## §3. Estimation of the Advanced TCP/IP Algorithms for Long Fat Network

Yamamoto, T.

ITER collaboration needs the wide bandwidth network for the fast data transfer. However, the throughput might be too small to transfer the bulk data, because the delay of the signal is not negligible. This problem comes from the mechanism of the communication protocol, TCP/IP. The effective throughput is equal to the bandwidth on LAN; however, the *window size* limits the throughput on WAN whose RTT is large due to the delay of the acknowledge packet. *Window size* is the data size which the sender can send to the receiver without acknowledgement in Fig. 1. The distance between Japan (Rokkasho) and France (Cadarache) is over 20,000 km. (via U.S. and U.K.) Round-Trip Time (RTT) of the signal is approximately 300 ms. This delay is very large compared with the RTT of LAN (less than 1 ms). The throughput using the ordinary personal computer (PC) might be under 10 Mbps.

The actual window size is defined as following equation;

$$win = \min(awin, cwin), \quad (1)$$

where *win* is the window size for the connection, *awin* is advertised window size which is advertised by the receiver, and *cwin* is congestion window size which is determined by the sender's algorithm. The *awin* reflects the receiver's condition. On the other hand, *cwin* reflects the network condition. The sender estimates the available bandwidth for the connection at all times. When the network begins to crowd, the sender should reduce the sending rate to avoid the massive packet loss. It is called congestion avoidance control.

In this study, long fat pipe network (LFN) environment was emulated by the Linux-based PC. The emulator could specify the delay and the ratio of packet loss. Figure 2 shows the schematic diagram of this emulation. The data was transferred from *PC: A* to *PC: B* through the emulator. The emulator was worked as bridge. *PC: A*, *PC: B* and the emulator are the commodity PC servers. The software of the emulator was composed of Linux and *netem* module. *Netem* is a network emulator developed for investigating the network protocol. The measuring tool was Distributed Benchmark System (DBS). DBS is a powerful tool for network evaluations.

The various TCP congestion avoidance algorithms were investigated with RTT of 300 ms which was estimated between ITER site in France and ITER satellite site in Japan. The sender's and the receiver's socket buffer size were enlarged enough to avoiding the limitation of *awin*. Table 1 summarizes the average throughputs with and without packet loss, where RTT = 300 ms. Each throughput had been measured five times and measurement time was 10 min. The throughput of CUBIC-TCP, Hamilton TCP and Scalable TCP were more than 600 Mbps with packet loss, $p_{loss} = 10^{-6}$. These algorithms would be suitable for LFN. Although the throughput of BIC-TCP was also more than 600 Mbps without packet loss, it decreased to under 500 Mbps with packet loss, $p_{loss} = 10^{-6}$. The standard algorithm, Reno, is not suitable for LFN. The advanced algorithm, CUBIC-TCP, Hamilton TCP and Scalable TCP can be more suitable for LFN such as the ITER international collaboration. [1)]

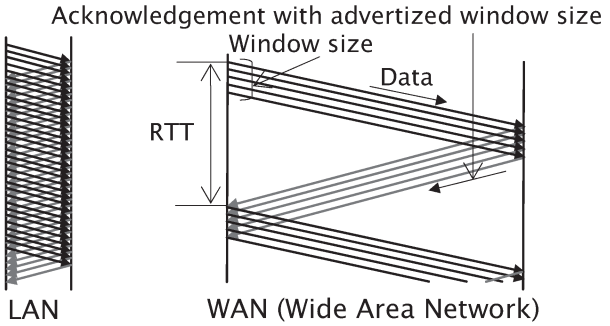


Fig. 1. TCP data handling on LAN and WAN.

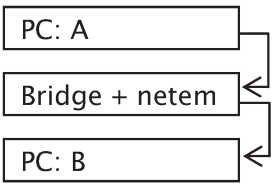


Fig. 2. The schematic diagram of the test environment.

Table 1.
The average throughputs (Mbps) for the various TCP congestion algorithms with and without packet loss, where RTT = 300 ms.

| The ratio of the packet loss | 0 | $10^{-6}$ |
|---|---|---|
| Reno | 320 | 150 |
| BIC-TCP | **710** | 490 |
| CUBI-TCP | **690** | **610** |
| Hamilton TCP | **840** | **610** |
| HighSpeed TCP | 460 | 160 |
| Scalable TCP | **870** | **740** |

The throughput which is more than 500 Mbps is emphasized with bold face.

1) Yamamoto, T., Fusion Eng. Des. **83**(2008)516.