

Historical Note on the Early Development of Packet Switching

There are two purposes in writing this note. One is to help the Science Museum project on the history of computing and the other is to provide material for the historical paper in the forthcoming IEEE Proceedings on Packet Switching. Some of the events described here are now more than ten years away and memories are fallible so it seemed worthwhile to go back to the earliest references I have and try to reconstruct, from a personal viewpoint, the part of the early history of packet switching of which I was aware.

In 1965, time sharing was very new and during a return journey from IFIP Congress in New York I took up a suggestion of Professor Shackel that NPL should organise a special seminar on time sharing and invite people from Project MAC later that year.

When that meeting took place in Autumn 1965¹ the inadequacy of data communication facilities for interactive systems was one of the things we talked about and Larry Roberts was among those who attended. The possibility of matching a system to the special traffic characteristics of time sharing systems was discussed often and I remember talking about it earlier that year with Richard Mills at MIT.

It was shortly after that meeting, ~~probably in September or October~~, that it struck me very forcibly that a store and forward system, if it was matched to very short message sizes, would have no problem in achieving transit times consistent with interactive systems.

I wrote a note dated November 10th 1965 entitled 'Remote on-line data processing and its communication needs'. This note described a store and forward system for short messages. It considered the use of a PCM-type channel of 1.5 Mbit/s but pointed out that all the keyboard traffic that would be generated in the London area would not saturate one such link. It proposed that the delay at any switching centre should not exceed 100 ms and consequently the lines in such a network should have a capacity not less than 10000 bit/s. It envisaged that such a 'short-message' data service would take over telegraph and telex services and make them more convenient.

This note also suggested that the control signalling requirements of the telephone network could be handled by the same mechanism more efficiently than by using individual telephone channels. The interaction between the new network and the telephone network would help to provide more elaborate interactions with the telephone system.

¹ Nov. 2, 3.

There was a short supplementary note on 16th November 1965 which among other things listed 12 special facilities which a new network could provide.

These papers were sent to a number of people who in their replies showed some interest. On 15th December 1965 I wrote a longer document entitled 'Proposal for development of a national communication service for on-line data processing'. This paper elaborated the earlier proposal and made an approximate calculation of delay based on the M/G/1 queuing system with constant service time. It concluded that each transmission path should have a capacity of at least 250 messages per second. This required 'a few PCM telephone channels' which meant around 100 K bit/s. It envisaged a switch which could handle 10000 messages per second.

The paper goes on to consider a 'character assembler and distributor' which is the device now known as a PAD. The capital cost per line of the whole system was estimated at £640.

A special kind of terminal was envisaged which is effectively a packet terminal and the use of these was thought to be more economic than providing the PAD function in the switch.

The paper went on to consider the interface with character terminals and with computers for which the communication system would act as a line concentrator. It then described the commands passing between the system and listed a dozen or so uses for a message communication network and ended by proposing that a pilot service should be set up in Central London.

I circulated this proposal widely among those who I knew were interested such as Banks and the UK Post Office. The reaction was generally appreciative but not enthusiastic.

During the next month or two I developed the design in more detail and in March 1966 gave a lecture at the NPL entitled 'The future digital communication network'. This was attended by more than 100 people of whom 18 were from the UK Post Office. By this time I had changed my maximum packet size to 128 bytes.

After the lecture, Dr. A. Llewellyn (now Director of the CAD Centre in Cambridge, UK) offered to send me the report by Baran Boehm and Smith of the work at Rand Corporation which, though it had been published by Paul Baran in 1964, had yet not come to our attention. By coincidence we had independently chosen 1.5 Mbit/s and 1024 bits as our main parameters. Our subsequent papers gave the 1964 paper as a reference. The distinctive features of our proposal were its matching to the short-message traffic of time-sharing systems, the packet

interface and PAD functions and the proposed ‘interface computer’ to mediate between the packet network and its terminals.

The contents of the lecture were later recorded in a paper which was circulated privately entitled ‘Proposal for a digital communication network’ dated June 1966. This document is the first one which uses the word ‘packet’, a word which I introduced to make a distinction between the message unit which the network employed and the message which the user wanted to send.

‘The unit in which information is carried must be distinguished from the message as understood by the user. This is like the distinction between “segment” and “page” in a multi-access computer system. ... Smaller units for transmission must be distinguished and we shall call them “packets”.’

The subheadings of the paper were as follows:

- Introduction
- Characteristics of the existing networks, contrasting with the proposed one
- The significance of the real time use of computers
- Real-time business computer systems
- The organisation of the multi-access computer system
- Properties desired in a digital communication network
- Overall description of the proposed implementation
- Varieties of interface
- Estimation of Response time
- The manipulation of packets in queues
- Function of the interface units
- Line Concentration
- Error Control
- Redundancy features, dependability
- The Command Language
- Control of congestion
- Cost estimate
- Problems in the implementation of the new network
- The future of the existing networks
- Summary

An aspect of the design which did not appear in the previous papers was the ‘interface computer’ which mediated between the interconnected system of switching nodes and the users,

both packet and character terminals. The design of a multiplexer for 500 character terminals was considered and a PAD for 500–1500 users was thought to be about similar in complexity to the node computer. As many as 100 of these PAD units were envisaged for one interface computer.

In the section on cost it was estimated that the charge per packet would be in the neighbourhood of 1 penny (then equal to 1/240 pounds Sterling) but there were errors in the calculation and an erratum slip was inserted in which the cost for the node computers was estimated as contributing 0.01 pence per packet. The calculation was intended to refute the charge that message switching was expensive but it was stated not to include the cost of the interface computer and local distribution system. The handling capacity of a switching node was estimated as 2500 packets per second, which was over-optimistic.

In the summary it was envisaged that the telephone network might ultimately operate as a store and forward system, but this would be a long way ahead and should not affect the proposal made at that date.

The Post Office reaction at the March 1966 lecture and at discussions which followed was surprisingly tolerant towards this proposal. People in my own Department, the Ministry of Technology as it was then called, were also interested but it was never considered seriously as a practical possibility.

During this time I had a number of discussions with Derek Barber, who worked in the same Division at the NPL, about the possibility of building a local network to illustrate the principles in a practical way. In those days we had a certain amount of flexibility in our research programme enabling us to begin such a project without undue formality.

When I became Superintendent of the Division in August 1966 I was able to initiate a programme of research. We decided on a local network in order to test out our ideas for an interface computer since we believed that the nodal system would be beyond the funds we could employ and was susceptible in any case to investigation by computer simulation. At this point Keith Bartlett, Roger Scantlebury and Peter Wilkinson came into a small team and began to develop the local network design.

The first open publication was at the ACM Symposium on Operating System Principles at Gatlinburg in 1967. This was an important meeting, attended by many people who subsequently became engaged in packet switching. The paper described the network proposal and recalculated the per packet cost as 0.1 penny for the high level network which is much closer to present day tariffs. Some of the early work on the design for a local network was described.

Roger Scantlebury presented the paper in Gatlinburg and his report on the meeting gave us our first indication of the proposal for an ARPA network project. He reported that its design was unlike ours but thought that they might take up some of the ideas expressed in our paper.

At the end of 1967 I was invited to attend the final meeting of Special Study Group A of CCITT in Geneva which was formulating the recommendations that would go to the Plenary in 1968. At this meeting Fred Warden of IBM gave a long lecture on the future of data communication from the industry point of view. The Chairman of the Study Group, Mr. Rhodes of the UK Post Office, knowing of my proposals for a new data network approached me during an interval and asked if I would like to give a short lecture on the principles of packet switching. I was thus able to introduce packet switching to CCITT. At that meeting a resolution was passed for the establishment of a Joint Study Group on New Data Networks in the 1968–72 Session of CCITT. Subsequently this Group, GM-NRD, produced the first recommendations for new data networks in the form of lists of facilities and included packet switching as a subject for study, though the detailed study had to wait for the period 1972–76.

The ACM meeting in Gatlinburg had been a small meeting of specialists. The first opportunity for presenting our work to a wider audience was the 1968 IFIP Congress in Edinburgh where I gave an invited paper on ‘Communication networks to serve rapid response computers’ and a series of four papers were given by NPL authors in which the design of the local network, then well under way, was described in some detail.

The local network began operation in 1971 and the software was subsequently rewritten and its full operation as a service began in 1973. It is still² in full-time operation and has upwards of 200 terminals. At one time its traffic reached 1 million packets per day.

The simulation study of mean networks was begun in 1970 and has chiefly concerned problems of flow control and routing. These studies have continued until the present day.

The essential features of the proposals which were put forward in the 1967 and 1968 papers were contained in the privately circulated ‘Proposal for a digital communication network’ of 1966 and this itself was derived from the March 1966 lecture.

² 1982