Full Capture System and Big Data Analysis: Connecting Businesses and Networks in IoT Era

● Hideyuki Chiba ● Shingo Kamuro ● Yuji Nomura

As the Internet of Things (IoT) and machine-to-machine (M2M) communications become prevalent, the number of devices that connect to networks is rapidly increasing. Data traffic via networks has thus grown to a significant volume. The connectivity to a network, and the quality of the network itself, are becoming prerequisites as opposed to added value that generates profits. Given this as a background, the service-oriented business models based on requests by upper layers, which have been around for some time, are expected to develop further and faster. To support such business models, Fujitsu has developed a real-time full capture system compatible with the 200 Gbps bandwidth, leveraging the Company's networking and big data analysis technologies. The system is designed to capture and analyze a large volume of high-speed data using virtual platforms. This paper briefly looks at user expectations for virtual platform services and challenges with regards to providing such services. Then it describes the characteristics of the developed system.

1. Introduction

In a society where everything is connected to a network by virtue of the Internet of Things (IoT), machine-to-machine (M2M) communications, telematics, and other technologies, a huge number of devices are adapted to using information technology.

As networks continue to progress, the unit prices of peripheral devices such as sensors inevitably drop. The operating cost for such networks is also being lowered. Against this background, telecom carriers are shifting their focus of revenues from conventional connection services to offering information obtained from connected devices.

For circulating and leveraging such information, the virtual platforms are the mainstream media, offering users scalability, easy provisioning, and flexibility in system configuration and modification. The virtual platform runs software that replaces functions hitherto realized by hardware, such as controlling and data transfer. The software programs themselves are transmitted via the network as information, and thus used as services, through IoT.

Leveraging the Company's networking and big data analysis technologies, Fujitsu has developed a real-time full capture system compatible with the 200 Gbps bandwidth. This system is designed to capture and analyze a large volume of high-speed data on a network by means of a virtual platform.

This paper describes our efforts to establish practical applications of the services this system can offer.

2. Requirements

In this section, we describe what is required of services that distribute large amounts of information on virtual platforms.

Basically, services are created on the upper layers such as a business layer, and attributed with certain specific intentions or significance. In other words, the services have their own "semantics (concepts/meanings)" that characterize them. The expected effects in business can be attained provided that the services are correctly operated as intended.

Take the launch of a new brand for example. The services created for this brand are in line with its product concept. There, the requirements include ensuring a means to realize such services and guarantees for their appropriate operations.

The network and applications are components of

services. However, service components in virtualization in recent years, namely, software-defined networking (SDN) and network functions virtualization (NFV), do not simply serve for reducing capital expenditure (CAPEX) or operational expenditure (OPEX). They also contribute to shortening the lead time before service launches and improvement, putting the top priority on increasing sales.

In order to sustain business-oriented services and increase profit, it is necessary to clarify the results yielded by these services. Also, further clarifications are needed to identify factors underlying such results, or reasons for failing to attain the expected results, and future measures for sustainable improvement. These aspects represent the concepts and meanings of services. This can be construed to be an equivalent of an existing practice of managing key quality indicators (KQI) and/or key performance indicators (KPI) by breaking down the service level agreement (SLA) or quality of service (QoS) in a top-down fashion.

Therefore, in order to shift towards a service-based profit system, it is necessary to identify the key factors to define the services by means of semantics analysis regarding them. Then, the components in the semantics structure thus identified must be maintained free from irregularities.

3. Challenges in virtualized services

As stated above, there have been operations revolving mainly around services based on virtual platforms. However, the significant increase in the number of connected devices and network bandwidths has a considerable impact on the operation models and systems.

Moreover, guaranteeing virtualized services on the network requires the following challenges to be overcome:

1) Completeness of data to be guaranteed

In cases where packet filtering and/or packet header slicing is employed in order to handle a large amount of data traffic, there are risks of missing out a portion of session data due to the filtering. The slicing discards information on payloads and upper protocol layers, and hence the completeness of the whole data is compromised.

2) Analysis of large amount of diverse data

Analyzing service concepts and meanings requires a large amount of diverse data. The same applies for analyzing the statuses and results of such services. Examples of such data include population statistics, traffic information, weather information, and other public information, as well as packet data communicated through the network. It is necessary to consider providing functions that are equipped with the capability to process these data integrally.

3) Improvement on the data processing speed

The increase in the data volume to be analyzed means a huge amount of time necessary for calculations as an appropriate portion of the data must be extracted for analysis. In the fields of telematics and medical services, the analysis results need to be made available immediately. Therefore, the real-time processing must overcome some issues; namely, that the real-time analysis depends on analysis speed, and that on-demand processing depends on the volume of accumulated data.

4. Solutions

Previously, dedicated hardware with special architecture and algorithms was used to handle issues regarding the analysis of large amounts of data, as stated above. However, considering the prospect of rapidly increasing and diversifying data to be analyzed, in both quantity and quality, certain limits are envisaged for the dedicated hardware in terms of scalability, functional flexibility, and extendability. Dedicated hardware also makes a cost increase inevitable. To offer a solution to this situation, Fujitsu has developed a real-time full capture system, designed to run on general-purpose IA servers and capture a large amount of high-speed data in full. The system is capable of analyzing such data in real time, and compatible with the 200 Gbps bandwidth (**Figure 1**).

The system makes it possible to observe real phenomena directly from packet information, as opposed to "anticipated" or "built-up" logs, services, or network information which are provided by virtual platforms. This is something existing networks hitherto have not been able to perform.

The system's architecture is unique in the sense that the services are realized by the software that runs on a general-purpose IA server, thus requiring no dedicated hardware. It makes the system introduction easier, with flexible variability according to the size of

operation required.

Because the systems or networks to be monitored must be detached from the platform, this system is fundamentally an external monitoring one. This means that there is no need for consuming the resources available on the target system, which enhances the credibility of the observation results. Furthermore, being an external system has an advantage of not affecting the monitored systems in the event of a system error.

5. Application example

The below is an example of applying this system.

5.1 Adaptability to virtual environment

Where there are cloud and other virtual environments involved in the network structure on the part of service providers or the network, information other than basic essentials is concealed. However, there is a possibility that information beyond the anticipated scope of errors or scaling will escape due to the fact that applications on the virtual platform, or other virtual infrastructure, automatically apply auto-healing or auto-scaling. Also, there is a chance of overlooking not only valuable information, but also information that may lead to errors. For these reasons, it is important to capture the information on the physical network in full before it undergoes concealment and virtualization.

Furthermore, the behavior of flows at upper layers is key indicators for network design of SDN, a component of services to virtualize the network. It is important to monitor the behaviors of layers 4 to 7 in order to guarantee normal flow behavior. If slicing is used, the upper layers will be cut out, making it impossible to ascertain flow statuses.

This can be understood as an equivalent of a policy control in networking, like the semantics in services.

5.2 Security features

Regarding the security features of this system, it employs a method to identify malicious attacks by the packet/traffic behaviors and content of payloads.

Presently, known threats are dealt with by various security appliances and firewalls. However, threats such as the zero-day attack still remain, and it is impossible to remove all threats completely. Given this, it is important to narrow down the intentions and scopes of attacks from their outcomes immediately, which requires real-time monitoring over any digressions, followed by an analysis of what was happening within the system. On this point, we consider that the system is capable of analyzing the affected domains, targets of

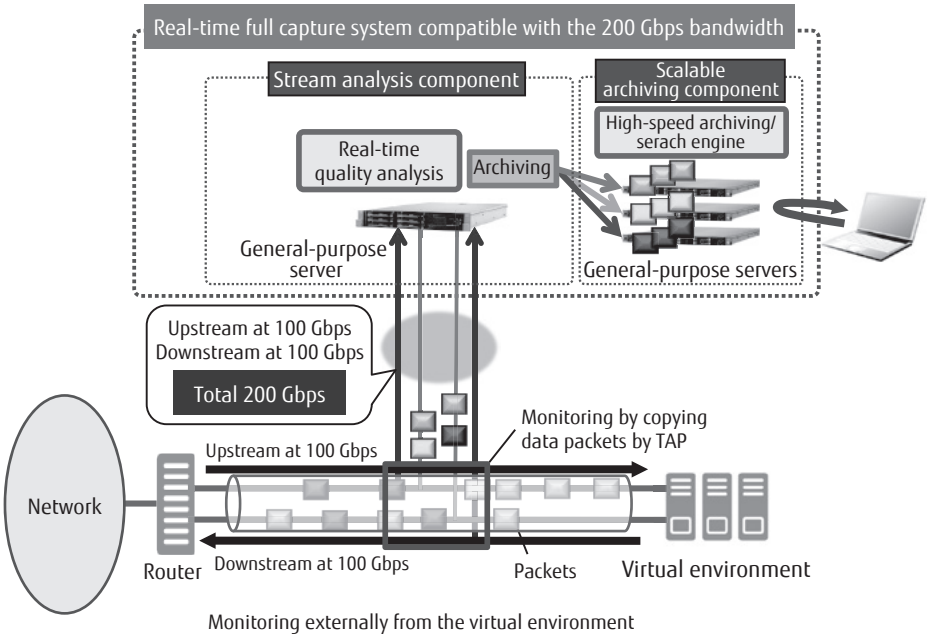

Figure 1
Real-time full capture system: configuration and coordination with virtual environment.

the attack, and sources of information, as the system monitors the entire packet information.

6. Conclusion

This paper described the ways in which services were provided in the virtual environment and the approaches in terms of the network layers to guarantee such services. It also explained the real-time full capture system that was compatible with the 200 Gbps bandwidth, developed by Fujitsu.

Given the increasing demand for upper-layer services such as IoT and M2M communications in society, providing a sound network is a prerequisite that must be guaranteed. Furthermore, such services will have importance in terms of ensuring the quality and quantity expected. They will achieve this by making continuous efforts to improve their quality and ensure sustained service provision. Therefore, increasing emphasis is also placed on the system guaranteeing their operations.

Hideyuki Chiba
Fujitsu Ltd.
Mr. Chiba is currently engaged in operation system development.

Shingo Kamuro
Fujitsu Ltd.
Mr. Kamuro is currently engaged in development of telecommunication solutions.

Yuji Nomura
Fujitsu Laboratories Ltd.
Mr. Nomura is currently engaged in R&D on ICT system operations management.