CERN-Data Handling Division
DD/79/2
J.M. Gerard
April 1979

The Development and Use of a High Speed Packet Switching Network in a High-Energy Physics Laboratory

(to be presented at the European Conference on Applied Information Technology (EuroIFIP79), London, 25-28 September 1979)

The Development and Use of a High Speed Packet Switching Network in a High-Energy Physics Laboratory

J. M. Gerard

CERN (DD Division)
1211 Geneva 23
Switzerland

ABSTRACT

During the past few years CERN has developed a general purpose, high performance packet switching network, called CERNET. Although certain design aims were fixed the actual usage has not always followed exactly these aims. The paper outlines the history and development of CERNET, its facilities, usage and future plans. Emphasis is placed upon the general aspects of the design and use rather than the particular techniques which have been used in the hardware and software.

1. INTRODUCTION

CERN, the European Nuclear Research Centre, is a high-energy physics research laboratory situated on the French/Swiss border near Geneva. In fact, the laboratory is divided geographically into two sites, one of which actually spans the border, whilst the second, newer site is completely in France a few kilometres away.

The actual research in CERN is carried out by teams of resident or visiting scientists who set up experiments in various regions of CERN. These regions are also widely scattered over both sites, thus giving rise to a general communications problem.

All of the experiments now use minicomputers in various ways. There is always a computer controlling the acquisition of experimental data and the storing on magnetic tape. Further facilities, including verification of the quality of the data, rejection of uninteresting data etc., depend upon the power and sophistication of the minicomputer(s).

The number of minicomputers in CERN currently exceeds two hundred and covers a wide range of manufacturers. CERN has tried to standardise wherever possible, so that most of these computers are Digital Equipment, Hewlett Packard or Norsk Data. However, since experimental groups may bring their own computer, or software, complete standardisation has proved impossible.

For its brute number-crunching power, CERN has a single computer centre located on the original site. This centre has gone through several phases, and is currently equipped with both CDC (a 7600 front-ended by a 6400 and a 6500) and IBM (a 370/168 and a 3032) mainframes.

The current trend towards increased complexity of experiments at CERN makes it necessary to have such large mainframes in order to process the data produced by these experiments, since the minicomputers in the experimental areas are unable to do so.

2. HISTORY

Practically ever since CERN began acquiring large mainframe computers it has been attempting to make their computing power available to the minicomputers in experimental areas. Early attempts to link directly on-line to the computer centre were made difficult by the inability of the mainframe operating systems to cater for the mixture of on-line real-time data analysis with off-line batch environment work and, later, terminal access. For this reason, the idea of on-line data acquisition and analysis in the computer centre was temporarily abandoned in favour of two alternative approaches, which were pursued in parallel.

One approach involved using the central computers only in batch mode. This was implemented in a system called FOCUS[1] which was in operation from 1968 until the end of 1978. Here a CDC 3000 lower series computer was equipped with data links to various experimental set-ups. By means of terminals connected to FOCUS the physicists could send data sample files to the 3000 file base, manipulate source files, initiate transfer of jobs (including the data sample files) to the central CDC computers and retrieve output for inspection or printing. At its peak (1970-1975) FOCUS was handling about 20 simultaneous terminal users and about 10 data links, plus three Remote Job Entry stations. However, its services were tending to become overstretched and it could not easily be extended to include the IBM computers.

In an alternative approach, for a very large experimental facility called OMEGA, a medium size CII 10070 computer was purchased in 1970 specifically to provide real-time data analysis and associated support facilities including terminals. The data communications function was performed by a network of PDP computers, knowns as OMNET[2]. The CII 10070 was logically in the centre of this network, with the terminals being connected to the various PDPs. This system also lasted until the end of 1978, at which time the CII 10070 was discarded as too old, expensive to maintain and not powerful enough. However, the PDP network was retained for integration into planned facilities.

During the mid-1970s it became clear that the lifetime of both FOCUS and OMNET was limited, so that both would need to be replaced fairly soon. With the sophistication of modern mainframe computers and operating systems and the acquisition in 1976 and 1978 of large IBM mainframes and mass storage facilities to complement the existing CDC mainframes, it was also felt that the various facilities could benefit by being reintegrated into the main computer centre. In addition, one had to take into account the growth, both inside and outside CERN, of other computer networks constructed for particular purposes. An example, inside CERN, is the very successful SPS network of Nord-10s[3] which is used to control the particle accelerator from a single control room. External to CERN there are many public networks being developed, such as EURONET, TRANSPAC.

The decision was thus made in 1975 to construct a general purpose data communications network (called CERNET) inside CERN, to be used for computer-computer communications. The performance should be such as to allow data transfer at speeds comparable to that of writing data onto magnetic tape. However, the network should also be able to handle lower speed traffic in an integrated way.

3. PROJECT DEVELOPMENT

As a result of CERN's previous work on high speed data communications there was already a high level of technical expertise on the construction and use of high speed links. Also there was already a large number of standard high-quality twisted pair cables over much of the site. Thus it was decided that, regardless of computer or type of network, the actual data links would be designed at CERN. This has been done, and the current links can run at 5 megabits/second over short distances, or 2.5 megabits/second over distances of several kilometres, in an asynchronous mode[4].

The type of method was chosen as a packet-switching one, rather than circuit-switching, because it was felt that packet-switching was both general purpose and very flexible. It was also hoped that packet-switching could be made to work sufficiently fast for the proposed types of application by a suitable choice of hardware and protocols: this hope has, so far, been realised. The general topology of the network was foreseen as mesh-type, with particular data links being inserted according to either traffic requirements or safety back-up needs.

The choice of computer for CERNET was, at the time, a choice between various minicomputers which could provide a variety of performance and hardware devices but in general no software specifically designed for high-speed packet-switching. After the usual type of evaluation, benchmarks, cost comparison, etc., it was decided to do the implementation on Modular Computer Services (Modcomp) Model II series computers. The only software to be taken was a basic communications-oriented operating system called MAXCOM.

The project was implemented in two phases. In the first phase the emphasis was on providing a network which could link together the mainframes, the minicomputers in the 'North Area' region of the second CERN site, the minicomputer software development laboratories and certain selected minicomputers on the first site. This latter group included a connection to the OMNET network, via a gateway, so as to give all OMNET subscribers access to the CDC/IBM complex in time for the removal of the CII 10070. The first phase was terminated at the end of 1978, at which time the configuration was as shown in Figure 1.

During the first phase the effort was mainly directed towards having a stable network and a reasonably complete range of user services. The actual throughput performance of CERNET was a secondary goal so long as an adequate data rate could be achieved. In practice, transfers of files took place at around 15 kilobytes/second, whilst special high-priority tests could get over 60 kilobytes/second. The limitations partly came from CERNET itself but also were much affected by the choice of priority level for the network interface software in the mainframes and the effective rate of disk accesses.

The second phase began in September 1978 and had as goals the extensions of CERNET services to the rest of CERN, links (via 'gateways') to other networks, both internal and external to CERN, and a general improvement in the performance and facilities offered by CERNET itself and the mainframe computers. Figure 2 shows how it is envisaged to make the extensions during the second phase.

The choice of the various levels of protocols was influenced by the hardware design and the speed requirements. At the time the development of standard protocols was at an early stage and it was unclear whether those in use, all of which tended to have small block sizes, would be capable of handling the data rates which were envisaged. The decision was thus made to define protocols specifically to fit with the hardware and speed requirements. In fact, CERNET is basically only a datagram network but with the special property of guaranteeing delivery in the same order in which the datagram packets enter the network. Thus, higher levels of protocol, namely program-to-program and file access protocol, are necessary for meaningful dialogue with the computer centre mainframes. The design work was greatly influenced by the protocols used in the Cyclades network[5].

For the software it was decided to avoid assembly language coding wherever possible, in favour of a high-level language. The ease of writing and debugging of programs far outweighs any loss in speed or increase in memory requirements, especially since speed can be attacked by isolating critical code, whilst memory expansion is becoming relatively cheap. The chosen high-level language was BCPL[6]. This is not specifically a systems programming language but is quite adaptable to writing systems programs and is extremely portable. The latter is very important since large amounts of software written for the CERNET nodes can also easily be modified for user machines. The overall software design assumed that CERNET would be a general mesh topology. Although a network control centre was planned, each node in CERNET was to be able to act independently. The path through CERNET for traffic between a particular pair of computers (subscriber or host) was to be fixed at any given time as a function of current topology. In the case of topology changes (link or node failures) CERNET was to automatically adjust to the new topology.

CERNET began genuine operation in March 1978 for a single user and mainframe (370/168). Expansion continued during that year so that by the end of 1978 there were more than a dozen minicomputer users plus both CDC and IBM mainframes. CERNET was also used for utility file transfers between the different mainframes.

4. USER-AVAILABLE SERVICES

However well a network may perform, its usefulness to the user is only as good as the services which can be obtained from it and the other computers connected to it. Thus, it is necessary to design and implement software for execution within the IBM and CDC mainframes to provide such services. These services of necessity involve a basic software module known (in standard networking jargon) as a Transport Manager, which handles multiple simultaneous conversations known as logical links or virtual calls.

The next level up from logical links is the transfer to and from the mainframe computer permanent file systems of various types of files. It was decided to implement this next level in a separate protocol, known as the File Access Protocol, which may be used by any minicomputer to talk to a File Manager program residing permanently in the mainframe computers. By this means the user can transfer data files in either direction without the need to write any software in the mainframes. Such files might be data sample files, utility programs, copies of a minicomputer file base etc.

The File Manager exists on both CDC and IBM mainframes. On the CDC it is combined with the Transport Manager into a single offering which can be run on the front-end 6400/6500 machines. This is convenient since these front-ends control the permanent file base but are not intended to do much serious number-crunching type of work.

In the case of the IBM the two are separate, so that the File Manager is simply a user of the Transport Manager, albeit with a high priority. However, it is quite acceptable for other user programs to make calls directly to the Transport Manager, via a set of interface routines which are held in the CERN program library. Such calls can request or accept logical links with other computers on CERNET, pass data in either direction and close the link at any time.

By means of the Transport Manager interface routines a sophisticated minicomputer programmer can permit a data analysis program executing in the IBM to take its input from a logical link to a minicomputer, analyse it and return results to the minicomputer. It also allows computer support programmers/analysts to write standard complete programs for the IBM which converse with the CDC File Manager to transfer files in either direction.

The minicomputer user himself, of course, requires software. The viewpoint which has been taken here is that if he has one of the standard minicomputers with a standard operating system then he can be given standard software. This software consists basically of a Transport Manager program, written in BCPL in such a way as to be as portable as possible[7], plus an interface package at the File Manager level. These interface packages are not necessarily written in BCPL but are always callable from the high-level language(s) normally used on the particular minicomputer involved.

5. CERNET USAGE

Despite the intention to provide real-time data analysis the first use made of CERNET by an experimental group simply involved using the IBM computer file base as back-up for the local minicomputer file base. The reasons for this are fairly clear:--

(a) the file bases on mainframe computers are well safeguarded against file corruption problems by reliable back-up systems;

(b) terminal facilities, including source editors, output scanning etc., may be much more sophisticated on the mainframes than on the terminal(s) which might be connected to the minicomputers;

(c) the programming of the minicomputers to send files to the mainframe computers, either by individual files or groups of files, was fairly straightforward.

Because of these reasons this first experimental group was soon regularly transferring several hundred files of various sizes and formats to the IBM. It has been invariably the case that other CERNET users have begun by doing the same thing.

Very shortly after this first use the computer centre programmers wrote a number of utility programs which executed on the IBM and transferred files of various types to and from the CDC 6400/6500. The end user for these programs is then normally someone who is logged in on the IBM terminal system Wylbur[3] and who invokes them by use of catalogued procedures known as 'EXEC' files. Thus, the actual usage is simple and much faster than the alternative method of going via magnetic tape.

By mid-1978 the original most enterprising experimental group did actually graduate to real-time usage. This usage involved locally selecting a particularly interesting unit of data (known as an event) and sending it to the IBM for immediate analysis, with the results being returned in a matter of seconds for superimposing on the original graphical representation of the data. In order to guarantee this response time a special job class was defined in the IBM specifically for this type of job. Initially the effect of this job class on other activities of the IBM was minimal, due to the relative infrequency with which the facility was used. However, it is clear that an abuse of this type of facility remains possible unless the job scheduler is capable of handling it.

The actual usage has, therefore, demonstrated that whilst relatively exotic applications are possible via CERNET, the bulk of the traffic is composed of mundane file transfer.

6. FUTURE PLANS

CERNET will inevitably grow in size and complexity as time goes on. It is the intention that this growth should increase, and not reduce, the overall reliability to the end user.

More important are the actual services offered by and via CERNET. It is these services that the user actually sees and uses, hence to him they are the network. In fact, the basic network should ideally be completely invisible and 100% available.

In the central mainframes it is planned to increase the capability of the File Managers to handle more types of files, including job and output files. Also an access method into the terminal system of these mainframes would allow the terminals of the minicomputers to act as terminals of the mainframes. However, terminal concentrators as such are not planned since CERN already has an extremely flexible hardware solution called INDEX[8].

CERNET itself is planned to include a control centre capable of monitoring the entire network and connected computers' status etc. Such a control centre could also itself provide services such as control of file transfer between other pairs of mainframes.

The services to be offered to minicomputers are very dependent on how much already exists in the minicomputer configurations. It is planned to cater for the 'worst' case of a minicomputer with very little in the way of peripherals and software by providing cross-assemblers, crosscompilers, link editors etc., normally on the IBM and doing downline loading of complete core-load or overlays. This is being done in a machine independent way[10] and in fact the first customers for such support are the CERNET nodes themselves.

Finally, it is hoped to offer standard packages to the standard minicomputers for a variety of applications. One such package would be for automatic file archiving and retrieval. Other possibilities are that some of the larger configuration minicomputers may implement a File Manager and thus offer services themselves.

ACKNOWLEDGEMENTS

The design, implementation and entry into service of CERNET has only been possible through the efforts of a large number of people over a long period. To mention only some individuals would be to risk offending others perhaps equally deserving; to mention all of them would be impractical. I, therefore, put on record my sincere thanks to everyone who has contributed as part of the 'CERNET team'.

REFERENCES

1 Ball D., Blackall P. M., Gerard V., Macleod G. R., Marcer P. J., Palandri E. M., FOCUS - A Remote Access File Handling System On-line to a CDC 6000 Series Computer, British Computer Society Journal, Vol. 14, No. 2 (May 1971).

2 Russell R. D., The Omega Project: A Case History in the Design and Implementation of an On-line Data Acquisition System, Proceedings of the 1972 CERN Computing and Data Processing School, p. 275.

3 Crowley-Milling M. C., et al, The Multi-purpose Control System for the CERN 400 GeV Accelerator, Proceedings of the Conference on Trends in On-line Computer Control Systems, IEE, 01-108 (1975).

4 Joosten J., Pieters R., Specification of Modcomp Links, CERN Data Handling Division Network Project Note DD/NPN/76/25 (Nov. 1977).

5 Pouzin L., Presentation and Major Design Aspects of the Cyclades Computer Network, Proceedings of the 3rd ACM/IEEE Data Communications Symposium, Florida, p. 80 (Nov. 1973).

6 Richards M., BCPL: A Tool for Compiler Writing and System Programming, Spring Joint Computer Conference (1969), p. 557.

7 Jacobs N. J. D., Nerdal J. I., The CERNET Portable Transport Manager, CERN Data Handling Division Report DD/78/4 (April 1978).

8 Wylbur/370 Reference Manual, Stanford University (Nov. 1975).

9 Slettenhaar H., INDEX: A Digital 'Telephone Exchange' System, CERN Data Handling Division Report DD/77/11 (Sept. 1977).

10 Montuelle J., Willers I. M., Cross Software Using a Universal Object Format CUFOM, submitted to Europ IFIP 1979.

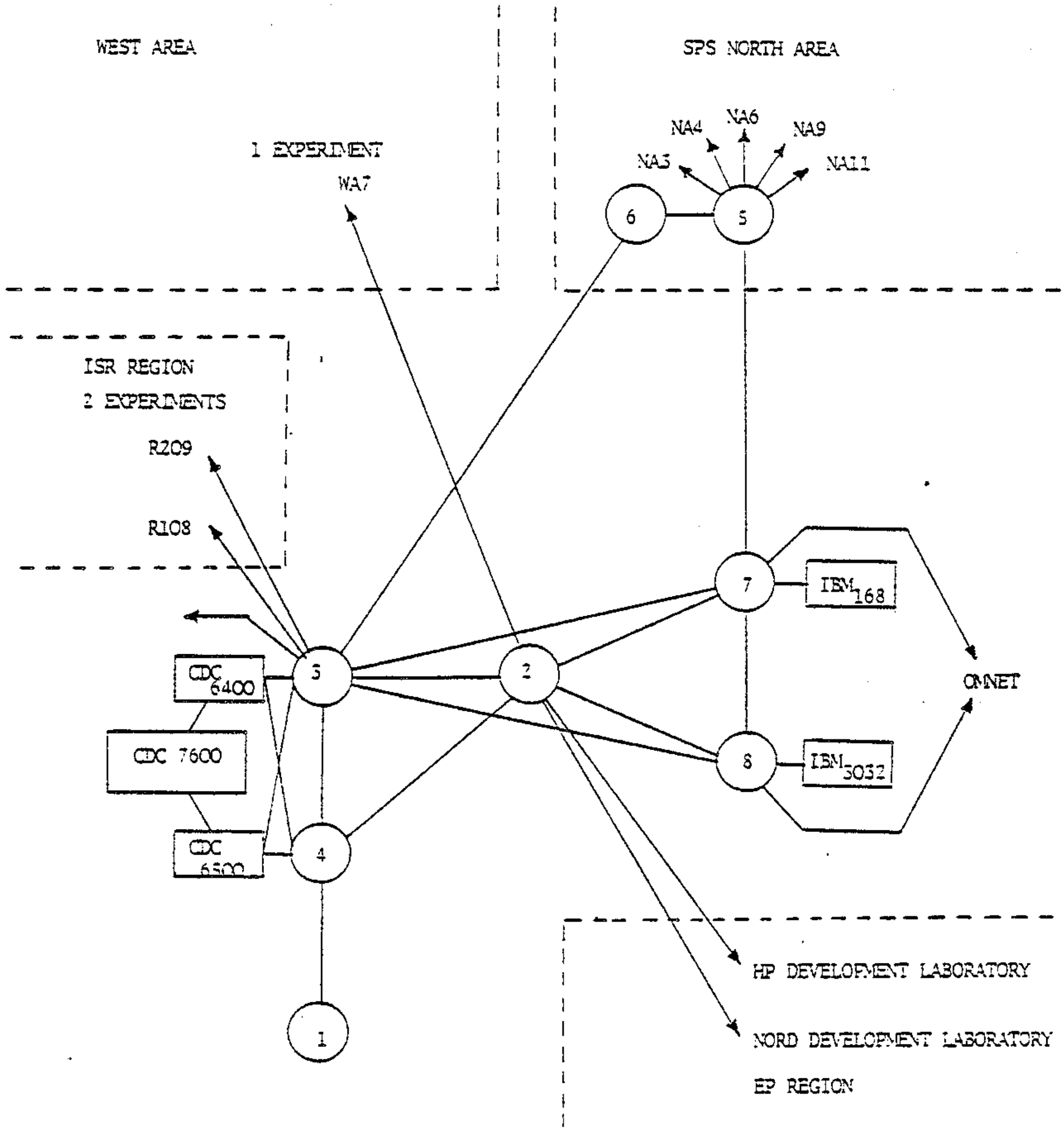

Figure 1: CERNET Configuration at the End of Phase 1 (December 1973)

Each number circle represents a Modcomp CERNET node computer. Their functions are as follows:

1.	Development machine
2.	General purpose concentrator
3, 4.	CDC concentrators
5, 6.	North Area Nodes
7, 8.	IBM concentrators

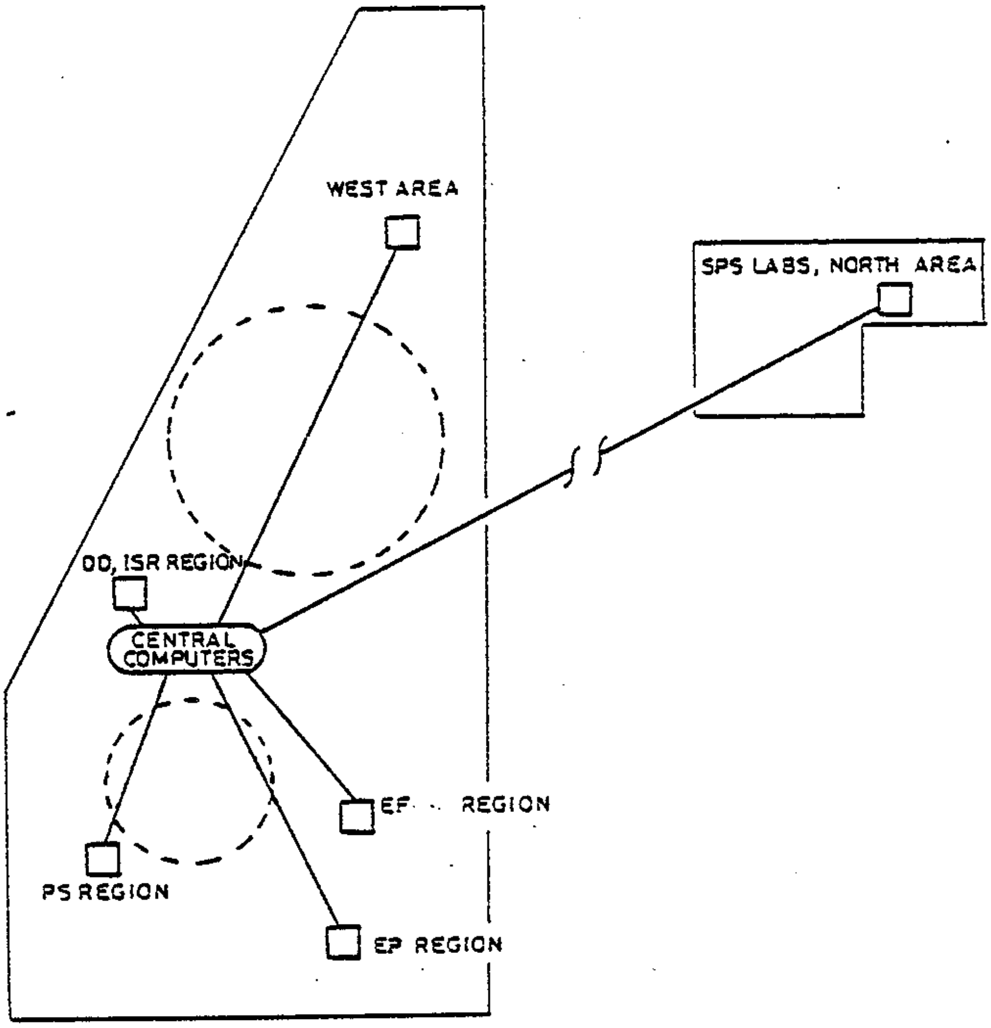

FIGURE 2: General Picture of the Phase II Extension of CERNET Facilities

Each square represents a connection-point servicing links to users' computers or experimental equipment in a number of regions of the site. Depending on the growth in data traffic from these regions, and the degree of availability required, each square may represent one or several CERNET node computers and each line to the central computers may represent one or several CERNET links.