A Comparative Reliability Study of Computer Network Configurations for Large-Scale Accelerator Systems

Edward W. Kamen and S. Simon Singh
kamen@ee.gatech.edu singh@ee.gatech.edu
School of Electrical & Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA

Abstract

A comparative study is given on the reliability of different possible network configurations for computer communications in large-scale accelerator systems, with the focus on bus, double ring, and switching hub networks. The measure of reliability is based on the computation of the probability that all nodes in the network are operative and can communicate with each other. Expressions for this probability are generated for each of the three network configurations, and then are evaluated to yield a definitive comparison in terms of reliability.

1 Introduction

The tremendous complexity of large-scale accelerator systems has resulted in the need for an information and control network that can meet very demanding requirements in terms of reliability, speed, capacity, cost, etc. Performance and cost issues are not only important in experimental physics systems - they are also of central importance in the use of accelerators for manufacturing materials such as the accelerator production of tritium (see [1]). In particular, for accelerators to be viable for manufacturing applications, the information and control network will have to possess a very high degree of reliability to insure that downtime is kept to a minimum. In building a new accelerator facility, or in upgrading an existing one, a fundamental question is how to design the information and control network to meet a very stringent reliability requirement. This paper is a first attempt to address this issue from an analytical standpoint, with the focus on the computer communication network that interconnects input-output controllers to terminals in a master control room. We do not consider the reliability of device-level networks (fieldbuses) interconnecting valves, actuators, PLCs, sensors, etc. However, the analysis developed here should be applicable to fieldbuses.

In the existing literature on reliability in computer communication networks, we have expressions for terminal-pair reliability. These expressions give the probability that a specific pair of nodes in a network can communicate with each other, but do not give a good "global measure" of reliability across a network. In applications, an important requirement is that failures in the network should not disrupt communications between operative nodes. Hence, a meaningful (global) measure of reliability is the probability that all operative nodes can communicate with each other. This probability, denoted by P(C), was studied by Ball [2] in the context of general computer communication networks. Ball developed a general procedure for determining P(C) based on a network decomposition scheme. Specific expressions for P(C) were generated by Lin and Silio [3] for single-ring and double-ring networks.

Another important global measure of reliability is the probability P(A) that all nodes are operative and can communicate with each other. In this paper, the primary objective is to generate expressions for P(A) for three different possible network configurations for computer communications in a large-scale accelerator system. These are bus, double ring, and switching hub networks. All three types of configurations have been used, or are under consideration for possible implementation, in accelerator systems. A switching hub configuration was proposed in [1].

In the next section, we describe the three different possible network configurations for an accelerator system, and then in Section 3, we generate expressions for P(A) for an N-node network. In Section 4 we evaluate P(A) for various values of the probabilities of failure of the nodes, links, and switches comprising the network. This yields a definitive comparison (in terms of reliability) of the different network configurations. Concluding comments are given in Section 5.

2 Network configurations

As is well known, a group of computers can be connected together in the three generic configurations of bus, double ring, and star (switching hub). In the application to accelerator systems, this threesome translates into the networks illustrated in Figures 1, 2, and 3. As noted in the Introduction, we focus only on connections of the input-output controllers (IOCs) to the terminals in the master control room (MCR). In Figures 1-3, the MCR is shown as a single block connected to the network, although in reality there will be a number of nodes in the MCR. We also show a gateway to other networks.

For the bus configuration illustrated in Figure 1, the network will go down if any node or link on the network fails, where a node consists of the bus interface and host attachment point. For the network illustrated in Figure 2, some robustness to link failure is built in as a result of the double ring structure of the network. In particular, communications can be diverted from the primary ring to the secondary ring as a result of a link failure. Robustness to node failures is also provided by the addition of bypass switches that switch off nodes that have failed.

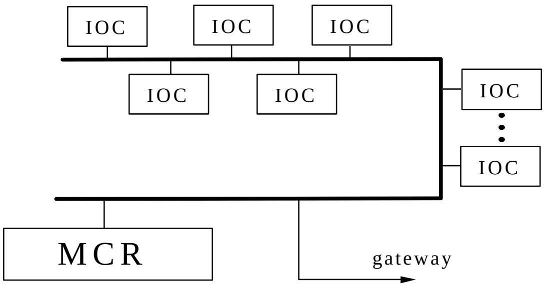

Figure 1 Bus configuration

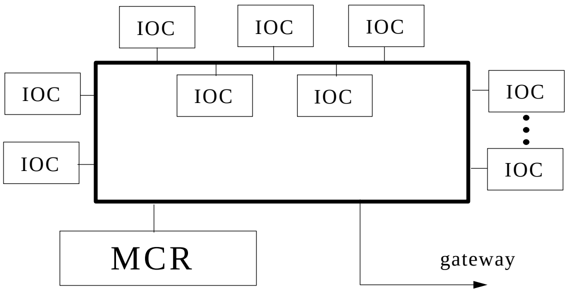

Figure 2 Double ring configuration

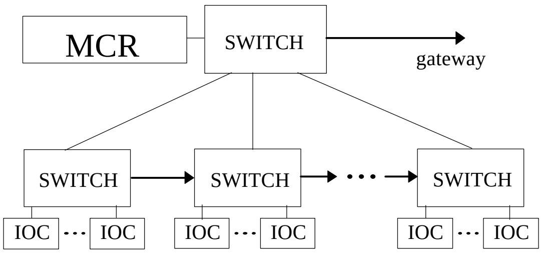

Figure 3 Switching hub configuration

As shown in Figure 3, the star configuration is given in terms of a central switching hub as the backbone of the computer network. Collections of IOCs are connected to local switching hubs, which in turn are connected to the central hub. In this framework, all nodes (host attachment points) on a switch are operative if the switch is operative. The local switching hubs shown in Figure 3 are also connected together to provide robustness to failures in the links from the local hubs to the central hub. This is the configuration proposed in [1].

3 Reliability analysis

Given an N-node computer communication network, let q denote the probability that, at a random point in time, a node fails. Assuming that failures of nodes are independent events, the probability that all nodes have failed is q^N, and the probability that all nodes are operating is $(1-q)^N$. The probability that only one node is in operation and all others have failed is $N(1-q)q^{N-1}$.

Now let C denote the event that all operative nodes of an N-node network can communicate with each other, and let P(C) denote the probability of C. To compute P(C), we first define the following events:

B_0 = event that all nodes have failed

B_1 = event that one node is operative and all other nodes have failed

B_i = event that i nodes are operative and can communicate with each other, where $i = 2,3,...,N$

Clearly, $C = \bigcup_{i=0}^{N} B_i$, and $B_i \bigcap B_j = \emptyset$, $i \neq j$,

and thus $P(C) = \sum_{i=0}^{N} P(B_i)$.

Using the values for $P(B_0)$ and $P(B_1)$, we have

$$P(C) = q^N + N(1-q)q^{N-1} + \sum_{i=2}^{N} P(B_i)$$

As noted in the Introduction, there exist papers on network reliability that are based on the computation of P(C). In this paper, the primary interest is on the computation of P(A), where A is the event that all nodes are operative and can communicate with each other. Clearly, P(A) = P(B_N), since by the above definition A= B_N. Obviously, P(C) is always greater than or equal to P(A). The computation of P(A) for each of the three configurations shown in Figures 1-3 is given below. In this computation, for the bus and ring networks it is assumed that failures of nodes, links, and switches are independent events. For the switching hub network, it is assumed that failures of links and switches are independent. The analysis given below does not include the possible failure of network interface cards (NICs), since in this paper a NIC is viewed as being part of a host, not part of the network.

3.1 Bus configuration

First, consider the N-node bus configuration illustrated in Figure 4. In this case, a node on the network consists of a bus interface and a host attachment point. Note that the network has two end-point terminations and a total of N+1 links. For any two operative nodes on the network to be able to communicate, it is necessary and sufficient that the two end-point terminations, all the links, and all the nodes be operative. Letting the probability of failure of a termination be denoted by T, the probability of failure of a link be denoted by DL, and the probability of failure of a node be denoted by q, we have that

$$P(A) = (1-q)^N (1-DL)^{N+1} (1-T)^2$$

Note that in this case, P(A) = P(C).

3.2 Double ring configuration

Consider the section of an N-node double ring configuration shown in Figure 5. In this case, a node on the ring consists of a ring interface and a host attachment point. As shown in Figure 5, each node on the double ring is connected to two bypass switches which short together

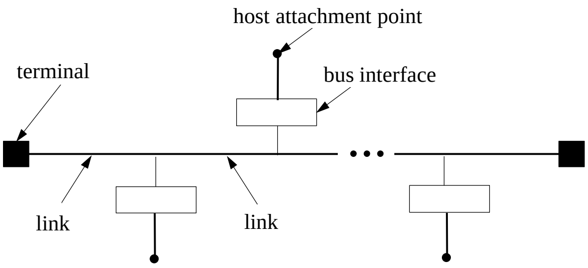

Figure 4 N-node bus network

the links on the bus and switch off the nodes in the case of a node failure. The probability of failure of a bypass switch is denoted by DS, and the probability of a link failure is denoted by DL.

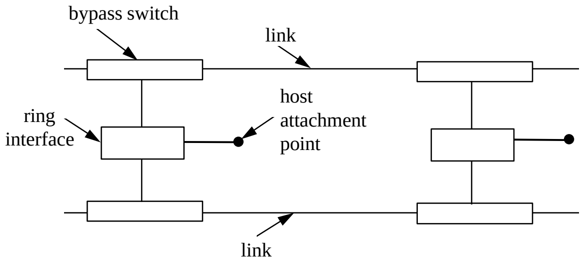

Figure 5 Section of double ring

The derivation of P(C) for a double ring with nodal bypass switches was carried by Lin and Silio [3]. The derivation of P(C) given in [3] can be modified to yield an expression for P(A): First, in the case when either the inner loop or the outer loop of the double ring is operative, the probability that all nodes are operative and can communicate with each other is given by

$$(1-q)^N \left[2(1-DS)^N(1-DL)^N - (1-DS)^{2N}(1-DL)^{2N}\right]$$

In the case when both the inner and outer loops are broken, the probability that all nodes are operative and can communicate is given by

$$\mathrm{N}(1-q)^N \left[(1-DS)^{2N}(1-DL)^{2(N-1)}(DL)^2\right]$$

P(A) is then given by the sum of the above two expressions; that is, P(A) is equal to

$$(1-q)^N \left[2(1-DS)^N(1-DL)^N - (1-DS)^{2N}(1-DL)^{2N}\right]$$

$$+\mathrm{N}(1-q)^N \left[(1-DS)^{2N}(1-DL)^{2(N-1)}(DL)^2\right]$$

3.3 Switching hub configuration

Consider the switching hub configuration shown in Figure 3, and suppose that there are M switches. Here the nodes are the host attachment points on the switches which are connected by interswitch links, each having a probability of failure equal to DL. Note that there are a total of 2M-3 links in the configuration shown in Figure 3. It is also assumed that all nodes on a switch are operative and can communicate with each other whenever the switch is working. The probability of failure of a switch is denoted by DS.

Now let E denote the event that all switches in the configuration shown in Figure 3 are able to communicate with each other. Then P(A) for this configuration is given by $P(A) = (1-DS)^M P(E)$ where P(E) is the probability of the event E.

To compute P(E), we first decompose E as follows. Let G_0 denote the event that all links from the central hub to the local hubs in the configuration in Figure 3 are operative, and the links between the local hubs are either working or not working. Obviously, when G_0 occurs all switches in the network can communicate with each other. For i = 1,2,...,(M-2), let G_i denote the event that all links from the central hub to the local hubs, except for i of them, are operative and all switches in the configuration can communicate with each other. Note that if all M-1 links between the central hub and the local hubs fail, not all of the switches can communicate.

Clearly $E = \bigcup_{i=0}^{M-2} G_i$ and since the G_i are mutually disjoint events, $P(E) = \sum_{i=0}^{M-2} P(G_i)$. It is easily seen that $P(G_0) = (1-DL)^{M-1}$, $P(G_1) = 2(M-2)(1-DL)^{M-1}DL$, and $P(G_i) = \alpha_i(1-DL)^{M-1}(DL)^i$, for $i = 2,3,...,M-2$ and for some positive real numbers α_i. Inserting these expressions for the P(G_i) into the expression for P(E) and factoring out $(1-DL)^{M-1}$ gives P(E) equal to $(1-DL)^{M-1}\left[1+2(M-2)DL+\alpha_2 DL^2+...+\alpha_{M-2}DL^{M-2}\right]$

Neglecting the second-order and higher-order powers of DL in the sum within the bracket, we have that $P(E) > (1-DL)^{M-1}\left[1+2(M-2)DL\right]$

and hence using the above expression for P(A), we have $P(A) > (1-DS)^M(1-DL)^{M-1}\left[1+2(M-2)DL\right]$

Thus we have a lower bound for P(A), and in fact, this bound is a fairly tight bound since the neglected terms in the derivation of the lower bound are second and higher-order powers of DL, which will be a small number. In the next section, we use the lower bound on P(A) to make comparisons with the P(A) values for the bus and double ring configurations.

4 Comparison of configurations

In our evaluation of the expressions for P(A) given in the previous section, we let N equal 206. This corresponds to using five nodes as connections to consoles of the MCR, one node for the gateway to another network, and 200 nodes for the input-output controllers. In the case of the switching hub configuration shown in Figure 3, we use a total of 21 switches so that M = 21. This corresponds to requiring a minimum of 26 ports on the central switch and a maximum of 13 on the local switches.

For the values of the failure probabilities of the network components shown in Table 1, the values for P(A) are plotted in Figure 6. It should be noted that the computation of P(A) for the switching hub is actually the lower bound given in Section 3. As is clearly evident from the plot, the switching hub configuration has a much higher degree of reliability in terms of the P(A) measure than the bus and double ring configurations.

Table 1 Scenarios for failure probabilities of network elements

	r1	r2	r3	r4	r5
q, T	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}
DL	10^{-3}	10^{-4}	10^{-4}	10^{-5}	10^{-6}
DS	10^{-3}	10^{-3}	10^{-4}	10^{-4}	10^{-5}

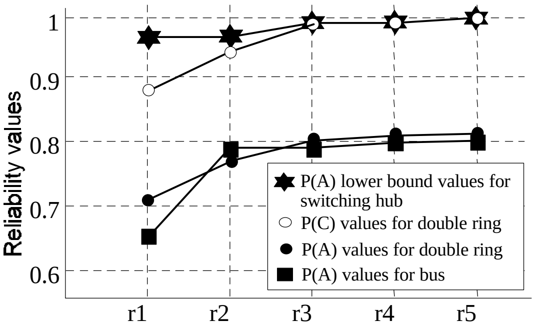

Figure 6 Network Reliability Values

It is also interesting to note from Figure 6 that the P(A) values for the bus and double ring are somewhat comparable. Hence in terms of the P(A) measure, there is not much of an increase in robustness as result of using a second ring in the ring network. However, as shown in Figure 6, the P(C) values for the double ring network are much better than the corresponding values for the bus. The reason for this is due to the fact that all nodes must be operative in the bus, so that P(A) = P(C) in this case. But for the double ring, not all nodes have to be operative since failed nodes can be switched out, and thus P(C) will be larger than P(A) for the double ring. Again looking at Figure 6, it is quite interesting to observe that the lower bound for the P(A) values of the switching hub are better than or comparable to the corresponding P(C) values for the double ring.

5 Concluding comments

Analytical expressions for the probability P(A) that all nodes are operative and can communicate were generated for the basic computer network configurations of bus, double ring, and switching hub. The approach developed here can be applied to more complicated networks consisting of mixtures of bus, ring, and switching hub configurations (see [4]).

Given cost constraints on the selection of network components and reliability values for the components, a problem of major interest is determining the configuration that yields the maximum possible value of P(A). In future work, we hope to be able to use the analytical formulation for P(A) to compute the increase in reliability that can be obtained by adding redundant links and switches.

In addition to reliability and cost, there is also the important factor of network performance under various traffic load distributions. With the move to very high bandwidth networks, such as Gigabit Ethernet, it appears that a very high degree of capacity will be readily obtainable for both broadcast-based and connection-based networks. An interesting question is whether or not a quantitative formulation can be developed for network performance under various traffic loads, and which could be integrated together with the quantitative measure of reliability given above in terms of P(A), or some other measure of reliability. We leave this an open area of research.

Acknowledgments

This work was supported in part by Westinghouse Savannah River Company.

References

[1] Kamen, E. and Gazarik, M. "APT Process Control and Information Network Architecture: Phase I --Scoping Analysis", School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 1996 .

[2] Ball, M. "Computing Network Reliability", Operations Research, vol. 27, no. 4, July-Aug. 1979, pp. 823-838.

[3] Lin, N. and Silio, C. "A Reliability Comparison of Single and Double Rings", Proceedings of IEEE INFOCOM'90: Conference on Computer Communications, pp. 504-511, San Francisco, CA, June 1990.

[4] Kamen, E. and Singh, S. "A Comparative Reliability Study of Computer Communication Networks," in preparation.