Introduction to InfiniBand Management

Ramón D. Acosta

June 2001

v 1.0

Executive Summary

The InfiniBand standard defines a management infrastructure that is the foundation for achieving multi-vendor interoperability in InfiniBand networks. Functions of the management layer include topology discovery, configuration, communication, and fault tolerance. These functions enable interoperability and integration with enterprise management tools in data center environments.

IBA's management model specifies the following managers and agents:

- Subnet Managers (SMs) perform topology discovery, configuration (including routing tables inside switches), and subnet maintenance
- Subnet Management Agents (SMAs) maintain management parameters for nodes and ports associated with SMs
- General Services Managers (GSMs) perform a variety of management functions related to performance, communication, and I/O devices
- General Services Agents (GSAs) maintain management parameters for nodes and ports associated with GSMs

After initial topology discovery and configuration, the SM periodically sweeps the subnet to determine if any changes have occurred and takes appropriate action to update the topology. The Master SM can also provide an interface to diagnostic frameworks for handling errors.

InfiniBand's comprehensive management model is a central feature of the architecture for enabling interoperability.

Introduction

The InfiniBand architecture is a switched-fabric architecture for next generation I/O systems and data centers. The InfiniBand Architecture (IBA) promises to replace bus-based I/O architectures, such as PCI, with a switched-based fabric whose benefits include higher performance, higher RAS (reliability, availability, scalability), and the ability to create modular networks of servers and shared I/O devices.

The InfiniBand standard defines a Fabric Management model as part of the architecture. This model incorporates mechanisms for topology discovery, configuration, administration, performance management, and device management. More importantly, the model defines a management infrastructure that is the foundation for achieving multi-vendor interoperability in InfiniBand networks.

InfiniBand Fabrics

The InfiniBand architecture connects servers and shared I/O systems. By moving I/O out of servers onto a switched network, InfiniBand technology results in what has been termed "deconstruction of servers" which is anticipated to be a foundational design principle of future data centers. The first release of the InfiniBand Architecture Specification took place last fall.[1] This architecture covers all layers of the standard, including Physical, Link, Network, Transport, and Management.

The InfiniBand architecture is a product of the InfiniBand Trade Association (IBTA, www.infinibandta.org), an industry group of over 200 companies led by Intel, Compaq, Microsoft, Hewlett-Packard, Dell, IBM, and Sun Microsystems. The IBTA was formed in the summer of 1999 as a result of merging the Next Generation I/O (NGIO, led by Intel) and Future I/O (led by Compaq) standards. The IBTA architecture definition work is performed by a series of Working Groups corresponding to layers of the architecture (e.g., ElectroMechanical, Link, Software, Management). There are also IBTA working groups focused on Marketing and Compliance and Interoperability.

Figure 1 depicts a prototypical InfiniBand environment. A single InfiniBand network is termed a subnet. Nodes of a subnet, which support queuing of messages, include:

- **Host Channel Adapters (HCAs)** – interfaces to processor nodes
- **Target Channel Adapters (TCAs)** – interfaces to I/O nodes (including storage and other networks)
- **Switches** – interconnection components for routing traffic within a subnet
- **Routers** – interconnection components for routing traffic across subnets or to non-InfiniBand networks

[1] "InfiniBand™ Architecture Specification, Volumes 1 and 2, Release 1.0," InfiniBand℠ Trade Association, October 24, 2000.

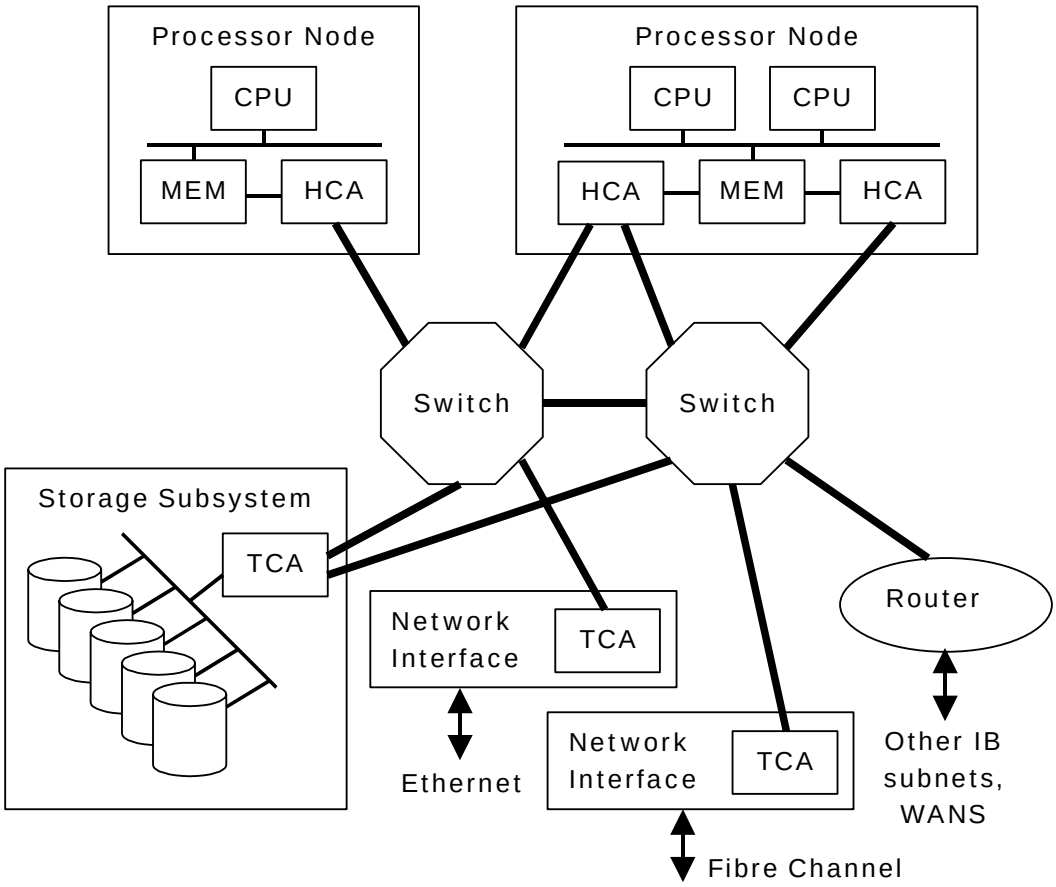

Figure 1. InfiniBand Subnet.

Each node contains one or more ports that are connected to other ports with point-to-point links. Three speeds of InfiniBand bi-directional interconnection links are supported: x1 (2.5Gb/s), x4 (10Gb/s) and x12 (30Gb/s).

The InfiniBand transport layer defines protocols for both unreliable and reliable messages, as well as Direct Memory Access (DMA) mechanisms. These protocols enable a variety of I/O and Interprocessor Communication (IPC) operations between end nodes. The interface for managing message traffic within a channel adapter is called a Queue Pair (QP). Each QP contains both a Send queue and a Receive queue. Communication channels between nodes are established by connecting their QPs.

InfiniBand networks support a number of mechanisms that enable logical control of components and traffic in a subnet. Partitions enable logical isolation of components sharing the same InfiniBand subnet. Virtual Lanes constitute a mechanism for having multiple virtual links share the same physical link. Virtual Lane 15, which is supported by all InfiniBand ports, is reserved for management layer traffic.

InfiniBand Management Model

The InfiniBand management model for a subnet supports managing components from multiple vendors in a single fabric. Functions of the management layer include topology discovery, configuration, communication, and fault tolerance. These functions enable interoperability and integration with enterprise management tools in data center environments.

IBA's management model specifies the following managers and agents (Figure 2):

- **Subnet Managers (SMs)** – perform topology discovery, configuration (including routing tables inside switches), and maintenance of a subnet
- **Subnet Management Agents (SMAs)** – maintain management parameters for nodes and ports associated with SMs
- **General Services Managers (GSMs)** – perform a variety of management functions related to performance, communication, I/O devices, etc.
- **General Services Agents (GSAs)** – maintain management parameters for nodes and ports associated with GSMs

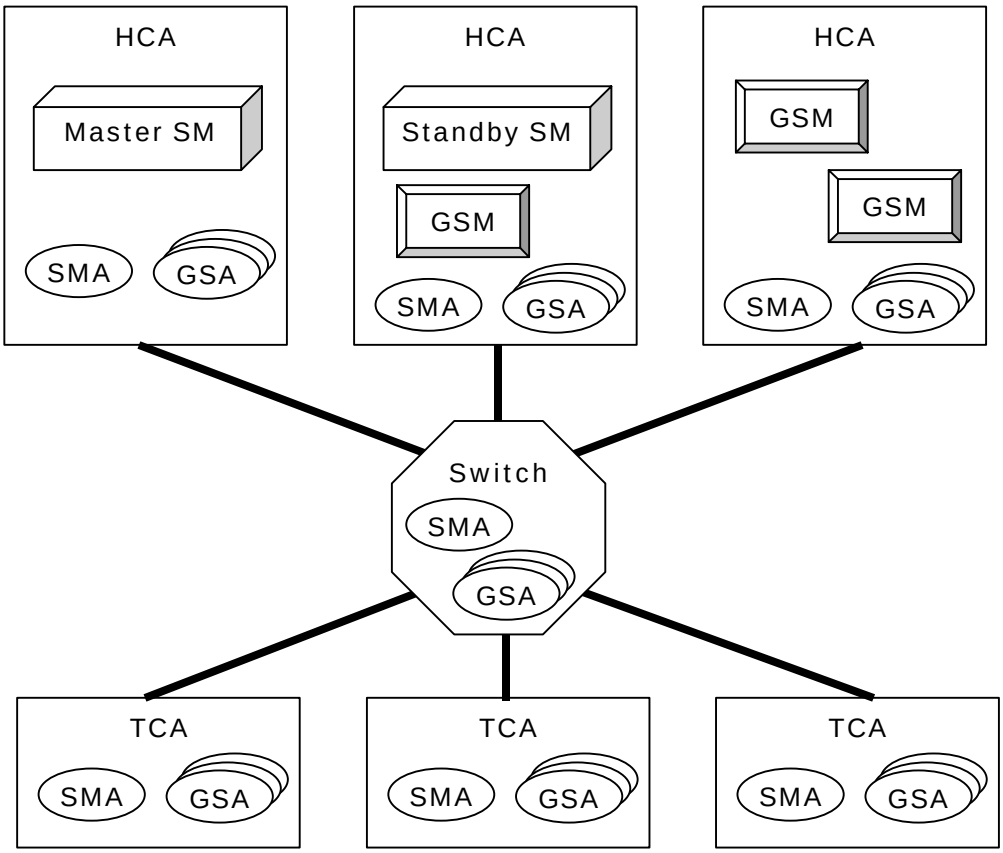

Figure 2. InfiniBand Management Model.

Every InfiniBand node has an SMA and possibly one or more GSAs. Managers and agents communicate via unreliable Management Datagrams (MADs). Two kinds of MADs are defined in the architecture:

- **Subnet Management Packets (SMPs)** – used for communication between SMs and SMAs; SMPs flow over Virtual Lane 15
- **General Services Management Packets (GMPs)** – used for communication between GSMs and GSAs

An InfiniBand subnet has a single Master SM and possibly one or more Standby SMs. The Master SM is responsible for maintaining the subnet, including definition of node/port addressing, routing tables, and partitions. After initial topology discovery and configuration, the Master SM performs periodic sweeps of the subnet to determine if any changes have occurred and take appropriate action to update the topology. The Master SM also supports interfacing to diagnostic frameworks for handling errors. When multiple SMs are present in a subnet, rules are defined by which it is possible to determine which SM will be the Master, with the other SMs taking on the role of Standby SMs.

General services provide infrastructure for management classes other than subnet management. These classes include:

- **Subnet Administrator** – information store for data collected by the SM that can be accessed by nodes
- **Performance Management** – monitor and report performance counters associated with components of the fabric, including ports on nodes
- **Baseboard Management** – power and chassis management for physical devices
- **Communications Management** – establish and manage communication channels between nodes
- **Device Management** – discover and manage I/O devices associated with channel adapters
- **SNMP Tunneling** –communicate between management agents and management applications using SNMP (Simple Network Management Protocol)
- **Vendor-specific Management** –framework for vendors to define management communications and operations outside of the IBA scope
- **Application-specific Management** - framework that can be used for defining application services that are outside of the IBA scope

Conclusions

The InfiniBand architecture, with its support for managing switched-fabric I/O, has far-reaching implications in future data center implementations. The comprehensive InfiniBand management model is a central feature of the architecture for enabling interoperability. The early development and integration of Fabric Management software and other management components will help guarantee the success of the InfiniBand technology.

Ramón D. Acosta, Ph.D., Vice President of Industry Initiatives

As Lane15's representative in standards organizations, Ramón D. Acosta plays a key role in ensuring that Lane15 solutions are both highly effective and widely adopted. Previously, he was Vice President of Engineering at Pervasive Software, where he steered the directions of Pervasive's database management and web development product families. His prior experience includes positions in research, development, and management with Scientific and Engineering Software (SES), International Software Systems Incorporated (ISSI), and MCC. Dr. Acosta has published over 20 papers on computer architecture and computer-aided engineering. He holds a B.S. and M.S. in Computer and Systems Engineering from Rensselaer Polytechnic Institute and an M.S. and Ph.D. in Electrical Engineering from Cornell University.

Lane15 Software

Lane15 Software is the leading developer of fabric management software for heterogeneous InfiniBand networks. The company's open, vendor-neutral products will be an early requirement for the development, testing, and deployment of InfiniBand hardware products as well as a key to the successful exploitation of InfiniBand in customer environments. InfiniBand hardware and software vendors will benefit from reduced time-to-market for new InfiniBand products, lower development costs for building required fabric management elements, ensured interoperability, and vastly accelerated customer and industry adoption of InfiniBand technology. Lane15 is headquartered in Austin, Texas. The company is a member of the InfiniBand^{SM} Trade Association. For additional information about Lane15, visit www.lane15.com