

1

Introduction

1.1 The rebirth of VoIP

In his famous book *Crossing the Chiasm*, Geoffrey A. Moore [B1] explains why so many innovative companies fail to turn their early successes into solid market positions and recurrent revenues. In an immature market, early adopters are eager to test new products and services, and they are willing to accept minor imperfections. When the market starts to mature, however, these ‘beta stage’ products do not sell as well, and many executives are led to believe that they do not innovate enough: instead of completing the product, they push even more ‘beta products’ to the market. This is the wrong decision: the key to massive product adoption in a mature market is *the whole product* (i.e., a product acceptable not only to technology enthusiasts and visionaries—the early market, but also to pragmatists, conservatives and even skeptics—the mainstream market). The whole product is not a marginal enhancement compared with the first prototypes, very often it requires as much effort and a lot of time to carefully analyze, understand and leverage the feedback of early users.

In many ways, the whole VoIP industry has just ‘crossed the chiasm’.

In 1999–2000, VoIP was one of the most successful buzzwords of the telecom bubble era. Every start-up company had a ‘new service’ or a ‘killer application’ that would change the landscape of the telecommunication industry for ever. The technology was evolving so fast that protocols introduced in 1998 were called ‘obsolete’ a year later ... in fact, every manufacturer claimed that his technology was so much better than that of competitors that interoperability was impossible.

From 2001 to 2003, VoIP faced a very tough reality check by pragmatist and conservative service providers. Many enthusiast ‘next-generation’ service providers that had spent billions on immature products realized that this expense did not find a mass market for them; in fact, the cost of sales of many of the ‘killer services’ exceeded any foreseeable

revenue. In a depressing climate with a start-up failing every month and even large service providers filing for the now famous ‘chapter 11’, the buzz for VoIP quickly disappeared.

Today the VoIP chiasm is behind us. Quite a few manufacturers and service providers survived, and are ready to participate in one of the most massive and disruptive technology changes ever faced by the telecom industry.

1.2 Why beyond VoIP protocols?

The new multimedia services will need to be much more than mere technology demonstrators. In order to build a demonstrator, engineers need only focus on the functional aspects: select a protocol, make sure it has the right service primitives, and combine these primitives into the desired functionality. The companion reference to this book, *IP Telephony: Deploying Voice-over-IP Protocols*, focuses on such functional aspects, presenting a high-level overview of packet media transport technologies, details on all three major VoIP protocols (H.323, SIP, and MGCP), and specific strategies to design services in the context of public networks where endpoints cannot be trusted and can be behind firewalls.

As its title implies, *Beyond VoIP Protocols: Understanding Voice Technology and Networking Techniques for IP Telephony* provides a broad overview of all the additional issues that need to be solved in order to deploy a multimedia service.

1.2.1 Selecting a voice coder

In the lab, almost any voice coder can be selected: there is plenty of bandwidth and hardly any packet loss. In a real network, however, even with the massive deployment of DSL, there is often a need to carefully select a voice coder that fits in the available bandwidth and provides the desired level of service. This is not an obvious choice, and it is necessary to have a deeper understanding on the internals of each voice coder in order to understand how each candidate coder may react to packet loss, for instance. During the bubble, the VoIP industry has generated many ‘magic coders’ which are supposed to outperform any other codec: a deeper understanding of codec technology helps separating true innovations from naive tricks. Finally, as the new generation of multi-rate adaptive coders appears for use by the 3G networks, it is important to keep in mind the fundamental differences between wireless and wired networks in order to evaluate which of the many innovations of AMR or AMR-WB coders may lead to significant improvements for Internet-based multimedia applications.

Chapter 2 “Introduction to Speech Coding Techniques” provides the necessary background to efficiently evaluate the candidate coders for a network and make the best compromise.

1.2.2 Providing ‘toll quality’ . . . and more

The first service providers who massively adopted VoIP were prepaid card vendors. Unfortunately, many of these service providers bet on the fact that most of their potential clients

would focus only on price and would have no means of complaining or asking for a refund if the voice quality was not acceptable. VoIP also had a lot of success among international transit carriers and arbitrage houses, and here as well voice quality is often not a prime concern. If you travel abroad and try to reach your voicemail, but cannot dial your DTMF access code correctly, chances are that your current service provider uses a VoIP network for international calls and never checked whether DTMF tones could get through.

Such bad experiences unfortunately backfired and created a perception among first-tier service providers that VoIP did not work. Most first-tier service providers conducted experiments between 2000 and 2002 in order to assess the elasticity of user voice quality acceptance levels relative to price. These studies aimed at designing tiered voice offers, with cheap, low-quality calls and more expensive toll-quality calls. To the surprise of everyone, these studies showed that there was only a willingness to pay more than toll quality for very high-quality calls (wideband coders); on the other hand, if the toll quality was perceived to be significantly lower than toll quality, there was no willingness to pay at all.

The consequence is that all post-bubble VoIP networks will need to provide a voice quality guaranteed to be comparable with toll quality, or better. Beyond the intrinsic quality of the voice coder, detailed in Chapter 2, Chapter 3 ‘Voice Quality’ discusses in detail how to control the most important parameters influencing end-users’ perception of voice quality: delay and echo.

1.2.3 Controlling IP quality of service

Peer-to-peer applications killed the idea that over-provisioning could solve quality-of-service problems on the Internet. Now that almost every DSL user is attempting to download his wish list of 10 or more 700-MB ‘DivX’ videos, it can be taken for granted, on the contrary, that most DSL links are permanently congested. The situation is likely to become even worse as some peer-to-peer telephony applications begin to use very aggressive redundancy techniques in order to get an unfair share of the best effort bandwidth. If everyone keeps throwing more packets to the best effort Internet, it will soon become very difficult to use this class of service for many applications that require a minimum level of responsiveness. In fact, the ‘best effort’ class of service is no longer usable for real-time applications, like telephony or videoconferencing, which cannot recover from packet loss.

Chapter 4 ‘Quality of Service’ discusses these issues from multiple points of view. At a low level, it explains the PGPS theory that makes it possible to provide differentiated levels of quality of service over all packet networks and helps understand the old ‘IP against ATM’ battles. It then presents the ‘DiffServ’ framework that today provides a simple, yet effective, way of marking IP packets with a desired quality-of-service level, and eventually downgrades to lower quality-of-service levels packets that are outside the agreed service-level agreement. There is a lot that can be done with DiffServ, but it must be used carefully, and the chapter also gives some guidelines on which types of traffic can be aggregated within a given service level and how to improve ‘fairness’ in a given class of service (even the best effort class).

Nevertheless, Diffserv does have some limitations, and it is likely that in the long run service providers will need to implement more dynamic ways of managing the service level of packet streams generated by end-users. The IntServ framework was initially presented as a direct application of the PGPS theory, and as such it is very powerful but also difficult to scale. Chapter 4 also discusses how a mix of IntServ and Diffserv could provide a good compromise for the future, and describes the current DQoS framework for cable networks, which is probably very similar to the techniques that will be used in the future on all public IP networks.

1.2.4 Dimensioning the network

From a dimensioning point of view, packet multimedia networks based on IP are unique compared with traditional telephony networks based on time division multiplexing (TDM), but also compared with other packet-based networks (e.g., ATM). The difference with TDM is obvious: bandwidth is no longer needed during silence periods, which makes it possible, when aggregating multiple streams, to save up to 50% of the bandwidth that would have been necessary if all voice channels were transmitting continuously. Unfortunately, this gain must be mitigated by the fact that the technique used by virtually all IP applications to transport media streams, RTP, is very inefficient in itself. In most cases, discontinuous transmission gains will be compensated by the overheads of the IP transport, and in the end the average capacity required by an IP transport with simple voice coders is comparable with what would have been required on TDM. It is possible to achieve further gains by using low-bitrate coders, but this has an influence on end-to-end delay and voice quality, or in specific circumstances by optimizing the IP/UDP/RTP transport layers.

The difference from ATM networks is less obvious. As there was no ‘best effort’ traffic on ATM, networks required a very strict dimensioning in order to minimize the chances of rejecting a new connection if the admission control failed. As we have seen above, the explosive growth of best effort traffic makes it impossible to use this class for interactive streams; but once a separate class of service is created for voice or video-conferencing, the fact that most of the network capacity is used for best effort traffic makes it a lot easier to dimension the real-time class of service. Within reasonable limits, the real-time class can ‘eat’ the capacity used by best effort users who have no service-level agreement and, in theory, never complain. In fact, hybrid voice and data networks are not only simpler to dimension, but will provide lower end-to-end transmission delays for voice streams, compared with a pure voice network. Chapter 5 ‘Network Dimensioning’ presents the traditional techniques that must be used to dimension a network (e.g., the Erlang laws that evaluate the number of simultaneous active calls for a given number of users, or the Poisson laws that can be used to evaluate the processing capacity of softswitches required to handle VoIP signaling). Chapter 5 also discusses the characteristics of VoIP streams compared with TDM voice channels and explains how to extrapolate the results of traditional network dimensioning theory, based on constant bitrate streams, to ‘on-off’ streams typical of VoIP when used in combination with voice activity detection.

1.2.5 Unleashing the potential of multicast

The IP network was initially designed and optimized to provide robust point-to-point connections, using ‘unicast’ packets. Unfortunately, not all applications work well with point-to-point connections: all broadcast applications do not scale well if they need to duplicate the information stream for every listener. Today most commercial IP networks are still ‘unicast’-only, and for this reason when you connect to TV station websites, you get only a small, low-quality image even if you have a DSL connection at home. The reason is that the TV station still uses unicast and therefore needs to send a copy of the TV channel to everyone. Today MPEG2 requires about 2–3 Mbit/s for TV-quality transmission over IP: sending a single channel to 1 million ‘IP-TV’ sets would require no less than 3 Tbit/s at the TV station!

Today many service providers are enhancing their IP networks to provide support for ‘multicast’. This supports their ‘triple-play’ strategy in the residential market, which requires high-quality TV transmission over IP and enables many services in the corporate market (e.g. large videoconferences—the equivalent of webinars with a video stream).

Chapter 6 ‘Multicast’ explains how the technology can turn an IP backbone into a optimized broadcast medium, and discusses the numerous issues that have delayed the introduction of multicast on commercial networks and continue to limit the scope of feasible applications today. As multicast does have a specific behavior related to network sizing, Chapter 5 also presents the impact of various multicast distribution tree configurations on the required capacity of each link.

For the engineering department of service providers, Chapters 5 and 6 together will provide much of the material required when designing a ‘triple-play’ offer combining voice, IP-TV, and Internet access over DSL.

1.3 Scope of this book

Beyond VoIP Protocols is a companion reference to *IP Telephony: Deploying Voice-over-IP Protocols* (for more details see the last page of this book). Both books have been written with the goal of supporting those involved in the design and deployment of multimedia VoIP projects, and provide invaluable references for most of the required technology.

Our idea is that during the execution of a project, an engineer frequently needs to have a fairly accurate view of ‘the complete picture’ in order to avoid fundamental mistakes or misunderstandings, and he/she will usually only need a complete, exhaustive reference for a small fraction of the overall technology involved. In our professional lives, we all have spent an enormous amount of time compiling this ‘complete picture’, and we hope that these two books will significantly reduce the time needed to assimilate the essential information and avoid many of those errors that can be made through only being aware of half the picture. That said, our intention was to provide a complete overview and not every detail on each subject. For instance, when introducing the audio-coding techniques in Chapter 2, we do provide some background on the ‘Z transform’ in order to give a feel for the power of this technique and become capable of understanding the

codec design diagrams. This background obviously does not replace a complete book on the Z transform if your intention is to specialize in codec design and audio-processing algorithms. Similarly, if you do design an H.323, SIP, or MGCP device, you will need to actually read the parts of the standards that relate to your specific application, but *Beyond VoIP Protocols* will help you skip through 90% of the text, as you will already have enough background.

We structured our reference texts as two books, because most projects involve two phases:

In phase 1, *IP Telephony: Deploying Voice-over-IP Protocols* can help you combine the various protocols for the target service, and complements the standards by discussing the most common issues that may result from incomplete protocol implementations, or architectures optimized for private networks which fail in a public environment. This first book focuses on the functional aspects.

In phase 2, you will need to know how many users the application will serve, you will need to select or configure an IP distribution network, you will need to build a business model for the service. In so doing, you will have to answer questions like: ‘How many gateway ports do I need for 150,000 users?’ ‘Is 128 kbit/s upstream sufficient to support VoIP for my DSL users?’ ‘Can this DSLAM support 5,000 users with 10% placing a phone call and 20% watching IP-TV?’ ‘Do I need an IP DSLAM, or will two ATM VCs per DSL user be sufficient to provide support for VoIP and video?’ ‘What is the cost per user of this softswitch, which is sold per simultaneous call?’

Beyond VoIP Protocols will not directly answer 100% of these questions, because every deployment is a unique challenge, but it will provide some useful tools that can be used as building blocks to help formulate a complete deployment strategy. For instance, every question that relates a number of users to a number of ports, calls, or aggregate bandwidth ultimately boils down to an Erlang calculation. The answer to most quality of service-related questions is a properly designed set of differentiated service levels which must obey certain constraints (detailed in Chapter 4), such as similarity of aggregated streams, proper support by the layer 2 transport level, etc.

1.4 Intended audience

The intended audience for *Beyond VoIP Protocols* is:

- Network planning teams, in charge of buying transport capacity, who need to guarantee acceptable end-to-end transmission delays.
- RFI/RFP technical teams who want to evaluate IP access devices (e.g., DSLAMs or BASs). The proper support for per-stream quality of service becomes fundamental in such equipment.
- Technical support teams for marketing departments who want to evaluate the cost side of ‘triple-play’ business models, which depend on the required bandwidth at the access level, as well as the sizing of control softswitches and gateways.

- National telecom regulators who want to evaluate the impact of IP wholesale prices on the viability of competitive VoIP service providers or want to assess the credibility of the threat of ‘virtual service providers’ (e.g., Vonage in the US), using an existing DSL local loop to provide telephony services without a license. Regarding numbering issues, the book provides strong arguments in favor of dedicating specific number ranges to VoIP (the solution adopted in Japan and many other countries), in order to avoid IP to TDM to IP connections, which obviously exceed acceptable delay constraints and introduce unwanted codec tandeming.
- Telecom students who want to understand how to use classic telecom tools in the context of VoIP. The book also provides an overview of many of the active research areas related to multimedia over IP and can help select a ‘hot’ topic for a thesis.

All the chapters in *Beyond VoIP Protocols* are relatively independent. They have been ordered from the most voice-centric to the most network-centric, but the reader can skip the chapters they are not interested in. For instance, if you are comfortable with considering voice coders as ‘black boxes’, you may skip Chapter 2 (or read the higher level description in Chapter 1 of *IP Telephony: Deploying Voice-over-IP Protocols*).

1.5 Conclusion

As with every technical book, despite careful proof-reading, there may be errors, typing mistakes, or you may find that some important new development are missing. We welcome your feedback, which can be sent by email to book@netcentrex.net. As technology is constantly evolving, we welcome all your suggestions for inclusion of new topics in future editions of the book.

Any updated material as well as a log of typos (should these be any) will be maintained at www.netcentrex.net/book. We will also publish there additional documents posted by our readers if we feel they are of interest to the intended audience.

We hope that you will find *Beyond VoIP Protocols* useful and that you will be convinced that, with appropriate design and planning, the technology is now mature enough to support massive deployments.

1.6 References

[B1] G.A. Moore. *Crossing the Chiasm*, HARPER COLLINS, 1991 (ISBN 0-06-662 001-3).