

Analysis of TCP performance over the PBB-TE during the protection switchover time

Hyung-Nam Kim ^a, Hyun-Pil Kim ^a, Jin-Seek Choi ^a

^a *Department of Electronics Computer Engineering, Hanyang University*

khn3182@hotmail.com, necro_hyunpi@naver.com, jinseek@hanyang.ac.kr

Abstract— Adapting to meet service requirements to users and growth of market in factory automation and health control equipments, network environment is rapidly changing into the reliable and high-quality Internet. TCP is expected to be a solution of the reliable transport protocol. The carrier class transport network is also expected as a solution of the underlying transport network which supports high availability with 50ms resilience time. This paper analyzes the impact of protection switching on TCP performance over the Provider Backbone Bridge Traffic Engineering (PBB-TE) network. Our experimental results suggest TCP protocol expected to be set Retransmission Timeout (RTO) to 50ms or depending on the circumstances of the underlying transport network.

Keywords—Performance analysis, PBB-TE, Protection, Switch over time, TCP RTO

I. INTRODUCTION

For factory automation and health care control, the transport network will require the most stringent Quality of Service (QoS) requirements. The carrier class transport network supports a few milliseconds delay and recovery time. In order to provide a synchronized service constraint, the request and response cycle time is maintained with a few milliseconds time even in network failure environments. Setting an appropriate restoration objective is crucial factor in safety-critical and hazardous environments such as automatic plant and hospital.

Carrier Ethernet provides with maintaining high reliability, Operation, Administration, and Management (OAM) and fault management in End-to-end Ethernet virtual Connection (EVC). Without using flooding, Ethernet Bridge-based Carrier Ethernet is used for transporting real-time traffic in Backbone networks. To achieve a high level of service availability, PBB-TE provides the recovery of less than 50ms latency like reliable SONET/SDH service [1]. Less than 50ms recovery time is important requirement. Internet backbones are obliged to meet the 50 millisecond restoration objective, although this factor was actually derived from the requirement of conventional telephone traffic [2]. Although carrier class network provides reliable environment, TCP is designed for unreliable network. For example, TCP requires a sub 5 s recovery time because the Spanning Tree Protocol (STP) triggers the topology recovery of the underlying network for network failure [3]. There is a difference between carrier class network recovery time and TCP recovery time. As a result, consideration of TCP is required to apply to PBB-TE

This paper identifies the cause of the decrease of TCP performance behaviour during the protection time over the reliable carrier class network. From the perspective of TCP stream, we take into the restoration factors to account for backbone links. We experiment TCP based application in the reliable PBB-TE network which supports 1:1 protection switching, and analyse TCP performance for transient-state that may cause packets losses.

Our results show that TCP performance is affected by RTO time instead of the protection time in the underlying network. Even if transport network provides the fast protection time to a few milliseconds time, the upper layer like TCP is not affected. During the switchover, therefore we raise the issue of long TCP RTO due to the packet loss in Carrier Ethernet network, such as the PBB-TE. Based on the investigation results, we propose 50ms RTO mechanism that increases the high precision in industry, and consequently improves TCP throughput for each source and destination node.

The rest of this paper is organized as follows. In the following section, we describe background how to react differently to the RTO in PBB-TE according to the cause. In section 3, we illustrate the implemented PBB-TE model to analyze transient performance during switchover time, and explain performance degradation. In section 4, we propose 50ms RTO mechanism for PBB-TE and observed two cases. Lastly, we conclude this paper in section 5.

II. BACKGROUND

A. PBB-TE

PBB-TE is the one of the Carrier class network technologies that is standardized by IEEE 802.1Qay. By setting up Ethernet Switched Paths (ESP) with the external agent, PBB-TE enables construction of active topologies. To ensure QoS in end-to-end ESP lines, Continuity Check Message (CCM) is sent to monitor the status of the lines. To ensure the reliability of a node or link fails, PBB-TE uses the protection switching mechanism to do switchover from the working line to the protection line. The biggest difference with IEEE 802.1ah PBB is the removal of flooding function for Ethernet frame which is not registered in Filtering Database (FDB) [4-5]. Instead of using source MAC learning and the STP to prevent loops, external agent creates Forwarding table with the optimal path.

PBB-TE is composed of Backbone Edge Bridge (BEB) and Backbone Core Bridge (BCB). By using MAC-in-MAC encapsulations which change independent the customer MAC

address to backbone MAC address, BEB is node for transporting to backbone network. BCB fast forward the MAC-in-MAC frame which is received from BEB or BCB to another node. Performance is monitored regularly through the Connectivity Fault Management (CFM) in ESP which is set between the BEBs.

B. Spurious TCP RTO

Recovery time of packet loss on TCP is entirely dependent on RTO. RTO of TCP is varied by measuring the delay of Round-Trip Time (RTT) sample in current network. Under certain sudden changes like the handoff in the wireless environment or routing changes or a rapid switchover during the switch over time, finding an appropriate RTO is limited. These sudden changes can be defined as a spurious RTO. There are algorithms, EIFEL (RFC 4015) and the F-RTO (RFC 5682), to solve these problems [6-7]. However, these algorithms only consider the delay spike where a problem has arisen at the point of connection to a service instance.

As recovery time of packet loss relies on RTO, too short RTO causes unnecessary retransmissions. In contrast to short RTO, a long RTO could reduce the throughput that causes a performance problem in the carrier class network. Surely, high reliability, guaranteed bandwidth makes the delay spike occur seldom in Carrier class networks; however, we focus on sudden changes in PBB-TE.

C. Restoration Objectives

If Reno TCP is on the transport layer, the 50millisecond target is too aggressive for file transfer rather 200ms protections time is recommended [2]. Because TCP is not created for dealing with network failures, there is no difference between network failure and network congestion from perspectives of TCP. It is a problem that RTO is triggered regardless of congestion [8]. RTO has great impact on a parameter in TCP congestion control. In the case of network failures, there is a considerable gap between RTO and 50ms recovery time for Internet backbone links especially PBB-TE network. In addition, it is possible that previously mentioned 200ms protection time can negative affect high precise industrial network.

III. EVALUATION OF TCP PERFORMANCE DURING THE PROTECTION SWITCHOVER TIME

A. Implemented model

In this paper, IEEE 802.1Qay protocol-based PBB-TE system is implemented as a Carrier class network. We built PBB-TE over the OpenWrt embedded system using Open Source Software (OSS) based on Linux platform. In contrast to Network simulator, performance assessment in Linux based test-bed helps to make practical outcomes and testing strategies for improving the reliability of the whole system.

PBB-TE is implemented to BEB and BCB as independent nodes. BCBs are composed of one for working path and another for protection path over a period of CCM interval that is equal to 10msec. Through the Q-in-Q modules, customers

can access to the BEB. At the user space of Server and client, WireShark and Iperf were installed to obtain results.

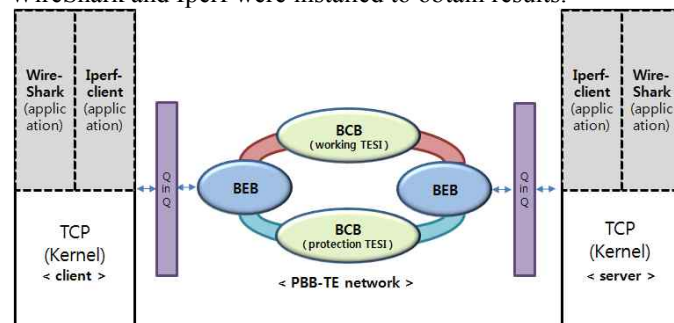


Figure 1. Implementation of PBB-TE technology

B. Analysis of transient performance during switchover

One of the key aspects of the transient performance is an evaluation of link failure on system. Under the abnormal conditions, the transient performance reflects performance degradation [9]. By using IPerf with maximum traffic load on PBB-TE network, we analyse the performance of TCP. After link failure, working TESI follows protection switching procedure during transient state.

1) **Packets loss of the TCP connections:** Shown in Fig.2 is a plot of the number of TCP and UDP segments when the client simultaneously sends segments to server with maximum bandwidth. Even though protection switchover time is supported within 27msec, our primary observation is that TCP stream has hold-off time during 295ms without any packets. The hold-off time equal to RTO is triggered by segments loss. After the RTO, the number of TCP segments is recovered as normal state. In contrast to TCP, UDP immediately retransmits after the protection switchover time.

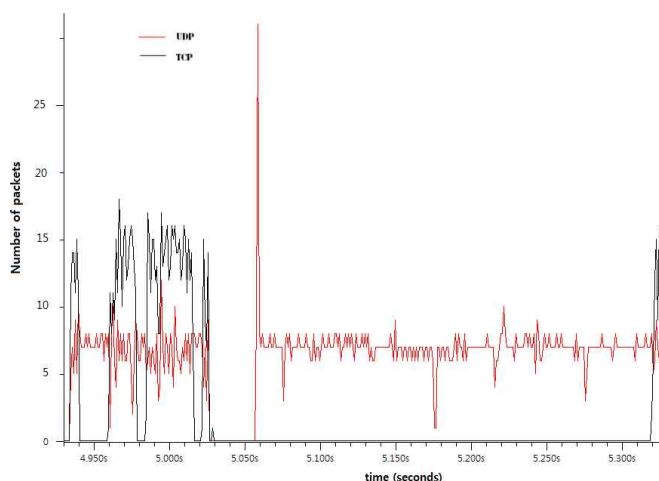


Figure 2. Number of packets UDP and TCP

2) **Change of hold-off time according to the protection switchover time:** For the purpose of looking over change of performance, we evaluate the performance varying the protection switchover time at transient state. We change protection switchover time using Manual switch function. Since the protection switchover time is inaccurate, we do not

change the CCM interval but use Manual switch function and a halt of Working TESI. If administrator commands Manual Switch to the BEB with halt time parameter through system call, BEB halts the working TESI while parameter's time and do switchover to designated Protection TESI. The Iperf client transmits TCP segments with 1420 Maximum Segment Size (MSS). As shown the Table 1, the following attributes exhibited that the fast retransmission is occurring not until 1ms. Even though the protection switchover time is faster (or less) than 50ms, TCP streams have the same average hold-off time until 3ms. It is difficult to avoid performance degrading at the transient state.

Table 1.

Protection time	max, min[Mbps] throughput	Average Lost packets of client	RTO	Average Hold-off time
150ms	156, 112	12	yes	291ms
100ms	156, 112	12	yes	291ms
50ms	156, 112	12	yes	291ms
27ms	156, 112	12	yes	291ms
11.5ms	156, 112	12	yes	291ms
3ms	156, 112	12	yes	291ms
1ms	156, 112	10.3	yes or no	219ms
0.1ms	156, 146	0.1	no	0.1ms

C. Performance degradation of TCP over PBB-TE

In PBB-TE networks, protection switching on the working lines can cause the loss of TCP segments during switchover time and force to retransmit after RTO. Even underlying network provide fast protection in less than 50ms, for the switchover time, TCP performance has not been improved.

The degree of degradation in TCP performance is depend on RTO time because the sender always waits an ACK until RTO to retransmit data. Although protection path is protected by backup path within a switchover time, the sender does retransmit segments after RTO is expired. As a result, the protection switchover time does not impact on the TCP performance. In the tests of uni-direction, bi-direction, and disable Nagle's algorithm, the TCP has the same performance depending on RTO. TCP requires fast retransmission mechanism for protection switchover time.

When the packet loss is invoked during the switchover time, TCP assumes that congestion is occurred in working line. TCP has no ability to distinguish the cause of RTO. RTO is a crucial factor in TCP congestion control. The congestion control is performed by TCP during switchover. Even if packet loss is not a target of congestion control, RTO is triggered due to the protection switchover. In particular, it is possible that TCP unnecessarily reduce the congestion window size due to the non-congestion losses. When TCP sends burst packets in switchover time, multiple packets loss or even one packet loss cause RTO to a decreasing performance which can be inevitable. If consecutive RTO occurs, the RTO timer will continue to increase twice that of before by increasing exponentially the back-off value.

IV. MODIFIED RTO MECHANISM FOR PBB-TE

While some TCP-based application can tolerate a sub 5s recovery time, Industrial Ethernet Network needs to provide reliable control for the plant with high precision without any network failure [3]. Due to the strict constraints and special purpose, each industrial protocol is designed to meet the requirements. There are many different requirements on the resiliency of Ethernet of constraints for a range of applications and their contingent requirements.

Most Carrier class Ethernet requires the stringent QoS requirements with a sub 50 ms recovery time. Taking into account performance degradation of TCP, we experiment 50ms RTO for PBB-TE.

We experiment the RTO mechanism in PBB-TE with 27ms protection time. As shown Fig.3, the result of usage of the RTO mechanism is divided into two cases due to the order of segment's sequence and shows lower performance degradation compared with TCP Reno.

A. Case 1 (in-order segment)

To prevent ACK segments from creating extra traffic, the receiver needs to delay sending an ACK segment if there is only one out in-order segment. If the server receives an in-order segment, the server delays sending an ACK segment during 40ms. As a result total 90ms, including 50ms RTO and 40ms delay, is needed to proper TCP services. After the client received the delayed ACK, the client's slow start mode begins.

B. Case 2 (out-of-order segment)

In other case, after 50ms RTO retransmission occurs, a retransmission segment with out-of-order sequence makes immediate ACK without delayed ACK. In the case TCP retransmission is serviced after the ACK.

If the retransmission makes consecutive three out-of-order segments, the client receives three duplicate ACKs and executes fast retransmission.

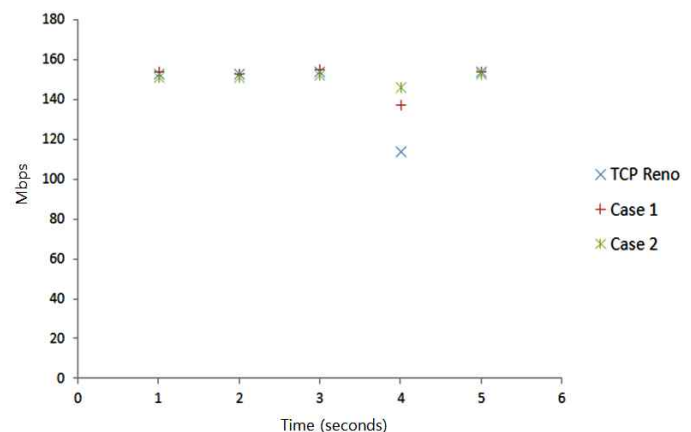


Figure 3. Performance of 50ms RTO

V. CONCLUSIONS

Different applications demand a certain resilience requirements to fit their quality of service requirements. One of the most stringent QoS requirements is needed in Industrial Ethernet Network. Controlling some high precision industrial

machinery needs a few milliseconds recovery time to control the plant in safety-critical and hazardous environments. The importance of fault detection and recovery is great impact on the Carrier class network services. On the other hand, TCP performance is affected by RTO time instead of the recovery time in the underlying network. Even if transport network provides the fast protection time, TCP is not affected. Depending on the circumstances of the underlying network, TCP RTO should be adjusted according to the network faults. The experiment results indicate that the use of the modified RTO did significantly improve TCP performance as long as the failures can be restored within the recommended period

REFERENCES

- [1] Bellcore, GR-499-CORE: Transport Systems Generic Requirements(TSGR): Common Requirements. December, 1998.
- [2] Y. Qiang, M.H. MacGregor, "Restoratoin Objectives for Internet Backbone links", CNS, pp.91-98, 2008.
- [3] M. Huynh, S. Goose, P. Mohaptra, "Resilience technologies in Ethernet", Computer Networks: The International Journal of Computer and Telecommunications Networking, Volume 54 Issue 1, pp57-78, Jan. 2010
- [4] Virtual Bridged Local Area Networks - Amendent 10: Provider Backbone Bridge Traffic Engineering, IEEE Std 802.1Qay, 2009.
- [5] Virtual Bridged Local Area Networks - Amendent 6: Provider Backbone Bridge, IEEE Std 802.1ah, 2008.
- [6] Ludwig, R. and A. Gurtov, "The Eifel Response Algorithm for TCP", RFC 4015, Feb. 2005.
- [7] P. Sarolahti, M. Kojo, "Forward RTO-Recovery (F-RTO): An Algorithm for Detecting Spurious Retransmission Timeouts with TCP", RFC 5682, Sept. 2009.
- [8] V. Paxson, and M. Allman. Computing TCP's Retransmission Timer, RFC Standards Track-2988, Nov. 2000.
- [9] Dongyan Chen, Sachin Garg, Kishor S. Trivedi, "Network Survivability Performance Evaluation: A Quantitative Approach with Applications in Wireless Ad-hoc Networks", MSWiM, pp.61-68, Sept. 2002.