LABORATORY FOR
COMPUTER SCIENCE

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

Analysis of Structures for Packet Communication

Computation Structures Group Memo 151
August 1977

Robert G. Jacobsen
David P. Misunas

A paper to be published in the Proceedings of the 1977 International Conference on Parallel Processing

This research was supported by the National Science Foundation under grant DCR75-04060 and by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under contract number N00014-75-C-06661.

545 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139

ANALYSIS OF STRUCTURES FOR PACKET COMMUNICATION*

Robert G. Jacobsen
David P. Misunas
Laboratory for Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract -- In a system utilizing packet communication techniques of message transmission, all communication between the units comprising the system is through discrete blocks of information conveyed in packets. Interconnection structures in such systems can range from bus and crossbar structures to complex routing networks. A comparative analysis of a number of interconnection structures for packet communication systems is presented and tradeoffs between the various structures in terms of cost and performance are analytically examined.

Introduction

The increasing popularity of multiprocessor systems and the corresponding necessity for efficient interprocessor communication means has spurred the study and development of communication paths for use in such systems. One means for interprocessor communication which is gaining popularity is that of packet communication. In a system with packet communication architecture, the units comprising the system communicate through the transmission of discrete information packets [2].

Classical approaches to the design of communication paths have included such structures as busses and crossbar switching networks. These structures are necessarily small, due to the small number of interconnected units and due to the speed requirements placed on the structure. As the number of interconnected units increases, these structures become cumbersome both in size and processing capability.

More recently, a new interconnection structure, the routing network, has been presented and used in the design of a new type of parallel computer [3]. This structure is capable of simultaneously conveying many packets to their destinations in the processor and has a slower growth rate than the crossbar structure.

The tradeoffs between the various interconnection structures are not clearly understood. In the case of the routing network, little analysis has been performed at all. Detailed studies have examined such structures as the bus and crossbar [5, 9]. Some network structures have been studied [1, 8], particularly in the context of telephone switching networks [6, 7, 8]. However, these studies have generally considered only fixed connection circuits, rather than packet switching circuitry.

In the analysis of the present paper, we examine the characteristics of three communication structures: the bus, the crossbar, and the routing network. The cost and performance of each structure is analyzed to yield results as to the various tradeoffs involved in the choice of one structure over another. The analysis of these interconnection structures is supported through simulation results obtained on a packet communication simulation facility.

System Architecture

The design of a system interconnection structure is a difficult and poorly-understood problem, generally relying heavily on the experience of the system architect. There are no rules or guidelines for one to follow in such an exercise, merely a few general philosophies. In the following paragraphs, we will examine this situation more closely in the context of a packet communication system.

A packet communication system generally has some structure similar to that shown in Figure 1. The units comprising the User of Figure 1 may be processors, memories, functional units, or any other devices capable of message transmission or reception. The Communication Network of the system provides a path between the various units of the User. This interconnection structure may provide a path from every unit to every other unit, from groups of units to groups of units, or from each unit to one or several of the others. For the purposes of this discussion, we will assume the most general case; that is, every unit of the User can communicate with every other unit through the Communication Network. Other interconnection schemes can be considered as being composed of a number of embodiments of this more general case.

*This research was supported by the National Science Foundation under grant DCR75-04060 and by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under contract number N00014-75-C-0661.

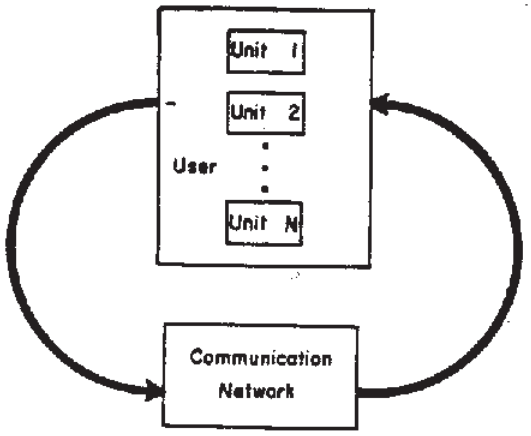

Figure 1. System Structure

Presumably, the designer of a packet communication system has an application area in mind for the system and has some idea of the amount of traffic which will pass over the communication medium. Thus, through some analysis, one should be able to generate a curve corresponding to the solid line of Figure 2. Such a user load curve expresses the number of packets generated as a function of the time required for an individual packet to transit the communication network and should always have a non-positive derivative, indicating that interunit communication will generally occur less frequently as the communication times increase.

On the other hand, the dashed curve of Figure 2 represents the load characteristics of the Communication Network and always has a non-negative derivative. The slope of the Communication Network load curve demonstrates that the load on the communication medium increases, the delay through the medium should eventually increase.

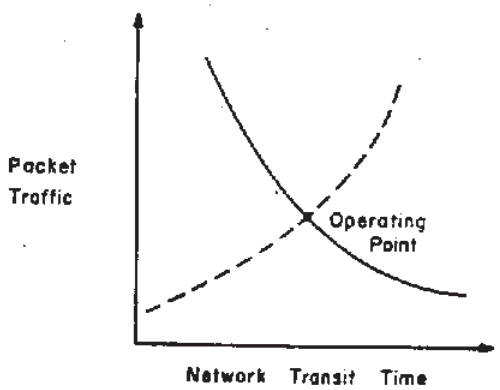

Figure 2. System Operating Characteristics

Generally, the two curves intersect at a point which will be the operating point of the system. Clearly, the system is only stable at the operating point and any digression from that point is countered by forces which tend to return the packet flow to the operating point.

Were it possible to empirically derive the User and Communication Network curves of Figure 2, the analysis and synthesis of packet communication systems would be greatly simplified. If there existed curves for the various types of interconnection structures, a designer need only develop the characteristic curve of his proposed User structure, choose a desired operating point on that curve, and match the appropriate Communication Network curve to yield the best cost/performance at that operating point.

Such a scheme may seem impractical, however, methods similar to this have been derived for many other branches of engineering, and there is no explicit reason why it is not possible to do so for aspects of computer design.

The remainder of this paper describes some preliminary results which were achieved while trying to generate load curves for various Communication Network structures. Whereas the achieved results do not yield rules for processor design, they provide a first step in that direction through the analysis of packet flow in the structures

Network Representation

The communication networks of the present study are formed of arbitration units and switch units. Each arbitration unit accepts the first packet to arrive at any input and passes the accepted packet to its output. In the case of conflict, one packet is arbitrarily selected and passed to the output before the other(s). Each switch unit transfers a packet on its input to one to its outputs, generally controlled by some switching specification contained in the packet.

The bus module of Figure 1 comprises an arbitration unit followed by a switch unit. Similarly, models for a crossbar and a routing network are shown in Figures 4 and 5. A network such as that of Figure 4 which is composed initially of switch units followed by arbitration units is called a distribution network, and a crossbar is one configuration of such a network. Similarly, a network which contains an initial stage of arbitration as that of Figure 5 is called an arbitration network.

The networks under study are structured as a number of stages connected in sequence. Each stage of a network is composed exclusively of either arbitration or switch units and is characterized by the log to the base N of the fanout/fanin ratio:

$$\log_N \frac{\text{(Number of Outputs)}}{\text{(Number of Inputs)}}$$

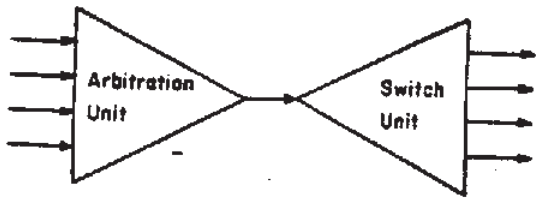

Figure 3. Structure of a Bus

This means of characterization has been chosen for two reasons. First, the size of the individual arbitration and switch units comprising each stage is clearly specified. Second, such a characterization represents a constant network architecture, regardless of the number of inputs and outputs.

The bus structure of Figure 3 (and all bus structures) is characterized by (-1, 1). Similarly, all crossbar structures are characterized by (1, -1). The "square-root" arbitration network of Figure 5 has the characterization (-1/2, 1/2, -1/2, 1/2).

Note that for an NxN communication network, the sum of all numbers in the network characterization must be equal to 0. Furthermore, in order for every input of a network to be able to communicate with every output, the sum of the absolute values of the numbers comprising the network characterization must be at least two. If the sum is greater than two, the network contains redundant paths.

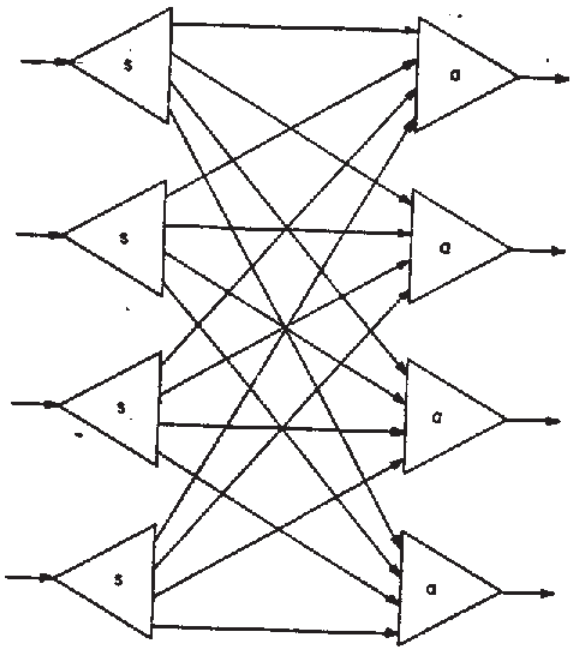

Figure 4. Structure of a Crossbar

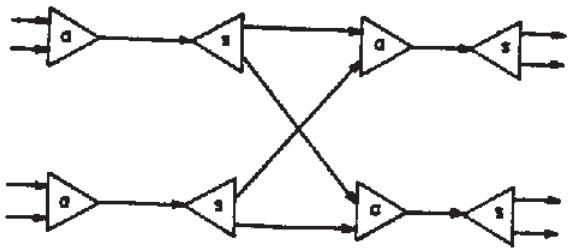

Figure 5. Structure of a Routing Network

At this point, we shall further restrict the networks under analysis to constant geometry NxN communication networks which can be characterized by a positive integer fraction f, where the network characterization is (-f, f, -f, f, ...) for an arbitration network or (f, -f, f, -f, ...) for a distribution network. The number of occurrences of f in each characterization is equal to the number of stages in the network, that is, to 2/f. Bus structures, crossbar structures, and simple power networks are examples of networks with such a characterization.

This restriction does not necessarily preclude the consideration in our model of networks which do not have alternating stages of arbitration and switch units. Without loss of generality, adjacent stages of the same type can be considered as one stage with a characterization which is equal to the sum of the characterizations of the two stages. However, the model described herein is only applicable to networks which can be characterized by a constant fraction f once reduction of identical adjacent stages has been performed.

Performance Analysis

For the purposes of finding the characteristic curve of a communication network, we need to make two simplifying assumptions. First, we consider the cost of a device proportional to the speed of the device times the number of wires connected to it. This assumption is not precisely accurate, but close enough for the purposes of this discussion.

Second, we assume that the packet distribution on the inputs of a communication network is even and Poisson and the distribution through any cross section of the network is even.

The communication networks under study are composed of an interconnection of one basic unit type, called a tie and consisting of an arbitration unit and a switch unit. The bus of Figure 3 is composed of one such tie. The network of Figure 5 can readily be seen to comprise a number of ties. Although the topology of a distribution network is slightly different than that of the networks in Figures 3 and 5, such a structure can be analyzed in a similar fashion.

We wish to examine two variables within each communication structure, a delay derater D and a loading representation F. D represents the average transit time for the network divided by the minimum transit time and can assume values ranging from one to infinity. D=1 signifies that the transit time through the communication network is only the hardware delay, whereas larger values of D indicate the presence of conflict in the structure.

F represents the fraction of the network that is not in use, that is, the free capacity of the network divided by the total capacity. In the following study, we examine D as a function of F to achieve each network characterization. The communication network load curve of Figure 2 represents a graphical depiction of a function similar to (1-F) vs. D. We have made this modification to the axis of the graph for the purposes of simplifying the analysis and the involved mathematics.

Representing the interarrival time on each input of an n-input tie by I and the service time by T, we find that a packet will arrive every I/n and hence:

$$F_{tie} = 1 - nT/I$$

Generalizing to all the units of a stage, a packet can be transmitted to the next stage at most every $T(n/N) = T(N^f/N) = T/N^{(1-f)}$. Thus:

$$F_{stage} = 1 - (T/N^{(1-f)})/(1/N)$$
$$= 1 - N^f T/I$$

Since all stages in this type of network are similary constructed:

$$F_{network} = F_{stage} = 1 - N^f T/I$$

The application of queueing theory techniques to the performance analysis of one tie, considering each tie as a queue and assuming Poisson arrival rates, yields the result:

$$D = 1 + (1-F)/4F$$

All ties in the network operate at the same F. Hence, overall, we can say:

$$D_{network} = 1 + (1 - F_{network})/4F_{network}$$

Simulation Results

Utilizing a packet communication simulation facility, a number of bus, crossbar, and routing network structures were simulated to see if actual performance followed the $D = 1 + (1-F)/4F$ formula. The simulation results are depicted in Figure 6.

The solid line of Figure 6 represents the graph of $D = 1 + (1-F)/4F$, and the points resulting from the simulation appear to observe this characteristic for the three structures under study.

The simulation modelled each network input as an independent source with a Poisson distribution and given interarrival time. The discrepencies of the simulation from the model for small values of F are due to the fact that the model contained infinite queues between the sources and the input ports, whereas such is impractical in the simulation, eventually causing the input queues to back up and affect operation of the sources.

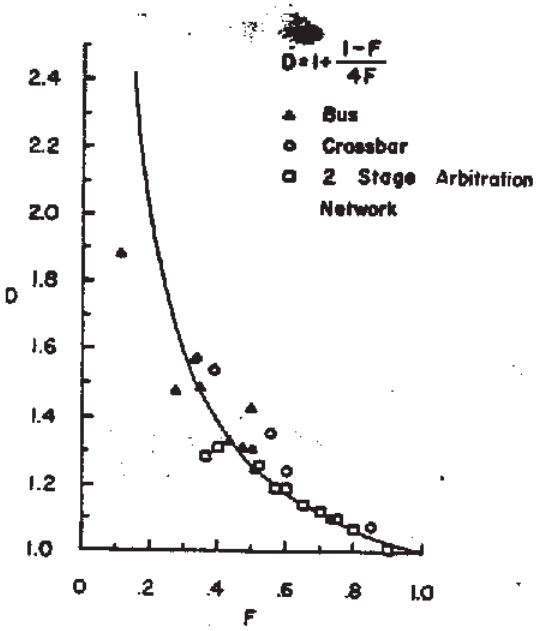

Figure 6. Simulation Results

Network Selection

The cost analysis for an arbitration network such as that of Figure 5 can be represented as follows, where C_{AN} is the cost of the network:

$$C_{AN} = (\text{number of stages})(\text{cost of each stage})$$
$$= (1/f)(\text{speed} * \text{number of wires})$$
$$= (1/f)([N^f/f] * N)$$
$$= N^{(1+f)}/f^2$$

In this case, speed is equal to N^f/f to maintain a constant average delay through the network with changes in f. The term N^f compensates for the increased loading of arbitration units due to the compression by N^f. The 1/f arises from the need for each stage to operate faster in networks with more stages.

In the case of a distribution network:

$$C_{DN} = (\text{number of stages})(\text{cost of each stage})$$
$$= (1/f)(\text{speed} * \text{number of wires})$$
$$= (1/f)(1/f) * (N^{(1+f)})$$
$$= N^{(1+f)}/f^2$$

A distribution network has a greater number of wires because each input wire of a stage of such a network is expanded to $N^{(1+f)}$ wires. Due to this

expansion, the component speed in a distribution network is only affected by the number of stages, that is, by 1/f.

Thus the linear cost assumption has led us to the conclusion that for some fixed performance, the arbitration network of Figure 5 costs the same as the distribution network of Figure 7. This result is non-intuitive at first, however, consider an arbitration network of complexity N. The units comprising this network have speed N due to the initial compression factor. The complexity of an equivalent distribution network is N^2, but the additional parallelism allows the network to be constructed of components with speed 1. Hence, the cost of the two networks is equivalent.

The minimum of the network cost $N^{(1+f)}/f^2$ occurs at

$$1/f = (1/2) \ln N$$

where 1/f is the number of stages. Hence, for the linear cost assumption of the model, the following structures are best suited for the specified number of inputs for either arbitration or distribution network:

N	Structure
7	1-stage networks (bus and crossbar)
50	2-stage networks
400	3-stage networks
3000	4-stage networks

An interesting result which arises from the performance computations is the determination of the optimal value of n, that is, the number of inputs to each arbitration unit and outputs of each switch unit. As we have seen, the minimum cost occurs when $f = 2/\ln N$. Thus, these expansion and compression ratios should be:

$$N^f = N^{2/\ln N} = e^2 \approx 7$$

To utilize the previously described results in the design of a packet communication system, one first determines the load curve of the units to be interconnected. The architecture of the communication network utilized in the system is specified by the number of units. With these specifications in mind, there are a number of design choices which can be made.

The load curves of the communication network consist of a family of curves which are parametric with cost. To design for a specific cost or technology, the intersection of that member of the family with the user load curve yields the performance which can be achieved.

Conversely, to structure the system for a specific performance, the desired operating point on the user curve is specified and the network curve which passes through that point determines the cost and speed necessary in the component parts.

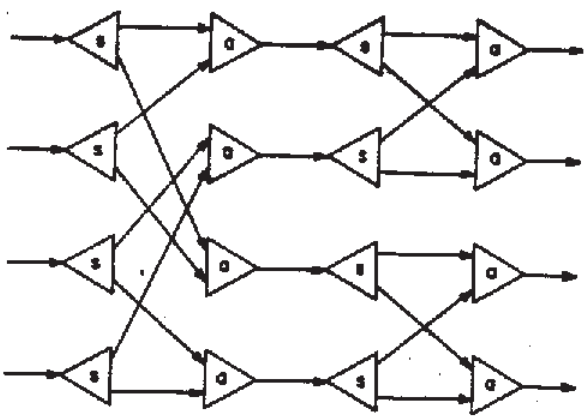

Figure 7. Structure of a Distribution Network

The choice of either an arbitration network or a distribution network must take into account important factors such as the available technologies. While these factors are not included in the model, they will dictate actual use of any results achieved therefrom.

Concluding Remarks

This attempt to probe the interconnection problem for packet communication systems has left many questions unanswered. The model utilized has a number of deficiencies and remains to be made more exact and extended to structures other than certain NxN power networks, such as asymmetric networks and concentration networks. Further refinement of the model and addition of other structures should provide much information useful in the synthesis of processor structures for packet communication. Despite its deficiencies, the model provides a first attempt to analyze such packet communication interconnection structures and yields some interesting insights into their behavior.

References

[1] Davidson, I. A., and J. A. Field, "Design Criteria for a Switch for a Multiprocessor Computing System," Proceedings of the 1975 Sagamore Computer Conference on Parallel Processing, IEEE, New York, (August 1975), pp. 110-114.

[2] Dennis, J. B., "Packet Communication Architecture," Proceedings of the 1975 Sagamore Computer Conference on Parallel Processing, IEEE, New York, (August 1975), pp. 224-229.

[3] Dennis, J. B., and D. P. Misunas, "A Computer Architecture for Highly Parallel Signal Processing," Proceedings of the ACM 1974 National Conference, ACM, New York, (November 1974), pp. 402-409.

[4] Marcus, M. J., New Approaches to the Analysis of Connecting and Sorting Networks, Research Laboratory of Electronics, M.I.T., Cambridge, Mass., Technical Report 486, (March 1972), 54 pp.

[5] Pearce, R. C., and J. C. Majithia, "Upper Bounds on the Performance of Some Processor-Memory Interconnections," preprint.

[6] Pippenger, N., The Complexity Theory of Switching Networks, Research Laboratory of Electronics, M.I.T., Cambridge, Mass., Technical Report 487, (December 1973), 51 pp.

[7] Pippenger, N., "On Crossbar Switching Networks," IEEE Transactions on Communications COM-23, 6 (June 1975), pp. 646-659.

[8] Thurber, K. J., "Interconnection Networks -- A Survey and Assessment," AFIPS Conference Proceedings 43, AFIPS Press, Montvale, New Jersey, (1974), pp. 909-919.

[9] Thurber, K. J., et. al., "A Systematic Approach to the Design of Digital Bussing Structures," AFIPS Conference Proceesings 41 Part II, AFIPS Press, Montvale, New Jersey, (Fall 1972), pp. 719-740.