ＰＣサーバ向け「ハイブリッドＲＡＩＤ」の開発（５）
2Ｆ－１１　～階層ＰＣＩバス構造マルチＳＣＳＩ方式～

八木沢 育哉，大枝 高，松並 直人，兼田 泰典，荒川 敬史
（株）日立製作所　システム開発研究所

１　はじめに

近年、PCの高性能化・低価格化に伴い、PCサーバの市場が立ち上がりつつある。一方、PCサーバの高性能化に対し、磁気ディスク装置の性能向上は緩やかであり、それに起因する性能ボトルネックが顕在化してきている。この解決策として、複数の磁気ディスク装置の並列動作と、冗長データであるパリティの付加により、高性能・高信頼を実現するRAIDが開発され、PCサーバに標準で搭載される事例も増えてきている[1]。しかし、RAID制御には専用のハードウェアが必要なため、低価格なPCサーバと比較して相対的に高価なものとなっている。また、PCサーバとのインターフェースであるSCSIバスが性能ボトルネックとなりRAIDとしての転送能力が制限されるという問題がある。

このような動向を踏まえ、主なRAID制御をソフトウェアで行ない、アクセラレータハードと併用することで低価格・高性能を実現する「ハイブリッドRAID」を開発した。本報告で述べるアクセラレータハードは、従来RAIDに比べ簡単な構成であり、PCサーバインターフェースのPCIバス化とマルチSCSIの実装により、低価格化・高性能化を図っている。

２　ＰＣサーバ向けＲＡＩＤの課題

RAIDは、磁気ディスク装置の並列動作と冗長構成により高性能・高信頼を実現する一方、主な制御であるアドレス変換とパリティ生成を行うための専用プロセッサやキャッシュ等が必要となり、低価格なPCサーバに適用するには、相対的に高価なものになってしまうという問題点がある。このため、RAIDをPCサーバに適用するには、低価格化が必須であり、RAIDの制御方式等を工夫することで、いかに効率よく部品点数を減らすかが重要となる。

また、RAIDは、PCサーバとのインターフェースにSCSIバスを用いているため、１つのSCSIデバイスとして取り扱うことができる反面、RAIDとしての転送性能がSCSIバスの最高転送性能以内に抑えられてしまうという欠点があり、SCSI処理のオーバヘッドも性能を低下させる原因となっている。したがって、SCSIバス性能以上の高速転送を実現するためには、PCサーバとのインターフェースを見直す必要がある。

３　ハイブリッドＲＡＩＤ

上記課題を解決するため、専用のハードウェアが行なっていたRAID制御の一部をソフトウェアで行い、構成の簡単なアクセラレータハードと併用する

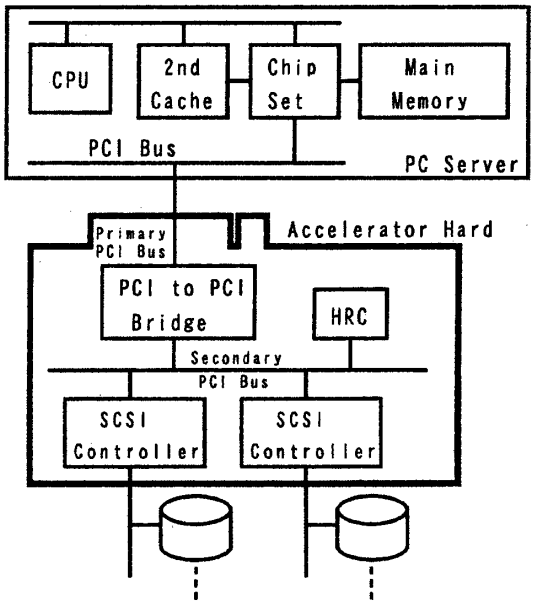

図１　ハイブリッドＲＡＩＤハードウェア構成

Development of 'Hybrid RAID' for PC Servers(5)
-Method of multi-SCSI with hierarchical PCI-
I. Yagisawa, T. Oeda, N. Matsunami, Y. Kaneda and H. Arakawa
Hitachi, Ltd. Systems Development Laboratory

RAID: Redundant Arrays of Inexpensive Disks
SCSI: Small Computer System Interface
PCI: Peripheral Component Interconnect
HRC: Hybrid RAID Controller (for parity)

ことで低価格化と高性能化を両立するハイブリッドRAIDを考案した。

ハイリッドRAIDは、低価格化を図るために、主なRAID制御をPCサーバ上のOSのデバイスドライバで行うことで、RAID制御専用のプロセッサを削除した。さらに、後述のPCIバス直結のメリットを生かし、PCサーバの主記憶をRAID制御用キャッシュの一部として使用することで専用部品を減らした。

また、SCSIバス性能を超える高速転送を実現するために、PCサーバへのRAIDのインターフェースをPCサーバ内部バスであるPCIバスとし、アクセラレータハードをPCIバス直結にした。さらに、アクセラレータハード上にマルチSCSIを実装することで、SCSIの性能ボトルネックを解消した。

今回、アクセラレータハードに上記のPCIバス直結とマルチSCSIを実装するために、次のような「階層PCIバス構造マルチSCSI方式」を採用した。

4　階層ＰＣＩバス構造マルチＳＣＳＩ方式

アクセラレータハードは、PCI拡張ボードとしてPCサーバのマザーボードに接続するが、PCI拡張ボード上にマルチSCSIを実装するには、次の課題があった。１つのPCI拡張スロットに接続できるPCIデバイスは１個に限定される、すなわち、拡張ボードのコネクタに直接接続できるボード上のPCIデバイスは１個というPCI規格の順守である。

この課題を解決するためには、マルチSCSIを一つのPCIデバイスに束ねる必要があるが、拡張スロットと複数のSCSIコントローラ間で高速なデータ転送を行うには、束ねる際の遅延時間の少ない方式が要求される。PCIインターフェースを持つ独自のバスコントローラを用いて、拡張ボード内部に独自のバスを配線し、このバスに複数の汎用SCSIコントローラを接続する方式もあるが、PCIバスを拡張ボード内部のバスとして使用した場合、束ねる際の遅延時間の影響はあるものの十分な転送性能を得られるので、PCI to PCIブリッジを用いた階層PCIバス構造を採用した。

この構造を用い、拡張スロット部のプライマリPCIバスを PCI to PCIブリッジを介して拡張ボード内に配線したセカンダリPCIバスに連結し、このセカンダリPCIバス上にPCIインターフェースを持つ複数のSCSIコントローラを接続した。

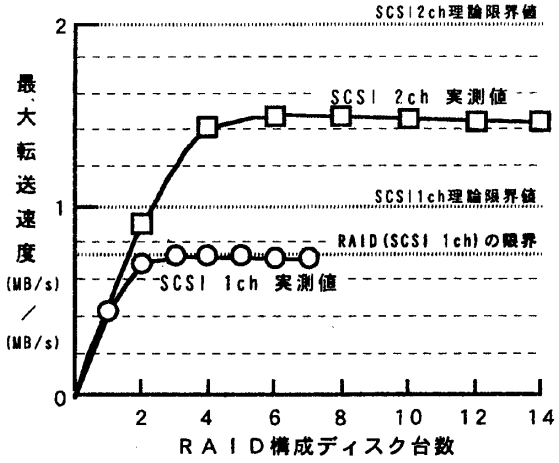

図２　シーケンシャルアクセス性能(64kByteリード)

以上のような構成を採ることで、SCSIチャンネル数に比例する転送性能を得ることができ、SCSI1chの性能によって制限されていた従来RAIDの限界転送性能を超えることができた(図２)。

5　まとめ

ソフトウェアによるRAID制御と、PCサーバ主記憶のキャッシュ化によりRAID制御専用のハードウェアを削減できた。この結果、従来RAIDの約1/3の価格でPCサーバ向けRAIDを実現した。

また、階層PCIバス構造マルチSCSI方式を採用したアクセラレータハードにより、PCIバス直結のインターフェースとマルチSCSIを実現し、従来RAIDの限界転送性能を超えることができた。

参考文献

[1] David A. Patterson, Garth Gibson, and Randy H. Karz: " A Case for Redundant Arrays of Inexpensive Disks (RAID)", Report no. UCB/CSD 87/391, Computer Science Division Department of Electrical Engineering and Computer Science, University of California, Berkeley, 1987.