WHITE PAPER

InfiniBand: The Production SDN

Background

The concept of Software Defined Networking (SDN) has emerged in recent years as an alternative to the Internet's distributed and proprietary control plane. The introduction of OpenFlow, as a first step towards this vision, allows for standard off-system control, but SDN is more than just the standard for device configuration. In his excellent presentation "*The Future of Networking, and the Past of Protocols*," (Shenker, et al. n.d.) Scott Shenker describes SDN as a three layer abstraction for the network control plane.

SDN networks promise a whole new range of network flexibility and scalability. While these standards are still evolving and developing, Mellanox Technologies has been implementing their concepts in practice for over 10 years now, providing existing data centers with a mature infrastructure for flexible, scalable and dynamic environments. Mellanox's implementation is standard, by InfiniBand Trade Association, and available as an open source project-OpenSM.

In this white paper we describe how the InfiniBand architecture and Mellanox products implement SDN successfully. InfiniBand SDN deployments are widely used today for HPC, Web 2.0, financial applications and many other markets.

The Promise of SDN

SDN networks promise:

- High-level "virtual" representation of the network
- Standard means to control its physical elements
- Scalable architecture that provides high performance even for large flat networks
- Quick additions of new networking features via open, industry standard interfaces

In order to deliver on these promises, a layered abstraction of the network control plane and separation of the control and data planes are required.

SDN abstraction for the network control plane, as presented by S. Shenker, is shown in Figure 1.

Network equipment is represented at the bottom of the diagram and data links are shown in green to complete the network. The control plane is depicted as blue arrows connecting the equipment to a software layer named "Network Operating System" (NOS).

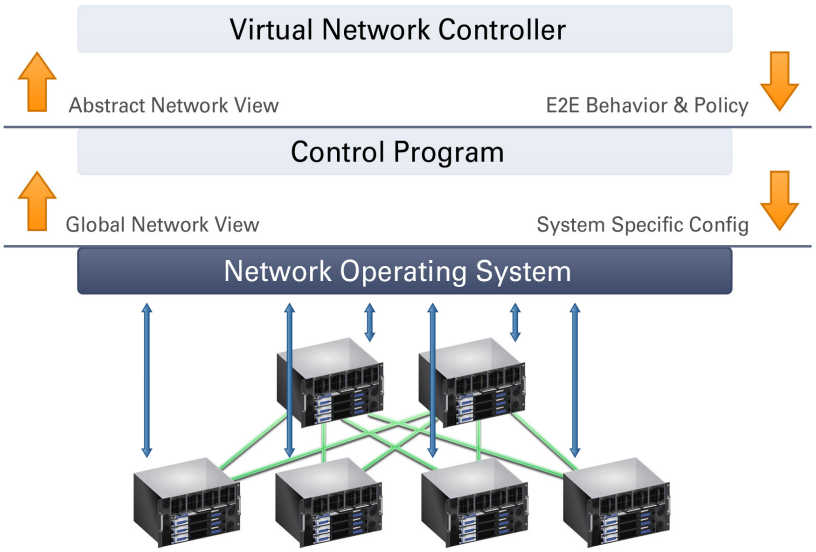

Figure 1. *SDN Abstraction*

These connections are the ones standardized by Ethernet SDN known as OpenFlow. On the one hand, the task of NOS is to represent a Global Network View, such as the connected graph of data links to the upper layers. On the other hand, NOS receives directives about the configuration to be applied to each system and performs the actual setting.

On top of NOS is a Control Program, sometimes called *Compiler*, which takes the directives provided as end-to-end behavior targets and converts them into specific system settings on top of the Global Network View graph.

A user interface program that converts the network manager's intent into high level virtual topology features is at the top of the abstraction and is called Virtual Network Controller.

Mellanox InfiniBand SDN

InfiniBand was specified with concepts similar to those of SDN. Data-path mechanisms (e.g. forwarding) were defined with a clear control plane interface. Every InfiniBand device is mandated to have an embedded control plane agent named Subnet Management Agent (SMA). The protocol and message format for in-band communication with the SMAs is well defined. Subnet Management Packets (SMPs) are used for control plane communication and thus SMP and their protocols are the equivalent to OpenFlow.

An InfiniBand based-SDN equivalent is described in Figure 2.

As shown, NOS equivalent functionality is provided by the Subnet Manager (software entity) which sends and receives SMPs to discover, configure and maintain the network. The Subnet Manager (SM) is available either as an open source project (OpenSM) or as part of Mellanox's Unified Fabric Manager™ (UFM™) product. It can operate as a standalone component or as multiple components working together for High Availability purposes.

The *Compiler* which converts end-to-end configuration and policy to topology-specific configuration is available as an integrated part of the UFM product. UFM also provides the user interface and API to define the network-wide policies at a virtual network level.

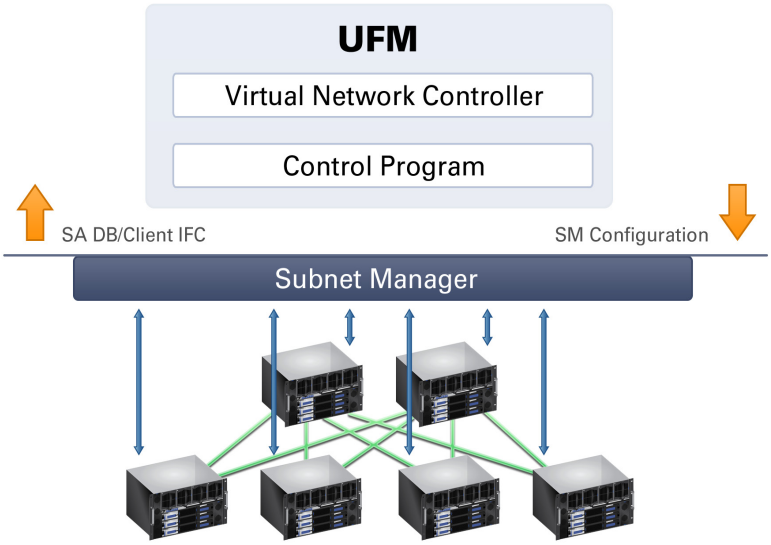

Figure 2. *Mellanox InfiniBand SDN Components*

The Mellanox SDN Advantage

The InfiniBand control plane protocol specification is mature and proven and has the following significant and unique advantages:

Works in-band so there is no need for a separate control network

"Remote" control plane management protocols require that some packet forwarding exists even before the forwarding plane is configured. InfiniBand is not dependent on other protocols to reside within the systems, and utilizes direct-routing for configuration. With such capability, it eliminates the need to build a separate management network and many real world installations rely solely on in-band management. This significantly reduces costs and complexity.

Defines a fast and standard mechanism for extracting the global network view

Topology discovery, the basis for building the Global Network View, is also fully standardized in InfiniBand and relies on SMPs. In contrast to other discovery methods in use today, which rely on higher level protocols such as LLDP and SNMP, the InfiniBand Subnet Manager directly addresses network discovery. It relies on InfiniBand in-band discovery protocols, which results in very fast and efficient fabric discovery. For example, the InfiniBand SM is able to extract a topology of a 20,000-node cluster topology in roughly 10 seconds.

Fast handling of topology faults or changes

It is not sufficient to discover and/or to plan the fabric routing scheme. In daily fabric operations, the ability to detect problems or changes is critical. Unlike other existing solutions, the InfiniBand standard has a specific, well-defined SMA feature that addresses that and provides the Subnet Manager with immediate notifications on fabric events and errors.

Standard global network representation

InfiniBand also provides a standard interface for providing the Global Network View. A set of standard network administration packets provides any authorized client with detailed fabric information.

Mellanox SDN Delivers on SDN Promises Today

Mellanox SDN solutions are already deployed in thousands of sites worldwide and have been in production for many years now. Mellanox networks are centrally managed by NOS software, OpenSM, which is fully controlled by UFM – Mellanox's Virtual Network Controller. Figure 3 shows how Mellanox products are used in the Mellanox SDN solution.

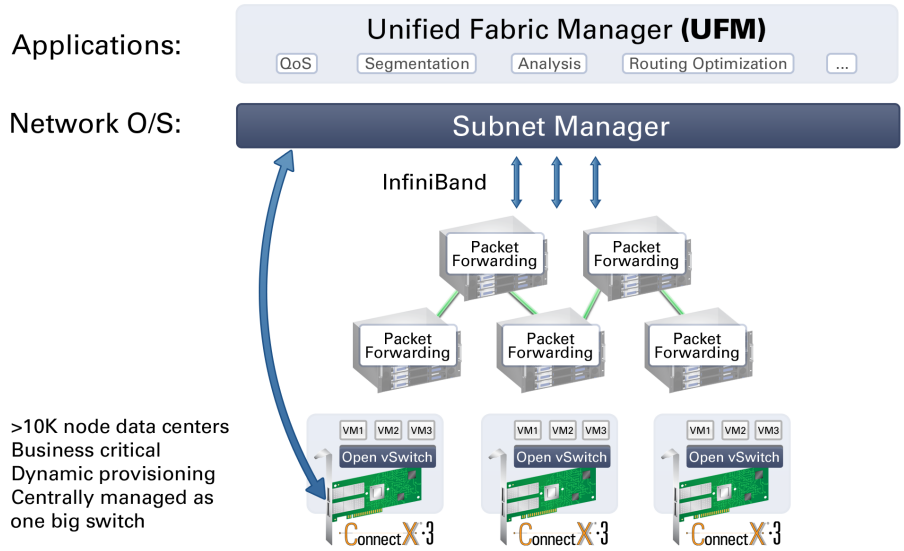

Figure 3. *Mellanox InfiniBand SDN Implementation*

The scale of InfiniBand deployments ranges from several nodes and up to >10K very fat networks (non-blocking and multi-pathing). Mellanox fabrics include High Availability and in-service upgrade, which provide the highest levels of uptime and application performance.

Mellanox SDN provides traffic engineering features like tenants' traffic-pattern driven routing, QoS and traffic isolation. All these features are defined in a topology agnostic manner. On top of that, Mellanox SDN also provides means to define topology-aware as well as topology-agnostic monitoring. This monitoring is provided today via UFM's enterprise-grade GUI and API, which can operate in direct-physical mode or in logical/service oriented mode, a mode that automates fabric provisioning according to business/application requirements. Moreover, UFM has a built-in Logical Model that abstracts the physical network layer as logical entities such as applications, logical services, various network tenants, etc. This provides the end-user with a business-oriented method to manage the network – monitoring and provisioning wise. Backed up by an extensive API, it enables a full set of tools to manage networks in the SDN era and seamlessly integrate in the customers' "big picture" of data center management.

Works Cited

Shenker, Scott, Martin Casado, Teemu Koponen, and Nick McKeown, n.d, T*he Future of Networking, and the Past of Protocols.*, http://www.slideshare.net/martin_casado/sdn-abstractions.

InifniBand Specification, http://www.infinibandta.org/content/pages.php?pg=technology_download

Unified Fabric Manager, www.mellanox.com/ufm

350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com