Chapter in *Encyclopedia of Multimedia Technology and Networking*, 2nd ed., M. Pagani (ed.), Idea Group Publishing, to appear 2009.

From Circuit Switched to IP-based Networks

Thomas M. Chen
Dept. of Electrical Engineering
SMU, Dallas, Texas

Introduction

Since the founding of the Bell Telephone System, the public switched telephone network (PSTN) has evolved into a highly successful, global telecommunications system. It is designed specifically for voice communications and provides a high quality of service and ease of use. It is supported by sophisticated operations systems that ensure extremely high dependability and availability. Over the past 100 years, it has been a showcase for communications engineering and led to groundbreaking new technologies (e.g., transistors, fiber optics).

Yet it is remarkable that many public carriers see their future in IP (Internet protocol) networks, namely the Internet. Of course, the Internet has also been highly successful, coinciding with the proliferation of personal computers. It has become ubiquitous for data applications such as the World Wide Web, e-mail, and peer-to-peer file sharing. While it is not surprising that the Internet is the future for data services, even voice services are transitioning to voice over Internet protocol (VoIP). This phenomenon bears closer examination, as a prime example explaining the success of the Internet as a universal communications platform. This chapter gives a historical development of the Internet and an overview of technical and non-technical reasons for the convergence of services.

Historical Perspective

The origins of circuit switching have been well documented (AT&T, 2006). A year after successfully patenting the telephone in 1876, Alexander Graham Bell with Gardiner Hubbard and

Thomas Sanders formed the Bell Telephone Company. In 1878, the first telephone exchange was opened in New Haven, Connecticut. For a long time, long distance switching was carried out by manual operators at a switchboard. Until the 1920s, the operator wrote down the number requested by the customer, and called the customer back when the other party was on the line. The route for a call was built link by link, by each operator passing the information to another operator who looked up the route for the call. Once a circuit was established, it was dedicated to that conversation for the duration of the call.

The national General Toll Switching Plan was put into effect in 1929. This established a hierarchical, national circuit switching network. Calls went to local offices connected to more than 2,000 toll offices and 140 primary centers, and up to 8 interconnected regional centers. Sectional centers were added in the 1950s. This hierarchical network, augmented with direct links between the busiest offices, continued up to the 1980s.

In the 1960s, transmission facilities were converted from analog to digital, such as the T-1 carrier (a time-division multiplexed system allowing 24 voice channels to share a 1.5-Mbps digital transmission link). Digital transmission offers the advantages of less interference and easier regeneration. Digital carriers used time division multiplexing (TDM) instead of frequency division multiplexing (FDM). In FDM, each call is filtered to 4 kHz and modulated to a different frequency to share a physical link. In TDM, sharing is done in the time domain rather than frequency domain. Time is divided into repeating frames. Each frame includes a number of fixed time slots. For example, 24 voice calls share the T-1 carrier which is 1.544 Mb/s. Each frame is 193 bits and takes 0.125 ms. In a frame, one bit is used for framing while the remainder is divided into 24 time slots of 8 bits each. The first time slot is used by the first voice call, the second time slot by the second call, and so on.

Automated electromechanical circuit switches appeared in the 1940s starting with a No. 4 crossbar switch in Philadelphia. These switches were able to understand operator-dialed routing codes or customer-dialed numbers to automatically route calls. In 1976, the first No. 4 ESS (electronic switching system) was installed in Chicago. Electronic switches were special-purpose computers with programmability.

Advances in digital circuits and computers not only improved the performance of telephone switches, but also changed the nature of network traffic. Starting in 1958, modems allowed computers to transmit digital data over voice-grade analog telephone circuits. In the 1960s, computers were still large and personal computers would not appear until the late 1970s. However, there was an evident need for computers to share data over distances. Digital data increased dramatically in the late 1970s with the proliferation of Ethernet local area networks and personal computers (e.g., Apple II in 1977, IBM PC in 1981).

The origins of the Internet began in 1969 with the ARPANET funded by the Advanced Research Projects Agency (ARPA, now DARPA). The ideas for a packet-switched computer network were supported by Lawrence Roberts and J. Licklider at ARPA, based on packet switching concepts promoted by Leonard Kleinrock and others (Leiner, et al., 1997). The first packet switches were made by Bolt Beranek and Newman (now BBN Technologies) and called IMPs (interface message processors). The first connected nodes were UCLA, Stanford Research Institute, University of California at Santa Barbara, and University of Utah.

The principles for IP along with a transport layer protocol called TCP (transmission control protocol) were conceived in 1974 by Vinton Cerf and Robert Kahn (1974). Cerf was experienced with the existing host-to-host protocol called NCP (network control protocol). TCP was initially envisioned to provide all transport layer services including both datagram and reliable connection-oriented delivery. However, the initial implementation of TCP consisted only of connection-oriented service, and it was decided to reorganize TCP into two separate protocols, TCP and IP (Clark, 1988). All information is carried in the common form of IP packets. The IP packet header mainly provides addressing to enable routers to forward packets to their proper destinations. IP was deliberately designed to be a best-effort protocol with no guarantees of packet delivery, in order to keep routers simple and stateless. IP can be used directly by applications that do not need error-free connection-oriented delivery. Above IP, TCP provides reliability (retransmissions of lost packets as needed) and flow control for applications that need perfectly reliable and sequential packet delivery. TCP/IP was adopted as a U.S. Department of Defense standard in 1980 as a way for other networks to internetwork with the ARPANET. Acceptance of TCP/IP was catalyzed by its implementation in BSD Unix.

The Internet may have remained an isolated network only for researchers except for two pivotal events. In 1992, the Internet was opened to commercial traffic, and later to individuals through Internet service providers. In 1993, the Mosaic browser introduced the public to the World Wide Web. The Web browser is a graphical interface that is far easier for most people to use than the command line used for earlier applications. The Web has become so popular that many people think of the Web as the Internet. Widely available Internet access and popular applications have driven data traffic growth much faster than voice traffic growth. Around 2000 or so, the volume of data traffic exceeded the volume of voice traffic.

Technical Differences in Circuit and Packet Switching

Circuit switching is characterized by the reservation of bandwidth for the duration of a call. This reservation involves an initial call establishment (set-up) phase using signaling messages, and the release of the bandwidth in a call termination (clear) phase at the end of a call. In modern digital networks, reserved bandwidth means periodically repeating time slots in time division multiplexed (TDM) links, as shown in Figure 1.

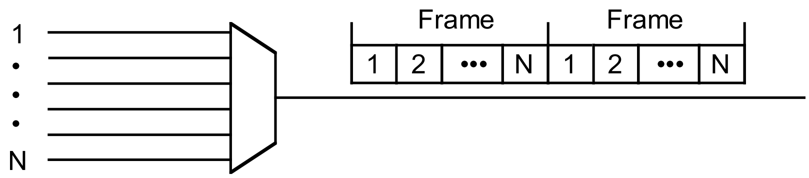

Fig. 1. Time division multiplexing.

Circuit switching works well for voice calls, for which it is designed. All voice is basically the same rate - 64 kb/s without compression, i.e., 8,000 samples/s and 8 bits/sample. TDM can easily handle bandwidth reservations of the same amount. Also, voice calls require minimal delay between the two parties because excessive delays interfere with interactivity. Circuit switching imposes only propagation delay through the network which is dependent only on the distance. Delay is considered one of the important quality of service (QoS) metrics. There is also minimal information loss, another QoS metric, because the reserved bandwidth avoids interference from other calls.

A number of drawbacks to circuit switching are typically cited for data. The first drawback is inefficiency when data is bursty (meaning intermittent data messages separated by variable periods of inactivity). The reserved bandwidth is wasted during the periods of inactivity. The inefficiency increases with the burstiness of traffic. The alternative is to disconnect the call during periods of inactivity. However, this approach is also inefficient because releasing and re-establishing a connection will involve many signaling messages, which are viewed as an overhead cost.

Another drawback cited for circuit switching is its inflexibility to accommodate data flows of different rates. Whereas voice flows share the same (uncompressed) rate of 64 kb/s, data applications have a broad range of different rates. TDM is designed for a single data rate and can not easily handle multiple data rates.

Asynchronous TDM or statistical multiplexing is an approach to improve the efficiency and flexibility of digital transmission, as shown in Figure 2. Instead of a rigid frame structure, time slots on a digital transmission link are occupied when there is data present in a FCFS (first-come-first-serve) order. It is possible to achieve a higher efficiency because time slots are not reserved. More flexibility is achieved because different (and variable) data rates can be accommodated. To resolve contention between simultaneous units of data, buffering is necessary. In practice, statistical multiplexing can not achieve total efficiency because it is necessary to label each slot to identify the connection using that time slot.

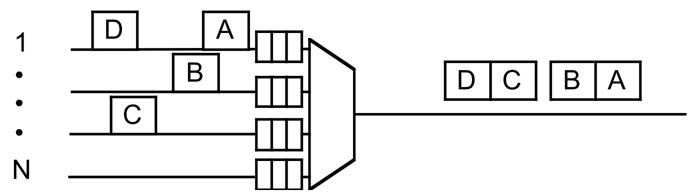

Fig. 2. Statistical multiplexing.

If the time slot label is viewed as a virtual circuit number, statistical multiplexing is the basis of connection-oriented packet switching. As shown in Figure 3, a virtual circuit number associates a flow entering a packet switch on a certain ingress port with an egress port. When packets are routed to the egress port, the incoming virtual circuit number is translated to a different outgoing virtual circuit number.

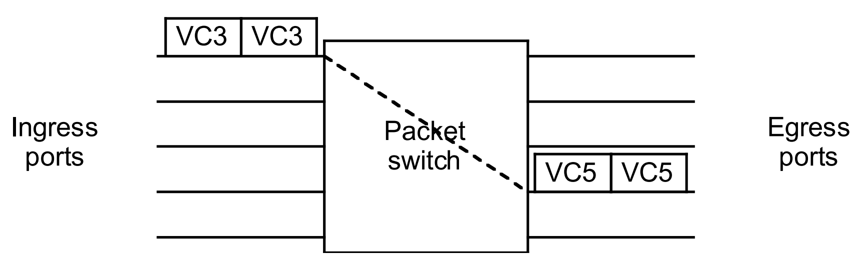

Fig. 3. Connection-oriented packet switching.

IP is connectionless and works a little differently. Instead of carrying a virtual circuit number, IP packets carry a destination address and source address. An IP router examines the destination address and decides on the egress port after consulting a routing table. The routing table lists the preferred egress port calculated from a routing algorithm (such as Dijkstra's algorithm) that usually selects the least cost route for each destination address. The great advantage of connectionless packet switching is the lack of state in the router. In connection-oriented packet switching, the switch must keep state or memory of every virtual circuit. A disruption of the switch will necessitate each virtual circuit to be reset, a process that will involve the hosts. In comparison, a disruption of an IP router will cause the dynamic routing protocol to adapt and find new routes. This will be done automatically within the network without involving the hosts.

IP offers a flexible and relatively simple method to accommodate a wide variety of data applications. Whenever data is ready, it is packetized and transmitted into the network. Routers are easy to deploy and dynamically figure out routes through standard routing protocols. Routers are simple by design, and costs have fallen dramatically with the growth of the Internet. It has been argued that the costs to switch packets has fallen below the costs of circuit switching (Roberts, 2000).

Packet switching may be more natural for data but involves its own challenges. One of the major challenges is quality of service. Statistical multiplexing handles contention by buffering, but buffering introduces random queueing delays and possible packet loss from buffer overflow. Thus, QoS in packet networks is vulnerable to congestion and degrades with increasing traffic load. A great deal of research in traffic control has highlighted a number of methods to assure QoS, including a resource reservation protocol (RSVP), weighted fair queueing (WFQ) scheduling, and differentiated services (Diffserv) (Jha and Hassan, 2002). However, there is disagreement in the community about the best approach, and consequently, support for QoS assurance has been slow to be implemented.

Philosophical Differences Between Internet and PSTN

From a perspective of design philosophies, the PSTN and Internet are a contrast in opposites. A few of the major differences are listed below.

- Circuits versus IP packets: As mentioned earlier, the basic building block of the telephone network is the circuit, a fixed route reserved for the duration of a call. The basic building block in the Internet is the IP packet (or datagram). Each IP packet is treated as an independent data unit.
- Reserved versus statistical: Related to the item above, the telephone network reserves a fixed amount of bandwidth for each call, thereby eliminating contention for resources and guaranteeing QoS. The Internet relies on statistical multiplexing. It is possible to congest the Internet which causes the entire network performance to suffer.
- Stateless versus stateful: By design, IP routers are stateless and keep no memory from one packet to another. Telephone switches must keep state about active calls.
- Homogeneous versus heterogeneous: The PSTN is a principal example of a homogeneous system. This is a result from the history of telephone networks run mostly by government regulated utilities. Also, there is a tradition of following comprehensive international standards. In contrast, the Internet designers recognized that computer networks would be heterogeneous. Different networks are owned and administered by different organizations but interoperate through TCP/IP.
- Specialized versus non-specialized: The PSTN is designed specifically and only for voice calls. The Internet is designed to be a general purpose network supporting a wide variety of applications.
- Functionality inside versus outside: One of the design philosophies behind the Internet is the end-to-end argument (Saltzer, Reed, and Clark, 1984). According to the argument, when

applications are built on top of a general purpose system, specific application-level functions should preferably not be built into the lower levels of the system. The IP layer takes a “lowest common denominator” approach. Applications add functions above the IP layer as needed. Intelligence and complexity are pushed out of the network and into the hosts. In comparison, telephone handsets are relatively simple, and telephone switches are complex and intelligent.

Transition to Voice over IP (VoIP)

VoIP is examined as an example of applications transitioning towards the Internet (Goode, 2002; Maresca, Zingirian, and Baglietto, 2004; Rizzetto and Catania, 1999; Varshney, Snow, McGivern, and Howard, 2002). Of all applications, voice might be expected to be the last to move to IP networks because it is served so well by the PSTN (Hassan, Nayandoro, and Atiquzzaman, 2000; Polyzois, et al., 1999). Examination of VoIP illustrates the forces behind the growing predominance of IP networking. Most of the advantages of VoIP over regular telephone calling is predicated on a couple of conditions that became true only recently: increasing broadband Internet access and ubiquitous computers with Internet connectivity.

- Lower costs: VoIP calls incur no additional costs above the monthly charge for a broadband Internet access. On the other hand, calls through the PSTN are billed by the duration of calls. Also, long distance and international calls are more expensive than local calls. VoIP calls are the same regardless of distance and duration.
- Little additional equipment: VoIP applications can be used in wireless phones or common PCs with a broadband Internet connection, sound card, speakers, and microphone.
- Abundant features: Common VoIP features include caller ID, contact lists, voicemail, fax, conference calling, call waiting, and call forwarding. Also, numerous advanced features are available.

- More than voice: IP can accommodate other media (video, images, text) as easily as voice. Also, VoIP applications may become integrated with e-mail or other computer applications.
- More efficient bandwidth utilization: More than 50 percent of voice conversations is silence. VoIP uses bandwidth only when a speaker is actively talking. Also, voice compression is easily done, increasing bandwidth efficiency.

Future Trends

VoIP is often highlighted as an example of communications convergence, that is, the migration of various different applications to a single common IP platform. Given broadband Internet access, the separate telephone service seems to be redundant and an unnecessary additional cost. Many people are envisioning that video services carried today through television networks will also migrate eventually to the Internet.

One of the ongoing challenges is QoS assurance. Researchers have investigated many potential solutions such as resource reservation protocols, differentiated services, and sophisticated packet scheduling algorithms (Chen, in press). However, universal agreement on the best methods is lacking, and it is practically difficult to evolve a large complex system like the Internet when different pieces are owned by separate organizations without central coordination.

A broader question is whether the current Internet needs a revolutionary change. The end-to-end argument has been evidently successful so far, but the Internet is being pushed by more demanding applications that can not be satisfied by simply more bandwidth and faster routers. These demands are causing Internet designers to rethink the original design principles (Blumenthal and Clark, 2001).

Conclusion

VoIP is often cited as an example of the migration from circuit switching to IP networks. Convergence of voice and video to the Internet will happen eventually for the reasons outlined in this article, but a number of obstacles must be overcome (Chong and Matthews, 2004). First, the existing PSTN is embedded and familiar to everyone. Many people are unfamiliar with VoIP. Also, today the QoS and reliability of the PSTN is unmatched. VoIP may offer savings and features, but they can not make up for the lack of QoS and reliability that people have come to expect.

References

AT&T. (2006) AT&T history: history of the AT&T network. Retrieved November 5, 2006, from http://www.attt.com/history/nethistory/switching.html.

Blumenthal, M., and Clark, D. (2001). Rethinking the design of the Internet: the end-to-end arguments vs. the brave new world. *ACM Transactions on Internet Technology*, 1(1), 70-109.

Cerf, V., and Kahn, R. (1974). A protocol for packet network interconnection. *IEEE Transactions on Communications Technology*, 5, 627-641.

Chen, T. (in press). Network traffic management. In H. Bigdoli (ed.), *Handbook of Computer Networks*. New York: John Wiley & Sons.

Chong, H., and Matthews, S. (2004). Comparative analysis of traditional telephone and voice-over-Internet protocol (VoIP) systems. In proc. of *2004 IEEE Int. Symp. on Electronics and the Environment*, 106-111.

Clark, D. (1988). The design philosophy of the DARPA Internet protocols. *Computer Communications Review*, 18(4), 106-114.

Goode, B. (2002). Voice over Internet protocol (VoIP). *Proceedings of the IEEE,* 90, 1495-1517.

Hassan, M., Nayandoro, A., and Atiquzzaman, M. (2000). Internet telephony: services, technical challenges, and products. *IEEE Communications Magazine*, 38, 96-103.

Jha, S., and Hassan, M. (2002). *Engineering Internet QoS*. Boston: Artech House.

Leiner, B., Cerf, V., Clark, D., Kahn, R., Kleinrock, L., Lynch, D., Postel, J., Roberts, L., and Wolff, S. (1997). The past and future history of the Internet. *Communications of the ACM*, 40(2), 102-108.

Maresca, M., Zingirian, and Baglietto, P. (2004). Internet protocol support for telephony. *Proceedings of the IEEE*, 92, 1463-1477.

Polyzois, C., Purdy, H., Yang, P-F., Shrader, D., Sinnreich, H., Menard, F., and Schulzrinne, H. (1999). From POTS to PANS: a commentary on the evolution to Internet telephony. *IEEE Internet Computing*, 3, 83-91.

Rizzetto, D., and Catania, C. (1999). A voice over IP architecture for integrated communications. *IEEE Internet Computing*, 3(3), 53-62.

Roberts, L. (2000). Beyond Moore's law: Internet growth trends. *IEEE Computer,* 33(1), 117-119.

Saltzer, J., Reed, D., and Clark, D. (1984). End-to-end arguments in system design. *ACM Transactions on Computer Systems*, 2(4), 277-288.

Varshney, U., Snow, A., McGivern, M., and Howard, C. (2002). Voice over IP. *Communications of the ACM*, 45, 89-96.

Key Terms

Circuit switching: a method of communication used in traditional telephony based on reserving an end-to-end route for the duration of a call.

Internet protocol (IP): the common protocol for packets in the global Internet.

Packet switching: a method of communication where data is encapsulated into separate routable units consisting of a packet header and data payload.

Public switched telephone network (PSTN): the traditional global telephone system.

Quality of service (QoS): the end-to-end network performance seen by an application, typically measured in terms of packet delay, delay variation, and loss probability.

Time division multiplexing (TDM): a method for multiple calls to share a single physical transmission link by allocating periodic time slots to each call.

Voice over IP (VoIP): the transmission of voice services over IP networks such as the global Internet.