VoIP As a Collaborative Tool on the Client PC

White Paper

January 2005

Revision 1.0

Document Number: 305460-001

† Hyper-Threading Technology requires a computer system with an Intel® Pentium® 4 processor supporting Hyper-Threading Technology and an HT Technology enabled chipset, BIOS and operating system. Performance will vary depending on the specific hardware and software you use. Seehttp://www.intel.com/info/hyperthreading/ for more information including details on which processors support HT Technology.

Contents

Revision History

Revision Number	Description	Revision Date
1.0	Initial Release.	December 2004

1 Introduction

1.1 Instant Teamwork: Collaboration on the PC

Instant Teamwork is a collaboration technology that lets any device, in any location, communicate and work with any other devices in various and potentially geographically diverse locations. This includes the platform of course – being able to share voice, data, and even video, but it also includes applications. Instant teamwork means integrating capabilities into major applications that allow for easy capture and summarization, integrated collaboration tools, and spontaneous collaboration, when needed, between devices.

As the PC plays a more and more crucial role in instant teamwork, VoIP has emerged as one of the central core capabilities. This allows for easy collaboration of voice, data, and, later, video. Corporations are becoming more and more geographically diversified entities, thus collaborative tools such as Voice over Internet Protocol can assist these highly dispersed groups of co-workers function as cohesive teams, no matter whether one's co-workers are 30 feet or 3000 miles away, and in several different directions. The huge cost savings promised by VoIP are most readily realized by the organization with multiple branch offices which have a preponderance of their calls being placed from one office to another.

Add to this that broadband subscribership is projected to grow from 188 million today in 2004 to 300 million in 2008, giving even the SMB segment the IP capabilities needed for the best VoIP experience. Enterprise VoIP deployments are seen to reach half of all new deployments by 2006, thus enabling improved business process improvements and reduced business operational costs.

This paper will look now at the various components that are needed to enable VoIP on a client PC, as well as the various standards and specifications that play a major role in enabling VoIP on the PC.

1.2 What Is VoIP?

Getting a clearer telephone connection has come a long way since you had to tap on the microphone portion of your handset to loosen the carbon crystals. Now the ubiquitous nature of the internet has opened the door to a change in telephony usage from the public switched telephone network (PSTN) to the use of an IP network to digitize and packetize voice transmissions – Voice over Internet Protocol.

The basic steps involved for VoIP are conversion of analog voice signals to digital signals, which are then compressed and translated into digital packets for transmission over the internet to a receiver, who decompresses and "depacketizes" the data into an analog signal – a voice in one's earpiece or speaker. Voice over Internet Protocol is emerging as a key ingredient in collaboration in the digital office. Thanks to the use of wideband codecs, these transmissions may be at higher levels of clarity than PSTN use.

1.3 Components Required

The basic requirement to enable VoIP on the PC is some sort of audio input/out device. There are a couple different routes one could take, either a handset or headset – a headphone/microphone combination, and tethered to the client or wireless.

1.3.1 Soft Phones

These are not really telephones, or telephone headsets. A soft phone is a software collaborative tool loaded onto one's PC, and is used to allow the PC to function as a telephone, and enables some of the desired telephonic features such as teleconferencing, multiple simultaneous call reception, and call forwarding. Most soft phones will be SIP compliant in that they will be defined as and will function as a SIP user agent, in other words, the soft phone will use SIP for call setup and for some of the control needs. There are soft phones which will work in a Microsoft Windows* OS environment, others for a Linux* OS environment. These often will enable other collaborative efforts such as address book maintenance, call history, and instant message transmission and reception; in some environments these may be sent between other soft phones, applications or devices. Some offerings also provide for, in addition to voice communication, integrated email/voicemail/fax handling.

1.3.2 Speaker/Microphone Pair

In an office environment where each individual may have his or her own office, the ecosystem is seen to be perfectly satisfied with a set of good stereo speakers and an array microphone. Most PC systems ship with good speakers which would easily cover the upper and lower ranges called for by wideband codecs, especially the speaker systems that come with a separate sub-woofer. The microphones, however, while normally capable of covering the high end, tend to bottom out at 100 Hz, 20 Hz to 50 Hz above the lower limit of the wideband codec definition.

It must be said, however, that in an office environment where there are either multiple people in one office or many people in a "cubicle farm" a more unobtrusive solution is minimally beneficial, or even required to maintain a proper working environment.

1.3.3 USB Handsets

These telephone handsets are connected to the host system via the USB port found on the front or back of a PC, with the driver software either native to the operating system or needing to be installed as provided by the headset vendor. Some vendors provide additional driver software that would be needed for integration of the keypad with a 3rd party soft phone. There are a number of vendors providing USB handsets, some VoIP vendors are giving less expensive (and thus low quality) USB handsets away as a promotional gift to new subscribers.

1.3.4 Digital Headsets

The digital headsets usually will connect to the host via USB (normally with a tethered 3-meter cable) and often will be found with an inline digital signal processor (DSP) which is used to increase the sound reproduction. Most of these will handle the upper frequency range required by wideband codecs. When used with wideband codecs, one needs to look at the lower range of frequencies covered by the headphones in that some will only go down to 80 Hz, while others will go as low as 20 Hz, thus being able to reproduce the vocal communication more clearly. However, the problem continues to be the microphones that often will only pick up sounds above 100 Hz or so, however most of these are noise-canceling microphones.

1.3.5 Analog Headsets

Most analog headsets connect to the host with a two-pronged connector, one for the microphone receptacle, one for the headset receptacle. On many PCs this will be on the front panel for desktop units or on the side for mobile units. Some, however, will only have connectors on the rear panel, thus making it harder to connect and disconnect the headset, especially from the desktop host system. Many of the lower-end products will tend to have a smaller spectrum of frequency coverage for both the headphones and the microphone (and thus do not come with the added inline DSP for increased signal cleanup as with the digital headsets). For example, some of the less expensive analog headsets will support 300 Hz to 3000 kHz, perfect for narrow band codecs, but unsuitable in a wideband environment. There are other higher-quality (more expensive) analog headsets which are rated for operation between 20 Hz and 20 kHz, perfect for wideband codec environments. Though many of these microphones are very sensitive to sound, most microphones again will bottom out around 100 Hz; very few are rated for lower frequencies.

1.3.6 SLIC Devices and Adapters

SLIC stands for Subscriber Line Interface Circuit, and describes a class of devices and adapters which enable standard POTS wired and wireless telephones to be connected to the PC. They differ from regular modem connections on the computer in that you plug the modem into the wall socket, while you plug the phone into the SLIC connector. The SLIC connector makes a telephone look like a regular speaker and microphone to the computer sound system. Often these products include on-chip ringing, integrated switching regulators and universal interfaces. The SLIC can be built into the PC, like a modem jack, or it can be external, connected to the PC via a USB connection. SLIC devices that connect directly to the Ethernet rather than to the PC are known as Analog Terminal Adapters, or just Terminal Adapters.

1.3.7 Headsets with Bluetooth** Wireless Technology

Bluetooth** wireless technology is becoming more and more prevalent in the digital office environment, and not surprisingly there are a number of headsets available that connect to the host PC system wirelessly via Bluetooth. These wireless headsets tend to be very lightweight (thus ergonomically friendly), with noise-canceling microphones which tend to reduce background noise, improving the call experience, and can permit the user to wander up to 30’ from the base during a call. Some headsets can allow for up to 8 hours of operation, however the bulk may only function for 3-4 hours before needing battery recharging. As with the connected headsets, careful attention needs to be paid when selecting a Bluetooth headset, the lower end frequency range will often not be covered by the microphone in the lower-end offerings.

1.3.8 IP PBX

Most modern PBXs allow some kind of dual mode operation; PBXs with a circuit switched backplane often have VoIP line cards available, and IP-PBXs can make use of media gateways to support legacy phones. PBXs are normally depreciated over seven (or more) years, so even though the 50/50 cross-over of circuit switched to VoIP technology in the enterprise occurs in 2005, there will be a substantial installed base of circuit switched phones for the next several years.

One of the benefits of VoIP over circuit switched technology is that the services offered by the phone system are implemented as software on a regular generic server. Since IP connectivity is ubiquitous this means that the IP PBX for any particular IP phone can be anywhere in the world. So service providers can offer cost effective PBX replacements without having to install any equipment on customer premises other than the phones. This service is normally called IP Centrex, after a similar offering in the circuit switched world. Corporate firewalls and NATs must be correctly provisioned to permit VoIP traffic to flow. This is one of the inhibitors to VoIP deployment, since corporate IT departments are sometimes reluctant to make these changes to their security systems.

1.3.9 Host PC System

When one is deploying VoIP-capable solutions, CPU selection becomes critical. The typical soft phone system requirements call for a minimum 400MHz CPU. This minimum is adequate to support single function devices like PDAs. When deployed in a productivity or collaboration environment, the VoIP application will be used in a simultaneous multi-tasking scenario, and the CPU minimum requirements should be considered incremental to the additional platform application requirements. In addition one must contemplate in a multi-tasking usage scenario that a processor supporting multi-threaded operation has the ideal horsepower to maintain a high quality experience. A key feature of PC-based soft phones is the ability to support multi-user conference calls. As the number of participants in a conference increases, the CPU requirement to support the additional users will scale, and CPU performance headroom to accommodate these additional users ought to be considered. The Intel® Pentium® 4 processor with HT Technology† based PC is an ideal choice for solutions looking to take full advantage of the multi-user collaboration applications made possible with PC VoIP. Note that when a conference has attendees from beyond the LAN, the number of participants may be limited by the bandwidth of the WAN connection.

1.3.10 Putting it All Together

There aren’t that many things that an OEM would need to assemble in order to have a VoIP-capable PC. The VoIP-capable environment would need an IP-PBX with the appropriate communications software included in the IP stack, a soft phone, and the audio device itself and its requisite drivers. Most of the hardware drivers for the headsets would tend to be native to the operating system, as it would see the headset or handset as a speaker and a microphone. In other environments, the physical hardware drivers for the headsets, handsets, speakers and/or microphones may need to be installed separately.

The SMB environment sometimes will not have need for an IP PBX as their people may connect directly to broadband. There are intelligent telephone adapters available to allow connection via either cable or DSL modems. The need for IP PBX in the small (and sometimes medium) business companies may vary from installation to installation.

As these components are assembled together, tested for compatibility and appropriate functionality, one then has as a result a powerful collaboration-capable environment which various hardware and software vendors could then use to develop, test, deploy, and support their collaborative productivity applications and products in the marketplace.

§

2 Standards and Specifications

The industry is in the midst of bolstering the methods used to handle data, video and voice traffic over the internet or over an intranet. There is significant movement toward addressing the issues surrounding network congestion, packet loss, reliability, and quality of sound for voice communication. What follows is a brief description of the major standards and specifications one must keep in mind while assembling the stacks for a VoIP-capable environment.

2.1 H.323

This is a real-time multimedia conferencing specification promulgated by the ITU. It is now generally accepted to be superseded by SIP (see below), a protocol with similar scope of applicability promulgated by the IETF. H.323 is an overall specification for voice, data, and video traffic, including requirements for call sequencing. It subsumes several related specifications, like H.225 and H.245 (see below). H.323 can be set up for audio only, audio and video, audio and data, and audio, video, and data communications. This basic stack definition specification for VoIP designs' strong point is compatibility. Any H.323 system can call any other H.323 system and it will simply work, even if one is audio-visual and the other is simply audio. In cases of mode mismatch, it will default to the lowest common denominator.

2.2 H.225

This specification is for call setup and breakdown, and communicates between the internet gateway's protocol translation section and the CP user's telephony PC. It describes how audio, video, data, and control information can be managed in an H.323 environment by communicating information about gateways, gatekeepers, and conferencing between two H.323 endpoints, and opens the call signal channel between an endpoint and the gatekeeper. As a part of H.323, it is destined to be superseded by SIP

2.3 H.245

This defines a number of call services, including the number of calls, the negotiation of which PC is the master, and which is the slave, the opening and closing of logical channels required for the transmission of media streams, flow control messages, and channel signaling. It also communicates the capability exchanges for this communication session. While H.225 handles call signaling, H.245 handles call management and control signaling in the RTP environment. It "competes" with SIP, both will not function correctly together. As a part of H.323, it is destined to be superseded by SIP.

2.4 G.711

G.711 is the international standard for encoding telephone audio on a 64 kbps channel as used in the PSTN networks. It is a pulse coded modulation (PCM) scheme operating at an 8 kHz sample rate, with 8 bits per sample. Theoretically an 8 kHz sample rate can encode voice signals between 0 and 4 kHz, but the PSTN network includes band-pass filtering to remove frequencies below 300 Hz and above 3,400 Hz. This narrow band audio codec has two main logarithmic algorithms for compression, the μ-law algorithm primarily used in the US, and the a-law algorithm used elsewhere in the world.

2.5 G.722

This specification defines a wideband digital audio codec which can achieve up to a 4:1 data reduction. It is capable of handling analog voice signals between 50 Hz and 7000 Hz instead of 300 Hz – 3400 Hz for traditional telephony, and is used for higher quality speech applications at a sampling rate of 16,000 samples per second and bit rates of 24, 32, and 40 kb/s. This codec allows for a more crisp delineation between numerous consonants which tend to be confused in a narrow-band environment (clearer delineation between "bale" and "vale" for example), thus improving the intelligibility and naturalness of speech significantly over narrowband codec environments.

2.6 G.722.1/G.722.2

The G.722.1 and G.722.2 are two more recent wideband codec specifications which also allow for an audio bandwidth of 50 Hz to 7 kHz, but operates at a different bit rate than the G.722 specification, and each utilizes a different coding/decoding scheme. Other differences are beyond the scope of this paper.

2.7 G.723.1

This narrowband codec specification was designed for voice and video conferencing, and can be optimized for real-time simultaneous encoding and decoding on some of the higher-end PCs. The G.723.1 codec samples analog input signals at 8 kHz. Because of the sample rate, because it compresses relatively aggressively, and because it is a relatively old codec it has a lower voice quality than some of the other codecs. Music or tones such as DTMF cannot be transmitted reliably with this CODEC, thus the use of G.711 or some out-of-band method.

2.8 G.729/G.729A

The G.729/G.729A specification is another narrow band voice codec that has been used in some VoIP applications. As a narrow band codec, it samples at 8 kHz, and operates on 16 bit per sample (same as G.723.1)

2.9 SIP

Session Initiation Protocol is a signaling protocol for the setting up, management, and breaking down of voice and video sessions across packet networks. It allows for multi-user sessions for voice and/or data transmission. The primary role of SIP is to allow a caller to find and ring the party or parties with whom the caller wishes to communicate. Unlike H.323, it has directory access protocol as part of its definition. Where H.323 is more LAN-minded, SIP was designed with the internet as its target; thus its syntax borrows heavily from HTTP and SMTP, so much so that even some of the HTTP (and SMTP) codes are re-used. This is not a session description protocol, however, thus does not provide for any sort of control of conference calls, but it can allow for them as it is designed to describe management of a session, not what sort of session is being managed. Being built as an RFC standard, it is designed to work with a number of other protocols (but not H.245), as a number of different protocols need to dovetail together – especially RTP, user authentication protocols, directory services protocols, and wideband (or narrow band) codecs. There are compatibility concerns however. For example a call between an audio-visual system and a pure audio system will default to audio in H.323, but will totally fail in SIP. Note that in a purely peer-to-peer environment, SIP is not required. In a SIP environment, one still requires a gateway to the PSTN and for proxy service; often these costs for connection to PSTN tend to be minimal, thus making this an attractive option for SMB. However, the trick to SIP is not the cost, but the knowledge of how to set up the firewall rules for the router required; this not an inconsequential task. Knowing this, many VoIP application providers who utilize SIP are offering guidelines and toolkits for putting the four or five firewall rules in place (without opening the barn door).

§

3 The Case for Wideband

For many decades now the upper frequency limit of digital telephony is generally around 3.3 kHz, and has been for some time, although with many long distance connections the network carries frequencies up to 2.7 kHz. On the bottom side, 220 Hz is a possibility in a pretty good connection, with 280 Hz to 300 Hz the norm. Normal FM radio and television audio bandwidth is generally from 30 Hz on the low side to 15 kHz on the upper bound. So what?

Intelligibility is the "so what." The sound of most consonants is carried in the higher frequencies, often at frequencies above the 3.3 kHz limit that defines the upper bounds of most narrowband codecs. When the teacher calls and says that your child is "sailing," you may be concerned because you might have heard, "failing." These two consonants, "S" and "F", are normally carried above 4 kHz, above the narrowband (and POTS) upper boundary, while in a wideband codec environment there would be little question of what was being communicated. Since the ability to determine what was stated decreases with a more constricted bandwidth, the use of wideband codecs can only improve the communication experience.

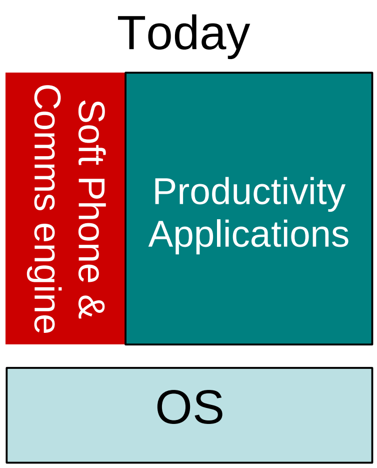
Within the business environment other factors compound the issue of speech intelligibility with narrowband. Most conferences are in rooms that echo quite nicely. Reverberation can be a greater problem with an increased distance between the speaker and the microphone. Studies have shown that increased bandwidth counteracts the problems of reverberation. Add to this the globalization of businesses; there is increased communication between people with widely disparate native languages, accents, and dialects. Comprehension of accented speech is far more difficult than speech native to both the speaker and listener(s). Some languages will even have trouble differentiating between consonants (substitution for "L" and "R" in some Asian languages, for example) or even local idiomatic phrases increase the difficulties.

The increase of bandwidth, and the requisite decrease of reverberation and noise, leads to the needed critical increased accuracy in speech communication, and probably decrease the fatigue factors in the listeners who are doing extra work just trying to figure out what was said, much less comprehend what was said. The suggested use of wideband codecs in a collaborative environment will obviously increase intelligibility, especially in today's business environment with its background noises, so that one could concentrate upon what was being communicated rather than upon what was being said; the added coverage of the upper and lower frequencies would cause the crispness and clarity communication experience to outshine the standard PSTN telephone call.

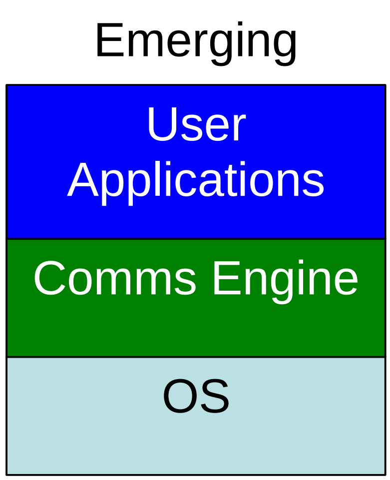
§

4 Moving Forward

As we move forward, voice quality issues are being addressed with the incorporation of wideband codecs. Interoperability concerns are being addressed at multiple levels, and issues are being resolved as software as well as PC and other hardware providers merge their efforts to provide solutions in this arena. VoIP-enabled PCs will, with increased performance, quality, and range of capabilities, enable businesses to concentrate on that which they do best, allowing the VoIP-enabled PC to be an ubiquitous, collaboration tool in the digital office. It will continue to enable standards-based business communication.



The higher-quality architectural elements are becoming more prevalent in the market. VoIP-capable handsets, headsets – both analog and Bluetooth – as well as speaker/microphone combinations are today on the market and ready to deliver communication at a better-than-POTS level. Specifications are pretty much settled out; hardware and software providers know what their products need to be compliant with. These elements may then be coupled with VoIP-enabled desktop and mobile PCs to empower businesses with affordable solutions and simplify the incorporation of VoIP and other communication functions into a seamless collaborative environment. However, as we see from the drawing on the left, today the ingredients are not necessarily a part of the whole.



Moving forward, as the next drawing illustrates, we will see these same ingredients becoming a seamless part of the communications engine. This engine would be composed of, but not limited to the codecs required, link(s) to the server, and some driver capabilities. Here is where opportunities lie for collaboration on the PC. In the previous example, the end-user has to go through the trouble of starting a separate application, while in this example, the user only needs to launch the collaboration tools from inside the primary application – the VoIP connections, in this case, simply take care of themselves, the process becomes very transparent and seamless.

Today, in some cases, and in the near future many collaborative capabilities are being drawn together seamlessly: physical presence detection, instant messaging, email, applications sharing, and multimedia presentations – some ideas include having voicemail be a .wav file in one’s email inbox, managing and maintaining a directory via a web interface rather than punching numbers on a keypad and so on. New video collaborative tools are beginning to be integrated into this mix; soon a collaborative audio/video/collaboration toolset will be a standard in the digital office. There is a growing sense of control over what, how, and when one may communicate, so that users may spend more time communicating and less time thinking about and managing that communication. To enable this, forward-looking reference designs are becoming more and more prevalent. Thus users, businesses, and service providers will derive through broad interoperability, innovative design and implementation an increased return on their investment in collaborative capabilities, providing an enterprise-class solution that merges with an existing PBX environment, drawing together desktops, laptops, PDAs, and others like Blackberry* devices. Though this is seen as a cost savings for, minimally, audio telephony, the possibilities for collaborative inventiveness are endless.

§

5 Glossary of Terms

Term	Definition
ATM	Asynchronous Transfer Mode – this is the universally accepted fast, cell-oriented transfer mode for most broadband integrated services digital networks. Packets are sent rapidly at a fixed length (53 byte cell units per packet), and are sent asynchronously relative to other related cells. It can handle voice, data, images, text, and video. It provides good bandwidth flexibility in both local area networks and wide area networks.
Circuit Switched	The technology used to transport calls on the PSTN. It allocates a fixed bandwidth across the network to a call for the duration of the call. It guarantees this bandwidth with a technique called "Time Division Multiplexing (TDM)," which logically divides a network connection into a collection of "time slots" which repeat 8,000 times per second. When all the time slots are allocated, no further connections are accepted on that connection.
DTMF	Dual-tone Multiple-Frequency – For each keypad on a standard touch-tone telephone there is a specific pair of tones assigned to be generated to facilitate switching within any digital PSTN.
GSTN	Global Switched Telephone Network or General Switched Telephone Network is the same as PSTN.
IP	Internet Protocol (although sometimes IP is used for "intellectual property") is the overall set of umbrella protocol definitions for data packets and addressing that are used for the internet.
ISDN	Integrated Services Digital Network consists of digital telephones and other data-transport services that regional telephone carriers may choose to provide. This involves the digitization of voice, data, text, graphics, music, and video so they may be transmitted over the standard telephone systems wires and cables.
Packet Switched	The technology used to transport data on the Internet, including Voice over IP. Compare to Circuit Switched.
PBX	Private Branch Exchange is the term for a private telephone system used within a medium to large sized company or enterprise. PBX users will share a number of lines for outgoing calls
PDA	Personal Data Assistants are handheld devices which combine various telephone, fax, we browsing, and calendar/personal organizer operations. Many will use a stylus for writing on a screen rather than a keyboard for data entry, although some now are enabled to respond to voice recognition technologies
POTS	The "Plain-Old Telephone Service" is synonymous with PSTN (see below).
PSTN	Public Switched Telephone Network is the world-wide interconnection of voice public telephone networks, both commercial and government owned. It is a conglomeration of circuit-switching telephone networks made up of primarily digital technology switching systems, although in some areas of the world the old-style banks of physical switches and relays still exist
RTCP	Real-time Control Protocol is based upon the periodic transmission of control packets to all those connected in a RTP session, and uses the same distribution mechanism as the data packets. Its primary function is to provide feedback regarding the quality of the transmissions.

Term	Definition
RTP	Real-Time Protocol, sometimes called “Real-Time Transport Protocol,” is a four-layer transport protocol which provides end-to-end delivery services for data with real-time characteristics, such as voice transmissions. It does not provide any mechanism to ensure delivery or provide quality of service guarantees, but it relies upon higher-level services and applications to provide these services.
SMB	This is small-to-medium business environments. These organizations often do not have their own in-house IT groups, but contract with a third party to provide these services
SIP	Session Initiation Protocol is discussed in the “Standards and Specifications” section above.
SONET	Synchronous Optical Network, a fiber-optic transmission system for high-speed digital traffic, is an international standard for high speed communication over fiber-optic networks. The SONET standard defines a hierarchy of interface rates that allow data streams at different rates to be multiplexed.
TDM	See Circuit Switched.

§