Software development within the CREWES Project

Darren S. Foltinek

ABSTRACT

This report gives a brief summary of the software development that is taking place within the CREWES project. The different techniques, technologies and development tools that are used are discussed, as are software development projects which are not included elsewhere in the CREWES Research Report.

INTRODUCTION

Software development within the CREWES Project can be classified into two broad categories: research programs and system utilities. Since the primary mandate of the CREWES Project is geophysical research, most of the software that is written falls into the first category. The primary goal of system utility software is to allow better and more productive use of our diverse computer systems by all users.

SOFTWARE DEVELOPMENT

Research Software

Research software will not be discussed in any detail in this report, since the individual authors all have papers elsewhere in the CREWES Research Report which discuss their algorithms and results. In general, software for the purpose of research is generally written and used solely by its author. The goal of most such programs is to implement and test a new algorithm or process. Because of this, the author generally expends a minimum of effort on issues such as the user interface, documentation, and code robustness. This should not distract from the value of the code, however, since these programs are, by definition, leading edge technology.

System Utility Software

System utility software is developed by the CREWES staff to assist in the effective and efficient use of the computer systems. The users of our computer systems cover a wide range of computer expertise levels, from beginners, with little more than application experience, to others who have many years programming experience. Because of this wide range of users, the high complexity of the computer system, and the variety of demands placed on it, there are many instances where a custom program will save a great deal of time and trouble for both the users and administrators of the system. This report only discusses the three major such programs which may be of interest to our sponsors. See the report by Henry C. Bland, "CREWES Computer Systems" for a complete listing of the variety of computer hardware and software in use.

Disk Space Management

Due to the generally large size of seismic data files, disk space on our computer systems is a resource that is always in high demand. On our network of Sun workstations we have over 5000 megabytes (MB) of disk storage spread over five disk drives which are used primarily for seismic data. The disk space requirements of individual users vary greatly, from those who need less than 50MB to others who are processing 500MB 3-D datasets. As most people know simply by looking at their own desk or garage, space requirements quickly expand to fill the available space. Currently disk space is allocated on an individual basis, whereby a user comes to the system administrator with a request for space to process data. The administrator then attempts to find the user the required space, which, for the sake of processing speed, must be located on a single machine. Often the administrator is forced to request that other users move or remove some of their data files. This is a continuous time expenditure for everyone involved. It was decided that a more automated system would be beneficial. After failing to find any commercial software that would fulfill our needs, the development started on the Disk Manager system.

When the Disk Manager software is complete, it will allow users to sign out disk space in the same way that a book is signed out of the library. A user allocates disk space which must then be returned to the system or renewed by a given date. Our current design also allows accounting information to be maintained, allowing the effective monitoring of usage and identification of users who are using unreasonably large amounts of space. Disk Manager is highly distributed, owning to the distributed nature of the available disk space. Each machine that has disk space allocated for seismic data will have a *server* process running on it. This server is responsible for maintaining a database of space available on its machine and answering requests for disk space. A *client* process, started up by a user wishing to allocate space, asks the user for the amount of space and the time period for which it is required. The client process then broadcasts a request for available disk space over the network. Any running server programs answer, detailing the space available on its machine. The user is then presented with a choice of machines on which their disk space may reside. After selecting a machine, the client process sends a request to the server, and the space is allocated. The Disk Manager system is thus a good example of a client-server based software package. The system, when operational, should improve both the efficiency and accountability of disk space allocation.

Plotter Drivers

Among the computer facilities of the CREWES project are two large plotters. These plotters are used to plot 24 inch black and white or 36 inch color seismic sections. It was realized that these plotters could also be used for the printing of large images of any nature. The production of posters used at the sponsors meeting, conventions, and tutorial days has always been a very time consuming and often expensive endeavor. It is now possible to design these documents on our Macintosh computers, render the resulting Postscript output files on a Sun workstation, and print the result to a plotter. Using the public domain Appletalk network software "CAP" (Kim & Schilit, 1988), the Postscript rendering software "Ghostscript" (Deutsch, 1992) from the Free Software Foundation, and the public domain bitmap manipulation library "PBM" (Poskanzer, 1989), the only missing piece was a device driver for the plotter. This results in significant savings in the time and expense required for the preparation of large format posters

SEG-Y File Utilities

The SEG-Y file format is one of the most common data formats used in seismic processing. While the published SEG-Y file format was precisely defined, there exist today several incompatible variations on the format. The major source of incompatibility is the different binary representations of floating point and integer data on different hardware platforms. The original specification used an IBM mainframe representation of the data. When seismic data is to be used on either the Sun workstations or PC compatible computers, the file must often be converted to the native data representation of the machine. The program we have developed, called *cvtsgy* (Bland, 1992) will convert data files between these basic formats and several other mutations of the original specification which we have encountered. There is also a related program, called *dumpsgy* (Bland, 1992) which allows the viewing of SEG-Y trace data and file and trace headers, from a SEG-Y file in any of the supported formats.

SOFTWARE DEVELOPMENT TECHNOLOGY

Hardware and Operating System Environment

Software development within the CREWES Project takes place primarily on the Sun workstations, which are running the OpenWindows X-window environment on top of Sun's variant of the Unix operating system. The resulting multi-tasking windowed operating system provides phenomenal improvements in productivity over the single-tasking DOS environment. Compared to DOS, these machines offer superior speed and stability, combined with the ability to simultaneously view source code, follow program execution in a debugger, and display its output. It is a combination thats advantages should not be underestimated by anyone considering a software development platform.

Programming Languages

Most of the research code that has been developed was written in Fortran 77. Despite being somewhat antiquated, it continues to be the programming language of choice among most scientific software developers, primarily due to its history, wide availability, relatively high level of standardization, and familiarity to most Geophysicists. When the complexity of code becomes high enough, however, the developers job is made more difficult by the constraints imposed by this simple language. A number of researchers at CREWES have thus begun using the C language, because it offers developers far greater power in its flexible handling of data and structuring of code. For the casual programmer, however, C can be a very difficult language to work with, and the power of the language is not usually needed until programs of significant data or structural complexity are written. For system utility software that we have developed, it is the language of choice due to its ability to work at the level of the operating system.

The Internet and Public Domain Software

One of the services that is only beginning to become widely known outside of academic computing circles is the international network of computers known as the Internet. The Internet has a long history (by computer standards), and today is coming to the forefront of modern computing. Among the many services if offers, one of the

in the world. This had lead to the establishment of extensive archives of public domain software, where packages that have been released for free public use are stored. The enormously wide variety, high quality, and easy availability of this resource has allowed us to keep the CREWES Project on the leading edge of computer technology. This has greatly benefited the end users of our computer systems, and made the lives of system administrators and software developers much easier. For a history and discussion of the Internet and some of its more popular services, interested readers are referred to the "Zen and the Art of the Internet", (Kehoe, 1992). See "CREWES Computer Systems" (Bland, 1992) for a list of some of the major public domain software packages we are using.

CONCLUSION

While software development is not an extensive component of the CREWES Project, the development taking place is vital to both the research being done and the efficiency with which it takes place. All of the software being developed within the CREWES Project is available to our sponsors, and packages which are already complete will be distributed at the 1992 sponsors meeting. We will also provide our sponsors, by request, with any of the public domain packages that we have found helpful.

ACKNOWLEDGEMENTS

First and foremost I would like to acknowledge the sponsors of the CREWES project for their continuing support, as well as the Department of Geology and Geophysics for the use of their computer systems. I would also like to thank the many authors of public domain software, whose excellent packages have assisted us greatly.

References

Bland, Henry C., 1992, CREWES Computer Systems, CREWES Research Report vol 4.
Bland, Henry C., 1992, "CVTSGY", "DUMPSGY", The CREWES Project.
Deutsch, L. Peter, 1992 "Ghostscript", 1992 Aladdin Enterprises.
Kehoe, Brendan P., 1992, "Zen and the Art of the Internet".
Kim, Charlie C. and Schilit, Bill, 1988, "CAP - Columbia Appletalk Package for Unix", 1988, The Trustees of Columbia University in the City of New York..
Poskanzer, Jef, 1989, "PBM - Extended Portable Bitmap Toolkit", 1989 Jef Poskanzer.