Next-Generation IT Strategy Based on ALL-IP Strategy

● Haruo Akimoto

(Manuscript received September 1, 2003)

The broadband IP network is accelerating the cooperation between IT systems such as servers and storage. Fujitsu's key strategy for next-generation IT systems is called the ALL-IP strategy. In this strategy, all IT resources, including servers, are interconnected over the IP network. This strategy provides high-performance access to IT systems at a reasonable cost. Autonomic functions will be needed to manage large, next-generation IT systems because they will have complicated management functions. Grid technology makes it possible to virtualize the components of IT systems. In this paper, we introduce our ALL-IP strategy and also discuss a 10 Gbit Ethernet switch, organic computing, and grid computing for realizing the ALL-IP system.

1. Introduction

The ALL-IP strategy is one of Fujitsu's key strategies for next-generation IT systems. In this strategy, all IT resources are connected by Ethernet; servers and storage are, of course, connected by Ethernet; and the servers' processor boards are connected by high-speed Ethernet. The broadband network makes it possible for IT resources to cooperate even when they are isolated. All IT resources are integrated through the IP network. The ALL-IP strategy enables users to obtain high performance at a reasonable cost, because Ethernet is designed for computer networks and its performance is improving rapidly. Fujitsu Laboratories will make every possible effort to realize its ALL-IP strategy.

From the viewpoint of hardware and middleware, research and development should be focused on three requirements: a 10 Gbit Ethernet switch, organic computing, and virtualization by grid computing. These are important requirements in the ALL-IP strategy, because a high-performance switch is required to make the best use of high-performance Ethernet and organic computing and a virtualization mechanism are required to keep a complex IT system in good working order. We will now briefly describe these requirements.

1) 10 Gbit Ethernet switch

Fujitsu has developed a single-chip 10 Gbit Ethernet switch as the key device for the ALL-IP strategy. This switch is mainly intended for the interconnection switch system of servers. This development project is supported by the New Energy and Industrial Technology Development Organization.

2) Organic computing

The organic server was the first server with an autonomous function that was made available to the public. Fujitsu has developed an organic server in a national project funded by the New Energy and Industrial Technology Development Organization. The word "organic" here means that the server has autonomic functions such as self-healing, metabolism, and evolution. We used the word "organic" because these are important functions of living organisms.

3) Virtualization by grid computing

Grid computing has proved to be very effective in large-scale simulations of practical situations. Grid computing is important not only in basic research for national projects, but also in practical business applications.

2. 10 Gbit Ethernet switch

In the future, we expect the following developments in computer systems:

1) Coexistence of SMP (Symmetric Multi Processing) and cluster computers.
2) Demands for low cost will accelerate the shift to cluster systems.
3) Storage systems will shift from high-cost Fibre Channel to low-cost IP networks.

There is a demand for faster IP networks, especially for interconnects. To meet this demand, we have developed a 10 Gbit Ethernet switch called AXEL, and this switch is the first single-chip 10 Gbit Ethernet switch in the world (**Figure 1**). We have also developed a switching unit for cluster systems and are developing new fields such as IP switches for storage systems and PCI Express switches.

In addition, we have developed a 10-gigabit switchboard that uses AXEL and multiple 1 to 10 Gbit aggregation boards. Experiments have shown that our 10 Gbit switch is suitable for interconnecting cluster systems such as organic servers.

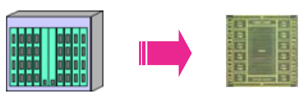
Moreover, we are developing an enhanced XAUI circuit that can interconnect cluster nodes separated by up to 20 meters.

3. Organic servers

We have been developing autonomic functions and a virtualization function for servers in a national project. In 2002, we developed a blade server with 200 blades that forms the basis of an organic server. In addition to its autonomic functions such as self-healing, metabolism, and evolution, this organic server has a dynamic load balancing function and is jointly operated by all of the blades cooperating with each other. Organic functions are important for a Web front-system and large-scale simulations such as biological simulations. We realized a 10 Gbit Ethernet interconnection for these servers in 2003 (**Figure 2**). Our next task will be to design and develop functionally differentiated blades that are more tolerant to failures than universal blades.

- Key features
  - 12-port 10 Gb/s Ethernet switch chip
  - Layer-2 switch with VLAN and QoS
  - On-chip high-throughput buffer memory
  - Integrated SERDES for XAUI
- Benefits
  - High density and low cost
  - Total aggregate throughput of 240 Gb/s
  - Low latency for cluster applications
- Enabling technologies
  - High-throughput memory for packet buffering
  - Buffer management for low latency
  - XAUI macro integration for low cost

SERDES: Serialzer Deserializer
GbE : Gigabit Ethernet



10 Gb/s SW Box　　Single-chip 10 Gb/s SW

- Focus on layer 2 and 10 GbE
- High-throughput buffer memory
- SERDES integration

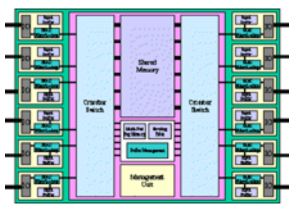


Figure 1
10 Gbit Ethernet switch chip.
World's first single-chip solution for 10 Gb/s Ethernet switches.

4. Organic storage

Fujitsu Laboratories has developed an organic storage system to absorb the massive and ever growing flow of data that is generated by human activity. The system has unlimited scalability, undergoes metabolic changes to provide data longevity, and enables quick data replication for disaster recovery and self-healing (**Figure 3**). A huge number of autonomic modules spreading over a wide-area network functions as a data grid that enables any data item to be accessed from any location and supports grid computing.

Today's storage system problems include the difficulty of ensuring expansion, limited mobility, and the need for quick maintenance when a malfunction occurs.

Regarding expansion, there is a need to replace existing large-storage systems such as SAN and NAS with bigger systems and then transfer the data to the new systems, which is a lengthy and complex task. However, organic storage systems can be expanded simply by attaching additional storage devices to the network; that is, the IP network. The additional storage devices are automatically recognized by the organic storage system, so the storage space is automatically expanded.

Regarding mobility, an organic storage system is implemented on the IP network, and all storage modules have organic control functions, so even if the required data is distributed throughout the system, it is transferred simultaneously. In a traditional RAID system, data transfer is done by a central RAID controller. Organic storage systems, on the other hand, do not have a central controller and data transfer is done simultaneously by the CPUs of each storage module.

Lastly, regarding maintenance, in a typical RAID system, for example, RAID-1, the secondary device keeps a copy of the data on the primary device, and when the primary device fails, operation is switched to the secondary device. An organic storage system also keeps a copy of each data item, but the data blocks are distributed throughout the storage system. When a storage device fails, the data block that has a copy of itself on the failed device recognizes the malfunction and creates another copy of itself in an empty stor-

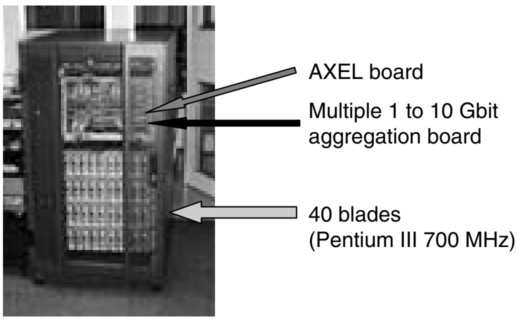


Figure 2
Blade server with AXEL.

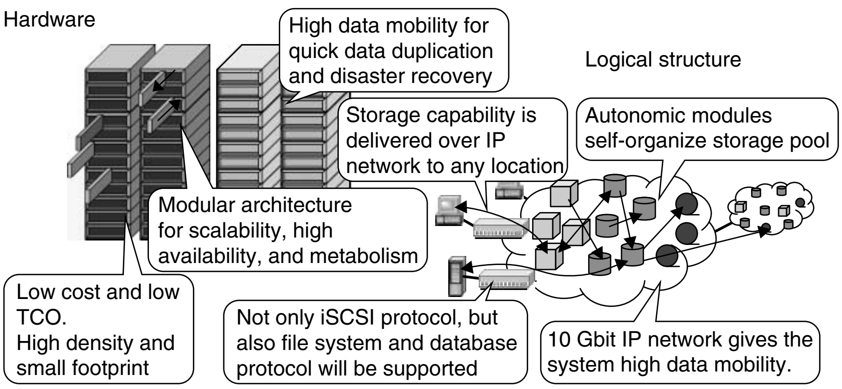


Figure 3
Organic storage system.

age space. Therefore, providing there is sufficient empty space in the system, there is no immediate need to change the failed device.

5. Organic network

We are focusing on a total system technology for organic networking that realizes adaptive resource control and automatic recovery from system problems in ISPs and enterprise network applications. This technology has the following features (**Figure 4**):

1) Robust networking
2) Basic systems for adaptive operation of organic servers and storage.

Accessibility over a network can fluctuate widely, so load balancing technology is important for IDC systems, especially in utility computing. In an organic network, there are two methods of load balancing. The first is to balance the load by predicting the amount of access and then adjusting the network's bandwidth accordingly. The second method is to look for a malfunction and then, if a malfunction is found, switch the network path that contains the malfunction to a reserved path.

6. Grid computing

Grid computing is a technology for virtualizing IT resources and realizing high-availability IT resources. The use of grid computing for scientific calculations has been researched in Europe and the US for a long time. Now, because of the expansion of Web services, grid computing technology can be applied to business applications by integrating it with Web services. Grid computing is also important for closer cooperation of IDCs.

6.1 Computational grid

We considered we should focus on massive computing to apply grid technology to enterprises. Therefore, we have developed a grid middleware called CyberGRIP (**Figure 5**) to solve the following problems:

1) The average usage rate of servers is under 30%.
2) Computer resources are only used by the working groups that own them.
3) Networks have a large pool of unused personal-computing power.

CyberGRIP determines the CPU capabilities, memory, storage, CPU load factor, and other in-

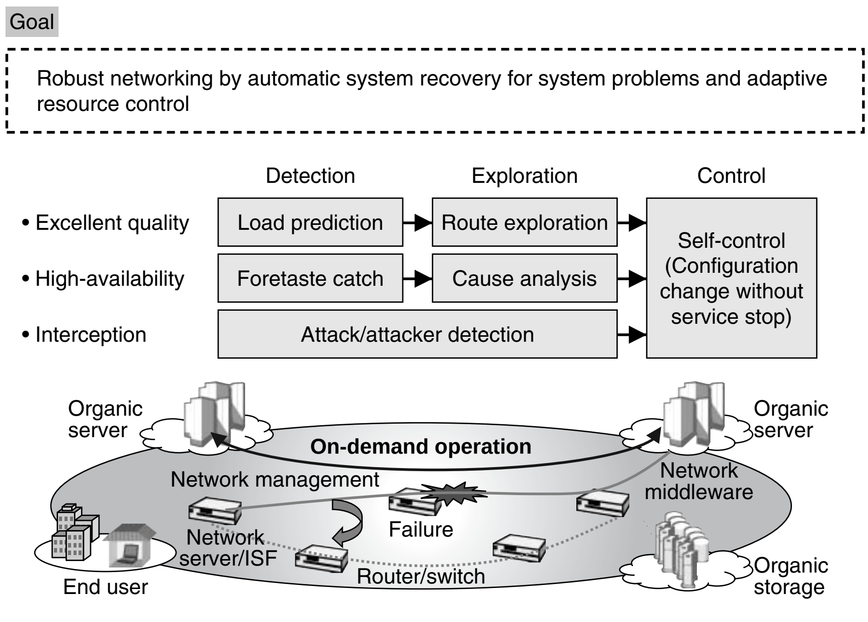


Figure 4
TRIOLE networks (organic networking).

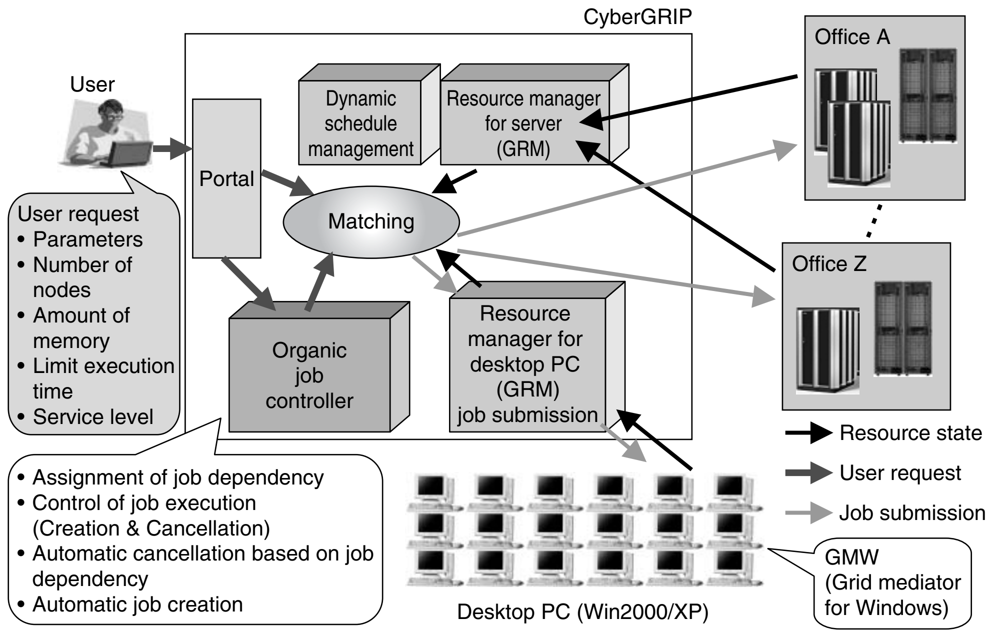


Figure 5
System configuration of CyberGRIP.

formation about the servers and PCs in a grid system. It then assigns the jobs submitted by users to a server or PC according to its ability to handle them (Figure 5).

Users define the structure of a large number of simulation jobs, then CyberGRIP uses a module called an “organic job controller” to submit jobs in the correct order and automatically set the appropriate parameters.

We tested CyberGRIP in an LSI CAD simulation that lasted 66 days. There were 4720 jobs, and the total calculation time (CPU time) was 31 000 hours. In a normal environment, a simulation of this size would take about 7 months, which means that CyberGRIP more than tripled the simulation speed.

6.2 Business Grid

The Business Grid national project was started in 2003. The mission of this project is to provide a reliable, next-generation IT infrastructure for massively distributed business computing. The function of the middleware for realizing Business Grid is to manage and effectively allocate heterogeneous resources. Business applications must be able to handle database systems and cooperate with other applications. This project supports standardization activities such as the Global Grid Forum for promoting the use and development of open-source software.

7. Utility computing

Utility computing is made possible by the high performance of broadband networks. It can be very beneficial to users because, in this type of computing, users are charged according to the amount of IT resources (e.g., bandwidth or CPU time) they use, which usually results in lower costs. In utility computing, it is important for the service providers to increase the efficiency of their networks’ IT systems, so they can reduce their charges and, therefore, their customers can reduce their IT expenses. IT vendors are expected to provide management tools for efficient utility computing.

One of the features of Fujitsu’s utility computing technology is that its response time is guaranteed by a three-tier linkage model

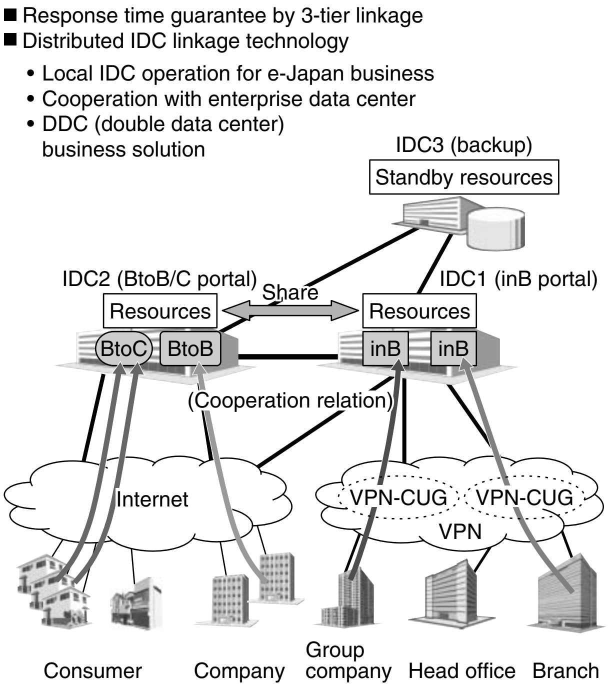


Figure 6
Fujitsu's utility computing.

(**Figure 6**). Web front-systems, applications, and data work in closer cooperation through the Web service mechanism. The utility computing system monitors the response time of Web front-systems, applications, and data access and then allocates IT resources in order to comply with the SLA (Service Level Agreement).

The other feature of this technology is it performs distributed IDC linkage. Fujitsu's IDC deployment strategy is to establish small, regional IDCs in every prefecture and large, central IDCs in several key locations. We have developed a load balancing technique that makes it possible to link regional IDCs with a central IDC. If a regional system becomes overloaded, the IDC linkage system determines the resources that are needed to process the burst, allocate servers at the central IDC, and redirect requests to the central IDC. By using this technique, central IDCs support the small IDCs in each prefecture and the SLA is guaranteed. Some large enterprises have multiple IDCs, and the IDC linkage technology can be applied to interlink them (Figure 6).

8. Conclusion

This paper discussed the ALL-IP strategy for next-generation computer systems. Broadband networks enable IT systems to cooperate with each other over the IP network. Fujitsu is focusing on developing its ALL-IP strategy. Next-generation computer systems will consist of large numbers of servers and storage systems, and these IT resources will be interconnected via high-performance Ethernet network systems. Fujitsu has developed key technologies for these next-generation computer systems, including a single-chip 10 Gbit Ethernet switch for interconnections in large PC-cluster systems. For such systems, organic functions are needed to manage the processing units.

This paper also discussed organic storage systems and the virtualization of IT resources through the use of grid computing. Autonomic functions are key requirements for organic computing, and virtualization is a key technology for next-generation IDCs. Utility computing should therefore be established based on autonomy and virtualization.

Acknowledgments

Part of this work—the work on the 10 Gbit Ethernet switch, organic server, and organic job controller—is supported by the Japanese New Energy and Industrial Technology Development Organization.

Haruo Akimoto received the B.E. and M.E. degrees in Electrical Engineering from Tokyo Institute of Technology, Tokyo, Japan in 1972 and 1974, respectively. He joined Fujitsu Laboratories Ltd., Kawasaki, Japan in 1974, where he has been engaged in research and development of character recognition, symbolic manipulation machines, artificial intelligence software, and research planning of computer systems. He is a member of the IEEE, Institute of Electronics, Information and Communication Engineers (IEICE) of Japan, and the Information Processing Society of Japan (IPSJ).