Server Virtualization Technology and Its Latest Trends

● Yoshihiko Oguchi ● Tetsu Yamamoto

(Manuscript received June 12, 2007)

IT systems have become increasingly larger and more complex, thus making it more difficult to build an optimal IT infrastructure in today's rapidly changing business environment. Server virtualization represents a base technology for addressing this problem. It enables the flexible construction of virtual servers with almost no hardware limitations, and consequently reduces the total cost of ownership (TCO) and makes it easier to use virtual servers in the changing business environment. Fujitsu supplies server virtualization technology in the form of the Virtual Machine Function for the "PRIMEQUEST" mission-critical IA server and "PRIMERGY" PC server. The Virtual Machine Function is based on Xen open-source technology. This paper describes the background behind and latest trends in server virtualization, and outlines Fujitsu's Virtual Machine Function and the Xen architecture. It then describes our development of various technologies for mission-critical systems. These technologies include large memory management, I/O processing optimization and enhancement, and RAS enhancement. This paper concludes by describing our future plans for server virtualization.

1. Introduction

IT systems that typify today's business infrastructure have become increasingly larger and more complex, and building an optimal IT infrastructure in today's rapidly changing business environment poses a challenging problem. Server virtualization represents a base technology for addressing this problem. This technology enables the flexible construction of virtual servers with almost no hardware limitations, and consequently reduces the total cost of ownership (TCO) and makes it easier to use virtual servers in the changing business environment.

Fujitsu supplies server virtualization technology in the form of a software product called the Virtual Machine Function[1),2)] for its "PRIMEQUEST" mission-critical IA server and "PRIMERGY" PC server. The Virtual Machine Function adopts Xen[3)] open-source technology. Fujitsu also contributes to Xen development as a leading member of the Xen community. In particular, Fujitsu has promoted the enhancement of functions and quality for mission-critical systems, based on its accumulated technologies for mainframe virtualization.

This paper describes the background behind and latest trends in server virtualization, and outlines Fujitsu's Virtual Machine Function and Xen. It then describes the extension of technology necessary for mission-critical systems. This paper concludes by describing our future plans for the proliferation of server virtualization technology.

2. Background and effects of server virtualization technology

Server virtualization technology is utilized to construct multiple virtual servers (i.e., virtual machines) on one physical server. An operating system can be operated and an application system can run on each of these virtual servers, just as on a physical server. Virtual machines can be built by dividing a physical server in terms of hardware and software. When the physical server is divided by hardware, server virtualization offers a special advantage: a hardware error or high load occurring in one divided section does not affect the other sections. In contrast, when the physical server is divided by software, server virtualization offers other advantages, such as allowing the CPU, memory, I/O devices, and other hardware resources to be assigned to virtual machines, and the resource assignment status changed even during operation. Moreover, when the physical server is divided by software (through server virtualization), the hardware resources can be shared by and assigned among multiple virtual machines, as well as being used exclusively by a specific virtual machine. The following mainly describes server virtualization by software.

Given the recent background of significant increases in application-system construction costs (in terms of actual expenses, time needed, locations, etc.), there are great expectations for server virtualization technology. In other words, there is an urgent need to quickly and flexibly satisfy a variety of changing business environment conditions, and systems should be built in service units and flexibly interlinked according to application conditions. Service Oriented Architecture (SOA) embodies such flexible design techniques. In this case, the construction of a total application system must usually consist of multiple systems. Building a total application system based solely on physical servers entails the expensive purchase of physical servers. Moreover, these physical servers must offer high levels of performance to accommodate peak loads during operation, but which may result in a low hardware usage rate and ineffective use of resources during other periods. At times requiring the high availability of system operation, a standby system must also be prepared to continue processing in case of an error occurring. Therefore, building and maintaining such a standby system that will be rarely used under ordinary circumstances also necessitate additional costs.

Introducing server virtualization technology can resolve such problems. When multiple virtual machines are built by software and exist on one physical server as shown in **Figure 1**, the installation cost and time needed can be dramatically reduced compared to a case where multiple physical servers are procured. Moreover, building multiple systems having low hardware resource usage rates (e.g., CPU usage rate) on the same physical server will result in more effective use of server resources. The ratio of resources distributed to low-load jobs and standby systems can also be kept low as in usual operation, though the distribution rate may abruptly increase in case of a higher load or error occurrence. Thus, server virtualization technology reduces the TCO and makes rapid system configuration changes possible under changing use conditions.

3. Outline of Virtual Machine Function

Fujitsu has supplied Physical PARtition (PPAR) and eXtended PARtition (XPAR) for PRIMEQUEST. PPAR and XPAR are technologies for virtualization by hardware. Fujitsu is now providing the Virtual Machine Function for the PRIMEQUEST and PRIMERGY servers. The Virtual Machine Function is the technology for virtualization by software, and was developed based on Xen technology — the emerging open-source server virtualization technology. Xen was developed at the University of Cambridge in the UK. The first version was released in 2003, and later evolved into open-source software. Various vendors subsequently participated in Xen development to enhance its functions and quality. Xen was also adopted in Red Hat and Novell Linux distributions, and has been applied to production runs. Similar to the development of Linux, open-source software is undergoing rapid enhancements of functions and quality, primarily because many developers are participating in the open-source community from various aspects.

Support and drivers were added to the Red Hat Enterprise Linux 5 virtualization software (including the Xen source) to create the Virtual

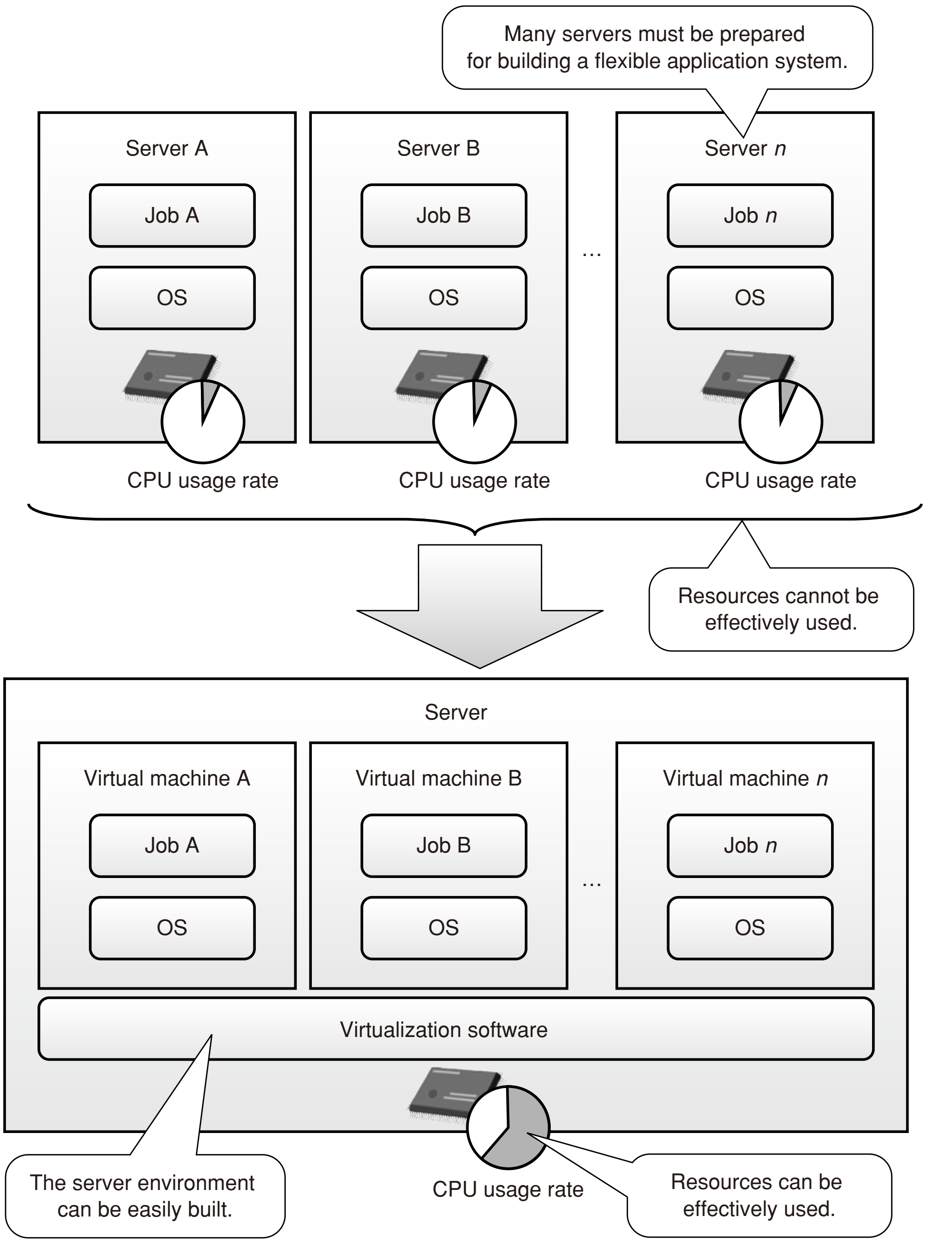

Figure 1
Background and effects of server virtualization.

Machine Function.

Figure 2 shows the architecture of the Virtual Machine Function. The virtualization software (called Hypervisor) positioned above the hardware provides server virtualization. In the Virtual Machine Function system, a virtual machine is called a "domain", with the special domain for managing the Virtual Machine Function called "Domain 0" and an ordinary domain for processing application called a "guest domain". In Domain 0, the Linux host operating system (host OS) operates and manages the guest domains by using the operation management tool. In the guest domain, the Linux or Windows operating system running on the physical server operates as the guest operating system (guest OS).

Server virtualization processing mainly consists of virtualizing the CPU, memory, and I/O devices. To virtualize the CPU, a virtual CPU is assigned to the domain, with correspondence of the real hardware CPU (physical CPU) switched to the virtual CPU according to the CPU scheduler. Physical CPU performance is distributed by time-sharing. To virtualize memory, a specific size is selected from real memory mounted on the hardware, and then the selected memory is assigned to each domain. The virtual device model virtualizes I/O devices such as disk drives as well as networks. Specifically, para-virtualized (PV) drivers are installed in the guest OS, and the I/O devices are accessed from the real device drivers via linkage drivers (back-end drivers) installed on the host OS side. In this virtual-

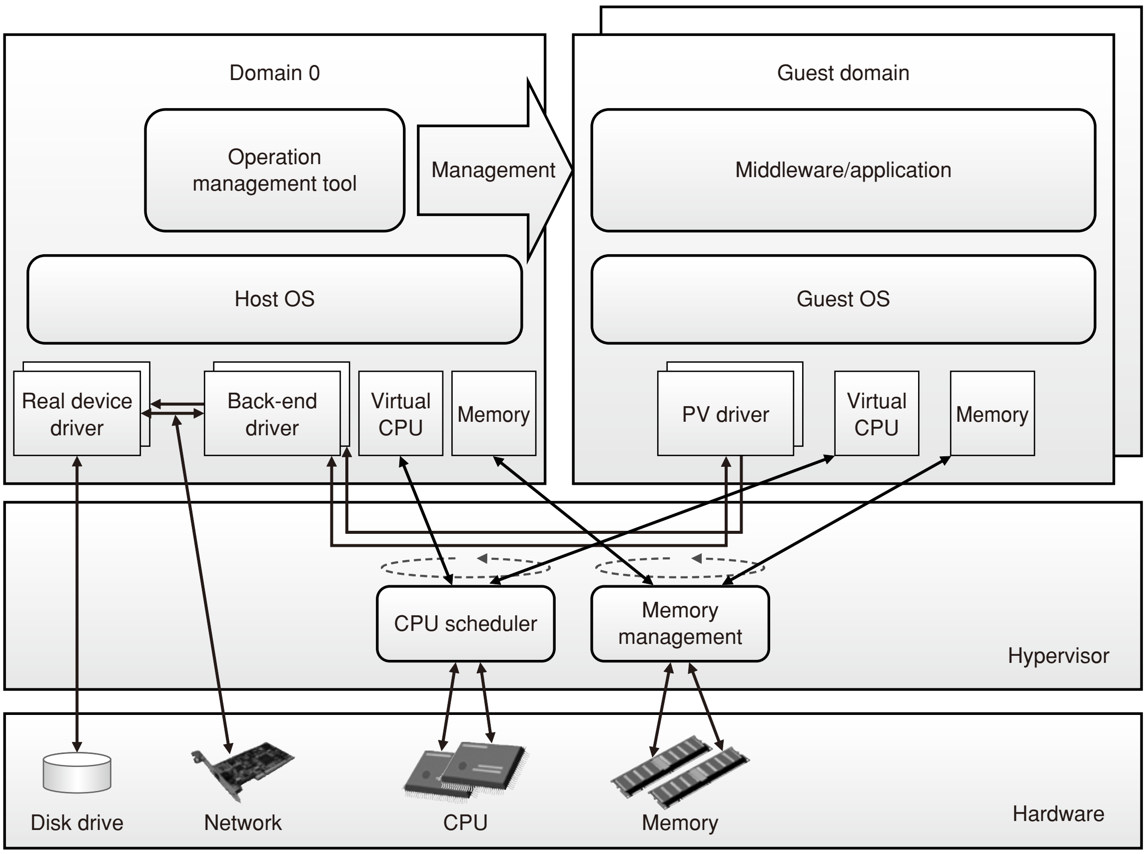

Figure 2
Architecture of Virtual Machine Function.

ization system, the real device drivers provided for Linux can be used without modification. As a result, the guest OS can share the various I/O devices supported by Linux. Since the PV drivers are already optimized for I/O processing on a virtual machine, sufficient performance is ensured in actual production runs.

4. Application to mission-critical servers

The development of Xen has rapidly advanced, but applying Xen to mission-critical servers is hindered by the problems of insufficient functions for large-scale hardware resource management, performance and reliability assurance, and maintenance. To address these problems, Fujitsu has continued development to improve both functions and quality for effectively applying Xen to mission-critical servers. These development results are now being adopted in the Xen community.

Products based on Xen have been developed for various platforms. The major products in this development are intended for the widely used x86 architecture. Conversely, the PRIMEQUEST server for mission-critical systems adopts the more reliable Itanium Processor Family (IPF) architecture. In this way, Fujitsu has enhanced functions mainly for supporting IPF in the Xen community. With the technical knowledge acquired from this development, Fujitsu is the first to support development results in the Xen community in the form of products based on Xen technology. The following outlines these development results.

4.1 Supporting a large-capacity memory management system

A server equipped with large-capacity memory may have discontinuous memory addresses in a given area (called a memory hole). Because conventional Xen development was conducted mainly for medium- or small-scale systems, memory problems such a memory hole specific to large-capacity memory were not be considered. Therefore, Fujitsu remodeled the memory management system for IPF Xen so that large-capacity memory could be utilized by handling memory holes correctly. As a result, the host OS, Hypervisor, and guest domains in this new system can utilize up to 2 TB of memory.

4.2 Improving I/O processing speed and extending I/O functions

PV drivers are always necessary for high-speed I/O processing. The IPF version of the Linux PV driver was developed based on the PV driver of the x86 version released in the Xen community. A Windows version of the PV driver was then newly developed.

Fujitsu is now developing virtual SCSI drivers so that a guest OS can control various functions of such SCSI devices as tape drives. Thus it becomes possible for the guest OS to directly utilize the SCSI device functions via these SCSI drivers in the same way as on a physical server.

4.3 Enhancing the reliability, availability, and serviceability (RAS) function

The RAS function required for applying products based on Xen technology to mission-critical servers has been enhanced as described below.

4.3.1 Developing error analysis tools

If a software error occurs, the cause must be analyzed and corrected as soon as possible. However, most errors in a mission-critical system occur at a special timing; therefore, conducting a reproduction test of these errors is difficult. For prompt error analysis, data such as the contents of memory and CPU registers must be dumped when an error occurs. Conventional Xen had no function, however, to dump and analyze the data of the total system. Therefore, Fujitsu developed the dumped-data analysis function for Xen (**Figure 3**). Since this function was developed by extending the standard dump function of Linux, the man-hours needed to develop the tool were kept low, while utilizing the technical knowledge about Linux error analysis. The data dump and dumped-data analysis functions are described below.

1) Data dump function

The data dump function for also collecting a dump at a crash of the host OS or Hypervisor was developed based on the kexec/kdump function — a standard data dump function of Linux. PRIMEQUEST employs the "sadump" function to collect dumped data more securely. The "sadump" function was also extended to collect dumped data in case the host OS or Hypervisor crashed.

2) Dumped-data analysis function

The dumped-data analysis function for analyzing the cause of error was developed based on the "crash" that was a standard dumped-data analysis tool of Linux. Conventional "crash" had no function for analyzing the Xen control structure (e.g., domain and virtual CPU information). To solve this problem, Fujitsu added a function to separate and analyze the domain information, virtual CPU information, and trace data from the dumped data. As a result, the Xen control structure and OS information collected at error occurrence were analyzed for more efficient error analysis.

4.3.2 Enhancing machine check architecture (MCA)

MCA is a RAS function of IPF. If a hardware error is detected, MCA attempts to correct the error, and reports the error informa-

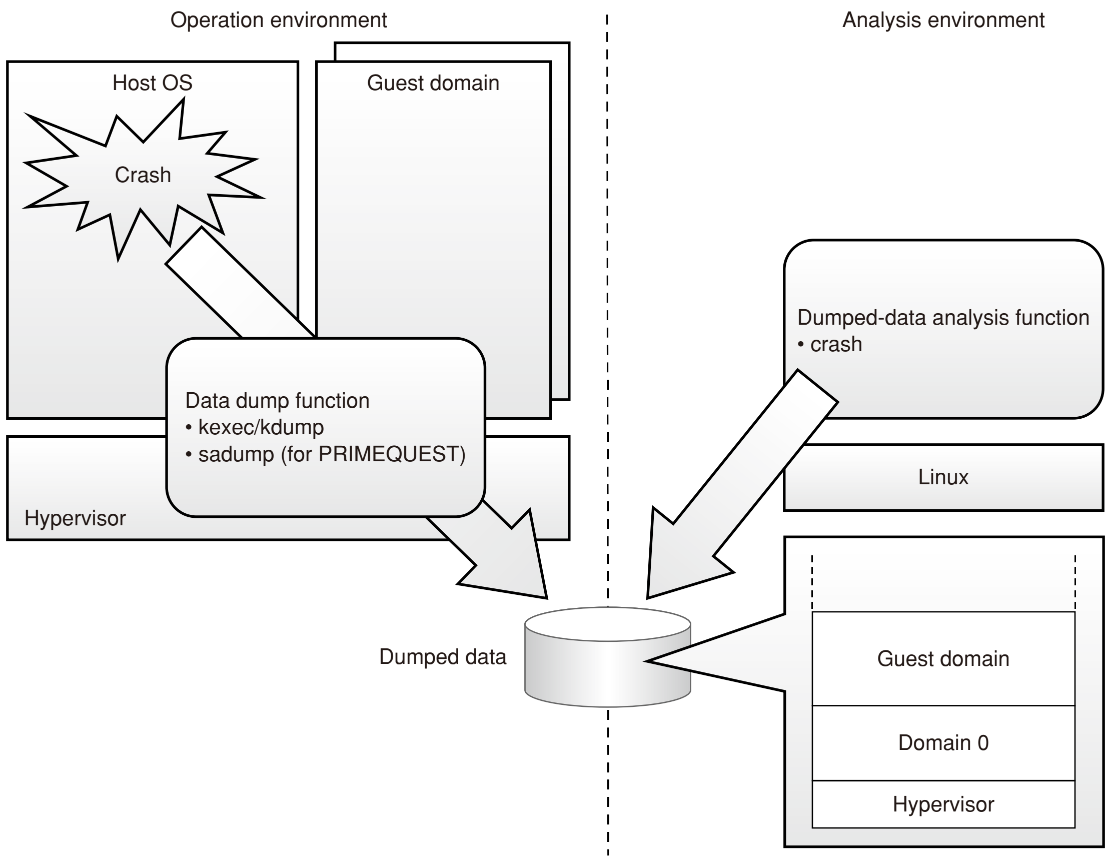

Figure 3
Error analysis tools.

tion to software for correct error handling. Xen of the IPF version has an additional function whereby the host OS saves the error information reported from MCA. As a result, the detection and automatic correction of a hardware error permits replacement of the related part(s) before an even more serious error could occur.

In the future, we will consider whether to develop a function for localizing errors detected in the operating hardware (that is, a function providing for degenerated operation in only the error-influenced guest domains and processes).

5. Future development

The use of server virtualization technology enables the easy cloning and restoration of a server environment, as well as migration between physical servers. Therefore, the server environment can be promptly saved in the standby system and restored at error occurrence, in order to continue processing at low cost. We believe that this technology will find wider fields of application by improving the efficiency of server operations as described above.

How to efficiently manage many virtual machines built at low cost will also pose an important challenge. To improve the operation management function of Xen, the open-source community is continuing the development of a GUI-based management tool known as "Virtual Machine Manager".[4] Fujitsu is also participating in this development as a leading member and contributing to more simplified operation management functions of virtual machines. In the future, Fujitsu will continue enhancing the operation management functions of the virtualized IT infrastructure through linkage with operation management middleware.

6. Conclusion

Server virtualization technology is always necessary for optimizing the IT infrastructure. This technology will become increasingly important in the future and be introduced for widespread use as a common technology for building the IT infrastructure. Also in the future, we will promote enhancement of the Virtual Machine Function by utilizing the advantages of PRIMEQUEST and PRIMERGY and continue virtualization product development for more efficient operation management of the IT infrastructure.

References

1) Fujitsu: PRIMEQUEST Virtual Machine Function. (in Japanese), April 2007. *http://primeserver.fujitsu.com/primequest/catalog/pdf/primequest-vm-wp200704.pdf*
2) Fujitsu: Fujitsu's Activities for Virtualization on IT Infrastructure TRIOLE. (in Japanese). *http://triole.fujitsu.com/jp/article/virtual/*
3) Computer Laboratory: Xen™ virtual machine monitor. *http://www.cl.cam.ac.uk/research/srg/netos/xen*
4) Virtual Machine Manager. *http://virt-manager.et.redhat.com/*

Yoshihiko Oguchi, *Fujitsu Ltd.*
Mr. Oguchi received the B.S. degree in Physics from the University of Tokyo, Tokyo, Japan in 1988. He joined Fujitsu Ltd. in 1988, where he has been engaged in research and development of virtual machine managers and operating systems for mission-critical servers.

Tetsu Yamamoto, *Fujitsu Ltd.*
Mr. Yamamoto received the B.E. degree in Information and Communication Engineering and the M.E. degree in Information Engineering from the University of Tokyo, Tokyo, Japan in 1998 and 2000, respectively. He joined Fujitsu Laboratories Ltd. in 2000, where he was engaged in research of Web service middleware. In 2006, he joined Fujitsu Ltd., where he has been engaged in research and development of virtual machines for mission-critical servers.