RAS Technology of the SX-9

KOBAYASHI Katsumi, YONEMURA Takashi, TAKAHASHI Kiyoaki, NAKASO Hiroko

Abstract

The Supercomputer SX-9 concentrates improvement technologies that have been refined over NEC's long years of experience in order to meet the requirements of high reliability and to provide support for the circuitry to the system levels. Reliability, Availability and Serviceability are thus attained at more advanced levels.

This paper introduces the RAS technology based on the latest innovations of the SX-9.

Keywords

RAS, Reliability, Availability, Serviceability, failure detection, auto recovery reconfiguration, maintenance diagnostics

1. Introduction

What is important in achieving a high degree of system availability is first of all, infrequency of failures, secondly optimum fault processing in the case of failure to maintain system availability, and thirdly prompt servicing to enable quick system recovery.

The RAS concept integrates such reliability improvement technologies and represents them by the initials of the three factors of Reliability, Availability and Serviceability.

The basis of the RAS technology is its infrequency of systems failures. This makes it important to improve the inherent reliability of each system component and to reduce the number of components. The Supercomputer SX-9 implements the CPU (Central Processing Unit) and the RCU (Remote Access Control Unit) on a single chip to increase the integration scale further than that of the previous SX-8. At the same time, the SX-9 has reduced the number of parts in order to achieve high reliability and to additionally provide the units and systems with countermeasures against eventual failures. The SX-9 has an auto recovery function that detects errors caused by failure and corrects the detected error or retries the failed operation. If auto recovery fails, the failure point is isolated and switched to an alternative system to continue the system operation, if this is available. If the system operation cannot be continued, the SX-9 restarts the system and its operation to offer improved availability.

The SX-9 also features a high maintenance diagnosis technology that identifies the failure point to enable quick recovery. The maintenance functions are integrated into the SVP (Service Processor) to improve serviceability by using maintenance tools and powerful information collection and implement integrated servicing procedures by means of remote maintenance.

The SX Series adopt the RAS technology cultivated via the Parallel ACOS Series of enterprise servers by optimizing its performance to support supercomputers. **Fig.** shows the outline of the RAS technology of the SX-9, which is a further advancement of the RAS technology used with the traditional SX Series products.

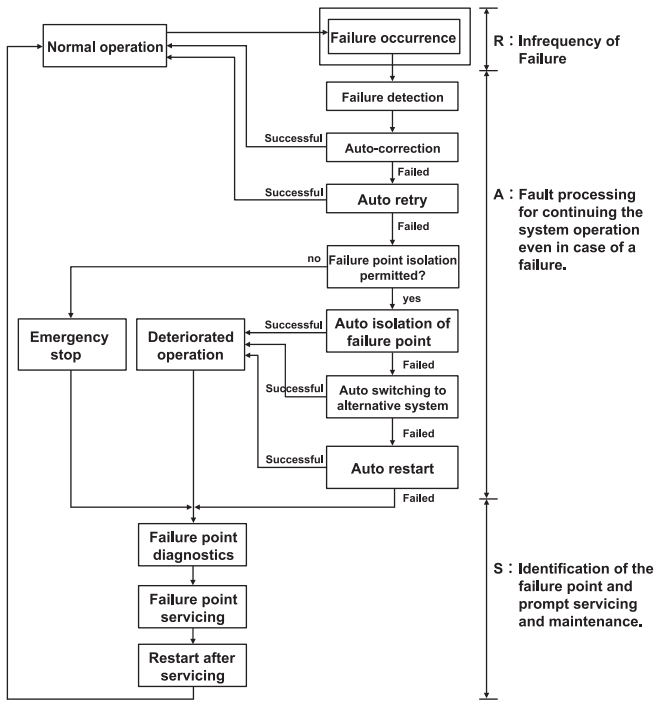

Fig. Failure recovery processing.

In separate sections in the following, we will describe the failure detection, auto recovery, re-configuration and maintenance diagnostic technologies provided by the SX-9.

2. Failure Detection

In order to detect errors, the SX-9 places the error detection circuits such as the parity check circuitry optimally at appropriate points in the system component units as well as in their internal circuitries. It also monitors the time taken and replies of the operations performed inside the units.

It incorporates the functions for the dependable detection of errors and for the prevention of their propagation in order to improve the error correction retrial rates. This issue is discussed in detail below.

In addition, the RAS processor that deals with system startup and stoppage and fault processing holds periodical communications to enable quick error detection even during system operation. This procedure minimizes the effects of errors on the system by preventing error propagation and by localizing errors in terms of both space and time.

3. Auto Recovery

The auto recovery technology can be classified into the technology for correcting errors using redundant hardware and that for retrying erroneous operations by aiming at correction using redundancy in terms of time.

3.1 Error Correction

The SX-9 uses mainly the MMU (Main Memory Unit) for the correction of 1-bit errors and the detection of 2-bit or longer errors are dealt with by means of S8EC (Single 8-bit Error Correction). For errors between units such as the CPU and MMU and those between nodes of a multi-node system, the SX-9 performs auto correction of 1-byte errors as well as the detection of 2-byte or greater errors.

3.2 Error Retry

Failures may be roughly classified into fixed errors that occur constantly and intermittent errors that occur from time to time. The ratio of intermittent failures has been increasing following advancements in speed and the integration of technologies. However errors that occur momentarily due to external disturbances, etc. (intermittent failure) can be corrected and the processing can be continued by retrying the operation affected by the error.

When an intermittent fault related to the CPU occurs, the CPU is initialized and included in the system again to resume the operation.

When a software error of RAM occurs, it is recovered by means of a correction with ECC or by the rewriting of the error word.

When an I/O processing-related error occurs, the OS (Operating System) retries the I/O instruction. If a retry fails, the OS switches the I/O path or switches the I/O system to an alternative system to retry by avoiding the error.

With a multi-node system, the communication between nodes is retried to attempt auto recovery from the error.

As a result, even in the case of a failure occurring, the processing can be continued and high availability maintained with little effect resulting from the error.

4. Reconfiguration

When a permanent failure (fixed fault) occurs and it cannot be recovered automatically and if the system adopts the redundancy configuration, the failed unit is identified and the system operation is continued in the degenerated mode.

When a fault occurs with one of the units in the main system (CPU, MMU and I/O units), the faulty unit is basically isolated to improve the system availability. When a fault is detected at the system startup, it is retried. However if auto recovery fails the operation is continued in the degenerated mode and only normal hardware is handed to the OS.

When a fixed fault occurs on a path to a peripheral processing device under the I/O system, the faulty path is degenerated and operation switches to an alternative path for continuing the system operation, thereby enhancing the failure resistance of the I/O processing.

Power supply redundancy is supported optionally. In case of a failure, only one power supply system is run to continue the system operation and improve the availability.

When a node fault occurs in a multi-node system, the faulty node is isolated. If a fault in the IXS (Internode Crossbar Switch) occurs, the path to the IXS from the ports of the RCUs of the nodes is isolated according to the extent of the fault. This is in order to minimize deterioration of the performance and to prevent the cluster outage (each RCU has a maximum of 2 ports

per lane, and the isolation of the path to the IXS is performed on a per-port basis). When it is required to isolate a lane due to the system configuration or the fault extent, the system is restarted automatically after the lane isolation in order to shorten the time until the restart of the operation.

5. Fault Processing Customization

Some supercomputer users do not tolerate even a slight deterioration in performance. To respond to these user needs, the SX-9 provides a fault processing customization function, that allows the user to select the priority between degeneracy and servicing of faulty points in case of a fault with a unit in the main system. This option makes it possible to stop the system when a fault occurs without degenerating the fault and to proceed immediately to servicing.

6. Maintenance Diagnostics

In general, when a failure occurs with a unit or the system becomes unavailable, it should be serviced promptly to recover normal operation. For this purpose, the SX-9 permits collection and analysis of fault information concurrently with the OS operation. Additionally, the remote maintenance capability enables total maintenance by requesting support to the maintenance technology expert at the maintenance center as required. In order to provide an even higher level of serviceability, the SX-9 introduces the integrated SVP to integrate the maintenance functions, improve the maintenance operation and shorten the time taken for maintenance.

6.1 Information Collection

The SX-9 logs the errors in the units in the main system and power supply faults, and the SVP is in charge of the integrated management of the error logs. The hardware is equipped with a hardware tracer function that collects the operational history of each unit until the fault occurrence in order to permit detailed tracing of hardware operations. The SVP operation history is also collected to facilitate the system-level analysis of the causal relations of fault occurrences.

6.2 Diagnostics

Identification of a failure point is essential for the recovery of a faulty unit but advancement in the integration scale of the technologies used has been increasing the share of intermittent failures. For this reason, the SX-9 adopts BID (Built-in Diagnostics), which indicate the failure point automatically and immediately uses the error log information at the moment the error is detected for the first time. In addition to the display of the order of suspect faults, the SX-9 identifies whether the detected failure is generated inside the local unit or propagated from another unit by using its enhanced error log analysis capability. It then displays the name of the unit to be serviced in order to enable quick and accurate servicing. It also indicates the failed LSI to enable effective servicing at the LSI level.

With a multi-node system, the SX-9 verifies the path connection between the IXS and single nodes by means of analysis and indicates the individual names of the error paths. This procedure enables effective servicing without oversight and thereby shortens the maintenance time.

6.3 Non–stop Maintenance

With a system in which the power supply is optionally duplicated, if one of the two power supplies fails, it can be serviced and returned into the system without stopping the system operation.

With a multi-node system, when the IXS fails and the units composing the IXS are isolated, the isolated units can be serviced individually.

If the paths from the lanes of the nodes are isolated on a per-port basis after servicing, the isolated paths are incorporated via the nodes without stopping the OS operation to restore the original configuration and performance. If it is the lanes that are isolated, the original performance can be restored by restarting the system at the timing permitted by the user.

6.4 Remote Maintenance

The direct connection of a telephone circuit to the maintenance center that holds a wide range of information for dealing with various faults ensures advanced and quick maintenance by experts. When a fault occurs in the system, the auto notification function notifies the maintenance center of the fault occurrences according to the level of the fault and sends the necessary fault information to the center.

In addition, in order to respond to strong recent requirements for computer systems security, the security of remote maintenance is enhanced by a callback function when the cir-

cuit is connected and by a switch that the user can control for permitting the circuit connection.

6.5 System Extension

The SX-9 permits upgrading from a single-node model to multi-node models as well as the expansion of new peripheral hardware that will be supported additionally according to user needs. Previously, the expansion of equipment and the extension of system scale used to necessitate long hours of system stoppage or recompilation of system configuration information. The SX-9 features a means for facilitating addition/deletion of system configuration information following equipment expansion or system scale extension. This function shortens the system stoppage time during equipment expansion.

7. Conclusion

In the above, we described the RAS functions of the SX-9. We are confident that these functions will be adequate to meet customer requirements for first-rate system reliability.

It is our stated intension for the future to positively incorporate user opinion in the pursuit of reliability for our products so that we may confidently offer dependable systems.

Reference

1) Kobayashi, K. et al.: “RAS Technology for SX-8,” NEC GIHO, Vol. 58, No.4, pp.33-36, 2005-7.
http://www.nec.co.jp/techrep/ja/journal/g05/n04/g050409.html

Authors' Profiles

KOBAYASHI Katsumi
Engineering Manager,
Computers Division,
1st Computers Operations Unit,
NEC Corporation

YONEMURA Takashi
Assistant Manager,
Computers Division,
1st Computers Operations Unit,
NEC Corporation

TAKAHASHI Kiyoaki
Engineering Manager,
3rd Solution Business Division,
NEC Software Hokuriku, Ltd.

NAKASO Hiroko
Engineering Manager,
Computers Division,
1st Computers Operations Unit,
NEC Corporation