RSAFE: A Robust Software-based Fault-Tolerant Scheme for Large-scale Ethernet Networks*

HOANG-ANH PHAM AND JONG MYUNG RHEE†
Department of Information and Communication Engineering
Myongji University
Gyeonggi-do, 449-728 Korea
E-mail: {anhph; jmr77}@mju.ac.kr

Most conventional software-based approaches for fault tolerant Ethernet (FTE) adopt a heartbeat mechanism for fault detection. However, due to the tradeoff between the number of nodes and the fault detection time, these conventional approaches do not provide scalability for large-scale networks. To solve this scalability issue, we previously proposed the SAFE scheme, which utilizes multiple subnets architecture. In this paper, we present an upgraded version of the SAFE scheme, the rapid SAFE (RSAFE) scheme, which adopts novel algorithms to significantly decrease the fault detection time while maintaining the advantages of the SAFE scheme. Simulation results using OMNeT++ show that the RSAFE scheme achieves excellent fault detection performance compared to the SAFE scheme. Furthermore, new analytical derivations and discussions of the key advantages of both the SAFE and RSAFE schemes are presented based on the multiple subnets architecture.

Keywords: fault-tolerant Ethernet (FTE), large-scale networks, failure switchover delay, heartbeat mechanism, SAFE, RSAFE

1. INTRODUCTION

Fault tolerance enables a system to continue operating when parts of it have failed. Incorporating fault-tolerance mechanisms greatly enhances the reliability and availability of a system. For most networked systems, the Ethernet has become a de facto standard but it did not originally support network fault tolerances. Therefore, a fault-tolerant Ethernet (FTE) is a common solution for fault tolerance in Ethernet-based systems. Furthermore, the amount of information flowing through communication networks proportionally increase according to the network size. This means that network fault tolerance becomes more challenging and important in large-scale networks, such as the naval combat system data network (CSDN). Various research studies have been conducted with the aim of enhancing FTE-based networks, and numerous developments and standardization efforts have been made [1-9]. However, few studies on FTE for large-scale networks have been reported, except our previous SAFE scheme [1].

At least two approaches exist for FTE implementation: software-based approaches and hardware-based approaches. In this paper, we address the software-based approach, in which multiple single-port network interface cards (NICs) are used to provide redun-

Received September 26, 2013; revised November 8 & December 30, 2013; accepted January 16, 2014.
Communicated by Xiaohong Jiang.
† Corresponding author: Jong Myung Rhee.
* This work was supported by the Human Resources Development program (No. 20134030200310) of the Korea Institute of Energy Technology Evaluation, Planning (KETEP) grant funded by the Korea Government Ministry of Trade, Industry and Energy.
dant communication paths. The FTE software called the redundancy protocol performs the algorithms for fault detection and fault recovery. The software-based approach is advantageous for using commercial-off-the-shelf (COTS) NICs without doing any modifications to the hardware and software drivers.

Most FTE software adopts a heartbeat mechanism, which periodically sends heartbeat messages for fault detection. For fast fault detection, the heartbeat messages must be sent more frequently. Furthermore, the number of heartbeat messages increases according to the number of nodes in the network. However, the heartbeat messages consume the network bandwidth, thereby limiting the network size and resulting in the network scalability problem. Due to this scalability problem, most conventional software-based approaches are not directly applicable to large-scale networks.

Recently, in [1], we proposed the SAFE scheme to solve the scalability issue in large-scale networks by adopting the multiple subnets architecture while still applying the conventional heartbeat mechanism for fault detection. Like the conventional heartbeat mechanism, a network fault is often determined on a data path between two nodes if there is no heartbeat message during two consecutive intervals of heartbeat repetition. Therefore, theoretically the failure switchover delay of the SAFE scheme takes, on average, twice the heartbeat repetition interval time. In this paper, we propose an upgraded version of the SAFE scheme called rapid SAFE (RSAFE). The RSAFE scheme adopts new algorithms in order to primarily improve the fault detection in terms of the failure switchover delay that is one of the most constraining factors in network fault tolerances.

For fast fault detection, two protocols, the parallel redundancy protocol (PRP) and the high availability seamless redundancy (HSR) protocol, can provide “zero” failure switchover delay [8, 9]. However, these two protocols are designed for automation applications within control information because the efficiency of these two protocols depends on the latency paid for monitoring and discarding the duplicates of the data in networks. Furthermore, the PRP and HSR can only protect against single point of failure. Therefore, these two protocols are not appropriate for large-scale networks.

The remainder of this paper is organized in the following manner. Section 2 presents the key characteristics and advantages of the SAFE scheme through new analyses and discussions that were not presented in [1]. Section 3 presents the fast fault-detection algorithms and further derivations in the RSAFE scheme. The simulations for faster fault-detection performance of the RSAFE scheme in terms of the failure switchover delay are presented in section 4. The last section discusses the conclusions and ongoing studies.

2. SAFE SCHEME

2.1 Review of the SAFE Scheme

To solve the scalability issue, the SAFE scheme divides a large-scale network into multiple subnets by limiting the number of nodes in each subnet. These subnets are interconnected in the form of a star topology as shown in Fig. 1. The network in the SAFE scheme can be extended by adding further surrounding subnets.

For FTE implementation, each subnet independently detects network faults by using the conventional heartbeat mechanism. Because a heartbeat message in the SAFE scheme is encapsulated in an Ethernet frame that cannot be transmitted to other subnets,
the SAFE scheme elects and manages master nodes in each subnet in order to exchange the subnet configurations for inter-subnet fault detection. Thus far, the SAFE scheme has been studied only on a star topology, in which the time frame for inter-subnet communication is predictable. The main drawback of the star topology is the vulnerability of the center subnet.

With regards to the performance of the SAFE scheme in terms of failure switchover delay, let $T_{fd}$ be the failure switchover delay of the SAFE scheme and $T_{hbinterval}$ be the heartbeat repetition interval time. Like the conventional heartbeat mechanism as aforementioned, theoretically the failure switchover delay of the SAFE scheme, $T_{fd}^{SAFE}$, averagely takes twice the heartbeat repetition interval time, $T_{hbinterval}$. However, through the detailed analysis in [1], the practical value of the failure switchover delay in the SAFE scheme can vary within the following boundary:

$$T_{hbinterval} < T_{fd}^{SAFE} < 3 \times T_{hbinterval}.$$  

(1)

From Eq. (1), it is also expected that the average value of the failure switchover delay, $T_{fd}^{SAFE}$, will be twice the heartbeat repetition interval time, $T_{hbinterval}$.

### 2.2 Further Analyses and Discussions

As discussed in [1, 5, 6], the number of heartbeat messages generated for fault detection is proportional to the number of nodes. Therefore, the number of heartbeat messages will increase according to the number of nodes. In practical terms, the heartbeat messages may cause congestion of data traffic. However, the multiple subnets architecture has an advantage in reducing the number of heartbeat messages generated. Additionally, in a large-scale network with thousands of nodes, there is a possibility of multiple link failures occurring in a given period. Available methods are ill-equipped to handle this issue. In contrast, the SAFE scheme increases the capability of the networks to overcome multiple points of network failures. The following subsections present new
analytical derivations and discussions in order to provide evidence of the aforementioned advantages of the SAFE scheme.

2.2.1 Reduction of network traffic generated by heartbeat messages

Let $S_{HBM}^1$ be the total number of heartbeat messages generated in a single subnet network of $N$ nodes at a heartbeat repetition interval time. $S_{HBM}^1$ is then determined as follows:

$$S_{HBM}^1 = 2N(N - 1).$$ (2)

Let $S_{HBM}^M$ be the total number of heartbeat messages generated in an $M$ subnets network of $N$ nodes at a heartbeat repetition interval time where each subnet $i$th of $K_i$ nodes independently adopts the heartbeat mechanism. The $S_{HBM}^M$ is determined as follows:

$$S_{HBM}^M = \sum_{i=1}^{M} 2K_i(K_i - 1) \quad \text{where} \quad N = \sum_{i=1}^{M} K_i. \tag{3}$$

Let $S_{HBM}^r$ be the reduction of heartbeat messages in the $M$ subnets network instead of a single subnet network. From Eqs. (2) and (3), $S_{HBM}^r$ is derived as follows:

$$S_{HBM}^r = S_{HBM}^1 - S_{HBM}^M = 2\left(\sum_{i=1}^{M} K_i\right)^2 - \sum_{i=1}^{M} K_i^2 = 4 \sum_{i=1}^{M} \sum_{j=i}^{N} K_iK_j. \tag{4}$$

From Eq. (4), the maximum reduction will be achieved when $K_i$ is equal to $K_j$ for any $i \neq j$. This means that the number of nodes in all the subnets is identical. Then, the reduction of heartbeat messages $S_{HBM}^r$ is represented as follows:

$$\text{Max}(S_{HBM}^r) = 4 \frac{M(M - 1)}{2} \left(\frac{N}{M}\right)^2 \left(\frac{N}{M}\right) = 2N^2\frac{(M - 1)}{M}. \tag{5}$$

Hence, for a given number of nodes, Eq. (5) implies that the number of heartbeat messages will be more reduced when the number of subnets increases. Let $R_r = \frac{M - 1}{M}$ be the reduction ratio obtained from Eq. (5), and then the heartbeat reduction is proportional to the reduction ratio for a given number of nodes. A numerical analysis to investigate the heartbeat reduction via the reduction ratio is depicted in Fig. 2. The graph shown in Fig. 2 is a useful reference for network designers to efficiently perform a subnet division. Furthermore, in the network design phase, the characteristics and functionalities of each node in the network are usually known in advance. Therefore, these parameters also serve as references for the subnet division and node distribution to reduce the latency due to inter-subnet communications.

The above analytical derivations only apply to the heartbeat message reduction in healthy scenarios. When a network fault is detected, the master nodes in the SAFE scheme generate additional messages for inter-subnet failure notifications. The number of additional messages depends on the location of the failure, the number of failures, and the
practical implementation of the two master nodes election algorithm. Therefore, it is quite complicated to generalize the heartbeat message reduction in faulty scenarios in the SAFE scheme. However, additional messages can be still accommodated by the network because these messages are not frequently generated.

2.2.2 Network capability enhancement to overcome multiple points of failure

Like the “divide and conquer” strategy, multiple link failures will be dispersed according to multiple subnets division. This means that the number of failures occurring in a subnet will decrease. Each subnet independently handles the failures, therefore increasing the capability of the whole network to overcome multiple points of failure. Fig. 3 (a) shows an example of an N-nodes network in the form of a single subnet network, in which three link failures are occurring in a given period: the link between NODE J and SWITCH A, the link between NODE K and SWITCH B, and the link between SWITCH A and B. It can be observed that the communication between NODE J and NODE K is interrupted. However, if these N nodes are divided into two subnets as shown Fig. 3 (b),
the communication between NODE J and NODE K is not interrupted.

2.2.3 Drawback of multiple subnets architecture

By adopting the multiple subnet architecture, the SAFE scheme not only solves the scalability issue for large-scale networks but also increases the capability of the network to handle multiple points of failure. However, the SAFE scheme is not able to handle one drawback of the multiple subnets architecture. Because the SAFE scheme adopts the conventional heartbeat mechanism in each subnet independently for fault detection, it is unable to detect the link failures between inter-subnet switches, which may incur the subnet isolation issue. For example as shown in Fig. 3 (b), the SAFE scheme will not detect the failures in the link between SWITCH A and C or in the link between SWITCH B and D. However, this drawback can be practically addressed by utilizing EtherChannel technology (IEEE 802.3ad) for multiple links between switches, as shown in Fig. 4. Although this drawback is not solved completely because this type of failure cannot be detected by the proposed approach, the subnet isolation issue can be avoided by incorporating EtherChannel technology.

![Fig. 4. Connection between two switches using EtherChannel.](image)

The extra switches and links required for subnet division increase the hardware costs in a multiple subnets architecture. The additional cost is a necessary tradeoff to enhance the network reliability of mission-critical large-scale systems.

3. RSAFE SCHEME

In this section, we describe the enhanced SAFE scheme called RSAFE. The RSAFE scheme is also based on multiple subnets architecture. Therefore, it inherits most advantages of the SAFE scheme. However, the RSAFE scheme adopts novel algorithms to improve the failure switchover delay and eliminates the master nodes in each subnet to reduce the latency in inter-subnet communications. Additionally, while the SAFE scheme has been studied only in star topology (Fig. 1), the RSAFE scheme is studied in both star and ring topologies (Fig. 5).

3.1 Novel Algorithm for Fast Fault Detection

To speed up the fault detection, the RSAFE scheme adopts two levels of heartbeat messages called the MAC-based heartbeat message (M-HBM), which is encapsulated in an Ethernet frame, and the IP-based heartbeat message (I-HBM), which is encapsulated in an IP packet. Each node checks the physical links from its two network interfaces that connect to switches before broadcasting the M-HBM as a conventional heartbeat mecha-
Fig. 5. Multiple subnets in the form of a ring top.

A SOFTWARE-BASED FAULT-TOLERANT SCHEME FOR LARGE-SCALE ETHERNET NETWORKS

nism. If there is a link failure or link recovery, the corresponding node will broadcast an I-HBM to the whole network instead of broadcasting an M-HBM inside a local subnet. The I-HBMs enable the other nodes to detect network faults soon after they occur. Consequently, the failure switchover delay will be shortened. The novel algorithms for fast fault detection, including heartbeat message sending (HBM-TX) and heartbeat message reception (HBM-RX), are described in the diagrams shown in Figs. 6 and 7, respectively.

Fig. 6. Novel algorithm for heartbeat message sending (HBM-TX) in the RSAFE scheme.

The two following extreme cases are considered to investigate how the failure switchover delay is improved in the RSAFE scheme. In the best-case scenario shown in Fig. 8, link failure or link recovery occurred shortly before NODE J was to send heartbeat messages to the other nodes in local subnet. By applying the HBM-TX algorithm, NODE J will broadcast an I-HBM instead of the usual M-HBM. When NODE K receives the I-HBM from NODE J, by referring to the HBM-RX algorithm, the link failure or link recovery at NODE J will be detected. It is obvious that the failure switchover delay is much less than the repetition interval time of one heartbeat, $T_{	ext{interval}}$.  

Fig. 6. Novel algorithm for heartbeat message sending (HBM-TX) in the RSAFE scheme.
In the worst-case scenario shown in Fig. 9, link failure or link recovery occurs soon after NODE J has broadcasted an M-HBM as usual. Therefore, the I-HBM will be sent during the next heartbeat interval. However, the failure switchover delay is still less than
A SOFTWARE-BASED FAULT-TOLERANT SCHEME FOR LARGE-SCALE ETHERNET NETWORKS

or equal to the repetition interval time of one heartbeat, \( T_{\text{hbInterval}} \).

Let \( T_{\text{subnet}} \) be the failure switchover delay for a local subnet. Based on the discussions on the two extreme cases above, the \( T_{\text{subnet}} \) can be bounded as follows:

\[
T_{\text{subnet}} \leq T_{\text{hbInterval}}. \tag{6}
\]

The diagram shown in Fig. 10 illustrates the time frame for the fault detection with the RSAFE scheme, including inter-subnet fault detection. When NODE J in SUBNET X detects a link failure or link recovery, it broadcasts an I-HBM to the whole network. NODE Q in SUBNET Y detects the failure upon receiving the I-HBM sent by NODE J. Let \( \Delta_t \) be the transmission latency between any two subnets. Then, the failure switchover delay for the RSAFE scheme \( T_{\text{RSAFE}} \) is derived as follows:

\[
T_{\text{RSAFE}} = T_{\text{subnet}} + \Delta_t. \tag{7}
\]

Fig. 10. Diagram for inter-subnet fault detection in the RSAFE scheme.

The transmission latency in Eq. (7) includes the propagation delay and the processing delay. Theoretically, these delays are much less than hundreds of milliseconds. Practically, however, they depend on the specific conditions and situations of the network.

Eqs. (6) and (7) indicate that the RSAFE scheme reduces the failure switchover delay when compared to the SAFE scheme in Eq. (1). By eliminating the master nodes in each subnet, the proposed algorithms eliminate the need for master node election and maintenance. Therefore, the effectiveness of inter-subnet communications will be improved.

3.2 Further Analyses and Discussions of Heartbeat Message Reduction in Faulty Scenarios

Unlike the SAFE scheme, the RSAFE scheme does not use master nodes for inter-subnet failure notification in faulty scenarios. Instead, when link failure occurs, the
corresponding faulty node broadcasts I-HBMs to every node in the local subnet, as well as in neighbor subnets. In the RSAFE scheme, the total number of heartbeat messages is a summation of the total number of M-HBMs generated by healthy nodes and the total number of I-HBMs generated by faulty nodes. However, the total number of I-HBMs generated by the faulty nodes is identical in both the single subnet network and multiple subnets network. Therefore, the heartbeat message reduction depends only on the reduction of the M-HBM generated by the remaining healthy nodes. Thus, the analytical derivations for heartbeat message reduction in the SAFE scheme are still applicable and valid for a number of healthy nodes in the RSAFE scheme.

The following derivations provide evidence for the above discussions about the heartbeat message reduction in the RSAFE scheme. Let \( F S_1^{HBM} \) be the total number of heartbeat messages, including M-HBMs and I-HBMs, generated in single subnet network of \( N \) nodes at a heartbeat repetition interval time when there are \( F \) failures. These \( F \) failures correspond to \( F \) faulty nodes. As each node has two network interfaces, if two failures occur at a node, it is concluded that this node has disappeared from the network. Therefore, the messages generated by this node will not be calculated. By applying the fault detection algorithm in the RSAFE scheme, the total number of heartbeat messages is a summation of the total number of M-HBMs generated by \((N - F)\) healthy nodes and the total number of I-HBMs generated by \( F \) faulty nodes. Each healthy node still broadcasts M-HBMs to the other \((N - 1)\) nodes on its two network interfaces. Meanwhile, each faulty node only broadcasts I-HBM to other \((N - 1)\) nodes on the remaining healthy network interface. Hence, \( F S_1^{HBM} \) can be determined as follows:

\[
F S_1^{HBM} = \left(2(N - 1)(N - F) + F(N - 1)\right) \frac{M - \text{HBMs}}{i - \text{HBMs}}
\]

Let \( F S_{HBM}^M \) be the total number of heartbeat messages, including M-HBMs and I-HBMs, generated by \( N \) nodes in the \( M \) subnets architecture. To achieve the maximum reduction, the number of nodes in all the subnets should be identical. Without loss of generality, let \( K \) be the number of nodes in each subnet, which means that \( N = MK \). Furthermore, because the single subnet network is divided into \( M \) subnets, the \( F \) failures are randomly distributed into \( M \) subnets. Let \( f_i \) be the number of failures in the subnet \( i \), in which each healthy node still broadcasts M-HBMs to other \((K - 1)\) nodes on its two healthy network interfaces. Meanwhile, each faulty node broadcasts I-HBMs to all \((N - 1)\) nodes, including the other nodes in the local subnet and the other nodes in neighbor subnets, on the remaining healthy interface. Hence, \( F S_{HBM}^M \) can be calculated as follows:

\[
F S_{HBM}^M = \sum_{i=1}^{M} \left[ \frac{2(K - 1)(K - f_i) + f_i(N - 1)}{M - \text{HBMs}} \right] \frac{M - \text{HBMs}}{i - \text{HBMs}}
\]

where \( MK = N \) and \( F = \sum_{i=1}^{M} f_i \).

Eq. (9) can be expanded and derived as follows:
\begin{equation}
\mathbf{S}_{\text{HBM}}^{M} = 2(K - 1)\sum_{i=1}^{M}(K - f_{i}) + (N - 1)\sum_{i=1}^{M}f_{i} = 2(K - 1)(N - F) + F(N - 1).
\end{equation}

Let $\mathbf{S}_{\text{HBM}}^{F}$ be the reduction of heartbeat messages in the $M$ subnets network instead of the single subnet network under scenarios of $F$ failures. From Eqs. (8) and (10), the maximum reduction of heartbeat messages in the faulty scenarios is derived as follows:

\begin{equation}
\text{Max}\{\mathbf{S}_{\text{HBM}}^{F}\} = \mathbf{S}_{\text{HBM}}^{F} - \mathbf{S}_{\text{HBM}}^{M} = 2(N - F)(N - K).
\end{equation}

Then, Eq. (11) can be represented as:

\begin{equation}
\text{Max}\{\mathbf{S}_{\text{HBM}}^{F}\} = 2(N - F)(N - \frac{N}{M}) = \frac{2N(N - F)(M - 1)}{M}.
\end{equation}

Eq. (12) implies that the maximum reduction of heartbeat messages for a given number of failures is also proportional to the reduction ratio, $R_{M}$. For non-failure scenarios ($F = 0$), Eq. (12) is definitely consistent with Eq. (5). This means that Eq. (12) is a comprehensive derivation for the heartbeat message reduction in both healthy and faulty scenarios.

4. SIMULATIONS AND EXPERIMENTS

The RSAFE scheme is implemented with OMNeT++, which is a component-based C++ simulation library and framework used in network simulators [10]. Fig. 11 shows a snapshot of a simulation, in which 64 nodes are uniformly distributed into 8 subnets that are interconnected in the form of a ring topology. The novel HBM-TX and HBM-RX algorithms for fast fault detection presented in Section 3.1 are implemented and validated in scenarios with various numbers of nodes and numbers of subnets.

Fig. 11. Snapshot of a simulation of 64 nodes divided into 8 subnets in a ring topology.
4.1 Experimental Descriptions

A fast Ethernet network was established to be used in the simulation. Various nodes, including nodes numbered 64, 128, 256, and 1024, are uniformly distributed into 8, 16, and 32 subnets to achieve the maximum reduction in the number of heartbeat message. The multiple subnets are interconnected in the form of a ring or a star topology as depicted in Figs. 1 and 5, respectively.

The failure switchover delay parameter is often included as a key requirement of capabilities for a mission-critical system, such as naval CSDN. For example in the naval CSDN, the typical design goal of the failure switchover delay is the one second that we have chosen in our study. Based on the analytical derivation in Eq. (1), the failure switchover delay in the SAFE scheme is less than three times the heartbeat repetition interval time. This means that the heartbeat repetition interval time should be less than one-third second in order to be sure that the failure switchover delay is less than one second as our design goal. Therefore, the $T_{thInterval}$ is arbitrarily selected as 300 milliseconds for our experiments.

Various experiments were conducted to validate the correct functioning of the network fault detection and to evaluate the performance in terms of the failure switchover delay when applying the RSAFE scheme. Each experiment corresponds to the number of nodes, the number of subnets, and the topology. For each experiment, 25 link failures were randomly generated, and there were, at most, 3 link failures that could occur simultaneously in a given period. The failure switchover delay was recorded when link failure was detected. Then, the failure switchover delays were recorded and compared to the derivations presented in Eqs. (6) and (7). Table 1 summarizes the experimental parameters that were applied in our experiments.

<table>
<thead>
<tr>
<th>Table 1. Experimental parameters.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
</tr>
<tr>
<td>Number of nodes</td>
</tr>
<tr>
<td>Number of subnets</td>
</tr>
<tr>
<td>Topology</td>
</tr>
<tr>
<td>Heartbeat repetition interval time $T_{thInterval}$</td>
</tr>
<tr>
<td>Link failure and recovery</td>
</tr>
</tbody>
</table>

4.2 Experimental Results

The experimental results in Tables 2 and 3 show that the failure switchover delays for local subnets, $T_{fsdT}$ are all less than the repetition interval time of one heartbeat. Thus, the analytical derivation in Eq. (6) is valid. The failure switchover delay still meets the design goal for various numbers of nodes and various numbers of subnets.

The experimental results for the inter-subnet failure switchover delay, $T_{fS}$ also agree with the derivation in Eq. (7) where $\Delta_s$ depends on the number of subnets and the network topology. For a star topology, a message can be sent from one node to other nodes with a predictable timeframe (e.g. a message travels through, at most, three subnets). This means that the $\Delta_s$ in a star topology does not depend on the number of subnets. However, for a ring topology, the $\Delta_s$ increases according to the number of subnets.
Table 2. Experimental results for star topologies.

<table>
<thead>
<tr>
<th>#Nodes</th>
<th>8 subnets</th>
<th>16 subnets</th>
<th>32 subnets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(ms)</td>
<td>(ms)</td>
<td>(ms)</td>
</tr>
<tr>
<td></td>
<td>$T_{f/sub}$</td>
<td>$T_{f/RSAF}$</td>
<td>$T_{f/sub}$</td>
</tr>
<tr>
<td>64</td>
<td>Min 4.9</td>
<td>5.1</td>
<td>Min 4.4</td>
</tr>
<tr>
<td></td>
<td>Max 285.7</td>
<td>285.9</td>
<td>Max 293.5</td>
</tr>
<tr>
<td></td>
<td>Avg 161.1</td>
<td>161.3</td>
<td>Avg 147.6</td>
</tr>
<tr>
<td>128</td>
<td>Min 27.6</td>
<td>27.8</td>
<td>Min 10.7</td>
</tr>
<tr>
<td></td>
<td>Max 280.4</td>
<td>280.6</td>
<td>Max 294.2</td>
</tr>
<tr>
<td></td>
<td>Avg 159.0</td>
<td>159.2</td>
<td>Avg 163.5</td>
</tr>
<tr>
<td>256</td>
<td>Min 18.7</td>
<td>18.9</td>
<td>Min 34.2</td>
</tr>
<tr>
<td></td>
<td>Max 280.9</td>
<td>281.1</td>
<td>Max 266.2</td>
</tr>
<tr>
<td></td>
<td>Avg 149.1</td>
<td>149.3</td>
<td>Avg 150.4</td>
</tr>
<tr>
<td>512</td>
<td>Min 10.9</td>
<td>11.1</td>
<td>Min 19.9</td>
</tr>
<tr>
<td></td>
<td>Max 289.4</td>
<td>289.6</td>
<td>Max 279.6</td>
</tr>
<tr>
<td></td>
<td>Avg 152.2</td>
<td>152.4</td>
<td>Avg 151.3</td>
</tr>
<tr>
<td>1024</td>
<td>Min 15.6</td>
<td>15.8</td>
<td>Min 39.6</td>
</tr>
<tr>
<td></td>
<td>Max 278.6</td>
<td>278.8</td>
<td>Max 293</td>
</tr>
<tr>
<td></td>
<td>Avg 160.7</td>
<td>160.9</td>
<td>Avg 159.9</td>
</tr>
</tbody>
</table>

Table 3. Experimental results for ring topologies.

<table>
<thead>
<tr>
<th>#Nodes</th>
<th>8 subnets</th>
<th>16 subnets</th>
<th>32 subnets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(ms)</td>
<td>(ms)</td>
<td>(ms)</td>
</tr>
<tr>
<td></td>
<td>$T_{f/sub}$</td>
<td>$T_{f/RSAF}$</td>
<td>$T_{f/sub}$</td>
</tr>
<tr>
<td>64</td>
<td>Min 32.5</td>
<td>32.8</td>
<td>Min 16.9</td>
</tr>
<tr>
<td></td>
<td>Max 287.4</td>
<td>287.7</td>
<td>Max 288.3</td>
</tr>
<tr>
<td></td>
<td>Avg 168.8</td>
<td>169.1</td>
<td>Avg 162.1</td>
</tr>
<tr>
<td>128</td>
<td>Min 14.6</td>
<td>14.9</td>
<td>Min 2.4</td>
</tr>
<tr>
<td></td>
<td>Max 290.6</td>
<td>290.9</td>
<td>Max 281.2</td>
</tr>
<tr>
<td></td>
<td>Avg 170.6</td>
<td>171.0</td>
<td>Avg 152.3</td>
</tr>
<tr>
<td>256</td>
<td>Min 17.9</td>
<td>18.2</td>
<td>Min 9.9</td>
</tr>
<tr>
<td></td>
<td>Max 278.0</td>
<td>278.3</td>
<td>Max 267.9</td>
</tr>
<tr>
<td></td>
<td>Avg 145.3</td>
<td>145.6</td>
<td>Avg 138.5</td>
</tr>
<tr>
<td>512</td>
<td>Min 6.2</td>
<td>6.5</td>
<td>Min 6.3</td>
</tr>
<tr>
<td></td>
<td>Max 278.3</td>
<td>278.6</td>
<td>Max 253.8</td>
</tr>
<tr>
<td></td>
<td>Avg 144.1</td>
<td>144.4</td>
<td>Avg 141.4</td>
</tr>
<tr>
<td>1024</td>
<td>Min 24.9</td>
<td>25.2</td>
<td>Min 6.5</td>
</tr>
<tr>
<td></td>
<td>Max 291.5</td>
<td>291.8</td>
<td>Max 283.6</td>
</tr>
<tr>
<td></td>
<td>Avg 155.3</td>
<td>155.6</td>
<td>Avg 150.1</td>
</tr>
</tbody>
</table>
5. CONCLUSIONS

In this paper, we presented the RSAFE scheme, a further development of the previous FTE approach for large-scale networks, the SAFE scheme. The RSAFE also adopts multiple subnets architecture, allowing it to maintain two key advantages of the SAFE scheme: (1) a reduction in network traffic generated by heartbeat messages and (2) an improvement in the capability of the network to overcome multiple points of failure. New analytical derivations and discussions were presented in some detail to clearly show these two advantages.

Unlike the SAFE scheme, the RSAFE scheme adopts novel algorithms, HBM-TX and HBM-RX, to improve the fault detection performance. The RSAFE scheme was successfully implemented with OMNeT++. The experiments were performed using various numbers of nodes such as 64, 128, 256, 512 and 1024. The experimental results in Tables 2 and 3 are consistent with the analytical derivations in Eqs. (6) and (7). They illustrate the ability of the proposed fault detection algorithms to significantly improve the failure switchover delay.

Further studies are needed on the practical implementation of the RSAFE scheme, for example, network protocol development and kernel programming. To completely validate the proposed scheme, further experiments also need to be performed in practical network platforms.

REFERENCES


Hoang-Anh Pham received his BEng in Computer Science and Engineering in 2005 from Ho Chi Minh City University of Technology (HCMUT), Vietnam. From 2005 to 2008, he was an Assistant Lecturer at the Faculty of Computer Science and Engineering, HCMUT. He received his MSc and Ph.D. in Communications Engineering from MYONGJI University, South Korea in 2010 and 2014, respectively. Currently, he is a Lecturer at the Faculty of Computer Science and Engineering, HCMUT. His research interests include fault-tolerant networks, data communications, internet of things, and mobile application development.

Jong Myung Rhee received his Ph.D. from North Carolina State University, USA, in 1987. After 20 years at the Agency for Defense Development in Korea, where he made noteworthy contributions to C4I and military satellite communications, he joined DACOM and Hanaro Telecom in 1997 and 1999, respectively. At Hanaro Telecom, which was the second largest local carrier in Korea, he served as Chief Technology Officer (CTO), with a senior executive vice-president position. His main duty at Hanaro Telecom was a combination of management and new technology development for high-speed Internet, VoIP, and IPTV. In 2006 he joined Myongji University and is currently a Full Professor in the Information and Communications Engineering Department. His recent research interests are centered on military communications, including ad-hoc and fault-tolerant networks. He is a member of IEEE, IEICE, and KICS.