

VIRTUALTALKER: AN ON-LINE MULTIMEDIA SYSTEM ON TOKEN PASSING NETWORKS

Jau-Hsiung Huang, Wei-Hsin Tseng, Ming-Jeu Ding, Yo-Song Su, and Lin-Chih Wu
Communications and Multimedia Laboratory
Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan

Abstract - Although video phone/conference applications on ISDN networks have been a popular product, a similar service on packet switched networks still faces many technical problems. In this paper, the implementation of a multimedia system on a token ring network connecting personal computers using commercially available consumer electronic devices is presented. This system allows people to interactively communicate with each other at different locations through video phone and shared white board services. Since the system transmits real-time traffic such as video and voice over the network, many design and technical issues such as the network protocol, encoding technique, delay jitter, packet size, synchronization of video and voice, and priority assignment are considered and presented in this paper.

1. Introduction

When two people at different locations have to communicate with each other, what are the most natural features required? To answer this question, we have to think about how people communicate face to face. When people communicate face to face, they certainly can see and hear each other; further, they usually use pencil and paper to illustrate hand-written diagrams or drawings which are difficult to orally describe. Based on such an observation, the previous question is answered. That is, the

required features are (1) the capability of video phone, and (2) the capability to draw and to see and modify the other person's drawing.

With the increasing power of modern computers and high speed network technologies, it brings the integration of multiple media into a single network application [1-5] possible. In this paper, we will present an implementation of such a system, named *VirtualTalker*, on personal computers connected by a token ring network. Moreover, we will discuss the design issues regarding different characteristics of various media.

VirtualTalker basically consists of two applications. One is the *VideoPhone* application which transmits video and voice simultaneously and the other provides the telephone service with a shared white board by using a tablet. We call the second application as the *ScratchTalker*. The display of VideoPhone on the monitor is shown in Figure 1 which contains two windows with a resolution 128x128. The left window shows the video output of the remote speaker and the right window shows the video output of the local speaker. The right window is needed for the speaker to adjust his/her appearance on the

monitor screen. The display of ScratchTalker on the monitor is shown in Figure 2. In ScratchTalker application, a tablet is provided to allow the user to draw on the screen. Various colors and functions are provided to the user for drawing his pictures. Note that in ScratchTalker, the voice channel is still open such that the users can still talk to each other across the network. Furthermore, VideoPhone and ScratchTalker are switchable by simply hitting a *hotkey* on the keyboard.

The following scenario can be a case of using VirtualTalker. John calls Mary up through VideoPhone to discuss his design of Garfield. During their conversation, John tells Mary that he needs her opinion about his design of Garfield; then, John hits the hotkey to turn the system into ScratchTalker where John loads the file containing the design of Garfield as shown in Figure 3. Note that the voice channel between John and Mary is still open so that the conversation between them is not interrupted. In ScratchTalker, Mary can modify the design of Garfield on her side using her tablet and this modification will be shown on John's screen immediately. Using ScratchTalker, John and Mary have the impression that they are drawing on the same monitor display. At the end, they can hit the hotkey again and the system is turn back to the VideoPhone application to continue their conversation.

Obviously, VirtualTalker provides a very friendly and powerful communication tool for people on the network. Such a system is built upon a token passing network connecting several personal computers or workstations in order to provide a low cost solution. There are many considerations in implementing such a system on packet switched networks. One of

the most important issues is the delay jitter since it may seriously degrade the quality of video and audio. In addition, the choice of network protocols, synchronization among media, priority assignment of packets are also important issues and will be discussed in this paper.

Section two describes the system architecture and section three contains the functions provided by VirtualTalker. In section four we describe the design and technical considerations about VirtualTalker. Conclusions are given in Section five.

2. System Architecture of VirtualTalker

VirtualTalker is a multimedia system based on the PC and/or workstation environment. This system has been developed on token ring and FDDI network running TCP/IP protocol suite. We first describe the hardware required by VirtualTalker in order to satisfy various demands of such a multimedia application.

2.1. Hardware Aspect

First of all, since the video application of the VideoPhone generates a huge amount of video data to be carried by the network in real-time; hence, a video compression board is required in order to swiftly compress the video data. Normally, the compression ratio for video phone applications ranges between 20 and 200, depending on the encoding technology and video quality required. Currently, there are many compression standards for image and video applications such as JPEG [6], MPEG [7], and CCITT H.261 [8].

For the implementation of VirtualTalker, a low cost JPEG compression board, which can compress video signals on the fly (30 frames per second), is used to encode the video frames. This board also provides functions of A/D (Analog/Digital) and D/A conversion of video, i.e., the *frame grabber* capability. In the video applications of VirtualTalker, the compression ratio of JPEG is around 15 (30 on workstation), which is acceptable for our system. The voice is coded using μ -law PCM coding with 8K sampling rate. The coding and decoding of voice are executed simultaneously for full-duplex voice communication in VirtualTalker application. Lastly, a set of microphone and speaker is required as the voice I/O devices and one video camcorder is used as the video input device. In the graphics aspect, we use the *tablet* as the graphics input device which is connected to the personal computer through the RS-232C interface. The equipments of the system, except the network, are illustrated in Figure 4.

2.2. Software Aspect

VirtualTalker is developed using C and assembly languages under DOS on PCs and UNIX on workstations. In DOS operating system, this system is implemented with TSR (Terminate and Stay Resident) program. It usually *sleeps* in the *background* once started, and becomes active when the user presses a designated *hotkey*. In UNIX system, this system is implemented as a daemon which waits for connections in the background. Once there is someone who wants to connect to this program to run VirtualTalker application, the program is brought to the foreground immediately. When the VirtualTalker

application is closed, the program executed previously will be brought up again and is not affected at all. To reduce the overhead of CPU, recording and playback of audio are performed by programming the DSP (digital signal processor) and DMA (Direct Memory Access) chips on the audio board and the mainboard respectively.

After the implementation, the video frame rate is about 8 frames per second with color and the compression ratio of video traffic is about 15. In the following, we will calculate the number of participants allowed in the use of VideoPhone. Since the resolution of the video window is 128×128 with 16 bits for each pixel, hence the data rate of the video before JPEG compression is $128 \times 128 \times 16 \times 8 = 2$ Mbps. For a compression ratio at 15, the transmitted video data rate is about 133 Kbps. The voice data encoded by PCM will require a bandwidth of 64 Kbps, hence, the total bandwidth required by one person is 197 Kbps.

If the token ring network has a bandwidth of 16 Mbps and we use half of it, i.e., 8 Mbps, for the VideoPhone application, then the system allows 40 people (20 pairs) to use VideoPhone simultaneously. Since only half of the bandwidth is used, the impact of the delay on the real-time applications can be significantly reduced and the other half of the bandwidth can be used to transmit data files as long as priorities of packets are properly assigned. If FDDI is used to replace the token ring network, then a similar calculation shows that 250 people can use VideoPhone simultaneously.

3. Functions Provided by VirtualTalker

The functions of the VideoPhone application is clear since it is just a telephone with video capability to see each other. Hence, we will only explain the functions of ScratchTalker. The *pen* (i.e., the stylus) of the tablet has two modes, the *writing* mode and the *erasing* mode. When the bottom of the pen is held, it is used as an *eraser*, and the region it scans is filled with the *background* color (i.e., erased). On the other hand, if the bottom is not pressed, it is a *pen* and the *foreground* color is used for drawing. The size of the eraser is much larger than that of the pen such that less efforts are needed to erase a wide range.

Within the screen display of the ScratchTalker, there is an editable scratch region together with its control panel located at the left-hand side of the display as shown in Figures 2 and 3. The upper part of the control panel consists of some *function icons*, and the lower part is the *color palette*. In the following we will describe the features of the function icons.

- 1) *Quit*: To terminate the application.
- 2) *Clear*: To clear all drawings on the screen.
- 3) *Reset*: To reset and restart the application when undesired problems happen.
- 4) *Save*: To save the drawings on the screen as a bit-map file.
- 5) *Load*: To load a previously saved scratch file.
- 6) *Thin/Thick*: To select the thickness of the pen.
- 7) *Print*: To print the current drawings on the screen.

The colors for the foreground and the background are chosen from the color palette. The setting of background color will not change the background of the scratch board immediately, but we can use the erasing mode or the *Clear* icon to change it. Current foreground color and the background color are displayed at the bottom-left corner of the screen.

4. Design and Technical Considerations

Many design issues have to be taken into account for the implementation of VirtualTalker. The topics include determination of packet sizes, choice of transport protocols, synchronization between media, resolving delay jitter, priority assignment of packets and choice of encoding schemes.

4.1. Determination of Packet Sizes

In developing VirtualTalker, the real-time requirement of video and voice packets is an important issue. According to our experiments, system interrupts, waiting for packet acknowledgements, and accessing the video data consume most of the CPU time. Hence, a larger packet size should be used to reduce the total number of interrupts. However, a large packet will introduce a large delay for video and voice applications; hence, the packet size should be carefully chosen. After many trials, the size of voice packets is chosen to be 700 bytes and that of video packets is chosen to be 1000 bytes for an acceptable quality.

4.2. Choice of Transport Protocols

The network protocols used in the system is the TCP/IP protocol suite; hence, we can either choose TCP or UDP as the transport protocol. Before determining the network protocol to be used in the system, we should first discuss the characteristics of multimedia traffic. In [9], it was pointed out that 100% error free transmission of voice packets and video packets are not necessary as long as they are properly encoded. For example, by losing a PCM encoded voice packet or an intra-frame encoded video packet will not significantly affect the quality of the application. It was pointed out in [10,11] that a loss of voice packets up to 2% is acceptable to human ears.

In [12], it was shown that of the total processing time of TCP protocol in file transferring, more than 70% of the time is spent in data checksum computation. That is, data checksum is the processing bottleneck of TCP. Since there is no data checksum in UDP; hence, if the transmission error rate is small and error free transmission is not required as mentioned above, UDP is a better candidate than TCP.

Moreover, since TCP is a connection-oriented protocol such that all packets should be delivered in sequence and error-free; therefore, any unreceived packet (may be due to transmission error or buffer overflow) will be retransmitted. Retransmission of packets incurs two side effects: 1) consuming extra bandwidth and 2) incurring extra delay. Since multimedia traffic is very time-sensitive in one hand and may tolerate some packet loss in another, retransmission of packets may not be desired. In this regard, UDP is a better

candidate than TCP. Based on reasons stated above, UDP is chosen as the transport protocol for the system.

4.3. Synchronization Between Media and Delay Jitter Problems

For VideoPhone application, it is important to synchronize voice and video data during the conversation. Since the video frame rate equals 8, that is, we have to send the video data of one frame every $1/8$ second. During $1/8$ second, there will be 1000 bytes of voice data generated if the sampling rate equals 8 KHz. Hence, after sending the video data of one frame, the system will transmit all voice data corresponding to that video frame in order to synchronize voice and video data.

However, since the data volume of each video frame is different from that of another after JPEG encoding; hence, the transmission time of one video frame data is also different from that of another. This situation creates delay jitter problem for voice traffic. Another source of delay jitter comes from the non-deterministic network access time for packet switched network. Since voice traffic is very sensitive to delay jitter, hence, the quality of voice will be badly damaged if the delay jitter is serious. Therefore, to resolve problems of delay jitter is very important.

Ideally, voice is transmitted and received isochronously as shown in Figure 5(a). The residual time in the output buffer of the speaker at the receiving side is shown in Figure 5(b) where a packet will always come in time whenever the buffer is empty. However, in packet switched networks, packet arrivals usually suffer delay jitters such that the arriving pattern may look like Figure 6(a).

Traditionally, buffers are used to smooth the delay jitter problem. Nevertheless, using buffers will introduce extra delay to packets as shown in Figure 6(b), which is undesirable for real-time traffic. Moreover, Figure 6(b) shows that the delay will never be stripped once it is created. Hence, the size of the buffer should be carefully chosen to be as small as possible.

To resolve the above problem, two techniques can be applied. One is that whenever the receiver detects a received packet being a silence packet (i.e., no active sound included) by measuring the power of the sound, this packet is dropped without feeding into the output buffer of the speaker. The alternative method is that the sender will not transmit the packet if it is a silence packet. Figure 6(c) gives an example which shows that suppose packets 6 and 7 are silence packets, then the delay will be stripped after packet 7 comparing to Figure 6(b). Clearly, the effect of delay jitter is much improved in this case. In a conversation between two people, there will be many silent packets; hence, the above techniques work well in resolving problems of delay jitter.

4.4. Priority Assignment

Since voice is more delay sensitive than video and other media, hence voice should be assigned the highest priority with video second. Clearly, the real-time hand-drawn images have the third highest priority and data file transfer should be assigned the lowest priority. The capability of supporting priority is the reason why token ring network is chosen as the network platform. Such a design allows other people to transfer files without hurting the quality of voice or video. Hence, this network

can still support other general purpose applications while someone is using VirtualTalker application.

4.5. Choice of Encoding Schemes

Since UDP is used as the transport protocol as explained earlier, it is possible that video packets or voice packets may be lost without retransmission. This phenomenon has an impact on choosing the techniques of video coding and voice coding. We first examine the coding technique for voice. The most popular voice coding techniques may be PCM, DPCM, and ADPCM. Since both DPCM and ADPCM encode the difference between voice samples to gain a higher compression ratio, therefore a correlation between voice packets exists. The disadvantage of such schemes is that by losing one voice packet will affect all following voice data (called error propagation). Hence, the voice coding technique employed in VirtualTalker is simply PCM.

Similarly, among three standard video coding techniques, JPEG is an intra-frame encoding scheme which encodes the video data frame by frame without any correlation between frames. On the other hand, MPEG and H.261 are inter-frame encoding schemes which normally have a higher compression ratio but will create correlations between frames. Hence, MPEG and H.261 also have the problem of error propagation. Therefore, JPEG encoding technique is employed in VirtualTalker to avoid error propagation although the compression ratio is slightly lower than the other two schemes.

5. Conclusions

As the advances of video compression techniques and higher network speed, it is possible to build a multimedia network, transmitting video and voice, by using personal computers and inexpensive commercial products. Such a system can improve the quality of communication for people at different locations and hence increases their productivity. Moreover, such a system can also be used in a Computer Support Cooperative Work (CSCW) environment. This paper presents the VirtualTalker system which is implemented in our laboratory using off-the-shelf products. As the prices of frame grabber boards and compression chips keep on dropping, the price of such a multimedia network system will soon fall within an acceptable range; then, the era of multimedia networking will come.

References

- [1] J. B. Postel, G. G. Finn, A. R. Katz, and J. K. Reynolds, "An experimental multimedia mail system," *ACM Trans. Off. Inform. Syst.* vol. 6, no. 1, pp. 63-81, Jan. 1988.
- [2] C. Nicolaou, "An architecture for real-time multimedia communication systems," *IEEE J. Select Areas Commun.* vol. 8, no. 3, pp. 391-400, Apr. 1990.
- [3] P. V. Rangan, and D. C. Swinehart, "Software architecture for integration of video services in the etherphone system," *IEEE J. Select Areas Commun.* vol. 9, no. 9, pp. 1395-1404, Dec. 1991.
- [4] W. H. F. Leung, T. J. Baumgartner, Y. H. Hwang, M. J. Morgan, and S. C. Tu, "A software architecture for workstations supporting multimedia conferencing in packet switching networks," *IEEE J. Select Areas Commun.* vol. 8, no. 3, pp. 380-390, Apr. 1990.
- [5] M. Arango et. al., "The Touring Machine System," *Communications of the ACM*, Vol.36, No.1, pp.68-77, January 1993.
- [6] G.K. Wallace, "The JPEG Still Picture Compression Standard," *Communications of the ACM*, Vol.34, No.4, pp.30-44, Apr. 1991.
- [7] *MPEG standard draft ISO-IEC/JTC1 SC29 on 22*, November 1991.
- [8] 1989 CCITT Study Group XV, TD 35. Draft revision of recommendation H.261: Video codec for audiovisual services at px64 kbits/s. *Image Communication*, pp.221-239, August 1990.
- [9] Jau-Hsiung Huang and Shen-Horng Lee, "MHTP: A Multimedia High-Speed Transport Protocol," *Proc. IEEE Globecom'92*, December 1992, pp.1364-1368.
- [10] T. M. Chen and D. G. Messerschmitt, "Integrated voice/data switching," *IEEE Communications*, Vol.26, pp.16-26, June 1988.
- [11] N. S. Jayant and S.W. Christense, "Effects of Packet Losses in Waveform Coded Speech and Improvements Due to an Odd-Even Sample-Interpolation Procedure," *IEEE Trans. Communication*, Vol.COM-29, pp.101-109, February 1981.
- [12] Jau-Hsiung Huang and Chi-Wen Chen, "On Performance Measurements of TCP/IP

and its Device Driver," Proc. 17th Annual Local Computer Network Conference, Miniapolis, Sept. 1992, pp.568-575.



Figure 1: The monitor output of the VideoPhone.



Figure 3: The display of Garfield of ScratchTalker.

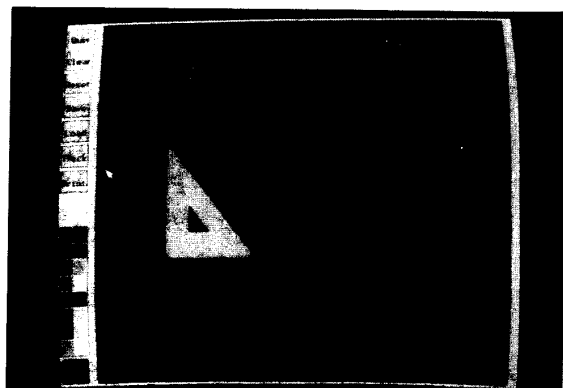


Figure 2: The monitor output of the ScratchTalker.

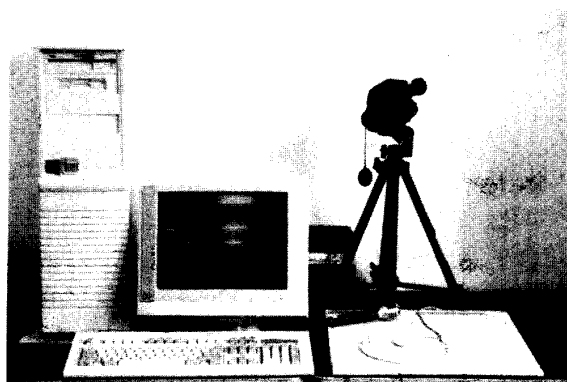
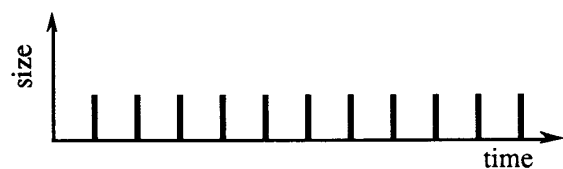
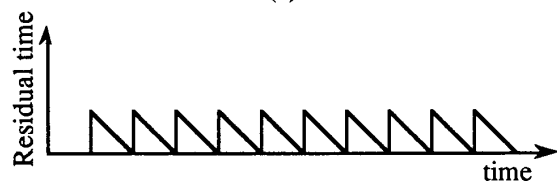


Figure 4: The equipment of each node of the VirtualTalker.

Biographies

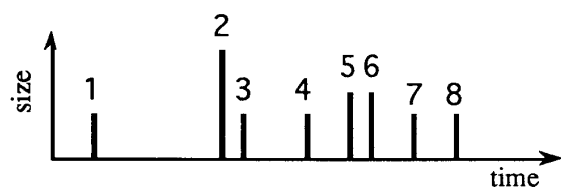


(a)

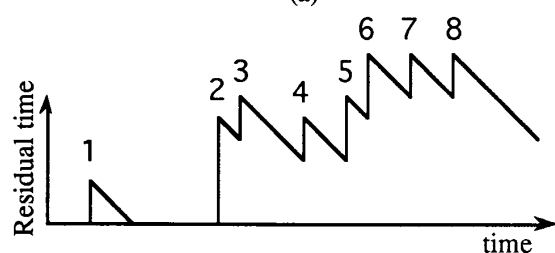


(b)

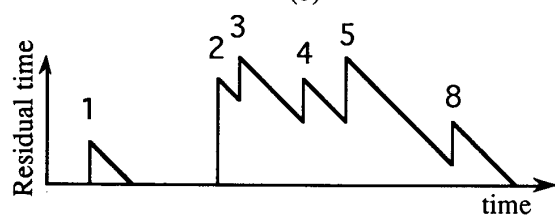
Figure 5: (a) The isochronous packets generated and received. (b) The residual time in the output buffer at the receiver.



(a)



(b)



(c)

Figure 6: (a) The packets with delay jitter. (b) The residual time if packets are buffered. (c) The residual time if silence packets are stripped.



Jau-Hsiung Huang received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1981 and the M.S. and Ph.D. degrees in computer science from the University of California, Los Angeles, Los Angeles, CA, U.S.A., in 1985 and 1988, respectively.

Since 1988, he has been a member of the faculty in the Department of Computer Science and Information Engineering (CSIE), National Taiwan University, where he is currently an associate professor. He has published over 40 technical papers in the areas of multimedia networking, high speed networking, parallel and distributed systems and performance evaluation of computing systems.



Wei-Hsin Tseng received his B.S. degree in CSIE from Tatung Institute of Technology, Taipei, Taiwan, in 1989 and M.S.

degree from National Taiwan University (NTU) in 1991. Since then, he has been a Ph.D. student in NTU. His current research topics include forward correction coding, admission control, and video modeling. He is an IEEE student member.



Ming-Jeu Ding received the B.S. degree in information science from National Chiao-Tung University, Hsin-Chu, Taiwan, R.O.C., in 1991. Since 1991, he has been a graduate student in the Department of CSIE, NTU. His current research work is focused on distributed systems and multimedia communication.



Yo-Song Su received his B.S. degree in CSIE from Tatung Institute of Technology, Taipei, Taiwan, in 1988. From 1988 to 1989 he was an Electronics Officer in the army. Since 1991, he has been a graduate student in the Department of CSIE, NTU. His current

research work is focused on design of the ATM switch and its performance evaluation.



Lin-Chih Wu received the B.S. degree in computer science & information engineering from National Taiwan University, Taipei, Taiwan, in 1991. Since 1991, he has been a graduate student in the Department of CSIE, NTU. His current research work is focused on performance evaluation, high-speed network protocols and the design of ATM switches.