

Infiniband

Infiniband

～ クラスタサーバの標準搭載である Infiniband ～

中尾 昌広, 児玉 憲造

Masahiro NAKAO, Kenzo KODAMA

Abstract: This paper shows the advantage and architecture of Infiniband which is a new standard replaced PCI. After several years all server machines are going to have Infiniband on board in order to adapt for the advanced network society.

1 はじめに

サーバの CPU と、ネットワークカードなどの周辺機器との間の I/O のスピード差は、ここ数年、課題とされてきた。この問題を解決するため、これまでも PCI-X などの規格が持ち上がったが、最も有力視されているのが Infiniband である。

現在の主流である PCI とは異なり、Infiniband では他の周辺機器との接続を共有せず、ポイント間での直接接続が実現することになる。また高速化以外にも、拡張性や信頼性の向上といったメリットがあるため非常に注目されている。

2 PCIの限界

2.1 PCIとは

現在の主流バス¹である PCI とは、米国 Intel 社などのハードウェアメーカーで組織される、標準化団体 PCI SIG が策定したローカルバス規格のことである。PCI バスは、32bit 幅、最大 133Mbps²のデータ転送が可能で、また、特定の CPU を選ばないこと、既存の多くのバスと共存可能であったことから、急速に標準化した。Fig. 1 に PCI を用いた入出力システムを示す。²⁾

2.2 PCIの問題点

PCI バスは長年にわたって標準規格として使われてきた。しかし、今日の要求であるサーバシステムの性能や RAS³の向上、スケーラブルなシステム構成など、大規模なシステムを構築する上でいろいろな問題点が出てきた。

現在のアーキテクチャでは、次のような問題が存在している。

- PCI では、それぞれの I/O コントローラがバスおよびアドレス空間を共有するため、局所的な障害が

¹CPU と各装置間の通信路はバスという共有されるデータの転送路を使って行われる。¹⁾

²64bit 幅、最大 533Mbps の規格もある

³Reliability Availability Serviceability(それぞれ信頼性、可用性、保守性)

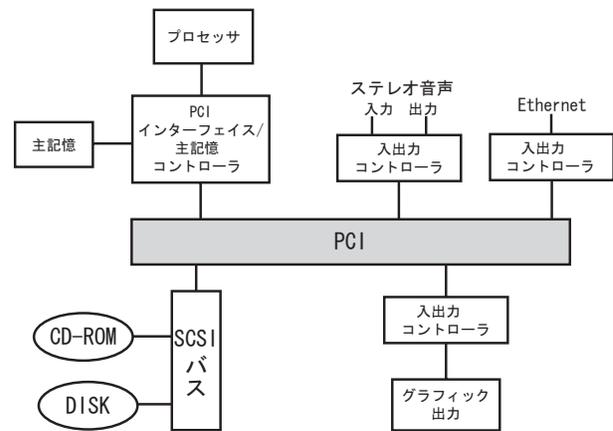


Fig. 1 入出力システムの構成

システム全体に影響を与える。

- バスを複数のアダプタで共有するため、各アダプタあたりのバンド幅が制限される。
- CPU が I/O コントローラを直接制御するため、I/O 動作により CPU 負荷が増加する。また、CPU は PCI バス上の障害の影響を受ける。
- 共有パラレルバスという性質上、バス速度と接続するアダプタ数の上限やバスの長さが制限される。

3 Infiniband

Infiniband⁴とはサーバの CPU と I/O 専用端末間の I/O 機能をオフロードすることにより、サーバ間のトランザクション速度を高速化させる新技術のことである。

Infiniband は、PCI のような「バス」技術とは根本的に異なっている。PCI のデータ経路は多数の装置間で共有される。一方、Infiniband は、他の装置と接続を共有することなく、例えば CPU とネットワークカードの場合のように、1 対 1 の接続を確立する。³⁾

⁴現在の使用では、1 チャンネルを 1 本のケーブルに収めたもの、4 チャンネルを 1 本のケーブルに収めたもの、12 チャンネルを 1 本のケーブルに収めたものの 3 種類が規定されている。10 数 m までの距離は銅線で、それ以上の距離では光ファイバを利用する見込み。

3.1 Infinibandが登場した背景

インターネットをはじめとしたネットワークの進化により、データセンタなどのサーバの重要性が高まるとともに、性能に対しさまざまな要求が出てきた。そこで、前述したような問題を解決して、これらの要求に応えるためのI/Oアーキテクチャとして登場したのがInfinibandである。

技術的にはスイッチを使ったネットワークによってサーバ・I/O間を接続しデータ転送を行なう。転送速度は2.5Gbpsを基本として、これを1、4、12本束ねて使用可能であり、最大転送速度は30Gbpsとなる。またDAFS(Data Access File System)、SDP(Socket Direct Protocol)、RNDIS(Remote Network Sriver Interface Specification)などのデータプロトコルをサポートしている。⁴⁾

3.2 Infinibandの歴史

次世代サーバ向けのI/Oアーキテクチャとしては、当初Intel社が中心となって開発していた「Next Generation I/O(NGIO)」と、Compaq/HP/IBMが中心となっていた「Future I/O(FIO)」があった。1999年8月に両者が歩み寄って「System I/O」という名称で統合を発表され、10月に名称を「Infiniband」と変更した。

規格の策定などはIntel社を中心に設立されたIBTA(InfiniBand Trade Association)で行われており、2000年10月には規格書*Infiniband Architecture Specification Release 1.0*が発表された。

3.3 Infinibandの利点

Infinibandはサーバクラスタを構築するうえで次のような利点を持っている。

- 現在のI/Oよりもはるかに広いI/Oバンド幅
- 高い信頼性と相互接続性
- スケーラブルな構成(1リンクあたり1、4、12bit)
- 物理媒体として銅線以外にも光ケーブルを使用可能

3.4 Infinibandアーキテクチャの概要

Infinibandの大きな特徴は、CPUとI/Oデバイスを接続するために共有メモリ/バスではなく、従来のメインフレームコンピュータで使われていたようなポイントツーポイントチャンネルを使用することにある。⁵⁾ Fig. 2に、Infinibandを使用したシステムの概略を示す。

CPUとI/Oデバイスは、Infinibandスイッチファブリック⁵⁾を経由して接続される。CPUはホストチャンネルアダプタ(HCA:Host Channel Adapter)、I/Oデバイスはターゲットチャンネルアダプタ(TCA:Target Channel

⁵⁾各装置間に複数の独立した経路を設けることによって、データの転送速度を高めるとともに、1ヶ所の障害の影響がシステム全体に及ぶのを防ぐ仕組み。

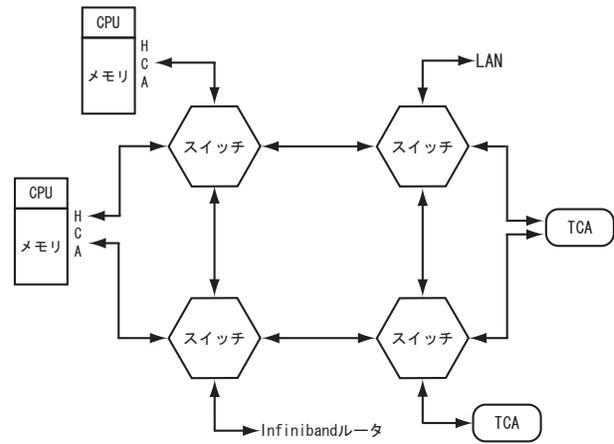


Fig. 2 Infinibandシステムの概略

Adapter)によりスイッチファブリックに接続される。このようなサブネットの構成は動的に管理され、システムを停止することなく各エレメントの交換や増設が可能である。また、I/Oのバンド幅や冗長性を高めるために、リンクやスイッチを増設することで対応できる。

4 将来の展望

Infinibandテクノロジーがまず最初に想定している用途は、サーバとリモート・ストレージ/ネットワーク機器との接続、および複数のサーバ間接続である。このほかの用途としては、パラレル・クラスタのプロセッサ間通信(IPC)用にサーバ内部で使用することも考えられている。これまでサーバのクラスタリングでは、従来は各サーバメーカーが独自規格によって接続していたが、Infinibandは初めてメーカーの壁を超えた統一規格でのクラスタリングが可能になるであろう。さらに、Infinibandのファブリックは数千ノードまで拡張でき、リンクも多重化が可能な仕組みになっており、可用性を高めていると思われる。

そしてそのさらに数年後には、CPUの速度と次世代のソフトウェアにおけるデータ量の増加に合わせるためにInfinibandが家庭用PCに標準搭載されることが予測できる。

参考文献

- 1) 計算機アーキテクチャ、坂井 茂、ソフトバンク株式会社、1995年
- 2) コンピュータの構成と設計・下、David A. Paterson、日経BP社、1999年
- 3) <http://japan.cnet.com/News/2000/Item/001026-5.html>
- 4) <http://storage-system.fujitsu.com/jp/term.html>
- 5) インターフェイス9月号、CQ出版社、2001年