Technologies of Enterprise Storage ETERNUS8000

● Jun Ishikawa

(Manuscript received May 31, 2007)

Recently, there has been a huge increase in the disk capacity of storage devices now being connected to multiple servers. Even when the servers undergo a scheduled stop, these storage devices must operate continuously for 24 hours a day year-round. This necessitates the development of highly reliable, high-performance devices that can be used without stopping even when a faulty part is replaced, a new firmware function added, or the system configuration changed. For this purpose, Fujitsu has developed technologies for nonstop operation by using an advanced redundant-hardware configuration, based on mainframe technologies acquired in the past. This paper introduces the technologies of the ETERNUS8000 — the latest enterprise storage device developed by Fujitsu. It first describes the hot-swapping technology that supports redundant copying, cyclic mirroring, and hot firmware swapping for nonstop operation. Then it describes such data guarantee functions as disk encryption and RAID migration, and the QuickOPC and SnapOPC backup functions. The paper concludes by explaining the remote maintenance functions that improve the maintainability of the ETERNUS8000.

1. Introduction

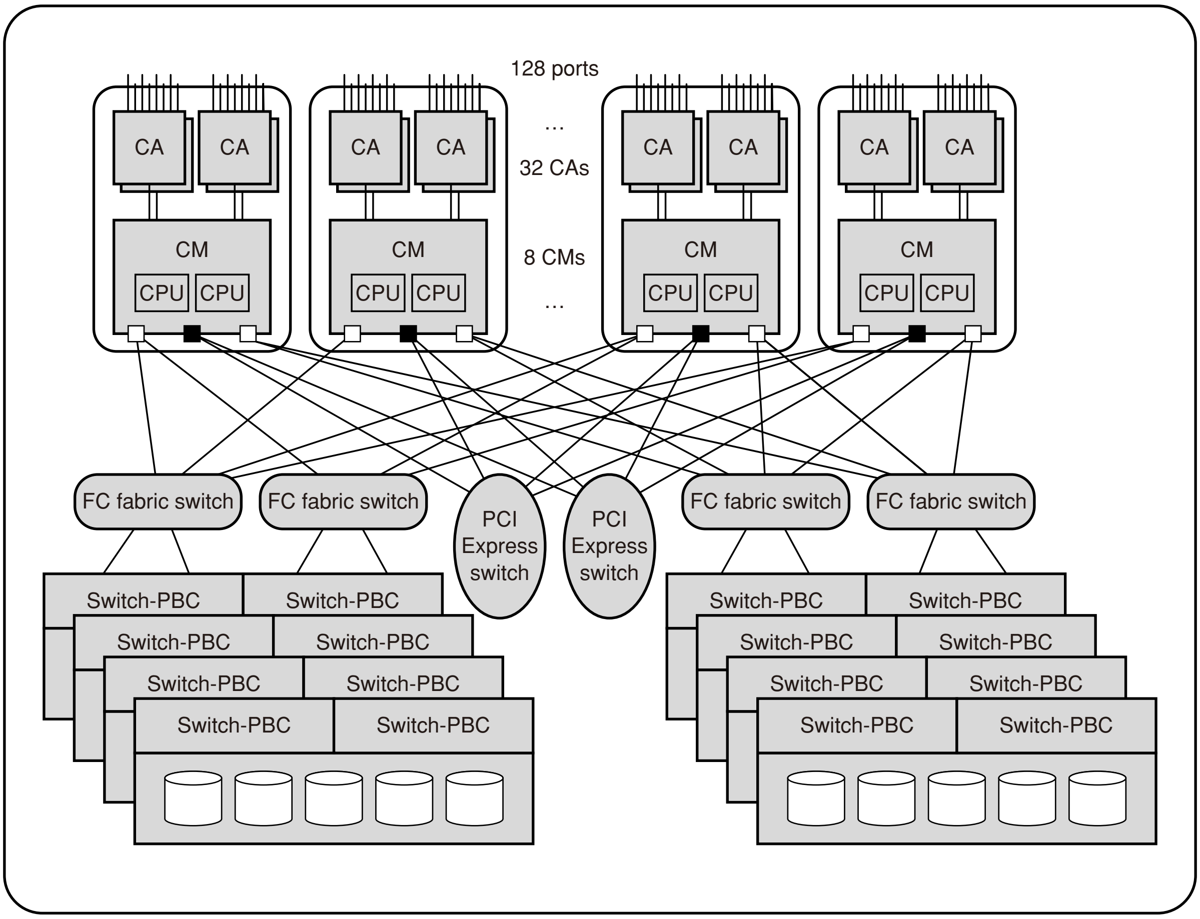
The IT systems used for business processing have become remarkably complicated and diversified based on recent changes in the business environment. The size of data being stored in system storage has also dramatically increased. Based on the advances made in open technologies, storage devices are now connected to multiple servers in various environments. Although servers can be intentionally stopped in many systems, the storage devices cannot be stopped.

The ETERNUS8000[1)] is Fujitsu's latest enterprise storage device, offering among the world's highest disk capacity. The maximum capacity of this device is 1.36 petabytes (PB), and stopping it during operation seriously affects business processing. Therefore, this storage device must offer very reliable high performance for reading and writing large amounts of data, and afford continuous operation without being stopped even when a faulty part or firmware function is replaced, or the system configuration changed.

This paper introduces the various advanced technologies of the ETERNUS8000; specifically, the hardware configuration for ensuring nonstop processing as much as possible under various conditions, the hot-swapping technology based on hardware technology, and the superior performance of data guarantee.

2. Hardware configuration of ETERNUS8000

All major components of the ETERNUS8000 such as the power supply and fan units have a redundant configuration. The high data transfer rate is made possible by a module structure that



CA : Channel adapter
CM: Controller module
FC : Fibre Channel
PBC: Port bypass circuit

Figure 1
Hardware configuration of ETERNUS8000.

contains controller modules (CMs or simply called controllers), channel adapters (CAs), and a device interface controller. Up to eight controllers can be connected via high-performance PCI Express switches provided as the standard interface based on the latest technology. These controllers are connected to disk drives via Fibre Channel (FC) fabric switches so that each disk drive can be accessed from any controller (**Figure 1**).

Each printed-circuit board (PCB) mounted in the ETERNUS8000 adopts diagonal interconnect wiring (**Figure 2**) to suppress signal attenuation and disturbance on transmission paths. This technique enables the connection of more controllers (than in a conventional system), and provides transmission paths for transferring data at a high speed of 4 Gb/s. This makes the development of high-performance, large-scale storage devices possible. The high-end model 2100 of the ETERNUS8000 Series featuring this hardware configuration is far superior to the ETERNUS6000 Series, one of the world's top-level products. Specifically, the ETERNUS8000 offers about double the high I/O processing performance and about triple the throughput of the ETERNUS6000.

We also thoroughly researched the data guarantee function. The data received from a

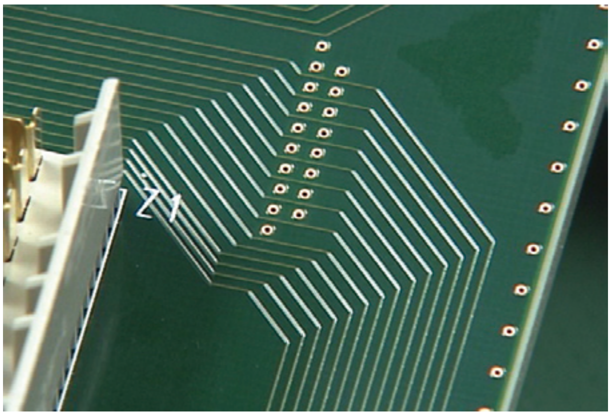


Figure 2
Diagonal interconnect wiring on PCB of ETERNUS8000.

server acquires a check code from the FC chip in a CA. During data transfer, the check code is used to confirm data integrity at several points. Check codes are also used in conventional systems to ensure data integrity. In the new system, however, check codes are added at a point nearer to the device input port for enhancing the function that detects data errors due to hardware errors, thereby improving system reliability.

3. Redundancy assurance and hot firmware replacement

This section describes the disconnection of parts to maintain disk redundancy, the automatic recovery of redundancy, and replacement of firmware during system operation.

3.1 Redundant copy function

If a disk error is detected, the faulty disk is disconnected and a standby disk (called a hot spare disk) is connected to recover disk redundancy. Should a second disk become faulty or disk data not be read correctly due to a medium error during the recovery of redundancy, however, the data is lost. The ETERNUS8000 resolves this problem in case a possible disk or medium error is detected by saving the disk data on a hot spare disk without disconnecting the anticipated faulty disk, and then disconnects that disk once the data is saved. Therefore, even if a disk or medium error occurs while data is being saved, the data can be recovered from another disk and thus minimize the possibility of losing any data.

3.2 Cyclic mirroring function

The major components of the ETERNUS8000 are made fully redundant based on reliable technologies so that operation can continue without stopping access from servers at any part of the system where an error occurs. For example, if a controller goes down, processing continues by using another controller, and the cyclic mirroring function disconnects the faulty part and automatically recovers data redundancy by using the remaining normal controllers (**Figure 3**). If an error is detected in one controller, the other controllers having a mirror relationship with the faulty controller recognize that data in the cache cannot be duplicated. Therefore, these controllers immediately save the cache data to a disk. The remaining controllers then recover the mirror relationship again. This technology is used to always duplicate the most important data of the storage device in the cache as well. After the faulty part is disconnected, a maintenance engineer replaces the faulty part with a normal part, diagnoses operation of the new part, and then connects it to the system. In this way, mirroring between all the controllers is recovered. Mirroring using three or more controllers is extremely more difficult than mirroring using two controllers. Fujitsu's storage products such as the GR800 Series and later adopt this mirroring technology. Moreover, this mirroring technology has been further improved for the ETERNUS8000 Series, which features the 2CPU configuration.

3.3 Preventive replacement

Since the main parts are fully duplicated or multiplexed, the device operates continuously even if an error occurs. Faulty parts, as well as batteries and other parts requiring periodic

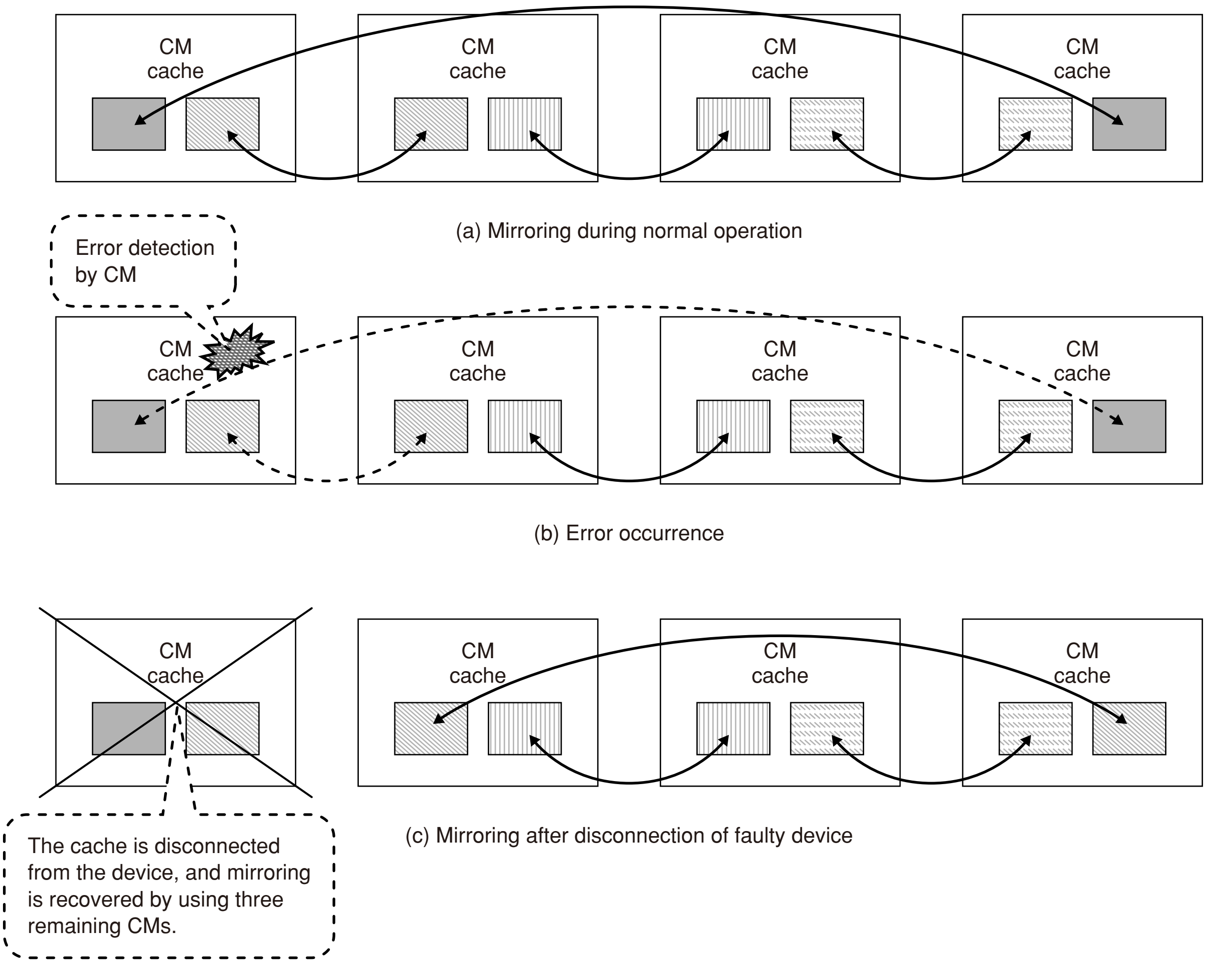


Figure 3
Cyclic mirroring by controllers of ETERNUS8000.

replacement, can be easily and safely replaced even during operation.

3.4 Hot firmware replacement

The hot firmware replacement technology provided for Fujitsu's F6495 Series storage devices and later has been further advanced with safer functions by employing an enhanced redundant configuration. When a maintenance program issues a firmware replacement request to a controller, the controller checks the device status. When all components of the device are normal, the controller divides the redundant components into two groups so as to retain the access path. While the controller checks the normality of the device, firmware is automatically replaced between the groups in turn. This replacement of firmware entails disconnecting the first-half components from the device, loading the new firmware, followed by the new program reconnecting the loaded components to the device [**Figure 4 (a)**]. Recovery of the redundant configuration and the normality of components are then confirmed, followed by the last-half components being disconnected and the new firmware loaded in those components [**Figure 4 (b)**]. Once the last-half components have been confirmed as being connected normally, firmware replacement is completed.

The use of this hot firmware replacement technology also enables a nonstop system to add new firmware functions and correct firmware

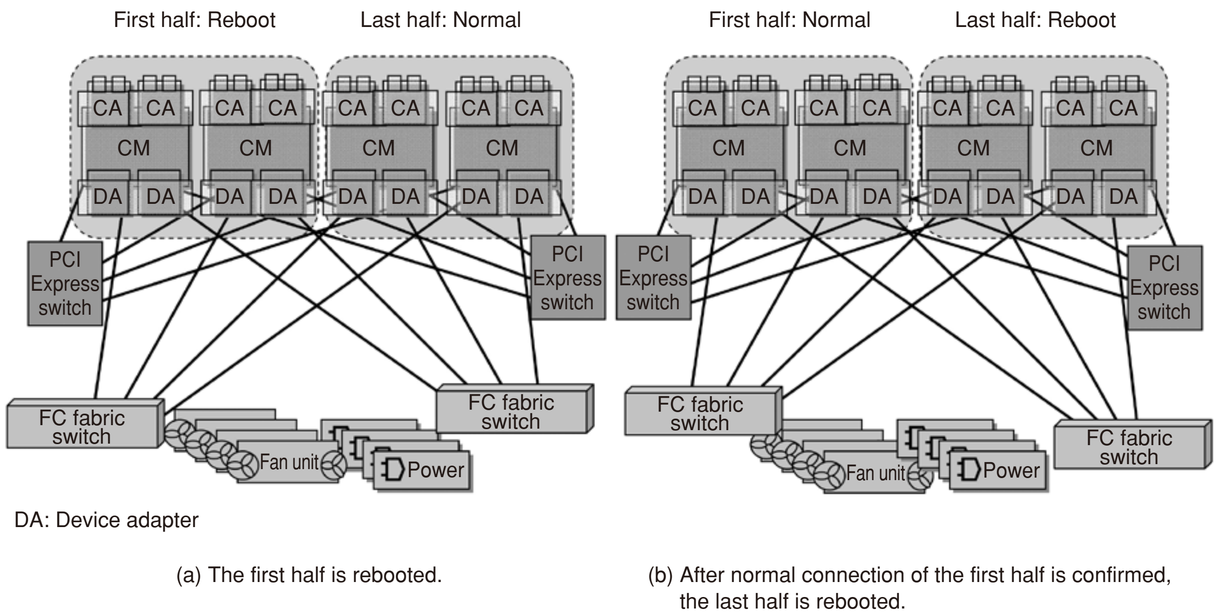


Figure 4
Hot replacement of firmware for ETERNUS8000.

errors during operation.

4. Data integrity, security enhancement, and safe operation

The ETERNUS8000 can contain a huge amount of data, including data for various purposes. This section describes the main technologies adopted for data security.

4.1 Disk encryption function

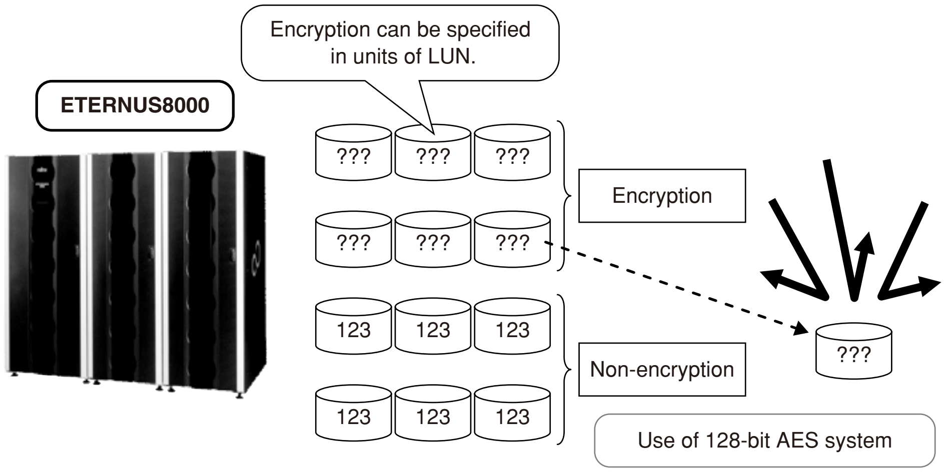
Certain volumes may contain customer data and confidential data. The system administrator must very carefully manage such secret data. In a conventional system, the system administrator must always confirm the complete deletion of secret data from the disk of a faulty disk drive that has been disconnected for removal from the data center. To eliminate this task, the ETERNUS8000 has a new function (**Figure 5**) by which the use of disk encryption can be specified in units of logical volumes (LUN). Therefore, the system administrator of the ETERNUS8000 need not consider the careless leakage of data.

4.2 Massive arrays of inactive disks (MAID)

Disk drives consume much electric power because disks continue rotating even when a disk data read or write operation is not being performed. For example, a disk provided to make a backup copy should only rotate while making a daily backup copy; however, it is continually rotating and consumes unnecessary power. The greater the number of disks, the higher the power consumption. To address this problem, the ETERNUS8000 adopts MAID technology for reducing power consumption of the entire device by scheduling disk rotation and stopping time. MAID reduces power consumption by 20 to 30% depending on the operating conditions, represents a unique function developed by considering the environment, and offers a major advantage in terms of operating cost (**Figure 6**).

4.3 RAID migration function

When disk space must be expanded to accommodate an increased amount of data or the response time delay shortened due to higher



AES: Advanced Encryption Standard

Figure 5
Disk encryption function of ETERNUS8000.

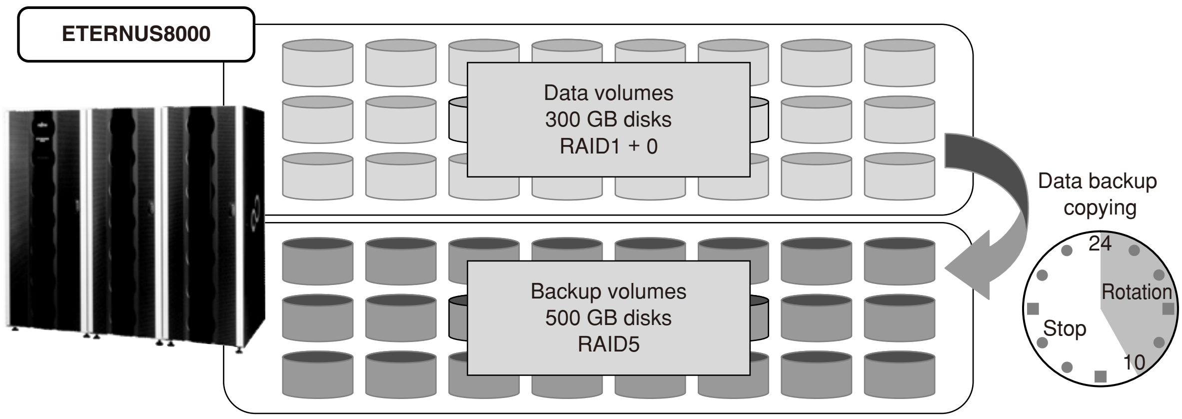


Figure 6
MAID function of ETERNUS8000.

access frequency, data must be moved to another disk of larger capacity or a more rapid RAID disk. In a conventional system, a backup copy of the data is made. This entails stopping the system, changing the configuration, restoring the backup data, and then tuning the system. The RAID migration function was developed to make such configuration change during system operation. This function also enables system tuning during system operation 24 hours a day throughout the year.

5. Backup functions and disaster recovery function

The ETERNUS8000 features an enhanced disaster recovery function and backup functions linked with ETERNUS SF Advanced Copy Manager — Fujitsu's storage management software.[2)]

The ETERNUS8000 supports various backup functions. When backup operation is instructed at the time required, the One Point Copy (OPC) backup function instantaneously reports the completion of backup to the upper-level device while the ETERNUS8000 continues the actual backup operation. When updating a volume whose backup copy is stored on a backup volume, the data is first saved to the backup destination volume, and then the data is actually updated. In this way, the backup data is assured. The ETERNUS8000 also supports QuickOPC and SnapOPC for various backup purposes. QuickOPC only copies the updated part; SnapOPC only copies the old data at updating.

In addition to the conventional function for remote-site copying using an FC, the ETERNUS8000 also supports a disaster recovery function using the Internet SCSI (iSCSI). Low-cost public lines can be utilized by using an iSCSI card having the encryption function.

6. Functions to improve maintainability

The ETERNUS devices are manufactured and equipped only with parts that have passed severe operation environment testing. Conversely, there is no way to reduce the number of faulty parts to zero, even when conducting the most demanding parts tests and mounting the most durable parts. Moreover, even when employing the most redundant configuration possible, parts failure will disrupt the redundant configuration. Leaving a disrupted redundant configuration unrepaired for a long time will result in a dangerous device condition.

Just like conventional devices, the ETERNUS8000 has a function to report errors to a remote site. Should a part error occur, the faulty part is disconnected from the device, and the remote maintenance function reports a failure event possibly related to an anticipated error detected by the diagnostic function to a remote maintenance center. In considering security, Fujitsu's remote maintenance policy inhibits external access to customer systems. Therefore, the device status must be steadily confirmed, faulty parts identified, and correct handling performed only according to reports sent from customers. A function that automatically sends internal device log data and device data for each failure event is supported for the ETERNUS6000 and ETERNUS3000 Series to improve the precision of faulty part identification. In a conventional system, detailed information about a faulty device cannot be confirmed until a maintenance engineer is dispatched to the customer's site and begins checking the faulty device. In the new system, however, the device status can be determined as the maintenance engineer travels to the site. Consequently, the device status can be confirmed more correctly in a much shorter time. The new system is also equipped with a function that periodically sends log data to the maintenance engineer. As a result, the maintenance engineer can easily check whether a device containing a replaced part is operating normally at a later time, and thereby anticipate another part failure. This reduces the system administrator's burden relative to administering a customer's system.

7. Conclusion

This paper introduced some of the functions and technologies provided by the ETERNUS8000. Many storage devices of open systems have adopted Fujitsu's technical knowledge based on highly reliable results acquired through mainframe development. The technologies for high-end storage devices are also applied to mid-range storage devices for providing devices of higher performance and reliability.

We fully expect storage capacity to continue expanding in the future and become increasingly more important. In that regard, Fujitsu's storage devices will continue providing new and improved functions to satisfy future system needs.

References

1) Fujitsu: ETERNUS8000 Storage Systems. *http://www.fujitsu.com/global/services/computing/storage/system/eternus8000/*
2) Fujitsu: ETERNUS SF AdvancedCopy Manager. *http://www.fujitsu.com/global/services/computing/storage/software/advance/*

Jun Ishikawa, *Fujitsu Ltd.*
Mr. Ishikawa received the B.S. degree in Physics from Chiba University, Japan in 1988. He joined Fujitsu Ltd. in 1988, where he has been engaged in the development of storage systems.