A Scalable and Fault Tolerant Network Structure for Tree Networks of Mission Critical Systems

Raheel Ahmed Memon, Yeonseung Ryu, Minho Shin
Department of Computer Engineering
Myongji University
Yongin, Gyeonggi-do, Korea

Jong-myong Rhee
Department of Information and Communication Eng.
Myongji University
Yongin, Gyeonggi-do, Korea

***Abstract*—In this paper, we propose a fault tolerant network architecture called Hierarchical Scalable Fault-tolerant Ethernet (H-SAFE) for mission critical systems. The proposed architecture is a tree based structure which provides multiple paths to route a packet. If any failure occurs in current path, then the alternative path recovers the failure and reroute the packet to destination without any information loss. In the failure state of any route, the fault detection and rerouting can be done within 750ms, which is a reasonable time for fault tolerance.**

Keywords- fault-tolerant network, tree structure, scalable

I. INTRODUCTION

Network fault tolerance is one of the most important capabilities required by mission-critical systems such as combat system data network (CSDN). Ethernet is becoming a de facto choice for open control network strategies. Ethernet, however, was not originally designed to handle network faults such as failure of network interface card (NIC), failure of hub/switch, or failure of cable connection. Various approaches for fault-tolerance network have been studied with four general goals such as no single point failure causes the loss of communication, fault detection and recovery within can be done within specified time, existing communication protocols must be supported and the use of existing commercial Off-The-Shell products (COTS) [1-5].

Generally there are two approaches for fault tolerance, first one is hardware and second one is software. In hardware approach, there is use of multi-port network interface card. Its switching time is fast but it requires developing a proprietary NIC. Second approach is software approach; it detects failure and executes the recovery algorithms. It uses the heartbeat mechanism for failure detection and recovery. However, it has scalability problem, because each heartbeat consumes some bandwidth and limits the number of nodes in a network.

In this paper, we propose a new fault tolerant network architecture called Hierarchical Scalable Autonomous Fault-tolerant Ethernet (H-SAFE). In proposed scheme, network architecture is a kind of tree architecture, dividing larger-scale network into subnets and connecting the subnets to core switches. Proposed scheme adapts heartbeat mechanism to detect faults and provides fast fault recovery.

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract UD070019AD.

II. ARCHITECTURE

In proposed H-SAFE scheme, we divide the larger scale network into several subnets and limit the number of nodes in each subnet. We can implement a large-scale network by adding additional subnets. Fig. 1 shows overall architecture of H-SAFE.

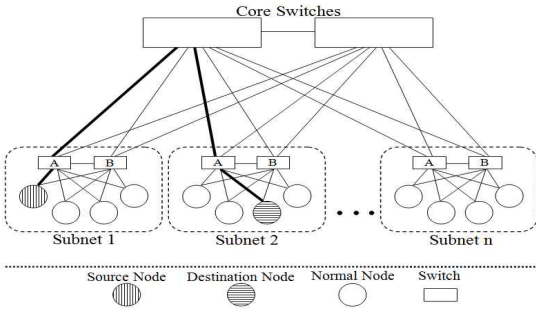

Figure 1. Overall Architecture

A. Subnet

A subnet contains limited number of nodes and two switches. Each node has two network interface cards to form dual connectivity for providing multiple paths, and both NICs are connected to the two switches of subnet. If one path become fails, then the alternative path will recover that fault. For detecting failures, we have implemented heartbeat mechanism in each node of the subnet. Each node in subnet sends the heartbeat message periodically on both two interfaces. The failure of any node can be determined if two consecutive heartbeat messages will miss.

B. Tree Structure of subnets

The switches of subnets are connected to core switches and forming a hierarchical architecture. The subnet switches and core switches of the architecture routes the packets using shortest path algorithm, core switches are having dual connectivity to subnet switches, and subnet switches also having dual connectivity to nodes for making several alternative paths.

As shown in Fig. 1, there are several alternative paths from source node to a destination node to prevent from failure and deliver the packet to destination in the worst conditions such as node, link, power, or switch failures.

III. FAULT TOLERANT NETWORK SCHEME

A. Fault Recovery Mechanism in a Subnet

Each node in a subnet has two network interfaces, interface A and B, and has software called FTE (Fault Tolerant Ethernet) which provides fault tolerant functionality. FTE takes place at layer 2 of the OSIRM, just below the IP and processes Ethernet frame based heartbeat messages. FTE periodically initiates transmission of heartbeats and processes the heartbeats that were sent by other nodes. The heartbeat is an Ethernet frame sent and received by nodes in a subnet.

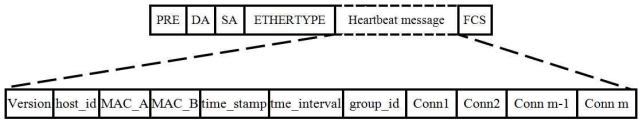

Figure 2. Format of Heartbeat Message

FTE sends heartbeat periodically every 320ms (heartbeat repetition interval time) and checks heartbeats received from other nodes every 50ms (detection latency). Fig. 2 shows the heartbeat message format and Table 1 shows description of the heartbeat message.

TABLE I. DESCRIPTION OF HEARTBEAT MESSAGE

Field	Description
host_id	IP address of sending node
MAC_A	MAC address of interface A of the node that sends this heartbeat
MAC_B	MAC address of interface A of the node that sends this heartbeat
time_stamp	time stamp taken at the moment of sending of this heartbeat
time_interval	time between subsequent heartbeats sent by the node
conn_*x*	Connectivity perception with other node.

By using this heartbeat they tell all other nodes that which heartbeats they receive. The heartbeats sent on A and B are sent to different multicast addresses. In normal situation (i.e., no connectivity failures), each interface processes its own multicast messages. Therefore, heartbeats are sent on interface A to multicast group A (via its native interface). Also, heartbeats are sent on interface B to multicast group B. A node listens to multicast group A heartbeats on interface A and listens to multicast group B heartbeats on interface B.

However, at start-up or when an interface fails to receive its heartbeat messages, a node sends and receives both A and B heartbeats on both interface A and B. As soon as heartbeats are received on a native interface then sending and receiving is restricted to the native interface only. The heartbeat is an Ethernet frame sent and received by nodes in a subnet.

FTE determines failure occurs on a path if it has not received consecutive 2 heartbeats from the path. When a path failure occurs in a subnet, FTE can detect it within 750ms.

B. Fault Recovery in Overall Network

Since FTE heartbeat mechanism is a layer 2 protocol, it works only within a single subnet. In order to exchange path information between multiple subnets, we propose a master node scheme which uses IP layer protocol. There are one or more master nodes in a subnet. Master nodes are responsible for maintaining path information between nodes in a subnet and exchanging the information with other subnet's master nodes.

At start-up time, all nodes elect master nodes for their subnet. The master nodes begin collecting path information by exploiting heartbeats. And then the master nodes broadcast IP packets to master nodes of other subnets in order to join the master node group. After joining the master node group, master nodes can exchange their network information. When the path information in a subnet is changed, the master node broadcasts updates to all master nodes in the master node group.

If a node wants to communicate with other node of other subnet, it has to get an admission from the master node. In order to get admission, the node sends to the master "query_path" message which contains the destination node address. The master node gives admission to the node by sending "answer_query" which contains the path information from the node to the destination node. After the node finishes communication, it has to notify to the master node by sending "finish" message.

IV. CONCLUSION

In this paper, we proposed a new fault tolerant network structure called H-SAFE for large scale mission critical systems. In H-SAFE, network is divided into several subnets and is formed a tree architecture of subnets. All nodes broadcast heartbeat messages periodically and detect faults on the path if a specified number of heartbeat messages are not arrived on the path. We also proposed a master node mechanism which is responsible for exchanging network status information between subnets. Our observation shows that fault detection and recovery process can be done within a specified time, which is a reasonable time to recover the faults.

REFERENCES

[1] J. Huang, S. Song, L. Li, P. Kappler, R. Freimark, and T. Kozlik, "An Open Solution to Fault-Tolerant Ethernet : Design, Prototyping, and Evaluation," Proc. IEEE International Performance, Computing, and Communications Conference, Feb. 1999, pp. 461-468.

[2] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu, "DCell: A Scalable and Fault-Tolerant Network Structure for Data Centers," Proc. ACM SIGCOMM 2008, Aug. 2008.

[3] P. Stelling, I. Foster, C. Kesselman, C. Lee, and G. Laszewski, "A Fault Detection Service for Wide Area Distributed Computations," Cluster Computing, Vol. 2, No. 2, 1999, pp. 117-128.

[4] J.B. Dugan, S.J. Bavuso, and M.A. Boyd, " Dynamic fault-tree models for fault-tolerant computer systems," IEEE Transactions on Reliability, Vol. 41, Issue. 3, Sep. 1992, pp. 363-377.

[5] S. Varadarajan and T. Chiueh, "Automatic Fault Detection and Recovery in Real Time Switched Ethernet Networks," Proc. IEEE INFOCOM '99, 1999, pp. 161-169.