ADVANCED TECHNIQUES FOR MICROPROCESSOR SYSTEMS

Edited by F.K. HANNA

(THE CAMBRIDGE RING - A LOCAL NETWORK)

A. Hopper

University of Cambridge

'ETER PEREGRINUS LTD.

CONTENTS

THE CAMBRIDGE RING - A LOCAL NETWORK

A. Hopper

University of Cambridge, UK

This paper deals with the design of local area computer communication networks and in particular with the Cambridge Ring. Such networks can be used to interconnect digital devices ranging from large mainframe machines to microprocessor systems.

INTRODUCTION

In the past few years a number of schemes have been developed for local communication networks. Such networks operate in a way which does not depend upon the characteristics of any particular device nor on the data being transmitted. They are capable of handling general purpose data link control protocols and are able to operate over a large range of propagation delays. They are designed at the bit rather than the character level, are capable of operating at high speeds and do not require sophisticated interfaces to the communicating devices. One of the most important design decisions is to what extent control functions are performed by the network itself and to what extent they are passed on to the attached host. Thus some hardware level protocols are defined which in turn are influenced by the traffic characteristics and the speed of the network.

Local Networks

The nodes in a local network can be interconnected in a number of ways (1). The completely connected network offers the lowest delays but is also expensive. A star network depends upon a fast central switch for routing but has fewer connections. The simplest is the ring network with no routing problems but with the largest theoretical delays (2,3,4,5). As local networks generally have to be cheap and simple the ring is often chosen as it fills these criteria well. Another type of network which does not perform any routing is the contention network which can be designed with carrier sense (e.g. Ethernet) or without carrier sense (e.g. Aloha) (6,7). Contention networks can be implemented cheaply and simply and as they can be used with radio as the transmission media are suitable for applications where the cost of laying a cable is high. A difficulty which all systems have to overcome is that of 'hogging'. This is a situation in which a particular station transmits so much that other stations are unable to do so. Ring networks overcome hogging by either passing around a 'permission token' or by insisting that empty slots make a complete revolution and are not used again immediately. Contention networks allow a clash to occur but prevent hogging by making sure the probability of the same two stations transmitting simultaneously again decreases with each clash. The spectrum of local network architectures is shown in Fig. 1.

Once the structure of the network has been chosen its mode of operation has to be defined. One of the most important areas is that of addressing. There are two basic addressing schemes; position addressing and code addressing. In the former each node has a number of permanently assigned slots and cannot use others. In the latter each packet is prefixed by an address which defines the user of the slot at that time. Code addressing has the advantage of being flexible but does take up extra bandwidth. On the other hand with fixed assignment of slots the characteristics of operation of the network are better known.

The communication process between two nodes will generally consist of the transmission of a number of packets. If this is known in advance then a call can be set up. This can be done in three ways; by using a 'start of call' - 'end of call' protocol, by specifying a 'start of call' and either a count number or a continuation marker in each packet. In most cases the destination is made deaf to all but the originator of the call for some period of time. It is now possible to shorten the address fields of packets as it is only necessary to distinguish between the maximum number of set up calls and the surplus bits can be used for data. This can be extended to include conferencing where the packets are read by the destination and readdressed somewhere else, or forwarding where they are only readdressed. An interesting addressing scheme has been developed where each node possesses a number of addresses which correspond to some entity in the host such as a process and can be addressed individually (8). This has the advantage that the location of a particular process does not have to be known in advance. However, long addresses are needed and such 'processes' are only likely to move between nodes if they correspond to simple entities such as file names.

In a simple network a source node is only allowed to transmit to one destination at a time. Such a point to point mode of operation can be extended to allow interleaved multiplexing of packets to different destinations. This affects the way acknowledgements are handled and the way that the control bits are used. If the source does not know the speed characteristics of the destination then the network should not allow it to transmit in such a way that congestion develops. One way of doing this would be for the destination to signal when it is ready to receive again. When a large amount of data is being transmitted between a pair of nodes then the acknowledgement traffic can be reduced by acknowledging on a per block rather than per packet basis. Even more complicated systems can be devised where more than one packet can be in flight to a particular destination in which case the acknowledgements have to contain sequence numbers and care has to be taken with respect to the build up of undelivered traffic.

For contention type networks, errors are assumed to occur frequently and powerful error detection facilities are provided whereas ring type networks can be made relatively error free. Thus for the contention network it is essential to be able to restart a packet at the source while this need not be done for the ring. If the transmitted packet is retained then the operation of the network can be made autonomous with respect to the host and most hardware errors can be hidden from the outside world.

In some applications the services of a particular node might be required extensively. If calls are set up then that node might be blocked from some sources for considerable periods of time. Under these circumstances it is advisable to implement an algorithm where transmission requests to a busy destination node are queued rather than ignored to be later arbitrated on a random basis.

Further decisions that have to be made include those between duplex and half-duplex operation and between synchronous and asynchronous transmission. The costs of changing software in the host machines have to be considered and if these are high then a transparent protocol might be used.

Hardware and Interface Design

The hardware design goals for a local network are that it should be reliable, cheap, have few wires and consist of distributed identical units. It should not be dependent upon any modulation system or clocking technique. For these reasons it is desirable to treat separately the repeater (or regenerative transmission) section of the node and the logical section which deals with the operation of the network. The interfaces between these have to be defined as well as the interface to the local host. The logical section will be referred to as the station and the logic between the station and the host as the access box. The structure of such a system is shown in Fig. 2, the interfaces being based on the Cambridge Ring.

The network interface to the repeater consist of input from and output to the medium wires which carry the signal between the nodes of the communications system. The interface to the station consists of data out, data in, one control line for 'external data in' as well as clock and power lines. This interface has to operate in a fail safe manner so that if the station malfunctions it can be disconnected and the repeater will continue to operate normally.

Let us now consider the interface between the station and the access box. The most important thing to note is that the two systems will generally be operating asynchronously. This means that under some circumstances enough time has to be allowed for flip-flops to settle before being sampled. Another possible constraint is the wire count which may make it necessary to multiplex data along a bus. Each station has an address (generally hardwired) and can supply it to transmitted packets. Thus, on the receive side, the interface will consist of the data/address bus and the control lines for ready to receive, packet received (from station), read source and read data. On the transmit side the control signals are load destination, load data, transmit packet and packet transmitted (from station). A second data/address bus is used if duplex operation is required. If transmitted packets return to the source with response information then further control lines are provided on the transmit side to read this data.

Having considered the interfaces, different modulation and clocking techniques will be briefly discussed. The modulation technique should allow the signals to travel a maximum distance between repeaters, not be dependent upon any particular clocking system and have signals available to make clock regeneration easy. It should have maximum timing tolerance between digits and not require complex logic to implement. In a simple system the clock can be transmitted separately from the raw data. This minimises the bandwidth required but introduces a DC component and does not allow the use of transformers. Phase modulation gives AC balance and allows the use of transformers for common mode rejection and isolation but requires twice the frequency. Transformers can further be used to provide a separate path to a node and for distributing power to the essential logic of the repeater. This is necessary as no assumptions can be made as to the availability of power from any node and the network has to operate in a fail safe mode. Other modulation systems can be developed; for example, a four wire system can be used if the wires available are of the type used for duplex operation of teletypes. Such a system is similar to phase modulation with the signals along each pair being one bit time out of phase with each other.

Clocking for a simple system can be provided by employing a central master clock. Such a clock is easily adjustable and can be made reliable. As it is passed from node to node it has to be reshaped. This may result in a system in which the data is not DC balanced and so does not permit the use of transformers. A better solution is to use a phase-locked loop (PLL) at each stage. This assumes that there are pulses on the ring at all times and that a suitable signal can be decoded from the input for controlling the PLL. The PLL has a certain pull in range and providing the incoming signal is within that range the PLL will look on. As a number of PLL are connected end to end the amount of jitter increases and when connected back on itself the signal might stray out of range. This is unlikely to happen for a small ring. There is a trade off between oscillator pull in width and jitter. Furthermore the performance of a PLL can be improved by increasing the PLL bandwidth, however this has the effect of increasing the effect of additive noise and so an optimum value has to be chosen for the open loop gain.

Another way of managing the clock is to provide each repeater with a separate monostable of uniform time constant. As the system is started a pulse train is introduced which has the effect of triggering the monostables one after the other at the introduced frequency. When the pulses have travelled once round the ring the pulse train input can be removed and the clock pulses will continue to circulate. This system keeps the clock one-shot period constant (and allows easy adjustment of the mark-space ratio for use in simple PM systems) but will probably introduce more jitter than the PLL and may be unworkable for a large number of repeaters without an 'elastic buffer'.

THE CAMBRIDGE RING

The data ring at Cambridge was designed to provide a high speed, low error rate communications path between computers and other devices in the Computer Laboratory. The primary uses of the ring are for equipment sharing, file sharing and dumping, and for research into distributed systems.

The original design was based on the register insertion principle where the packet to be transmitted is placed in a shift register which is inserted in series with the ring at the appropriate moment in time. As the delay in inserting a register and thus transmitting is at most one packet time, hogging does not occur and bandwidth is distributed to all nodes symetrically.

In due course it was realised that a more attractive system would be one based on the empty slot principle. In its simple form the empty slot system suffers from hogging. This defect can be overcome if each packet makes a complete revolution of the ring and is not marked empty until it has passed the original source. With this scheme the interaction with the ring at each node is minimised and reliability is improved. As performance characteristics of the two systems are very similar the empty slot system was adopted as the basis for the Cambridge ring.

The structure of a node is as shown in Fig. 2. In addition to repeaters, stations and access boxes there is a unique station called the monitor station. This station is used for setting up the slot structure during turn on, for monitoring the ring and clearing lost packets, and for accumulating some error statistics. Immediately preceding the monitor station there is an error logging station which uses a normal station and receives packets containing error information sent from active ring stations and the monitor station.

The packet structure is shown in Fig. 3 and is chosen to allow the maximum timing tolerance and minimum delay at the transmitter and receiver. The leading bit is always a one and is followed by a bit to indicate whether the slot is full or empty. Now follows a control bit used by the monitor station to mark as empty packets which are circulating indefinitely due to an error in the full/empty bit. This is followed by 4 eight bit bytes the first two of which are used for destination and source addresses and the last two for data. Finally there are two control bits used for acknowledgement purposes and a parity bit used for ring maintenance.

When a station has a packet ready for transmission in its shift register it waits until the beginning of the next slot. It now reads the full/empty bit and at the same time writes a one at the output. If the full/empty bit was a zero it transmits the packet, if however the full/empty bit was a one the slot is already occupied and the algorithm is repeated for the next packet. This scheme minimises the delay at each node.

The transmitted packet makes its way to the destination where the control bits are set on the fly to indicate accepted, busy, or rejected. It now returns to the source where the slot is marked empty. If the packet returns with the control bits unchanged it was not recognised by any destination. Each station knows the total number of slots in the ring and can thus clear the full/empty bit immediately.

Thus on transmission the packet is delayed until an empty slot is found but then the transmission is rapid. This is in contrast to the register insertion scheme where the delay round the ring can be large but the initial delay before the register is inserted is small. As the destination does not explicitly signal when it is ready to receive the next packet the ring can easily become clogged with packets returning marked busy when devices with varying speed characteristics are being interconnected. In order to overcome this the following algorithm is incorporated in the hardware to reduce the number of busies. If a source transmits a packet and it returns marked busy then it is not allowed to retransmit it until some time later. This additional delay is dependent on ring loading and is approximately the time to acquire the next empty slot. If any further retransmissions are attempted the extra delay is increased to about 16 times the original delay. Thus the number of busies is decreased and performance is improved.

Each station possesses a station select register which is initialised by the host. This register can be set to accept or to reject all packets, or to receive from one source only. When combined with a time out mechanism it can be used to allocate resources on the ring.

Maintenance and Error Recovery

There are no end-to-end CRC or parity checks on transmitted packets; however, a copy of the information is retained at the source and is compared with the returning packet. This provides a powerful error detection facility but does not indicate that the packet was correctly copied at the destination.

If one of the SOP bits is corrupted or the full/empty bit becomes full then this will be detected and corrected by the monitor station. If full becomes empty then the packet might be ignored at the destination but this will be detected by the source. Similarly the transmitter will detect if the monitor station bit becomes corrupted in such a way that the slot is marked empty. An error in the address fields may cause the packet to be delivered incorrectly or be assigned to the wrong source. An error in the response bits might have a more serious effect as it will not be detected by the transmitter, which might repeat the packet or assume it was received correctly when this was not the case. Errors are generally detected by the source or monitor station within one ring delay.

The Cambridge Ring provides powerful maintenance facilities for localizing transient faults and ring breaks. This is done by using the parity bit at the end of the packet. Each station continually computes the parity of every passing packet and empty slot. This parity is written into the parity field, which is simultaneously sensed, thus not increasing the delay. If the new parity does not match the old, a fault has occurred. Since the correct parity is inserted at each station, the fault must have occurred since the last active station. Each station requires two parity circuits, one to check the incoming packets, and one to produce a parity for an out-going (and perhaps changed) packet.

Thus, the ring is continuously monitored for errors and each detected fault is located to the nearest active downstream station. The error information is then transmitted to the monitor station by inserting into the next empty slot a packet which is entirely zero apart from the source field and the full bit. Thus the monitor station bit is also cleared so that this packet can be sent independently of the transmission shift register and is removed by the monitor station. The fault message itself may become corrupted, giving rise to further valid fault messages; nevertheless, the indicators reaching the monitor station will at least be correct for the nearest fault.

The above scheme is also used to detect ring breaks. This is done by arranging that the phase-locked loop at each repeater continues to operate in the centre of its frequency range when there is no data at the input. This is interpreted as a string of zeros, and the station is made to give a repeated fault message packet to which other repeaters downstream synchronise. Thus, when the ring is completely broken and unable to be used for data transmission the forward data path is used to send fault localizing messages.

Hardware

The ring is built using TTL technology and operates at 10 MHz with a maximum distance of 100 meters between repeaters. Higher data rates would be readily attainable with faster logic. The signals are transmitted along twisted pairs of the type normally used for duplex operation of teletypes. Transformers are used throughout for isolation and common mode rejection. As the repeaters have to operate reliably and whether they are connected to a station or not, they are powered directly from the ring. This power is injected into the system at a number of independent points.

Each station is fully duplex so that it can transmit, and receive, concurrently and independently. The number of bits delay at a station is a fraction of a bit and the minimum ring delay is about 5 microseconds.

Several modulation techniques were considered. A four wire scheme was chosen as it is suitable for a pair of twisted wires and has no ambiguity about the start of a digit (unlike phase modulation). A change on both pairs indicates a one and a change on only one pair indicates a zero, each pair being used alternately.

A number of different access boxes have been developed. Their complexity varies according to the speed of the host, and according to the level of buffering. Access boxes can be grouped as implementing three types of interfaces; polled, interrupt and direct memory access (DMA). A polled interface is the simplest and is suitable for attaching microprocessors to the ring. Each control signal at the hosts/station interface is memory mapped in the microprocessor and data transmission and reception consists of writing and reading

from memory. Such microprocessors have been developed in the Cambridge Ring as printer controllers, VDU concentrators or where a simple data processing function is required. An interrupt interface is the simplest way of connecting a minicomputer. If the interrupt is at the macro-rogram level the data transmission rate may be low. However, this may be remedied by servicing the interrupt in the micro-program. DMA operations can be performed by a simple microsequencer or by a more sophisticated device which buffers a number of ring packets and writes them to store in one block. This is suitable where high performance is required, and a system of this kind has been developed based on the 8X300 micro-processor.

Discussion

The most important thing to note about the Cambridge Ring is the high bandwidth and low error rate. At present the ring contains some fifteen repeaters and twelve active stations and is approximately 600 meters long. The system bandwidth excluding overheads is 4 Mbits/sec and the maximum point-to-point transmission rate is 1 Mbit/sec. The error rate is of the order of 1 error in $5x10^{11}$ bits which allows protocols to be simple because faults are infrequent and can be handled by repeating at a high level.

The maximum point-to-point bandwidth is a strong function of ring size and in particular of the number of circulating slots. Because of the round robin protocol and synchronisation considerations a station can transmit only once per n+2 slots (ring size = n slots). As extra nodes are inserted n may increase and thus the point-to-point bandwidth may decrease. It is interesting to note that increasing the basic data rate, for example from 10 Mbits/sec to 20 Mbits/sec, does not improve the point-to-point performance, since the number of circulating packets will double although the system bandwidth will increase by a factor of two.

To conclude, it is interesting to consider additions which could be made to the basic design to enhance its specification. One such option is the number of data bytes in a packet. By increasing this number the point-to-point bandwidth may be increased at the expense of additional delay to other transmissions. Another enhancement would be to include a number of user settable control bits in each packet. These would serve to identify control packets which at present are marked by a (non unique) bit pattern in the data field. Another option could be a broadcast feature where each station recognises a particular address. However, this may lead to problems because it is difficult to determine which stations copied the data. Finally, the maintenance features could be upgraded so that maintenance packets are sent when a station turns on or turns off, when the number of slots changes,when there is a modulation error or when there is fault in the bit-by-bit comparison of the transmitted and returned packet. These options are being considered and may be included in a new version of the ring based on LSI.

ACKNOWLEDGEMENT

The material in this paper is based on references 2,3,4,5. The author would in particular like to thank Professor M.V. Wilkes and Professor D.J. Wheeler for numerous and exhaustive discussions on the subject of local area computer networks.

REFERENCES

1. D.C. Clark, K.T. Pogran and D.P. Reed, "An introduction to local networks" Proc. IEEE vol. 66, pp. 1497-1517, Nov. 1978.

2. M.V. Wilkes and D.J. Wheeler, "The Cambridge digital communication ring", Local Area Communication Networks Symposium, Mitre Corp and National Bureau of Standards, Boston, May 1979.

3. A. Hopper, "Data ring at Computer Laboratory, University of Cambridge", Report of workshop on local area computer networking, National Bureau of Standards, Aug. 1977.

4. A. Hopper, "Local area computer communication networks", Tech. Rep. 7, Univ. of Cambridge, Computer Lab., Apr. 1978.

5. A. Hopper and D.J. Wheeler, "Maintenance of ring communication systems" IEEE Trans. on Comm. Vol. Com-22, No. 4., Apr. 1979.

6. N. Abramson, "Another alternative for computer communications", FJCC, 1970, pp. 695-709.

7. R.M. Metcalfe and D.R. Boggs, "Ethernet: Distributed packet switching for local computer networks", CACM, Vol. 19, pp 395-404, July 1976.

8. P.V. Mockapetris, M.R. Lyle and D.J. Farber, "On the design of local network interfaces", NCC, Aug. 1977.

9. H. Sullivan and T.R. Bashkow, "A large scale, homogeneous, fully distributed parallel machine", Comput. Arch. News, vol. 5, March 1977.

10. A. Hopper and D.J. Wheeler, "Binary routing networks", IEEE Trans. on Com-uters, Vol. C-28, No. 10, Oct. 1979.

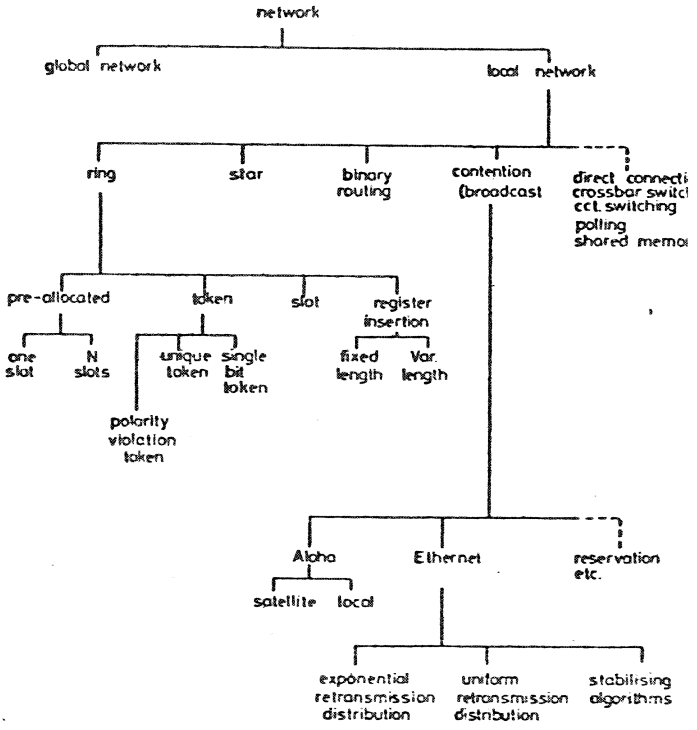

Fig.1 Local network achitectures

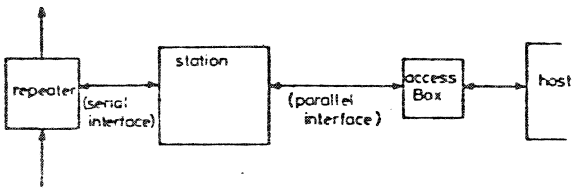

Fig.2 System structure

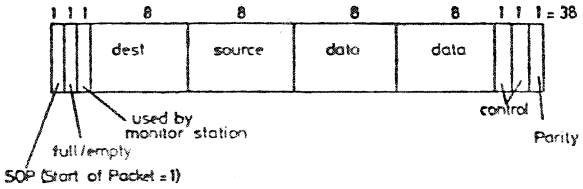

Fig.3 Cambridge ring packet format

Published by Peter Peregrinus Ltd., Stevenage, UK, and New York

ISBN: 0 906048 31 1

Printed in Great Britain by A. Wheaton & Co. Ltd., Exeter