SOFTWARE-DEFINED NETWORKING ON INFINIBAND FABRICS

Ariel Cohen

April 7th, 2016
GOALS

- **Virtual Ethernet networks over IB**
  - Communication between applications which are not IB-aware
  - Communication between IB-connected servers and Ethernet-connected servers
  - Connect to multiple physically isolated Ethernet networks using gateways

- **Security and isolation**

- **Fault tolerance**

- **High-performance**
  - Offloads on the servers, low-latency Ethernet gateway

- **Para-virtualized I/O, SR-IOV, and non-virtualized I/O**

- **Network services: firewall, load balancer, IP routing, VPN**

- **Software-defined**
  - Dynamic creation, removal, assignment, monitoring with a management tool
  - OpenStack support
SOLUTION

- Implement virtual networks as overlay networks on IB
- Encapsulate Ethernet packets in IB messages
  - Use the UD transport (RC an option)
  - Support a large MTU (10KB) on IB to accommodate jumbo Ethernet packets on UD
- Map the broadcast domain of a virtual network to an IB multicast group
- Perform learning and aging in the server drivers and the gateway
- Ability to assign virtual networks to IB partitions
- Provide offloads in the HCA
  - Checksum, large send, RSS/TSS, etc.
- Security enforcement in the HCA
  - Anti-spoofing, L2-L4 filtering
VIRTUAL NETWORK WITH NO UPLINK

Server 1

Bonding

vNIC1  vNIC2

Server 2

Bonding

vNIC1  vNIC2

Server 3

Bonding

vNIC1  vNIC2

IB Multicast Group
VIRTUAL NETWORK WITH AN UPLINK

Server 1
- Bonding
- vNIC1
- vNIC2

Server 2
- Bonding
- vNIC1
- vNIC2

Gateway 1

Gateway 2

Multicast Group 1

Multicast Group 2

Bcast

Ethernet
ADDRESS RESOLUTION

- Server vNIC driver maintains a forwarding table (FT) to map destination MAC+VLAN to IB address handle of another server or gateway
  - The FT is populated based on ingress traffic learning
  - Aging is applied to the FT

- Similarly, the gateway maintains a forwarding table based on packets received from the servers
VIRTUAL NETWORK SERVICES

- On-demand network services
- Multiple services – single virtual appliance
- High availability through VRRP
- Supported Network Services
  - L3–L4 Firewall
  - Static and Dynamic NAT
  - L4 Load Balancing
  - IPSEC VPN
  - Static routing & gateway services
HARDWARE

HCA (EDR – 100Gbs)

Leaf switch (EDR) +
Ethernet gateway
2x40GE ports
8x10GE ports

Fabric Interconnect
12 I/O modules
Quad 40GE (16x10GE) QSFP
Quad 10GBaseT
Network Services Module
FABRIC MANAGER

- Manages physical and virtual network elements
  - Switches and routers
  - Server network and storage interfaces
  - Network services
  - Virtual networks & Fabric I/O

- Single console for fabric administration
  - Provisioning & Orchestration
  - Server I/O, Network Services, and Virtual I/O Profiles
  - Monitoring for Debug and diagnostics
BUILDING BLOCKS OF SDN

Virtual Networks
Overlay isolated Ethernet networks upon the physical infrastructure

Virtual Network Services
Provision overlay networks with Virtual Network Services such as Firewalls, VPNs, and Load Balancers

Virtual Server I/O
Connect VMs to Virtual Networks with virtual InfiniBand, Ethernet, and Fibre Channel adapters

Virtual Fabric I/O
Connect dynamically to multiple physically isolated LANs and SANs using gateways
SECURE MULTI-TENANCY

- Isolated tenant virtual networks
- Network security services protecting tenant domains
  - Virtual firewalls and VPNs
- Isolated administrative domains
  - Cloud tenants are secured from the administrative actions of other tenants
PRIVATE CLOUD APPLIANCE (PCA)
SDN over IB in an Engineered System

Oracle Sun x86 Servers

- Oracle Fabric Interconnect F1-15
- Oracle Sun Data Center InfiniBand Switch 36
- Oracle Switches ES1-24
- Oracle Sun Data Center InfiniBand Switch 36
- Oracle Fabric Interconnect F1-15

Oracle Sun x86 Servers

- Oracle Sun x86 Servers
- Oracle ZFS Storage Appliance
THANK YOU

Ariel Cohen
ariel.x.cohen@oracle.com