SCIENCE AND ENGINEERING RESEARCH COUNCIL
RUTHERFORD APPLETON LABORATORY

COMPUTING DIVISION

D I S T R I B U T E D C O M P U T I N G N O T E 5 1 0

VISITS
Notes on a visit to Dr K H Bennett
University of Keele, 13 November 1981

issued by
Dr D A Duce

17 November 1981

DISTRIBUTION:
R W Witty
F Chambers (Logica Ltd)
D A Duce
Miss G P Jones
Investigataors/Bennett

OVERVIEW

The purpose of the visit was to introduce Fred Chambers to the Keele project and discuss progress.

Keith Bennett gave an overview of the project. In 1978 the University Computer Centre was seeking a machine to act as a basis for an interactive computing environment (batch facilities being provided by the North West Universities network). GEC proposed a number of 4080's for this purpose. Several members of the computer science department were worried by GEC's glib dismissal of the distributed filestore problem and this led first to a feasibility study funded by DCS (1978/9) and then to the present grant (1979-82).

Two RA's, Pearl Brereton and Paul Singleton, are employed on the project and a number of PhD students are associated with the work, Ken Lunn (finishing off), Pete Trueman and Jerry Collins (taking over from Ken Lunn). In addition Phil Peake (University Computer Officer) works part-time on the project. Phil is the Unix guru.

One aim of the project is to provide a Pascal compiler and execute system from personal workstations accessing the distributed filestore. Particular concerns are reliability and the reliability versus performance tradeoff. Reliability is achieved through redundancy.

One application being constructed is a filestore spread over two or more machines accessible from a number of user workstations, communicating over a Cambridger Ring. The pilot implementation is based on a Pascal Microengine and filestore distributed over 2 LSI-11s. On a good day this system can be demonstrated!

In this system, explicit transfers have to be made between the remote and local filestores. Consideration is being given to integrating the distributed system in a conventional operating system (Unix). It is hoped to have this operational by June 1981.

Aims of the project are a more quantitative understanding of trading reliability and performance, and of the nature of the interface to the distributed filestore. The understanding of the interface to the filestore is a crucial issue. Currently the interface is at the filename level.

Paul Singleton is looking at a more abstract, theoretical approach to the problem, though since this is very new work little progress has been made so far.

Peter Trueman has been considering what happens when a processor goes down. Consideration is being given to whether to build or simulate a prototype system.

Experience with the Cambridge Ring has shown that performance is determined by software and hardware interfaces. Performance figures obtained with Qbus PI access logics are 1.5 Kbytes sec^{-1} (interface driven through UCSD Pascal run interpretively) rising to 6 Kbytes/sec with hand coded assembler software. The group are writing a paper on experience with the Ring.

The filestore software running on the 11/02, is written in Modula. Some limitations have been found in Modula for this task, especially lack of process priority and control over scheduling. Modula-2 overcomes some of these problems. Ian Wand has been given feedback on their experiences with Modula.

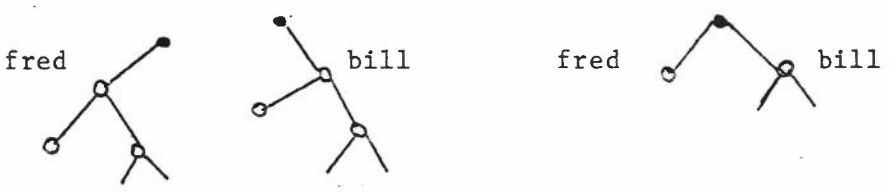
The filestore design assumes the existence of a high speed local area network. In principle the design should be independent of network technology, but in practice there are problems with how to handle timeouts in long blocks, which leads to the receiver code being written with a knowledge of the properties of the sender.

FILESTORE DESIGN

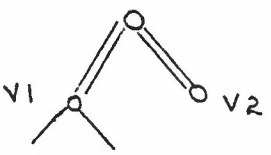
Ken Lunn gave a more detailed presentation of the filestore design.

The design assumes a totally reliable local area network and presents a hierarchical Unix-like directory structure.

Each disc holds a directory tree:



These can be overlaid:



Ken has developed a set of algorithms which allow overlaying and ensure consistent access.

The set of volumes on which a given directory occurs is termed the set of associated volumes. Root must be present on all volumes. Suppose the assiciated volumes of fred are V1 and V3, then any files underneath fred must be replicated on V1 and V3. The set of associated volumes of directories beneath fred must be contained in [V1, V3].

Directories are manipulated by a single directory server. An algorithm has been devised to enable a workstation to locate the directory server.

Clients talk to the directory server to insert, find, lock or protect files in the directory scheme.

When a directory is created one can specify how many volumes it will reside on.

Algorithms have been developed to automatically restore consistency when a volume is brought on line.

A client writes only to a single scratch file so that if a server goes down one can recover back to any chosen atomic action. The filestore provides atomic read/write to disc and atomic update to the directory server. The directory server also provides locks so that only one update can be in progress at any time.

To read a file one issues a find request to the directory server which returns pointers to each copy of the file in the system. The client then issues a read request for the selected copy of the file. This mechanism is provided so that the client can perform same scheduling of requests to particular servers, load balancing etc. It is envisaged that this interface will not be visible to the end-user of the system, but will be hidden by some intermediate layer (shell).

Jerry Collins is currently learning about the system, it is proposed that he will develop applications programs using the system to get a practical feel for how it behaves.

Thought is being given to performance measurements of the system.

It is hoped by the end of the project to have produced a set of design rules and performance data based on experience with this approach.

FUTURE WORK

The present grant expires in December 1982. Keith has no firm plans for continuing the work yet, but would not like to see the project split up.

Ken Lunn is taking up a post with the University Microprocessor Applications Unit when he finishes his PhD. The Unit offers a consultancy service to local industry.

MISCELLANEOUS

The department has a PDP11/10 with large disc stood idle, for want of a Cambridge Ring interface. Use of this machine would considerably aid evaluation of the filestore - DAD suggested Keith request the loan of a Ring interface from the pool for this machine.

Keith has a paper design for a Z80 based hardware basic block machine to interface to Qbus. Cost to build would be £6-700 in components.