Technologies of ETERNUS VT600 Virtual Tape System

● Takashi Miyako ● Tatsuya Ohmido

(Manuscript received October 20, 2005)

Faster tape processing and a lower total cost of ownership (TCO) are required for IT applications that use a magnetic tape system. The ETERNUS VT600 virtual tape system was developed to meet these needs. It achieves major improvements in speed, reliability, and functionality by combining an innovative tape virtualization technology with the latest Storage Area Network (SAN) technologies and hardware components. This paper introduces the architecture, functions, and advantages of the ETERNUS VT600.

1. Introduction

Fujitsu has developed and provided high-end tape library subsystems since 1985. However, customers now require faster data processing, more efficient use of tape cartridges, and a smaller installation space.

The ETERNUS VT600[1] (hereafter also "the VT600") is a control device that provides virtualized high-end tape subsystem resources for server applications by acting as a virtual tape library (VTL). It is Fujitsu's latest high-performance virtual tape system and incorporates the CentricStor VTL control technology developed by Fujitsu Siemens Computer (FSC).[2]

The ETERNUS LT160, LT270, and LT130 tape libraries, which have high-end Linear Tape-Open (LTO) drives, can be connected to the VT600 as back-end physical tape libraries. Fujitsu's highly reliable and efficient ETERNUS3000[3] RAID is used for the tape volume cache (TVC). The VT600 can also emulate Fujitsu's F6473K tape drive (36-track tape drive) for host servers.

In this paper, we describe the basic architecture, functions, and advantages of the VT600.

2. ETERNUS VT600 virtual tape system

2.1 Development background

Tape libraries for automating tape operations have been developed for more than 15 years. However, although there have been dramatic reductions in the physical size of servers, disk devices, and other equipment, the large amount of space that is still required to install a tape library has become a major problem. On the other hand, tape drive technology has evolved, and the capacity of cartridge tapes has been increased from 200 MB (18 tracks) to 10 GB (128 tracks). In addition, the large capacities enabled by LTO technology have become available for open systems — one notable example being the recently released third-generation LTO, which has a capacity of 400 GB. Therefore, technologies that effectively and automatically enable the use of large-capacity tape cartridges have become more important. Moreover, the cost per megabyte of RAID hard disk drives (HDDs) has been decreasing, and large-capacity HDDs that can store hundreds of gigabytes of data have become available. The VT600 combines tape virtualization technology, massive-capacity tape libraries used as back-end tape storages, and open system RAIDs used as TVCs. Compared to conventional tape library systems. They are physically much smaller and provide more efficient data storage and faster tape application processing.

2.2 Basic concept of VT600

The VT600 is based on a software concept that can combine a fast online disk cache with up to 256 virtual drives in conjunction with modular expandable storage hardware (**Figure 1**). The system combines industrial standard products with Fibre Channel (FC) switch and LTO tape technology.

By using the following functions of the VT600, tape data processing can be done on tape devices that are virtually generated and configured in integrated channel processors (ICPs).

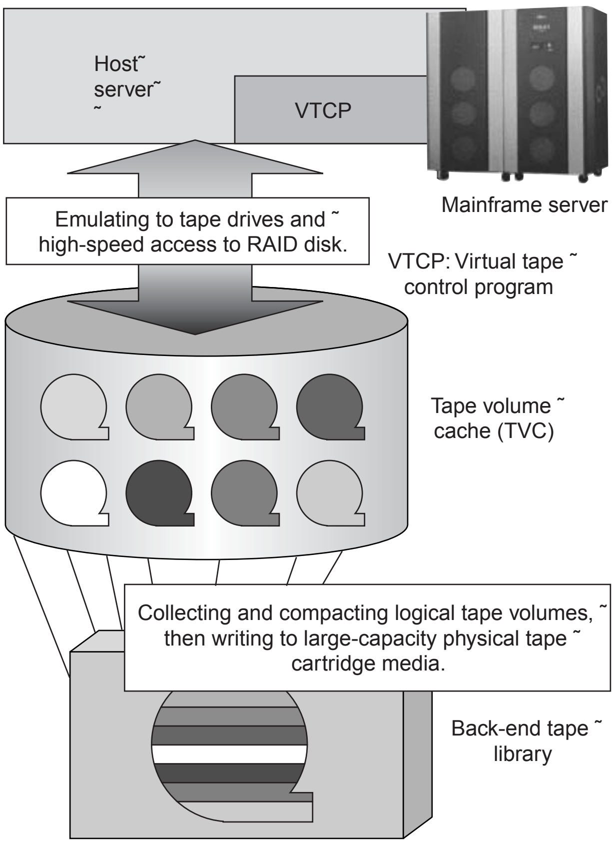


Figure 1
Basic concept of ETERNUS VT600.

1) Tape emulation for host server

Emulates the F6473K 36 track cartridge tape drive.

Up to 256 drives are virtually available on the VT600.

2) Compression of logical volumes

Data from the host is compressed and then stored in logical volumes (LVs) in the disk array. In addition to reducing the size of data, compression also reduces the amount of data transfer to the disk array and physical drives, which improves the performance of the entire system.

3) Volume stacking

Multiple LVs are stacked onto a single cartridge tape. As a result, tapes are used almost to their full capacity, so automatic, efficient use of capacity is achieved.

4) Immediate migration

When an LV is unloaded, it is written to a real cartridge tape as soon as possible. The target LV is not deleted from the TVC as long as no error occurs in the TVC area, even after writing to a physical volume (PV) is completed.

2.3 System configurations

The VT600 is a control device that provides a virtual high-end tape subsystem resource for applications on a host server by connecting with the host server interface and creating a VTL system for it. Virtual tape control program (VTCP) is a newly developed host server software for controlling the VT600. VTCP controls the VT600 subsystem via a LAN. **Figure 2** outlines the configuration of the host (Fujitsu Global Server) and the VT600.

The following software programs are installed on the host server, and tape operations are performed in the VT600.

1) Magnetic tape control program (MTCP)

A utility software that initializes and copies magnetic tapes

2) Virtual tape control program (VTCP)

Software that controls the VT600

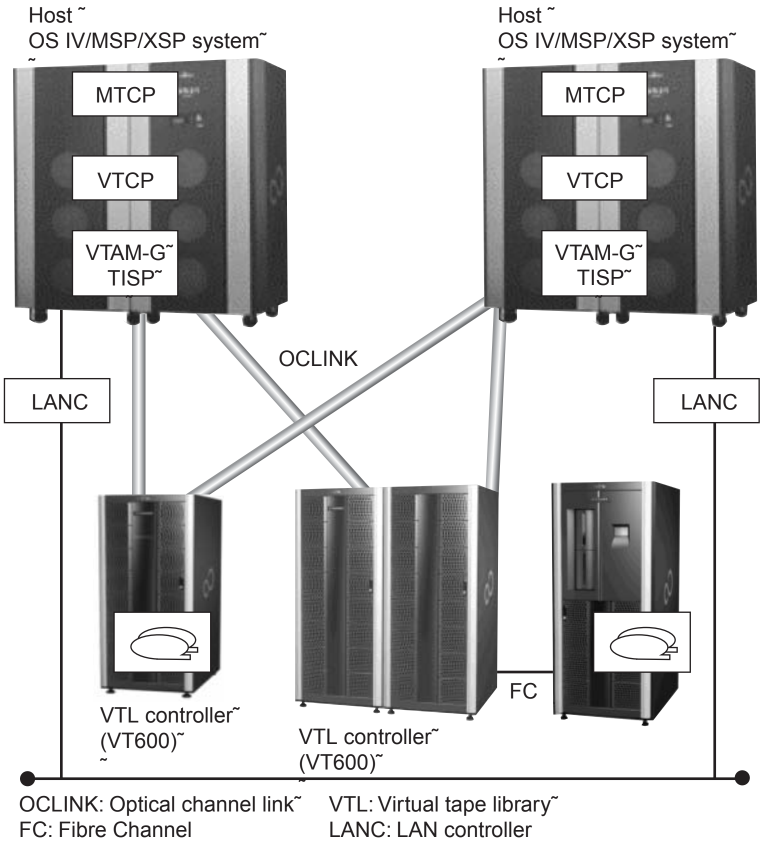


Figure 2
System configuration.

3) VTAM-G TISP (VTAM-G TCP/IP Support Program)

Software that supports the TCP/IP telecommunication function.

3. ETERNUS VT600 functions

3.1 Volume groups

A PV that contains the data of an LV can be controlled by grouping the PV with the LV. As a result, PVs can be prevented from receiving specific LV data and efficient capacity management becomes possible (**Figure 3**).

3.2 Dual save and multiple back-end tape library functions

Based on the volume group concept, the VT600 provides a dual save functionality by which LVs are written to two different PVs by linking them with dual physical volume groups (PVGs). Dual PVGs can be selected from the same library (local dual save) or a remote library (remote dual save) (Figure 3).

If a read or write error occurs on a PV, the PV is registered as faulty, and all LVs from that PV are reorganized so they once again exist on two PVs in the two specified PVGs.

3.3 Import/export function

Multiple LVs can be copied to a PV, which can then be ejected and stored in a safe place outside the VT600. When a disaster occurs on the system, the data can be recovered by remounting the PV onto the VT600 and importing the LVs it contains. A PV can be ejected from the VT600 simply by specifying its volume name and other information from the host (**Figure 4**).

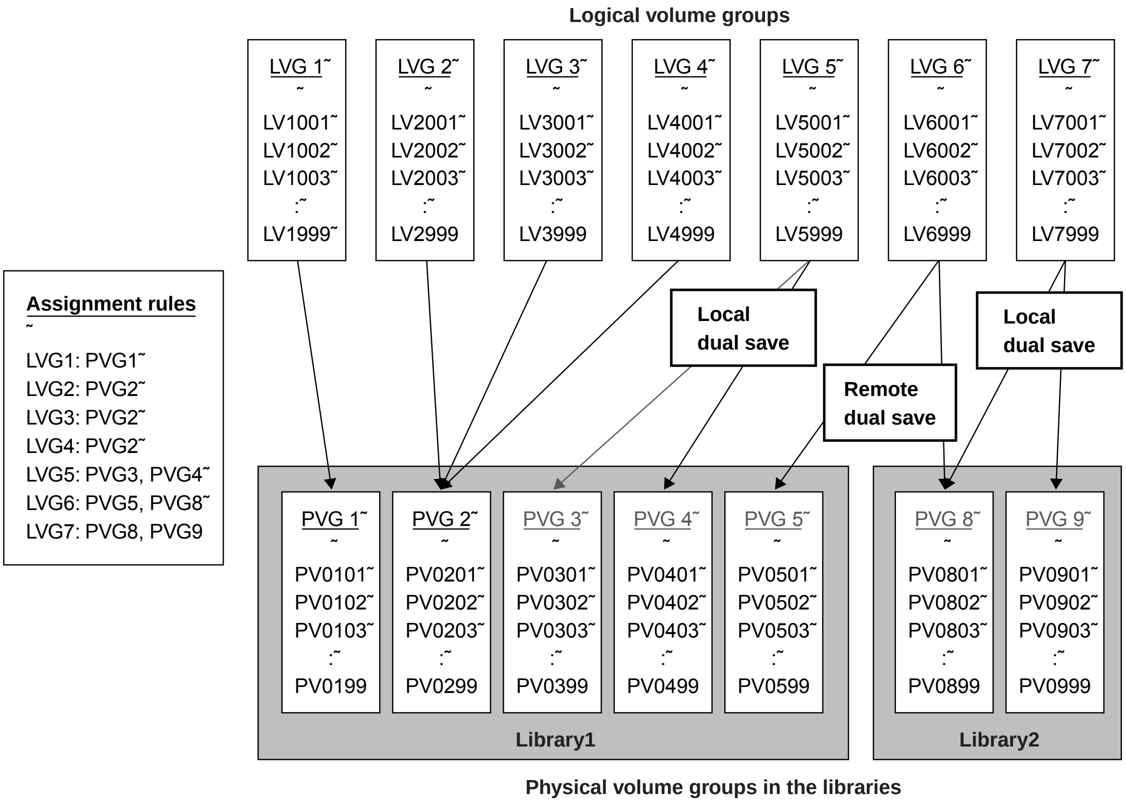


Figure 3
Volume groups of ETERNUS VT600.

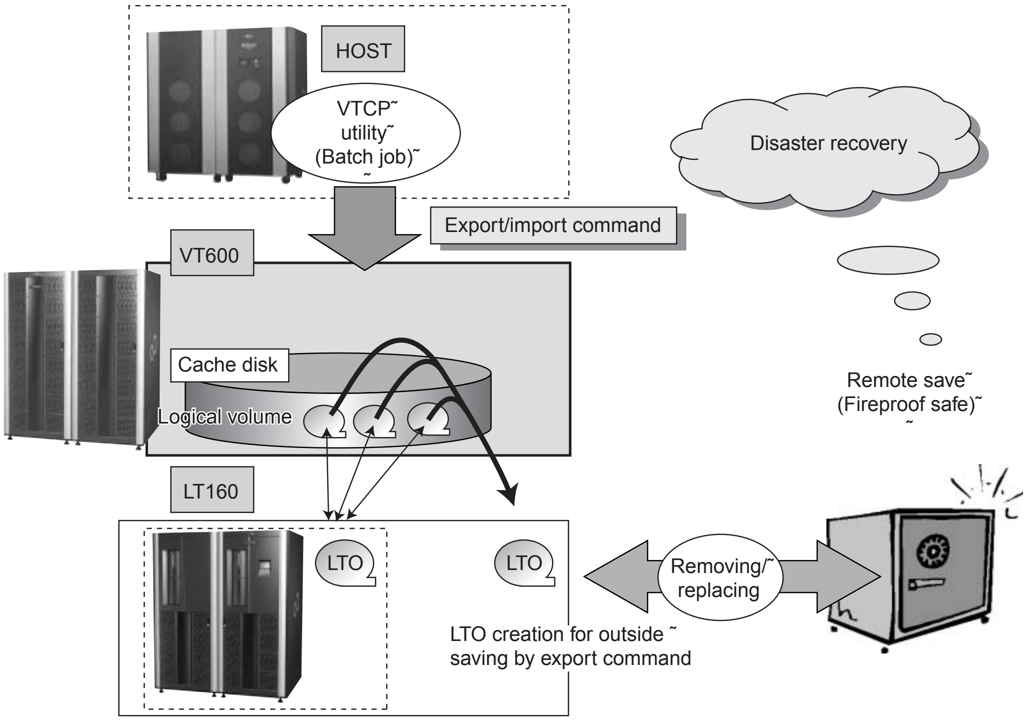


Figure 4
Import/export function.

3.4 Tape volume cache (TVC) partition function

This function enables individual cache file systems to be reserved for exclusive use by particular LV groups (**Figure 5**). LV groups that are not assigned to a cache file system are distributed to the remaining caches (FLOATING setting).

LVs are stored on the cache file system so they can be quickly accessed from the host. However, if there is insufficient space on the cache file system to contain a new LV, data that has not been accessed recently is removed so the LV can be stored.

This function enables specified LV groups to remain on the cache file system, even if they are not accessed for extended periods.

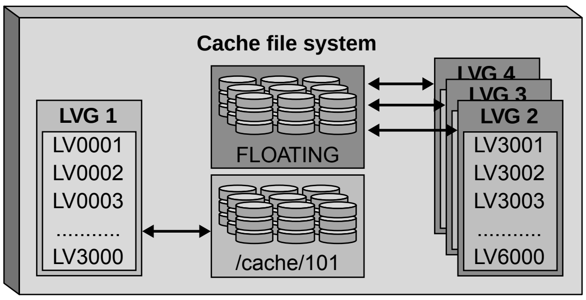
3.5 Preload function

A migrated LV can be pre-staged (recalled) in the TVC before its associated job is executed so there is no recall processing delay when the job is executed.

3.6 Virtual library processor (VLP) failover

Two sets of the VLP are built in, and usually one of them controls the VT600's internal operations (active/standby architecture).

The standby VLP continuously monitors the operating VLP, and when it detects an error, VLP processing is automatically failed-over to the standby VLP. As a result, operational continuity of the virtual tape system is maintained.



In this example the LV group LVG1 is assigned the cache file system /cache/101.

The LV groups LVG2, LVG3, and LVG4 are distributed to the remaining caches (FLOATING).

Figure 5
Example of exclusive use of cache file system by LVG.

3.7 Reorganization

When an LV is updated, the new LV data is added to the end of a PV, and the data that was written to the PV before the update becomes invalid. Therefore, as PV data update is repeated, the invalid area on the PV increases in size and the number of empty PVs in the PVG decreases.

To cope with this problem, only the data of an effective LV is aggregated from one PV to another. After aggregation, the original PV is reinitialized and freed, which improves the usability of the virtual tape system's PVs.

4. ETERNUS VT600 configuration

4.1 Hardware components

The VT600 is composed of servers, the software on those servers, RAID disks for caching tape volume data, and a back-end tape library. The hardware components are interconnected by an FC network and a LAN with redundant configurations as shown in **Figure 6**.

4.1.1 Integrated channel processors (ICPs)

The ICPs communicate with the hosts via an OCLINK or FC interface and access the TVC for tape volume data.

4.1.2 Integrated device processor (IDP)

The IDP communicates with the tape drives in the LT160 tape library. It stores the LV data on the TVC to the PVs of the tape library and restores the LV data to the TVC by reading it from the PVs of the library.

4.1.3 Virtual library processor (VLP)

The VLP communicates with the host via a LAN and controls and monitors the VT600 hard-

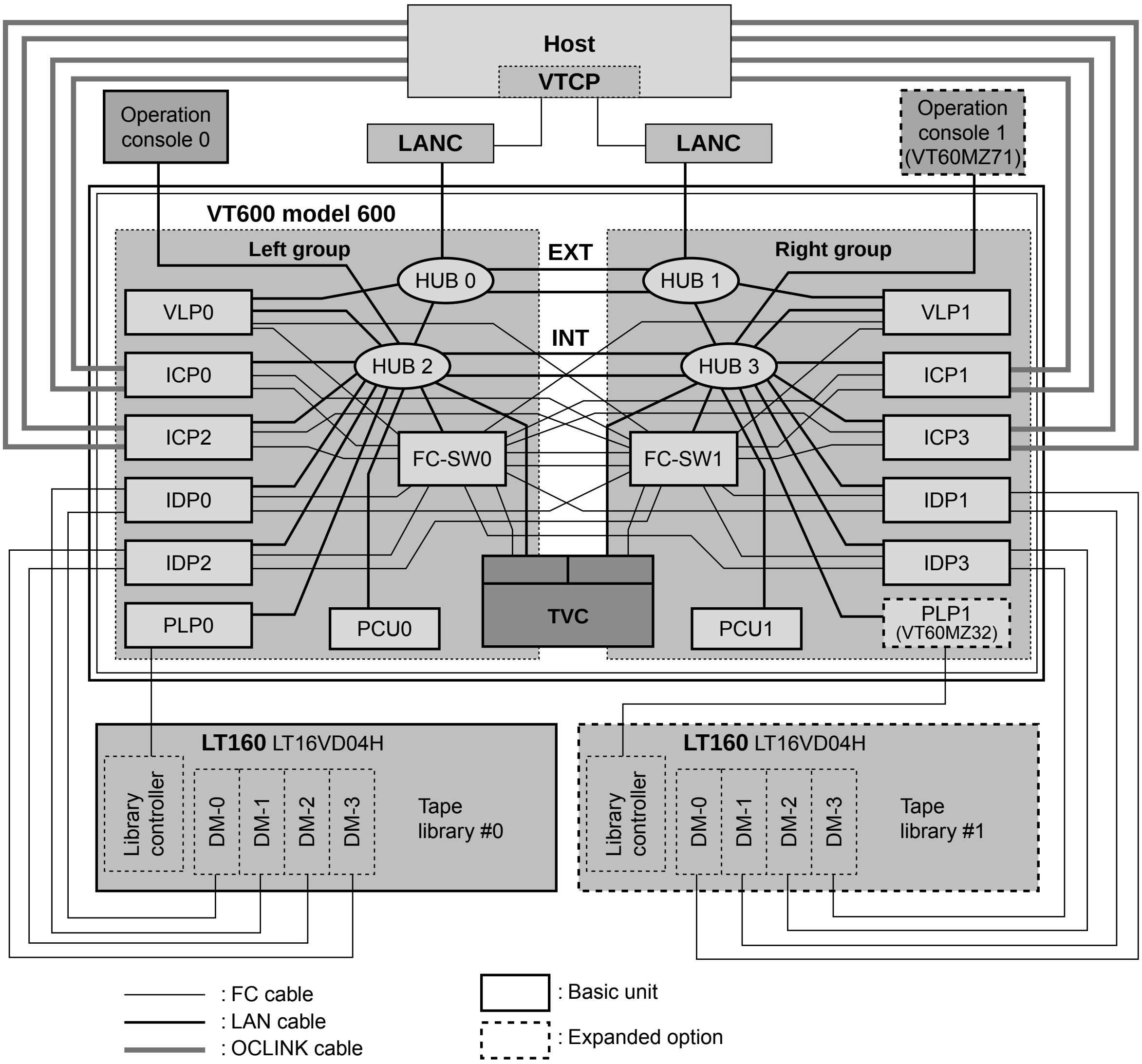


Figure 6
Hardware components.

ware and software components. It also receives mount requests from the host and then manages the mounting of LVs onto the virtual tape drive. The VLP also manages the information database for the LVs and PVs.

4.1.4 Tape volume cache (TVC)

The high-reliability, high-performance ETERNUS3000 RAID consists of a tape volume data cache and buffer equipment.

4.1.5 Back-end tape library

The ETERNUS LT160/LT130/LT270 tape libraries can be connected as back-end tape library components. All of the LV data is backed up in the back-end tape library.

4.1.6 Physical library processor (PLP)

The PLP communicates with the VLP and controls the robots of the back-end tape library.

4.1.7 Operation console

The operation console is used to monitor the operating status of each server of the VT600, for the registration of LVs, and other tasks.

4.1.8 Power control unit (PCU)

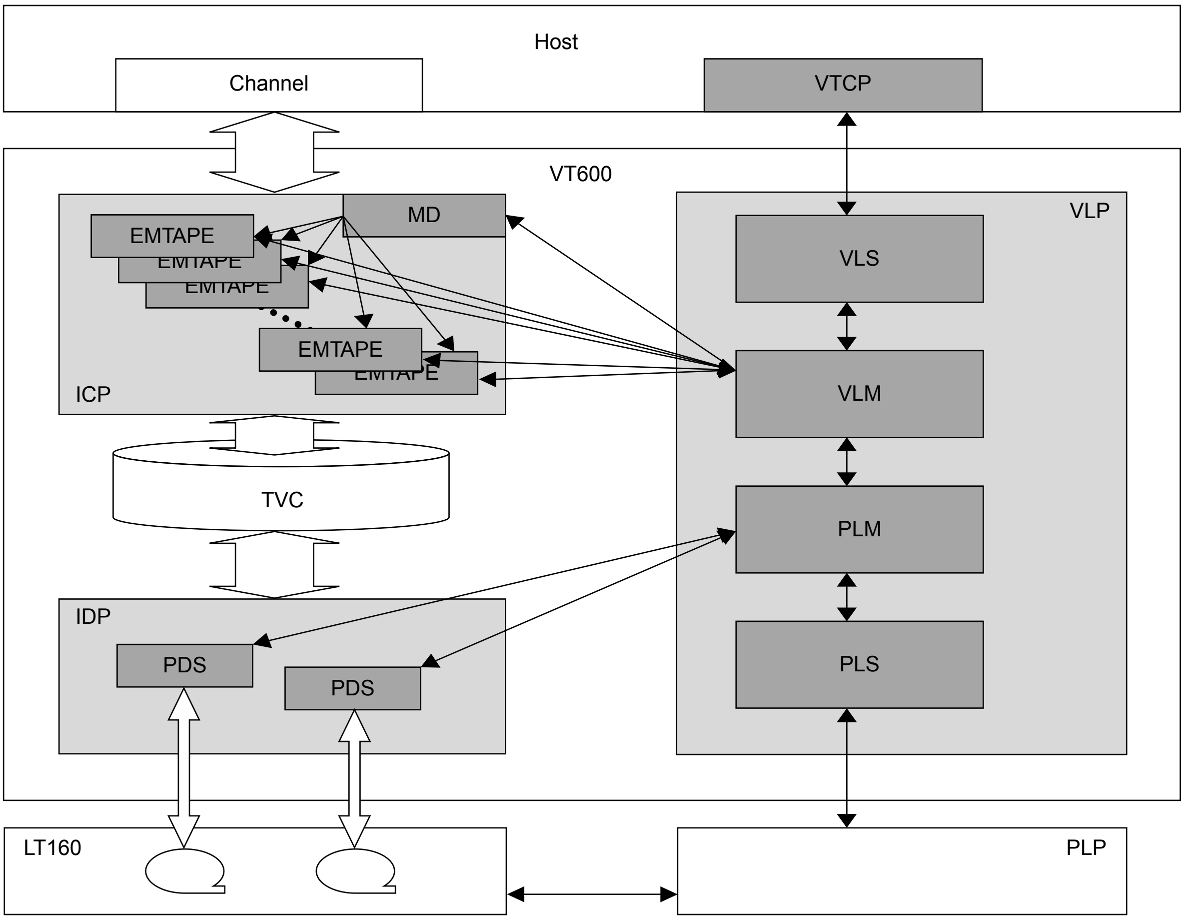
The PCU is connected to the power supply control interface (source code control interface [SCCI] and remote cabinet interface [RCI]) with the upper-level host servers. It orders power on/off of the power distribution units (PDUs) connected to the AC power line of each hardware unit on the VT600 system and power on/off of the units connected to the internal interface. The PCU, therefore, manages the power on/off of the entire VT600 system.

4.2 VT600 software modules and components.

The VT600 consists of the software mounted on each server and a virtual tape subsystem (**Figure 7**).

4.2.1 ICP software components

Depending on the type of host system used, multiple channel interface cards of different types can be installed on the host side of each ICP. The ICPs have a dual FC interface board for connec-



ICP: Integrated channel processor
IDP: Integrated device processor
EMTAPE: Tape emulation
MD: Mount daemon
PDS: Physical device service
PLM: Physical library manager
PLP: Physical library processor
PLS: Physical library service
TVC: Tape volume cache
VLM: Virtual library manager
VLP: Virtual library processor
VLS: Virtual library service
VTCP: Virtual tape control program

Figure 7
VT600 software modules and components.

tion to the TVC of the ETERNUS3000 RAID system.

The main task of the ICP is to emulate logical tape drives for the connected host systems. The ICP stores LVs in compressed form in the TVC.

4.2.2 IDP software components

The IDPs are responsible for communication and data transfer between the TVC and the physical drives in the tape library. For external connections, the IDPs are equipped with two SCSI controllers (LVD or HVD) or an FC controller (two ports).

A physical device service (PDS) is generated for each physical drive connected to an IDP. PDSs are subordinate processes for the physical library manager (PLM) described below.

The main task of the PDSs is to transfer the data of LVs between the TVC and PVs.

4.2.3 VLP software components

The VT600's internal processes and operations are coordinated by the virtual library manager (VLM) and the physical library manager (PLM), which communicate closely with each other via a small number of commands such as save and restore LV. Additionally, there are one or more virtual library services (VLSs) and one or more physical library services (PLSs) that subordinate processes to the VLM and PLM.

1) Virtual library service (VLS)

For a logical drive in an ICP connected to a host system, the host application must issue a logical mount command via a LAN to the VLP before access to the logical drive via FC can be granted.

VLS emulates different tape libraries so that new interfaces do not need to be installed on the host's operating systems.

The most important commands that are common to all interfaces are the mount, unmount, and query commands. The purpose of the VLS is to translate the different library interface commands into common commands for the VLM.

2) Virtual library manager (VLM)

The main task of the VLM is to manage the status and location of logical drives (LDs) and LVs in the TVC of the RAID system. The information about the LVs is located in the TVC and the VLM database. The VLM database is also located on the RAID system, and a standby copy of it is kept on the local RAID disks of the VLM so the VLP can be swapped to the standby VLP in case of a failure.

From a connected host system, the VLM registers each mount request for an LV via the VLS (described above). The VLM then checks whether the LV is in the TVC, and if it is (cache hit), the system reserves the required disk space based on the maximum volume size of 800 MB. If the TVC has insufficient capacity, the system deletes the data of the oldest and largest volumes from the TVC according to a modified Least Recently Used (LRU) method. Then, the VLM tells the tape emulation the location of the LV data in the TVC. Next, the tape emulation searches for the file and tells the VLM that the LV is loaded when it finds the file. The VLM then acknowledges the mount to the host via the VLS, which then starts tape processing on the LV. This procedure, therefore, mounts the LV onto the logical drive.

If the requested LV is not in the TVC (cache miss), the VLM sends a restore request to the PLM (described below) requesting that the LV be read from the PV in the tape library and returned to the reserved space on the TVC.

When the host application wants to finish operations on the LV, the host issues an unload command and unmount command to free the LV and LD for the next mount process.

In addition, the VLM must check whether the LV in the TVC has been modified. If it has been modified, the VLM asks the PLM to copy the modified version to a PV in the library. Even though the save request is completed immediately, the copy to tape ordered by the PLM is performed after the host receives the unmount LV acknowledged by the VLM. Therefore, the host can continue operation almost immediately after the unmount of the LV, and no PV or library latency are involved during this process.

3) Physical library manager (PLM)

The PLM coordinates all peripheral-related functions. It manages the physical drives (PDs) and PVs in the connected tape libraries. When PVs are added to a PVG, the tapes are subsequently labeled. Therefore, the tapes must already exist in the physical tape library.

The PLM gets its jobs directly from the VLM, which sends the following requests to the PLM:

- Save an LV from the TVC to a PV.
- Restore an LV from the PV to the TVC.
- Delete an LV. The PLM registers the LV in the tape directory of the PV as invalid and completely purges the LV from the PLM database.

In the first two cases, the PLM selects a PD and PV and initiates, via the respective PLS, the mounting of the PV on which the active LV is or will be stored. It then orders the corresponding PDS process to transfer the LV between the TVC and PV.

The third case applies when a host is no longer permitted to use an LV and the administrator subsequently wants to remove the LV from the CentricStor database with the appropriate VLM command. In this case, the VLM sends an internal delete command to the PLM.

4) Physical library service (PLS)

The PLS handles communication between the PLM and the physical library interface. It translates mount and unmount commands from the PLM into the correct library interface commands and sends them to the physical library. If the communication runs over TCP/IP (network-attached), the distance between the VLP and the libraries is unrestricted.

4.2.4 PLP software components

The PLP software manages the PVs contained in the back-end tape library, processes the mount requests from the VLP, and controls the robots of the back-end tape library. It also manages the number of mounts for each PV and periodically cleans the heads of PVs in the back-end library.

4.2.5 TVC control software modules

The RAID disk storage system is the heart of the entire virtual library system. As mentioned earlier, the data of the LVs to be read or written is stored in the (non-volatile) TVC on RAID disks. The maximum file size of a cached volume is 800 MB; these values can be two to three times higher when compression on the ICP is used. An important point is that data transfer between a host and drive runs exclusively on the TVC. Because the VT600 internal design is based on the Storage Area Network (SAN) concept, any VT600 internal server (i.e., ICP, IDP, or VLP) can access the TVC. The amount of time that data stays in the TVC is governed mainly by factors such as the cache size, size of LVs, and pattern of host accesses to the data. The metadata of an LV and its first 128 kB are maintained permanently in the TVC, even when the rest of the LV has been migrated to a PV. This first 128 kB is called the LV stub and includes the volume label. As a result, pure “bookkeeping” functions can be performed very quickly because they require no action by the physical tape library.

Each ETERNUS3000 RAID system contains a TVC, and the VT600 can be connected to multiple ETERNUS3000 RAID systems using tape volume cache extensions (TVCEs) to increase the total capacity and throughput performance of the VT600.

The first RAID system of the VT600 always contains the metadata, which mainly consists of the databases of the VLM and PLM. A copy of the metadata resides on the internal local RAID disks of the VLP.

Each TVC has a distributed tape volume file system (DTVFS), which is mounted on all VT600 internal servers and in which the LVs are stored.

The VLM is responsible for the administration of the LVs in the different TVCs, and as the master of the DTVFSs, the VLM allows tape emulations and the PDS processes to access its LVs.

4.3 Virtual tape control program (VTCP)

VTCP is the host server software that controls the VT600 (**Figure 8**).

1) Mount/unmount function

The VTCP uses this function to check the mount/unmount requests of a virtual tape for the virtual drive generated at the host and makes requests to the virtual tape drives.

When the VTCP intercepts a mount request for a logical volume, it issues a mount request to mount the volume onto a virtual tape drive using the volume-request serial number assigned to the volume.

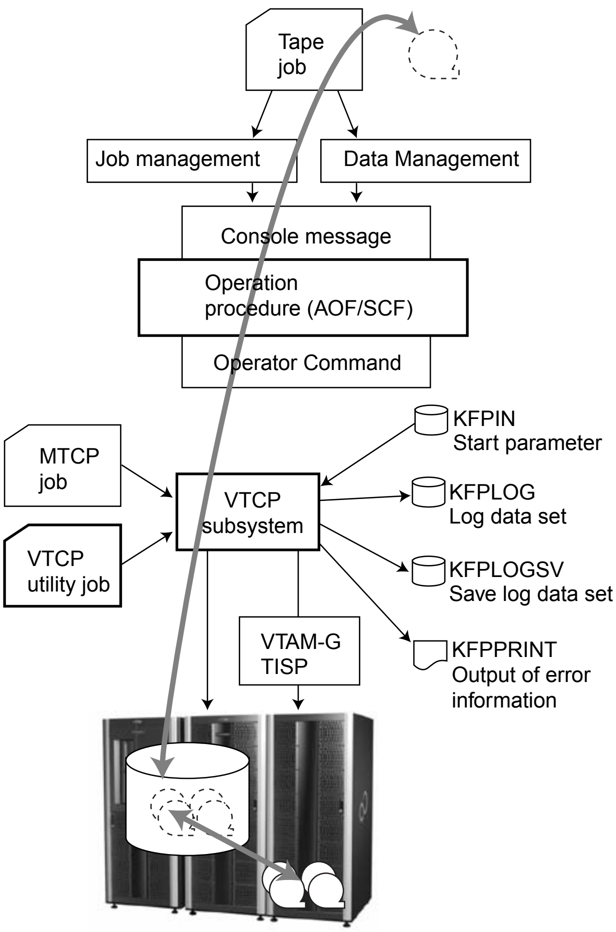


Figure 8
VTCP operation.

2) Scratch pool select function

The scratch pool select function assigns virtual tapes for requests for the scratch volume (unspecified volume) specified in the job DD statements (FD statements). These requests do not specify a volume serial number.

Other functions include the assignment select function of a virtual drive and write protect function of a virtual tape.

5. Benefits and advantages of ETERNUS VT600

The VT600 brings the following major benefits and advantages to tape application processing in customers' IT systems.

1) Faster data processing

Mechanical operations (mount, unmount, load, unload, and rewind) of cartridge tapes become unnecessary due to the tape emulation, which increases the speed of tape operations (**Figure 9**).

2) Efficient use of cartridge tapes

Each LV is written to a cartridge tape of the back-end tape library continuously (streaming mode). As a result, the efficiency of cartridge tape operations is increased.

3) Smaller installation space

Large-capacity, high-speed LTO drives are used for the back-end tape devices, and as mentioned above, the cartridge tapes are used efficiently. As a result, the VT600 requires much less installation space than other large-scale tape

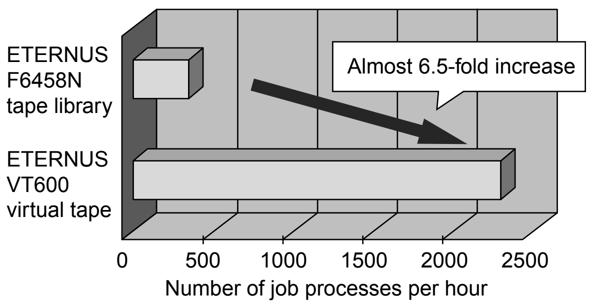


Figure 9
Number of job processes per hour (50 MB file).

libraries, for example, the F6455, F6457, and F6458 (**Figure 10**).

6. Enhancement of ETERNUS VT600

6.1 Open-system connectivity

The VT600 was initially introduced to Fujitsu's mainframes to realize efficient tape operations and improve the customers' TCO.

There was no immediate need to introduce virtual tapes as there was with mainframes, which currently cannot use the larger-capacity cartridges that have become available. However, order-of-magnitude increases in cartridge capacity and transfer rates required new methods of using tapes that conflicted with operational and technical needs. The VT600 is an open-system solution that satisfies a broad range of requirements in heterogeneous environments containing mainframes and open systems (**Figure 11**). Some examples of these requirements are:

1) High performance and cost effectiveness (media usage up to 100% and consolidation of devices)
2) Optimal usage of latest technologies
3) Reduced administration costs
4) Increased application robustness and availability

6.2 Encryption for export function of ETERNUS VT600

It is important to provide security for exported data, and the VT600 does this by placing an encryption device between itself and the back-end library. First, the host sends data to the VT600. Then, the VT600 prepares the data for backup and sends it to the encryption device. The encryption device then encrypts the data and exports it to

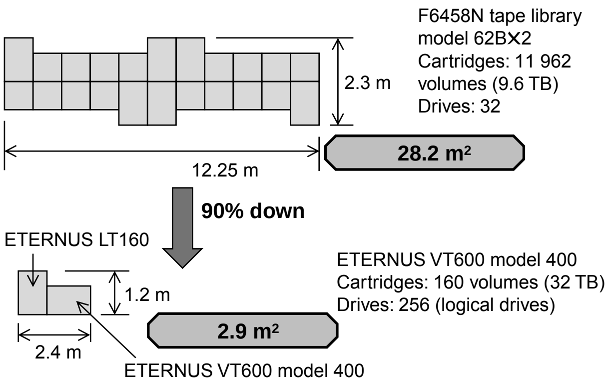


Figure 10
Installation space.

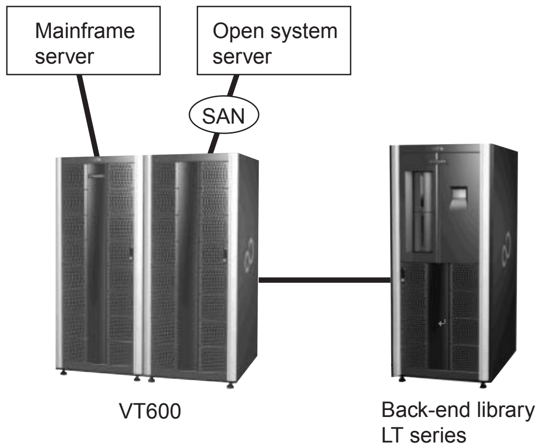


Figure 11
System connectivity.

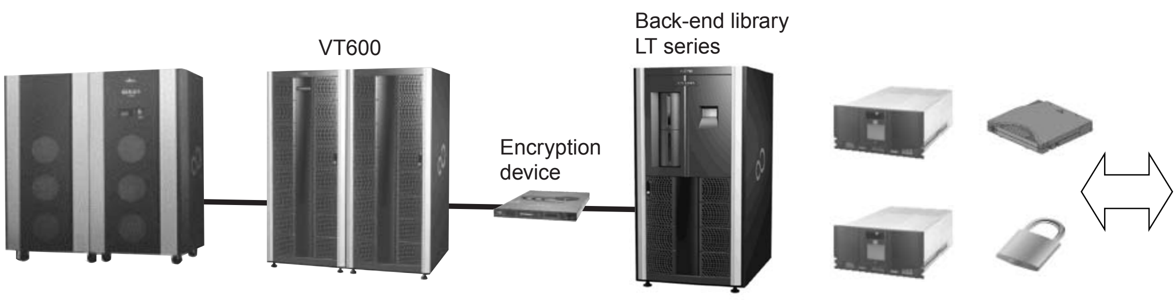


Figure 12
Encryption for export function.

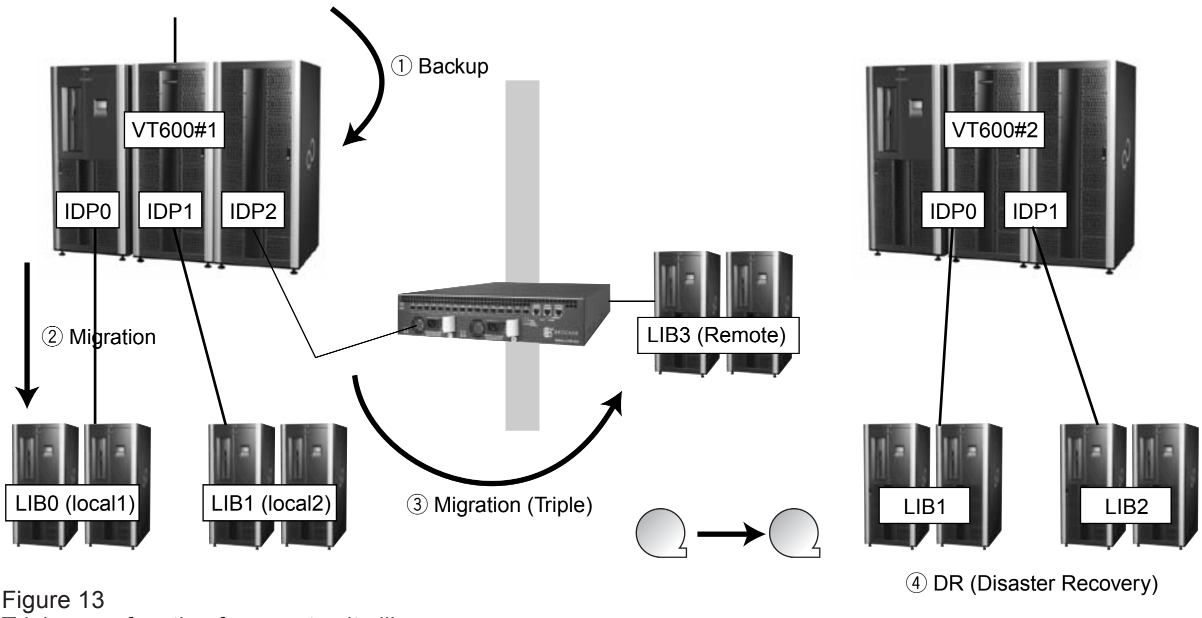


Figure 13
Triple save function for remote site library.

the external cartridges (**Figure 12**).

6.3 Triple save function for remote library

The VT600 provides a dual save function for multiple back-end libraries. To protect against disasters, one more data saves are required to migrate data to a remote back-end library. (**Figure 13**) The VT600 will soon provide a triple save function that configures two migration volumes in the local library and one migration volume in a remote library. This function will satisfy customers' disaster recovery requirements for their data.

7. Conclusion

The ETERNUS VT600 is a newly developed VTL system for Fujitsu's mainframes and open systems. It increases the speed of tape data processing; realizes efficient, automated use of the latest large-capacity physical tape devices; and saves customers a significant amount of installation space. As a result, the TCO of customers' IT systems can be immediately improved. Enhancements desired by customers such as comprehensive tape application consolidation under heterogeneous server environments and stronger data protection and disaster recovery will become available in the future.

References

1) ETERNUS VT600. (in Japanese).
http://storage-system.fujitsu.com/jp/products/virtualtape/vt600/
2) FSC: CentricStor.
http://www.fujitsu-siemens.com/products/storage/centricstor/index.html
3) ETERNUS3000.
http://www.fujitsu.com/global/services/computing/storage/system/eternus3000/

Takashi Miyako received the B.S. degree in Electronics Engineering from Yokohama National University, Yokohama, Japan in 1982. He joined Fujitsu Ltd., Kawasaki, Japan in 1982, where he has been developing virtual tape library systems since 2003.

Tatsuya Ohmido received the B.S. degree in Computer and Information Sciences Engineering from Iwate University, Morioka, Japan in 1986. He joined Fujitsu Ltd., Kawasaki, Japan in 1986, where he has been developing virtual tape library systems since 2003.