# Successful High-precision QoS Measurement in Widely Deployed 10-Gbit/s Networks Based on General-purpose Personal Computers

*Kenji Shimizu*†, *Katsuhiro Sebayashi, and Mitsuru Maruyama*

**Abstract**

In this article, we describe techniques for building 10-Gbit/s wire-rate precise network measurement systems that can measure the quality of service (QoS) of high-speed networks precisely and quickly, even with non-sampling measurements. We describe our 10-Gbit/s PCI (peripheral component interconnect) network interface card for general-purpose personal computers and present experiment results that verify the effectiveness of our measurement systems.

## 1. Network measurement in high-speed application era

The spread of high-speed networks has made it easy to use broadband applications such as high-quality video streaming applications via the Internet. If we consider video streaming services, the bit-rate depends on the video format and quality, varying from some tens of kilobits per second to several megabits per second. Therefore, considering multiple client accesses in Internet service providers' networks, the total amount of traffic could reach over 1 Gbit/s. In such an environment, network operators must know the network status in detail in order to provide stable service for users. That is why we have developed a precise, high-speed network measurement technique that can be applied to a wide range of networks.

† NTT Network Innovation Laboratories
Musashino-shi, 180-8585 Japan

## 2. 10-Gbit/s PCI network interface card with network measurement extensions

NTT Laboratories has developed a 10-Gbit/s PCI (peripheral component interconnect) network interface card (NIC) with network measurement extensions (measurement NIC). It works both as a normal NIC for a 10-Gbit/s network and as a measurement NIC. Its extensions enable precise measurements of the network status and quality of service (QoS). It will let us develop a reliable network measurement system based on inexpensive general-purpose personal computers (PCs).

For example, the headers of all the packets flowing in a 10-Gbit/s network can be captured without any sampling and used for analysis. Furthermore, precise timestamps based on timing signals using GPS (global positioning system) can be appended to packets at both the sender and receiver sides. These timestamps are useful for calculating network delays and jitter. Furthermore, while acting as a streaming server's NIC, our card can append precise timestamps to the streaming packets entering and leaving the server. It makes possible novel application serviceability eval-

| | |
|---|---|
| Line interfaces | Three protocols are supported:<br>- 10GbE LAN-PHY<br>- 10GbE WAN-PHY<br>- OC-192c POS (packet over SONET/SDH) |
| Connector, cable | - LC connector<br>- Single-mode fiber |
| Bus interface | PCI-X (rev.1.0a)<br>64 bits/133 MHz |
| Dimensions | 126 × 266 (mm) |
| External clock input | 10 MHz ×1, 1 PPS ×1,<br>SMB connector |

10GbE: 10 Gigabit Ethernet
LAN: local area network
PHY: physical layer
WAN: wide area network
SONET/SDH: synchronous optical network, synchronous digital hierarchy
LC: small form-factor fiber optic connector
PPS: pulses per second
SMB: subminiature version B, a coaxial RF (radio frequency) connector

Fig. 1. 10-Gbit/s PCI network interface card with network measurement extensions.

uation systems that perform network measurement while simultaneously providing streaming services. A photograph of the card is shown in **Fig. 1**, which also lists its specifications.

## 3. Classification of network measurement techniques

Network measurement can be classified into passive and active measurement. Although each technique is usually chosen for a different purpose, our measurement NIC has hardware-assisted extensions to support both of them.

- Active measurement: a system sends measurement-specific packets (probe packets) into the network under test. Then, by analyzing their status such as inter-packet gap (IPG) and arrival time at a receiver, it obtains the network characteristics, including network delay and maximum available bandwidth.

- Passive measurement: a system monitors the traffic, which is split by utilizing a router's mirroring port or optical splitters. It obtains network usage in detail, such as the usage rate of network bandwidth by each flow and protocol distribution.

## 4. Challenges for a high-speed network

The following methods are usually used to measure high-speed network characteristics such as a bandwidth of 10 Gbit/s.

(1) Sampling/non-sampling

Sampling is a nondestructive technique used in commercial routers for collecting information about the amount of traffic travelling through the router. If only one packet out of 1000 packets, for example, is examined, then the amount of collected data can be reduced, so the measurement system can easily analyze the data at the lower load. However, detailed characteristics cannot be obtained if the sampling rate is very low. On the other hand, in non-sampling measurement, where all information is acquired, the system load for collecting all of the packets is very high. Thus, it has been difficult for network routers, switches, and inexpensive general-purpose PCs to perform such measurements.

(2) Header and payload analysis

In header analysis, a system examines only a small amount of data from the head of each packet while ignoring the rest. As a result, the system's load is low because there is less data to analyze. On the other hand, payload analysis examines the entire contents of each packet: this is useful for detecting computer viruses and for intrusion detection systems.

## 5. Network measurement extensions of the measurement NIC

We have implemented the following hardware-assisted extensions in our measurement NIC to make the precise measurement of high-speed networks easy (**Fig. 2**).

(1) Wire-rate packet handling function

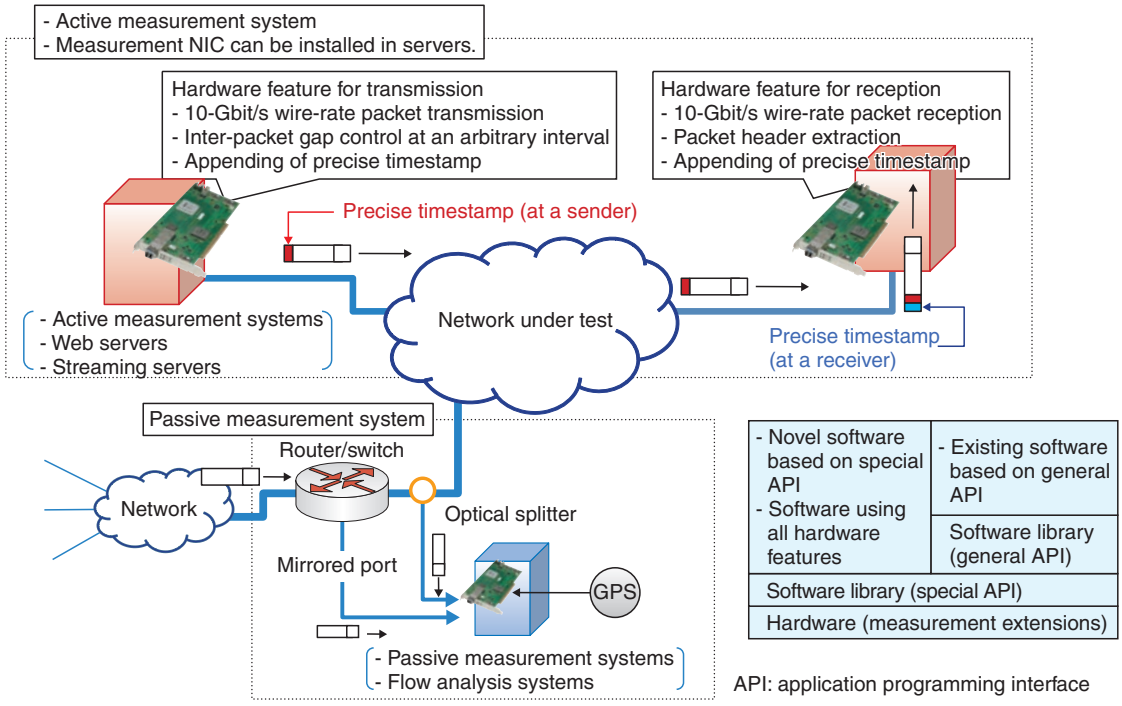


Fig. 2. Hardware-assisted features.

This function achieves 10-Gbit/s non-sampling analysis even with a system based on a general-purpose PC. It includes hardware-assisted header examination (called header extraction) and efficient DMA (direct memory access) data transfer mechanisms to fully utilize the limited bandwidth of the PCI bus.

(2) Header extraction function

This function can extract only the header, which reduces the data analysis amount. Since the extraction size can be set to an arbitrary value, our measurement NIC can also be useful for payload analysis.

(3) Precise timestamp function

Our measurement NIC can generate precise timestamps based on external timing signals from GPS receivers, for example. Timestamps are appended to packets at both the sender and receiver sides. They enable precise measurement of network delays and jitter. Even when a GPS signal is not used, the traffic burstiness can be precisely measured by using the internal clock on our measurement NIC.

(4) Packet capturing software library

The packet capturing software library has application programming interfaces compliant with the de-facto standard library "libpcap". By using this library, existing network measurement software based on libpcap can easily take advantage of our measurement NIC's hardware-assisted functions.

(5) Traffic playback function

The hardware-assisted traffic playback function can utilize stored characteristics (from previously captured traffic) to generate test traffic and inject it into a network under test while emulating the stored characteristics precisely [1]. In detail, protocol fields, IPGs, and packet lengths can be exactly matched thanks to the hardware-based IPG control with 5-ns resolution. This function is useful because it lets network operators measure the network quality when an application's traffic flows through a network without deploying actual servers in each site.

(6) Network measurement and traffic controls in servers

Our measurement NIC can work as a normal NIC in end-node servers such as streaming servers while its measurement features are simultaneously being utilized. For example, it can append timestamps to packets carrying audio and video signals for application-specific traffic-QoS measurement.

## 6. Development of precise measurement systems for 10-Gbit/s networks

We have developed a precise measurement system

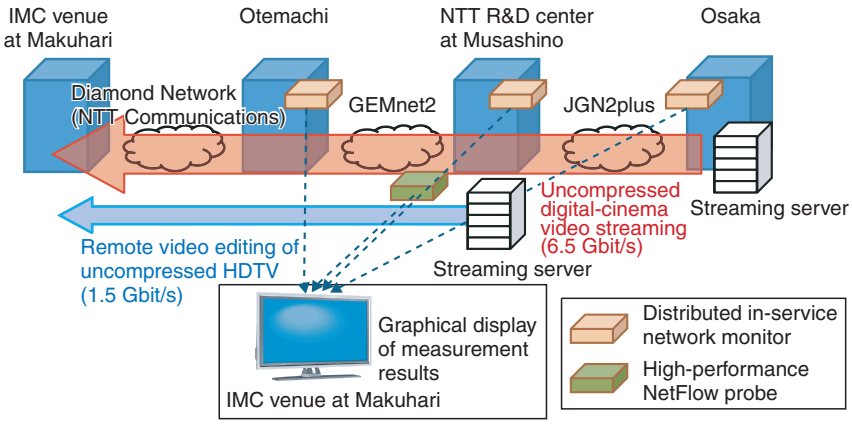


Fig. 3. Network configuration of the experiment at IMC 2008.

for high-speed networks by using these network measurement extensions.

### 6.1 Improved measurement software for NetFlow analysis

We have modified existing open-source software so that it can take advantage of the ability of our measurement NIC to generate NetFlow packets from captured packet headers. Such software is called NetFlow probes. NetFlow is a well-known protocol used to collect traffic flow information that is used to obtain the amount of data, duration time, and other properties of each flow of packets. The improvement was achieved by making slight modifications that changed the referred software library to a modified one. As a result, the NetFlow probe can generate NetFlow packets without sampling any packets from the traffic flowing at the wire-rate speed of 10 Gbit/s, although the performance depends on the number of flows processed.

### 6.2 Distributed in-service network monitor

The systems can be deployed at multiple measurement sites where they capture all the packets and append timestamps to them. Using these timestamps at every site enables network delays, jitter, and traffic burstiness to be obtained. The measurement results are then sent to the central management software to graphically show the characteristics of the traffic. This feature is useful in comparing the statuses of different measurement sites and determining the congestion points when network trouble, such as packet dropping, occurs.

## 7. Applicability evaluation in 10-Gbit/s high-speed networks

We evaluated the applicability of deployed systems at IMC Tokyo 2008 Interop Media Convergence held in Japan in June 2008.

### 7.1 Overview of IMC Tokyo 2008

At IMC Tokyo 2008, various kinds of leading-edge products and technologies toward high-quality media contents service were exhibited. Held at the same venue as Interop 2008, it let visitors collect information about trends for next-generation audio and visual media services in broadband IP (Internet protocol) networks.

### 7.2 Experimental aims

At the IMC venue, many broadband applications transmitted, stored, and distributed large amounts of traffic such as high-quality video streams over high-speed networks. We aimed to evaluate the applicability of our measurement systems in such conditions where mixed traffic characteristics were observed simultaneously in 10-Gbit/s networks.

### 7.3 Experimental setup

We connected four measurement sites (the IMC venue at Makuhari, NTT R&D Center in Musashino, and NTT sites in Osaka and Otemachi) using three different networks: 1) JGN2plus[*1] operated by NICT (National Institute of Information and Communica-

[*1] JGN2plus: R&D testbed network for next-generation network, administrated by NICT.

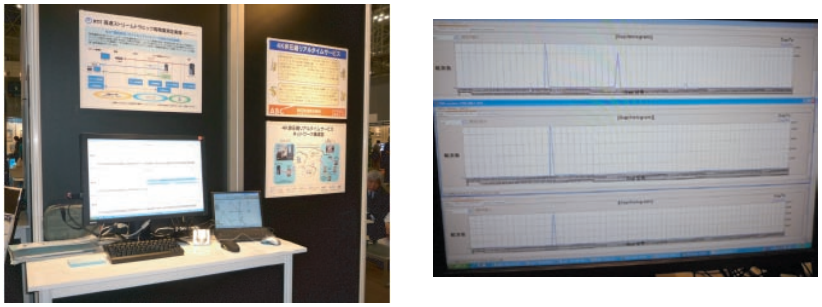

Fig. 4. Demonstration of measurements made at IMC 2008 venue.

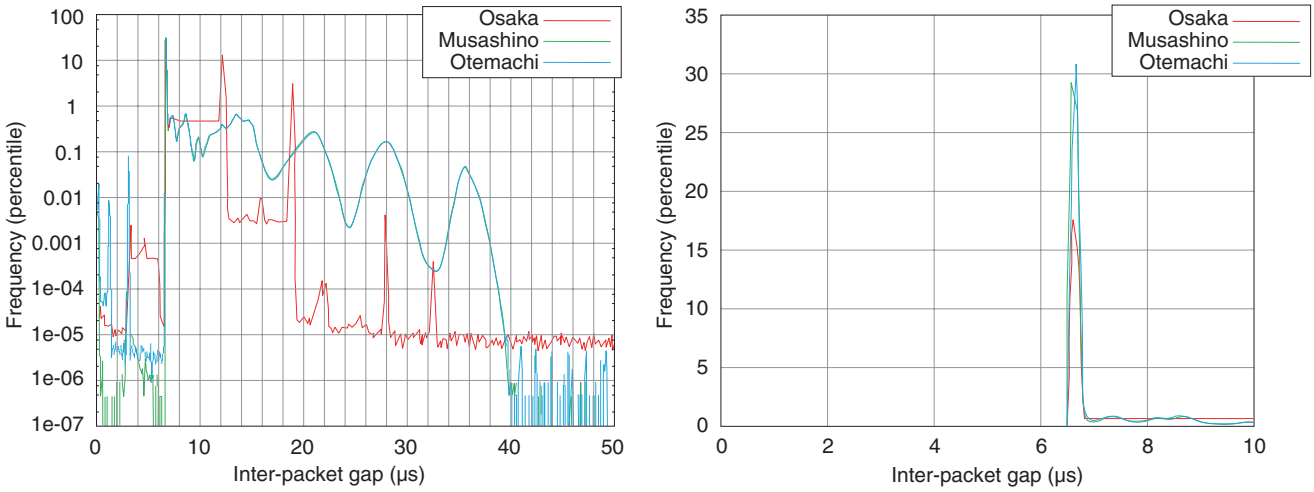


Fig. 5. Observed IPG characteristics (left: in a wide IPG range, right: magnified view).

tion Technology), 2) GEMnet2[*2] operated by NTT Labs., and 3) Diamond Network[*3] operated by NTT Communications. These networks are currently based on 10-Gbit/s Ethernet technologies (**Fig. 3**). We deployed video streaming servers at the Musashino and Osaka sites. These servers can deliver and store high-quality video streams including uncompressed high-definition television (HDTV) and 4K digital-cinema-quality video over IP networks. Using remote video editing work lets us simultaneously handle video playback streams and editing flows with different traffic characteristics over the networks. Photographs of our exhibition booth and measurement results at IMC Tokyo 2008 are shown in **Fig. 4**.

## 8. Evaluation items

From the streaming server in Osaka, an uncompressed 4K digital-cinema-quality video stream was delivered to the IMC venue at Makuhari at a bandwidth of 6.5 Gbit/s. At the same time, uncompressed HDTV video content was remotely retrieved from Musashino and delivered to the IMC venue by the video editor there at 1.5 Gbit/s. The total amount of traffic was around 8 Gbit/s, which is extremely high compared with existing streaming applications. In the experiment, commonly used existing network measurement software was also used by NTT Communications to verify its applicability to high-speed networks. We, NTT Labs., performed the following measurements.

*2 GEMnet2: Ultrahigh-speed R&D testbed network, constructed and administrated by NTT Laboratories.

*3 Diamond Network: Reliable and ultrahigh-capacity network constructed by using simple network facilities, administrated by NTT Communications. The network service is currently on trial.

### 8.1 Traffic QoS measurement at multiple measurement sites

Distributed in-service network monitoring systems were located at the Musashino, Osaka, and Otemachi sites, where they captured the streaming traffic and measured the traffic burstiness (IPG) in real time. In this configuration, we checked whether we could detect a fluctuation in traffic characteristics induced by passage through the network.

### 8.2 High-performance NetFlow analysis

By using our high-performance NetFlow probe software, we checked whether it could handle up-to-8-Gbit/s ultrabroadband traffic even when the software was mostly based on open-source software.

## 9. Results

The IPG distributions at the three sites are shown in **Fig. 5**. In Osaka, some peaks were observed at around 6-, 12-, 19-, 28-, and 32-μs IPGs, but these peaks disappeared after the packets traversed the network to Musashino and Otemachi. From the distribution pattern around the peak values there, the IPG values were dispersed into various values which, we think, may have been caused by queueing in Ethernet switches and routers in the network. We have previously reported that the IPG distribution pattern in a congested switch became broader when the switch congestion became worse [2].

Figure 5 also shows a magnified view of the IPG distribution at around 6 μs. Since 9-kbyte jumbo packets were transmitted in the experiment, the observation of 6-μs IPGs revealed that the traffic flowed at 10-Gbit/s wire-rate speed with very fine time resolution. Such bursty traffic can sometimes cause network trouble such as packet dropping when intermediate network routers overflow. Furthermore, the number of 6-μs IPGs observed at Musashino and Otemachi increased to 30% compared with 17% in Osaka, which reveals that the traffic burstiness severely worsened. These results support the applicability of our systems to QoS measurement in 10-Gbit/s networks. We successfully clarified the fluctuating behavior of traffic characteristics influenced by the network configuration in real time.

In the high-performance NetFlow analysis, we did not detect any failures to capture packets of 8-Gbit/s traffic even in non-sampling measurements. Although the number of flows was less than 10, the experiment confirmed the improvement in NetFlow probe performance.

## 10. Conclusion

Through experiments, we have achieved successful realtime measurement of detailed traffic characteristics by using our measurement NIC based on general-purpose PCs, which have previously been difficult to utilize in 10-Gbit/s networks. By widely deploying the measurement systems, network operators can obtain useful information about network status to maintain their service quality.

In this article, we mentioned only some of the many measurement features implemented in our measurement NIC. We are planning to further develop advanced network measurement systems including packet capturing systems incorporating a large amount of storage and sophisticated active measurement systems such as ones providing available bandwidth estimations based on a hardware-assisted traffic generator.

**Acknowledgment**

This research is partially funded by the National Institute of Information and Communication Technology (NICT).

## References

[1] K. Shimizu, K. Sebayashi, T. Kawano, and M. Maruyama, "The network measurement system for 10-Gbit/s stream traffic with precise traffic playback functions based on a general-purpose PC," BS-8-1, IEICE Society Conference, 2008 (in Japanese).

[2] K. Shimizu, T. Ogura, T. Kawano, H. Kimiyama, M. Maruyama, and K. Koyanagi, "Application-coexistent Wire-Rate Network Monitor for 10 Gigabit-per-Second Network," IEICE Trans. Inf. & Syst., Vol. E89-D, No. 12, pp. 2875–2885, 2006.

**Kenji Shimizu**
Research Engineer, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.E. and M.E. degrees in electronics engineering from Sophia University, Tokyo, and the Ph.D. degree in engineering from Waseda University, Tokyo, in 1998, 2000, and 2007, respectively. He joined NTT Network Innovation Laboratories in 2000, where he has been studying processing system architectures and Internet systems. His interests include high-speed protocol processing, traffic monitoring, and content delivery network technologies. He is a member of IEEE and the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.

**Katsuhiro Sebayashi**
Senior Research Engineer, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.E. degree in electrical engineering from Tokyo Denki University, Tokyo, in 1990. He joined NTT Laboratories in 1990. He is currently working on a policy-based network control architecture for network security and a network-wide traffic measurement architecture.

**Mitsuru Maruyama**
Group Leader, Senior Research Engineer, Supervisor, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.E. and M.E. degrees in electrical engineering and the Dr. degree in computer science from the University of Electro-Communications, Tokyo, in 1983, 1985, and 1999, respectively. He joined NTT Laboratories in 1985 and has been engaged in R&D of a high-definition videotex system, video-on-demand systems, and an IP-based realtime video transmission and archiving system. He is currently studying fast-protocol processing system architectures and multi-agent systems. He is a member of the IEEE Computer Society and Communications Society, IEICE, and the Information Processing Society of Japan.