OPEN NETWORKING REVOLUTION

Aftab Siddiqui
Internet Society
Technical Engagement Manager - Asia Pacific
siddiqui@isoc.org
Open Compute Project

OCP is a collaborative and community focused initiative on redesigning hardware technology to efficiently support the growing demands on compute infrastructure.

Projects:
• Storage
• Networking
• Server Design
• Open Rack
• Certification
• Hardware Management
• Data Center
# Traditional Networking

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>Feature 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proprietary Network OS</td>
<td>• Proprietary Features</td>
</tr>
<tr>
<td>Proprietary System</td>
<td>• Few APIs Available</td>
</tr>
<tr>
<td>Proprietary Silicon</td>
<td>• Locked-in and complex support models</td>
</tr>
<tr>
<td></td>
<td>• Mostly proprietary ASICs</td>
</tr>
</tbody>
</table>
### Open Networking

<table>
<thead>
<tr>
<th>Operating System</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Operating System Icon" /></td>
</tr>
</tbody>
</table>
| • Cumulus Linux  
• PicOS  
• IPIInfusion  
• SwitchLight OS  
• MS Sonic  
• FBOSS  
• Etc….  |

<table>
<thead>
<tr>
<th>Open Standard Hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Open Standard Hardware Icon" /></td>
</tr>
</tbody>
</table>
| • Dell ON  
• HP Altoline  
• EdgeCore  
• Mellanox  
• Etc…..  |

<table>
<thead>
<tr>
<th>Merchant Silicon</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Merchant Silicon Icon" /></td>
</tr>
</tbody>
</table>
| • Broadcom  
• Spectrum  |
### What are my options?

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Model</th>
<th>ASIC</th>
<th>Ports</th>
<th>CPU</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell</td>
<td>S4048</td>
<td>Trident II</td>
<td>48x10G SFP+ 6x40G QSFP+</td>
<td>Intel Atom C2338</td>
<td>2GB</td>
</tr>
<tr>
<td>Dell</td>
<td>S6000</td>
<td>Trident II</td>
<td>32x40G QSFP+</td>
<td>Intel Atom S1220</td>
<td>4GB</td>
</tr>
<tr>
<td>Dell</td>
<td>Z9100</td>
<td>Tomahawk</td>
<td>32 x Multirate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accton/EdgeCore/HP</td>
<td>5712</td>
<td>Trident II</td>
<td>48x10G SFP+ 6x40G QSFP+</td>
<td>Intel Rangely C2538</td>
<td>8GB</td>
</tr>
<tr>
<td>Accton/EdgeCore/HP</td>
<td>6712</td>
<td>Trident II</td>
<td>40G x 32</td>
<td>Intel Rangely C2538</td>
<td>8GB</td>
</tr>
<tr>
<td>Mellanox</td>
<td>SN2700</td>
<td>Spectrum</td>
<td>32 x Multirate</td>
<td>Spectrum</td>
<td></td>
</tr>
<tr>
<td>Mellanox</td>
<td>SN2100</td>
<td>Spectrum</td>
<td>16 x Multirate</td>
<td>Spectrum</td>
<td></td>
</tr>
</tbody>
</table>
Facebook 6-Pack….too complicated?

Facebook 6-Pack Modular Chassis

Fabric Card

Line Card
[16 x 40G]
Network Operating System

- Hardware specification of switches are more or less similar but tough task is to pick the right Network Operating System

- Best way is to find major requirement and check the features (no brainer)

- If there is any virtual environment available for those NOS then deploy and test.

- CumulusLinux and IPInfusion both provide virtualized versions of their NOS and without any significant restrictions

- CumulusVX is the most user friendly VM available to-date. It supports all major environments

- All NOS are evolving quite rapidly, keep checking.
All the Network Operating Systems are based on Linux and therefore offer the same kind of CLI... CumulusLinux, OcNOS, ONL, Dell OS10 etc
Every NOS offer different port naming convention

Some NOS offer utilities to simplify command line config such as auto complete
First Deployment Experience

**Target:** Establish Connectivity between multiple PoPs.

**Media:** Dark Fibre

**Service Offering:** Backhaul (Layer 2) and Transit services.

**Budget:** Challenging (we were forced to look into white box switches)

**Project Timelines:** Weeks rather months

**POC:** All services (layer 2 and transit) from 4 PoPs
First Deployment Experience

- Network Design [Dell S4048-ON switches, CumulusLinux], VXLAN overlay
Not that Simple

Troubleshooting and configuration isn't very simple.

Solution: Ansible
Basic Automation

Variables Required:

- **hname** = HostnameA
- **swp** = Number of Interfaces
- **lip** = Loopback IP
- **bint** = Bond interface name (e.g. bond0)
- **bslaves** = Member of bond interfaces (separate with , or space)
- **bip** = Bond IP
- **localasn** = Local ASN (Eg: 420000XXXX where XXXX is POP-ID)
- **remotehname** = HostnameB
- **nip** = Neighbour IP for BGP (e.g. 30.10.0.2)
- **remoteasn** = Remote ASN (eg: 420000XXXX where XXXX is POP-ID)
- * - Populated Automatically on selection of Remote Host
Standard Configuration

- Create User Credentials
- Enable Routing (e.g. Quagga)
- NTP
- DNS
- MOTD
- SWP Interfaces
Basic Automation

**POP-A:**
- auto swp4
- iface swp4
- mtu 9216
- auto vni-600030
- iface vni-600030
- vxlan-id 600030
- vxlan-local-tunnelip 30.30.0.1
- vxlan-remoteip 30.10.6.1
- auto br-vl104
- iface br-vl104
- bridge-ports swp4.104 vn1-600030
- bridge-stp on

**POP-B:**
- auto swp4
- iface swp4
- mtu 9216
- auto vni-600030
- iface vni-600030
- vxlan-id 600030
- vxlan-local-tunnelip 30.10.6.1
- vxlan-remoteip 30.30.0.1
- auto br-vl104
- iface br-vl104
- bridge-ports swp4.104 vn1-600030
- bridge-stp on
First Deployment Experience

• Whitebox Switches are good even in enterprise and ISPs as well. You do need 10G/40G 25G/50G switches in your network.

• CumulusLinux worked well for “almost” everything we needed But review your requirement before selecting NOS.

• VXLAN can solve many problems to help de-clutter layer 2 network. MTU can be a killer though (50 extra Bytes to accommodate) and it breaks LACP and LLDP.

• Operationally simple and economical deployment IF you have proper automation.
Thanks

Any Questions
First Deployment Experience

• VXLAN Packet Captures

- Frame 13: 209 bytes on wire (1672 bits), 209 bytes captured (1672 bits) on interface 0
- Ethernet II, Src: Dell_01:4b:30 (14:18:77:01:4b:30), Dst: Dell_01:5d:30 (14:18:77:01:5d:30)
- Internet Protocol Version 4, Src: 1.1.1.1, Dst: 2.2.2.2
- User Datagram Protocol, Src Port: 37103 (37103), Dst Port: 4789 (4789)
- Virtual eXtensible Local Area Network
- Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.2
- Internet Control Message Protocol

- Frame 2: 568 bytes on wire (4544 bits), 568 bytes captured (4544 bits) on interface 0
- Internet Protocol Version 4, Src: 30.0.0.1, Dst: 30.0.0.2
- User Datagram Protocol, Src Port: 10123 (10123), Dst Port: 4789 (4789)
- Virtual eXtensible Local Area Network
- 802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 100
- Internet Protocol Version 4, Src: 60.0.0.2, Dst: 60.0.0.1
- Internet Control Message Protocol