

Introduction to the InfiniBand Interoperability Challenges

Ramón D. Acosta

June 2001

v 1.0

Executive Summary

The IBTA specification contains compliance statements that specify all required and optional component characteristics and protocols. Although following the compliance statements should result in product interoperability, the reality is that full-scale InfiniBand interoperability will be a tedious and arduous process that must overcome the following challenges:

- The specification contains inconsistencies and features which are "under-specified"
- Initial features and services will be a mixture of optional and required features
- Vendors will go beyond the standard to provide value-added features
- Defects in early release products
- Compliance testing does not guarantee interoperability

In addition, there are interoperability challenges specific to management components:

- Master SM negotiation and handover
- Multi-vendor SMAs
- Service resolution
- Management across different subnets
- Multi-network and multi-management environments

InfiniBand interoperability is a complex problem that requires adherence to the standard, flexible solutions, and extensive testing.

InfiniBand Fabrics

InfiniBand is a switched-fabric architecture used for interconnecting 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 IBA is being defined by the InfiniBand Trade Association (IBTA, <http://www.infinibandta.org>), an industry group of over 200 companies led by Intel, Compaq, Microsoft, Hewlett-Packard, Dell, IBM, and Sun Microsystems. 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.

Interoperability Challenges for InfiniBand

The IBA specification defines requirements for devices to be compliant with the architecture. This is done by means of compliance statements specifically formulated to specify all required and optional component characteristics and protocols for HCAs (Host Channel Adapters), TCAs (Target Channel Adapters), switches, routers, cables, connectors, etc. As an example, the following compliance statement from the specification says that whenever a Get() MAD (Management Datagram) is received by a channel adapter, switch or router requesting a management attribute, the device receiving the MAD shall respond using a GetResp() MAD.

C13-10: In response to a valid Get(), the responder shall generate a GetResp() which consists of one or more response MADs – the number is a function of class-specific requirements for the requested attribute.

In an ideal world, strict adherence to compliance statement requirements would result in products capable of interoperating at all layers of the architecture. The practical reality, however, is that full-scale heterogeneous InfiniBand interoperability involves a tedious and arduous process. This process requires dedicated resources (time, people, equipment) from multiple vendors cooperating to achieve interoperability.

Why is this so? What are the key challenges to be overcome for InfiniBand interoperability?

- **Inconsistencies and “under-specification”.** Release 1.0 of the IBA specification took place just at the end of 2000. As a brand new specification, and a rather lengthy and involved one (almost 1500 pages), there are bound to be either inconsistencies or under-specification of features and protocols. In fact, many of the IBTA’s technical Working Groups have been dedicating time to identifying and developing fixes for errata in the specification. Meanwhile, vendors have been developing products with the published standard and early device implementations are just beginning to appear in Q1 of 2001. Looking to the future, the specification will never be perfect for an architecture as complex and all-encompassing as InfiniBand. Marketplace realities require that vendors overcome deficiencies in the IBA specification by cooperating to build and test interoperable products.

¹ “InfiniBand™ Architecture Specification, Volumes 1 and 2, Release 1.0,” InfiniBandSM Trade Association, October 24, 2000.

- **Optional features in the specification.** The IBA specification defines a number of features and services that are optional for minimal InfiniBand device implementations. However, most product implementations are supporting optional features to build truly usable InfiniBand environments. For example, the only required transport service type is Unreliable Datagram with an MTU (Maximum Transfer Unit) size of 256 bytes. Yet practical device implementations for I/O intensive environments must support optional RDMA operations over either Reliable Connection or Reliable Datagram services with larger MTU sizes. The challenge is to determine which products can interoperate effectively based on the features and functions of the architecture they support.
- **Value-added features and higher-level protocols.** Beyond features that are defined as optional in the standard, there are a number of areas in which vendors will seek competitive advantages by going beyond the standard to provide value-added features. In fact, the IBA specification points out a number of architectural boundaries with the expectation that functionality outside of these boundaries is left up to implementations (e.g., routing and partitioning policies, vendor-specific and application-specific management, higher-level protocols sitting on InfiniBand, integration with application and enterprise management environments). As this value-add functionality is developed, defining and validating interoperability becomes more complex. Additional interoperability issues will be encountered as some of this value-add functionality becomes standardized in the form of higher-level protocols (e.g., SNMP, IP-over-IB, SCSI-over-IB).
- **Defects in early release products.** Despite InfiniBand being a new technology, the competitive drive of vendors to get InfiniBand products to market is high. As with any new standard of this level of complexity, software and hardware implementation errors will undoubtedly be made in early product implementations. Areas that may prove to be problematic in early products include timing, handshake protocols, and platform-independence (e.g., little-endian vs. bit-endian byte ordering). Depending on the severity of these defects, vendors may choose to proceed with productization to get early leader competitive advantages. Interoperability will play a very important role in determining the production quality and evaluating defect severity for early release products. Releasing InfiniBand products that have weak interoperability profiles will be detrimental not only to individual vendors, but to the entire InfiniBand movement.
- **Compliance testing vs. interoperability testing.** A clear distinction needs to be drawn between compliance testing and interoperability testing of InfiniBand components. The IBA specification calls out requirements by means of compliance statements. Consequently, compliance test suites target systematic coverage of compliance statements. Rigorous compliance testing is a prerequisite for building quality products. One should view compliance testing as one of the key steps of the interoperability testing process, but it is not the only step and should not be considered a replacement for achieving broader interoperability goals.

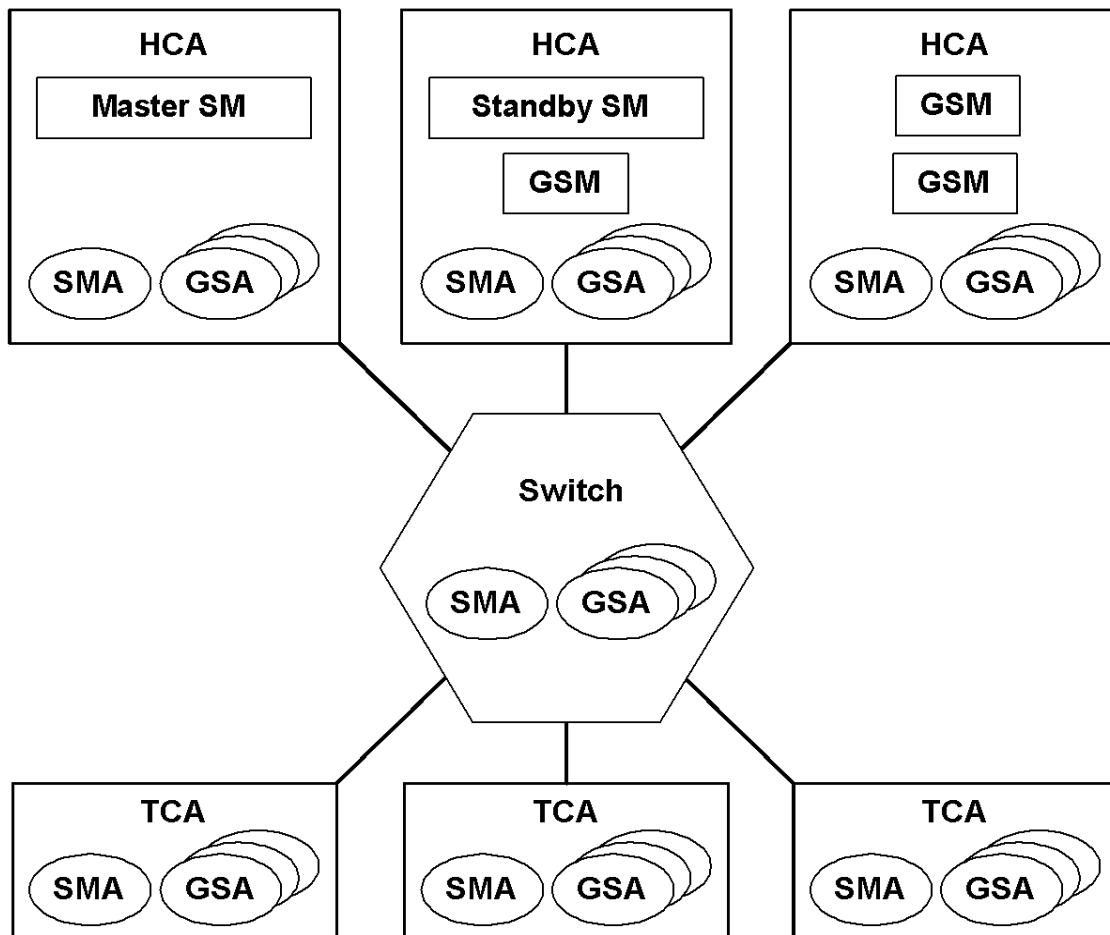
Management Model

InfiniBand technology defines a management model for a subnet that 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.

The IBA management model defines the following set of managers and agents that communicate with each other via MADs.

- **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

The following diagram shows a graphical representation of the Infiniband Management Model.



To achieve interoperability, the IBA specification describes managers, agents, and MADs in terms of vendor-independent interfaces. Implementations must conform to interface protocol requirements to achieve interoperability. This applies to SMs, SMAs, GSMs, and GSAs.

In the case of the SM and SMAs, a properly working management layer is an essential precondition for almost any kind of interoperability testing since the first packets to flow over InfiniBand links when you bring up a fabric are SMPs (Subnet Management Packets). Beyond the SM and SMA, the Communication Manager is the next “critical path” component for interoperability as it is involved in establishing and tearing down communication channels between queue pairs on node ports.

Management Layer Interoperability

All of the challenges for InfiniBand interoperability described earlier in this section apply to management components as well. Specific issues that represent added complexity for management layer interoperability include:

- **Master SM negotiation and handover.** The IBA specification describes rules for determining who is the Master SM, but “mastership” negotiation and handover mechanisms to a Standby SM are left up to vendor implementations. Complications are almost guaranteed to arise with merging of subnets, especially in heterogeneous environments with multiple operating systems and policies (e.g., routing, partitioning).
- **Multi-vendor SMAs.** Environments with components from different vendors will very likely involve SMAs from different vendors, and a Master SM from yet another vendor. Careful compliance to the IBA spec will be a precondition for interoperability. Careful interoperability testing will be necessary since diagnosis and recovery from transient errors in the field will be very difficult.
- **Service Resolution.** The IBA defines several mechanisms for resolving service names into addresses of nodes/ports that provide such services. These mechanisms include Subnet Administration Service Records and Service ID resolution protocol for Communication Management. In practice, the IBA service resolution mechanisms need to be integrated with higher-level directory services in order to be usable by and interoperable with applications and management environments.
- **Management across different subnets.** While the IBA specification rigorously defines mechanisms for managing nodes within a subnet, the specification of multi-subnet management mechanisms and routers is addressed in much less detail. Achieving interoperability across subnets can thus be anticipated to create additional complexities in bringing up InfiniBand fabrics. Again, Fabric Management is at the center of supporting multi-subnet interoperability.
- **Multi-network environments.** Success of InfiniBand requires smooth integration in data center environments where multiple networking technologies coexist today (e.g., Ethernet, Fibre Channel, Storage Area Networks). It is important to extend Fabric Management SM and GSM mechanisms to enable InfiniBand integration into Internet and Enterprise data centers. For these environments, interoperability transcends InfiniBand vendors and products into the realm of multi-protocol networks, including associated network processors and routers.

- **Multi-management environments.** Perhaps even more critical than multi-network systems is the problem of systems management. Data center administration is widely recognized as one of the largest expenses associated with IT. The typical approach to reducing administration expenses and increasing efficiency is to deploy systems management infrastructures. Some systems management hooks have been incorporated into InfiniBand (e.g., SNMP Tunneling) and there are efforts underway to build a richer set of integration mechanisms. These integration mechanisms will provide additional management interoperability problems to attack at higher levels of the architecture.

Conclusions

The InfiniBand Architecture, with its support for managing switched-fabric I/O, has far-reaching implications in future data center implementations. InfiniBand interoperability is a complex problem and requires adherence to the standard, flexible solutions, and extensive testing. Given that the IBA specification was released in the Fall of 2000 and that devices have just recently been announced, now is the time to lay the foundations for interoperability.

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 InfiniBandSM Trade Association. For additional information about Lane15, visit www.lane15.com

InfiniBand^{TM/SM} is a trademark and service mark of the InfiniBand Trade Association. All names and brands are property of their respective owners.