Software Technology Offers Real-time Analysis of Super-fast Telecommunications at 200 Gbps

● Yuji Nomura ● Masahisa Tamura ● Toshihiro Ozawa

Smartphones and tablets have become widely and readily available, and the expanding scope of applications for data center use has stimulated the advancement of cloud services. Thus, it is becoming crucial to improve the quality of IT-network-based services. Meanwhile, ever-increasing volume of data traffic and further sophistication of system structures are leading to service disruptions such as slow data propagation speeds and difficulties in establishing network connections. For early detection and swift recovery of such service disruptions, it is essential to analyze and distinguish between their causes to know whether they are due to network (connection) quality or lie in application programs. This requires detailed, real-time behavioral analysis of each communication packet. For complicated cases, evidence of the errors may need to be accumulated for later in-depth analyses. In order to meet these requirements, Fujitsu Laboratories has developed technology that facilitates real-time analyses of communication packets on super-fast communication networks, as well as high-speed data accumulation and retrieval. The technology has been realized in the form of software that requires no expensive, dedicated hardware, and it is installed on a versatile system. This paper describes the software technology that enables analysis, accumulation and retrieval of data from high-speed communications, and it explains the system structure.

1. Introduction

As the communication traffic volume flowing through networks grows year after year, monitoring of communication quality and the detection of delays have become major challenges. Conventionally, for communication lines with communication speeds exceeding 20 Gbps, it has not been possible to analyze the quality of all communications even with expensive hardware, nor has it been possible to offer convenient and stable services to all users who make use of communication services.

Fujitsu Laboratories has developed packet collection technology that is suitable for multi-core CPUs, quality analysis technology for transport layers such as TCP and application players such as HTTP, and technology for high-speed accumulation and retrieval of collected packet data,[1),2)] and has implemented these technologies as software in systems. As a result, real-time analysis of 200 Gbps communication quality has become possible with just one general-purpose server and software. Further, accumulation and retrieval at 40 Gbps can be achieved by configuring a system combined with a general-purpose server that accumulates packet data.

Convenient and stable service can be offered to many users through application of this system to quality monitoring of carrier networks and data center networks.

2. System configuration

As shown in **Figure 1**, the system that implements this technology consists of a stream analysis unit that collects packet data for super-fast communication with speeds of up to 200 Gbps, and a scalable data accumulation unit for accumulating the collected packet data. The stream analysis unit consists of a single general-purpose server so that the arrival times of packets can be managed with a single clock. The scalable data accumulation unit consists of several general-purpose servers to allow storage capacity expansion.

Conventionally, high-speed packet collection depended on the performance of the CPU and other hardware components, and the maximum number of packets that could be received per unit time would often be limited. Moreover, improvement of processing performance was difficult owing to the fact that sufficient memory copying speed among the processes of packet collection, network quality analysis and application quality analysis, which all depend on memory-access performance, could not be achieved.

This time, network quality analysis and application quality analysis, along with high-speed packet collection at 200 Gbps, have been achieved using only software and without costly dedicated hardware. Network quality analysis allows the detection of the detailed traffic volume of each service and user on the network, as well as communication quality problems caused by the network, such as packet loss, network delays, and so on. Application quality analysis allows calculation of traffic volume of each application, as well as detection of problems that are caused by response delay and the like of applications that offer services. Integration of these different types of quality analysis makes it possible to detect the causes and locations of service delays.

3. Features of the developed technologies

This section describes the features of the developed technologies. **Figure 2** shows the acceleration technology for packet collection and memory access.

1) Acceleration technology for packet collection

Packet reception (collection) processing at the OS level imposes a high processing load and burdens the entire system. For the acceleration of reception processing in particular, efficient reduction of interrupt processing with high load is important. This time, technology that reduces the number of processing times by aggregating the interrupts generated for each packet arrival at the time of packet reception has been developed. Moreover, by implementing this technology in the form of a driver, reduction of the processing load of the OS was achieved. Packet reception processing performance was improved even further through load distribution allocating interrupt processing among multiple CPU cores.

2) Acceleration technology for memory access

Current memory bandwidth in general-purpose server is on the order of 200 Gbps to 400 Gbps, which falls short of the bandwidth requirement for the analysis of packets at 200 Gbps through repeated read and write operations. This is a major reason for the high processing load of communication quality analysis. Traditionally, packet analysis processing has been done by executing in sequence three different types of processing, namely packet collection, network quality analysis and application quality analysis, and copying the packets after each processing to pass it on to the next. Under this method, copying 200 Gbps packet analysis three times, for example, requires 600 Gbps of

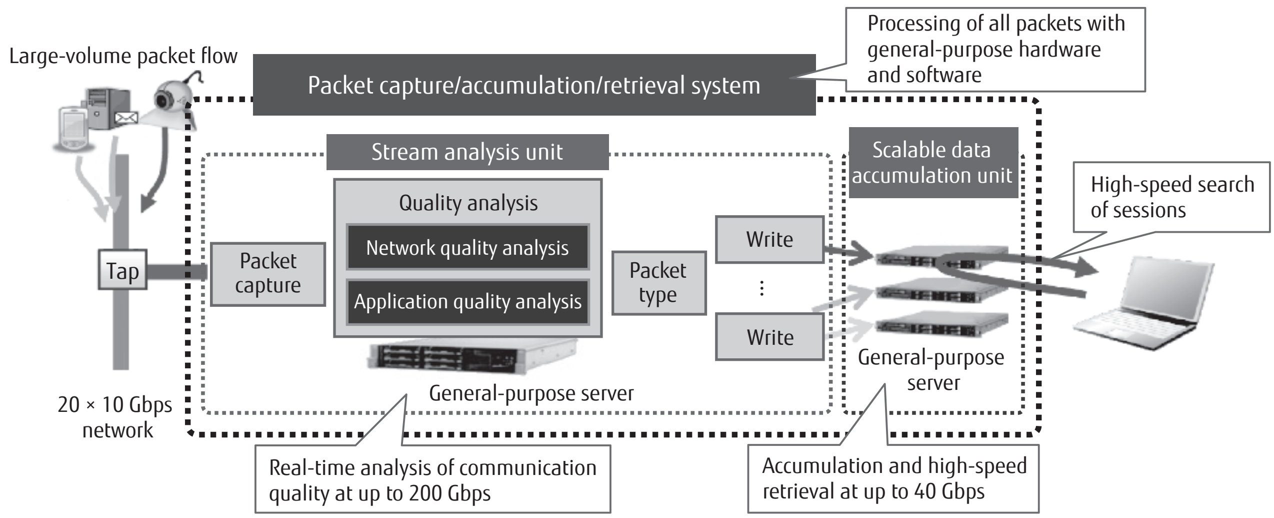

Figure 1
System configuration implementing the developed technology.

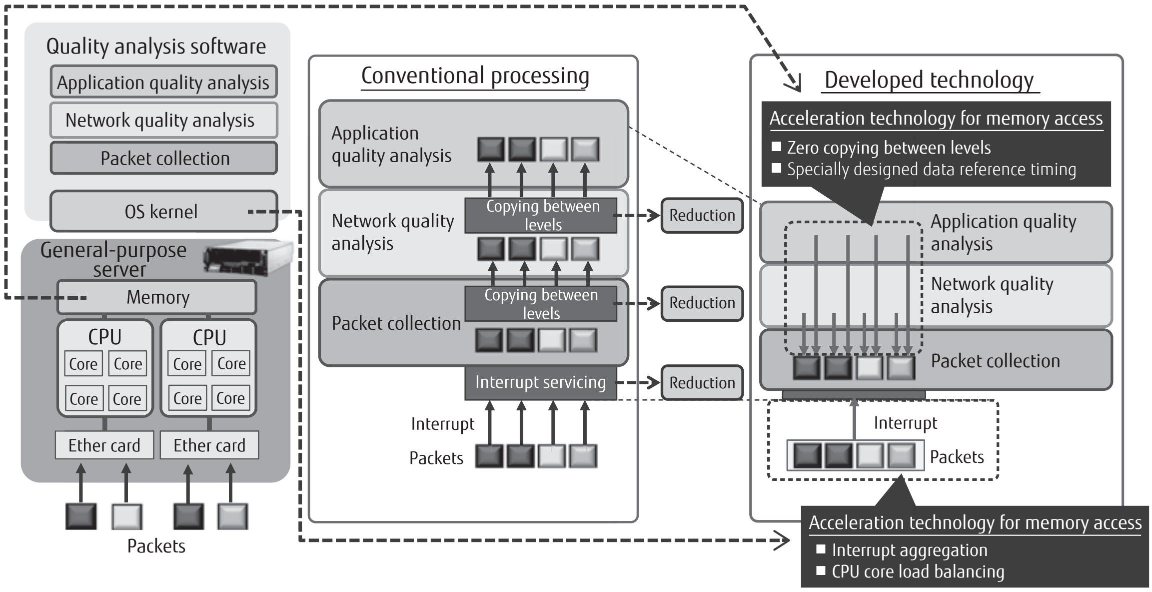

Figure 2
Acceleration technology for packet collection and memory access.

memory bandwidth. Thus limited memory bandwidth availability of 200 Gbps precludes such packet analysis. To overcome this problem, the method and timing of data referencing between packet collection processing and quality analysis processing were innovated, preventing simultaneous writing and writing to areas that are being referenced, and making referencing possible without copying of packets and analysis data. As a result, the need for exclusive control of copying and processing was eliminated.

3) Parallel processing technology

Figure 3 shows parallelization processing technology for the effective utilization of multiple CPU cores.

Today's CPU cores are very fast, but not as fast as the packet analysis processing of 200 Gbps can be execute by a single CPU core. Therefore, how to efficiently use multiple CPU cores to execute packet analysis processing in parallel is an important question. There is the issue that when executing a large number of packet analysis processes in parallel, the analysis processing performance does not rise proportionally with the number of CPU cores. This is due to the fact that, conventionally, exclusive processing such as the use of exclusive access control locks is required when sharing the same data structure among multiple analysis processes, and that concurrent analysis processing is not possible. To solve this problem, we developed technology that allows access to the same lock-free ring buffer[note)] from a large number of analysis processes running on multiple CPU cores, without need for exclusive control. Concretely, as shown in Figure 3, we devised a data structure such that exclusive control is not performed in the case that multiple analysis processes do not refer to the same memory area. This prevents conflicts among multiple CPU cores and can improve analysis processing performance proportionally with the increase in the number of CPU cores.

4. Scalable high-throughput data accumulation mechanism

Along with technology to collect communication packets of ultra-high-speed networks and analyze them in real time, technology that allows the accumulation and fast retrieval of packets at a rate of up to 40 Gbps has been developed. This technology, which uses only general-purpose servers and switches, realizes high scalability and reliability by operating in a coordinated manner multiple general-purpose servers.

---

note) One of the data storage methods that allows overwriting of data is the ring buffer. The lock-free ring buffer is a type of ring buffer that employs a buffer access method that eliminates the need for exclusive control.

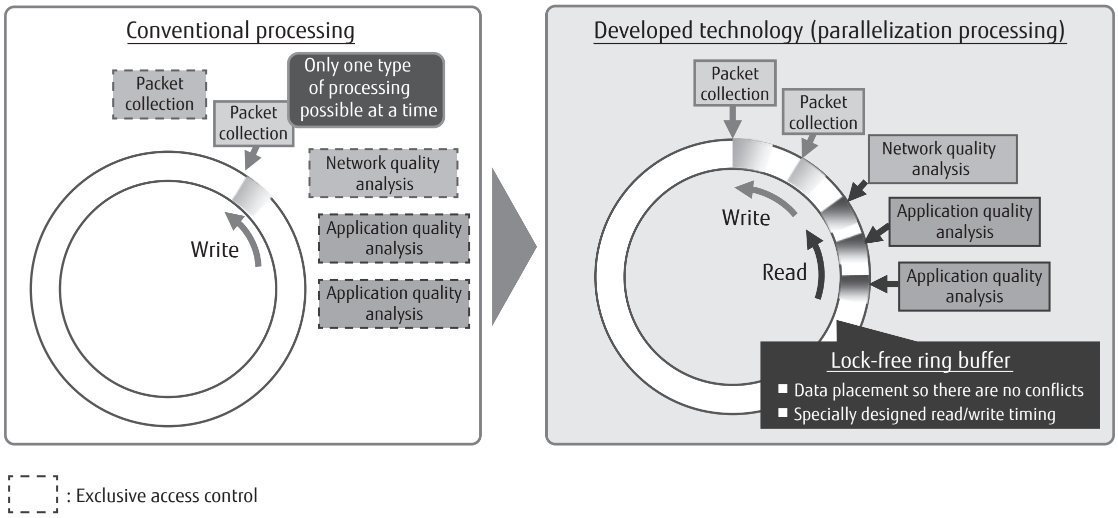

Figure 3
Parallelization processing for effective utilization of multiple CPU cores.

The scalable data accumulation unit developed this time is configured as distributed storage achieved by connecting and operating in a coordinated manner multiple servers in a network in order to secure disk capacity scalability and parallel writing performance. Therefore, the captured data is accumulated via that network.

- High-speed accumulation

Accumulation of all communication packets requires keeping up with the pace of the continuous incoming data flow without delays. The accumulation performance of hard disk drives (HDD) is highest when data is written sequentially. To combine this with high-speed retrieval, which is described later, data is written in batches on the order of a few Mbytes according to the packet analysis results, instead of being written in the order in which packets arrive.

The accumulation of packets is done by accumulating new data while deleting old data. If access occurs to update the HDD's sector management information while old data is being deleted, this disrupts the sequentiality of data writing and diminishes accumulation performance. Therefore, the HDD's sector management information is maintained in the management space on the memory. The processing of this management space is simplified by keeping the size of the write data a fixed length. The allocation of sectors on the HDD when writing data is done toward increasing addresses, regardless of the distribution of free space. When the end is reached, memory space allocation is done from the beginning again so as to maintain sequentiality.

- High-speed retrieval

High-speed retrieval of the desired data from the large number of accumulated communication packets is also important. To allow high-speed retrieval, data accumulation is done by processing the data according to search conditions in addition to the arrival time of each packet. The time and source address are among the search keys typically used in packet analysis. To optimize search by source address, packets are grouped along two axes, namely the appropriate range of sources addresses (we call this the "source group") and the range of packet arrival times, rather than writing all packets in chronological order. This makes it possible to narrow down the search scope to a given group when performing searches by source address and/or time, and to speed up searches by performing searches in parallel for each group.

Prior to grouping by source address, packet groups are extracted as sessions based on matching of the packet attribute information items of destination/source address, destination/source port number, and protocol number. The attribute information of the extracted sessions is held on a session basis. This greatly reduces storage capacity requirements compared with the holding of attribute information on a packet basis. Through searches based on this session attribute information, for example in the logical unit of communication sessions per terminal during calling, high-speed retrieval can be achieved.

5. Conclusion

This paper introduced technology for analyzing communication packets in real time and accumulating packets as evidence, allowing high-speed retrieval to enable high-speed analysis at a later time.

The technology developed this time allows the realization of real-time analysis of network quality and application quality while monitoring a large volume of packets coming in on an ultra-fast network at 200 Gbps, ten times the traditional level of performance, by using only general-purpose hardware and software. As a result, causes and specific locations of service quality deterioration can be determined at low cost. Further, as historical data of sources and the like is retained, the sources of cyber attacks and the affected servers can easily be determined. Moreover, knowledge of the traffic volume required by each application allows construction of appropriate infrastructure in line with actual needs.

All the above means that the proposed technology can be applied to good effect for improving network quality, more stable operation of data centers, enhanced security, early resolution of troubles, fast service delivery, and so on.

Going forward, we will continue to aim for further improvements in communication data accumulation and retrieval processing performance and carry out demonstration experiments.

References

1) M. Tamura et al.: Proposal of Distributed Storage for Accumulation and Usage of Packet Data in a high-speed network. (in Japanese) 6th Forum on Data Engineering and Information Management (DEIM 2014), D2–3, March 2014.
2) M. Tamura et al. : Distributed object storage toward storage and usage of packet data in a high-speed network. Asia Pacific Network Operations and Management Symposium (APNOMS).

Yuji Nomura
Fujitsu Laboratories Ltd.
Mr. Nomura is currently engaged in research and development of operations management for ICT systems.

Toshihiro Ozawa
Fujitsu Laboratories Ltd.
Mr. Ozawa is currently engaged in research and development of next-generation storage systems.

Masahisa Tamura
Fujitsu Laboratories Ltd.
Mr. Tamura is currently engaged in research and development of next-generation storage systems.