Mellanox TECHNOLOGIES

Connect. Accelerate. Outperform.™

InfiniBand FAQ

Rev 1.3

Last Updated: December 22, 2014

MICROSOFT EXHIBIT 1034
Microsoft v. Partec
IPR2025-00318

Petitioner has added numbering to page 1. Otherwise, it leaves the original page numbering.

NOTE:
THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT ("PRODUCT(S)") AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES "AS-IS" WITH ALL FAULTS OF ANY KIND AND SOLELY FOR THE PURPOSE OF AIDING THE CUSTOMER IN TESTING APPLICATIONS THAT USE THE PRODUCTS IN DESIGNATED SOLUTIONS. THE CUSTOMER'S MANUFACTURING TEST ENVIRONMENT HAS NOT MET THE STANDARDS SET BY MELLANOX TECHNOLOGIES TO FULLY QUALIFY THE PRODUCTO(S) AND/OR THE SYSTEM USING IT. THEREFORE, MELLANOX TECHNOLOGIES CANNOT AND DOES NOT GUARANTEE OR WARRANT THAT THE PRODUCTS WILL OPERATE WITH THE HIGHEST QUALITY. ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL MELLANOX BE LIABLE TO CUSTOMER OR ANY THIRD PARTIES FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING, BUT NOT LIMITED TO, PAYMENT FOR PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY FROM THE USE OF THE PRODUCT(S) AND RELATED DOCUMENTATION EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Mellanox Technologies
350 Oakmead Parkway Suite 100
Sunnyvale, CA 94085
U.S.A.
www.mellanox.com
Tel: (408) 970-3400
Fax: (408) 970-3403

Mellanox Technologies, Ltd.
Beit Mellanox
PO Box 586 Yokneam 20692
Israel
www.mellanox.com
Tel: +972 (0)74 723 7200
Fax: +972 (0)4 959 3245

© Copyright 2014. Mellanox Technologies. All Rights Reserved.

Mellanox®, Mellanox logo, BridgeX®, ConnectX®, Connect-IB®, CORE-Direct®, InfiniBridge®, InfiniHost®, InfiniScale®, MetroX®, MLNX-OS®, PhyX®, ScalableHPC®, SwitchX®, UFM®, Virtual Protocol Interconnect® and Voltaire® are registered trademarks of Mellanox Technologies, Ltd.

ExtendX™, FabricIT™, Mellanox Open Ethernet™, Mellanox Virtual Modular Switch™, MetroDX™, TestX™, Unbreakable-Link™ are trademarks of Mellanox Technologies, Ltd.

All other trademarks are property of their respective owners.

Contents

1 What is InfiniBand (IB)? 4
2 How is InfiniBand Different from Traditional Network Protocols? 4
3 In What Ways Is InfiniBand Different than TCP? 5
4 Is InfiniBand Strictly an I/O Fabric? 5
5 Is InfiniBand a Layered Protocol? 5
6 What Are the Advantages of InfiniBand? 6
7 What Are the Available InfiniBand Data Rates? 6
8 What Is RDMA and What Are Its Benefits? 7
9 What Are the Primary Elements of the InfiniBand Architecture? 8
9.1 What Is a Host Channel Adapter (HCA)? 9
9.2 What Is a Switch and How Does It Work Within an InfiniBand Fabric? 9
9.3 What Is a Subnet Manager (SM)? 9
9.4 Is a Router a Necessity In an InfiniBand Fabric? 10
9.5 What Is a Gateway and How Does It Work Within an InfiniBand Fabric? 10
10 What Is Virtual Protocol Interconnect® and How Does It Relate To InfiniBand? 10
11 What Are LIDs, GIDs, and GUIDs? 11
12 Does InfiniBand Support IP Traffic? What Is IPoIB? 11
13 What Are Reliable and Unreliable Transport Modes? 12
14 Does IPoIB Support Bonding? 12
15 Does InfiniBand Support Multicast? 12
16 Does InfiniBand Support Quality of Service? 13
17 Is InfiniBand a Lossless Fabric? 13
18 How Does InfiniBand Deal With Security? 14
19 How Does Credit-based Flow Control Work? 14
20 Does InfiniBand Have Spanning Tree? 14
21 What Are Verbs? 15
22 How Can I Monitor the Bandwidth, Congestion, and Health of My InfiniBand Fabric? 15
23 Where Can I Find More Information About the InfiniBand Protocol and Mellanox Applications? 16

Mellanox Technologies Confidential

1 What is InfiniBand (IB)?

InfiniBand is a network communications protocol that offers a switch-based fabric of point-to-point bi-directional serial links between processor nodes, as well as between processor nodes and input/output nodes, such as disks or storage. Every link has exactly one device connected to each end of the link, such that the characteristics controlling the transmission (sending and receiving) at each end are well defined and controlled.

InfiniBand creates a private, protected channel directly between the nodes via switches, and facilitates data and message movement without CPU involvement with Remote Direct Memory Access (RDMA) and Send/Receive offloads that are managed and performed by InfiniBand adapters. The adapters are connected on one end to the CPU over a PCI Express interface and to the InfiniBand subnet through InfiniBand network ports.. This provides distinct advantages over other network communications protocols, including higher bandwidth, lower latency, and enhanced scalability.

The InfiniBand Trade Association (IBTA), established in 1999, chartered, maintains, and furthers the InfiniBand specification, and is responsible for compliance and interoperability testing of commercial InfiniBand products. Through its roadmap, the IBTA has pushed the development of higher performance more aggressively than any other interconnect solution, ensuring an architecture that is designed for the 21st century. Mellanox is proud to be an active member of the IBTA steering committee.

For more information on InfiniBand, please see:

https://cw.infinibandta.org/document/dl/7268

2 How is InfiniBand Different from Traditional Network Protocols?

InfiniBand is designed to enable the most efficient data center implementation. It natively supports server virtualization, overlay networks, and Software-Defined Networking (SDN).

InfiniBand takes an application-centric approach to messaging, finding the path of least resistance to deliver data from one point to another. This differs from traditional network protocols (such as TCP/IP and Fibre Channel), which use a more network-centric method for communicating.

Direct access means that an application does not rely on the operating system to deliver a message. In traditional interconnect, the operating system is the sole owner of shared network resources, meaning that applications cannot have direct access to the network. Instead, applications must rely on the operating system to transfer data from the application's virtual buffer to the network stack and onto the wire, and the operating system at the receiving end must have similar involvement, only in reverse.

Mellanox Technologies Confidential

In contrast, InfiniBand avoids operating system involvement by bypassing the network stack to create the direct channel for communication between applications at either end. The simple goal of InfiniBand is to provide a message service for an application to communicate directly with another application or storage. Once that is established, the rest of the InfiniBand architecture works to ensure that these channels are capable of carrying messages of varying sizes, to virtual address spaces spanning great physical distances, with isolation and security.

3 In What Ways Is InfiniBand Different than TCP?

InfiniBand is architected for hardware implementation, unlike TCP which is architected with software implementation in mind. InfiniBand is therefore a lighter weight transport service than TCP in that it does not need to re-order packets, as the lower level link layer provides in-order packet delivery. The transport layer is only required to check the packet sequence and deliver packets in order. Further, because InfiniBand offers credit-based flow control (where a sender node does not send data beyond the "credit" amount that has been advertised by the receive buffer on the opposite side of a link), the transport layer does not require a drop packet mechanism like the TCP windowing algorithm to determine the optimal number of in-flight packets. This enables efficient products delivering 56 and soon 100Gb/s data rates to applications with very low latency and negligible CPU utilization.

4 Is InfiniBand Strictly an I/O Fabric?

No, InfiniBand offers much more. At the lowest layer, InfiniBand provides a high performance, low latency, reliable switch fabric to serve as a scalable I/O interconnect. However, InfiniBand offers higher layers of functionality that enable application clustering, virtualization, and SANs (Storage Area Networks).

5 Is InfiniBand a Layered Protocol?

Yes. The InfiniBand specification defines the protocol in modular layers loosely based on the OSI 7-layer model and covers layers 1-4. The specification defines interfaces between a given layer and the one immediately above and below. Thus, the lowest physical layer interfaces only to the link layer above. The InfiniBand link layer defines an interface to the physical layer below and another interface to the network layer above.

Mellanox Technologies Confidential

6 What Are the Advantages of InfiniBand?

InfiniBand's primary advantages over other interconnect technologies include:

- Higher throughput – 56Gb/s per server and storage connection, and soon 100Gb/s, compared to up-to 40Gb Ethernet and Fibre Channel
- Lower latency – RDMA zero-copy networking reduces OS overhead so data can move through the network quickly
- Enhanced scalability – InfiniBand can accommodate theoretically unlimited-sized flat networks based on the same switch components simply by adding additional switches
- Higher CPU efficiency – With data movement offloads the CPU can spend more compute cycles on its applications, which will reduce run time and increase the number of jobs per day
- Better ROI – higher throughput and CPU efficiency at competitive pricing equals higher productivity with lower cost per endpoint

7 What Are the Available InfiniBand Data Rates?

InfiniBand serial links can operate at various signaling rates, and can then be bundled together to achieve higher throughput. The raw signaling rate is coupled with an encoding scheme, resulting in an effective transmission rate. The encoding minimizes the error rate for data sent across the copper or fiber wires and adds some overhead (for example, 10 bits transmitted for every 8 bits of data).

The typical implementation is to aggregate four-link units (4X). Currently, InfiniBand systems offer the following throughput rates (in each direction per link):

Table 1: InfiniBand Data Rates

Name	Abbreviation	Raw Signaling Rate	Applied Encoding	Effective Data Rate	Aggregated (4x) Throughput
Single Data Rate	SDR	2.5 Gb/s	8b/10b	2 Gb/s	**8 Gb/s**
Double Data Rate	DDR	5 Gb/s	8b/10b	4 Gb/s	**16 Gb/s**
Quad Data Rate	QDR	10 Gb/s	8b/10b	8 Gb/s	**32 Gb/s**
Fourteen Data Rate	FDR	14.1 Gb/s	64b/66b	13.64 Gb/s	**54.5 Gb/s**
Enhanced Data Rate	EDR	25.8 Gb/s	64b/66b	25 Gb/s	**100 Gb/s**
High Data Rate	HDR	51.6 Gb/s	64b/66b	50 Gb/s	**200 Gb/s**
Next Data Rate	NDR	TBD	TBD	TBD	**TBD**

For more information on InfiniBand data rates, please see:
http://www.infinibandta.org/content/pages.php?pg=technology_overview

Mellanox Technologies Confidential

8 What Is RDMA and What Are Its Benefits?

InfiniBand uses Remote Direct Memory Access (RDMA) as its method of transferring the data from one end of the channel to the other. RDMA is the ability to transfer data directly between the applications over the network with no operating system involvement and while consuming negligible CPU resources on both sides (zero-copy transfers). The application on the other side simply reads the message directly from the memory, and the message has been transmitted successfully.

This reduced CPU overhead increases the network's ability to move data quickly and allows applications to receive data faster. The time interval for a given quantity of data to be transmitted from source to destination is known as latency, and the lower the latency, the faster the application job completion.

Figure 1: Traditional Interconnect

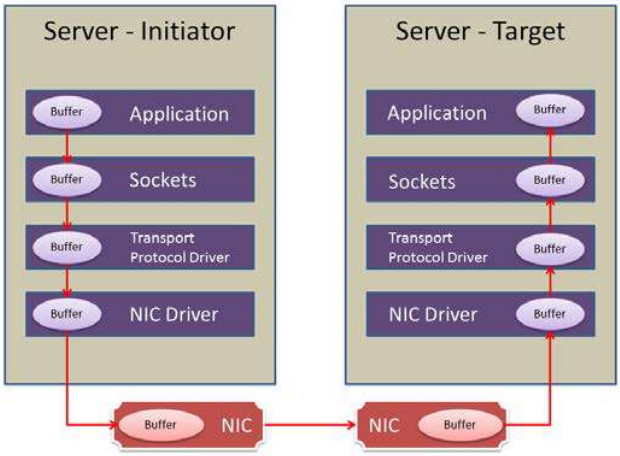
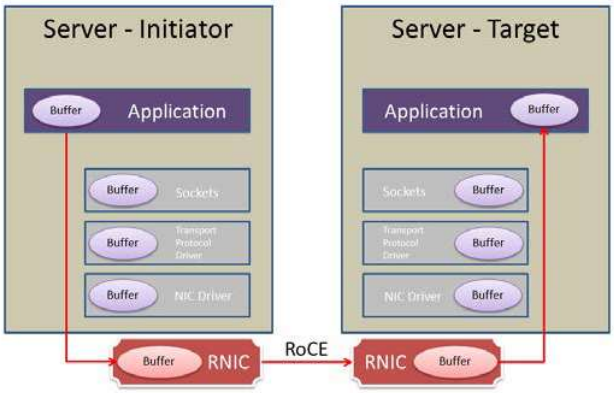


Figure 2: RDMA Zero-Copy Interconnect



Mellanox Technologies Confidential

Mellanox FDR InfiniBand has achieved latency as low as 0.7 microseconds, far and away the lowest latency available for data transmission.

For more information on RDMA, please see:
http://www.mellanox.com/blog/tag/rdma/

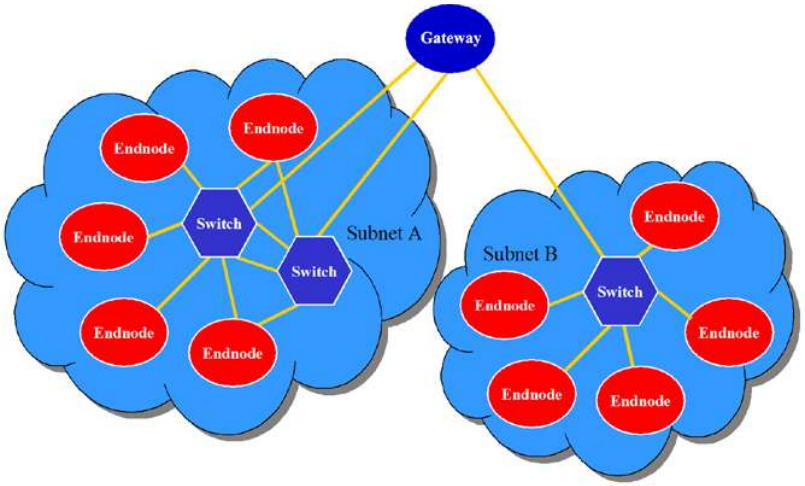
9 What Are the Primary Elements of the InfiniBand Architecture?

The basic building blocks of an InfiniBand network are:

- Host Channel Adapters (HCA)
- Switches
- Subnet Managers (SM)
- Gateway

Each end node must have a Host Channel Adapter to set up and maintain the link with the host device. Switches contain more than one InfiniBand port and forwards packets from one port to another to continue the transmission of the packet within a subnet. A router is used to forward packets from one subnet to another (if necessary). Subnet Management is handled through Software Defined Networking, which controls the network's physical elements and introduced traffic engineering features, often via open, industry-standard interfaces.

Figure 3: Basic InfiniBand Architecture



For more detailed information on InfiniBand architecture, please see:
http://www.mellanox.com/related-docs/solutions/deploying-hpc-cluster-with-mellanox-infiniband-interconnect-solutions.pdf

Mellanox Technologies Confidential

9.1 What Is a Host Channel Adapter (HCA)?

The Host Channel Adapter (HCA) is an interface card or controller that bridges between the InfiniBand wire and the host system bus. Each end node must have an HCA, which sets up and maintains the link between the host device and the rest of the entities on the network.

HCAs provide port connections to other devices. An HCA can be connected to another HCA, to a target device, or to a switch.

9.2 What Is a Switch and How Does It Work Within an InfiniBand Fabric?

A switch is used to physically connect devices within a network and forward incoming data traffic toward its destination. Switches have multiple ports that process and forward data across cables to the specific device for which it is intended, regulating the flow of traffic within the network.

In an InfiniBand fabric, the switch is a crucial piece of the architecture. In fact, InfiniBand is called a “switched fabric” or “switch-based interconnect” because when traffic is forwarded there is a logical connection from one port to another, similar to the old-style telephone switchboards. As more switches are added to the system, there are even more possible paths between devices.

9.3 What Is a Subnet Manager (SM)?

The Subnet Manager (SM) is a software entity that configures its local subnet and ensures its continued operation. It sets up primary and secondary paths between every end point so that traffic flow forwarding decisions are preprogrammed and data arrives in the least amount of time. There must be at least one SM present in a subnet in order to manage all switch and router setups, and to reconfigure the subnet when a link goes down or a new link comes up. The SM can reside on any of the devices within the subnet.

All devices in the subnet must contain a dedicated Subnet Management Agent (SMA), which is used by the SM to communicate with the various InfiniBand components.

There can be multiple SMs in a subnet, as long as only one is active at any moment. Non-active SMs, known as Standby Subnet Managers, keep copies of the active SM’s forwarding information and verify that the active SM is operational. If an active SM goes down, a standby SM takes over responsibilities to ensure that the entire fabric continues with its operation.

Software-Defined Networking (SDN) has emerged as a primary means of subnet management, not only controlling the network devices and the data traffic between them, but also adding network flexibility and scalability.

Mellanox Technologies Confidential

9.4 Is a Router a Necessity In an InfiniBand Fabric?

A router is used to physically connect devices between separate computer networks. When data reaches a router, it reads the address information in the packet and determines the ultimate destination.

In Ethernet networking, large subnets are not efficient due to the flooding mechanism and the spanning tree. A router is therefore required to bridge between various smaller subnets.

These issues do not exist in InfiniBand, in which large subnets of up to approximately 40,000 nodes can run efficiently. There is no need to divide into smaller subnets, so a router is not a necessity.

However, with the ever-growing demands on data centers, InfiniBand router technology, such as Switch-IB™ by Mellanox, is available should there be a need for a subnet of more than 40,000 nodes.

9.5 What Is a Gateway and How Does It Work Within an InfiniBand Fabric?

A gateway is a device within a subnet that serves as a bridge between two protocols. This allows, for example, an InfiniBand cluster to access Ethernet network management or storage interfaces. This allows a company to implement an InfiniBand network with a dedicated interface to existing legacy equipment that runs on other protocols. Gateways are not requied to connect InfiniBand clusters with Ethenet networks. Some or all of the compute nodes can have Etherent access by utilizing Ethernet adapters in addition to their InfiniBand HCAs or by using HCAs with Virtual Protocol Interconnect®.

For more information on InfiniBand gateways, please see:
http://www.mellanox.com/page/gateway_overview

10 What Is Virtual Protocol Interconnect® and How Does It Relate To InfiniBand?

Virtual Protocol Interconnect® (VPI) is a Mellanox technology that offers its customers the best available interconnect while ensuring scalability affordably. VPI allows any port within an HCA or switch to run either InfiniBand or Ethernet, and to alternate between the two protocols as necessary. For servers and storage systems, this allows the highest connectivity flexibility. For the switch fabric, this creates a perfect gateway out-of-the-box, enabling integration between InfiniBand and Ethernet fabric and clusters.

Mellanox Technologies Confidential

Mellanox VPI offers both port flexibility and financial flexibility. Instead of pay-as-you-grow, you can adapt ports as needed to allow for growth. Moreover, Mellanox VPI provides all the benefits of InfiniBand while maintaining easy connectivity to an existing Ethernet fabric. Most of all, this is achievable with no performance penalty whatsoever, as Mellanox ensures the highest bandwidth and lowest latency in both its InfiniBand and Ethernet offerings.

For more information on Mellanox VPI, please see:
http://www.mellanox.com/related-docs/case_studies/CS_VPI_GW.pdf

11 What Are LIDs, GIDs, and GUIDs?

All devices in a subnet have a Local Identifier (LID), a 16-bit address assigned by the Subnet Manager. All packets sent within a subnet use the LID as the destination address for forwarding and switching packets at the Link Level. The LIDs allow for up to 48,000 end nodes within a single subnet. When a subnet is reconfigured, new LIDs are assigned to the various endpoints within the subnet.

Routing between different subnets is done on the basis of a Global Identifier (GID), a 128-bit address modeled after IPv6 addresses, which allows for InfiniBand's essentially unlimited scalability. GIDs identify an end node, port, switch, or multicast group.

Global Unique Identifiers (GUID) are 64-bit definitions for all the elements within a subnet, including chassis, HCAs, switches, routers, and ports. The GUID never changes, and is used as part of the address for creating a GID.

GIDs and GUIDs are independent of LIDs and are therefore immune to subnet reconfiguration.

12 Does InfiniBand Support IP Traffic? What Is IPoIB?

Internet Protocol (IP) packets can be sent via an InfiniBand interface by encapsulating the IP packets in an InfiniBand packet via a network interface. This is known as IP over IB (IPoIB). As long as the InfiniBand network has the necessary driver installed, it creates an interface for each port using partition keys (PKEYs) and can then transport IP packets across the InfiniBand network seamlessly.

For more information on IPoIB, please see the Mellanox OFED User Manual or Mellanox WinOF VPI User Manual.

13 What Are Reliable and Unreliable Transport Modes?

When packets are transmitted from one node to another, they can be sent either reliably or unreliably. Reliable transport is exactly as it sounds – there can be no loss or duplication of packets, the packet must be delivered in order, and the receiving node provides acknowledgment (positive or negative) back to the sender that the packet has been received correctly. By assuring delivery of data to the remote peer via the hardware protocol, the application itself is relieved of such responsibility.

Unreliable transport does not provide such acknowledgment, and will make a best effort to deliver the packet (known as a datagram when using unreliable transport) in order.

Reliable transport requires that additional resources be reserved to ensure that packets are sent and received correctly, as well as to handle the acknowledgment of such. Therefore, there are situations in which a data center might opt to use unreliable datagrams instead of reliable connections in order to allocate those resources elsewhere.

IP over InfiniBand (IPoIB) can run over either Unreliable Datagram (UD) mode for unreliable transport, or Connected Mode (CM), which allows for reliable transport. It is possible to switch between the two modes at runtime with a single command. The default InfiniBand transport mode is UD, but it is possible to add a script that automatically configures new interfaces as CM for reliable transport.

14 Does IPoIB Support Bonding?

Bonding refers to the process of joining two ports together for use by a single application, offering it more flexibility in sending data. Bonding of IPoIB interfaces is only available in Linux, and can be accomplished in the same manner as is bonding of Ethernet interfaces, that is, via the Linux Bonding Driver. Windows does not yet support bonding.

15 Does InfiniBand Support Multicast?

Switches can be configured to forward either unicast packets (to a single location) or multicast packets (to multiple devices). These fully hot swappable connections are managed by the Subnet Manager and can include sending to all systems on the subnet or to a subset of those systems. InfiniBand's high bandwidth provides the backbone for such multicast abilities without the need for a secondary interconnect link.

Mellanox Technologies Confidential

16 Does InfiniBand Support Quality of Service?

Quality of Service (QoS) is the ability of a network to provide different priorities to applications and to guarantee a certain level of performance in the flow of data to those endpoints.

InfiniBand supports QoS by creating Virtual Lanes (VL). These VLs are separate logical communication links that share a single physical link. Each link can support up to 15 standard VLs (designated as VL0 through VL14) and one management lane (designated as VL15). VL15 is given the highest priority and VL0 is the lowest.

A data packet is defined with a Service Level (SL) to ensure its QoS level. The SL provides each link with a desired priority of communication. Each switch or router along the communication path has a mapping table that is set by the Subnet Manager to translate SL to VL in order to keep the proper priority within the VLs supported on that link.

On the host side, QoS parameters can be assigned to each flow (queue pair), essentially providing 16 million QoS levels per endpoint.

When the bandwidth demand for an application reaches the defined QoS limit, the resources used by the application on the host will no longer grow. Lowering the QoS limit for an application that had reached its defined limits will free up resources on the host, benefitting other workloads on that host. This has high value for traffic sharing.

For more information on Quality of Service, please see the Mellanox OFED User Manual or the Mellanox WinOF VPI User Manual.

17 Is InfiniBand a Lossless Fabric?

A lossless fabric is one in which packets are not routinely dropped. Ethernet is considered a lossy fabric because it frequently drops packets. The TCP transport layer detects the lost packets and adjusts for them.

InfiniBand, by contrast, uses link-level flow control to ensure that packets are not dropped in the fabric. This lossless flow control leads to very efficient use of bandwidth within a data center, and makes InfiniBand highly suitable for long-range communication between data centers.

Mellanox Technologies Confidential

18 How Does InfiniBand Deal With Security?

InfiniBand's adapters offer support for advanced virtualization features such as Network Functions Virtualization (NFV). Together with SDN-optimized InfiniBand switches, it is possible to build the most efficient, scalable and high performance networks. InfiniBand solutions allow data to be moved and analyzed faster than with any other solution, thereby providing data center managers the capabilities to collect network traffic information, to analyze traffic behavior (such as the detection of anomalies), and to secure it. The solutions enable building infrastructures for firewall and intrusion detection solutions. Furthermore, network traffic control over InfiniBand solutions is flexible and re-programmable in real-time.

For an example of an SDN-based security solution with SR-IOV adapters, see the following blog post.

19 How Does Credit-based Flow Control Work?

Flow control is used to manage the flow of data between links in order to guarantee a lossless fabric. Each receiving end of a link (Virtual Lane) supplies credit to the sending device to specify how much data can be received without loss of data. A dedicated link packet manages the credit passing between devices in order to update the number of packets the receiving device can accept. No data is transmitted unless the credit that is announced by the receiving end indicates sufficient buffer space to receive the entire message.

20 Does InfiniBand Have Spanning Tree?

Spanning Tree is considered one of the best and one of the worst elements of Ethernet. It allows Ethernet networks to diagnose when a loop exists between nodes and disables redundant or parallel links to prevent packets from being sent repeatedly.

While this was extremely useful 10-15 years ago, spanning tree is now considered a liability, in that it restricts the ability of network engineers to build high performance networks because it places limits on how many parallel paths there can be between nodes.

Infiniband, on the other hand, enables massive fabrics of links with full bandwidth between all endpoints by managing traffic via a central agent that sees all the paths in the network. Today, Software-Defined Networking (SDN) serves as this central subnet management, distributing and allocating traffic to make use of all the parallel links between endpoints. Better yet, there is no performance impact of this management because SDN determines the possible paths within the configuration at connection setup.

Mellanox Technologies Confidential

21 What Are Verbs?

A verb is a method by which an application requests an action from InfiniBand's message transport service. The set of verbs are semantic descriptions of the various actions by which applications interact within an InfiniBand network. The verbs are fully defined within the InfiniBand software transport interface specification.

These verbs are the basis for specifying the APIs that an application uses, but the InfiniBand architecture does not define APIs. Other organizations, such as the OpenFabrics Alliance, provide a complete set of APIs and software that implements the verbs to work seamlessly with the InfiniBand hardware.

For more information on the use of InfiniBand verbs, please see:
https://cw.infinibandta.org/document/dl/7268

http://www.mellanox.com/related-docs/prod_software/RDMA_Aware_Programming_user_manual.pdf

22 How Can I Monitor the Bandwidth, Congestion, and Health of My InfiniBand Fabric?

The OpenFabrics Enterprise Distribution (OFED) is an open-source stack of software drivers, core-kernel code, middleware, and user-level interfaces in support of InfiniBand fabrics. Mellanox OFED for Linux and Mellanox OFED for Windows (WinOF) includes various diagnostic and performance tools that monitor the health of an InfiniBand network, including monitoring the bandwidth of transmissions and monitoring for congestion within the fabric.

Mellanox also offers its Unified Fabric Manager® (UFM®) software, which is a powerful platform for managing InfiniBand computing environments. UFM enables data center operators to efficiently provision, monitor, and operate their fabric, while boosting application performance and ensuring that the fabric is up and running at all times.

For more information on monitoring an InfiniBand fabric, please see:
Mellanox OFED for Linux:
http://www.mellanox.com/page/products_dyn?product_family=26&mtag=linux_sw_drivers

Mellanox WinOF:
http://www.mellanox.com/page/products_dyn?product_family=32

UFM:
http://www.mellanox.com/page/products_dyn?product_family=100&mtag=unified_fabric_manager

23 Where Can I Find More Information About the InfiniBand Protocol and Mellanox Applications?

InfiniBand Trade Association: http://www.infinibandta.org

Mellanox Website: http://www.mellanox.com

Mellanox Community: https://community.mellanox.com/welcome

Mellanox Academy: http://www.mellanox.com/academy/

Mellanox Webinars: http://www.mellanox.com/page/webinars

Mellanox Blog: http://www.mellanox.com/blog/

InfiniBand White Papers: http://www.mellanox.com/page/white_papers

Mellanox Technologies Confidential