**LETTER**

Performance Evaluation of TCP/IP over SCTP/IP Parallel Networking with Parallel Route Transmission*

Shota YAMADA, Student Member, Daisuke ISHII, Member, Satoru OKAMOTO, and Naoaki YAMANAKA, Fellows

**SUMMARY** In this paper, Transmission Control Protocol/Internet Protocol (TCP/IP) over Stream Control Transmission Protocol (SCTP/IP) parallel transmission system is proposed to realize large TCP/IP throughput. The proposed system enables SCTP/IP connection between switches by protocol stacking. The proposed system is implemented on a software switch to evaluate its performance. The evaluation result indicates that the proposed system can achieve 90% throughput compared with serial transmission when the delay difference among parallel routes is 20 msec.

**Key words:** SCTP, parallel data transmission, encapsulation

1. Introduction

With appearance of cloud computing, large amount of data is transmitted at high speed nowadays. 40 Gigabit and 100 Gigabit Ethernet (40 GE and 100 GE) are expected to be introduced in the data center networks near future. However, the cost of wide area network (WAN) configured with long reaching 40/100 Gigabit interfaces will be an emerging problem because 100 GE private line is much expensive than that of 10 GE.

Parallel transmission, such as link aggregation between layer 2 switches and link bundling between layer 3 routers, is an effective solution to realize huge bandwidth transmission between two network nodes at a reasonable cost. These link-level aggregation techniques can be extended to network wide aggregation and bundling. Network wide parallel transmission requires a method of aggregating data transmissions over multiple links/routes into a single logical channel (a big pipe). One of the network wide aggregation methods for implementing a big pipe is round-robin aggregation. Parallel transmission with round-robin aggregation has been reported [1]. As reported in [1], if delay difference among parallel routes exceeds a threshold value which is determined by a window size of the TCP stack, TCP packet retransmission will occur frequently by packet reordering. As a result, TCP throughput decreases drastically.

In this paper, TCP over SCTP parallel transmission system is proposed. Novel transport stack TCP/IP over Stream Control Transmission Protocol (SCTP) [2]/IP is introduced in the proposed system to improve the TCP/IP throughput. TCP/IP is used because it is generally used by almost all data transmission applications. SCTP provides better performance than TCP [3] when packet reordering occurs. With these technologies, single TCP session over multiple SCTP/IP connections with round-robin aggregation is realized. Throughput of the proposed parallel transmission system is evaluated.

2. The Proposed Parallel Transmission System

2.1 Conventional TCP/IP over Round-Robin Aggregation Methods

To figure out the characteristics of TCP/IP delay difference among parallel routes, TCP/IP throughput with round-robin aggregation is experimentally evaluated. Figure 1 shows a schematic diagram of the experiment system. Both of the end host (PC1/PC2) are connected to software based Layer-2 (L2) switches (SwitchA/SwitchB) via 1 Gbps Ethernet link. The round-robin aggregation algorithm is implemented to both L2 switches. L2 switch transmits every frame from 1 Gbps input link to several 100 Mbps output links in round-robin manner. Every frame from 100 Mbps links are multiplexed into 1 Gbps link. Under this environment, TCP/IP throughput between hosts is measured with SACK. Link delay of one 100 Mbps link is changed from 0 ms to 10 ms to emulate actual delay differences among routes in WAN.

Figure 2 shows the throughput versus delay characteristics. 2/3 Parallel means two/three parallel routes between Switches.

As shown in Fig. 2, throughput decreases drastically when the delay exceeds 2 msec. Additionally, TCP through-
Delay [msec]

Fig. 2 Throughput versus delay, link delay of one link is changed from 0 ms to 10 ms.

put of three parallel links is worse than two parallel links. This means that retransmission control of TCP/IP had worked frequently due to the packet reordering. In this experiment, 100 Mbps links were connected as parallel routes. If a high-speed link such as 1 Gbps or 10 Gbps is applied for parallel links, value of delay difference which throughput rapidly decrease is 1/10 or 1/100 of 2 msec by simple scaling. There is a crucial problem to use TCP over round-robin parallel transmission.

2.2 TCP/IP over SCTP/IP Parallel Transmission System

In this paper, Transmission Control Protocol/Internet Protocol (TCP/IP) over Stream Control Transmission Protocol (SCTP)/IP parallel transmission system is proposed to realize large TCP/IP throughput. The proposed system enables SCTP/IP connection between switches by protocol stacking and it is applied between L2 switches. The proposed system is implemented on a software switch to evaluate its performance. As shown in Sect. 2.1, round-robin aggregation with large delay difference among routes causes drastic throughput degradation in TCP/IP communication environment. SCTP has a great feature that it provides better performance than TCP [3] when packet reordering occurs. If SCTP/IP can be applied as a data transport protocol at end hosts, throughput of SCTP/IP with round-robin aggregation will be superior to TCP/IP with round-robin aggregation. However, because SCTP is a new developing protocol, there is a problem that SCTP can not deal with almost all conventional applications. In this paper, to avoid this problem, TCP/IP over SCTP/IP parallel transmission system is proposed. The proposed TCP over SCTP parallel transmission system works as follows. Figure 3 shows a schematic structure of the proposed system.

First of all, source node transmits data to Gateway 1 by TCP/IP as well as the usual communication. When IP packet (over Ethernet frame) is received, Gateway 1 encapsulates the received Ethernet frame into an SCTP/IP packet. After frames are transmitted over parallel transmission (with round-robin aggregation) by using SCTP/IP, terminal Gateway 2 decapsulates frames from received SCTP packets and transmits frames to a destination node. Finally, the destination node receives IP packet as usual, and the communication is completed. As a result, end user can transmit TCP/IP data with parallel transmission without any application modification.

3. Implementation and Throughput Evaluation of the Proposed System

3.1 Experimental Setup

The proposed system is implemented to software L2 switches to examine delay-throughput characteristics. SCTP protocol stack and TCP/IP over SCTP/IP functions are also implemented to L2 switches. L2 switch acts as a gateway shown in Fig. 4. Single SCTP connection is established between gateways. In this implementation, lksctp-2.6.24-1.0.8 [4] is used as SCTP protocol stack. Figure 4 shows a schematic diagram, and Fig. 5 shows a protocol stack of the experimental system. Network bandwidth of parallel links between gateways is scaled down from 1 Gbps to 100 Mbps, since main purpose of this experiment is to verify the system concept. In the future, parallel link will be extended to 1 Gbps or 10 Gbps or more.

3.2 Experimental Results

TCP/IP throughputs are measured with three kinds of TCP/IP transmission systems. First is a TCP/IP over 300 Mbps serial link transmission system where link bandwidth is narrowed down from 1 Gbps to 300 Mbps, second is a TCP/IP over TCP/IP (ToT) parallel routes transmission
system [5], last is the proposed TCP/IP over SCTP/IP parallel routes transmission system. ToT is a technique which constructs a virtual connection by aggregating flow between two nodes like PC or router, and transmit as a single TCP connection [5]. Original delay of all links in this environment is 0.5 msec and SACK is used in every TCP/IP transmission. Figure 6 shows throughput versus delay characteristics, where delay is inserted only at one link and changed from 0 msec to 20 msec. Throughput of ToT decreases gradually when the difference of delay exceeds 5 msec, and it falls below to the half when the difference of delay exceeds 10 msec. On the other hand, in the proposed method, throughput is sensitively decreasing as compared with ToT. As reported in [3], in the case of packet loss, SCTP clock out new data without the change of its congestion window where this behavior is not found in TCP on detection of packet loss. Therefore, decrease of throughput can be reduced by using TCP over SCTP when only the link delay of one link is extremely large. Throughput of the ToT system and the proposed system saturates between 0 msec and 4 msec because of the encapsulation overhead caused by software processing. In this implementation, each packet is encapsulated by modified VTun [6] which is usually used as a virtual private network (VPN) tool. Therefore, it is necessary to modify and improve existing VTun programs if the experiment as shown above is done under the high speed environment. Throughput of the proposed system between 0 msec and 4 msec are 60% of the serial transmission which packet reordering does not occur. However, difference of throughput becomes small as delay increases. It achieves up to about 80% in 0 msec, and about 90% in 20 msec. Thus, the proposal system can achieve throughput equal to serial transmission at a low cost especially in the global scale network environment such as long-distance communication. In addition, 10 msec delay difference in 100 Mbps correspond to 100 msec in 10 Gbps by simple scaling.

4. Conclusion

TCP/IP over SCTP/IP parallel transmission system with round-robin aggregation system was proposed. An evaluation of its throughput-delay characteristics indicated that the proposed system can achieve throughput equal to serial transmission at a low cost in the global scale network environment.

Acknowledgment

This work is partly supported by “Lambda Access Project” funded by the National Institute of Information and Communication Technology (NICT).

References