SESSION II

COMPUTER NETWORKS IN DATA GATHERING AND DATA ANALYSIS

Michael Yost, *Presider*

The use of computer networks in data gathering and data analysis

MICHAEL YOST and FRED BREMNER
Trinity University, San Antonio, Texas

In our research, the data are gathered in the laboratory on a DEC PDP-11/24 minicomputer. Once the data are gathered, they are analog-to-digital (A/D) converted and sent by the PDP to an IBM 370-3031. The data are submitted to a Fast Fourier Transform (FFT) on the IBM computer. Once the FFT is completed, the data are accessed by an IBM PC-XT that is equipped with an IRMA system. The PC is used in its emulation mode to access the data in the mainframe. In the mainframe, SPSS or SAS is used to do the initial statistical analyses. When necessary, subsets of the data are transferred to the PC, where other computer programs are used to analyze the data. In many instances, they are returned to the mainframe from the PC, where they are submitted to multivariate statistical analysis. Finally, all of the data are stored on floppy disks, which are prepared on the PC.

The gathering, handling, and analysis of EEG and EKG data have always been difficult. The sensitivity of the instruments, the rate at which the data are gathered, the transformation of the data needed prior to statistical analysis, and the number of data gathered from each subject make the entire experimental process one which is filled with tasks that must be carried out with great care. Each step of the experiment must be conducted with care for the results to be valid. The power and sophistication of computers now assist us in the gathering and managing of complex data.

The gathering and handling of EEG and EKG data consist of a series of complex tasks. The initial gathering and storage of the data require a computer to work with analog data. Also, a computer that can electronically control the delivery of a stimulus as a part of the experiment is frequently needed. Once the data are gathered, they must be transformed from their original analog form to a digital form. The performance characteristics of the computer used to transform the data from one form to another are typically different from those used to control experiments and gather the data. Finally, the data must be reviewed, edited, and statistically analyzed, which usually requires the use of another computer. This third type is usually a large, powerful computer designed to do mathematical calculations. Through networking or interfacing, the data

can be passed from one computer to another without having to be reentered. This article describes the review, analysis, and decision-making process involved in developing the computer network used in our laboratory to gather and analyze EEG and EKG data.

COMPUTER NETWORK

A network of three computers is used to conduct our research. Figure 1 contains a graphic description of the network. The data are gathered and the experiments are controlled by a DEC PDP-11/24 computer located in the neuropsychology laboratory. Basically, the PDP-11/24 is configured as a process-control computer. Using this computer in conjunction with a Vetter analog tape recorder, we can gather up to seven channels of data at a sampling rate of 256 data points per second. The PDP computer and Vetter tape recorder can be used to gather these data while the computer simultaneously controls the delivery of stimuli to subjects.

The data are gathered in analog form. Prior to analysis, the data are A/D converted to digital form by the PDP computer at a rate of 256 points per second per channel. A dedicated coaxial line connects the PDP computer to the campus mainframe computer, an IBM 3031 with 12 MB of virtual memory. A program called KERMIT is used to send the digitized data from the PDP computer to the University IBM. The process of identifying, acquiring, and using KERMIT is described elsewhere in this

The authors' mailing address is: Department of Psychology, Trinity University, San Antonio, TX 78284.

Figure 1. Computer network.

issue (Yost, Holcombe, & Eddy, 1985). Once the data are sent to the mainframe, they are stored on tape and then submitted to a computer program that performs a Fast Fourier Transform (FFT) on them. To perform the FFT, either a program developed by Benignus (1969) or the Statistical Analysis System (Ray, 1982) is used. Either can be used with an IBM 3278 terminal attached to the mainframe. These terminals are available at a number of locations on campus. The editor and control functions of the Conversational Monitoring System (CMS) are used to operate these programs.

The principal advantage of using the mainframe to perform the FFT is speed. Many data are generated when we gather from multiple channels at a rate of 256 data points per second for as much as 20 sec. A large, powerful mainframe is required to perform this mathematical task. Usually we submit the data to the FFT on the IBM computer at nonprime times.

The output from the FFT program is the data used in the statistical analyses, which, prior to analysis, often must be reviewed, and, in some instances, edited. Because of the loads on the IBM mainframe and the availability of terminals during the working day, it is best to take the data from the IBM mainframe to a microcomputer for editing and initial analyses. To do this, we pass the data from the IBM mainframe to an IBM PC-XT that has been networked with the IBM mainframe.

SELECTING AN INTELLIGENT WORKSTATION

A number of different networking options were evaluated prior to selecting the IBM PC-XT. Our first attempt was to use a dumb terminal with a 1200-baud modem. This configuration gave us access to the mainframe to perform our data-analysis tasks, but due to the load on the mainframe, its use during prime-time hours was often limited. There were several other limitations to the use of this configuration. When our IBM mainframe was accessed through a telephone line, the user was limited to the use of the EDIT editor and could not use the XEDIT editor. The XEDIT editor allows full screen editing, whereas the EDIT editor allows only one-line-at-a-time editing. EDIT is more cumbersome and is slower than XEDIT. Also, a 1200-baud transmission rate seems to be acceptable at first, but after a very short time becomes very slow, and the terminal appears to lack responsiveness. Finally, even with a hard disk attached to the dumb terminal, the process of storing and retrieving data is difficult and cumbersome.

Our second attempt at networking was to use a microcomputer in conjunction with a modem as an intel-

ligent workstation. This configuration allowed a little more flexibility but it still proved to be impractical. The speed with which we could work was about the same as with the dumb terminal, and we still could use only the EDIT and not the XEDIT editor. The fact that this was an intelligent workstation provided some added flexibility over the dumb terminal, in that the microprocessor could be used as a stand-alone unit when not connected to the mainframe. Passing files between the microprocessor and the mainframe was a complex task and one that was time consuming. The data-transfer rates between the two computers was very limited for the size of the data files that we use in our analyses.

In a final attempt at networking, we chose to acquire the IBM PC-XT computer with 512K bytes of core and a 10-MB hard disk. In addition, we purchased an IRMA communications card and the accompanying software programs to use in our research. The combination of the PC-XT, the hard disk, the IRMA card, and the IRMA software seems to satisfy our research needs very well. The IRMA card is installed inside the PC-XT, and the software is installed both on the hard disk and in one of our CMS machines on the IBM mainframe. The IRMA card is hardwired directly into the mainframe. The IBM PC-XT is a very powerful machine in itself and is capable of performing a great deal of work. We keep several statistical packages: a graphics package, a database manager, and a work processing package resident on the disk for immediate use at all times. We also have a printer, a plotter, a color graphics monitor, and a monochrome monitor as peripherals on the PC-XT. These peripheral devices serve to make the research efforts more efficient and provide independence from the computer center.

The use of the PC-XT as a terminal is very simple. Once the computer is booted, it automatically seeks out and connects itself to the mainframe. To use the mainframe, the user simply signs on and gives the necessary identification and account information. The transfer rate between the PC-XT and the mainframe is 9600 baud. Because of the way that the PC-XT is connected to the mainframe, it is possible to use the XEDIT editor. One of the powerful assets of the IRMA communications system is that the PC-XT can be used as a microprocessor and a terminal to the mainframe simultaneously. That is, it is possible to boot the PC-XT, sign on to the mainframe, and submit several jobs to be processed, and, while they are being processed, switch back to the PC-XT and use it as a microprocessor. Switching back and forth between systems is done by pressing two keys simultaneously. Transferring files between the mainframe and the PC-XT is a menu-driven process. The files are transferred directly between a CMS machine and the PC-XT. The process is started by executing a program on the PC-XT that brings a menu on the screen. Simple nonprogramming responses to the menu are required to transfer programs or data in either direction between the two computers. Large data sets can be transferred between the two computers in reasonable amounts of time. It is easier to store large data

sets on floppy disks with the PC-XT and the IRMA system than it is to store them on tapes at the computer center.

In 1983 the IBM Corporation announced the availability of a 3278/79 emulation system. This system is available at present for the IBM PC and the IBM PC-XT. The system requires installation of a board inside the computer, and installation of software in both the mainframe and the microcomputer. The function, use, and operation of the IBM system are almost identical to the IRMA system. Both of the systems cost approximately \$1,100.00.

We have been successful in developing a method that permits the use of the same data sets with programs on both the PC-XT and the mainframe. The data on the mainframe are accessed with either SAS or SPSS. Both of these programs have options that permit the user to generate data files within the mainframe. We begin work on the mainframe by developing a data set that we want to analyze. Standard data-set generation techniques are used to insert the common delimiters and end-of-file marks required by the programs on the PC-XT. Once the file has been generated, it is transferred to the PC-XT, where it is read into the directory of our Total Information Management System (TIM), a database management program. We then use the utility programs within TIM to transfer

the data to the directory containing the statistical program or the directory containing the graphics program. The entire process can be reversed when data are transferred from the PC-XT to the mainframe. When the data are in the mainframe, they are in an EBCDIC form; in the PC-XT, they are in an ASCII form. When the IRMA system is used to transfer the data from one computer to another, the data are automatically transformed from one form to the other.

At the present, we feel very comfortable with our ability to use three computers to conduct research. Each of the three computers is used to perform the tasks that it does best, and there are no problems in passing the data from one computer to another. This provides a very efficient and effective research working environment.

REFERENCES

- BENIGNUS, V. (1969). Computation of coherence and regression spectra using the FFT. *Common*, pp. 45-63.
- RAY, A. (1982). *Users guide: Statistical analysis system*. Chicago: Statistical Analysis Institute.
- YOST, M., HOLCOMBE, J., & EDDY, D. (1985). Microcomputer, minicomputer, mainframe interfacing. *Behavior Research Methods, Instruments, & Computers*, 17, 163-164.