Hitachi’s Gigabit Switch 4000/3000 Series for Implementing Guaranteed Networks

Kentaro Hashimoto
Tomoaki Kouzu

OVERVIEW: There has been tremendous increase in the volume and diversity of data as computer technologies have become more advanced in recent years, and this has created demand for networks that can carry larger quantities of data at higher throughputs. This has fueled a migration to broadband networks, and an active effort to integrate different kinds of data transmissions into one network to reduce costs. However, the inherent behavior of broadband networks is to make a best effort to deliver transmissions. Since best effort provides no absolute guarantees, this has led to a demand for different QoS (quality of service) classes to accommodate different traffic priorities and guaranteed networks that can ensure real-time delivery of data even if a network device or line fails. In order to meet this need, Hitachi, Ltd. has developed and deployed gigabit switch 4000/3000 series that guarantee bandwidth availability, that continue to support traffic even when a device or line fails, and that provide traffic volume monitoring and a host of other advanced functions. The combination of these capabilities support guaranteed networks.

INTRODUCTION

RECENTLY we have seen remarkable progress in IP (Internet Protocol) and Ethernet* based networks and the scale of network platforms as well as transmission speeds have both shown a dramatic increase. Penetration of Ethernet technology has not just been confined to in-house LANs (local area networks) but is now manifested as wide-area Ethernet to provide fast, cost-effective WAN (wide area network) services that is further hastening the migration to broadband networks. Meanwhile, applications are also becoming more diversified with the proliferation of multimedia services including audio and video content and the emergence of IP-based SAN (storage area networks) systems.

Up to now, companies have used highly reliable

* Ethernet is a registered trademark of Xerox Corp.

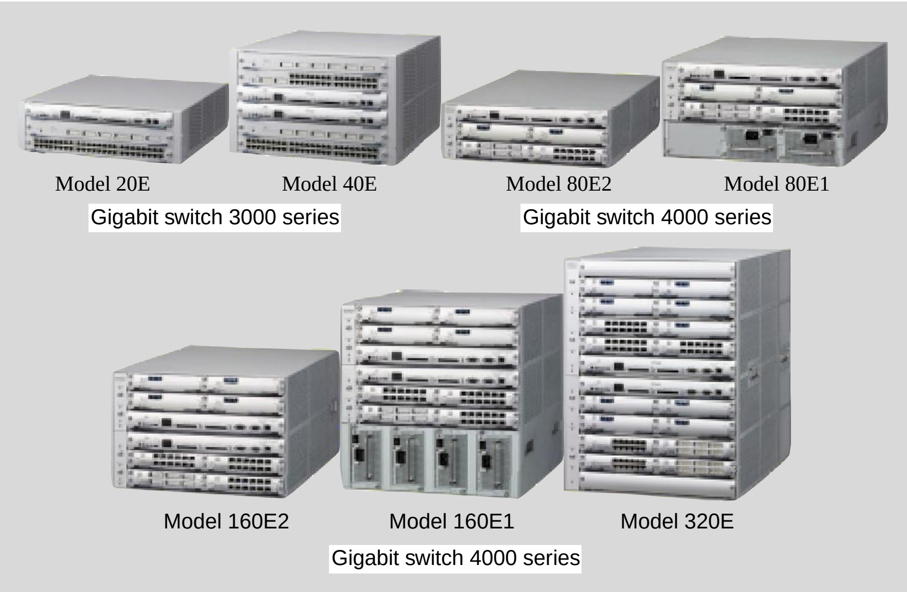

Fig. 1 — Hitachi’s Gigabit Switch 4000/3000 Series.
The 3000 series of this gigabit switch has two models: 20E and 40E for small- and mid-size networks. The 4000 series of the gigabit switch has three models for large-scale networks, i.e. 80E, 160E, and 320E.

but costly leased lines for internal corporate networks that carry mission-critical business transactional data that demands the highest standards of reliability. Moreover, while companies typically use dedicated phone networks for in-house extension phone systems, they use either IP or Ethernet for file transfer, web browsing, and other data communications applications, and separate networks are implemented for different applications. Unfortunately, the costs to build and to operate, and to administer separate networks used for different purposes becomes increasingly burdensome. This has led to recent efforts to combine these various networks on one integrated IP-Ethernet to reduce the operating and administrative burden, and the number of actual implementations of this integrated approach is growing. In this paper, we describe one such implementation based on gigabit switch 4000/3000 series developed by Hitachi (see Fig. 1).

REQUIREMENT FOR GUARANTEED NETWORK

Ethernet that is implemented in most corporate LANs was developed specifically to provide a fast but economical shared transmission medium. Given these characteristics, networks based on the Ethernet standard are called best-effort networks. Best-effort networks generally provide near maximum performance if one has exclusive access to the lines implementing the network, but as the number of users increases, the transmission medium is divided up among the users and performance deteriorates. Best-effort networks are thus not suitable for real-time communications such as voice communication or video conferencing that require minimum guaranteed bandwidth even if mixed on the same transmission medium.

With its inherent advantages for building fast, low-cost networks, Ethernet is a good choice for reducing the costs associated with network systems. But if Ethernet is used to implement an integrated network combining applications and services with different quality requirements — mission-critical business transactions, file transfer, web browsing, and other data communications applications, voice, etc. — then satisfactory QoS (quality of service) cannot be obtained unless the various problems outlined above are resolved. Furthermore, if various kinds of traffic and applications are integrated on one network, then all applications are vulnerable to a network failure. For system configurations of network, enhanced reliability and redundant technologies are thus required to keep the system operational. With these considerations in mind, Hitachi developed and deployed the 4000 series and 3000 series of the gigabit switch for implementing guaranteed Ethernet-based networks.

FEATURES OF GIGABIT SWITCH 4000/3000 SERIES

Quality of Service

Hitachi's gigabit switch 4000/3000 series incorporate features capable of converting best-effort Ethernets into guaranteed Ethernet-based networks with granular and precise QoS controls.

The switches use dedicated hardware to detect different kinds of traffic, divide the traffic into different classes, and determine output scheduling at very high speed. Each interface is equipped with shaping capability that shapes the output traffic volumes at specified bandwidths, and queuing functions are implemented within the specified bandwidths including priority queuing and round-robin queuing that processes in sequential order.

The gigabit switch 4000 series can be equipped with a hierarchical shaper based technologies acquired in developing the ATM (asynchronous transfer mode) switch and gigabit router. The hierarchical shaper can flexibly provide bandwidth for up to 1,024 users individually within the bandwidth provided by the physical line, and it can schedule output according to priority queuing, weighted fair queuing or other policies in four traffic classes within the bandwidth allocated to each user.

Redundancy Technologies

Redundancy capabilities between devices

Hitachi developed a gigabit switch redundancy protocol, a proprietary protocol enabling redundant configuration between two switches at Layer 2. It ensures stable continuous operation without a loop developing by first making sure that a failed active switch has actually been set to standby status before the standby switch becomes active status.

Redundancy between Layer 3 switches is implemented with the industry standard VRRP (virtual router redundancy protocol) that is compliant with RFC2338. VRRP permits the definition of a virtual router address that actually consists of multiple routers. By setting up a virtual router address as the default gateway, network devices can automatically switch over to the standby router if the master router fails and continue to support communications.

Ordinary VRRP routers are only able to monitor

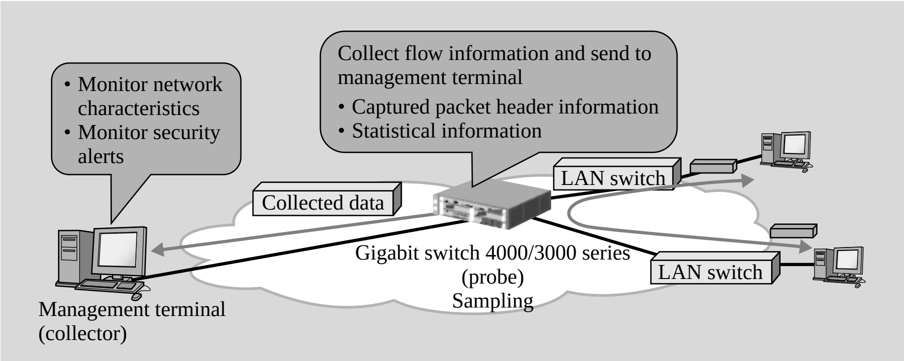

Fig. 2—Scheme for Collecting Statistical Traffic Data with sFlow Monitoring System.
Header information of packets passing through network equipment is collected and network traffic trends are analyzed.

the transmission paths to which they are directly connected, and therefore cannot detect communications failures that exist across multiple routers. To address this potential problem, Hitachi developed and incorporated a proprietary VRRP polling function. This capability periodically sends out a polling signal (a monitoring packet) to IP addresses specified in the master router (e.g. the adjacent router). If it fails to receive an acknowledgment, this indicates a line failure, and it causes the system to automatically switch over to a backup system, the same as if the master router failed. With this capability, WAN redundancy is supported in addition to router redundancy.

Support for physical line redundancy

Physical line redundancy is implemented by the IEEE 802.3ad link aggregation standard. Link aggregation means grouping multiple physical links and treating them as one logical higher bandwidth capacity link. If one of the physical links configuring the logical link fails, communication is nevertheless sustained using the remaining physical links.

In the link aggregation implemented in the gigabit switch 4000/3000 series, up to 16 physical links can be grouped in one logical link, which can be combined across an interface card slot. What traffic is allocated to which physical links is determined by addresses in the data frame, application information, etc. This permits traffic to be sequentially guaranteed and well balanced among the physical links, also makes it easy to add additional bandwidth capacity. We have also implemented two unique capabilities: a standby link function that prevents any decline in bandwidth capacity by automatically switching over to a standby physical link if an active physical link fails in the logical link, and a function limiting the number of ports that can be disengaged that can be set to shut down a logical link if even one physical link in the logical link fails.

Statistical Data Collection Function

The ability to measure the quantity of data traffic flowing through the network is essential to design future networks. The ability to monitor sudden upsurges in traffic volumes is also necessary to detect DoS (denial of service) attacks, viruses and worms, pinpoint the source of an attack, and prevents the network from being affected with the damage of such attacks in large scale.

The gigabit switch 4000/3000 series support sFlow monitoring, a technology for collecting statistical traffic data in networks that is compliant with the RFC 3176. The sFlow monitoring system collects and analyzes traffic data using probe and collector functions. The probe samples the flow of packets at a fixed interval, extracts header information from the packets, and sends the information to the collector. The header information is stored in the collector and used to analyze traffic trends. Both gigabit switch 4000/3000 series support the probe function.

The sFlow monitoring system is capable of analyzing a wide range of network traffic parameters on all ports from Layer 2 up to Layer 4, and we expect to see more sFlow-compliant equipment and applications in the years ahead (see Fig. 2).

EXAMPLES OF APPLICATIONS TO IP STORAGE SYSTEMS

Here we will describe several applications of the gigabit switch 4000 series to an iSCSI (Internet small computer system interface) storage system. These scenarios are based on a series of trials interconnecting a RAID (redundant array of independent disks) system with the gigabit switch 4000 series conducted in March 2004 by Hitachi's RAID System Division and IP Network Division.

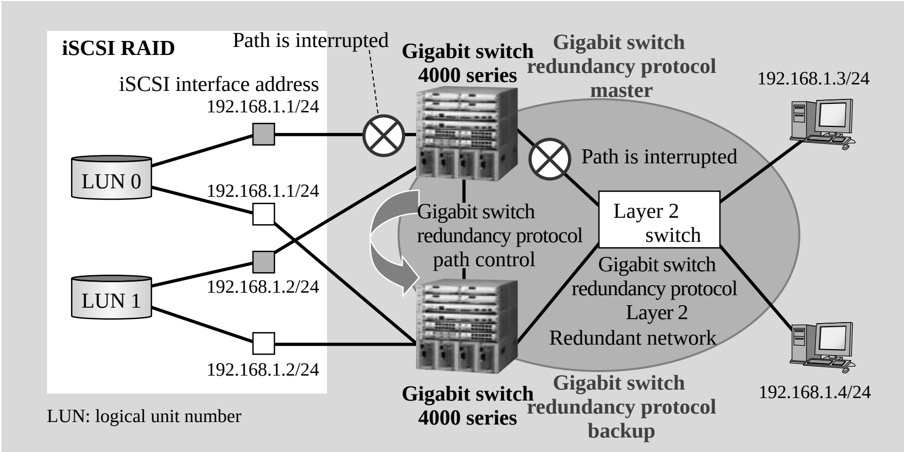

Fig. 3—Example of iSCSI Storage Redundancy Implemented with Hitachi's Gigabit Switch Redundancy Protocol.
The gigabit switch redundancy protocol functionality permits automatic switching to a backup system if a transmission path or storage interface fails without the host being aware.

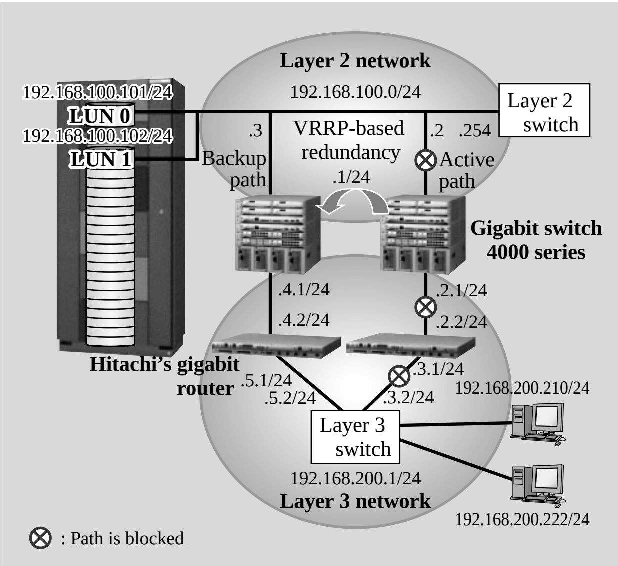

Fig. 4—Example of Storage Redundancy Implemented with VRRP.
Redundancy is achieved for all paths using the VRRP polling function implemented on the gigabit switch 4000 series.

Example of iSCSI Storage Network Redundancy Implemented with Gigabit Switch Redundancy Protocol

Fig. 3 illustrates how redundancy is implemented in transmission lines and in a storage system using the gigabit switch redundancy protocol. Here gigabit switch 4000 series redundancy is achieved using the protocol, and communication to/from the backup gigabit switch 4000 series is normally blocked. The two interfaces mapped to the LUNs (logical unit numbers) targeted by the iSCSI store the same IP address. Normally IP addresses must be unique and communications cannot be guaranteed if duplicate IP addresses exist. But in this trial configuration, communication with the backup system is blocked by the gigabit switch redundancy protocol, so the gigabit switch 4000 series are not recognized as the duplicate IP addresses from the hosts illustrated in the right side of Fig. 3. With this setup, the iSCSI target achieves redundancy by sharing the same IP address over multiple interfaces.

When the system fails and the redundancy protocol causes switching, the target iSCSI initiator once shuts down the current connection, then establish a reconnection. While there is a very brief interruption until the reconnection is established, disk accesses can continue after the reconnection is made. The system automatically switches back once the failed link is restored, and the transmission path is returned from the backup system to the master system.

Example of Storage System Redundancy Implemented with VRRP

Fig. 4 illustrates a storage redundancy solution using VRRP polling. In this configuration, a gateway address (192.168.200.1) is set on the initiator host side as the VRRP polling destination of each gigabit switch 4000 series switch configured by the VRRP. Since failures across multiple routers can be detected by the series's VRRP polling, communications can be restored by VRRP path switching regardless of which segment on the path fails.

Example of Storage Network Redundancy Implemented with Link Aggregation

Fig. 5 illustrates how storage redundancy can be achieved using link aggregation. Reduction of bandwidth capacity can be effectively prevented with a combination of the gigabit switch redundancy protocol and link aggregation described earlier. In this example, three physical lines have been combined to form one logical line. Here the redundancy protocol

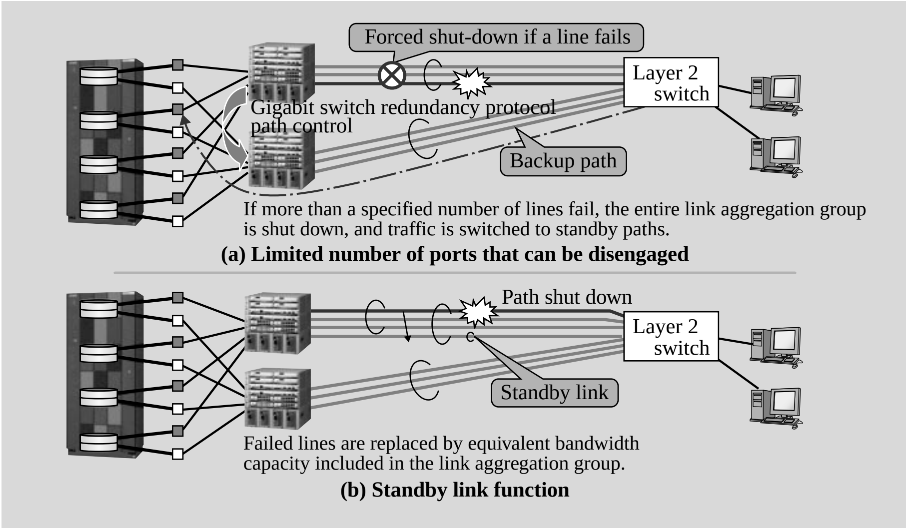

Fig. 5—Example of Storage Redundancy Implemented with Link Aggregation.
Unique capabilities of the gigabit switch 4000 series prevents any decline in bandwidth capacity even when link aggregation fails.

has been set up to switch to a backup system if the logical line implemented by link aggregation fails.

If the function limiting the number of ports that can be disengaged is used in this environment, the entire active link is suspended if even one physical link configuring the logical link fails. When this occurs, the redundancy protocol detects that the active logical path of one of the switches is blocked and automatically switches over to the backup path. The backup path supports communication over all its physical lines, so this effectively prevents any reduction of available bandwidth [see Fig. 5 (a)].

Fig. 5 (b) shows an example using the standby link function. Here, two of three physical lines are defined as a logical link while the remaining physical line is defined as a standby link. Then if one of the physical lines in the logical link fails, the standby physical line is automatically incorporated into the logical link thereby ensuring that the same amount of bandwidth is available as before the line failed.

CONCLUSIONS

This paper gave an overview of the advanced functions of Hitachi's gigabit switch 4000/3000 series, and introduced some of the capabilities with application examples.

The series feature a range of advanced functions so they can be used to implement guaranteed networks. Although not addressed in this paper, the gigabit switch 4000 series are also scalable. They deploy interfaces supporting throughputs up to 10 Gbit/s and thus can readily accommodate future network expansion.

Given the robust performance of its gigabit switch 4000/3000 series, Hitachi will continue finding new applications for the switches while working to further enhance their functional capabilities and reliability.

REFERENCE

(1) S. Ono et al., "Gigabit Routers Supporting IPv6 Commercial Networks," *Hitachi Hyoron* 84, pp. 387-390 (May 2002) in Japanese.

ABOUT THE AUTHORS

Kentaro Hashimoto
Joined Hitachi, Ltd. in 2000, and now works at the Network Center, the Enterprize Server Division, the Information & Telecommunication Systems. He is currently engaged in customer technical support work. Mr. Hashimoto can be reached by e-mail at kehashi@itg.hitachi.co.jp.

Tomoaki Kouzu
Joined Hitachi, Ltd. in 2003, and now works at the Network Center, the Enterprize Server Division, the Information & Telecommunication Systems. He is currently engaged in product promotion technical support. Mr. Kouzu can be reached by e-mail at tomoaki.kouzu@itg.hitachi.co.jp.