SOLUTION BRIEF

ASAP2
Accelerated Switching & Packet Processing

Mellanox ASAP^2
Accelerated Switching and Packet Processing

Overview

The pursuit of business agility, operational simplicity and infrastructure efficiency has spurred a slew of new technologies such as cloud, Software Defined Networking (SDN), Network Function Virtualization and Cloud-Native Computing (container-packaged, dynamically managed, and micro-services-oriented computing environment for modern distributed workloads). One of the common themes of these technologies is that all of them are pushing the envelope with respect to how many application instances can be packed efficiently, securely and quickly onto a certain physical infrastructure footprint. The co-existence of multiple micro-services, multiple applications, or even multiple tenants necessitates an efficient underlying network fabric. Virtual switching and routing software normally deployed in servers is one of the key elements of this network fabric, but it is challenged with poor performance and high CPU overhead.

Mellanox Accelerated Switching and Packet Processing (ASAP^2) solution combines the performance and efficiency of server/storage networking hardware along with the flexibility of virtual switching software to deliver software-defined networks with the highest total infrastructure efficiency, deployment flexibility and operational simplicity.

Background and Challenges

Virtual switching was born as a consequence of server virtualization, as hypervisors need the ability to enable transparent switching of traffic between Virtual Machines (VMs) and with the outside world. One of the most commonly used virtual switching software solutions is Open vSwitch (OVS) which is targeted at multi-server virtualization deployments. OVS is commonly deployed in cloud (such as OpenStack), SDN and network function virtualization (NFV) environments which are often characterized by highly dynamic endpoints, as well as high data communication performance requirements. This is especially true for NFV where the applications themselves are the Virtualized Network Functions (VNFs), such as firewall, virtual Evolved Packet Core (vEPC), Deep Packet Inspection (DPI), etc.

Similarly, with the proliferation of server virtualization, especially in hyper-converged infrastructure where network and storage traffic traverse the same network, virtual switching provides a simple solution for VM-to-VM communication, but comes with similar drawbacks and complications as well.

In spite of the flexibility of virtual switches, some of the top challenges they face include:

- Poor I/O performance
- Unpredictable application performance
- High CPU overhead

Poor I/O Performance

Cloud builders are not using the same old servers with 100Mbit/s or 1Gbit/s any more. According to Crehan Research[1], the industry is reaching the inflection point where combined high-speed Ethernet over 10Gbit/s is going to exceed 50% of overall server-class NIC (network interface card) shipments. Cloud service providers are leading the adoption of 25, 40, 50 and even 100Gbit/s server NICs to enhance overall infrastructure efficiency.

With high-speed server I/O, multiple packets can arrive every microsecond and vanilla OVS just can't keep up. Without acceleration, OVS is achieving about 500,000 packets per second (pps) on a 10Gbit/s link where theoretical maximum packet rate can be 15 million pps. In certain real application scenarios where telco VNFs are deployed and traffic is dominated by small voice packets, OVS can only achieve 1/80th of bare metal I/O performance over a 10Gb/s interface.

[1] Server-class Adapter & LOM/Controller Long-range Forecast Tables up to Calendar 2020, published Jan. 2016

Unpredictable Application Behavior

Even before OVS reaches a complete stop and can't forward any further packets, things can slow down significantly. Queues build up, latency skyrockets and packets can be dropped. This, reflected in some real-time applications, such as VoIP, would affect customer experience in the form of sound quality degradation, pauses or dropped calls. This uncertainty in application performance is particularly problematic for cloud and service providers as it means they are unable to deliver on strong service level agreements to their customers.

High CPU Overhead

Once upon a time, all packet processing was done in the so-called slow path (aka CPU) in Cisco's routers. But no router or switch from any reputable networking vendor today is doing packet forwarding with the CPU any more. Instead, packet processing and forwarding are offloaded to a hardware fast path, normally implemented in ASICs or network processors. As the network edge being pushed to server hosts, bare metal servers can achieve much higher packet I/O performance because the majority of packet forwarding can be offloaded to the NIC. However, with compute and network virtualization, because of the path that packets need to traverse within a server, and because packet formats are changed, not all NICs can perform the offload required. When packet processing (including checksum/CRC calculation and encapsulation/de-capsulation) is performed by the CPU, multiple CPU cores now need to shift from application processing to packet processing. This packet processing overhead can make the processors grind to a screeching halt running other workloads. This results in significantly degraded application performance, and reduces the overall efficiency of the infrastructure.

The Solution - ASAP^2

Mellanox supports accelerated virtual switching in server NIC hardware through the ASAP2 feature in ConnectX-5 NICs along with BlueField SmartNICs, and upcoming 200Gb/s NICs. What enables this unique feature is an embedded switch (eSwitch) in the hardware that implements switching between virtual NICs, i.e. vNICs. With a pipeline-based programmable eSwitch built into the NIC, enabling them to handle a large portion of the packet processing operations in hardware. These operations include VXLAN encapsulation/decapsulation, packet classification based on a set of common L2-L4 header fields, QoS and Access Control List (ACL).

Built on top of these enhanced NIC hardware capabilities, ASAP^2 provides a programmable, high-performance and highly efficient hardware forwarding plane that can work seamlessly with the SDN control plane. It overcomes the performance and efficiency degradation issues associated with software virtual switching implementation. The ASAP^2 feature will be further enhanced and broadened in future generations of ConnectX NICs, such as ConnectX-6, and higher.

There are two main ASAP^2 deployment models: ASAP^2 Direct and ASAP^2 Flex.

ASAP^2 Direct

In this deployment model, VMs establish direct access to Mellanox's NIC hardware through Single Root IO Virtualization (SR-IOV) Virtual Function (VF) to achieve the highest network I/O performance in virtualized environment.

One of the issues associated with legacy SR-IOV implementation is that it bypasses the hypervisor and virtual switch completely, and the virtual switch is not aware of the existence of VMs in SR-IOV mode. As a result, the SDN control plane cannot influence the forwarding plane for those VMs using SR-IOV on the server host.

ASAP^2 Direct overcomes this issue, by enabling packet processing operations to be offloaded from the virtual switch to the eSwitch forwarding plane, while keeping intact the SDN control plane. In the case of OVS, the SDN control plane remains the same: forwarding table and policy information are communicated from a corresponding SDN controller through OVS vSwitchd running in user space. Then depending on user-defined rules, the forwarding table entry can be programmed into the ConnectX NIC's eSwitch instead of OVS kernel module.

Virtual machine (VM) instances connect to ConnectX NIC through SR-IOV, and directly send and receive data packets to/from the NIC itself.

ASAP^2 Flex

In this deployment model, VMs run in para-virtualized mode and still go through the virtual switch for its network I/O needs. But through a set of open APIs such as Linux Traffic Control (TC), or Data Path Development Kit (DPDK), the virtual switch can offload some of the CPU intensive packet processing operations to the Mellanox NIC hardware, including VXLAN encapsulation/decapsulation and packet flow classification.

ASAP^2 Solution Benefits

Enhanced Cloud Networking, NFV and Hyperconverged Infrastructure Performance

ASAP^2 Direct offers excellent small packet performance beyond the raw bit throughput. Benchmarks show that on a server with a 25G interface, OVS accelerated by ASAP^2 Direct achieves 33 million packets per second (mpps) for a single flow with near-zero CPU cores consumed by the OVS data plane, and about 18 mpps with 60,000 flows performing VXLAN encap/decap. These performance results are three to ten times higher than DPDK-accelerated OVS. ASAP^2 offers the best of both worlds, software-defined flexible network programmability, and high network I/O performance for the state-of-art speeds from 10G to 25/40/50/100/200G.

Total Infrastructure Efficiency

By letting the NIC hardware take the I/O processing burden from the CPU, the CPU resources can be dedicated to application processing, resulting in higher system efficiency.

Deployment Flexibility

All changes made to support ASAP^2 will be open-sourced and up-streamed into Linux, OVS and other communities. In addition, ASAP^2 is completely transparent to applications – VNFs stay hardware independent and don't need to change. This makes ASAP^2 deployment easy and flexible to use.

EXPLORE FURTHER

Learn Three Ways ASAP^2 Beats DPDK for Cloud and NFV

http://www.mellanox.com/blog/2016/12/three-ways-asap2-beats-dpdk-for-cloud-and-nfv/

Watch ASAP^2 Direct/OVS Offload Demo Video:

https://www.youtube.com/watch?v=LuKuW5PAvwU

Learn more about CloudX™ Benefits

http://www.mellanox.com/solutions/cloud/benefits.php

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