Kcs p156

BOOKLET TA – 4

ifip congress 71
ljubljana – august 1971

HARDWARE AND SYSTEMS

CONTENTS

STAR-RING: A COMPUTER INTERCOMMUNICATION AND I/O SYSTEM

J.N. Potvin, P. Chenevert, K.C. Smith and P. Boulton
Computer Systems Research Group
University of Toronto, Toronto, Canada

This paper describes a novel hardware and software system which facilitates interconnection of several computers and peripheral devices. The system provides communication between programs in separate computers and also allows each machine to access the peripheral devices of any other. The latter is accomplished entirely from the computer requesting such foreigh I/O service, without the need for programming of the foreign computer.

special

The hardware and the program structure are not limited to the particular context of the prototype system, and with appropriate device interface boards and program translation, can be used for almost any system of interconnected machines.

INTRODUCTION

The Star-Ring System was originally conceived and built as a means to facilitate the access by one digital machine to the peripheral devices of another. However, once the Star-Ring concept had been developed it became apparent that the idea had much more general applicability.

For example Star-Ring provides a convenient mechanism for expansion of the I/O system of an established machine. The need for expansion may arise simply to accommodate the attachment of additional peripherals beyond the normal complement. A more likely possibility is the desire, often based on economic arguments, to connect a variety of devices foreign to the original system. In either case Star-Ring provides a unified approach to solution of both the hardware and software problems which result.

the

MOTIVATION OF THE ORIGINAL PROBLEM

Many computing centers have more than one computer in their facility, each with its assorted peripheral devices. Some of the computers may be very small. For such machines as the PDP8/1, a reasonable collection of peripherals usually costs several times more than the central processor. However, in the environment, there may exist one or more larger machines for which the overwhelming peripheral cost is better balanced by a complex CPU.

Currently, the peripherals of one machine in no way enhance the power of any other. Thus, duplication of facilities and large costs result if each machine is to have complete I/O capability including discs and card readers for example.

The purpose of this paper is to introduce Star-Ring, a unified hardware-software system, which enables the I/O capability of one machine to be utilized by all other machines in the system. Through the symmetries inherent in the single connection of each machine and its peripherals to the system, each is able to communicate with every other.

POSSIBLE APPROACHES TO A SOLUTION

1. Individual Connection

There exist many possible schemes for the connection of a small computer (B) to a larger one (A). For example, the large computer may simply see the small computer as a standard peripheral in its I/O subsystem (see Fig. 1). If a second digital device (C) is to be connected to the larger computer (A), a second interface must be built [1,2].

This is wasteful in two ways. First, each additional device (C) interfaced to the larger computer (A) has access to B only through the larger computer (A). The direct connection of C to B, shown in Fig. 1, avoids this problem but introduces a second. The resulting number of interfaces needed for a reasonable collection of heterogeneous devices may become wastefully large.

A general solution to both problems exists classically. This solution reduces the number of interfaces to one per machine:

2. Bus System

In this scheme, each machine is interfaced to a common bus shared by all machines (see Fig. 2). This arrangement has the advantage that each machine needs only one interface to the common bus in order to be interfaced to all other machines. Also, it is possible for each machine to communicate directly with every other machine.

However, only two machines can communicate at any one instant, and several problems thus arise. First, the total combined speed of data transfer is limited by the I/O speed of the slowest device, in addition to the inherent delay in the distributed I/O system. The latter results from propagation delay time, address decoding time, etc., along the common bus cable running from machine to machine.

Secondly, because of this limited speed during connection, the average data rate from any one machine to another is reduced when all machines are communicating in a multiplex mode on the bus. For example, if five conversations were under way, then the speed of each conversation is reduced to one fifth of what it would be if the connection were direct and not time-shared on the bus [3].

by

A further problem is lock-up. It occurs when two machines address each other and find each other busy. This problem can be solved by a cyclic transfer of control along the bus. This transfer of control must occur, however, at bus speed. This reduces further the overall system speed.

The following scheme solves the three problems while maintaining the advantages of a bus organization:

3. Star-Ring System

The Star-Ring system shown in Fig. 3 maintains the advantages of the previous systems, while discarding their disadvantages and introducing no others.

Interfacing complexity is ~~reduced~~ kept low, since only one interface for each machine is used. Through this interface, all other machines are accessable. The machine need only present on the parallel bus a destination address, a few bits of destination-oriented control if necessary and a few bytes of data. It must then signal the control unit that this information is set up in certain buffers. The control unit transfers the data and signals when the operation is complete.

At the receiving end, the control unit accepts data from the sending control unit and places it in buffers. It signals the destination digital machine that (4 bits of) source address, (4 bits of) control information, and (16 bits of) data are ready in certain buffers. The digital machine then takes the data from the buffers in words whose size depends on the word size of the machine.

Lock-up situations are avoided by rotational transfer of control around the ring. Control is transferred asynchronously from each control unit to the next when it has completed its control task. Since the ring is arranged to be physically small, high speed control transfers as well as data transfers are possible [4].

Data rate is kept very high by dividing the system into slow and fast areas of data transfer. The Ring can be very fast. In one parituclar implementation it transfers 24 bits of informat-ion in about 100 ns from one control unit to another. The branches of the Star are as slow as the I/O system of the digital machine and the necessary cable transmission system.

Unlike the normal bus system, slow branch data rates do not interfere with overall operation. The path that is shared by all branches (i.e., the Ring) is high speed.

Since each word transferred occupies the Ring for only about 100 ns, 8 digital machines in the system will permit a 1 MHZ word rate from one control unit to another. Thus an 8 unit system introduces a delay of about 1 microsecond into data transmission where data rates in the I/O subsystem are less than 2 Mbyte/sec. The upper bound of 2 Mbyte/sec is not unreasonable for even the highest speed digital I/O systems available today.

With this organization, non-overunnable devices can operate in a completely duplex basis with any other machine in the system. It is recognize, however, that overrunnable devices, such as a disc, etc., once data transmission is started, will need assurance of contiguous access to the input system of the receiving digital machine, without other devices on the ring interfering. A preselected mode (similar to the IBM I/O burst mode) of operation is possible in which the destination machine is temporarily dedicated to reception of data from a particular source machine. This connection is established when both machines concur and not until. Thus, even unbuffered digital devices may be connected to the system.

THE PROTOTYPE SYSTEM

A prototype system has been constructed along the lines of the previous general arguments. It consists of an 8 arm Star-Ring to which are presently connected a large IBM 360/44, A DEC PDP8/I, a Calcomp microfilm Plotter and an Applied Dynamics hybridizable analog computer. Though the applications to which this configuration is adapted are many, two distinct hybrid computing arrangements emerge. The first of these is a major machine incorporating the connection between the 360/44 and the AD4 to be used for very large problems [5]. The second consists of the connection between the DEC PDP8/I and the AD4 to be used for initial setup of fundamentally analog problems by the PDP8/I, for digital diagnosis of the analog machine's faults and for problems where the simple decision capability of the PDP8/I will suffice. Of course by virtue of the Star-Ring connection and a special program arrangement in the 360/44 (to be described subsequently) the PDP8/I has access to the storage and I/O resources of the 360 while exercising the AD4 analog machine.

Each machine, for both of its directions of transmission, normally connects to the Ring through three cards, only one of which is specially designed to constitute the unique interfacing done for any machine. The other two cards of which there are 5 pairs (of a possible complement of 8) in the prototype system are the Buffer Board and the Ring Control Unit.

Ring Control

To obtain high speed, the 3 (8 bit) byte message exchanged between machines is transferred on the Ring as parallel data with control and status information necessary for interlocked operation. Figure (4) shows schematically the connection active in the transfer of data from a s source of data through the Ring Sender, along the bus to the Ring Acceptor and from there to the sink of data.

Close examination of Figure (4) will reveal that an active Sender does not simply transmit 3 bits of address to be detected by an appropriate sender, rather the active Sender decides to which Acceptor it must send, and captures it by pulling down on one select line of the 8 uniquely coupled to the Acceptors of the prototype system. No conflict is possible because only one Sender is active on the bus at any one time. This technique which is fairly economical for a limited number of devices and a short Ring bus, was found to minimize the time slot necessary for a single Sender-Acceptor exchange.

In order to eliminate conflict on the shared bus it is imperative that only one Sender-Acceptor exchange occur at a time. This is arranged by the special daisy-chain connection between Senders shown in Figure (4) which arranges to transfer control around the ring. Upon receipt of a ICI signal from the previous sender an internal decision is made to transmit or to merely transfer control. If data is ready for transmission a Sender-Acceptor exchange occurs and about 140 nsec elapse before control is sent to the next Sender via the ICO signal. In the event that no data is available the activated Sender will pass control in about 60 nsec via ICO. Accordingly the data rate for buffer to buffer data transfers through the Ring depends on the activity of the devices connected to the system. It ranges from 2.5 megabytes per second if all 8 devices demand half duplex service to 6.5 megabytes per second if only one Sender is active.

In addition to the straightforward transfer of data in the multiplex mode implied by this description a burst mode may be presented for those devices without the logic capability of a gneeral purpose computer. Though in this mode the 3 byte bursts of data to the preselected device are interspersed on the ring with bursts to other devices, no other Sender may communicate with the selected Acceptor. It is apparent for this and other reasons including physical disconnection or the powered-down condition of a connected device that an attempt to transfer data via the Ring will not always succeed. Such conditions are detected and appropriate action taken by virtue of 3 status bits always returned to the Sender by the Acceptor during each attempt at data transfer to an Acceptor. The Flow Chart in Figure (5) shows the entire decision process required for a complete Sender-Acceptor exhange.

Data Buffering

For each device interfaced to Star-Ring a Single Buffer board suffices for full duplex data buffering. Data is gated onto the internal Ring bus using open collector NAND drivers. Data is intercepted from the bus using the data input terminals of standard TTL D flipflops. A total of 40 flipflops suffice for single buffering for both directions of data transfer.

Ring Control Unit

The flow of data into and out of the data Buffer is controlled by the Ring Control Unit Board. The RCU is composed of two functional parts: The Sender and the Acceptor.

The Sender Control

The Sender in turn consists of 3 parts: the Sender Control, Sender Status and Sender Addressing sections. The Sender control utilizes an asynchronous decision technique to determine, at the time the Sender is given control, whether the data in the buffer is ready or not. If ready, an interlocked exchange of data and status may be initiated. If not ready, control is shifted to the next unit in the Ring.

The Sender Addressing section decodes the 3 bit address from the source device to select one of the 8 lines each connected to the Select input of one of the 8 Acceptors. At the same time that data is gated from the buffer onto the Ring bus the Sender presents its own address on 3 Ring bus lines to indicate the source of data to the selected Acceptor.

The Sender Status section receives status from both the destination Acceptor and the Source device. Depending on these inputs and others it generates its own internal status and outputs Sender Status to the destination Acceptor, to the Sender Control and to the Source Device as necessary.

The Acceptor Control

The Acceptor consists of 3 parts corresponding directly to those of the Sender. Acceptor Control determines the reaction of the Acceptor to selection by a Sender on the Ring bus. This reaction is conditioned by the nature of the Sender's request as well as the state of Acceptor Status. Acceptor Status manages the record of the activity of the Acceptor and the connected Sink device. In particular it maintains the record of preselection for burst access by a suitably coded Sender.

Device Interface

Device interfacing has been made quite convenient by virture of a standardized interface to the Control Unit and Buffer boards. Though the interface to a simple machine such as the PDP8/I occupies part of one 5 x 6 inch board, the complexity of the I/O convention of the IBM 360/44 necessitates the use of two such boards. ~~For~~ programming convenience the Star-Ring port has been interfaced so as to masquerade as a combination 2501/1403 reader/printer.

The particular interface design chosen uses totally separate facilities for Sender and Acceptor to utilize maximally the IBM Standard Interface and obtain full duplex operations.

SOFTWARE ASPECTS OF MACHINE INTERCOMMUNICATION

In a system of several interconnected computers, intercommunication is practical between the machines only if the software of each computer provides appropriate support. Unfortunately, due to the extreme variability in machine language ~~capabilities~~ and operating system characteristics among a collection of computers, the solution to the software support problem is not as elegant as the solution to the hardware problem.

The prototype system described earlier is an adequate example of this problem. Consider the PDP8/I. This is a small computer and in normal use, operates under the control of a relatively simple operating system. The machine is not multi-programmed and hence at any time only one program in the PDP8 would be interacting with the Star-Ring hardware. Compare this to the IBM 360/44. This machine is large, operating under control of OS/360 MFT, a complex operating system which (in this case) allows simultaneous processing of two problem programs. It is obviously desirable to allow both these programs to communicate simultaneously via Star-Ring. As another extreme, consider the Calcomp micro-film plotter. This is not really a computer, since it has no decision-making capabilities. Obviously it requires no software support of any kind.

Finally, it should be noted again that one aim of the Star-Ring system was to allow a program in one computer to 'acquire' use of peripheral devices connected to any other machine.

For this to be feasible, it must be possible without the necessity of user action at the computer which controls the devices of interest. Any procedures which require the user to simultaneously execute programs in both machines are intolerable.

The Hardware Solution to the Software Problem

A first solution to the communication problem is provided by the Star-Ring hardware itself. Since this was specifically designed to appear as a standard I/O device to each host machine, it follows that the standard I/O software of each machine can read or write to the Ring. By this simple means, a user program in the PDP8, for example, can prepare "~~card~~ images" in tape "Calcomp plotter language" which can then be written on the Ring which, to the PDP8, is a standard ~~card~~ punch. paper

This technique is quite appropriate to the smaller machines on the Ring, especially in communicating with non-programmable devices such as plotters if these are directly connected to the Ring. However it provides little assistance in the problem of communicat~~ions~~ing between two computers, and none in the problem of communica~~tions~~ting between a program and a peripheral device connected to another computer.

The Software Solution

Because of the dependence on the detailed characteristics of the operating system involved, the software solution can only be provided in a given context. The context chosen was obviously that of the prototype system, and in particular, the most complex computer of this system, the 360/44, was the base machine for the software design and implementation.

Of course the general structure and logic of the software has been designed to be appropriate to any other multiprogrammed machine which might be connected to the Ring. Indeed the lgoic is such that if two multiprogrammed identical machines were connected via Star-Ring, each would be equipped with identical software and could then communicate effectively. Two problems must be solved by the software:

1. Provision of access by any Star-Ring port to any 360/44 I/O device without operator or user intervention at the 360.
2. Provision of access by any 360 problem partition to any Star-Ring port, independent of any possible simultaneous use of the Ring by any other 360 problem partition.

Access to 360/44 I/O Devices

In order that this access can be provided without intervention, it is necessary that the Star-Ring software be permanently resident in core and that external computers pass requests for device access to this program via the ring.

Permanent residence is readily provided by using a small (10k byte) partition under OS MFT. All Star-Ring software resides in this partition, being loaded once at each system cold start.

This program continuously monitors the 360 Star-Ring input, awaiting requests for service. Such requests, which of course originate at other computers on the Ring, arrive in the form of a simplified version of the 360 Job Control Language statement known as the DD (Data Definition) statement. The DD statement is normally used by 360 programmers to request I/O devices and data storage space and was chosen primarily because of its familiarity. Any such statement that may arrive at the 360 from the Ring is analysed by the permanently resident Ring program to determine what service is being requested. Assuming the resources requested are available, they are acquired from O/S 360 and the data transfer with the Ring is initiated and serviced to completion by the Ring program. If the resources are not available, the service request is refused (to avoid system deadlock possibilities). The Ring program is capable of supporting 28 simultaneous services of this type with the seven other Ring ports.

It is readily apparent that the above description indicates that dynamic resource allocation of devices to the Ring partition is required. However, OS 360 MFT will not provide this (to avoid deadlock possibilities) [6]. Examination of OS 360 indicated that provided only one partition was granted dynamic resource allocation, deadlock could not occur. A mechanism was developed which allowed this dynamic allocation, with no changes of any kind required in OS/360.

Problem Partition Access to the Ring

It is relatively straightforward to provide a service routine in the permanent Ring program which handles transfer of data between a problem partition of the 360 and any Ring port. Such services must be initiated by the problem partition.

However, this requires inter-partition communication which is also prohibited by OS 360 (in order to protect one user from another). This prohibition is readily circumvented by placing the Ring program in supervisor mode. Then the Ring program can override all protection and accomplish the desired data transfers.

SUMMARY

This paper has described a novel hardware and software system which facilitates interconnection of several computers and peripheral devices. A prototype system has been ~~constructed~~ to interconnect a Calcomp ~~microfilm~~ plotter, an Applied Dynamics AD/4, a DEC PDP8/1, and an IBM 360/44. Tests of the prototype system have proven entirely satisfactory.

The hardware and the program structure are not limited to the particular context of the prototype system, and with appropriate device interface boards and program translation, can be used for almost any system of interconnected machines.

References

[1] R.P. Bianchini, "An Interface Between a Control Data 6000 Series Computer and a Honeywell 16-Bit Series Computer", Mathematics & Computers NYO-1480-119, Courant Institute of Mathematical Sciences, New York University.

[2] Beattie & Penstone, "Design of an Interface Between System/360 and PDP/8 Computers", CIPS Bulletin, Summer, 1968.

[3] J.M. Bennett et al, "A Grafted Multi-Access Network", Information Processing 68, North-Holland Publishing Co.
[4] Atkins & Nordmann, "Supplementary Material on Control point Logic Design", 69 Internal Document, DCS, University of Illinois, Urbana, 1969.
[5] "Input/Output Interface - Channel to Control Unit" - IBM Form A22-6843-3.
[6] IBM System/360 Operating System: Job Management, Program Logic Manual. Form Y23-6613.

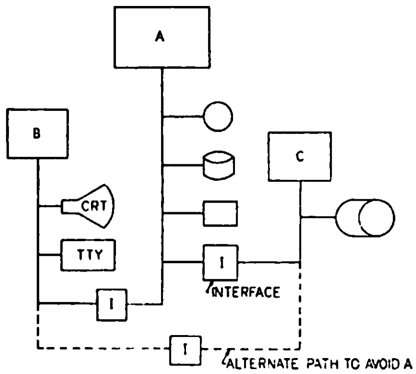

Fig. 1. Direct Connection

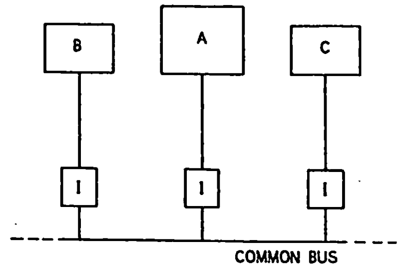

Fig. 2. Bus System

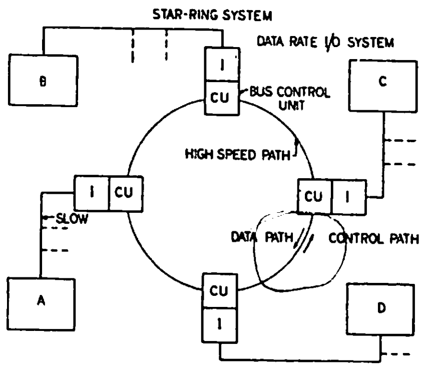

Fig. 3. STAR-RING System

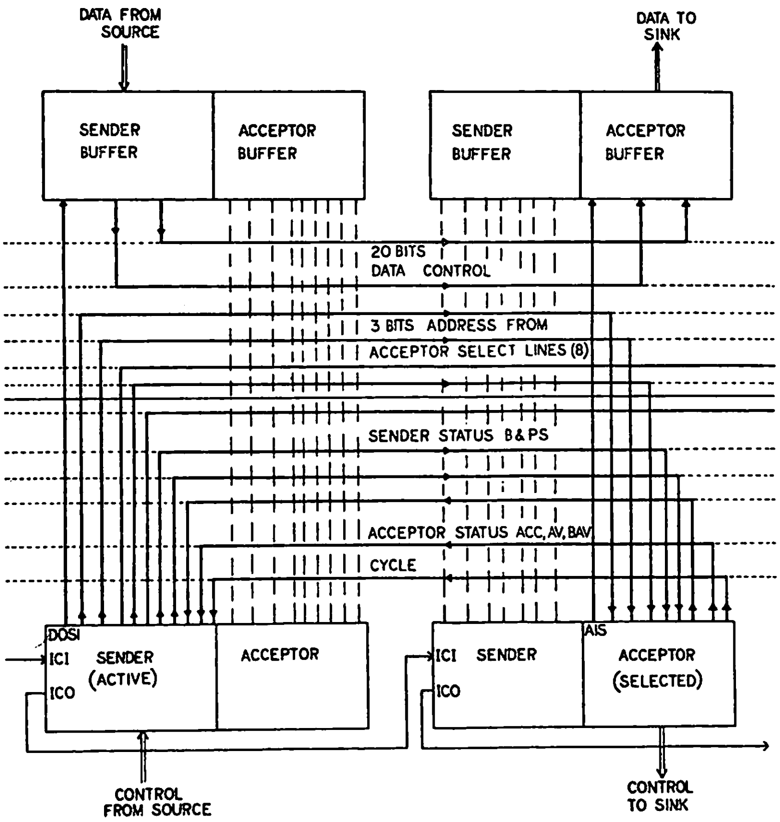

Fig. 4. STAR-RING Data Transfer

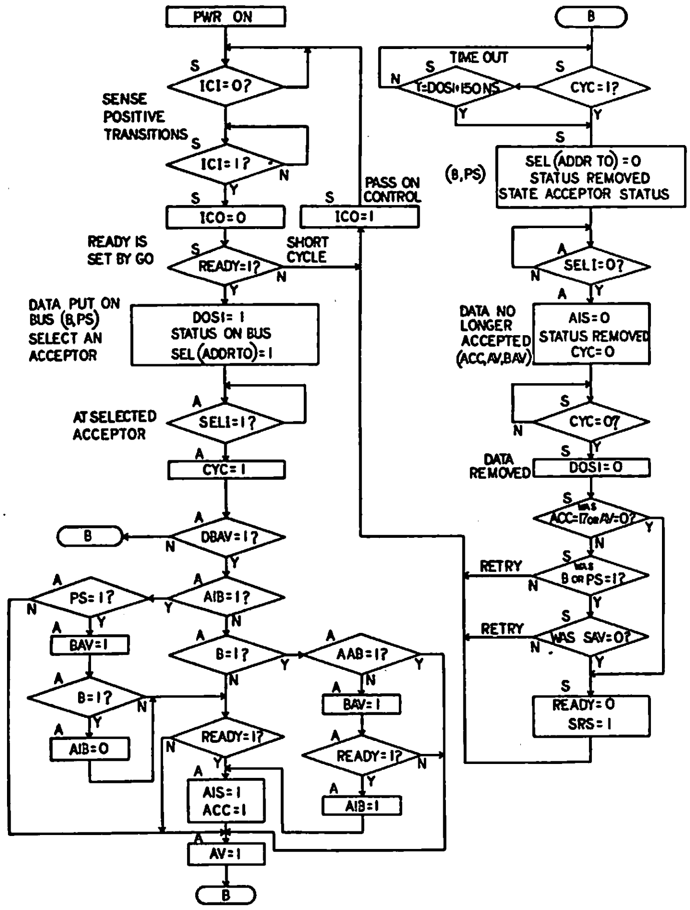

Fig. 5. Data Transfer Flow Chart