# Wireless TCP for High-speed Mobile Communications

*Kazunori Yamamoto, Hideharu Suzuki,*
*Aya Hokamura and Katsumi Sekiguchi*

*The speed of mobile communications networks is increasing rapidly along with progress being made in wireless technology. Wireless TCP technology has been developed to maximize the use of radio bearer speed and play an important role in providing users with high-quality service in high-speed mobile communications networks.*

Technology Reports

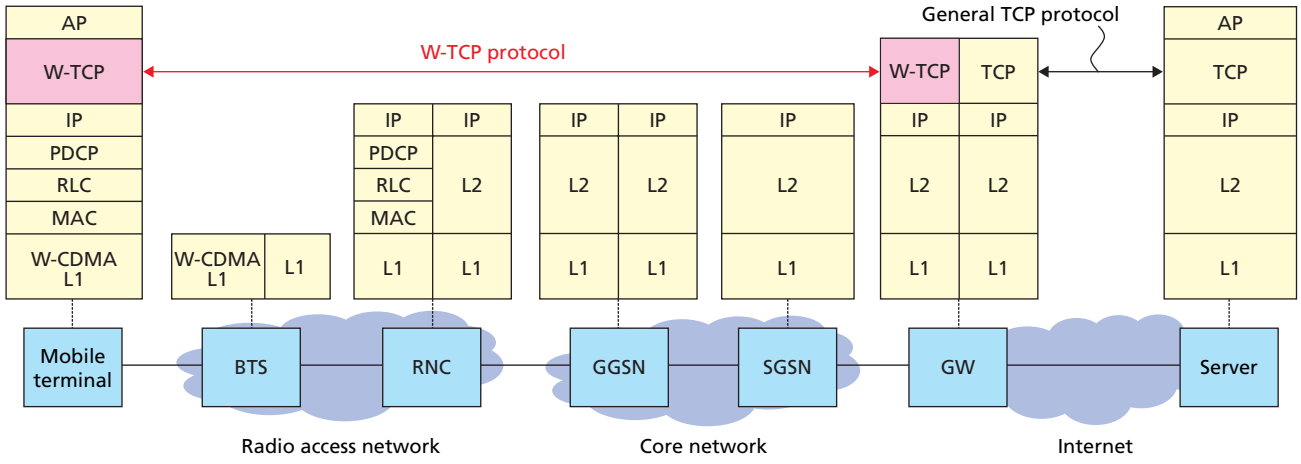
## 1. Introduction

TCP is positioned in the transport layer of the OSI basic reference model, and is a protocol providing reliable data communications in end-nodes. DoCoMo uses TCP as a basic technology to provide primary services such as i-mode, i-motion, and full browser services.

Since the R&D of TCP originally focused on wired networks, performance tends to be insufficient when applying conventional TCP to mobile communications networks, which have different communications characteristics than wired networks. As a case in point, a sudden increase in delay in mobile communications networks results in a TCP retransmission timeout, thereby reducing throughput[*1]. DoCoMo has developed wireless TCP (W-TCP) technology [1] optimized for mobile communications networks to provide high-quality services.

**Figure 1** shows the network configuration for DoCoMo's 3G and beyond mobile communications systems. Mobile communications networks consist of radio access networks and a core network, and are connected to external networks (such as the Internet) through gateways. TCP connections are terminated in gateways, and W-TCP is applied to TCP connections between gateways and mobile terminals. New functions are constantly being added with parameters optimized to gateways

*1 Throughput: Effective amount of data transmitted without error per unit time.




AP: APplication
BTS: Base Transceiver Station
GGSN: Gateway General packet radio service Support Node
GW: GateWay
MAC: Medium Access Control
PDCP: Packet Data Convergence Protocol
RLC: Radio Link Control
SGSN: Serving General packet radio service Support Node

**Figure 1 Network configuration of 3G mobile communications systems**

and mobile terminals for ensuring high-quality service over rapidly evolving wireless networks.

This article describes the issues and solutions of W-TCP, as well as DoCoMo's development work in the field of high-speed mobile communications networks.

## 2. Issues and Solutions of W-TCP in High-speed Mobile Communications Networks

### 2.1 Depletion of Send and Receive Buffers

TCP send and receive buffers in gateways and mobile terminals become depleted when bandwidth delay product[*2] increases due to higher radio bearer speed. It is generally known that the sizes of send and receive buffers must exceed the bandwidth delay product to make the most of network communications speed. To maximize the use of radio bearer speed with W-TCP, the sizes of send and receive buffers are optimized according to the maximum radio bearer speed.

However, the radio bearer speed varies with wireless quality and congestion in a cell, and the maximum radio bearer speed is therefore not always available to the user. When radio bearer speed is low, the TCP window size[*3] becomes excessive, and packets are retained in the intermediate node, giving rise to the following issues related to service quality and efficiency in network resources use.

- Packets issues in the intermediate node increase the end-to-end delay, so that response time of real-time applications such as interactive Web applications and Push-to-talk over Cellular (PoC)[*4], deteriorates.
- The efficiency of buffer use deteriorates due to an increase in low-speed user packets accumulated in the buffer space of the intermediate node. When the number of low-speed user packets at the intermediate node increases, the necessary buffer space is not allocated to other users and throughput may deteriorate.
- After a user cancels a download (or upload), unnecessary packets equivalent to a single window size are passed through the wireless link until TCP transmission terminates. For example, if the user cancels a download with the window size expanded to 64kbytes, 64kbytes of unnecessary packets pass through the wireless link.

The number of TCP data packets sent must be controlled according to the radio bearer speed in order to solve these issues. Many schemes have been proposed to achieve this end, and may be classified based on the allocation of functions into the following 3 approaches:

---

*2 Bandwidth delay product: An indicator of network characteristics. Given by the product of communications speed between end-nodes and signal round-trip time.
*3 Window size: The size of the buffer space used for TCP flow/congestion control. Determines the maximum number of packets that may be transmitted without receiving the ACKs.
*4 PoC: Mobile phone service in IP network. Standardizing is underway by the Open Mobile Alliance (OMA) and other organizations.

1) End-node Approach

The number of data packets sent is controlled in the end-nodes terminating a TCP connection. In the DoCoMo mobile communications system (Fig. 1), the gateway or mobile terminal is primarily responsible for control. Since the number of data packets sent is controlled directly at the transmission source, it can be accurately adjusted; however, network status cannot be accurately understood with this method.

2) Intermediate Node Approach

The number of data packets sent is controlled in the intermediate nodes in the network. Congestion control applying Random Early Detection (RED)[*5] is used in the radio bearer controller, and the number of data packets sent may be adjusted according to the radio bearer speed [2]. On the other hand, since TCP congestion control is activated indirectly, the number of data packets sent cannot be finely adjusted.

3) Hybrid Approach

The end-nodes terminating a TCP connection and intermediate nodes are used together in controlling the number of data packets sent. Since the advantages of the end-node and intermediate node approaches are utilized in combination, the number of data packets sent can be optimally controlled. Conversely, since development is required for multiple types of nodes, the scale of development work required is considerable.

DoCoMo currently uses the intermediate node approach to handle the issues at low communications speeds, but is investigating the end-node and hybrid approaches in view of the need to provide a higher-quality service [3].

## 2.2 Spurious Timeouts

A sudden increase in delay occurs in mobile communications networks due to such factors as increased layer 2 retransmissions due to deterioration in wireless quality, processing delays associated with switching between wireless channels, and handovers. TCP retransmission timeouts occur due to increased delays, even though no packets are lost. This phenomenon is referred to as a 'spurious timeout', and when it occurs, unintended congestion control reduces the number of packets sent, and thus throughput deteriorates.

This issue can be avoided by using W-TCP with Forward-Retransmission Time Out recovery (F-RTO) [4] in the Charging and Protocol Conversion Gateway (CPCG)–the i-mode gateway. F-RTO is an algorithm that detects spurious timeouts and prevents unintended convergence control.

Other algorithms are available to detect spurious timeouts [5][6]. F-RTO differs from these algorithms in that it does not use the TCP option, and is therefore able to detect spurious timeouts without increasing traffic on the network. Consequently, F-RTO is considered the most suitable method for mobile communications networks with limited wireless resources.

**Figure 2** shows the results of evaluating F-RTO at a spurious timeout by using a W-CDMA emulator to simulate the wireless environment. The graph shows the average throughput when a single spurious timeout occurs during a TCP slow start and download of a 2Mbyte file. When F-RTO was applied, throughput was increased by 6%, 57%, and 92% at radio bearer speeds of 384 kbit/s, 3.6 Mbit/s, and 14 Mbit/s, respectively. These results clearly show that the use of F-RTO has a positive effect on high-speed mobile communications networks [7].

## 2.3 Asymmetry in Radio Bearer Speed

When radio bearer speed is increased in only one direction in mobile communications networks, such as with High Speed

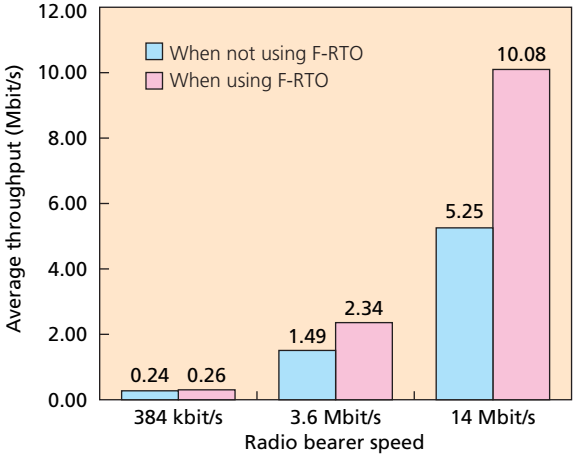



**Figure 2 Results of F-RTO evaluation at spurious timeout**

*5 RED: An algorithm used to stochastically drop packets accumulating in the router to avoid congestion, or to notify congestion to end-nodes.

Downlink Packet Access (HSDPA)[*6], radio bearer speed is referred to as being 'asymmetrical.' Asymmetry in radio bearer speed results in a bottleneck in communications speed in the ACKnowledgement (ACK) direction, and restricts TCP throughput when the communications speed in the ACK direction is extremely low compared with that in the data direction. DoCoMo has avoided this issue by raising the maximum radio bearer uplink speed to 384 kbit/s in association with increased downlink speed using HSDPA, and theoretically can accommodate throughput of 28.8 Mbit/s with TCP.

In addition to increasing radio bearer uplink speed in this manner, such methods as header compression, ACK filtering, and ACK convergence control [8] have also been proposed to solve this issue. Since these methods reduce the volume of ACK traffic itself, they are effective in the use of wireless resources. When considering that the volume of ACK traffic increases in association with increased radio bearer speed, these technologies are expected to become even more necessary in the future.

## 3. Conclusion

This article has described the issues and solutions of W-TCP in high-speed mobile communications networks, as well as the current state of DoCoMo's W-TCP development work. We will further investigate the technical solutions described above as part of efforts to improve the quality of service provided over high-speed mobile communications networks.

### REFERENCES

[1] H. Inamura, G. Montenegro, R. Ludwig, A. Gurtov and F. Khafizov: "TCP over Second (2.5G) and Third (3G) Generation Wireless Networks," RFC 3481, Feb. 2003.

[2] Y. Morihiro, Y. Kato and Y. Ishikawa: "A Study of TCP Transmission Window Control Based on W-CDMA System," Proc. IEICE Gen. Conf. '03, B-5-118, Mar. 2003.

[3] K. Yamamoto, H. Suzuki, N. Ishikawa, M. Miyake and H. Inamura: "A TCP Flow Control Scheme for 3G Mobile Communication Networks," Proc. ICCCN'06, pp. 229-236, Virginia, Oct. 2006.

[4] P. Sarolahti and M. Kojo: "Forward RTO-Recovery (F-RTO): An Algorithm for Detecting Spurious Retransmission Timeouts with TCP and the Stream Control Transmission Protocol (SCTP)," RFC 4138, Aug. 2005.

[5] R. Ludwig and M. Meyer: "The Eifel Detection Algorithm for TCP," RFC 3522, Apr. 2003.

[6] E. Blanton and M. Allman: "Using TCP Duplicate Selective Acknowledgement (DSACKs) and Stream Control Transmission Protocol (SCTP) Duplicate Transmission Sequence Numbers (TSNs) to Detect Spurious Retransmissions," RFC 3708, Feb. 2004.

[7] K. Sekiguchi, A. Hokamura, K. Yamamoto, H. Suzuki, N. Ishikawa and O. Takahashi: "Evaluation of Spurious Timeout Detection Algorithms over 3G Mobile Packet Access Network," IPSJ Technical Report, 2006-MBL-36, Feb. 2006.

*6 HSDPA: A high-speed downlink packet transmission system on W-CDMA. Maximum downlink transmission speed under the 3GPP standard is about 14 Mbit/s. Optimizes the modulation method and coding rate according to the radio reception status of the mobile terminal.