〈SX-8のハードウェア〉

SX-8のRAS技術

RAS Technology for SX-8

小林勝美* 竹村功夫** 中曽浩子*
Katsumi Kobayashi Isao Takemura Hiroko Nakaso

要　旨

SX-8は高信頼化の要求に応えるため，回路レベルからシステムレベルに至るまで，長年培った高信頼化技術を結集し，より高度なレベルで，信頼性（Reliability），可用性（Availability），保守性（Serviceability）を実現しています。

本稿では，最新の技術を駆使したSX-8のRAS技術について紹介します。

The SX-8 adopts the state-of-the-art hardware architecture and technologies in order to cope with the requirement to much higher availability of computer systems and realizes RAS (Reliability, Availability and Serviceability) feature on a high level.

This paper describes the RAS technologies of the SX-8.

1．まえがき

システムの高い可用性を実現するためには，まず故障が少ないこと，次に故障しても適切な障害処理によりシステムの稼働を保つこと，さらに故障を速やかに修復してシステムの迅速な復旧を図ることが重要です。

RASという概念は，こうした高信頼化技術を総合的に捉えたもので，Reliability，Availability，Serviceabilityの頭文字で表しています。

RAS技術の基本は，まず，装置やシステムが故障しにくいことです。装置を構成する部品の固有信頼度を上げ，部品点数を削減することが重要です。スーパーコンピュータSX-8では，前機種SX-6からさらに集積度を高めたLSIチップを採用するとともに，使用部品数および部品間の配線数の削減を図り，高い信頼性を実現した上で，万一の故障に備えて装置やシステムで対策を講じています。故障により生じた誤りを検出し，次に検出した誤りを訂正，または，再試行を行うことにより自動回復を試みます。自動回復に失敗した場合は，故障部分を切り離し，代替装置があれば切り替えを行い，システムの運用を続行します。システムの運用が続行できない場合は再始動でシステムの運用を再開し，可用性を向上しています。

また，高い保守診断技術により，故障箇所を指摘し，速やかな修復を可能としています。保守機能はサービスプロセッサ（Service Processor：SVP）に一元化しており，保守ツールや強力な情報収集方式の採用による保守性の向上，リモート保守による統合的な保守の実現を図っています。

SXシリーズでは，エンタープライズサーバであるパラレルACOSシリーズで培ったRAS技術をスーパーコンピュータ向けに最適化しています。従来のSXシリーズのRAS技術をさらに発展させたSX-8のRAS技術の概要を**図**に示します。

また，以下にSX-8で実施している故障検出，自動回復，再構成，保守診断の各技術について順に紹介します。

2．故障検出

SX-8では，パリティチェックなどの誤り検出用回路を，システムを構成する装置内の随所に，また装置内の回路に最適になるように配置し，誤りを検出します。また，装置内動作の時間やリプライを監視しています。

確実に誤りを検出するとともに，誤りの伝搬を防ぐ機能を備え，後述の誤り訂正率や再試行率を高めています。

さらに，運用継続型マルチノードシステムにおいては，待機系IXS，および，IXSとシングルノードの接続経路を任意の間隔で診断することにより，IXSとノードの故障の速やかな検出を可能にしています。

誤りの伝搬を防ぎ，空間的にも時間的にも誤りを局所化することにより，誤りがシステムへ及ぼす影響を小さくしています。

3．自動回復

自動回復は誤りを冗長化したハードウェアにより訂正する技術と，時間的に冗長化する再試行の技術とに分類されます。

＊ コンピュータ事業部
Computers Division

＊＊ NECソフトウェア北陸　第三ソリューション事業部
NEC Software Hokuriku, Ltd.

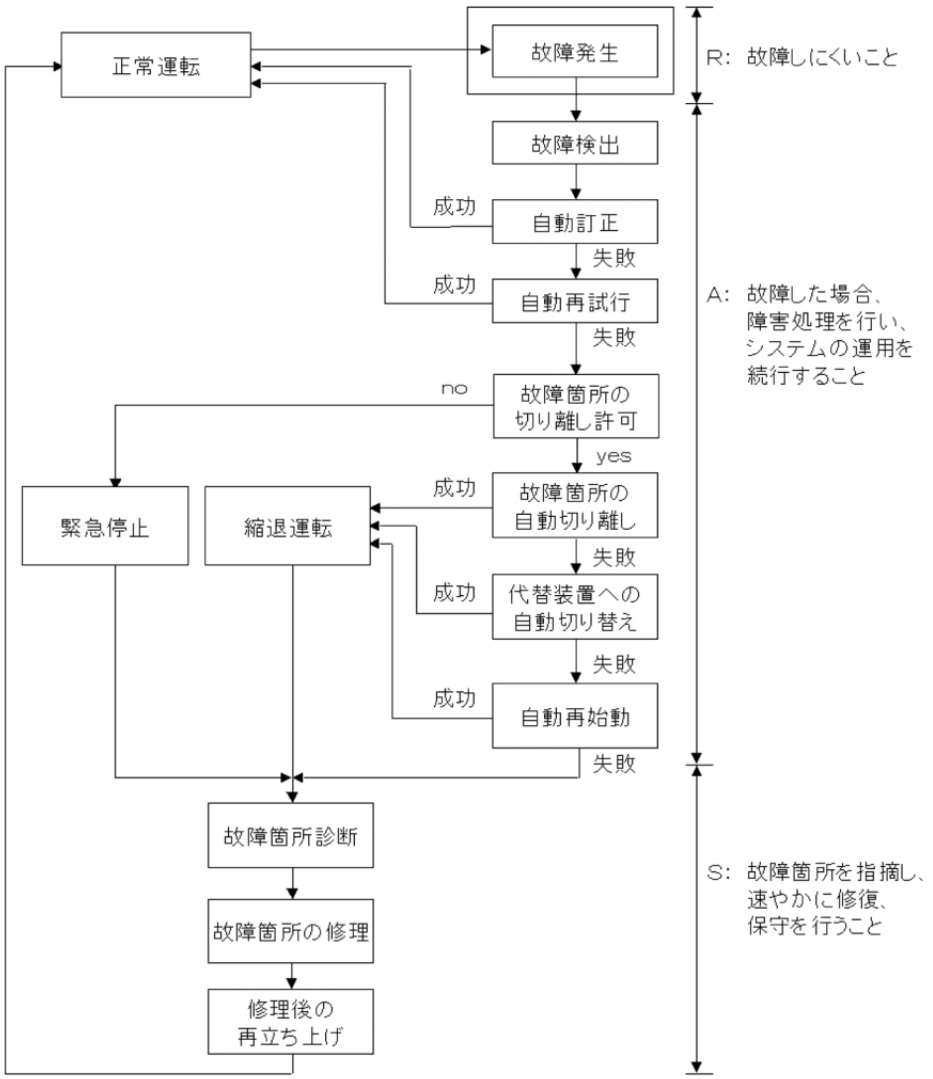

図　故障修復処理
Fig.　Failure recovery flow.

3.1　誤り訂正

SX-8では，主記憶装置（Main Memory Unit：MMU）において1ビットエラー訂正，2ビット以上のエラー検出を行っていますが，これにはSEC/DED（Single bit Error Correction/Double bit Error Detection）とS8EC（Single 8 bit Error Correction）の2通りの手法を採用しており，主記憶容量に応じて，誤りの検出と訂正が最適に行える方を選択して使用しています。また，CPU，MMUなどの装置間や，マルチノードシステムにおけるノード間のデータ転送では，1バイトエラー訂正，2バイト以上のエラー検出を行います。

3.2　誤り再試行

故障は大別して，固定的に発生する固定故障と間欠的に発生する間欠故障とに分類されます。テクノロジの高速化，高集積化に伴い，間欠故障の比率が高まっていますが，外乱などによる瞬時的な故障（間欠故障）であれば，影響を受けた動作のやり直しにより誤りを除去し，処理を続行できます。

CPUにかかわる間欠障害については，CPUを初期化し，再度システムに組み込んで運用を継続します。

また，RAMのソフトエラーについては，ECCによる訂正，エラーワードの再書き込みなどにより救済しています。

入出力処理にかかわる誤りに対しては，オペレーティングシステム（OS）により入出力命令が再試行されます。再試行が不成功に終わった場合は，OSが入出力経路を切り替えたり，代替装置へ切り替えたりして，誤りを回避し再試行します。

マルチノードシステムでは，ノード間の通信を再試行することにより，誤りからの自動回復を試みます。

万一故障しても再試行に成功すると誤りの影響をほとんど受けずに処理を継続でき，高い稼働率を維持できます。

4．再構成

故障が恒久的なもの（固定障害）で自動回復できなかっ

た場合，冗長化構成を採っている装置では，故障した装置単位に切り離し，縮退した形でシステムの運用を継続します。

本体系装置（CPU，MMU，入出力装置）は，基本的に障害が発生した装置単位に切り離し，システムの稼働率を高めています。また，システム立ち上げ時に障害を検出した場合にも，再試行を試み，自動回復できなかった場合は縮退し，正常なハードウェアのみをOSに引き渡すようにしています。

入出力装置配下の周辺処理装置へのパス上に固定障害が発生した場合は，障害パスを縮退し，代替パスに切り替えてシステムとしての処理を継続することにより，入出力処理の耐故障性を高めています。

電源は，オプションで二重化をサポートし，故障時には片側の電源のみでシステムの運用を継続することにより，可用性を向上しています。

マルチノードシステムでは，お客様の運用形態に合わせて2つのシステム構成を用意しています。現用系と待機系の2系統のIXSを備える運用継続型と，諸元を半分にして運用を再開する性能重視型のいずれかの構成を採用することにより，クラスタダウンを防いでいます。

5．障害処理のカスタマイズ

スーパーコンピュータの場合，わずかな性能低下も望まないお客様もいます。SX-8では，このようなユーザニーズに応えるために，本体系装置ごとに障害箇所の縮退あるいは修理のどちらを優先するかを選択でき，固定障害が発生した場合，縮退を行わずに即座にシステムを停止し，速やかに修理に取りかかれるように障害処理をカスタマイズする機能を有しています。

6．保守診断

一般に，故障が発生した装置やシステムが使用不可となった場合には，速やかに修復して正常に稼働させる必要があります。そのため，OS運用と並行して障害情報収集と解析を行えるようにしています。また，リモート保守の採用により，必要に応じて，保守センタの保守技術専門家に支援を求めて総合的な保守が行えます。さらに高度な保守性を提供するために，SX-8では統合SVPを導入し，保守機能を一元化するとともに保守操作性を向上し，保守時間の短縮を図っています。

6.1 情報収集

SX-8では，本体系装置および電源障害のエラーログ収集を行っており，これをSVPによって一元管理しています。ハードウェアにはハードウェアトレーサを備え，障害発生時点までの装置内の動作履歴情報を収集して，ハードウェアの動作が詳細にトレースできるようにしています。また，SVPのオペレーション履歴も収集しており，障害発生の因果関係をシステムレベルで分析することも可能にしています。

6.2 診　断

障害装置の復旧には，故障個所を指摘する必要がありますが，使用されるテクノロジが超高集積化されるにつれ，発生する故障も間欠故障が多くなってきています。そのため，最初に誤りを検出した時点のエラーログ情報を用いて，故障箇所を自動的に即座に指摘するビルトイン診断（Built-In Diagnostics：BID）方式を採用しています。被擬優先順位に従った表示に加えて，SX-8ではさらにエラーログの分析能力を強化し，検出した故障が自装置内で発生したものか，他の装置から伝搬したものかを切り分けた上で，修理すべき装置名を表示し，迅速かつ的確な保守を可能としています。また，故障LSIを指摘しており，LSIレベルでの確実な修理が行えます。

マルチノードシステムでは，IXSとシングルノードのパス接続を診断で検証し，誤っている場合は個別のパスを指摘することにより，修理ミスを見逃すことなく確実な修理を行い，保守時間を短縮しています。

6.3 無停止保守

電源をオプションで二重化したシステムでは，一方の電源が故障した時にシステムの稼働を継続したまま，電源の修理および修理後の組み込みが可能です。

マルチノードシステムにおいては，ノード間のデータ転送で故障が発生し，交換対象の装置が1バイトエラー訂正可能な範囲に収まる場合，クラスタ運用中のシステムに擾乱を与えることなく，IXSを構成する装置を個別に修理することができます。

さらに，マルチノードシステムではシステム構成によっては，クラスタ運用中のシステムに擾乱を与えることなく，IXSを修理できます。修理後，運用継続型ではそのまま待機系として組み込み，元の冗長構成に復旧します。性能重視型では，お客様の許すタイミングでシステムを再始動することにより，元の性能への復旧が可能です。

6.4 リモート保守

様々な障害対応に豊富な情報を有する保守センタと電話回線で直結することにより，専門家による高度で迅速な保守を実施します。また，システムで障害が発生したときに，その障害のレベルに応じて自動的に障害発生を通報するとともに，必要な障害情報を保守センタに送る自動通報機能を有しています。

さらに昨今のコンピュータシステムのセキュリティに対する強い要求に応えるために，回線接続時のコールバックや，お客様自らの操作による回線の接続許可を与えるスイッチを設け，リモート保守を行う際のセキュリティの強化を図っています。

6.5 システムの拡張

SX-8はシングルノードモデルからマルチノードモデルへのアップグレードや，ユーザニーズに応え追加サポートされる新しい周辺系ハードウェアの増設が可能です。これまで，機器の増設，あるいはシステム規模の拡張を行う場合

は，システムを長時間停止してSVP内にあるシステム構成情報の作り直しが必要でした。SX-8では機器の増設やシステム規模を拡張する場合に，システム構成情報の追加/削除を容易にする手段を備え，機器増設時のシステム停止時間を短くしています。

7．むすび

以上，SX-8で備えているRAS機能について説明しました。これらはお客様各位の高信頼度システムの要望に十分に応えられるものであると確信しています。

今後もさらにお客様の意見を積極的に取り入れ，よりいっそう信頼性を追求した確かなシステムを提供していきたいと考えています。

参考文献

1） 三箇山ほか；「SX-6シリーズのRAS技術」，NEC技報，Vol. 55, No. 9，pp. 31～33，2002-9.

筆者紹介

Katsumi Kobayashi
小林　勝美　1985年，NEC入社。現在，第一コンピュータ事業本部コンピュータ事業部システム技術部技術エキスパート。

Isao Takemura
竹村　功夫　1984年，NECソフトウェア北陸入社。現在，第三ソリューション事業部技術マネージャー。

Hiroko Nakaso
中曽　浩子　1986年，NEC入社。現在，第一コンピュータ事業本部コンピュータ事業部システム技術部技術エキスパート。