directly from a VoIP telset (or an analog telset plus a Vonage-like box) in one office, out through the firewall, through the Internet, then back into another office via a firewall. One problem with this arrangement is that any kind of direct office-to-office VoIP, without the call passing through a PBX, requires that the caller knows, or can look up, the IP address of the IP telset of the called party. However, in most companies the actual IP addresses of all devices connected to the LAN are “non-routable addresses”, such as ‘192.168.64.123’. A non-routable address is an address that is not recognized by routers in the Internet. Also, the actual IP address of each device may change from time to time where dynamic address assignment is in operation, using a DHCP (Dynamic Host Configuration Protocol) server. When a user is surfing the public Internet, the company’s firewall translates the current non-routable address of his or her PC into a routable address on outbound packets, and translates this routable address back into the non-routable address on inbound packets. The routable address used for this purpose is selected from a pool of routable addresses owned by the company. (This technique is called “Network Address Translation”, or “NAT”.) The association of the routable address with the non-routable address of the user’s PC is a short-term association, which expires after a short period of time. Thus, the routable address that is temporarily associated with the user’s PC exposes the PC to the public Internet only while the user is actively surfing the Internet. A short time after the user stops, the routable address is returned to the pool of available externally-valid addresses, so that it can be later associated with another user’s PC. Any inbound packet addressed to that routable address after this point in time will be either discarded or, if the routable address has been re-used, sent to another user’s PC.

If a user wants to be able to receive VoIP calls via the Internet direct to an IP telset, then (a) his or her telset would have to be assigned a fixed non-routable address, by disabling DHCP, and (b) this fixed non-routable address would have to be permanently mapped (in the company’s firewall) to a fixed routable address (an arrangement referred to as a “conduit”). *In practice this does not, and should not, happen.* Network security staff go to a lot of trouble to make sure that incoming Internet traffic cannot get to any device other than the company’s website servers. They are particularly careful about keeping inbound traffic (other than “response” traffic from websites that a user is visiting) away from office LANs and PCs.

The upshot of this is that the use of direct-to-the-telset VoIP via the Internet cannot be enabled in a business environment. (The recent appearance on the market of products that are specifically designed to bypass corporate firewalls, to enable desktop-to-Internet VoIP, should be viewed with alarm.)

The second possible approach to VoIP via the public Internet is to route the calls to and from the telset via a pure-IP PBX which, in turn, is connected to the Internet. This avoids the problem of keeping track of IP addresses: the PBX either keeps track of the current non-routable address of each telset or establishes ethernet communication directly using the telset’s MAC address (Media Access Control address). Instead, the two PBXs will establish communication from the calling extension to the called extension using the called extension’s telephone number. This is a more workable solution to VoIP via the Internet. However, the connection of pure-IP PBXs to the Internet gives rise to another security serious problem: it exposes the PBX to external attacks. Also, if at any point the ethernet to which the telsets are connected is intermingled with the office PC LAN, it provides a potential path between the Internet and the PC LAN that bypasses the main firewalls that protect the internal network.

While vendors, such as Cicso, may state that connecting a pure-IP PBX to the Internet has many benefits, such as allowing the system to handle telephone calls, voicemail, and email in an integrated manner, such a direct connection (even with firewall functions built into the PBX) places a company’s communications infrastructure at risk.

In summary, both approaches to using the Internet as a means of cheaply connecting telephone calls between two business offices are fraught with security problems and should be avoided.

VoIP over a corporate IP network (WAN or VPN)

A much less problematic application of VoIP is over a private corporate IP network – which may be a WAN, using leased circuits between one router and another, or some form of Virtual Private Network (VPN), where the site-to-site connectivity is provided via a semi-private shared network (in which the company's traffic is securely isolated from other companies' traffic on the network). Assuming that the company's firewalls are set up properly, the corporate network is safe from outside attack.

The case for putting voice traffic on the corporate IP backbone network, rather than using the PSTN, either at discounted long distance rates or under a voice Software-Defined Network (SDN) arrangement, rests on the assumption that the corporate WAN or VPN represents a fixed cost for the company. This may appear to be true in the short term. But in the longer term it is likely that companies who put VoIP traffic on their existing IP networks will find themselves adding bandwidth, and upgrading routers, in order to (a) get acceptable VoIP performance during data traffic peaks, and (b) maintain data performance as VoIP traffic is added to the network. VoIP equipment vendors will encourage potential purchasers to think of bandwidth for VoIP on their existing IP backbones as "free", knowing that it will be many months before anyone gets around to calculating the incremental cost of bandwidth additions and router upgrades.

One way to make VoIP on corporate networks more reliable (and less likely to interfere with business-critical data traffic) is to segregate data-carrying and voice-carrying facilities – ideally to the point where there is effectively a separate network for voice. However, this goes against the "free bandwidth" sales pitch used by the vendors. What may happen, in companies that go for VoIP in a big way in their corporate networks, is that they will start off sharing the corporate backbone, but will end up with segregated networks within two years.

Having an almost completely separate IP network for corporate voice communication brings into clearer focus the economics of VoIP for site-to-site voice communication. If the costs of ownership of all the IP network infrastructure and bandwidth that has been added to achieve VoIP is objectively assessed, the resulting cost is very likely to exceed what it would have cost to complete the calls via the PSTN, using the previously existing PBX infrastructure. (Long distance charges, at corporate-discount rates, or under a voice SDN tariff, may be as low as 2 cents/minute or less.)

Unfortunately, this kind of objective economic analysis is unlikely to be done, at least in the near future. A VoIP project is most commonly initiated by the corporate IT group, or corporate data networking group, after hearing a vendor's pitch for VoIP. The vendor either presents a very limited cost analysis, using unrealistic assumptions that favor VoIP, or simply sells VoIP as "the way of the future" (the gospel according to John Chambers).

VoIP as a virtual remote concentrator for a PBX/ACD

While carrying voice traffic over WANs or VPNs, in an attempt to save on long distance call costs, may not make economic sense when objectively analyzed, VoIP as a way of creating a "virtual remote concentrator" for a small branch office is a much more practical and economic proposition. If a PBX or ACD at a large office location is equipped with an IP interface (to make it IP-ready), then a small branch office can be equipped with IP telsets connected via ethernet switches to the corporate IP network that links the branch office to the larger office. The voice traffic will then ride the IP network, along with data traffic, to the larger office. There it will enter the PBX/ACD at the IP interface. In this case there is no pretence that the bandwidth used by the VoIP traffic is free. Rather, the network between the branch office and the larger office would be correctly engineered for the voice/data traffic mix, and the cost of doing this would be clearly understood.

The attraction of this approach comes from the fact that the total cost of the VoIP arrangement – the IP telsets plus the ethernet switches to which they are connected, plus the incremental bandwidth and router capacity – is less than the total costs of ownership of a new PBX/ACD for the branch office, and possibly cheaper than installing a remote PBX switching module, controlled by the main PBX. The IP telset approach can be economic even in the absence of any data traffic, where the IP network is provided exclusively for voice.

The savings are particularly significant where the main switch is acting as an Automatic Call Distribution system (ACD) and the remote site is a call center. In this case, software modules for ACD functionality can be shared between the central location and one or more remote call centers. Such an arrangement for call centers has operational advantages as well:

- It allows enhancements to ACD functionality at the central site to immediately be available at the remote call center or centers, and

- Where calls answered by agents at the remotes sites are transferred to a third destination, the transfer switching action can easily be done at the central site, under the control of the remote agent – an improvement over having the transferred call "tromboned" through a switch at the remote site, or using complex ACD-to-ACD signaling to achieve the central-site switching.

Remote VoIP call centers, connected to a central IP-ready ACD, have been successfully implemented in both domestic call centers and in offshore call centers. In the case of an offshore call center, the main ACD is in the country being served (for example, the USA). It acts as the concentration point for the inbound '800' number traffic from customers. This traffic is distributed to the VoIP-equipped call center (or call centers) offshore. (The central ACD may also support a domestic call center that handles certain types of customer enquiries or transactions, and acts as a fall-back call center in the case of any protracted network problems.) Offshore call centers that were originally using a local ACDs, connected to the ACD in the country that they are serving via a non-IP voice network, have reported improved reliability and clearer connections after changing to VoIP.

The central PBX or ACD in this type of arrangement is most commonly a traditional switch (such as an Avaya G3R) that has been made IP-ready. This is a low-risk approach, as will be explained later. In a few cases, the switch is a pure-IP switching system, such as a Cisco ICS7750. Using a pure-IP switch has some advantages, such as making some CTI solutions for call centers easier to implement. However, pure-IP switches are less stable and lack some software features that have been standard in PBXs and ACDs for over twenty years.

Introduction of IP telsets (with an IP-ready PBX)

In this section the conversion to IP telsets of an existing office, with its own PBX, will be considered as a "stand-alone proposition". In other words, the conversion of one or more floors in the building from traditional telsets to IP telsets will be considered, without paying any attention to whether the existing PBX is being made IP-ready (by the installation of IP interface cards) or is being replaced by a pure-IP PBX. The costs and logistics of what happens in the PBX room are left out of this description entirely.

IP telsets operate over the same type of twisted pair cable that is used for ethernet LANs. This means that, if a floor is being fitted out for the first time, or if wiring is being re-run as part of a renovation project, the same type of cable (for example, CAT 5 or better) would be used for the LAN connections and the telsets. In fact, it has been a common practice, for several years, to use the same cable for LAN and telephone wiring, because the cost savings from using a slightly lower grade cable (for example, CAT 3) for PBX connections are barely worth the trouble of buying the two different types of cable. Also, it is useful to be able to re-assign voice cables as LAN cables for unexpected requirements in certain cubicles and rooms, without having to worry about whether the cable is of the right type. Where all the cabling is already CAT 5 or better, no change would be needed to the wiring to convert to IP telsets. The only change that might be needed would be at the termination points (on the "tech plate" in the cubicle/office/room, and in the wiring closet), depending on what kind of socket had been used originally. IP telsets will typically present an RJ45 plug to insert into an RJ45 socket in the tech plate, whereas traditional telsets (including digital PBX telsets and analog telsets) use RJ11 plugs.

One way of avoiding the need to change the sockets or wiring is to use the option, which most IP telsets have, of sharing a single physical connection (from the cubicle/office to the wiring closet) between voice and data. Telsets with this feature contain a "mini ethernet hub" and provide a secondary ethernet socket into which the ethernet connection from the PC can be plugged. This means that the PC can be plugged into this secondary socket and the telset can then be plugged into the RJ45 socket in the tech plate that the PC was previously plugged into. This is cabled to the wiring closet (where it is typically patched into a port on an ethernet switch). The secondary ethernet connection on the telset is pinned as a HED (Hub-like Ethernet Device), so the PC, which is a TED (Terminal-like Ethernet Device), can be plugged into it using the existing straight cable that was previously plugged into the RJ45 socket in the tech plate. (There is no need to change the cable to a cross-over cable). The main ethernet connection on the telset (the one that goes via the tech plate to the ethernet switch in the wiring closet) is pinned as a TED.

Although using this option has the advantage of avoiding the need for any wiring changes, it has a number of drawbacks. In particular:

- It introduces more points of failure into the PC's network connection: the physical connection into the telset, the telset itself, and the telset's power connection.

- It increases the likelihood of the telephone and data connections adversely affecting one another. In particular, the voice path may be impaired when a large volume of data is being transmitted to or from the PC (for example, when a large document is being sent to a shared printer).

- It makes it difficult to position the telset in the most convenient location on the desk because it now has two somewhat-stiff four-pair cables attached to its rear.

For the rest of this section it will be assumed that the telephone wiring is kept separate from the PC LAN.

In the wiring closet the ethernet connections that are used for voice will need to be patched to an ethernet switch. Technically this switch can be the same ethernet switch that is used for the office LAN. However, there are a number of reasons for keeping the voice traffic completely separate from the PC LAN. The most important of these is the need to preserve voice communication in the event of something going wrong on the LAN, such as a "packet storm" caused by a virus infecting many PCs and/or servers, or by a faulty device somewhere on the LAN. In any case, it is unlikely that the wiring closet will contain enough spare ethernet switches or spare ports to accommodate all the telset connections, and so additional ethernet switches will need to be purchased.

Note that, regardless of whether a dedicated or shared ethernet switch is used to serve the IP telsets, the switch must be connected to UPS-protected power, otherwise the entire floor will lose telephone service in the event of a building power failure. This contrasts with conventional PBX telsets which typically draw their power from the PBX and have no active components in path between the PBX and the telset.

In summary, in the best-case scenario, a conversion to IP telsets *might* be accomplished without buying any new ethernet switches, and without installing new cables or changing RJ11 sockets to RJ45 sockets. (The only items to be purchased in this case would be the telsets themselves and, for a traditional PBX, the IP cards.) But a more likely scenario is that it would be necessary to also purchase one or more ethernet switches, and/or replace CAT 3 cabling with CAT 5 or better, and/or change the RJ sockets in the tech plates and in the distribution panels in the wiring closets.

Reports from sites that have been equipped with IP telsets (used with traditional PBXs fitted with IP cards) have shown that the IP telsets, by themselves, do not have any drawbacks from a user's point of view; but neither do they have any significant advantages from the user's point of view. This being so, the case for such a conversion must be based on costs versus operational advantages. These costs are roughly as follows:

- In the case of a straight conversion from existing telsets to IP telsets, the required investment is about $350 per telset for the new IP telsets, plus the cost of the ethernet switches (at an average per-port cost of about $100), plus re-cabling costs if the existing cables are not at least CAT 5.

- In the case of a floor being equipped for the first time, the incremental cost of IP telsets over traditional digital PBX telsets is about $100 per telset, plus $100 per ethernet switch port. (If a traditional PBX exists, but does not have any spare line cards to support additional telsets, the costs of buying IP telsets can be somewhat offset by the savings achieved by buying a small number of IP line cards instead of a larger number of traditional line cards, since each IP line card can support up to about 96 IP telsets.)

The operational benefits that could be used to justify these investments might be based on the ability to physically move telsets between any two cubicles or offices, without requiring a conventional PBX "move". Provided that the PBX software identifies the IP telset based on its ethernet MAC address, that telset will retain its extension number and its features no matter where it is plugged into a tech plate (within the floors or zones equipped for IP telsets). But there is a downside to this. If a telset is damaged, then a step is now needed that is *not* needed with conventional telsets: the PBX has to be told to recognize the replacement telset as being assigned to the extension number of that user. (By contrast, a conventional telset just needs to be plugged into the right socket.)

This feature of IP telsets (that they are each unique as regards MAC address) raises a security issue which needs to be considered. Anyone who steals (or borrows) a telset can assume the extension number and class of service of its normal user. They can make calls (which may be expensive or of a harassing nature) from a location of their choosing, where they cannot be observed. Until the removal of the telset is detected, and action taken, these fraudulent or offensive calls appear to have been made by the normal user of that telset. The bad guys can also receive calls intended for the normal user of the telset: they may be able to pose as the normal user and obtain confidential information by so doing; or they might say things to callers that will embarrass the normal user. The bad guys might return the "borrowed" telset before the normal user returns from a meeting, business trip, or vacation, so that nobody is immediately aware of what has happened. They might even put another telset in place of the one that they borrowed so that its absence is not noticed by other staff while they are using the borrowed telset from elsewhere in the building.

Powering IP telsets

A further characteristic of IP telsets is important to note. Unlike traditional basic digital PBX telsets and analog telsets, the first-generation IP telsets do not draw power from their connection to the switch (in this case, the ethernet switch). They require a separate power supply that is plugged into the 110/240-volt AC supply. The addition of yet another of these black cubes (or "wall warts" as they are sometimes called), under a user's desk, is something of an inconvenience. (Power sockets generally seem to be in short supply, as evidenced by the number of power strips that can be found under desks in most offices.) The requirement for external power also means that the telset is vulnerable to the wall wart becoming accidentally unplugged.

Where UPS-protected power is available at users' desks, this should be used for the telset power, so that telephone service is not lost in the event of a failure of street power. Where individual UPS units are provided for users' PCs, the telset power should be derived from these, provided that there is a spare socket. Where there is no UPS power available then the vulnerability of the telephone service to power failures should be taken into account in any contingency plans.

If the option is used of sharing an ethernet connection between the telset and the PC (which, as mentioned above, can be problematic), users should be made aware that their PC's network connection will be lost if the power supply to their telephone becomes unplugged. Helpdesk scripts should include a question about the telset power supply in cases where a user reports "I have lost my Internet connection, I can't print, and my telephone is dead".

In the future, the problem of supplying power to IP telsets will be addressed by "Power Over Ethernet" (POE). A POE standard, known as IEEE 802.3af, was finalized in June 2003. (Products based on the draft standard were announced even before the standard was finalized.) However, the technology should be approached with some caution. Some vendors rushed to come out with products before the standard was finalized. Even with the standard now finalized, it will take some time for all of the implementation problems to be addressed and for a number of reliable, fully-compatible products to emerge. Early purchasers of POE devices may find that the products of one vendor do not interwork with those of another. They will therefore become locked into a single vendor for the closet-end equipment and the IP telsets. Also, vendors will be charging a premium for POE IP telsets (over wall-wart-powered IP telsets, which are already more expensive than conventional PBX telsets.) Taking into account the extra cost of the POE IP telsets, plus the POE ethernet switches, the economics of IP telsets start to look less attractive.

The 802.3af POE standard calls for power to be fed to the user-end ethernet device using 48 volts d.c. power to deliver a minimum of 12.95 watts to each end device. Thus, a 100-port ethernet switch capable of providing power to all its ports concurrently would be delivering up to about 2 x 100 x 12.95W = 2.6kW of power. (The power supplied to the end devices is multiplied by 2 because up to 50% of the power will be dissipated in the cables.) This is a significant amount of heat dissipation for a wiring closet. If the closet is not fully air-conditioned then it would become necessary to at least increase the amount of ventilation. In some cases it would mean the installation of full air conditioning in the closet.

As an alternative to replacing ethernet switches with POE switches, some vendors (such as PowerDsine) sell “midspan” units. These are connected between the ethernet switch ports and the cables, in order to inject the POE power. This avoids the need to replace existing ethernet switches; but it still raises the issue of heat dissipation in the wiring closet.

When you start to examine all the issues surrounding POE, the rationale for POE IP telsets starts to look shaky. Consider the chain of events:

- You start by considering IP telsets as an alternative to conventional telsets. You find that IP telsets are more expensive than conventional telsets. Also, you have to buy additional ethernet switches if you do not have enough spare ethernet ports on the existing switches in the wiring closets, or if you want to maintain a separation between voice and data traffic. Any new ethernet switches are a further additional cost, on top of the IP telsets.

- Next you find that IP telsets need wall warts to power them, which is inconvenient, so you look at buying the even more expensive POE IP telsets instead.

- To be able to use POE IP telsets you need to install POE ethernet switches in order to feed power to the telsets. You find that these switches are more expensive than conventional ethernet switches.

- Then you find that POE ethernet switches dissipate a lot more heat than traditional ethernet switches, so the wiring closets will get too hot, risking switch failures. You therefore face extra costs for the installation of closet air conditioning, or at least increased ventilation.

In other words, as you go down the path of POE, extra costs are piled on top of extra costs. Clearly, POE should be approached with extreme caution and careful financial analysis.

It should also be noted that POE is of limited use beyond powering IP telsets; so you should not believe a POE vendor’s sales pitch about POE having “many other possible uses in the future”. All other ethernet devices in use today around the office already have their own power (PCs, shared printers, shared scanners, and so on), and will continue to do so. The only other serious

POE end-device product that has been announced is a small security camera. (In a misguided attempt to illustrate the flexibility of POE at trade shows, PowerDsine has exhibited an electric shaver, modified to run off a POE-equipped ethernet connection. This gimmick serves only to highlight how limited the applications of POE really are.)

A pure-IP PBX as a straight PBX/ACD replacement

In the last three sections we looked at the following uses of VoIP technology:

A. PBX-to-PBX VoIP via a corporate IP network.

B. Equipping remote sites with IP telsets that work with a central site IP-ready PBX.

C. Conversion of a main site to IP telsets.

For each of these, the PBX needs to be capable of supporting IP telsets. This can be achieved with a conventional PBX that has been made IP-ready (by the installation of one or more IP interface cards). It can also be achieved by using a pure-IP PBX. It is not necessary to replace a traditional PBX with a pure-IP PBX in order to do any of the above three things. It follows that a proposal to replace a traditional PBX by a pure-IP PBX must be considered on its own merits. In particular:

- Since these three applications of VoIP (A, B, and C above) do not require a pure-IP PBX, any proposal made by a vendor, or by an IT group, for converting from a traditional PBX to a pure-IP PBX, is flawed if it states that a pure-IP PBX is a prerequisite for achieving the claimed benefits of VoIP (as in arrangements A, B, and C). Such a proposal should be viewed with suspicion.

- The case for converting from a traditional PBX to a pure-IP PBX must be based on the economics and intrinsic capabilities of the system itself, and not on benefits that can be achieved by simply making the existing PBX IP-ready and introducing IP telsets.

- When a pure-IP PBX is introduced, all existing telsets served by the PBX have to be converted to IP telsets, because traditional telsets will not work with a pure-IP PBX. Therefore, when making a comparison between replacement of a traditional PBX by a pure-IP PBX, versus leaving everything just as it is, all the extra costs described earlier have to be added to the cost of the pure-IP PBX, including IP telsets, ethernet switches, new wiring, and RJ45 sockets. (As mentioned before, a common trick, when trying to show a VoIP conversion in a favorable economic light, is to book the costs of adding ethernet switches, and upgrading the wiring, as an “IT expense”.)

Bearing the above points in mind, we will now look at the case for replacing a traditional PBX with a pure-IP PBX.

The most well known pure-IP PBX is the Cisco ICS7750. This product is being deployed in pilot sites by a growing list of customers. Accounts of these deployments differ considerably. The primary determinant of whether the assessments are positive or negative seems to be whether the person talking is an “IT person” or not:

- IT staff who have sponsored a pure-IP PBX conversion project seem to be reporting good results. They say that the users like the additional features provided by the system, such as the ability to look up a telephone number in Microsoft Outlook and dial the number on their telephone by clicking on the entry. They say that there have been no major problems with outages.

- Staff responsible for traditional voice telecommunications functions seem to be reporting disappointing results, such as intermittent voice-path problems and occasional total system outages. They also seem to be reporting that the workload on the staff who support the telephone service has increased, not decreased, and that the problem of ongoing support has been dropped in their laps without adequate information from the vendor.

The intermittent speech-path problems reported by users include:

- One-way transmission.
- 100% return echo heard by external callers into the PBX.

These problems are frequently reported when a pure-IP PBX is used, but much less often when IP telsets are used with an IP-ready traditional PBX. The exact rate at which these problems occur has not been established by any published study. It seems to be high enough to halt pilot deployments that do not have the weight of a CIO or CTO behind them. One major bank has had a pilot that has been halted and restarted several times over a period of two years because of problems like this, whereas another major bank has endorsed VoIP as its strategic direction for voice and is pressing ahead with a multi-site roll-out, in spite of continuing reports of problems at its pilot site.

Clearly, there is a difference in expectations between the two cultures (IT and traditional voice telecommunications). IT staff seem to regard the occasional crashing of servers, the need to periodically reboot servers, the surfacing of application bugs, and a “background level” of user complaints, as part of their daily experience. Many IT staff will assign greater weight to the benefits of new functionality than to the disadvantages of a less resilient system. (This valuing of new functionality over stability is, after all, how companies came to vote with their purchasing dollars for Microsoft’s software, and for Microsoft’s approach of turning the user community into an extension of its QA department.) By contrast, voice telecommunications staff expect complaints about service, if any, to arise from problems in the execution of moves and changes, not on something as basic as the day-to-day telephone service. They regard an incident in which a large number of users lose dial tone as a rare and terrifying event.

The designers of software-controlled PBXs learnt, in the 1970s, how to build highly reliable systems. This was done by means of a redundant architecture, typically using main and backup control processor cards, and N+1 sparing of other less critical components, combined with software that was very thoroughly debugged over many years. By contrast, vendors like Cisco are new to the field of voice telecommunications. Very few of their data products have had any kind of redundant architecture and Cisco had to create the software to control systems like the ICS7750 in a relatively short space of time. It would be a source of surprise and wonder if Cisco had managed to reach the level of resilience that one finds in a Nortel or Avaya PBX in the time that Cisco has had available since entering the voice market.

But more important than the “call processing” software that controls the setting up of calls is the basic IP routing software. There is a fundamental difference between a router and a conventional TDM PBX or public network switch. In a TDM switch, the switching functions are hardware-based. The controlling software that runs on the switch's control processor is only involved in the interpretation of dialed digits, external signaling, the setting up of speech paths, and the clearing down of speech paths. Once a call is underway, the software leaves the call alone. The movement of bit-streams through the switch is handled by hardware. By contrast, a VoIP switch relies on software that manipulates each individual packet throughout the duration of the call. Bugs in this software therefore affect individual telephone calls that are in progress. The IP routing software, as used in, for example, the Cisco ICS7750, is a version of router "IOS" software that has undergone many large changes in the last 8 years. IOS has had a history of bugs being added, with each new feature-enhanced version, faster that prior bugs could be fixed. Even before Cisco entered the voice market, Cisco's software was widely criticized by companies, such as banks, that needed their data networks to be reliable. Industry observers have described IOS as "tens of millions of lines of spaghetti code".

The complete software suite in a product like the Cisco ICS7750 can thus be viewed as consisting of two parts:

- Call processing software. This software is about 3 years old and is roughly equivalent to the software that runs in the CPU of a traditional PBX. It is considerably less resilient than the call processing software in a traditional PBX (software which has undergone debugging over the last 25 years).

- Packet-switching software. This software is about 8 years old and, when used in a pure-IP PBX, is roughly equivalent to the hardware-based TDM switching matrix in a traditional PBX. This packet-switching software is enormously less resilient than the hardware-based logic of a TDM switching matrix, which is for all practical purposes bug-free.

The software of the ICS7750 is thus a combination of a part which is considerably less resilient that that of a traditional PBX and a part which is enormously less resilient than the switching matrix of a traditional PBX. It should come as no surprise, therefore, that users of pure-IP PBXs report problems with both individual telephone calls and overall service resilience.

In summary, there appear to be very good reasons for not contemplating an acroos-the-board change to a pure-IP PBX at present. However, as explained above, the shift of the decision-making about voice strategy from traditional voice telecommunications groups to IT groups, in response to Cisco's massively-funded pro-VoIP publicity machine, means that such decisions may not be based on logic or objective economic analysis.

On this last point – economic analysis – it is hard to find any published findings. Cisco and others try to keep discussions away from detailed cost figures; and the traditional PBX vendors seem unwilling to go on the offensive against products like the ICS7750 for fear of being branded "anti-VoIP". However, there have been verbal reports of total life-cycle costs for pure-IP PBX installations running as high as three times those of traditional PBXs.

The future of VoIP in the PSTN

Clearly, VoIP is not going to go away anytime soon. There are advantages to using IP telsets to equip remote branch offices, remote call centers, and work-from-home locations that make this approach attractive, independent of the pros and cons of pure-IP PBXs versus IP-ready traditional PBXs.

The advance of VoIP in the private-network market leads some observers to suggest that, ultimately, the PSTN will migrate to VoIP technology. This would certainly be Cisco's wish. With so many people believing that the PSTN will migrate to VoIP technology, we certainly cannot dismiss the idea without a strong argument against it.

If the PSTN were to migrate to VoIP technology then it would strengthen the case for buying pure-IP PBXs: these would eventually be able to connect to the PSTN via high-speed IP connections, instead of via one or several T1 or E1 voice-TDM channels. With a few rare exceptions, the traffic passing through a PBX is predominantly inbound and outbound PSTN traffic, not extension-to-extension traffic. Therefore, anything that reduces the cost of handling traffic from and to the PSTN has a significant impact on the cost of the PBX.

Conversely, if the PSTN is going to remain a circuit-switched network, this weakens the case for buying a pure-IP PBX. It means that a pure-IP PBX, or a corporate network of pure-IP PBXs, would forever be a "VoIP island in a circuit-switched PSTN sea". The complexity of any PBX derives mainly from its need to interwork with the PSTN and handle functions such as cost management (class of service restrictions, route-based carrier selection), call accounting (CDR), external numbering (direct inward-dialing, switchboard functions), and caller ID pass-through. Any change in technology that makes the internal architecture of the PBX diverge from the standards of the PSTN is unlikely to reduce the cost of the PBX and is more likely to introduce new operational problems. So, if the PSTN is going to remain circuit-switched, it would be impossible to build a strategic case for buying a pure-IP PBX based on it being "ready for when the PSTN converts to VoIP".

Because of its bearing on the choice between making a traditional PBX IP-ready and buying a pure-IP PBX, the question of whether the PSTN will migrate to VoIP technology will now be examined.

A useful starting point is to consider what has happened during the last 30 years in data communication. We can then consider what the implications of this are for voice communication. As described at the start of this paper, the economics of data communication have turned out, historically, to favor circuit switching over packet switching. Organizations that have gone into the data communications business thoroughly understanding all their short-run and long-run costs, starting with PCI, Graphnet, Telenet, and Tymnet in the early 1970s, and continuing through to the ISPs of the late 1990s, have used a blended packet/circuit approach (such as X.25 or frame relay) wherever possible in their networks. In the case of the NORDIC network and ISDN, pure circuit switching was adopted. In terms of circuit-miles, the Internet is today based predominantly on blended packet/switching; only at the Internet's core is traffic switched on an IP basis. Packet switching, as a network-wide approach, has appeared to be cost-efficient only in situations where a network operator, such as Arpanet, was facing circuit prices that vastly exceeded true underlying transmission costs, as a result of a leased-circuit provider (like AT&T) exploiting its monopoly position.

IP switching is, of course, an essential part of the Internet because, in the beginning, it was what defined Arpanet and now, in many people's minds, defines the Internet. However, as explained earlier, ISPs remain competitive by buying the minimum number of routers that they can get away with, and continue to make extensive use of X.25 and frame relay in their "access networks" – between customers' sites and their main operating centers, and also between their dial-access infrastructure and their main operating centers.

As applications such as full-motion video start to dominate Internet traffic, there will be growing pressure on the cost/performance characteristics of IP routers. Extensions of IP such as MPLS (Multi-Protocol Label Switching) introduce blended packet/circuit-switching protocols into IP networks, in order to handle high-bandwidth delay-sensitive traffic like video. MPLS has so far been deployed mainly in private and semi-private IP networks. However, within a few years we may well find that the IP core of the Internet is gradually migrated towards blended packet/circuit-switching technology, using MPLS or a further evolution of MPLS. (See Appendix 1, “Could The Internet Really Work Without IP?”.)

The pressures that will drive ISPs and other network operators to consider blended packet/circuit switching (or perhaps even pure circuit switching) for data communication will increase because of three trends:

- *The cost of bandwidth is falling faster than the cost of switching hardware*. This trend, which established itself in the late 1950s, is likely to continue for two reasons:
    - The rate of cost/bandwidth improvement in optical fiber systems is higher than the rate of cost/performance improvement in routers. This is because the rate of cost/bandwidth improvement in optical fiber systems depends more on improved modulation and multiplexing techniques than on chip prices, and also because the rate of improvement in cost/performance of a finished product like a router is lower than that of its components.
    - The need for the Internet to carry more bandwidth-hungry applications, like video, tends to drive up the cost of routers non-linearly, whereas the cost of bandwidth is a linear function of the bandwidth.

- *The cost of router software is growing.* The software licensing and software support part of the total cost of ownership of software-intensive products like routers is growing. Vendors are charging more each year for their software licenses and for ongoing software support, because they find that the size of their software development and maintenance groups is growing each year as their products become more complex.

- *The cost of staff to manage the configuration of IP networks is growing.* Network operators, such as the ISPs, find that they need to have large and growing numbers of highly trained, well-paid staff to deal with the complexities of setting up router networks – staff with a level of skills that is not required to administer a circuit switched network like the telephone network.

As a result of these trends, the ISPs’ attention will become increasingly focused on the cost of the IP core of their networks: they will seek cheaper switching options in order to remain profitable. History shows that heavier traffic, cheaper bandwidth, and the need to reduce the cost of switches, drive designers away from the pure-packet-switching end of the switching spectrum towards the circuit-switching end of the spectrum.

Given all this, and given that telephone traffic consists of a continuous stream of delay-sensitive bits (which a packet-switched network is not very good at dealing with), it is highly unlikely that an economic case can ever be made for converting the PSTN to VoIP. On the contrary, the Internet will, over time, start to move towards a blended packet/circuit-switched core.

If this prediction turns out to be correct then it follows that a pure-IP PBX, or a corporate network of pure-IP PBXs, will not only be a “VoIP island in a circuit-switched PSTN sea” but will also come to be regarded as technological dead-end.

Conclusion

The current prominence of IP, and the perception that IP is the king of all protocols, is largely an accident of history. The Internet, in its 1993 form, was “in the right place at the right time” when Tim Berners-Lee’s work on HTML led to the first practical browser (Mosaic). The Internet of 1993 had evolved from Arpanet. Arpanet had been based on pure packet switching because of its origins as a DARPA-sponsored effort to develop a resilient data network that could survive a first-wave nuclear attack. Arpanet remained packet-switched in the 1970s and early-1980s because optimal use of monopoly-priced bandwidth remained an important requirement. When bandwidth costs started to fall, IP was too firmly entrenched to be reconsidered.

Network operators (like Telenet, Tymnet, PCI, and Graphnet) who were able to obtain bandwidth at nearer-to-cost prices chose blended packet/circuit switching, or (in the case of the NORDIC network) pure circuit switching, as a way of minimizing the cost of the switching components of their networks. The Internet today is, in large part, a blended packet/circuit-switching network, with IP used only at its core.

In the core of the Internet, which is currently using IP, two forces are at work that will drive the ISPs to search for ways to carry more traffic with lower incremental investments in routers: (a) bandwidth costs are continuing to fall, making switching costs (that is, routers) a greater proportion of total costs, and (b) the use of video applications is starting to demand the handling of much higher packet rates without any variability in packet delays. The ISPs will demand router software that supports a blended packet/circuit protocol in parallel with (but ultimately in place of) IP. MPLS (which is starting to be widely used in corporate WANs and VPNs) incorporates a blended packet/circuit switching is a step towards the dethrowning of IP.

While services like Vonage can benefit from the low marginal costs of Internet bandwidth enjoyed by high-speed Internet users, a conversion of the entire telephone network to VoIP would increase total costs considerably and is therefore highly unlikely.

Given this background, VoIP as a switching technology (in the form of pure-IP PBXs) is likely to be a technological dead end. Investments in this technology should be approached with great caution.

By contrast, use of IP telsets, in combination with IP-ready traditional PBXs, represent an easily-deployed and fairly-stable technology that can be useful for remote offices, call centers, and work-from-home VoIP platforms.

My recommendations are therefore as follows:

- Make existing PBXs IP-ready, by installation of the appropriate interface cards, and deploy IP telsets in situations where there are clear cost and/or operational advantages in doing so.

- Do not mix PC LAN traffic with VoIP traffic: use ethernet switches to serve IP telsets that are separate from those serving the PC LANs.

- Do not waste your money on POE telsets and POE ethernet switches.

- When doing economic analyses of VoIP projects, make sure that all costs arising from introduction of IP telsets are included.

- Do not accept, without rigorous economic analysis, the proposition that putting voice traffic on a corporate IP network, using VoIP, will generate savings versus public network telephone calls (which can cost as little as 2 cents/minute) when all costs are taken into account.

- By all means, try a pure-IP PBX pilot in a building that is not involved in mission-critical operations or customer contact; but do not commit to a major deployment of pure-IP PBXs until the future of VoIP in particular, and IP in general, become clearer.

- Closely involve the voice networking staff at an early stage in any IP-telset or other VoIP project, so that all the required changes to procedures for telset moves, adds, and changes, and all the changes to the operational-support databases used for these activities, can be properly thought through before the conversion takes place.

- Do not purchase any VoIP product that does not support full-quality (G.711) digital voice encoding, and configure all VoIP components (telsets, WAN interfaces, and so on) for G.711 operation.

APPENDIX 1

COULD THE INTERNET REALLY WORK WITHOUT IP?

We tend to regard IP as a defining characteristic of the Internet, even though much of what we regard as the Internet is based on blended packet/circuit switching components, and only the central core consists of routers. Devices that are connected to the Internet, such as PCs, “talk IP” to the central core via the ISP’s X.25 or frame-relay “access” network.

If we examine what is happening on those access networks, we see at least three protocol layers in operation, for example:

- Top layer: TCP
- Middle layer: IP
- Bottom layer: frame-relay

Supporters of IP sometimes refer to the way that IP works as “connectionless”. They claim that connectionless operation is, in some philosophical way, better than a network based on any kind of “connection” or “session”, such as circuit switching or blended packet/circuit switching (using SVCs). However, far from operating in a connectionless mode, browsers in fact operate in a connection-oriented manner:

- Browsers establish one or more TCP “connections” in order to transfer the HTML files, graphics files, and the executable applets, that make up a web page. TCP is a connection-oriented protocol layer that sits on top of the connectionless IP layer.

- On top of the TCP layer, many websites use further protocols and techniques to “sessionize” the communication between the user’s browser and the website. These include the use of SSL (Secure Sockets Layer) to establish a secure session between the browser and the website, and the use of cookies to associate the user’s viewing of a web page with his or her previous viewing of other web pages on the same site, so that the series of web pages that the user views can be tracked by the website as a session.

Thus, the activity that takes place when a user checks his or her bank account balance, or orders a DVD, is very much a “session” – in exactly the same sense that the communication between the user of a 1970s-style “dumb terminal” and a mainframe was a session. It is only the underlying IP layer of the protocol stack which is connectionless. The upper layers of the stack are used to hide the connectionlessness of the IP layer from the user.

The operation of the IP layer is hidden from users in other ways too. The widespread use of DNS hides, from the user, another defining characteristic of the Internet’s IP core, namely the use of IP addresses, by replacing them with DNS names like “www.amazon.com”.

Given that the Internet is nowadays used almost entirely in a connection-oriented mode, as described above, and given that almost all connections are set up using DNS names instead of IP addresses, it would be an easy matter to reproduce the behavior of the Internet using a circuit-switched or blended packet/circuit-switched network in place of the IP switching layer.

As an illustration of this, imagine that:

- A DNS-like service is used to translate domain names (for example, “www.amazon.com”) into X.121 addresses (as used in an X.25 network), or into ISDN telephone numbers. The user’s browser, acting on the results of the DNS translation, then instructs the PC to set up SVCs or ISDN calls to the webserver.

- Next, the user’s browser uses a variant of TCP (or a similar protocol), over the SVC or ISDN call, in order to retrieve the HTML files, graphics files, and applets that constitute the web page.

- Wherever the HTML code of the web page specifies that part of the material for the page should be retrieved from another address (for example, from “www.imdb.com”), the domain name would be translated, using the DNS-like service, to an other X.121 address or ISDN number. The browser would then instruct the PC to set up another SVC or ISDN call to get the files from the site at that address. (A single physical connection to an X.25 network or the ISDN can be used to initiated two or more calls at once, so the transfer of several files can take place concurrently, as it does with browsers operating in IP mode today.)

The result of all this, from the user’s point of view, would be indistinguishable from the operation of a standard browser that retrieves web pages via the IP-based Internet. In other words, *nothing in the HTML + DNS model, developed by Tim Berners-Lee in 1989, requires the protocol layer underneath TCP to be IP*.

Like the HTTP protocol that operates on top of TCP to retrieve web pages, the email protocols SMTP, POP3, and IMAP also operate on top of TCP and are therefore connection-oriented. The other two commonly-used protocols, FTP (File-Transfer Protocol) and telnet (a session-based interactive terminal-to-host session protocol), operate on top of TCP and are connection-oriented. All five of these top-layer protocols could easily be used on top of a version of TCP adapted to operation over an SVC or an ISDN call.

In summary, the IP layer of the Internet is not a pre-requisite for the majority of services that use the Internet. These services could operate just as well over an X.25 network, a frame-relay network, or the ISDN. Similarly, other blended packet/circuit switched protocols can be substituted for IP without affecting users’ experience of the Internet.

This is not to say that IP is likely to disappear quickly. It is firmly entrenched and is supported by powerful vendors like Cisco. Also, the new generation of CIOs and CTOs that came to power on the crest of the Internet wave has learnt to think of IP as the revolutionary force that swept away the ancient technologies of IBM, DEC, AT&T, and others. However, over time, IP could, and probably will, be superseded by a blended packet/circuit-switching protocol.

The first step in this direction was the introduction of MPLS (Multi-Protocol Label Switching). MPLS is basically a blended packet/circuit-switching protocol, interwoven with IP. MPLS was introduced in order to meet the needs of delay-sensitive traffic such as video and voice. It revamps the techniques used in frame-relay and ATM. The use of the term “label” (rather than a familiar term like “frame”) is an attempt to hide, from IP-enthusiasts, the fact that very little in MPLS is really new.

MPLS is already gaining widespread acceptance in corporate networks and semi-private VPNs. The major VPN providers, like Equant, have already implemented MPLS throughout their networks and many large companies have MPLS projects underway.

The way that MPLS implementations operate today interweaves the MPLS and IP layers, so that the network can carry normal IP traffic as well as MPLS traffic. However, over time, MPLS may be simplified by being detached from IP, thus creating a straightforward packet/circuit-switched network. Pockets of IP-less MPLS could grow so that, eventually, IP is driven out of the Internet.