White Paper

Centec's SDN Switch – Built from the Ground Up to Deliver an Optimal Virtual Private Cloud

Table of Contents

Virtualization Fueling New Possibilities – Virtual Private Cloud Offerings

Over the past five years, organizations large and small have been moving to the cloud to take advantage of the efficiencies and scalability it offers. It's an attractive value proposition – instead of having to invest in, build out and manage all the infrastructure required to support all the different applications and services an organization needs, they can turn to a provider to quickly and cost-effectively provision it all for them.

The popularity of software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS) solutions is a testament to how eager organizations are to simply consume a service, without having to worry about all the underlying complexity required to enable and support it. This is why organizations are increasingly interested in Virtual Private Cloud (VPC) offerings, as they look to "consume" the underlying capacity they need to deliver all their different business applications and services. Cloud Providers are struggling, however, to deliver VPCs that offer the performance, agility and cost efficiencies they expect from a cloud offering – the challenge is in the underlying network.

Virtualization is the key to enabling these new offerings, but while compute and storage architectures have been virtualized for some time now, the network has lagged behind. Only recently, with the introduction of software defined networking (SDN) and network function virtualization (NFV) is the network able to support the fast provisioning, flexibility and scale of the compute and storage architectures it is connecting. SDN and NFV solutions separate the management and services, respectively, of the network from its traffic forwarding capabilities, however, how it's done can make all the difference in the results.

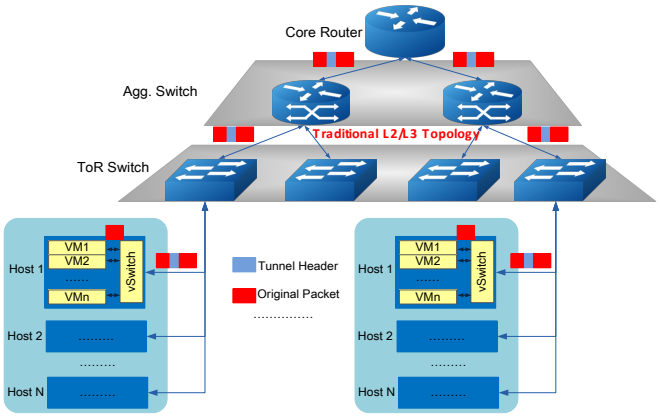
Cloud Providers need to ensure the SDN solution they choose to support their VPC offerings can enable them to quickly and easily isolate tenants and support all the value-added services and applications their customers need to run their business. This paper reviews the different approaches to network virtualization and the advantages of Centec's hybrid switch solution.

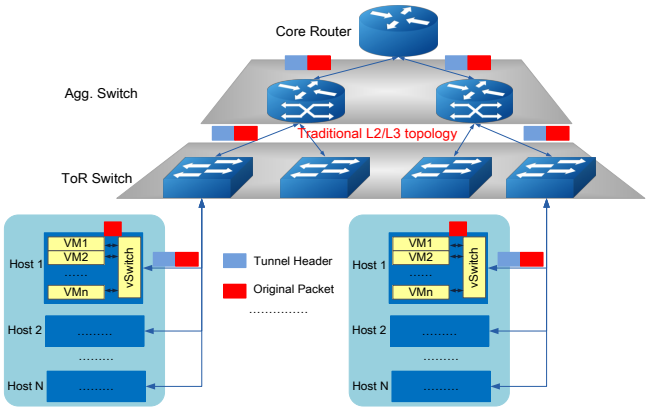
Current Approaches to Network Virtualization Force Compromises

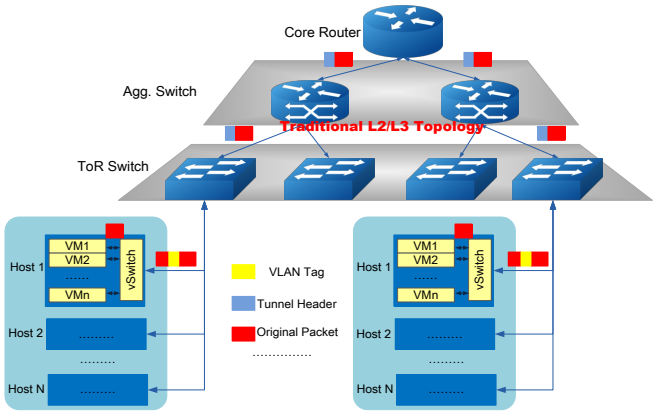
There are a number of ways Cloud Providers can virtualize the network to provision a VPC offering, each with its own benefits and challenges. At a high level, they can be split into two different approaches:

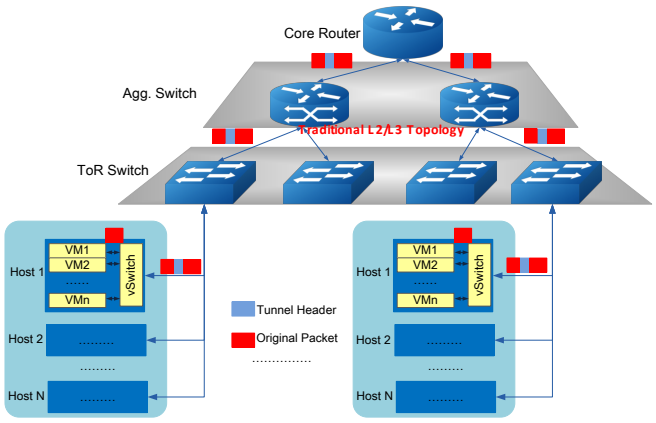
1. **A software overlay approach,** where everything is done via software encapsulation, mostly using virtual switches, and the hardware network is generally unaware of virtual networks.
2. **Direct fabric programming**, where each switch (both hardware and virtual) is programmed directly to operate as they should to handle multiple virtual networks.

There are, of course, various implementations of each approach, which are detailed in the table that follows.

	Overview of Approach	Pros	Cons
Software Overlay			
Virtual Local Area Network (VLAN)	 Traditional implementation – private network is delivered by encapsulating traffic at Layer 2	• Simple deployment and high stability. • Supported by all networking equipment today. • Minimal packet overhead (inserts only 4 bytes of VLAN tag). • Can support both virtualized nodes and physical nodes. • Moderate visibility – original packets visible to management tools.	• Scalability limitations of 4K VLAN. • Difficult to realize the automatic control of the original physical devices. • MAC address of VMs need to be visible to physical switches, hence the table size can become the bottleneck of the network.

Tunnel Overlay based purely on Software	
	Most common implementation - private network created entirely via the virtual switch
		• Hardware independence.
		• High flexibility.
		• Network operation and maintenance team can be separated.
		• High scalability of network.
			• vSwitch easily overloaded – can impact performance.
			  ○ Inefficient lookup on huge flow table.
			  ○ Network Interface Cards (NICs) can't support fragmentation offloading of TCP because doesn't get the complete TCP packets.
			  ○ Each server can be seen as a network node – can overload cloud controller.
			• Poor network visibility- management tools can't see original packets.

Direct Fabric Programming			
Proprietary Hardware Fabric – Tunnel Overlay based on Hardware	 Network seen as a Fabric, connecting all nodes by a tunnel overlay that applies different policies to different access services.	• Excellent performance and high reliability. • High visibility of the network - including original packets. • Good enough scalability and flexibility.	• Closed system – poor compatibility with other devices/vendors. • Extremely expensive to build out.

OpenFlow Network	 Entire network of OpenFlow swtiches (OpenFlow is an open standard that supports innovative routing and switching protocols).	• Excellent performance. • High visibility of the network - including original packets. • Simplified management – both the virtual and physical network is controlled by the same management platform.	• Closed system – poor compatibility with other devices/vendors because not all devices support OpenFlow. • Extensive upgrade – all devices (from aggregation to the core) needs to be upgraded to support OpenFlow. • Labor Intensive – a lot of development required to migrate; a lot of complex rules to manage and maintain.

All of these approaches force some sort of compromise: the software approaches offer a lot of flexibility, but have scale, visibility and performance issues; the direct fabric programming approaches deliver scale, visibility and high performance, but are expensive and inflexible – often locking customers into using a single vendor's solutions.

Requirements for a VPC

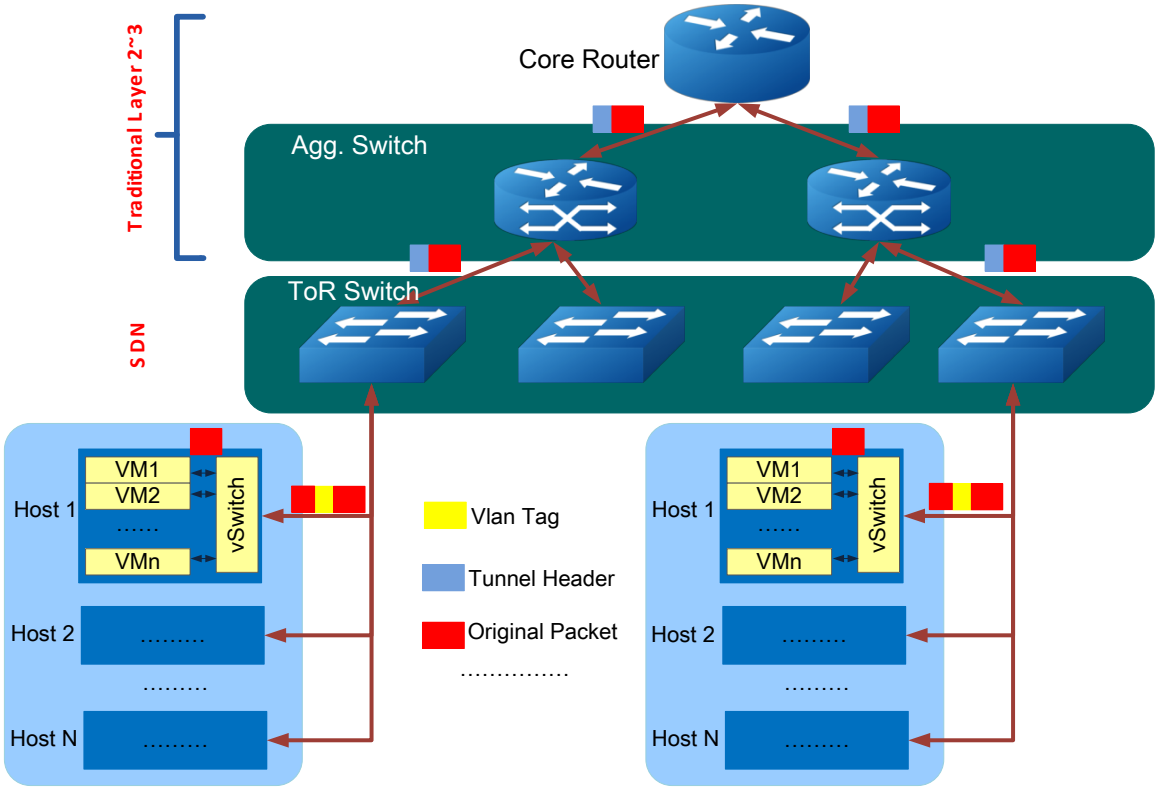
What is really needed is a hybrid solution that combines the best of both software and hardware approaches to make it easy to deploy and manage a solution that deliver:

- **Flexibility**
- **Visibility**
- **High Performance**
- **Scale**

Centec took the challenge and developed a solution that allows Cloud Service Providers to leverage network virtualization in a way that takes advantage of both software and hardware benefits to support truly successful VPC deployments.

Centec – A New Way to Approach VPCs

Centec offers a new way to build multi-tenant networks in a SDN VPC platform that brings the best of both software and hardware worlds together, without compromise, by using a ToR-Offload approach. The ToR-offload hybrid solution combines specialized top-of-the-rack (ToR) switches (hardware) and an integrated OpenStack Cloud Controller (software), to provide the benefits of a unified network fabric – tunnel overlay - that delivers both the performance and flexibility providers need.



The distributed design of the Centec solution offloads many of the process intensive operations from the vSwitch to the ToR switches. The ToR switches operate as super NICs, using Centec's next-generation switching silicon to deliver optimal performance and offload the vSwitch. The Controller, which is a Neutron plug-in, manages all the ToR switches, via a standard OpenFlow interface, treating them as part of a hypervisor.

Centec's ToR-offload solution requires the Open vSwitch instance on the compute node only maintain local VM information to keep its flow table small. All remote VMs not running on the compute node would be managed and maintained by the ToR switch.

The role of the OvS instance is to forward packets within the local system; packets not matching local VMs will be forwarded via the NIC to the ToR switch (pre-fragmenting it if necessary). The ToR switch will use multiple table lookups to perform the following functions:

- Identify the appropriate network instance ID and tenant instance ID
- Rate-limit packets, if required
- Track statistics per VM/network/tenant

- Perform security checks
- Determine whether the packet will be forwarded via Layer 2 bridging or Layer 3 routing
- Find the destination ToR and send it:
    - Directly to that ToR.
    - Via a ToR-to-ToR tunnel. In this instance, the destination switch will conduct similar functions - de-encapsulating the packet and forwarding it to the appropriate destination OvS, which will send it to the right VM.

To support the virtual distributed Layer 3 gateway, the ARP proxy in each ToR switch will tell the VMs that their gateway's MAC address is attached in the ToR switch. This scheme solves the bottleneck issues of Layer 3 gateways in current OpenStack platforms. More information on the multi-table implementation is covered in the Appendix.

Flexible Deployment Modes

The Centec ToR-Offload SDN solution doesn't rely on private protocols or have any special requirements for the aggregation and core devices in the network, so it can easily integrate within any cloud environment. There are two primary ways to deploy the solution, based on a Cloud Provider's requirements. They can deploy using an:

1. **External Controller**: Centec can integrate with any external Controller that is compliant with OpenFlow protocols. With this architecture, each ToR switch is controlled by the Controller, via an OpenFlow interface. The ToR switches are only responsible for forwarding traffic, with the application layer above the OpenFlow Controller taking over operations, such as translating the messages into the flow table.

   Communications between the Controller and ToR switches include OpenFlow and open virtual switch database (OVSDB) messages, which set up the private network (tunnels). Only Query messages from the OVS Agent in the compute node are transmitted to the Centec Plug-in to lookup the VLANID/TunnelID. Except that, there is no any other network message between compute node and the Centec Plug-in.

Powered by Centec's next-generation switching silicon, with innovative N-Flow™ architecture to deliver:

- Application-oriented flow tables, with programmable match fields and actions
- 32K flows, without expensive, inefficient external TCAM
- Interoperability with leading SDN controllers and switches, due to OpenFlow support

2. **Internal Controller**: In this architecture, the ToR switches take on some networking operations to reduce processing required by the control nodes. The Cloud Platform sends abstract data to the ToR switches, which have a Cloud Agent, with a DHCP Proxy dedicated to translating all the broadcast DHCP packets into unicast packets to be sent to the DHCP Server. They also have an integrated ARP Proxy responsible for replying to the ARP requests from VMs to significantly reduce the amount, and impact, of broadcast traffic on the network nodes. There is another purpose about ARP Proxy, which can be used to assist virtual distributed l3 gateway support.

The Centec Difference

The Centec ToR-offload solution uses tunnels to abstract the network and easily isolate tenants, while offloading the compute intensive operations (e.g. encapsulation/de-capsulation) to the ToR switch to ensure optimal performance. By reducing pressures on of the control plane, the network can be quickly scaled to address changing needs. The solution offers:

- **Flexibility** – requiring no special networking equipment, beyond the ToR switches, and supporting open, standards based protocols, ensures the solution can easily integrate into a variety of environments and meet the specific needs of different cloud deployments.

- **Simplified Management**- Abstracting the complexity of the network (including table mappings) enables Cloud Providers to quickly and easily manage the VPC offerings and roll out value-added network services to meet customer demands. It provides visibility into the network (including the original packets) to make it easy to understand what is going and quickly troubleshoot any issues.

- **High Performance and Scale** – offloading intensive network processing functions, such as flow table lookup, tunnel encapsulation/de-capsulation, and quality of service (QoS) to ToR switches using Centec's next-generation silicon ensures optimal performance and scale. By minimizing broadcast, multicast and unknown unicast traffic throughout the network, the solution eliminates unnecessary bandwidth consumption and potential disruptions to maximize uptime.

- **Lower Costs** –the solution is easy to deploy – there is no need to change the original network architecture – and manage, leveraging a cost-effective blend of hardware and software that drives down both CAPEX and OPEX costs.

Conclusion

The Centec hybrid ToR-offload solution gives you the flexibility of a software overlay, with the performance of hardware. Abstracting the complexity, the Centec solution makes it easy for Cloud Providers to deploy and manage VPC offerings for their customers. With OpenFlow ToR switches the solution performs all the necessary acceleration functions in hardware and manages all the coordination and complexity to make it easy to make adjustments as needs change. With Centec, Cloud Providers can support their VPC offerings, quickly and easily isolating tenants and supporting all the value-added services and applications their customers need, to increase satisfaction and loyalty and drive new and recurring revenue.

Contact Information

Address: Suite 4F-13/16, Building B, No.5 Xing Han Street, Suzhou Industrial Park, Jiang Su Province, China,
Postal Code: 215021
Tel: 86-512-62885358
Fax: 86-512-62885870
Business: sales@centecnetworks.com
Support: support@centecnetworks.com

Appendix

Centec's use of its own merchant silicon in its switches allows it to support advanced capabilities that enable new features in today's data centers. Unlike typical merchant silicon solutions, Centec is able to provide lower-cost solutions that support advanced features, such as multi-level flow tables that scale to meet a cloud service provider's needs. Here we cover some of the highlights of Centec's hybrid SDN solution for VPCs and provide more details around flow table management and controller orchestration.

Multi-Level Flow Table

Centec supports customization, with programmable match fields and programmable actions to meet different requirements. This example depicts the multi-stage flow table, which works to provide 10K tenants up to 32K VMs.

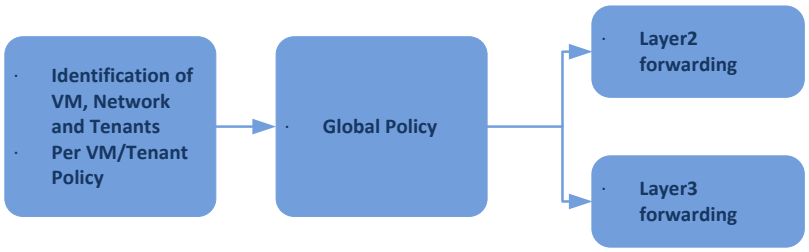


Table ID	Feature
0	• VM Identification • Network Identification • Tenant Identification • Tunnel Identification • Statistics and QoS operation-based on VM/Tenant
1	• Security or QoS based on VM, subnet or Tenant • To determine the next stage Table -Table 2 or Table 3
2	• Layer2 forwarding
3	• Layer3 forwarding

Operational details:

a) On packet entry into the ToR, the switch will perform lookup in the 1st flow table (table ID #0) with the look up key of (a) VLAN (identifying the tenant), (b) MACSA (MAC Source Address which identifies both the VM and tenant) or (c) port + MACSA (identifying the VM and tenant by checking the VM and port bindings).

b) The metadata of the lookup result will be used as network instance ID and tenant instance ID. The lookup result may include a rate-limit pointer and a statistics pointer to perform rate-limit or collect statistics per VM/network/tenant as well as perform a per VM/network/tenant security check.

c) In the next step, the look-up operation is performed on the 2nd flow table (table ID #1). Global network security checks are performed. As well, the look-up result will decide that whether Layer 2 bridge forwarding or Layer 3 routing will be performed. If it's Layer 2 case, it will go to the 3rd flow table (table ID #2). Otherwise the 4th flow table (table ID #3) will be used for the next step.

d) After the look-up operation on the 3rd or 4th level of flow tables, the packet may be sent to a remote ToR switch via a ToR-to-ToR tunnel. If needed, the packet will be encapsulated in the tunnel and sent to the physical network. Before the encapsulation, the original VLAN tag will be removed if it ever existed.

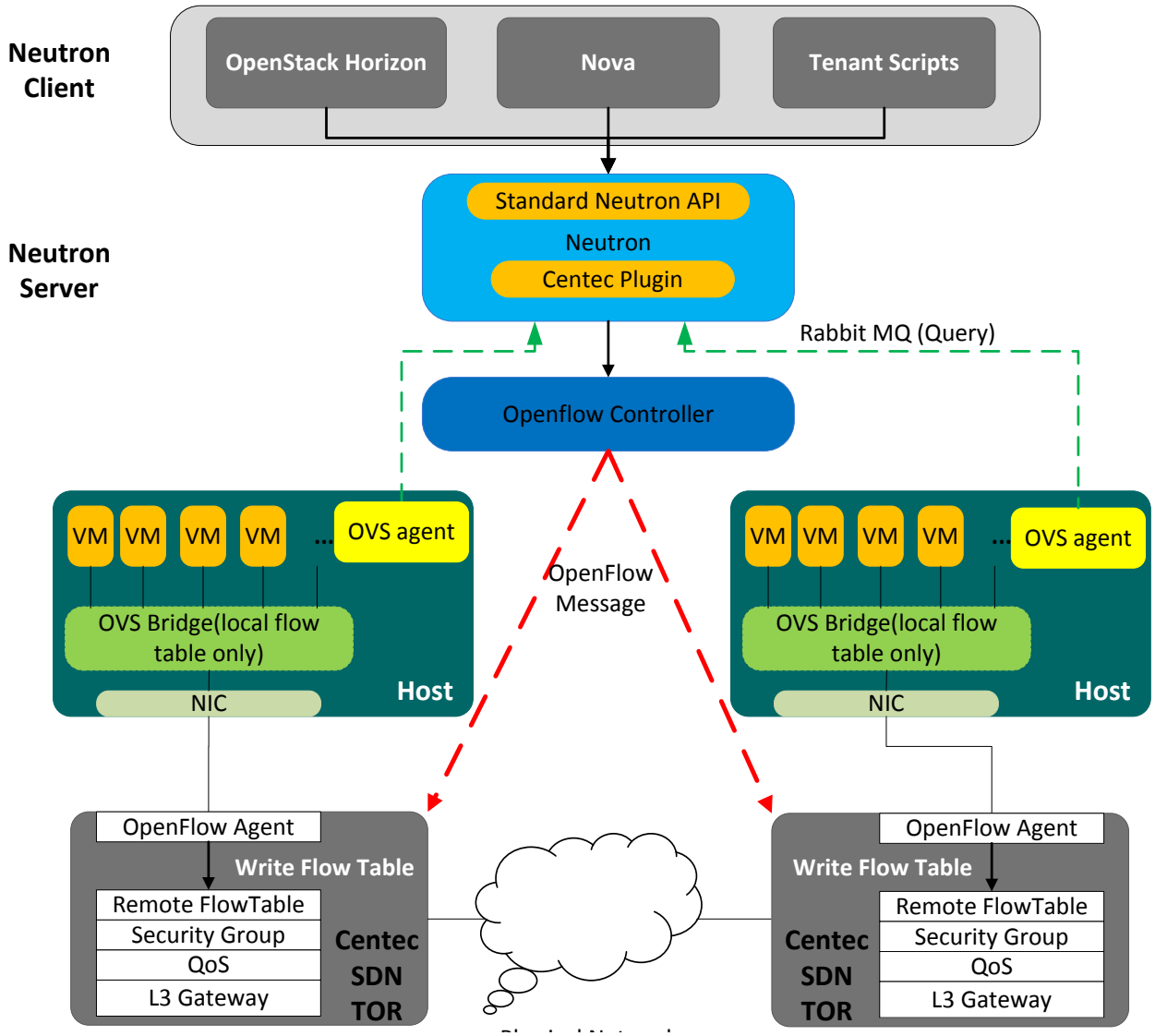
e) After the encapsulated packet arrives at the remote ToR switch, the remote ToR switch will look up the 1st flow table with a key of the tunnel IP (source+dest) and encoded tunnelId. Then the network instance ID and tenant instance ID will be retrieved. The same process will be applied to

the next 2-4 flow tables. After the whole process, the tunnel header will be stripped. Optionally, a VLAN tag may be inserted.

f) Finally, the inner packet will be sent to the target compute node. Inside the target compute node, the packet will be sent to the final VM after the look-up operation in the internal flow table of OVS.

External Controller Architecture

The solution can be integrated with an external Controller, which manages the ToR switches via an OpenFlow interface. The following depicts this architecture:



Internal Controller Architecture

When deployed using the internal Controller, the ToR switches are used to offload network operations and minimize broadcast messages. The following depicts this architecture: