Beyond Third Generation Telecommunications Architectures: The Convergence of Internet Technology and Cellular Telephony*

Randy H. Katz
randy@cs.berkeley.edu
EECS Department, University of California, Berkeley, USA

I. Introduction

The telephony network has evolved over the last 120 years into one of the most ubiquitous and complex technological systems known to humankind. Over one billion of the five billion inhabitants of the planet have made a telephone call, an astounding number that continues to grow as the developing world increases its investment in telecommunications infrastructure.

Yet at the end of the 20th Century, the engineering design principles that underlie this great technical achievement are being challenged. The fastest growing sector of the telecommunications industry is cellular services, introducing new demands for user and handset mobility in the future network. In contrast to wireline telephones, cellular handsets are sophisticated electronic devices able to perform advanced digital signal processing, extensive user interface support, and will ultimately provide many personal information management functions (address book, electronic mail, personal messaging, etc.).

Following a similar explosive growth trend is the number of Internet subscribers. Today, their dominant application is data, with emerging capabilities for telephony services within the context of the Internet. Internet usage counters many of the engineering assumptions about the telephone network, which has been carefully engineered to deliver outstanding performance for real-time voice. The rise of handheld devices, Personal Digital Assistants, integrating ever more extensive computing and communications capabilities in a small formfactor, will further increase the importance of data applications within the network.

II. Comparison of Telecommunications and Data Communications Industries

A. The Telecommunications Industry: Challenges and Opportunities

Telecommunications is perhaps the largest service industry in the world. The total worldwide telecommunications revenue was approximately $900 billion at the end 1996. This is expected to grow by 50% to $1300 billion by the year 2000.

Cellular telephony remains the fastest growing sector of telecommunications. In the United States, the largest worldwide market for cellular services, subscriber growth remains on an exponential curve. At the end of the 1996, the number of U.S. subscribers grew to 44 million[1]. The number of cellular sites in the U.S. grew by 25% in one year, to 30,000 cell sites. It is predicted that by the Year 2001, the number of worldwide mobile telephone users will have grown to 600 million. By the same time, the number of Internet users will have grown to 400 million. Both trends remain on the exponential growth curve well into the next century[2].

The rapid rise of *d*igital cellular systems is an important related trend. Digital technology not only enables high quality voice and more efficient spectrum utilization. It provides a critical foundation in the wide-area for ubiquitous data-oriented applications as well[3].

B. Data Communications Industry: Opportunities and Challenges

The portable computer market is the most rapidly growing sector of the personal computer market. By the year 2000, it is expected that this will represent a $50 billion market[4]. Telecommunications equipment providers are increasingly aware of the rising importance of data services. Nokia has stated that it expects 20-30% of its revenue to come from data-oriented equipment and services by the end of the century.

An emerging market is for computing devices that have been called Personal Digital Assistants (PDAs). These are small "palmtop" devices that integrate many personal information management functions, like a personal calendar, notes, and an address book. The U.S. Robotics PalmPilot is the first such device to become a commercial success, selling over one million units since its initial product launch in 1996. The PalmPilot Professional executes the TCP/IP network stack, thus making such devices full fledged nodes on the Internet one a dial-up connection is established. The popularity of this device indicates that a market exists for integrated computing and communications capabilities that can fit in a pocket. This could be the shape of things to come for future mobile telephones or computer-telephone combinations.

In terms of yearly revenue, the telecommunications industry is considerably larger than the data communications industry. Part of the imbalance is due to relative size of service revenue compared with equipment revenue (the data communications industry consists of equipment vendors; the telecommunications industry is a mixture of both operators and equipment providers). Nevertheless, as indicated by predicted future growth, the latter is growing much more rapidly. The telecommunications industry is expected to grow from about $900 billion in 1996 to $1300 billion by the year 2000, a growth rate of 50%. In the same time period, the data communications industry is expected to grow from $29 billion in revenue to $72 billion, a growth rate of 300%.

III. Is the Future of the Network Voice or Data?

Beyond cellular telephony, the second trend affecting future telecommunications infrastructures is the dramatic rise of the Internet and the World Wide Web. Reuters reports that that there are 82 million PCs connected to the Internet today, and they expect this number to grow to 286 million by the Year

*This paper is based on the keynote address given by the author at the ACM Mobicom '97 Conference, Budapest, Hungary, September 1997

[1] CTIA Web Page

[2] Information from Ericsson Radio Systems

[3] Information from Ericsson Radio Systems

[4] Arthur D. Little, Industry Estimates, quoted in The Economist Magazine, 14 May 1994.

2001. CommerceNet's Internet Demographic Survey reports 55 million Internet users in the United States and Canada alone. It is reasonable to double this number to represent a worldwide Internet population of about 100 million users. Thus, the number of Internet users is perhaps 10% of the 1 billion strong worldwide telephone population.

Internet access changes the fundamental assumptions about the underlying network traffic. It has been reported that more than 50% of the telecommunications traffic in the San Francisco Bay Area is already data. The much longer connection times associated with data calls have overloaded the local telephone switching infrastructure, causing excessive call blocking during peak usage periods. The network was provisioned for short duration voice conversations. It does not represent the best design for computer sessions. For data-oriented applications like Internet access, the appropriate network design metric is data throughput, not voice quality. Data support within the network is becoming increasingly important.

IV. A Comparison of Internet and Telephone Technology

A. Strengths and Weaknesses of Internet Technology

1. *Strengths*

A key underlying assumption of the Internet is that end nodes are intelligent and have the ability to execute the TCP/IP protocol stack. Until recently, this implied a fairly complex (and expensive) end device, such as a laptop computer. Today's personal digital assistants, costing a few hundred dollars, are now sufficiently powerful to execute these protocols. While existing telephone handsets are quite a bit dumber (and less expensive) than this, it should be noted that cellular telephones do possess embedded microprocessors and have the capability to run more sophisticated software than is currently the case for the telephone network.

The Internet achieves its robust communications through packet switching and store-and-forward routing. It is not necessary to create a circuit between end points before communications can commence. There is no state in the network, in particular, there is no connection state in the switches. A switch can fail, and since it contained no critical state, the network can adapt by rerouting the packet stream around the lost switch. Routing information is distributed among the nodes, with no centralized node controlling critical functions such as routing. This helps enhance the network's resilience to failure.

One the great successes of the Internet is its ability to operate over an extremely heterogeneous collection of access technologies, despite large variations in bandwidth, latency, and error behavior. The key advantage this provides is that the network needs to make few assumptions about the underlying link technologies.

2. *Weaknesses*

The Internet also has some serious weaknesses. First, it provides no differential service. All packets are treated the same. Should the network become congested, arbitrary packets will be lost. There is no easy way to distinguish between important traffic and less important traffic, or real-time traffic that must get through versus best effort traffic that can try again later. Second, there are no control mechanisms for managing bottleneck links. The third weakness lies in one of the Internet's strengths: store-and-forward routing. The queuing nature of store-and-forward networks introduces variable delay in end-to-end performance, making it difficult to guarantee or even predict performance. A fourth weakness arises from another one of the Internet's strengths: decentralized control. In this case, it is very difficult to introduce new protocols or functions into the network, since it is difficult to upgrade all end nodes and switches. The last in our list of weaknesses comes from the Internet's assumption of a cooperative routing infrastructure. Without a truly trusted infrastructure, the Internet as it now exists suffers from well-known security problems.

B. Strengths and Weaknesses of Telephone Technology

1. *Strengths*

A fundamental assumption of the telephone network is that the end nodes require little or no intelligence. This makes it possible for the network to support a diverse collection of low cost end devices, ranging from telephone handsets, cordless telephones, headsets, and fax machines. The telephone network has been highly optimized to provide excellent performance for voice. It performs this job extremely well, even in the face of large latencies associated with international telephony. Finally, the network is a true utility. It has demonstrated its ability to operate successfully through power outages and in the face of natural disasters. This is achieved, in part, through the sophisticated, robust, and hierarchically arranged switches controlled and managed by the service providers.

2. *Weaknesses*

The phone network achieves much of its advantageous voice performance through a significant overallocation of bandwidth resources. For example, the modest 3.4 KHz audio voice band signal is converted to a 64 kbps pulse code modulation (PCM) digital voice encoding. Internet audio applications and cellular telephony have demonstrated more sophisticated encoding formats that operate at rates as low as 8 kbps, with even lower rates becoming available. Unfortunately, this representation for digitized voice permeates the entire telephone network. Even data bandwidth is provided in 64 kbps increments! The vagaries of the voice data type have influenced the entire network design. For example, some bits in the audio encoding are more important than others, and thus are more carefully protected. Another example comes from interleaving schemes within the voice signal. Acceptable latencies in human speech (approximately 250 ms) have led to channel interleave schemes that create serious delay problems for data transmission, especially at the transport layer. A second weakness is that the telephone network's switching infrastructure has been determined by the statistics of voice call traffic. As the frequency of long duration data calls increase, the network operators have been discovering the inadequacy of their existing network design. A third weakness is the difficulty of introducing new services into the so-called "Intelligent Network," which is the advanced service architecture of the existing phone network. The problem arises from complex service interactions. Some strengths also lead to weaknesses when viewed from a different perspective. The phone network's approach to robustness, through highly engineered, fault tolerant switches, is very expensive.

C. ATM: The Grand Convergence?

Asynchronous transfer mode (ATM) technology has been trumpeted as a Grand Unification, combining the best attributes of packet-switching technology with the best attributes of circuit-switched technology. The network can support either real-time data like voice through its fine-grained multiplexing capability, or more time-insensitive data transmissions. The controversy surrounding ATM is whether it will become a dominant end-to-end technology, or one high speed point-to-point technology among others.

1. *Strengths*

One of ATM's key strengths is its virtual circuit concept, with call set-up in advance of data transmission. This is critical for ATM's ability to manage scarce resources and achieve its evolving model of quality of service (QoS) guarantees. A second strength is its use of fixed, small size "cells" to enable fast switching. Such an organization makes it possible to build switches that can quickly examine incoming streams and dispatch them to the appropriate output port. A third strength is the incorporation of sophisticated statistical multiplexing mechanisms to support a variety of traffic models.

2. *Weaknesses*

The fundamental weakness of ATM is its strong connection-orientation, inherited from its telephony ancestors. Every cell in a transmission stream must follow the same path, in strict sequence, and this path much be established in advance of the first cell transmission. Thus, even short communications sequences require an end-to-end set-up, adding to communications latency. If a switch fails, the connection must be torn down and re-established, implying that either switches must be made very robust or the network will not perform well in the event of switch failures. Furthermore, providing adequate support for end-node mobility in the connection-oriented model is difficult, and remains an area of intensive research investigation.

D. Next Generation Internet

An alternative effort to build an "integrated services" network, that is, a network equally good at carrying delay sensitive transmissions as delay insensitive ones, based on Internet technology, has been called "Integrated Services Packet Network" (ISPN).

The next generation Internet has several attractive features. It will have ubiquitous support for multipoint-to-multipoint communications based on multicast protocols. Recall that multiparty calls were the number one requested additional functionality by cellular telephone subscribers. The next generation of the Internet protocols, IPv6, has built-in support for mobility and mobile route optimization.

To achieve good real-time performance, the Internet Engineering Community has developed a signaling protocol called RSVP (ReSerVation Protocol). This enables the creation of a weaker notion of connections and performance guarantees than the more rigid approach found in ATM networks. Performance is a promise rather than a guarantee, so it is still necessary to construct applications to adapt to changes in network performance. RSVP is grafted onto the Internet multicast protocols, which already require participants to explicitly join sessions. Since receivers initiate the signaling, this has nice scaling properties.

A key element of the ISPN's approach to providing reasonable performance is the Real Time Protocol (RTP) built on top of the underlying routing protocols. RTP supports application level framing. This places the responsibility onto applications to adapt to the performance capabilities of the network. The protocol includes a control protocol component that reports to the sender received bandwidth at the receiver. Thus, if the network cannot support the sender's transmission rate, the sender reacts by sending at a lower rate. The protocol has the capability to gently probe the network at regular intervals in order to increase the sending rate when the network can support it.

This approach stands in contrast to the ATM model. The latter has a more static view of performance in terms of guarantees. Guarantees simplify the applications since they need not be written to be adaptive. But it places the onus on the network to police the behavior of the traffic so that the guarantees can be achieved. This puts more state into the network, which now requires set-up before use and which makes the network sensitive to failures.

A second critical advantage of the ISPN architecture is the ease with which new services can be introduced into the network, using the so-called "proxy architecture." Proxies are software intermediaries that provide useful services on behalf of clients while shielding servers from client heterogeneity.

E. Internet Telephony

A recent development has been the rise in the use of the Internet as an alternative infrastructure to exchange spoken or other telephone data, such as fax. Bypassing the traditional telephone infrastructure makes it possible to circumvent existing telephone tariff structures, thus achieving much lower calling costs.

The so-called "gateway architecture" even makes it possible to support existing handsets and local telephone access. A user calls a local number of a gateway server, that translates the analog voice in digital packets that are sent out on the Internet to another gateway in the country being called. This server dials the local number of called party, converting the packet voice back into analog signals. The process is reversed in the opposite direction.

The pricing for such services can be very attractive. A one-minute call from San Francisco to Germany using AT&T and Deutsche Telecomm costs approximately $1.20 per minute. The same call made through a gateway provider has been advertised as costing as little as $0.28 per minute.

An often-stated reservation about Internet telephony is that the audio quality is not good enough. However, there are trends at work that are likely to improve the quality Internet audio. The first will be the deployment of independent long-haul IP-networks by the telecommunications operators. Operators will interconnect with each other at well-determined access points. Such networks will be carefully sized to meet the needed delay and bandwidth demands. The second is Moore's Law. Faster hardware will help to reduce latencies in the gateway servers. Scalable processing architectures like networks of workstations (NOWs) will make these servers more scalable. The final trend is the emergence of new standards that will help improve the quality of Internet audio. These include RSVP, the H.323 coding scheme, new encoding strategies that allow lost packets to be reconstructed (for example, MPEG audio), and software implementations of voice coding that will operate at 8 kbps.

The final comment is that many leading telecommunications companies are already using the Internet to provide low cost, international services. Internet FAX services have been around for several years. And many local operators in the United States view the Internet as an inexpensive way to enter the long distance service market without a major capital investment.

V. Implications Beyond the Third Generation

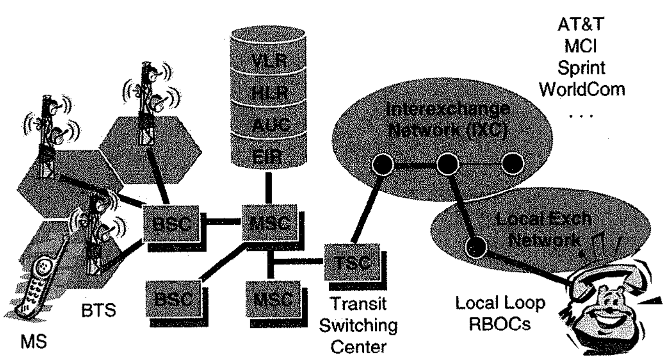
A. Introduction

The third generation is understood to provide universal multimedia information access with mobility spanning residences, businesses, public/pedestrian, mobile/vehicular, national, and global regions. Such systems are expected to support data rates in the wide-area in the range of 512 kbps, to 155 mbps. The first cellular systems were based on analog transmission techniques, such as AMPS in North America. The second generation is digital cellular, such as IS-54 (TDMA-based), IS-95 (CDMA-based), and GSM (an alternative TDMA system).

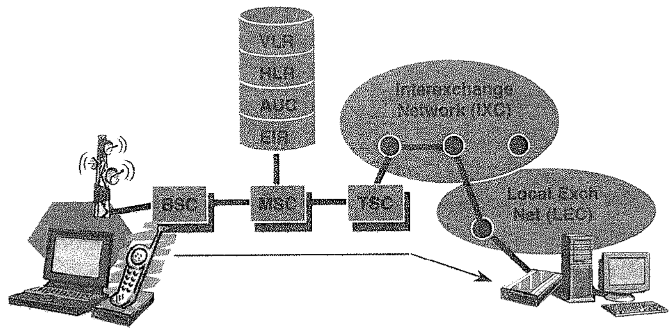
The details of the third generation are still being defined, but it is known that it will embrace multiple radio access technologies suitable for local, wide-area, satellite, etc. and able to achieve higher bandwidth than existing airlinks.

In the rest of this section, we will present one possible vision for how the GSM cellular and wireline telephony infrastructure might evolve into the "fourth generation," based on Internet technology.

B. The Cellular Telephone Network

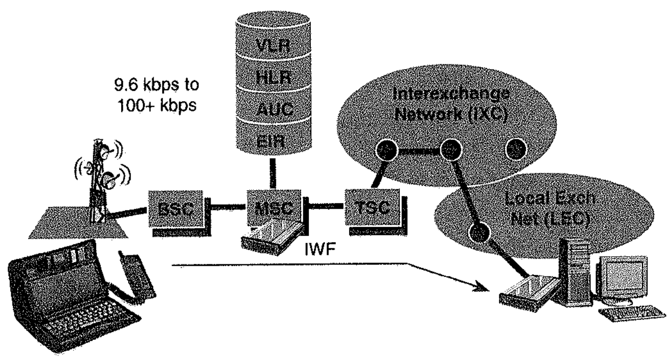


The GSM cellular telephone infrastructure consists of Base Station Controllers (BSC), Mobile Switching Centers (MSCs), the mobility and authentication databases (Visitor Location Register, VLR; Home Location Register, HLR; Authentication Center; AuC; and the Equipment Information Register, EIR), and the Transit Switching Center (TSC). The TSC provides the interconnection to the Interexchange Network (IXC) provided by the long distance carriers. A local exchange network (LEC) provides the connectivity between IXCs and a local user.



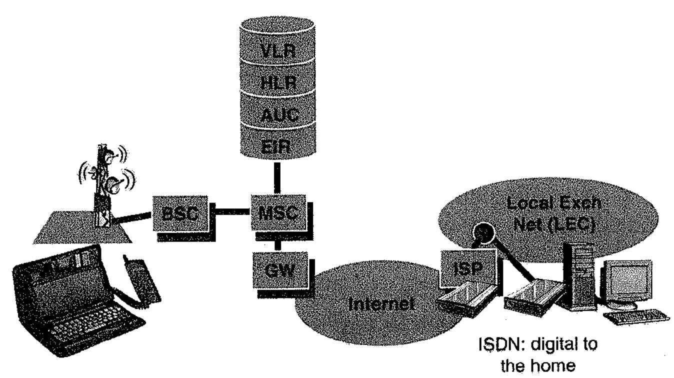
Now consider a data connection established across the cellular network. The originator has a laptop computer, a modem, and an analog cellular telephone. A connection is made from the laptop to the telephone network, to a modem at the dialed-up computer. In following this connection path, it is interesting to note the progression from digital to analog encodings, and back again. These conversions and translations have much to do with the high latencies and limited bandwidths seen on dial-up lines.

C. Data over Digital Cellular

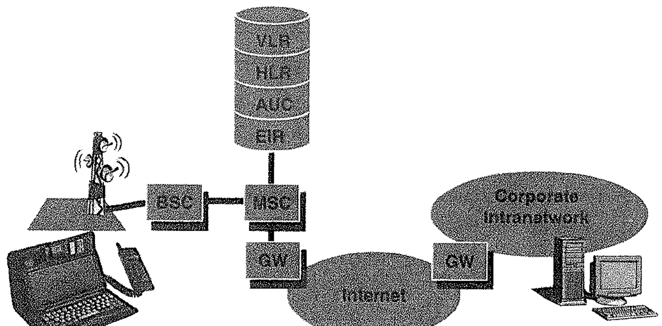


Associated with the MSC is a special function called the IWF, or Interworking Function. This is really nothing more than a modem, converting the digital stream back to analog for conversion and carriage via the IXC. The multiple conversions from digital to analog and back again occur once again. The only difference is the relocation of the modem from the end node into the mobile switching infrastructure.

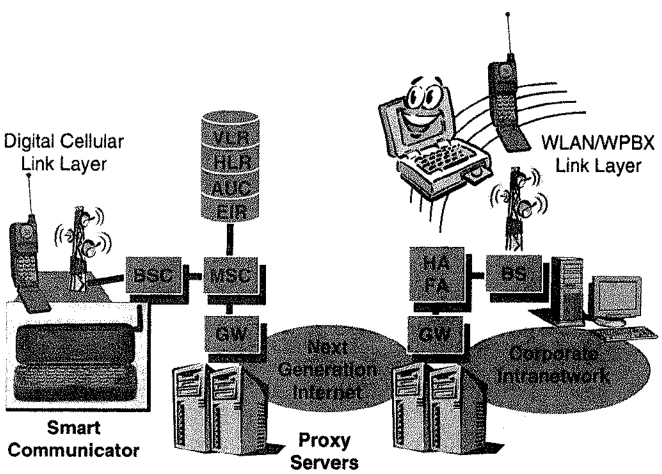
D. Internet-Based Architecture



Associated with the MSC is an IP-gateway, with direct connectivity to the Internet. Voice streams are routed through to the TSC and the IXC. Data streams are forwarded to this gateway, and sent out over the Internet. This has the significant advantage of converting a circuit-switched data call into a packet-switched transmission that bypasses the existing telephone network.



The rest of the path depends on the way in which the computer on the other end of the data call is connected to the Internet. If it is connected to an Internet Service Provider (ISP) via ISDN, then a local ISDN circuit between the ISP and the end computer is used to complete the path. If the end computer is on a corporate intranetwork with Internet connectivity, then all of the routing is done via the Internet, with no need for modem dial-ups of any kind.



In its ultimate form, we expect to see proxy servers associated with the gateways to the Internet. These could be used, for example, to packetize the GSM-encoded voice for ultimate delivery to a corporate user on her desktop computer attached to a corporate network. Such an architecture offers the possibility of complete integration between cellular handsets and laptops computers in the wide-area, and wireless PBXs and wireless LANs in the local area.

VI. Summary and Conclusions

Data access in telecommunications networks in general, and cellular telecommunications networks in particular, will accelerate in importance. Data applications will come to dominate. The Internet's emerging capabilities for real-time traffic, multipoint communications, and broadcast-based information dissemination make it a compelling technology to use for the next generation of multimedia systems.

The secret of the Internet's success is threefold. First, the network can be dumb because the intelligence is embedded in the end points. Moore's Law tells us that for a given cost, end points will get ever more powerful at an exponential rate. Second, the relatively simple and inexpensive infrastructure components that underlie Internet technology give a tremendous economic edge to IP-networks in comparison with existing telephone network technologies. And third, the easiest way to gain performance in IP-networks is by throwing bandwidth at the problem. This is a reasonable approach, given the trends towards ever increasing bandwidth in the wide area.

Finally, the economies of scale favor the Internet. Hundreds of thousands to millions of units of IP-infrastructure components are sold per year, compared to a few thousand for telephony network infrastructures.

These trends present a powerful case to rethink the architecture of the telecommunications infrastructure beyond the third generation.

Scanning the Literature

edited by **Bo Li**
boli@acm.org
Hong Kong University of Science and Technology

Local and global handovers for mobility management in wireless ATM networks

Marsan, M., Chiasserini, C-F., Lo Cigno R., Munafo, M. and Fumagalli, A.
IEEE Personal Communications, V. 4, N. 5, August 1997

Fixed path virtual circuit is one of the key technologies used in ATM networks, moving to a mobile wireless environment, this poses a challenging issue. This paper addresses the virtual circuit (VC) management in wireless ATM, i.e., how to handle the dynamic re-establishment of the ATM virtual circuits

with the short span of the mobile terminal (MT) handover from once cell to another. The goal is to ensure the in-sequence and loss-free delivery of the ATM cells in order to guarantee the QoS requirement of the ATM connections. A number of approaches have been proposed in literature, including 1) full establishment requiring a new VC setup each time a MT moves; 2) connection extension which extends the original VC from the old base station (BS) to the new BS that MT moves into; 3) incremental re-establishment that requires a partial establishment of a VC while reusing a portion of the original VC; and 4) multicast establishment approach which preallocates resources in the network portion surrounding the cell where the MT is located. Of them, the incremental re-establishment approach seems to be most appealing in that it avoids the potential resource waste in multicast establishment approach while also cutting down the portion of the VC which needs to be re-established. This paper presents a incremental re-establishment scheme, and examines a number of design issues, especially in-band signaling required and buffer management. It articulates that the incremental upgrade of the wireline ATM network to handle the mobile users is essential.