"Generation Free Platform," Technology for the Dependable Network Platform

YOSHIKAWA Takashi, KAMI Nobuharu

Abstract

Network equipments must have high availability in order to compose a node of the dependable network. We have developed a conceptual prototype of the dependable system-platform named the Generation-Free Platform. Some technologies are installed and operate together to demonstrate reliable and scalable service deployment over the generation of services and technologies. The long-distance and protocol-free connectivity of optical interconnects enable us to configure a redundant group over plural chassis. Then the automatic availability management framework allocates resources for a certain service on the redundant resources. In that process, the physical information such as location of each resources, shared risk group, and failure rate is considered. The evaluation result of a prototype system shows the proposed framework achieves no complete system down unless multiple-point failure case. This corresponds to the same level of reliability as legacy manual risk management while the resource utilization efficiency is twice as high.

Keywords

dependability, redundancy, virtualization, re-configuration, extendibility, platform

1. Introduction

The nodes that form dependable networks are generally required to achieve availability as high as 99.999%. Since it is impossible to achieve such a level of availability by using individual devices, redundant configurations such as dual redundancy are used for this purpose. The redundancy makes it possible to switch an active system to a standby system if the active system should fail, but what is important here is to properly identify the unit of switching according to its function, version changeability and failure rate, and to modularize each system accordingly.[1] On the other hand, while increasing the redundancy increases the availability, it also results in increasing wasteful resources so a careless implementation of redundancy leads to a degradation of the resource usage rates. This condition therefore makes it also important to decide how to configure a node by combining the modularized resources.

This paper focuses mainly on the latter point, i.e., the method of combining modules for building a high-availability node by referring to a prototype called the Generation Free Platform.

2. Redundant Configuration Utilizing Virtualization in Lower Layers

What is important in considering the availability is not to avoid breakdown of individual devices but to avoid an event that stops the network services. Traditionally, each service was tied to the device providing it in a 1-to-1 relationship. In this case, the percentage that one of the services stop is equal to the failure rate of the device, that is, the product of the components forming the device. One of the methods for untying such a 1:1 relationship between a service and a device is virtualization. This is achieved by installing an abstracting layer between the services and physical resources so that the services are not executed directly on the physical resources but are executed on the logically created virtual node. This virtualization technology allows the nodes for service provision to significantly improve their flexibility and extendibility.

Nevertheless, since the virtualization technology has hitherto belonged to computer technology as represented by grid computing, it cannot be applied directly to the nodes that belong to network technology. This is because the computer-oriented virtualization is packaged at the software level as seen in dividing the processes and assigning them to multiple servers. On the other hand, when data is processed without using the CPU as in the network node, virtualization at the software level is out of the question and it is necessary to invent a virtualization mechanism for the physical resources at the lower layer or in the interconnection layer that binds them together.

We therefore began a study into a device platform at the physical resource level that features lower layer re-configurability at the component level to enable virtualization. When services are allocated on this platform, a high-availability node may be created by allocating the services in consideration of their availability and resource utility. As a result, we were able to develop a prototype platform and called it the GF-PF (Generation Free Platform).

3. Optically Connected Modular Hardware Platform

The GF-PF has multi-slot hardware as shown in **Fig. 1** (the photo shows a 4-slot model). Each slot accommodates a service card, switch card, line card, etc., which are combined to implement a service. The characteristics of the GF-PF is that the unit of modularization is not only the card in the slot but that there are several grades including the "sub-module (mezzanine)" below the card and the "cabinet" above the card, and that the modules are connected equivalently via an optical interconnection system.

The optical interconnection improves the physical interchangeability by allowing even the modules with a high-speed interface above 10Gbps/line to be replaced easily by simple disconnection and connection via the optical connector. In addition, the use of broad bandwidth and the long distance capability of optical fibers make it possible to extend an internal bus seamlessly outside the cabinet. This feature enables redundancy across modules distributed in multiple cabinets so that the dependability as a system can be improved. As shown in **Fig. 2**, the optical interconnections are provided by NEC-original PETIT modules that were adopted as the model of the standardized PT model. Optical connection is enabled at any point of a board instead of the ordinary optical interconnection with which connections are always made at the edges of boards. The system contributes thus to the implementation of a modular configuration of various grades from the on-board mezzanines to the cabinets.

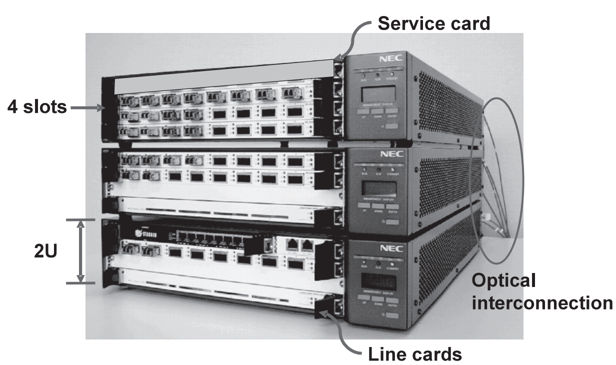

Fig. 1 Generation Free Platform.

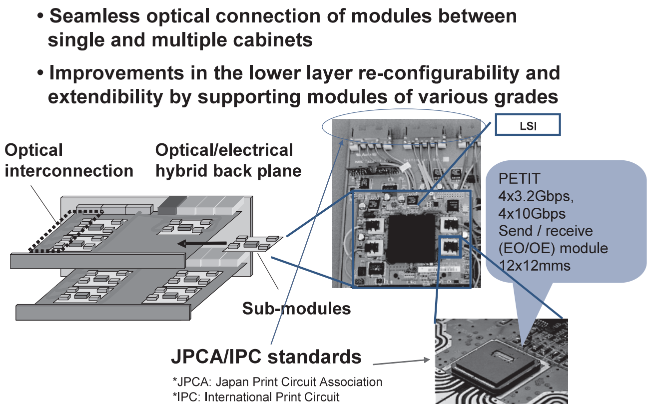

Fig. 2 Seamless connection from the board to cabinet using standardized optical modules.

4. Framework of Availability Management

Now let us consider the services packaging method for modularized physical resources from the viewpoint of availability. We become aware that "availability management can be implemented only when factors are watched in the physical layer, which is the lower level layer as shown in **Fig. 3**.

One of the information items related to the physical layer is its failure rate. With modularized hardware, the MTBF (Mean Time Between Failure) varies between modules. To form a virtual node by combining the modules and guaranteeing its availability, it is required to calculate the overall availability of the virtual node system from the availabilities of the individual modules and to effectively allocate services to them. In addition, it is also necessary to consider the notion of the SRG (Shared Risk Group) that is dependent on the physical information of module installation locations. The SRG refers to a group sharing a single risk of failure, for example a group of modules that stops operation when the power supply shuts down. Redundancy aiming at availability improvement would be meaningless if an Active/Standby configuration is set thoughtlessly for the modules belonging to the same SRG. As described above, the physical resource information is critical

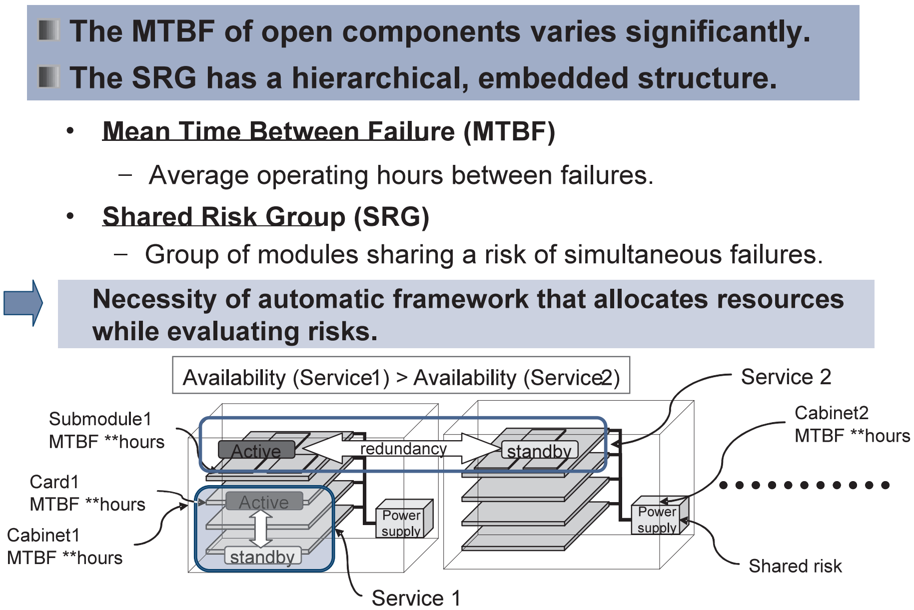

Fig. 3 Low-layer factors associated with availability.

in availability management and a mechanism for the acquisition and management of the relevant physical resource information is required in a low layer.

We therefore compiled the availability management framework as described below. It is composed of three major blocks as shown in **Fig. 4**.

1) Redundancy Management Interface

This is the interface for the manager that configures the virtual node. It can be used for settings of the amounts of resources used for each service, the type of redundancy (1+1, 1: N...) and whether or not SRGs are separated.

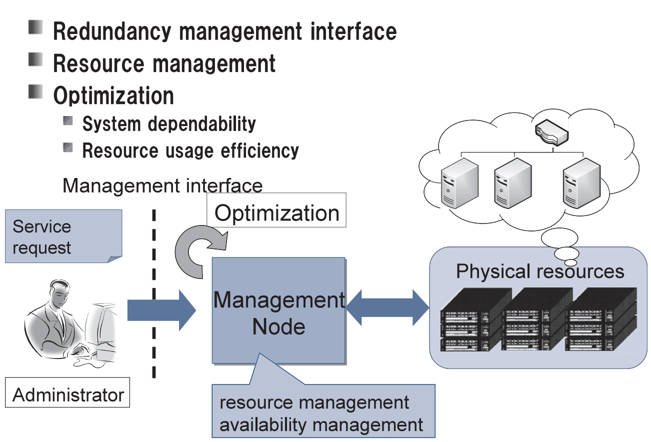

Fig. 4 Availability management framework.

2) Resource Management

This is the module for determining the physical resources that can satisfy the settings in 1) for each service based on calculations considering the resource usage rates, their MTBF and the SRG, as well as for managing these physical resources and controlling them.

3) Optimization

This function monitors the system availability and resource usage efficiencies and feeds back the information for optimization.

5. Quantitative Evaluation of Availability Management

We evaluated the effectiveness of the above-described availability management framework using the model shown in **Fig. 5**. The evaluation formula is shown in **Fig. 6**. Here, E is the expected value of how modules executing services under the virtualized module service allocation break down due to failures that are produced with failure probability p. The grades of modularization are considered in three ways: chips, cards and cabinets. A Failure is identified when both the Active and Standby nodes break down and stop the service completely due to a failure of a module. The resource manager, described above, is run based on an algorithm that optimizes the resource allocation so as to maximize the resource usage efficiencies

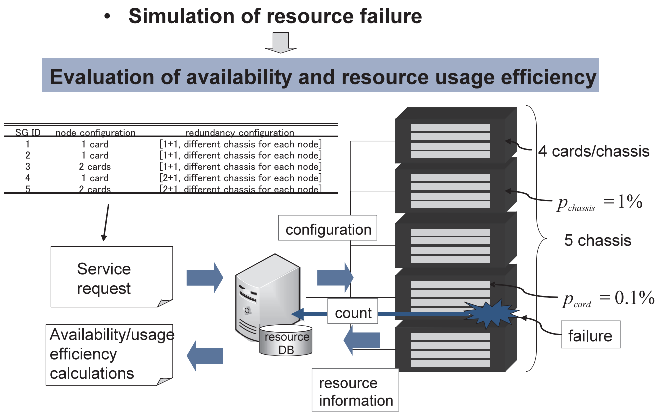

Fig. 5 Quantification model.

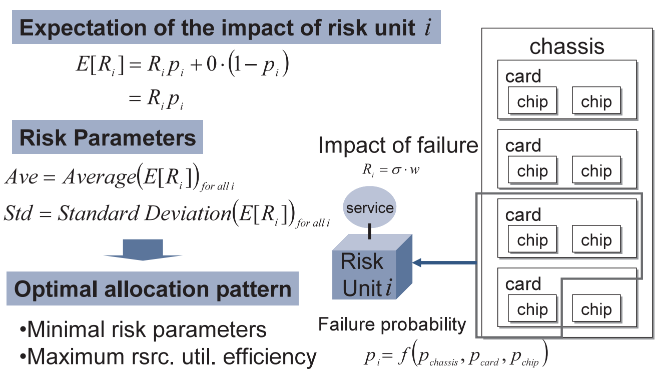

Fig. 6 Formula for the evaluation of service allocation validity of virtual modules.

while minimizing the risks. The hardware is modeled after the above-described Generation Free Platform and the virtual node is configured using a choice of five chassis each of which has four slots for the service cards. The failure probability is assumed to be 0.01% with chips, 0.1% with cards and 1% with cabinets.

We selected the models shown in **Fig. 7** as the targets for comparison with the present system described in the above.

① The present system, in which resources are allocated in the virtual node so as to minimize the failure risk and maximize the resource usage rates based on consideration of the MTBF and SRG.

② Random allocation system, in which resources are allocated randomly in the virtual node so as to maximize the resource usage efficiencies, without being concerned about the failure risk.

③ Manual allocation system, in which all of the modules are provided in dual redundancy in order to avoid the risk of failure.

Fig. 8 shows the results of simulations. “1” shows the ratio between the number of times a service was stopped because both the virtual active and virtual standby nodes broke down due to module failures and the number of times the service was

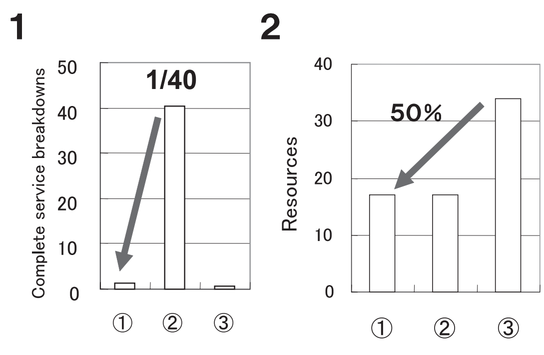

Fig. 8 Availability and resource usage rate.

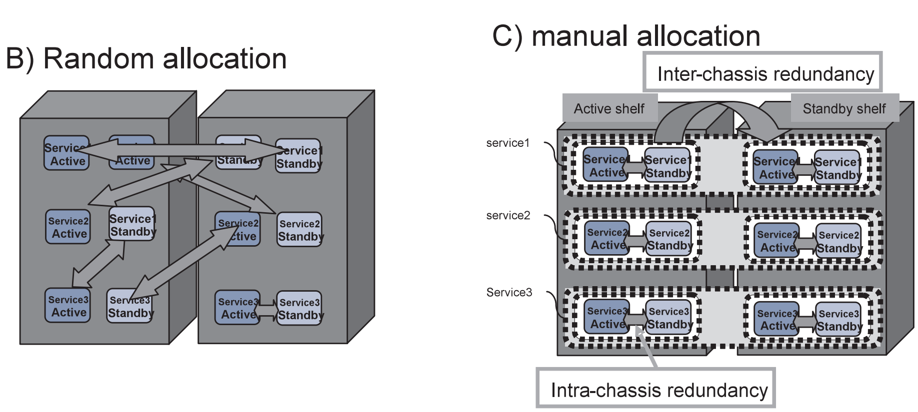

Fig. 7 Random/manual allocation system models for comparison.

continued because the node broke down but the standby node was active. With the present system, the number of times the service was stopped was equal to that of the manual allocation system in which all modules are provided with dual redundancy, and was only 1/40th that of the random allocation system thanks to the continuance of operation of the service using the virtual standby node. It was thus confirmed that the automatic resource allocation was applied very effectively. On the other hand, "2" shows the comparison of resource usage efficiencies assuming the same availability for all systems. In the case of the full-dual manual allocation system ③, the required resources are simply doubled and many wastefully unused resources are generated. Compared to this, the present system ① was able to achieve the equivalent resource usage efficiency as system ② with which services were allocated randomly. This is because the automatic resource allocation of the present system allocates the resources to maximize the resource usage rates while minimizing the breakdowns due to failures and achieves higher resource usage efficiencies than the manual allocation system ③ with which all modules are provided with dual redundancy based on human perception. In this way, the present system allows the physical characteristics of individual modules to be taken into consideration and can configure a high-availability system even when open modules are used.

6. Conclusion

In the above, we introduced the "Generation Free Platform" technology as a device configuration technology. Dependability is thus enabled by allocating services by taking into account the physical resource information of the hardware that is modularized for re-configurability to the lower layers.

References

1) C. Y. Boldwin and K. B. Clark, "Design Rules," translated by H. Ando, Toyo Keizai Inc., 2004.
2) N. Kami, et al., "Scalable and Reliable Platform for Service-oriented Networking and Computing Systems," MILCOM2004, Monterey, CA, 31 Oct., 2004.
3) T. Yoshikawa, et al., "Optical Interconnection as an IP macro of a CMOS Library," Hot Interconnects 9, 2001, Stanford, CA, 22-24 Aug. 2001.

Authors' Profiles

YOSHIKAWA Takashi
Principal Researcher,
System Platforms Research Laboratories,
Central Research Laboratories,
NEC Corporation

KAMI Nobuharu
Assistant Manager,
System Platforms Research Laboratories,
Central Research Laboratories,
NEC Corporation