CHAPTER

1

Introduction

The telecommunications, television, and information technology (IT) network industries are all transformed by the Internet. The transformation is driven by the need for growth based on new services, more complete global coverage, and consolidation. In this chapter, we will explore some of the problems and solutions for end users and every type of business because of the profound disruptions caused by the Internet.

Problem: Too Many Public Networks

Before the emergence of the Internet, users and service providers were generally accustomed to thinking in terms of four distinct network types: Networks for IT (data), networks for voice, mobile networks, and networks for television. Each of these dedicated network types could, in turn, be divided into many incompatible regional and even country-specific flavors with different protocol variants.

Thus, we find many types of telephony numbering plans, signaling, and audio encodings; several TV standards; and various types and flavors of what the telecom industry calls *data networks*—all of them incompatible and impossible to integrate into one single global network.

The mobile telephone networks have converged on a smaller number of standards in the second generation (2G) networks and in the emerging third generation (3G) mobile networks. It may turn out, however, that with the proliferation of new radio technologies for the so-called 4th generation (4G), such as Wi-Fi and WiMAX, all modern mobile networks will become just a wireless access mechanism to the Internet, where all public communications, entertainment, and applications will reside anyhow.

Data networks that originated in the telecom industry came in many forms, such as digital private lines, X.25, Integrated Services Digital Network (ISDN), Switched Multimegabit Data Service (SMDS), Frame Relay, and Asynchronous Transfer Mode (ATM) networks. These so-called data networks were mostly inspired by circuit-switched telephony concepts. Their names are meant to suggest that they were not designed primarily to carry voice.

Voice networks are still used for data and fax because of their general availability, though less and less so. However, these networks have come to the end of their evolution, since they are fundamentally optimized for voice only. TV networks were designed and optimized for the distribution of entertainment video streams.

Needless to say, all network types (data, voice, TV, and mobile) have specific end-user devices that cannot be ported to other service providers or network types, and most often cannot be globally deployed.

The impact of the Internet has made the wired and wireless phone companies and the TV cable companies look for new business models that can take advantage of Internet technologies and protocols, among them the Session Initiation Protocol (SIP) for real-time communications, such as Voice over IP (VoIP), instant messaging (IM), video, conferencing/collaboration, and others. Examples of the various categories and their business models are illustrated in Table 1.1. We assume that most readers are familiar with the acronyms used in the table, and we also explain these acronyms and terms in the book. They can also be found in the index.

Table 1.1 Internet Communications in 2005 with Examples from North America

CATEGORY	WHO	PROTOCOLS	STRENGTHS	WEAKNESSES
Open IM services with VoIP voice (competing islands)	Pulver FWD, Gizmo/ SIPphone, Damaka, Ineen	Standard SIP	Internet Presence Video User gets SIP URI On Net is free	Limited financing

CATEGORY	WHO	PROTOCOLS	STRENGTHS	WEAKNESSES
Closed IM islands with VoIP	Yahoo, MSN, Google, AOL, Skype (the most innovative)	SIP or other	Internet Presence Video On Net is free PSTN gateways	Nonstandard Walled gardens
PSTN over IP	Most "VoIP" companies	SIP	Internet Anywhere Video (Packet8) On net is free	Low-cost PSTN No new services Compete on price Costly infrastructure
Telephony over cable	TV cable companies	Everything from PSTN to MGCP to SIP with "P-" extensions	Broadband Internet Access to 80%+ households	Large investments in PSTN and older VoIP flavors
Wireless walled gardens	3G mobile operators	SIP for IMS with "P-" extensions	Strong financing	Central control inhibits innovation IP network cost
Wireline emulation of IMS: TISPAN	Wireline phone companies "NGN"	SIP with "P-" extensions		Duplicate IMS & NGN services

The proliferation of isolated communication islands as shown in Table 1.1 makes them less useful as their number keeps increasing (think of many more communication islands all over the world). Building communication islands (also called "walled gardens") is in conflict with Metcalfe's law that the value of the network increases by the square of the number of connected endpoints. Last, but not least, in case of an emergency, having many networks that cannot communicate directly is not very helpful.

Closed networks are an impediment for innovation, since innovators must work (technology and legal agreements) with every closed network separately to bring a new service or product to market. By contrast, the Internet extends the reach for new applications and services instantly to the whole world.

Another observation from Table 1.1 is that the strongest financing available is at present for closed networks (walled gardens), the ones that are most limited in reach and usefulness. This raises business issues and regulatory questions (what are the public interest obligations, if any?) that are beyond the scope of this book.

Incompatible Enterprise Communications

Enterprise communication systems are often an even greater mix of incompatible and disjoint systems and devices:

- Proprietary PBX and their phones. Phones from one PBX cannot be used by another.
- Instant messaging is a separate system from the PBX.
- Various IM systems don't talk to each other.
- Voice conferencing and web-based collaboration use yet other systems.

Maintaining various incompatible and nonintegrated proprietary enterprise systems is quite costly and reduces the overall productivity of the workforce.

Network Consolidation: The Internet

The Internet has benefited from a number of different fundamentals compared to legacy networks, such as the tremendous progress of computing technology and the open standard Internet protocols that define it. This progress can be attributed to the expertise of the research, academic, and engineering communities whose dedication to excellence and open collaboration on a global basis have surpassed the usual commercial pressure for time-to-market and competitive secrecy.

The result is an Internet that uses consistent protocols on a global basis, and is equally well suited to carry data, transactions, and real-time communications, such as instant messaging (IM), voice, video, and conferencing/collaboration. Actually, the Internet is the "dumb network," designed for any application, even those not yet invented. This is in stark contrast to the isolated "walled gardens" with central control of all services illustrated in Table 1.1.

Voice over IP

Although the Internet has quickly established itself as the preeminent network for data, commercial transactions, and audio-video distribution, the use of voice over the Internet has been slower to develop. This has less to do with the capability of the Internet to carry voice with equal or higher quality than the telephone network but rather with the complex nature of signaling in voice services, as you will see in Chapter 6, "SIP Overview."

There are various approaches for voice services over the Internet, based on different signaling and control design. Some examples include the following:

- Use *signaling* concepts from the telephone industry—H.323, MGCP, MEGACO/H.248.
- Use *control* concepts from the telephone industry—central control and softswitches.
- Use the Internet-centric *protocol*—Session Initiation Protocol (SIP), the topic of this book.

The movement from such concepts as telephony call models to discovery/rendezvous and session setup between any processes on any platform anywhere on the Internet is opening up completely new types of communication services.

The use of SIP for establishing voice, video, and data sessions places telephony as just another application on the Internet, using similar addressing, data types, software, protocols, and security as found, for example, on the World Wide Web or e-mail.

Separate networks for voice are no longer necessary, and this is of great consequence for all wired and wireless telephone companies.

Complete integration of voice with all other Internet services and applications probably provides the greatest opportunity for innovation. The open and distributed nature of this service and the "dumb" network model will empower many innovators, similar to what has happened with other industries on the Internet and the resulting online economy.

Most IM systems on the Internet already have voice and telephony capability as well, though if it is proprietary, they cannot intercommunicate without IM gateways, although IM gateways inevitably cannot translate all the features from one system to another. IM gateways are also transitory in nature, since any changes to a proprietary IM protocol may render the gateway close to useless. By contrast, SIP-based communications offer a global standards-based approach for interoperability for presence, IM, voice, and video, as we will show in the following chapters.

Presence—The Dial Tone for the Twenty-First Century?

Unsuccessful telephone calls are a serious drag on productivity and a source of frustration, since both parties waste time and talk to voicemail instead to each other. Also, the timing of the phone call may not be appropriate or not reach the called party in a suitable location. The advent of presence, so well-known from IM systems, can provide much more rich information before trying to make a call in the first place, compared to just hearing the dial tone. Another convenience of SIP and presence is that many contact addresses may reside beneath a buddy icon, so the caller need not to know or worry about picking the right phone number or URI. Presence may, therefore, replace the dial tone used in telephony for well over 100 years.

The Value Proposition of SIP

SIP is not just another protocol. SIP redefines communications and is impacting the telecom industry to a similar or greater degree than other industries. This has been recognized by all telecom service providers and their vendors for wired and wireless services, as well as by all IT vendors. Chapter 2 will provide an overview of how the Internet and SIP are redefining communications.

SIP Is Not a Miracle Protocol

As discussed in Chapter 2, "Internet Communications Enabled by SIP," SIP is not a miracle protocol and is not designed to do more than discover remote users and establish interactive communication sessions. SIP is not meant to ensure quality of service (QoS) all by itself or to transfer large amounts of data. It is not applicable for conference floor control. Neither is it meant to replace all known telephony features, many of which are caused by the limitations of circuit-switched voice or to the regulation of voice services. And such a list can go on.

Various other Internet protocols are better suited for other functions. As for legacy telephony, not all telephone network features lend themselves to replication on the Internet.

The Short History of SIP [1]

By 1996, the Internet Engineering Task Force (IETF) had already developed the basics for multimedia on the Internet (see Chapter 14, "SIP Conferencing") in the Multi-Party, Multimedia Working Group. Two proposals, the Simple Conference Invitation Protocol (SCIP) by Henning Schulzrinne and the Session Initiation Protocol (SIP) by Mark Handley, were announced and later merged to form Session Initiation Protocol. The new protocol also preserved the HTTP orientation from the initial SCIP proposal that later proved to be crucial to the merging of IP communications on the Internet.

Schulzrinne focused on the continuing development of SIP with the objective of "re-engineering the telephone system from ground up," an "opportunity that appears only once in 100 years," as we heard him argue at a time when few believed this was practical.

SIP was initially approved as RFC [2] number 2543 in the IETF in March 1999. Because of the tremendous interest and the increasing number of contributions to SIP, a separate SIP Working Group (WG) was formed in September 1999. The SIP for Instant Messaging and Presence Leveraging (SIMPLE) was formed in March 2001, followed by SIPPING for applications and their extensions in 2002. The specific needs of SIP developers and service providers have led to an increasing number of new working groups. This very large body of work attests both to the creativity of the Internet communications engineering community, and also to the vigor of the newly created industry.

We will shorten the narrative on the history of SIP by listing the related working groups (WG) in chronological order in Table 1.2. We have listed for simplicity the year of the first RFC published by the WG, though the WG was sometimes formed one to two years earlier. Years denote a new WG that has not yet produced any RFC.

Table 1.2 History of SIP-Related Working Groups

NAME	FIRST RFC	CHARTER
avt	1996	Real-time transmission of audio and video over UDP/IP: RTP
mmusic	1998	Internet conferencing and multimedia communications: SIP, SDP, RTSP
iptel	2000	Routing and call processing for IP telephony: TRIP, CPL, tel URI
sip	2000	Development of the SIP protocol: SIP methods, messages, events, URI

(continued)

Table 1-2 *(continued)*

NAME	FIRST RFC	CHARTER
enum	2000	DNS-based use of ITU-T E.164 telephone numbers
sipping	2002	Applications and extensions to SIP
simple	2004	Use of SIP for Instant Messaging (IM) and Presence
xcon	2005	Centralized conferences
behave	(2005)	Behavior for Network Address Translation (NAT) for use with SIP, RTP
ecrit	(2005)	Emergency communications (such as 911, 112)
p2psip	(2005)	Peer-to-peer SIP (not yet a formal WG)

The growth of SIP-related standards in the IETF is illustrated and discussed in Chapter 21, "Conclusions and Future Directions."

References in This Book

Because of the multiple developments on the Internet, SIP is being used in ever-more services, user software, and various user devices (such as in SIP phones, PCs, laptops, PDAs, and mobile phones). This is, in effect, a new industry and its participants keep making new contributions to the core SIP standards, mainly in the area of new services and new applications. This book reflects SIP developments up to and including the 64th IETF in November 2005.

We have included, by necessity, many Internet drafts that are designated *work in progress*, since they are the only reference source for this particular information. Some of these drafts may become standards by the time you are ready to use them; some may be a work in progress and have a higher version number than quoted as of this writing; and still others may be found only in an archive for *expired drafts*.

The SIP WG drafts that are work in progress can be found online at the IETF web site:

```
http://ietf.org/html.charters/sip-charter.html
```

Additional individual submissions and Internet drafts from other working groups can be found at the following site:

```
http://ietf.org/ID.html
```

SIP-related drafts that have *expired* (older than six months) can be found on several archives. As of this writing, following are some of the sites:

```
www.cs.columbia.edu/sip/drafts
www.softarmor.com/sipwg
```

Readers may also perform a web search, such as Google, for any IETF SIP-related topic or for any Internet draft or RFC.

Several books have been published on Internet multimedia, Voice over IP, and SIP, some of which are listed here. [3], [4], [5] They focus mainly on how SIP works. This book is less about explaining how SIP works, but rather what it does and the new communications and services it enables.

We have reproduced some of the exciting services and features discussed in the IETF SIP WG and its main offsprings, the SIPPING and SIMPLE Working Groups. Also included in our discussion are some drafts from Bird of Feather (BOF) sessions that have not even made it to an accepted WG charter, such as the peer-to-peer (P2P) SIP group. [6]

Many of these expired proposals may not develop into IETF standards for various reasons, but represent good work, often backed up by running code. The references to such expired Internet drafts are intended to make you aware of these ideas that may otherwise remain buried in an archive. Such references are clearly marked as expired, so as to distinguish them from accepted work in progress items of IETF WGs that are on the path toward acceptance as standards.

SIP Open Source Code and SIP Products

There is an ever-increasing amount of open source code for SIP, and it is increasing in quality. Most or many commercial SIP products are actually based on open source SIP code. An authoritative list of SIP open source code is available from the SIP Forum:

```
www.sipforum.org
```

The SIP Forum is also an excellent source for finding commercial SIP software and products for the enterprise, for consumer products, for service providers, tools, and others.

Excellent lists of SIP products are also maintained on the web sites of `Pulver.com` and `Ubiquity.com`:

```
www.pulver.com/products/sip
www.sipcenter.com
```

References for Telephony

We assume throughout this book some understanding of telephone services and of telecommunication protocols. There is a vast literature pool available on telephony and telecommunications. We refer you to *Newton's Telecommunications Dictionary* [7] to brush up on various terms that will be used in the following chapters.

Summary

This chapter has discussed some of the problems and solutions to the communications industry by the Internet, and also a brief history of the SIP protocol.

During the migration from circuit-switched telephony to IP-based communications, there are too many isolated wired and wireless communication networks, even though most (but not all) are converging on SIP. SIP has undergone a 10-year development as a standard and in implementations in the marketplace.

By adopting the Internet as *The Network* with wired and wireless access, and SIP as the standard protocol, rich global communications are taking shape.

The old dial-tone in telephony may well be replaced by *presence* information, and rich multimedia will replace the narrowband voice communications used in circuit-switched telephony.

References

[1] The authors would like to thank Professor Dr. Jörg Ott, co-chair of the SIP WG and early contributor to the MMUSIC WG for helping with data on SIP history.

[2] RFC stands for Request for Comments and many of them are Internet standards.

[3] *SIP: Understanding the Session Initiation Protocol*, 2nd Edition, by Alan B. Johnston, Artech House, 2003.

[4] *SIP Demystified* by Gonzalo Camarillo, McGraw-Hill, 2001.

[5] *SIP Beyond VoIP* by Henry Sinnreich, Alan B. Johnston, and Robert J. Sparks, VON Publishing, 2005. `www.vonmag.com/books`.

[6] See the web site for P2P SIP at `www.p2psip.org`.

[7] *Newton's Telecommunications Dictionary*, 17th edition by Harry Newton, CMP Books, March 2001.