OPTIMAL LOOP TOPOLOGIES FOR DISTRIBUTED SYSTEMS*

C. S. Raghavendra and M. Gerla

UCLA Computer Science Department
University of California
Los Angeles, CA 90024

ABSTRACT

Double loop network architectures offer higher performance and reliability than single loop networks. In this paper, a double loop network which is optimal among all double loops is described. This network topology consists of a loop in the forward direction connecting all the neighboring nodes, and a backward loop connecting nodes that are separated by a distance $\lfloor \sqrt{N} \rfloor$, where N is the number of nodes in the network. We show that this network is optimal in terms of hop distance between nodes, delay, throughput, and terminal reliability. The paper includes derivation of closed form expressions for the maximum and average hop distance between nodes, number of distinct routes between two farthest nodes, and throughput. The effect of node and link failures on network performance is also considered.

1. INTRODUCTION

Local area networks have been extensively used in recent years to support distributed processing. One of the key design issues is the identification of a network topology which is best suited for a variety of different applications. Loop or ring architectures are quite popular, since they require simple control software and provide high data transfer rates. Many loop network architectures have been described in the literature, some using centralized control, others using distributed control [FARM 69, PIER 72, OHLI 77]. In this paper, we consider a loop operated in the "check and forward" mode [GRNA 80]. Referring to Figure 4, the destination address of each received message is checked to determine the output channel (either to the local host, or to a neighboring node) onto which the message should be forwarded. If the channel is free, the message is forwarded "on the fly", without buffering. Otherwise, the message is stored in a delay insertion register, waiting for the channel to become free.

A major disadvantage of a unidirectional **single** loop system is vulnerability, since any link or interface failure partitions the network. Another disadvantage is the fact that some messages must travel along the entire loop before reaching their destinations. This affects the delay performance.

In order to improve **reliability** and **performance** of loop networks, double loop architectures have been proposed. Liu proposed a Distributed Double Loop Computer Network (DDLCN) [LIU 78, WOLF 79a], which is designed as a fault-tolerant distributed system, and claimed to have achieved better reliability as well as better communications performance than any other loops. More recently, another double loop network, called daisy-chain loop, was proposed by Grnarov et. al. [GRNA 80]. This network was shown to be superior to DDLCN, in terms of reliability and communications performance in both fault free and fault mode of operation [GRNA 80].

In this paper, we present a generalized double loop network, which is optimal among all double loop topologies. The DDLCN and daisy-chain become special cases of the proposed generalized network. We present the analysis of the optimal network and derive closed form expressions for maximum number of hops, average number of hops, and throughput. We also study the effects of node and link failures and show the improvements obtained with respect to the daisy-chain loop.

2. OPTIMAL DOUBLE LOOP NETWORK

The optimal double loop consists of a forward loop connecting all the neighboring nodes and a backward loop whose interconnection pattern depends on N, the number of nodes in the network. As shown later, for optimum performance and reliability the skip distance of the backward loop is $\lfloor \sqrt{N} \rfloor$, where skip distance is the hop distance (measured on the forward loop) between two nodes which are adjacent in the backward loop. Except for the fact that skip distance is a function of N, all other features are similar to those of the daisy-chain loop [GRNA 80]. The proposed topology for a 15-node network is shown in Figure 1, where the skip distance in the backward loop is 3. For comparison

* This research is supported in part by the ONR contract N00014-79-C0866
(Research in Distributed Processing)

CH1694-9/81/0000/0218$00.75 © 1981 IEEE

purposes, a 15-node daisy-chain loop network is shown in Figure 2, and a 15-node DDLCN is shown in Figure 3.

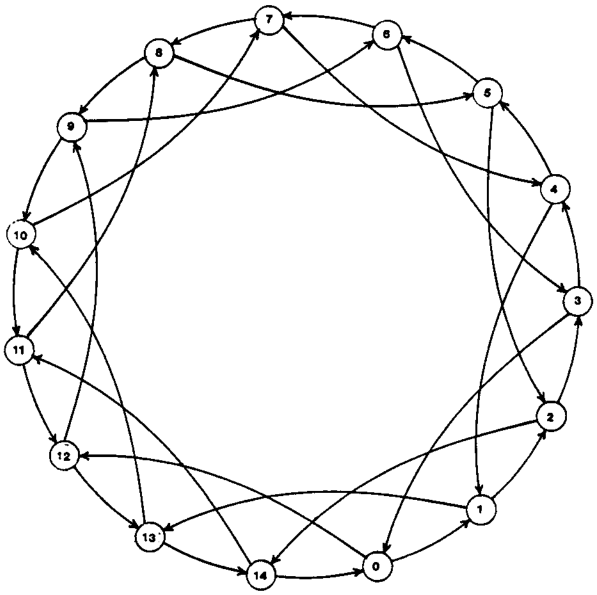

Figure 1. A 15-node Optimal Loop Topology

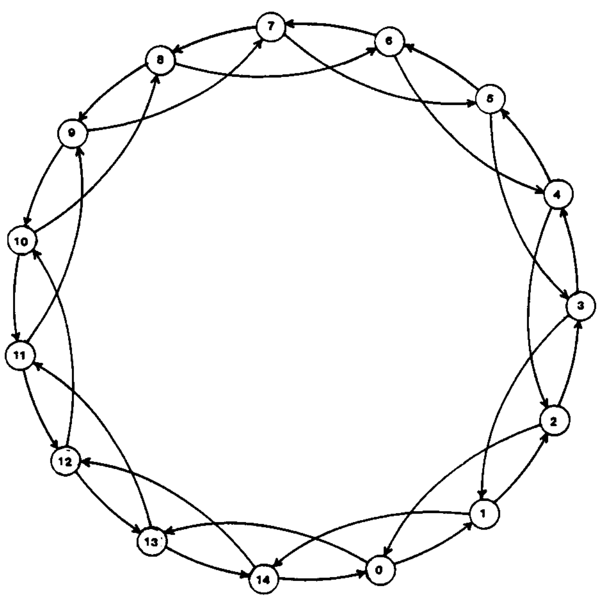

Figure 2. A 15-node Daisy-Chain Loop Network

In the optimal loop network, both forward and backward links are active, and there exist several paths from a source to a destination. The network can tolerate several link and nodal interface failures before becoming partitioned. Intuitively, delay and reliability are improved since the skipping of several nodes creates "short cuts", and also provides more alternate paths. We will prove this more rigorously later in the paper.

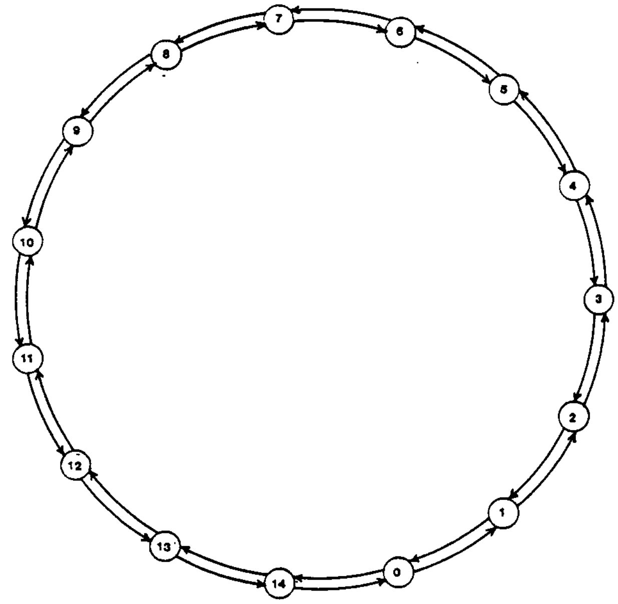

Figure 3.
A 15-node Distributed Double Loop Computer Network

The functional organization of a loop interface is shown in Figure 4. An interface receives messages from the host it is connected to. It also receives messages relayed by the neighboring interfaces. Conflicts of the simultaneous arrivals of messages from the three streams are resolved in the same way as in DDLCN [WOLF 79b], i.e., by delaying incoming relayed messages in variable-length shift registers located in the loop interface. To achieve better reliability, a loop interface is split into two identical modules which have separate control and separate line driver/receiver (D/R). Both modules share transmitter (T) and receiver (R) for communications with the host (H) connected to this interface.

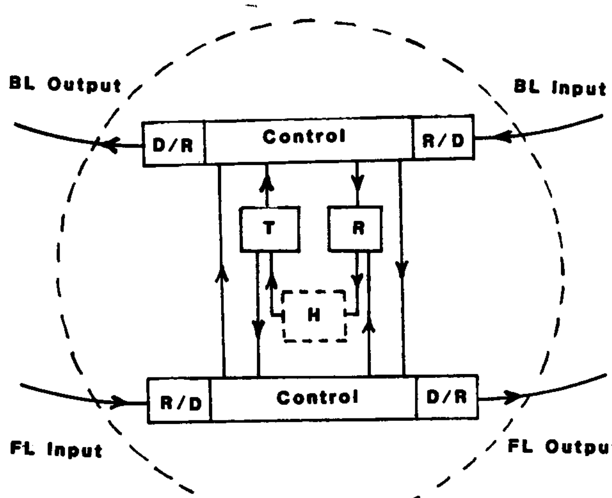

Figure 4. Functional Organization of a Loop Interface

In the message format shown in Figure 5, the destination address field is used in selecting the path. If the destination address matches with the interface address, the message is removed from the loop and routed to the host. Otherwise, the message is forwarded one "short" step (on the forward loop) or, when appropriate, one "giant" step (on the backward loop). Messages circulate on the loops until they reach their intended destinations. An adaptive routing algorithm can be used to route messages when there are failures in the network.

Flag	Destination address	Origin address	Control	Data	CRC	Flag

Figure 5. A Message Format

3. MAXIMUM AND AVERAGE NUMBER OF HOPS

Recall that both forward and backward loop are unidirectional, with the forward loop connecting adjacent nodes and the backward loop connecting nodes at hop (i.e., skip) distance $h > 1$. The shortest route between a source and a destination may therefore involve a combination of forward and backward hops. For each node, there is some other node which is diametrically opposite, i.e., it is the farthest away in the hop distance sense. Indeed, in designing the backward loop topology, the skip distance h will be selected such that the diameter d, i.e., the maximum over all node pair distances is minimized. Clearly, the hop distance from a node to all other nodes will vary from 1 to d. In the following we derive expressions for maximum and average number of hops and find their optimum values.

Let N be the number of nodes in the network;
h be the skip distance in the backward loop;
and d be the diameter.

We want to choose h so as to minimize,
d = diameter = max d_{ij} over all i,j.
where d_{ij} is the shortest hop distance from node i to node j.

The nodes in the network are numbered from 0 to N-1. To get the shortest path between i and j we use backward hops as long as it is advantageous. For two diametrically opposite nodes, backward hops are advantageous as long as d is greater than (N-bh), where b is the number of backward hops used. In addition to the backward hops, h-1 forward hops will be required to reach the farthest node. The relationship between maximum number of hops d and skip distance h is,

$$d = N - (d - h + 1)h + (h - 1)$$

$$d = N - dh + h^2 - 1$$

$$d = \left\lfloor \frac{N}{(h+1)} \right\rfloor + (h - 1)$$

For d to be minimum, the optimum value of h is $\lfloor\sqrt{N}\rfloor$.

If h = 1, we get Liu's DDLCN of diameter d = $\lfloor N/2 \rfloor$;
If h = 2, we get daisy-chain of diameter d = $\left\lfloor \frac{N}{3} \right\rfloor + 1$.

The average number of hops is defined as,

$$\frac{1}{N}\sum_{i=0}^{N-1}\left[\frac{1}{(N-1)}\sum_{j=0}^{N-1} d_{ij}\right]$$

where $d_{ii} = 0$.
As the network is symmetric, the term in parentheses is the same for all nodes i, thus:

Avg. no. of hops = $\frac{1}{(N-1)}\sum_{j=0}^{N-1} d_{ij}$ for an arbitrary i.

Thus we need to calculate the sum $S = \sum_{j=0}^{N-1} d_{ij}$ of the distances to all other nodes in the network. Note that we always consider the shortest distance between any node pair. Recall that the maximum number of backward hops used in any route is,

$$b = \left\lfloor \frac{N}{(h+1)} \right\rfloor$$

Using this result we have,

$$S = [1 + 2 + 3 + ... + (N - bh - 1)] + [1 + 2 + ... + h] + [2 + 3 + ... + (h + 1)] \; .... + [b + (b + 1) + ... + (b + h - 1)]$$

Each term in the sum corresponds to a route with 0, 1, 2, ..., b number of backward "giant" hops (i.e., steps on a backward loop) respectively.

Manipulating the previous expression, we obtain:

$$S = [1 + 2 + 3 + ... + (N - bh - 1)]$$

$$+ h(1 + 2 + ... + b) + b(1 + 2 + ... + h-1)$$

which simplifies to,

$$S = \frac{(N - bh)(N - bh - 1)}{2} + \frac{bh}{2}(b + h)$$

The average number of hops is given by S/(N-1), i.e.,

$$AVG = \frac{1}{(N-1)}\left[\frac{(N - bh)(N - bh - 1)}{2} + bh\frac{(b + h)}{2}\right]$$

Substituting the value of b,

$$AVG = \frac{1}{2(N-1)}\left[\left(N - \left\lfloor\frac{N}{h+1}\right\rfloor h\right)\left(N - \left\lfloor\frac{N}{h+1}\right\rfloor h - 1\right) + \left\lfloor\frac{N}{h+1}\right\rfloor h\left(\left\lfloor\frac{N}{h+1}\right\rfloor + h\right)\right]$$

For the sake of comparison, we report DDLCN and daisy-chain results. These are:

For DDLCN, AVG = $\frac{1}{(N-1)}\left\lfloor\frac{N^2}{4}\right\rfloor$

For daisy-chain,

$$AVG = \frac{1}{2(N-1)}\left[\left(N - \left\lfloor\frac{N}{3}\right\rfloor 2\right)\left(N - \left\lfloor\frac{N}{3}\right\rfloor 2 - 1\right) + \left\lfloor\frac{N}{3}\right\rfloor 2\left(\left\lfloor\frac{N}{3}\right\rfloor + 2\right)\right]$$

Next, we want to minimize the average number of hops, i.e, we want to minimize the following expression:

$$A = \left(N - \left\lfloor\frac{N}{h+1}\right\rfloor h\right)\left(N - \left\lfloor\frac{N}{h+1}\right\rfloor h - 1\right) + \left\lfloor\frac{N}{h+1}\right\rfloor h\left(\left\lfloor\frac{N}{h+1}\right\rfloor + h\right)$$

For optimality,

$$h + 1 = \frac{N}{(h+1)}$$

That is, $h = \lfloor\sqrt{N}\rfloor$

Note that the same optimum value of $h = \lfloor\sqrt{N}\rfloor$ minimizes both maximum number of hops and average number of hops. In Table 1 we summerize maximum and average hop results for DDLCN, daisy-chain, and optimal loop for typical values of N.

N	DDLCN Max	DDLCN Avg	Daisy-Chain Max	Daisy-Chain Avg	Optimal Loop Max	Optimal Loop Avg
10	5	2.80	3	2.33	3	2.33
15	7	4.00	6	3.21	5	3.00
30	15	7.76	11	5.69	9	4.65
50	25	12.75	17	9.00	12	6.14
100	50	25.25	34	17.33	18	9.09

Table 1. Maximum and average hop distances for three different loop architectures.

4. THROUGHPUT ANALYSIS

In this section we perform the throughput analysis to determine the total message traffic that the network can handle before saturation occurs. When link utilization reaches unity, the network is saturated. The total message arrival rate at which this occurs is the maximum traffic the network can handle, and is defined to be the throughput. For the analysis we make the following assumptions [KLEI 76]:

1) Poisson message generation with total aggregate rate γ, (msg/sec)
2) Uniform traffic pattern, and
3) Saturated throughput conditions, i.e., message traffic on each link equals link capacity.

We compute the maximum traffic in a way similar to Wolf's [WOLF 79b] relating each link utilization to the total message arrival rate and setting the maximal utilization to one.

We assume that shortest routes only are used for message transmissions. If there is more than one shortest route, traffic will be **equally** divided among the alternate routes. The total traffic on a forward link may be different from the total traffic on a backward link. As the average number of hops AVG is computed by considering the shortest routes from a node to all other nodes, the total network traffic caused by distributing γ/N units of traffic from one source to all other destinations is:

$$T_f + T_b = AVG^*\gamma/N$$

where T_f is the contribution to total traffic on a forward link and T_b is the contribution to total traffic on a backward link. By using suitable policies for routing traffic when there is more than one shortest route between a node pair, it is possible to balance the flow on forward and backward links. That is,

$$T_f = T_b = \frac{1}{2}AVG^*\gamma/N$$

When all sources are active, saturation is reached if:

$$(T_f + T_b)N = 2N\mu$$

where μ is the link capacity (msg/sec)

Substituting for T_f and T_b we obtain:

$$\gamma_{SAT} * AVG = 2\mu N$$

or:

$$\gamma_{SAT} = \frac{2\mu N}{AVG}$$

Finally, substituting for AVG we obtain:

$$\gamma_{SAT} = \frac{2(N)(N-1)\mu}{\frac{1}{2}\left[\left(N - \left\lfloor\frac{N}{h+1}\right\rfloor h\right)\left(N - \left\lfloor\frac{N}{h+1}\right\rfloor h - 1\right) + \left\lfloor\frac{N}{h+1}\right\rfloor h\left(\left\lfloor\frac{N}{h+1}\right\rfloor + h\right)\right]}$$

γ_{SAT} is maximum if average number of hops is minimum. Therefore, γ_{SAT} is also maximized when $h = \lfloor\sqrt{N}\rfloor$

We may compare this result with DDLCN and daisy-chain results.

For DDLCN, $\gamma_{SAT} = \frac{2(N)(N-1)}{\left\lfloor\frac{N^2}{4}\right\rfloor}\mu$

For daisy-chain,

$$\gamma_{SAT} = \frac{4(N)(N-1)\mu}{\left[\left(N - \left\lfloor\frac{N}{3}\right\rfloor 2\right)\left(N - \left\lfloor\frac{N}{3}\right\rfloor 2 - 1\right) + \left\lfloor\frac{N}{3}\right\rfloor 2\left(\left\lfloor\frac{N}{3}\right\rfloor + 2\right)\right]}$$

Since γ_{SAT} is proportional to the inverse of AVG, and AVG is minimized by our topology, the throughput is also optimal for our topology.

As a special case, when N is the square of an integer, we get the following closed form expressions for maximum number of hops, average number of hops, total flow, and throughput.

$$\text{Diameter d} = 2(\sqrt{N} - 1)$$

$$\text{Avg. no. of hops} = \frac{N}{(\sqrt{N} + 1)}$$

$$\gamma_{SAT} = \frac{2(N-1)}{(\sqrt{N} - 1)} \mu$$

For example, when N = 64 we have,

Diameter d = 14
Avg. no. of hops = 64/9
$\gamma_{SAT} = 18\,\mu$

When there is a link failure, the maximum and average number of hops will increase. However, the performance of the optimal loop can be shown to be still superior to that of DDLCN or daisy-chain loop in similar failure conditions.

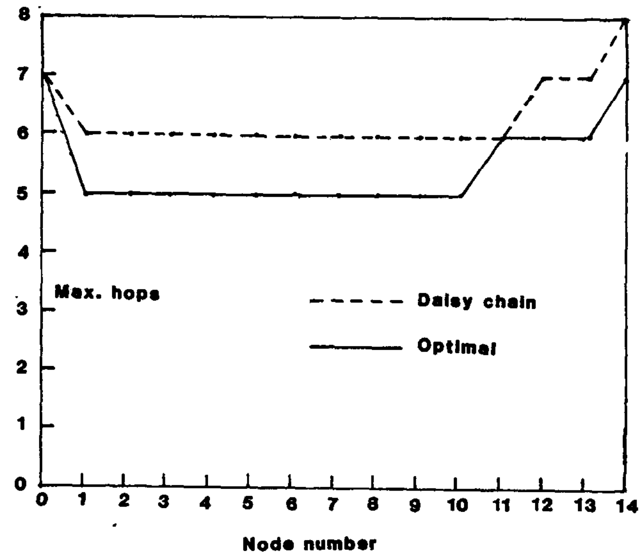

Figure 7.
Effect of a Link Failure on Maximum No. of Hops

5. FAULT TOLERANCE

The optimal double loop architecture provides good protection against node and link failures. As many alternate routes exist between any node pair, paths can be selected adaptively in case of failures. The effect of a node failure is to increase the maximum number of hops for some nodes and increase the average number of hops slightly. Figure 6 shows the effect of node #0 failure on optimal loop and daisy-chain loop, and Figure 7 shows the effect of link 0→1 failure on optimal loop and daisy-chain loop.

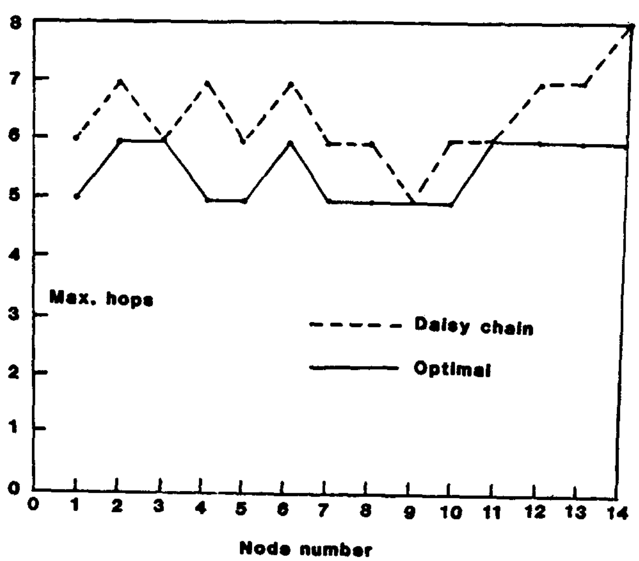

Figure 6.
Effect of a Node Failure on Maximum No. of Hops

A useful measure of fault-tolerance is to compute the terminal pair reliabilities. We consider two farthest nodes for terminal reliability analysis. In the optimal double loop network there exist many alternate routes between any pair of nodes, and therefore we expect high terminal pair reliability. For the two farthest nodes, one can intuitively see that the following properties hold:

Terminal reliability α No. of alternate routes; and

Terminal reliability $\alpha \; \frac{1}{\text{diameter}}$

We have already shown that the diameter is minimized when $h = \lfloor\sqrt{N}\rfloor$. Also, for this value of h, the number of alternate routes between the two farthest nodes will be higher in the optimal loop than in the daisy-chain loop, as shown in the following derivation. Therefore, we can conclude that the optimal loop network gives higher terminal pair reliability than the daisy-chain loop.

Now we compute the number of alternate routes between any two farthest nodes in the network. There is one path using only forward links, which may not be a shortest path. There are several additional paths which use backward links and forward links. All these routes are shortest routes between that node pair. Each of these routes consists of $b = \left\lfloor \frac{N}{h+1} \right\rfloor$ backward hops and (h-1) forward hops. We get different routes depending on the relative positions of "giant" steps and forward hops along the route. Therefore, the number of alternate routes between the two farthest nodes is the number of combinations of b out of (b+h-1) objects, that is:

$$S = R - 1 = \binom{b+h-1}{b} = \binom{b+h-1}{h-1}$$

The expression for the number of distinct routes is given by,

$$R = 1 + \binom{b + h - 1}{h - 1}$$

For Liu's loop R = 2;

For Daisy-chain R = 1 + $\binom{b+1}{1} = 2 + \left\lfloor \frac{N}{3} \right\rfloor$;

For the optimal loop, with $h = \lfloor \sqrt{N} \rfloor$,

$$R = 1 + \binom{\left\lfloor \frac{N}{h+1} \right\rfloor + h - 1}{h - 1}$$

For example, a 15-node network has h=3, b=3, and the number of distinct routes between any two farthest nodes is, $1 + \binom{5}{2} = 11$.

If we maximize R with respect to h, we find that the optimum value of h is $\lceil \sqrt{N} \rceil$. However, for most values of N, $h = \lfloor \sqrt{N} \rfloor$ also gives the same maximum value for R. Therefore for most values of N, and for N the square of an integer, the optimal loop will also maximize the number of distinct routes between any two farthest nodes.

6. CONCLUSIONS

We have presented an optimal double loop topology for locally distributed networks. This network is shown to provide the optimum performance with respect to maximum number of hops, average number of hops, throughput rate and terminal reliability. In particular, better performance is achieved with respect to previously proposed topologies such as DDLCN and daisy-chain loop. The cost of this topology in length of loop cable is $N(\lfloor \sqrt{N} \rfloor + 1)$, as compared with 2N for DDLCN and 3N for daisy-chain. Nodal interface costs are the same for all these architectures. One should realize that in a local network cable cost represents only a small fraction of total system cost. Therefore it appears that the additional cable cost is well compensated for by the performance benefits of the optimal loop topology.

This type of interconnection scheme among host processors may find applications in tightly coupled multiprocessor system. A separate host processor can perform the reconfiguration in case of failures. The reconfiguration of interconnections under failures, and addition of new nodes can be done easily if the interfaces are brought to one place. Interestingly, if all the links are bidirectional, the optimal double loop interconnection is precisely the same as the Illiac IV structure.

ACKNOWLEDGMENTS

The authors wish to thank Prof. Avizienis for his encouragement and support during this research and Ms. Rita Hall for her help in the preparation of the final manuscript.

REFERENCES

[BARN 68] G. H. Barnes et. al., "The Illiac IV Computer", IEEE Transactions on Computers, C-17, No. 8, August 1968, pp 746-757.

[FARM 69] D. Farmer, E. Newhall, "An Experimental Distributed Switching System to Handle Bursty Computer Traffic", Proc. ACM Symposium on Problems in the Optimizations of Data Commns., October 1969, pp 1-23.

[GRNA 79] A. Grnarov, L. Kleinrock, M. Gerla, "A New Algorithm for Network Reliability Computation", Proc. Computer Networking Symposium, NBS, Gaithersberg, December 1979, pp 17-20.

[GRNA 80] A. Grnarov, L. Kleinrock, M. Gerla, "A Highly Reliable Distributed Loop Network Architecture", Proc. 1980 Int. Symp. Fault-Tolerant Computing, Kyoto, October 1980, pp 319-324.

[KLEI 76] L. Kleinrock, Queueing Systems Vol. II : Computer Applications, Wiley Interscience, New York, 1976.

[LIU 78] M. T. Liu, "Distributed Loop Computer Networks", Advances in Computers, Vol. 17, Academic Press, 1978, pp 163-221.

[OHLI 77] Y. Oh, M. T. Liu, "Interface Design for Distributed Control Loop Networks", Proc. 1977 National Telecommunications conference, Los Angeles, December 1977, pp 31:4(1)-31:4(6).

[PIER 72] J. Pierce, "Network for Block Switches of Data", Bell Syst. Tech. Jour., Vol. 51, No. 6, July/August 1972, pp 1133-1145.

[WOLF79a] J. Wolf, M. T. Liu, B. Weide, D. Tsay, "Design of a Distributed Fault-Tolerant Loop Network", The 9th Annual Int. Symp. Fault-Tolerant Computing, Madison, June 1979, pp 17-24.

[WOLF79b] J. Wolf, B. Weide, M. T. Liu, "Analysis and Simulation of the Distributed Double Loop Computer Networks", Proc. Computer Networking Symposium, NBS, Gaithersberg, December 1979, pp 82-89.