DECspin: A Networked Desktop Videoconferencing Application

1 Abstract

The Sound Picture Information Networks (SPIN) technology that is part of the DECspin Version 1.0 product takes digitized audio and video from desktop computers and distributes this data over a network to form real-time conferences. SPIN uses standard local and wide area data networks, adjusting to the various latency and bandwidth differences, and does not require a dedicated bandwidth allocation. A high-level SPIN protocol was developed to synchronize audio and video data and thus alleviate network congestion. SPIN performance on Digital's hardware and software platforms results in sound and pictures suitable for carrying on personal communications over a data network. The Society of Technical Communication chose the DECspin Version 1.0 application as a first-place recipient of the Distinguished Technical Communication Award in 1992.

2 Introduction

In late 1990, we began to design a software product that would allow people to see and hear one another from their desktop computers. The resulting DECspin Version 1.0 application takes digitized audio and video data from two to eight desktops and distributes this data over a network to form real-time conferences. The product name represents the four major communication elements that unite into one cohesive desktop application, namely, sound, picture, information, and networks. The overall technology is referred to as SPIN. This paper first presents an introduction to conferencing and gives a brief overview of the framework on which SPIN was developed. The paper then details SPIN's graphical user interface. Although the high-level protocol (which is the application layer of the International Organization for Standardization/Open Systems Interconnection [ISO/OSI] model) that SPIN uses to synchronize distributed audio and video is proprietary, a general discussion of how SPIN uses standard data networks for conferencing is presented. Performance data for DECspin Version 1.0 running on a DECstation 5000 Model 200 workstation with DECvideo and DECaudio hardware follows the discussion of network considerations. Finally, the paper summarizes the future direction of desktop conferencing.

3 Introduction to Conferencing

When the SPIN project started, standalone teleconferencing products were available but not for desktop computers. Typically, the products offered cost as much as $150,000, required scheduled conference rooms and operators, and needed leased telephone lines. These systems did not operate as part of a corporate computer data network but instead required

dedicated, switched 56-kilobit-per-second (kb/s), T1 (1.5-megabit-per-second [Mb/s]), and T3 (45-Mb/s) public telephone components in order to operate. Originally designed as two-way conference units, these teleconferencing products later included hardware to multiplex several equally equipped systems. In addition, the enhanced systems included custom logic to implement a hardware compressor/decompressor (codec) that reduced digital video data rates sufficiently to use leased telephone lines.

During the last several years, other conferencing systems have been demonstrated. The Pandora research project by Olivetti Research resulted in an excellent desk-to-desk conferencing system. Although the Pandora system was expensive per user and did not use existing network protocols, it did prove the viability of using a digital conferencing system from one's office and demonstrated the natural progression from room conferencing to office conferencing. This system served as a good example for our own emerging desktop model, DECspin Version 1.0.

Throughout this same period, several compression standards suitable for video capture and playback have evolved and been implemented. The Joint Photographic Experts Group (JPEG) industry-standard algorithm results in intraframe compression of frames of high-quality video (on the order of 25 to 1).[1,2] This algorithm is well suited for either single-frame capture or motion-frame capture of video information. This form of compression is most appropriate for real-time video capture and playback where low (i.e., frame-by-frame) latency is required.

The Motion Picture Experts Group (MPEG) standard results in interframe compression of motion video.[3] This algorithm is well suited for motion-frame capture of video because only the differences between successive frames are stored. Interframe compression is appropriate for video capture and playback where real-time low latency is not required.

The H.261 standard results in interframe compression of motion video that is most responsive to the demands placed on capturing live video for dissemination over low-bandwidth public telephone networks.[4] This compression is suitable for video capture and playback with reasonable latency but is not quite real-time in nature. H.261 is the standard used most in the teleconferencing systems on the market today.

Finally, the last few years have also witnessed the emergence of dramatic new base computer and network technologies. Reduced instruction set computer (RISC)-based workstations supply the needed processing power and I/O bandwidth to process large and continuous amounts of data, and fiber distributed data interface (FDDI) technology results in 100-Mb/s local area networks for the desktop. Consequently, the SPIN development project got under way to provide a novel and innovative software application that could take advantage of the powerful new systems and networks.

4 Overview of Underlying Hardware and Software

We came up with the SPIN project in response to the question: How can we communicate easily with graphics, video, and audio on the desktop as well as over both local and wide geographical area networks? Video help documentation, textual help, and audio help are used on the desktop to communicate how the application works. Sound, picture, graphics, and network elements are all woven together to provide better communication among conference participants.

Early in 1991, we received our first prototype of the DECvideo TURBOchannel frame buffer, which included the necessary hardware to input and capture an analog video signal, to digitize the signal, and to display the pixel information on the screen. The frame buffer was special in that it displayed 8-bit pseudocolor, 8-bit gray-scale, and 24-bit true-color graphical data simultaneously. This feature allowed captured video data to be displayed without data dithering.

Dithering is the process of converting each pixel of video data to a form that matches a limited number of available colormap entries. Most workstation frame buffers are 8-bit pseudocolor. Hence, digitized, 24-bit true-color video data for display would need pixel-by-pixel conversion. Algorithms exist that could be used to accomplish this conversion. However, a better SPIN conference, in terms of frame rate and picture quality, was achieved by performing no software dithering, thus relying on the ability of the DECvideo hardware to display 24-bit true-color video or 8-bit gray-scale video.[5] In addition, the DECvideo hardware could scale down the incoming video image in real time so that fewer pixels (i.e., less data) represented the original image.

Concurrently, SPIN used a DECaudio TURBOchannel card that could sample an input analog audio signal from a microphone and deliver an 8-kilohertz digitized audio bit stream. The DECaudio hardware could also convert a digital audio stream for output to an analog speaker or external amplifier. A DECstation 5000 Model 200 with DECaudio and DECvideo components provided the core hardware capability used in SPIN development work.

In addition to these new hardware capabilities, the SPIN effort needed new underlying base software capabilities. The DECvideo hardware required the Xv video extension to the X Window System to allow for the display and capture of video data. (The Xv extension was jointly developed by base system graphics and MIT Project Athena teams.) The DECaudio component used the AudioFile audio server, developed by Digital's Cambridge Research Laboratory, to capture and play back digital audio data.

A prototype software base was created to make fundamental measurements of video and audio data manipulation within the workstation and over a network. Testing the prototype over a 100-Mb/s FDDI network and a 10-Mb/s Ethernet network demonstrated that a conferencing product running over existing network protocols was possible.

5 The Spin Application

SPIN is a graphical multimedia communications tool that allows two to eight people to sit at their desktop computers and communicate both visually and audibly over a standard computer data network. The user interface employs a telephone-like "push" model that allows a user to place an audio-only, video-only, or audio-video call to another desktop computer user. Here, the term "push" means that SPIN conference participants control all aspects of the digitized data they send onto a network. Thus, users can feel confident about the security of their audio and video information. A caller initiates all calls to other users, and a call recipient must agree to accept an incoming SPIN call. Because all data is in the digital domain, this model makes it almost impossible to use SPIN to eavesdrop on another person. Placing a wiretap on a person's call would involve intercepting network packets, separating data from protocol layers, and then reassembling data into meaningful information. If the network data were encrypted, interception would be impossible. SPIN also provides other communication services, such as an audio-video answering machine, messaging, audio-video file creation, audio help, and audio-video documentation. Figure 1 shows a screen capture of a SPIN session in progress, using the DECspin Version 1.0 application.

NOTE

Figure 1 (Sample SPIN Session) is unavailable.

The product is easy to learn and to use. The graphical user interface is implemented on top of Motif software. Motif provides the framework for the SPIN international user interface. A model was chosen in which all actions taken by a user are implemented by push buttons that activate pop-up menus. The SPIN application does not use pull-down menus, because they require language-specific text strings to identify the purpose of an entry and thus require translation for different countries. Also, pull-down menus are intended for short-term interaction, and SPIN menus usually require more long-term interaction. All push-button icons are pictorial representations of the intended function. For example, the main window has a row of five push buttons, each of which activates a specific function of the application and is shown in Figure 1.

In the main window, the first button from the left contains a green circle with a vertical white bar, the international symbol for exit. This button appears in the same location in each of the pop-up windows. It is used to exit the window or, in the main window, to exit the application.

The second button from the left is labeled with the communication icon. This button is used to select the call list shown in Figure 2. The call list contains the various buttons and widgets used to place a call to another user, to create and play back SPIN files, and to display a list of received SPIN messages, if any exist. The list provides a way to play and manage audio-video answering machine messages. For example, to place a call to another user on the network requires just three steps.

1. Enter the computer network name of the machine and user into SPIN's phone database as "user@desktop." A string representing something more understandable to a novice is also allowed, e.g., "user@desktop1.dec.com" becomes "user@desktop1.dec.com Firstname Lastname at Digital Equipment Corporation."

2. Select whether the call is to be sound only, picture only, or both. The toggle push buttons under the large note icon control audio select; those under the large eye icon control video select. Once the call is established, these buttons can be set or unset by clicking a mouse or using a touch-screen monitor and are useful for muting the audio portion or freezing frames of the video portion.

3. To establish a two-way network connection, press the call push button under the connection icon (which is labeled with two arrows going in opposite directions) that appears next to the desired call recipient. If the person called is logged on, a ring dialog box appears on the call recipient's screen and a bell rings. If the call recipient is not available, a dialog box appears on the caller's screen asking whether the caller wishes to leave a message. The caller can then choose to leave a message or not.

   Depending on the individual settings, users can see and hear one another in multiple windows on the screen. To connect all conference participants in a mesh, press the "join" push button, which has a triangular icon.

NOTE

Figure 2 (SPIN Call List Pop-up Window) is unavailable.

Returning to the main window, the middle push button is the SPIN control button. As shown in Figure 3, the SPIN control pop-up window contains slide bars that, from top to bottom, allow the caller to set maximum capture frame rate, hue, color saturation, brightness, contrast, speaker output volume, and microphone pickup gain. At the bottom of the control window are buttons for selecting compression and rendering.

NOTE

Figure 3 (SPIN Control Pop-up Window) is unavailable.

To the right of the control button in the main window is the status icon button. Pressing this button causes the status pop-up window shown in Figure 4 to appear. The status window displays, below the camera icon, the active size of the captured video area in pixels. Beneath these dimensions is a vertical slide bar that indicates the average frames-per-second (frames/s) capture rate sampled over a five-second interval. To the right of the camera icon is the connection icon, under which appears the number of active connections. Below this number are the sound and picture icons, under which appear the number of active audio connections and the number of active video connections, respectively. The second slide bar indicates the result of sampling the average outgoing bandwidth consumption (measured in

Mb/s) of the application on the network. This measurement is also updated every five seconds.

NOTE

Figure 4 (SPIN Status Pop-up Window) is unavailable.

Finally, the fifth push button (on the far right) in the main window is the information button. By pressing this button and selecting the type of on-line information desired, the user can access the documentation pop-up windows, as illustrated in Figure 5. Within each documentation window are several topics and two columns of toggle push buttons that can be used to obtain either textual documentation or video documentation. The video documentation comprises short videos that contain expert help about the operation of the application.

NOTE

Figure 5 (SPIN Information Pop-up Windows) is unavailable.

As a final level of help, all push buttons and widgets within the application have associated audio tracks that tell the user what the buttons and widgets do within their context in the application. To activate the audio tracks, the user must first select the button or widget and then press the Help key on the keyboard.

6 Network Considerations

SPIN uses standard data networks to transport the information that composes a conference. Data networks are usually private networks that a user community maintains. Such networks often include a number of individual networks joined together by bridges and routers. Unlike public telephone networks, which are most frequently used for phone calls, private networks are used for a variety of computer data needs, including file transfers, remote logins, and remote file systems. However, telephone networks often provide the long-distance lines used to make up private wide area data networks.

The use of data networks allows conferencing data to be treated as would any other type of data. SPIN requires no special low-level networking protocols to transmit its data and uses the transmission control protocol /internet protocol (TCP/IP) or the DECnet protocol. Also, SPIN requires no changes to existing operating systems. When performing the prototype work for the SPIN application, we were not certain whether the real-time nature of conferencing could be accomplished on inherently non-real-time networks and operating systems. Consequently, we developed a special high-layer synchronization conferencing protocol, called the SPIN protocol, that uses existing data networks. This protocol is responsible for the synchronization of audio and video information. The SPIN protocol monitors the flow of data to the network in order to alleviate network congestion when detected. As the network becomes congested, the protocol makes the decision to withhold further video data, since video is the largest consumer of network bandwidth. This withholding of video data is a key feature of the SPIN protocol, because it allows a conference to vary the video frame rate on a user-by-user basis. Thus, video bandwidth can scale to the lesser of either the bandwidth available or the number of frames/s of video bandwidth that a given platform can sustain.

If the withholding of video corrects the network congestion, video data is once again allowed in the conference. If not, the SPIN protocol delays audio data and stores it in a buffer until the network is able to handle this data. If the network outage lasts approximately 10 seconds, audio data is lost. Periods of audio silence are used as a means of recovery from periods of network congestion. Thus, variable video frame rates along with this treatment of audio data allow for the graceful degradation of a conference as the network becomes busy.

SPIN has been demonstrated over a variety of public and private data networks including Ethernet (10 Mb/s), FDDI (100 Mb/s), T1 (1.5 Mb/s), T3 (45 Mb/s), cable television (10 Mb/s, more correctly, Ethernet running over two 6-megahertz cable television channels), switched multimegabit data service (SMDS) (1.5 or 45 Mb/s), asynchronous transfer mode (ATM) (150 Mb /s), and frame relay (1.5 or 45 Mb/s). Some of these networks are local or metropolitan area technologies, i.e., local area networks (LANs), whereas others are wide area technologies, i.e., wide area networks (WANs), as illustrated in Figure 6.

Each type of network provides SPIN with different latency and bandwidth characteristics. SPIN makes corresponding adjustments to a conference to account for these differences and does not require a dedicated bandwidth allocation to carry on a conference. If a given network supports bandwidth allocation, this feature only enhances SPIN's ability to deliver video and audio information.

WANs may use a router to interconnect two or more LANs. SPIN has been tested on a number of routers with mixed results, i.e., some routers correctly handle SPIN's bidirectional traffic pattern whereas others do not. Since some routers do not correctly handle bidirectional data traffic without packet loss, wide area routers must be individually tested with SPIN to verify proper operation. Some router problems were traced to the use of old firmware or software. Consequently, SPIN acted like a diagnostic tool in pointing out these problems. For example, running the SPIN application with audio only, across Digital's private IP network, yields varied results. Digital's IP network is an example of an open network, with routers from most router vendors. We traced most instances of poor SPIN performance to old or obsolete routers (some in service for the last six years without upgrades). These routers usually dropped packets when routing between adjacent Ethernet networks that were only 10 percent busy. After these routers were upgraded to the DECNIS family of routers, the SPIN application functioned correctly, even on congested networks.

To demonstrate daily use of SPIN, we created a metropolitan area network (MAN). Figure 7 shows the network topology, which spanned the states of New Hampshire and Massachusetts. The test bed allowed us to demonstrate our FDDI products, including end-station FDDI adapter cards, multimode FDDI wiring concentrators, and single-mode FDDI wiring concentrators. SPIN was used in 30 workstations, two of which were attached to large-screen projection units in conference rooms.

7 Performance

The conference quality achieved when running the SPIN application depends on many factors. The available network bandwidth, the processor speed, the desired frame-rate specification, the compression setting, the picture size, and how the pictures are rendered all affect the quality of the conference. Table 1 contains performance data for DECspin Version 1.0 at various combinations of settings for these factors.

As shown in Table 1, we tested SPIN performance using two basic picture sizes: 256 by 192 pixels and 160 by 120 pixels. The tests were performed over both Ethernet and FDDI networks for black-and-white and color cases. Also noted in the table is whether or not software compression was enabled for a specific test case. The far right column shows the frame rate achieved for the different combinations and also summarizes the network bandwidth consumed in each test. The table is presented primarily to give a sampling of the frame rate and, hence, the level of visual quality achieved for a specific combination of parameters. Frame rates affect an observer's ability to detect change within a sequence of frames. With a slow frame rate, the resulting video sequence may appear choppy and incomplete, whereas a normal frame rate (24 to 30 frames/s) leads to a smoothly varying video sequence with even continuity from one sequence to another. The frame rates in Table 1 below about 6 to 7 frames/s are considered low quality. Those in the 8-to-19-frames/s range are considered good quality, and those in the 20-to-30-frames/s range are high-quality video. The best cases in Table 1 are those that used software compression to deliver a pleasing frame rate with the least amount of network bandwidth consumed together with some degradation of individual frame quality. The software compression was tuned to provide nearly the same frame quality as the uncompressed case.

Table 1 also shows performance data measured using a DECNIS router. As noted earlier, wide area usage of SPIN depends on a router with correct algorithms for handling of bidirectional continuous stream traffic. The DECNIS family of routers can supply the full T1 bandwidth when presented with bidirectional SPIN traffic. Other routers on which SPIN was tested typically delivered only 25 to 50 percent of the T1 bandwidth. Note that this was only true on the particular routers we tested and that routers other than DECNIS routers may also be able to deliver full T1 bandwidth for this particular traffic pattern.

Hardware compression technology mentioned in the section Overview of Underlying Hardware and Software reduces the bandwidth requirements for conferencing. Experimentation with motion JPEG compression (using the Xv extension with compression functions on an Xvideo frame buffer board) has shown that at a resolution of 320 by 240 pixels, true-color frames can be used at 15 to 20 frames/s at a bit rate of just under 1.0 Mb/s. This bit rate produces a good- to high-quality conference with very low latency. H.261 and MPEG technology result in similar frame rates and picture size at about one-half the bandwidth but higher overall latency. Using motion JPEG as the example, high-quality conferences require about 1 Mb/s per connection. If all conferences are to be high quality, this bit rate allows 1 two-party conference on a T1 connection, 5 two-party conferences on an Ethernet segment, and 50 two-party conferences on an FDDI network. Using GIGASWITCH FDDI switches, more than 500 two-party conferences can take place simultaneously on a network. More users could be supported on T1, Ethernet, or GIGASWITCH networks, if lower quality conferences are acceptable.

8 Conclusion

It became clear during the development and deployment of SPIN that high cost per user limits the widespread use of the application. The cost of video for DECspin Version 1.0 adds about $8,000 to the price of a workstation. Audio for Version 1.0 adds about $2,000 per workstation. These costs, which are prohibitive to most potential users of the technology, do not include the network cost impact.

Digital's Alpha AXP family of computers come with audio input and output hardware as part of the base workstation. In spring 1993, Digital released to the Internet community a version of DECspin that uses this hardware to carry on audio-only conferences and shows the user a voice waveform instead of a video image. This version eliminates the add-on hardware cost for audioconferencing. A new low-cost video option would go far to reduce the add-on cost for video and facilitate a wider use of the SPIN application.

The SPIN application and its associated protocol have been demonstrated on Digital and non-Digital computers, operating systems, and networks. In particular, SPIN has been shown on SPARC workstations running Solaris software. Additionally, SPIN has been demonstrated on a personal computer using the Microsoft Multimedia Extensions (MME) to Windows software. This platform provides a very large user community of potential SPIN users and dramatically drops the price per user compared with the original product. Interoperability among platforms and a common user interface give Digital a leadership position in this fast-forming market.

Today, high-quality conferencing can scale to hundreds of seats on a LAN with lower-quality conferencing scaling to larger, more geographically dispersed networks. Several factors will lead to the widespread use of this technology: better and less-expensive hardware, programmable codecs, and higher-speed and less-costly cross-country networks. Less-expensive video hardware allows many users to upgrade their systems to include video, while programmable compression technology allows users to achieve improvements in picture quality, compression transcoding, and lower network needs. Higher-capacity and less-costly cross-country networks allow more users to access conferencing services. Ultimately, even homes will have better computer connectivity and bandwidth. As these changes occur, and we believe they will, desktop conferencing can become the interactive telephone of the twenty-first century.

9 Acknowledgments

The authors wish to acknowledge and thank all the members of the close team that worked on the SPIN project and made the DECspin Version 1.0 product a success. Key individuals in this effort were Diane LaPointe, Beverly Oliphant, Jonathan George, Garrett Van Siclen, and Jack Toto. We would also like to thank early supporters of the product efforts, including Jim Miller, Karl Pieper, and Jim Cocks. In addition, we extend our thanks to Walt Ronsicki, Videhi Mallela, Nathalie Rounds, and the rest of the team who established the FDDI test bed; to Dick Bergersen, who handled the quality assurance for DECspin Version 1.0 and gave the team excellent feedback on the product; and to Tom Levergood and the other members of Digital's Cambridge Research Laboratory who gave us support and assistance in regard to the AudioFile audio server. Finally, we would also like to offer thanks to our management, particularly Bill Hawe and John Morse, who are strong advocates and supporters of our product efforts.

10 References

1. Digital Compression and Coding of Continuous-Tone Still Images, Part 1, Requirements and Guidelines, ISO/IEC JTC1 Committee Draft 10918-1 (Geneva: International Organization for Standardization /International Electrochemical Commission, February 1991).

2. Digital Compression and Coding of Continuous-Tone Still Images, Part 2, Compliance Testing, ISO/IEC JTC1 Committee Draft 10918-2 (Geneva: International Organization for Standardization/International Electrochemical Commission, Summer 1991).

3. Coding of Moving Pictures and Associated Audio, Committee Draft Standard ISO 11172, ISO/MPEG 90/176 (Geneva: International Organization for Standardization, December 1990).

4. Video Codec for Audiovisual Services at Px64 Kb/s, CCITT Recommendation H.261, CDM XV-R 37-E (Geneva: International Telecommunications Union, Comité Consultatif Internationale de Télégraphique et TélTéphonique [CCITT], August 1990).

5. R. Ulichney, "Video Rendering," Digital Technical Journal, vol. 5, no. 2
   (Spring 1993, this issue): 9-18.

11 General References

L. Palmer and R. Palmer, "Desktop Meeting," LAN Magazine, vol. 6, no. 11 (November 1991).

TURBOchannel Hardware Specification (Palo Alto, CA: Digital Equipment Corporation, TRI/ADD Program, 1990).

Open Software Foundation, Inc., OSF/Motif, Programmer's Reference, Release 1.1 (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1991).

R. Scheifler, J. Gettys, and R. Newman, X Window System C Library and Protocol Reference (Bedford, MA: Digital Press, 1988).

12 Trademarks

13 Biographies

Lawrence G. Palmer Larry Palmer is a principal engineer with the Networks Engineering Architecture Group. He currently leads the DECspin project for the PC and has been with Digital since 1984. Larry is one of three software developers who initiated the PMAX software project for the DECstation 3100 product by porting the ULTRIX operating system to the MIPS architecture. He has a B.S. (highest honors, 1982) in chemistry from the University of Oklahoma and is a member of Phi Beta Kappa. He is co-inventor for five patents pending on enabling software technology for audio-video teleconferencing.

Ricky S. Palmer Ricky Palmer is a principal engineer with the Computer

Systems Group. He joined Digital in 1984 and currently leads the DECspin project. Ricky is one of three software developers who initiated the PMAX software project for the DECstation 3100 product by porting the ULTRIX operating system to the MIPS architecture. He has a B.S. (high honors,

1980) in physics, a B.S. (1980) in mathematics, and an M.S. (1982) in physics, all from the University of Oklahoma. He is co-inventor for five patents pending on enabling software technology for audio-video teleconferencing.