PACKETIZED VOICE FOR SIMULATED COMMAND, CONTROL, AND COMMUNICATION

Thomas L. Gehl
International Business Machines Corporation
Manassas, Virginia

ABSTRACT

To improve its military readiness in today's budget environment, the Department of Defense needs advanced techniques that provide effective command, control, and communication (C3) training of its personnel with fewer resources. Government, industry, and academia are working to specify the distributed interactive simulation (DIS) environment, which consists of protocol data units (PDUs) that contain information about the simulated entities, a communication architecture that provides the necessary services for networked simulation, image generation databases to represent the physical environment, and performance measures for evaluation of both the simulation and training processes. This paper discusses an innovative method of integrating voice for command, control, and communication into the DIS environment. The radio communication and digitized voice characteristics that affect the C3 training architecture will be discussed. A packetized voice architecture will be proposed that provides functional radio capabilities such as (1) selecting channels of a radio communication device, (2) receiving and listening to multiple voices on one radio channel, (3) selecting filters to emulate the radio communication signatures, and (4) providing environmental effects on voice communications. The performance issues of prototyping the packetized voice architecture and a proposed DIS PDU for packetized voice will be presented.

INTRODUCTION

Historically, the emphasis of DoD simulators has been on training the student to functionally operate tactical equipment. The simulator contained the man-machine interface of the tactical device and processed responses to the functional selections of the device. For the most part, the training scenarios were the dynamic effects of peer participants in the tactical exercise. Since the individual simulators contained the tactical scenarios, few external connections to other simulators were needed. Replication of all of the tactical possibilities derived from human interaction was too costly and complex to provide realistic collective training on a large scale. Through the use of networking technologies, simulators can be linked together to provide the force-on-force engagements in a combined arms environment. C3 functions also need to be provided to further enhance the tactical training environment.

There has been much work recently, through the Workshops for the Interoperability of Defense Simulators, on draft standards for DIS. DIS is "an exercise involving the interconnection of a number of simulation devices in which the simulated entities are able to interact within a computer generated environment. The simulation devices may be present in one location, interconnected by a Local Area Network (LAN), or may be widely distributed on a Wide Area Network (WAN)."[1] The current DIS draft standard specifies the data structures or PDUs that are communicated over a LAN and WAN to multiple simulation applications to provide state information of the simulated world. This paper will discuss how distributed interactive simulation can be embraced in providing simulated communications for C3 training.

RADIO COMMUNICATION SYSTEMS

Communication is fundamental in enabling commanders to effectively command and control their forces in any battle environment. Communication not only enables commanders to convey orders that result from the decision making process but also provides commanders with the information they need to make good decisions. To provide effective communication for battle environments, operational communication systems must be survivable, flexible, reliable, secure, and interoperable. Capabilities that operational communication systems provide to the C3 process must be understood to create an effective combined arms training environment.

A general communication system consists of information source, transmitter, communication channel, receiver, and listener (Figure 1). The information source produces a message that is either written, voice, or formatted data. The transmitter converts the message into a signal format that is suitable for the communication system. The communication channel provides the medium over which the signal is transmitted. The receiver accepts the incoming signal and converts it back to the form of the original message and presents it to the listener. The full set of characteristics of a radio communication system must be understood and simulated correctly to ensure effective and realistic training for command, control, and communication.

Typical radio transmitters frequency modulate and amplify the incoming voice messages prior to the message being fed to an antenna. The method in which a particular radio modulates, amplifies, and transmits a signal creates a signature that is unique to that particular radio transmitter. The radio broadcasts the messages to all receivers on the same communication channel. The radio receivers must demodulate the incoming signal and tune to the signal frequency that corresponds to the radio channel selected by the listener. Like the transmitter, the receiver can add noise to the signal, depending on the frequency stability of the receiver. The receiver's ability to tune onto a signal is affected not only by the fidelity of its

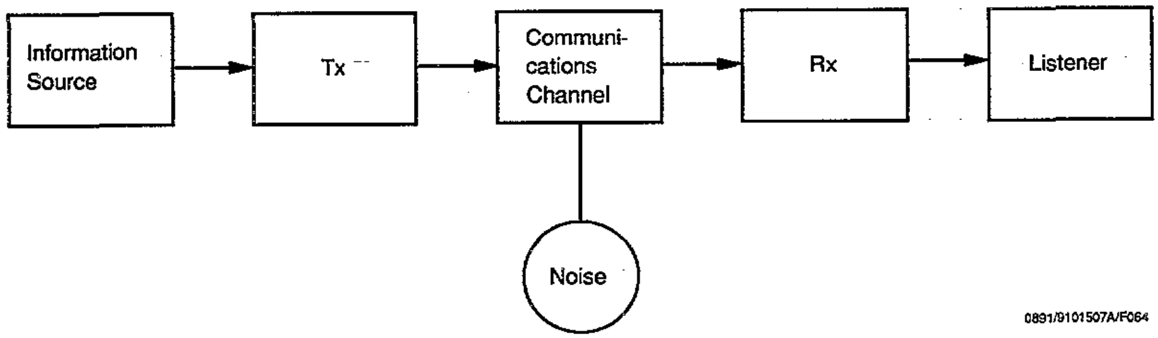

Figure 1. Communication System

tuner but also by the transmitted signal-to-noise ratio (SNR) and the propagation effects added to the signal. Receivers also provide automatic gain control that adapts to gradual changes of the signal strength from propagation attenuation, thus removing the need for continuous adjustment by the listener.

In simulating a radio communication system, we must take into consideration the communication medium's propagation effects in addition to the transmitter and receiver processing. The propagation effects of the communication medium are frequency dependent. Very High Frequency (VHF) radios, which are predominant in military applications, use a line-of-sight method of communication between the transmitter and receiver. For line-of-sight communication, the terrain and obstacles between the transmitter and receiver must be included in determining the propagation effects. Because of the frequency of VHF radios, the signal attenuation can be calculated by the spreading factor of 1/(4*PI*R**2), where R is the distance the signal travels. Most VHF radios are omni-directional, creating multi-path signals in addition to the line-of-sight signal. When a radio transmits a signal, the line-of-sight signal will combine with the multi-path signal at the receiver (Figure 2). The method in which the multi-path signal combines with the direct path signal is determined by the distances traveled by the two signals and the reflection amplitude and phase of the multi-path signal, which in turn depends on the dielectric constant of the ground.

The parameters identified in the preceding paragraphs all contribute to communications fidelity and are important in defining the architecture of the simulated communication system. Some of the parameters that are static, such as the fidelity of a particular radio, can be stored at the receive node and referenced by a radio identifier while other parameters, such as transmitter location, must be communicated with the voice information. Another consideration of the architecture is the difference between the operational and simulation communication mediums. The operational communication medium uses the atmosphere, while the simulation communication medium uses digital computer networks such as LANs.

CHARACTERISTICS OF PACKETIZED VOICE

The DIS environment is based on the interaction between simulation devices, through Local Area Networks (LANs) and Wide Area Networks (WANs), using the DIS PDUs to communicate state information of the simulated world. The DIS data is put into packets that are sent over LANs using the necessary communication protocols. One method of integrating the simulated radio communications between distributed simulation devices would be to use many of the ideas inherent to a DIS architecture. The position put forth here is to include a Voice PDU in the DIS PDU set and to utilize this new PDU to communicate packetized voice over the DIS network. In discussing how simulated communications would be provided, consideration must be given to not

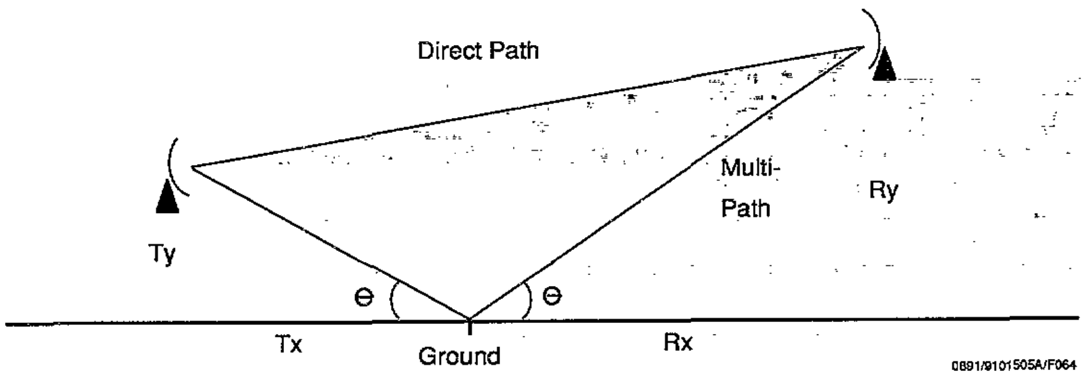

Figure 2. Multi-Path Effects of Phase and Amplitude

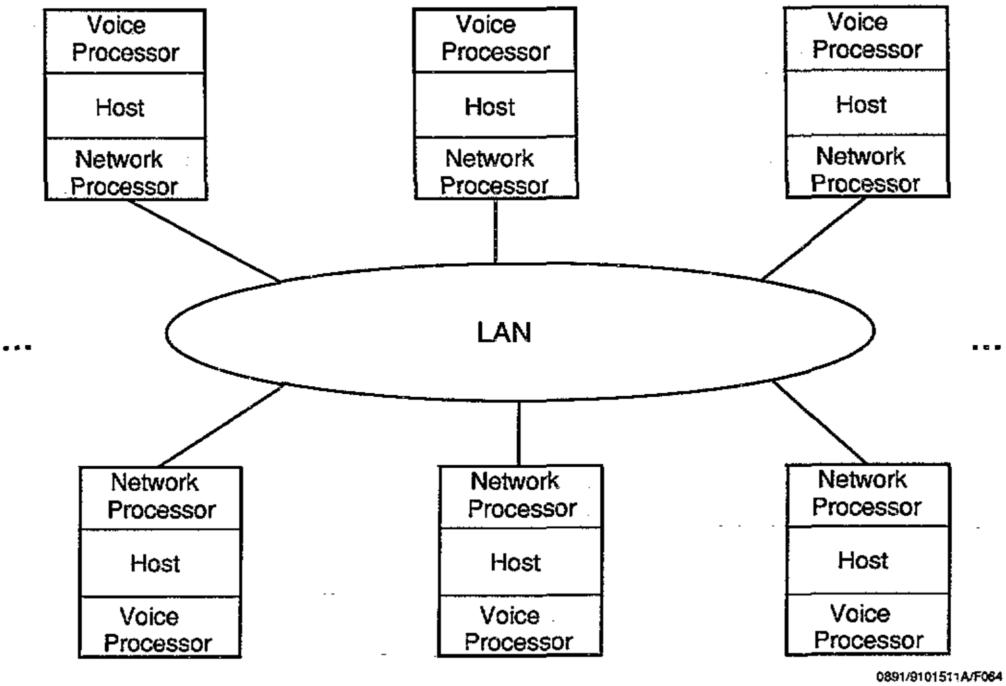

Figure 3. Packetized Voice Architecture

only the radio communication characteristics discussed earlier, but also to the packetized voice requirements.

To communicate voice information in LAN packets, we must first digitize the analog voice. Typical voice bandwidth is approximately 3.3 kHz. After adding sideguards to avoid interference between voice samples, the voice bandwidth is approximately 4 kHz. The Nyquist Theorem states that the sampling rate is required to be twice the signal bandwidth to capture all of the signal information. Thus, the required sampling rate for voice information is 8 kHz, which is the rate at which an analog to digital (A/D) converter must sample the analog voice information. The representation of the amplitude of each voice sample is determined by the number of bits used per sample. For telephone quality voice, 8 bits are used as the sample size. To provide 8 bits per sample for an 8 kHz sampling rate, the data rate for voice information must be 64 kbps. Recently, high-quality voice has been digitally communicated using 16 kbps of digital bandwidth. This reduction of digital voice bandwidth uses compressions algorithms that take advantage of redundancies in voiced vowels and consonants.

Many papers have been written recently on acceptable interspersion delays (jitter) between successive voice packets, the amount of voice information communicated per voice packet, and the end-to-end delay incurred by the voice packet. To keep the interpacket delays relatively consistent, there must be a certain level of synchronization between distributed voice nodes in communicating the voice packets. Acceptable interspersion delays between voice packets are approximately 20 milliseconds (ms). The variance in interpacket delays can be alleviated by using *play-out protocols* to smooth jitter between successive voice packets.[2] The amount of voice information per packet affects the allowable processing delay between packets and the rate at which voice packets are communicated. It is recommended that voice packets do not contain more than 50 ms of voice information per packet, and many designs involve about 20 ms of voice information per packet.[3] The overall loop delay or total delay betwen voice nodes is recommended not to be greater than 250 ms or it will affect the listener.

Digital voice does not require a very low error rate, due to the redundancy in voice information. The acceptable loss rate for voice packets through the communication medium is approximately 1-2%. This requirement is easily met by most physical communication mediums and probably is only a concern when either the receive node cannot handle the throughput requirements or when frequent collisions occur due to loading of collision avoidance protocols, such as Ethernet. Many of the above requirements not only affect the voice digitization design but also the network protocols that communicate the voice information between voice nodes.

PACKETIZED VOICE ARCHITECTURE

The voice architecture for simulated communications consists of three functional areas: (1) digitizing the analog voice to meet the minimum voice requirements, (2) processing the digitized voice at the transmit and receive nodes to provide radio communication effects, and (3) communicating the radio communication information in a method that takes into consideration both the voice and simulated communication characteristics.

The functional architecture is designed to enable integration of the simulated radio communication with the

Transmitter Signature	Transmitter SNR	Transmitter Location	Communications Channel	Compression Algorithim ID	Sample Rate	Data Sample Size	Voice Packets

0891/91001523/F064

Figure 4. Voice Packet for DIS

current DIS state information for tactical training, and consists of three functional processors: one for voice digitization, one for radio simulation, and one for network interface (Figure 3). The computational requirements and the interaction between the functional processors are affected by the complexity of providing the simulated communication effects and satisfying the voice characteristics within the DIS environment.

To digitize the analog voice, A/D converters (transmitters) are designed to have a sampling rate of 8 kHz and a sample size of 8 bits. Thus, the converters are producing 64 kbps of data to digitally represent the analog voice signal. The digitized voice is buffered in sample memory. The voice packet size for this architecture has been selected to be 2048 bits, which equates to 32 ms of voice information per packet.[4] A digital signal processor (DSP) is included on the voice adapter to control the sampling rate of the digital converters, add real world sounds to the digital voice, and process algorithms for voice compressions, if desired.

The voice processor or DSP interfaces to the host through shared memory. When a voice packet is complete, the DSP interrupts the host; subsequently, the host services the interrupt. Servicing the interrupt involves the host preparing the simulation radio packet and communicating the simulated radio packet through the network adapter. A method of preparing the simulated radio packet is to add data fields, containing state information, to the voice packets (Figure 4). This method is analogous to the DIS concept. The following information has been selected to be communicated in the data fields with the digital voice packets:

- Transmitter Signature
- Transmitter SNR
- Transmitter Location
- Communication Channel
- Compression Algorithm ID
- Sample Rate
- Data Sample Size.

In communicating the simulated radio packet, the host calls a service mechanism that is supported by the network adapter. The service mechanism, provided by the protocol on the network adapter, is dependent on both the voice and simulation needs. To simulate radio communication, the protocol needs to be able to broadcast the radio information to receivers that have the same selected channel. A subset of a network broadcast capability is the multicast capability that uses a group-addressing technique that enables the receive nodes to filter information that does not match its group address. Multicast can be used to broadcast voice packets to receive nodes with the same selected radio channel. Due to the transmission rate of voice packets, there is a need to reduce the number of packets that the receive host must process. As the number of transmit nodes increases, this problem becomes more apparent and can cause overloading of the receive node. Therefore, the group-addressing technique is important to reduce the number of receive interrupts into the host processor. As a result, multicast is the primary network requirement for DIS.[5]

To enable effective voice communications, either the network adapter or the application process must ensure relatively constant arrival rates of the successive voice packets that meets the voice requirements. Fiber Distributed Data Interface-II (FDDI-II) provides an isochronous capability that is a slotted bandwidth for 125 microsecond (us) periodic voice traffic, in addition to the synchronous capability that is supported by FDDI. An isochronous capability enables communications of voice packets every 125 us, which is the voice sampling rate. The FDDI synchronous transmission capability guarantees a maximum delay between voice packets, but does not provide a constant delay as does the isochronous capability. Isochronous communications is ideal for voice traffic while synchronous communications can be designed to ensure reliable voice communications by the network adapter. If the only communication mechanism available is asynchronous, the host can provide a form of synchronization by performing a polling loop that ensures the communication of the voice packets on a pseudo-periodic interval.

Once the network adapter receives data that is destined for its host, it interrupts the host. As mentioned, the host can use the group-addressing technique to filter messages on the network adapter that are not for the selected radio channel. The host receives the incoming data and correlates the header information relative to any additional information it has stored concerning the terrain, environment, and radio characteristics. By having the receive nodes contain static information, the simulated radio packets need to only contain dynamic information, thus reducing the network bandwidth requirements. From the data fields, the host will determine what radio effects are needed and either compute them or select filters on the voice adapter to add the effects.

After the host interprets the information, it interrupts the voice adapter and selects the effects that need to be provided to the voice signal by the DSP. The effects, which are selectable by the host, can be provided by DSP algorithms, depending on the fidelity of the radio simulation desired. Once these effects are added to the signal by the DSP, the digital voice is converted to analog by a digital-to-analog (D/A) converter. The analog signal that is heard will contain the original voice signal plus any effects due to the simulation of the radio communication.

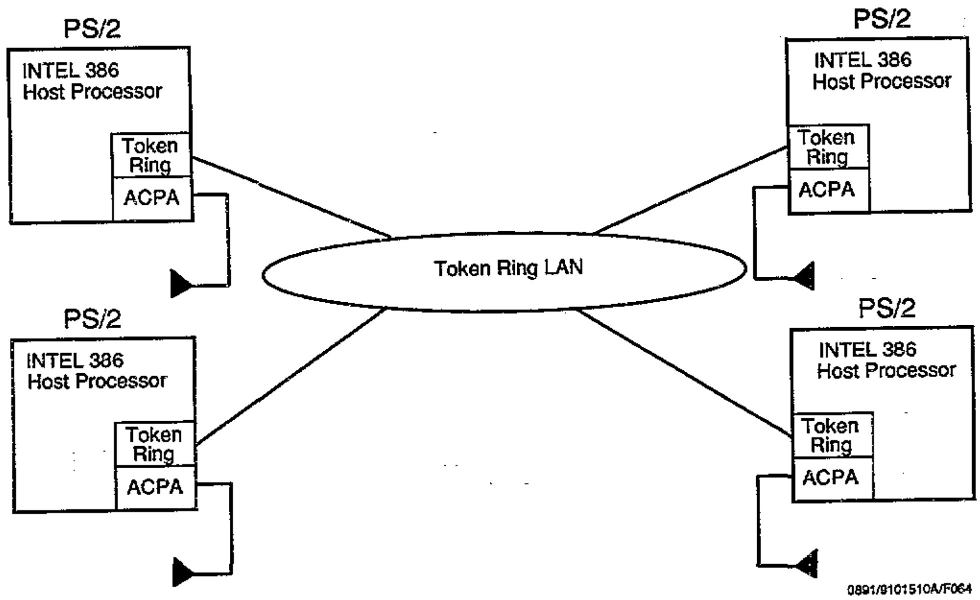

Figure 5. Packetized Voice Prototype

An example of some of the calculations required to ensure accurate radio simulations are:

- Selecting the strongest signal that results from the combination of the line-of-sight and multi-path signals
- Selecting and processing one or more DSP algorithms to replicate transmitter signature and environmental effects
- Adding signals that are received simultaneously, if the radio being simulated allows multiple voices to be heard.

PACKETIZED VOICE PROTOTYPE

The packetized voice prototype consisted of a Personal System/2 (PS/2) Model 70 hosting an Audio Capture and Playback Adapter (ACPA) and an IBM Token Ring adapter (Figure 5). The ACPA performed the voice digitization function and the token ring adapter communicated the radio communication information. In prototyping the packetized voice architecture, the simulated radio communication was integrated with the DIS state information.

The processing performed in the prototype occurred on a priority basis, using interrupt service routines to satisfy the voice interspersion requirements. The voice interrupts occurred periodically every 23 ms from the ACPA and were serviced as highest priority. Incoming simulated radio packets from the token ring adapter were serviced as the next highest priority. And the DIS state information was calculated and communicated in the time remaining between the voice interrupt service routines.

For rapid prototyping, the ACPA — a readily available multimedia device — was used as the voice adapter. The ACPA was designed to record and play back high-quality stereo sounds that contain 22 kHz of bandwidth, thus requiring a sample size of 16 bits for stereo quality and a sampling rate of 44 kHz for stereo bandwidth. The ACPA was overspecified for our voice application, but it provided the functional characteristics needed to determine performance requirements for a simulated communication system. Using the ACPA's DSP, a 61 tap lowpass filter was performed on the 44 kHz sampled data with a 4:1 decimation, providing 11 kHz sampled data without aliasing. Simulating radio communications effects would require about the same amount of processing that it took to perform the 61 tap lowpass filter. The resulting data rate of the prototype was 176 kbps per node, which is 2.75 times the bandwidth needed for a packetized voice implementation. In processing the voice signal, 256 samples of voice information was buffered per packet. Taking into consideration the 11 kHz sampling rate, the 256 voice samples resulted in 23 ms of voice information per packet transfer, which meets the voice requirements.

After the voice packets were created, the ACPA interrupted a PS/2 processor, which initiated an interrupt service routine. During the interrupt service routine, the host received the data from shared memory, appended channel information to the voice information, and issued the "SEND_BROADCAST_DATAGRAM" command using the Network Basic Input Output System (NetBIOS) software to send the voice data to the token ring network adapter. The send broadcast mechanism took approximately 3 ms to service in this prototype. The interrupt service routine needed to be completed within the 23 ms period of the ACPA interrupts, which was easily accomplished for the transmit function. The time remaining between processing the interrupt service routine and the 23 ms ACPA period was allocated to any additional processing required to integrate the DIS state information.

After the communicated packet was received onto the network adapter, the host was interrupted to receive the voice information. The host serviced this interrupt by receiving the voice information and simulating any communication effects before sending the voice information to the ACPA card. For simulated effects, incoming voice signals were added so that multiple voices can be heard when talking simultaneously on the same channel. To add multiple voice channels, three buffers were designated in the host random access memory consisting of *new data*, *old data*, and *shared memory*. When a new packet was received, the host verified the channel of the packet and discarded the packet if it was not the channel currently selected by the host. If it was the correct channel, the host moved the packet into the new data buffer, added the new data buffer to the old data buffer, and moved the added data back into the old data buffer. After the shared memory buffer was received by the ACPA, which occurred every 23 ms, the data in the old data buffer was moved into the shared memory buffer. The addition of new data with old data was performed on multiple receive frames within the 23 ms period, depending on the number of simulated radios transmitting on the same channel. Any time available between the 23 ms voice period was used to process incoming DIS state information. This function was one example of how the simulated effects could be created.

In prototyping the integration of simulated radio communications and the DIS state information, lessons were learned about the preferred capabilities of the voice and network adapters. First, the primary concern in this implementation was not so much the bandwidth utilization of the physical network but the queuing problems due to processing the frequent receive interrupts. In fact, as the number of stations were increased, the receive node became overwhelmed in processing the voice interrupts that occurred. Also, depending on the computational complexity of the effects that were provided, the host must share its resources between the processing of the application and the servicing of the voice interrupts. The computational threshold of a 20 MHz 386 machine was approached by having to process the interrupt service routines of 4 concurrent voices from the LAN while simultaneously performing simple DIS maneuvering calculations. At this time, the loading of a 4 MHz token ring was at about 40%, predominantly voice traffic that concurred with the expected traffic for 4 voices broadcasted every 23 ms over a 4 MHz token ring. In past research, it has been demonstrated that the error rate versus loading of token ring architectures, like FDDI, is good for high loading situations. Considering the 100 MHz bandwidth of FDDI, the computational loads on the host in processing the simulated radio communications were a much greater concern than the bandwidth loads of available LAN technologies.

To enhance future designs of the packetized voice architecture, the following additional capabilities would be very useful. The sample rate of the A/D and D/A converters must be at least 8 kbps with 8 bits per sample, and it would be preferred if the sample rate was selectable to simulate applications that contain higher frequency content. To drastically reduce the number of voice interrupts, voice packets which do not meet the minimum amplitude threshold due to talk spurts of normal speech pattern should not be transmitted. Papers have been written to show that thresholding voice packets can reduce the number of interrupts by at least a factor of 2. Also, it would be preferable if the packet size could be varied relative to the system error performance. The longer that the voice can tolerate between voice packets, the less interrupts the host must process. In fact, it would be very useful if the voice adapter was a bus master so that it could communicate directly to the network adapter without having to interrupt the host. In this case, the host only would be interrupted when the voice adapter needed any information to provide the effects of the radio simulation. In addition, a multicast communication service is needed to eliminate receive interrupts for voice packets that do not match the selected receiver channel.

This paper recommended an architecture for integrating simulated radio communication with distributed interactive simulation. It also revealed some lessons learned in prototyping the packetized voice architecture and desired hardware and software capabilities to develop a DIS training system that supports simulated radio communications.

CONCLUSIONS

In response to today's limited resources, the defense community must determine innovative and cost-effective methods of ensuring military readiness. The DIS environment will enable tactical training that will enhance our military readiness in a cost-effective manner. Communications in the DIS environment, both voice and data, is a key to providing effective combined arms training and mission rehearsal. Because of the recent advances in commercial technologies and distributed systems, the DIS environment has become a reality that will bring training systems into the 21th century. Innovative utilization of these networking technologies, such as integrated packetized voice, will add training capabilities that were not feasible in the past. Many more training innovations will be possible as the technologies concerning networking and distributed systems become commercially available.

REFERENCES

[1] Military Standard (Draft), "Entity Information and Entity Interaction in a Distributed Interactive Simulation,"Contract Number N61339-89-C-0043, PM TRADE, DARPA, January 25, 1991.

[2] Cohen, D., "Issues in Transnet Packetized Voice Communication," presented at 5th Symposium Data Communication, 1977.

[3] Lines, B.M., and Hughes, D., "Performance Criteria for the Transmission of Real Time Interactive Speech Over Local Area Networks Employing the Manufacturing Automated Protocol," *Polytechnic Southwest*, Plymouth, UK.

[4] Friedman, E., Student Member IEEE, and Ziegler, C., Member IEEE, "Packet Voice Communications Over PC-Based Local Area Networks," *IEEE Journal on Selected Areas in Communications*, Vol. #7, No. 2, February 1989.

[5] Gehl, T.L., Brann, J.J., "Network Requirements for Distributed Tactical Training," 12th Interservice/Industry Training Systems Conference, Orlando, FL.

ACKNOWLEDGMENTS

The author wishes to thank Joe Brann and Jon Wist for their support in reviewing the paper and his managers, Dave Dreyer and Linda Marshall, for their support. Also, I would like to thank Kathy Kelliher for her help in getting the paper out the door.

ABOUT THE AUTHOR

Thomas Gehl is a systems engineer in the IBM training systems organization. He is responsible for an Independent Research and Development project involving real-time networking of distributed simulation. He actively participates on the Communication Architecture subgroup of the Standards for Interoperability of Defense Simulations. He has supported Naval Ocean Systems Center (NOSC) efforts in determining enhancements for the Battle Force Inport Training system. He has established a Cooperative Research and Development Agreement (CRDA) with the Aircrew Training Research Division's Armstrong Laboratory at Williams Air Force Base, which involves network research for multiship training.

Mr. Gehl holds a master's degree in Systems Engineering and a bachelor's degree in Electrical Engineering from Virginia Polytechnic Institute and State University. He presented a paper entitled "Network Requirements for Distributed Tactical Training" at the 12th Interservice/Industry Training Systems Conference (I/ITSC). Mr. Gehl was formerly responsible for the design, integration, and test of digital signal processing and beamforming algorithms for the AN/BSY-1 submarine sonar system.