PureApp System - Connecting to Customer's Network

Disclaimer

The information presented here is not meant to be exhaustive, and it does not claim to be current. Rather, it is meant to provide a beginner an overview of the interconnection between the PureApplication rack and the customer's network and the various networking technologies involved in the interconnection. Network misconfiguration can lead to an unwelcome disruption to the customer's network. Seek the guidance of IBM experts and experienced staff-members if the information collected through the TDA document is either incomplete or does not provide all the information that you need as an installer of the PureApplication rack at a customer's data-center. The installer of the PureApplication rack must be confident that the risk of customer-network disruption is low enough for one to proceed with the SGEN process.

(The latest TDA template can be found at: <insert>)

Introduction

Installation of PureApplication System rack at the customer's site primarily involves connecting the rack to the customer's computer network at the site. Both an understanding of networking in general and an understanding of how the PureApp rack connects to the customer's network is essential to being successful at this task. It is hoped that the information on this page will provide the requisite background knowledge needed.

PureApp Connections to Customer's Network

External connections from the PureApp rack fall into two types, based on primary use:

- Customer Management Network: The PureApp UI is accessed through this network connection. All user interaction with the System Console and the Workload Console is through the Customer Management Network.
- Customer Data Network: Virtual machines of all deployed workloads exist on the Customer Data Network. User transactions with all virtual applications exposed through the workloads is through the Customer Data Network.

Management Network has low network traffic. Data Network traffic will be generally higher, with volume being dependent on transaction volumes. The recommendation is for these two networks to be logically separate (different VLAN) and also physically separate (connecting to different network devices). Such an isolation is common in enterprise networking, and will be available at most if not all of our customer's data-centers.

Fault-tolerance, High availability

Generic understanding of redundancy and high-availability in networking is necessary before delving into the details of Management and Customer networks.

PureApp racks use a pair of TOR switches for the rack, these are BNT G8264 switches. Download a copy of the Application Guide for the G8264, its a good reference for many of the terms one will encounter. http://www.bladenetwork.net/userfiles/file/G8264_AG_6-8.pdf

Switch Redundancy and Spanning Tree Protocol (STP)

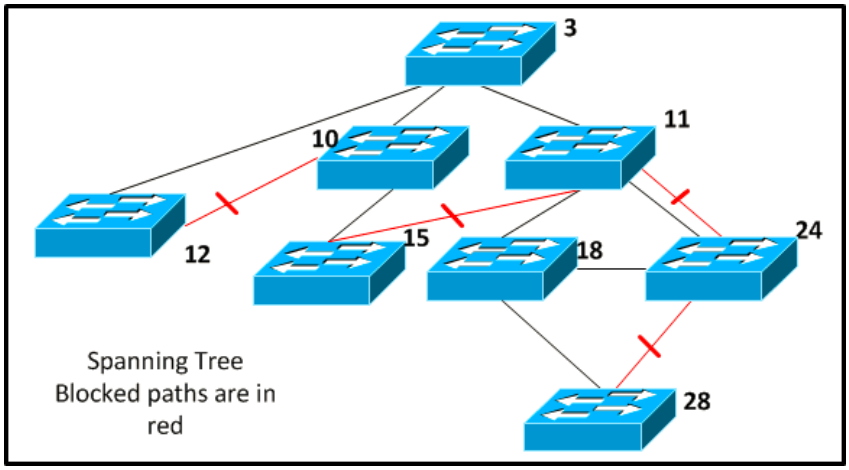

Diagram 1: Generic Switch Network

Diagram 1 is a generic switch-network diagram with multiple paths being set up between switches for redundancy and fault tolerance. STP (Spanning Tree Protocol) is enabled on the switches and this prevents network loops from happening by blocking all but one of the redundant paths. If an active path fails for any reason, the blocked path becomes active. If Switch 10 fails,the currently blocked path from Switch 11 to Switch 12 becomes active, and traffic from Switch 3 can still reach switch 15. Also note that only one of th etwo links between Sitch 11 and switch 24 is active.

Look for more information on the internet to learn about bridge priority, and BPDU (Bridge Protocol Data Unit), an other details of Spanning Tree Protocol.

STP has been enhanced from the original specs to RSTP (Rapid STP), which introduced algorithms and behavior for faster convergence in the event of topology changes. Cisco introduced PVST (Per VLAN Spanning Tree) and PVST+ (Per VLAN Spanning Tree Plus), these are proprietary to Cisco. The BNT G8264 TOR Switches offer PVRST (Per VLAN Rapid Spanning Tree) which is fully interoperable with Cisco's PVST+.

Link Aggregation Group (LAG)

As seen in the discussion above, while multiple paths between the same network switches can exist, STP allows only one of those to be active. While allowing fault-tolerance against port failure, throughput is limited to what one link can provide. Link aggregation provides a means of bundling or aggregating multiple network connections together to provide higher throughput as well as fault-tolerance.

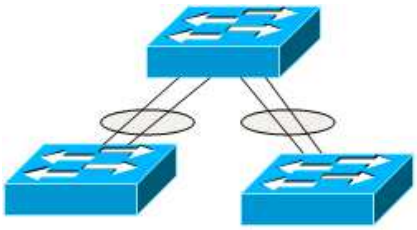

Diagram 2: Link Aggregation

PureApp rack TOR switches support two types of link aggregation: Cisco's Etherchannel (proprietary) and IEEE LACP (Link Aggregation Control Protocol). LACP sends LACP packets down all the links that have the protocol enabled. If a device on the other end of the link also has LACP enabled, the two units detect multiple links between themselves and combine them into a single logical link. Failover is also included in LACP. Because of these features, LACP is considered a "dynamic" link-aggregation protocol.

Muli-Chassis LAG (MC-LAG)

Multi-chassis Link Aggregation Group is a generic term for a LAG whose links are connected to different nodes (network devices) which are interconnected using some sort of peering protocol that allows the MC-LAG feature. While a LAG provides fault-tolerance for single-link failure, an MC-LAG provides fault-tolerance in the event of node failure.

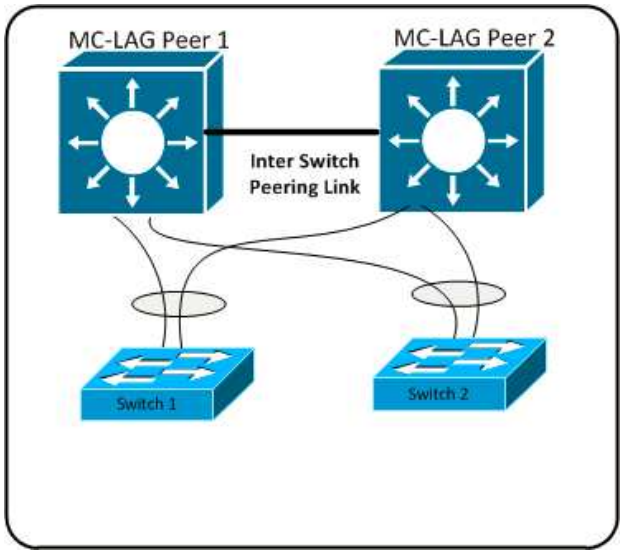

Diagram 3: MC-LAG

In Diagram 3 above, a failure of one of the MC-LAG peer devices does not affect network connectivity for devices connected to the access switches.

The mechanics of peering between the peered switches is implemented differently by different vendors: Cisco VSS, Cisco vPC, BNT VLAG, HP IRF etc. are some examples. Hence, the peered switches must be identical and from same vendor for the peering to work. The aggregated links can interconnect between different equipment, provided they run the same protocol (Etherchannel, LACP).

VLAN, Access, Trunk, Native VLAN, Tagging etc.

An excellent discussion on the topics listed in the heading of this section can be found at http://www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/configuration/guide/cli_rel_4_0_1a/AccessTrunk.html. A summary is presented here.

VLANs

A router limits broadcasts to transmission within the subnet originating the broadcast. A LAN (Local Area Network) is by definition all devices in a broadcast domain. With no VLANs defined, a switch places all its ports into a default VLAN-1 and forwards broadcasts onto all its ports.

A VLAN (Virtual LAN) creates a broadcast domain typically made up of one or more ports of a switch. The physical switch is divided up into multiple "virtual" LANs or VLANs.

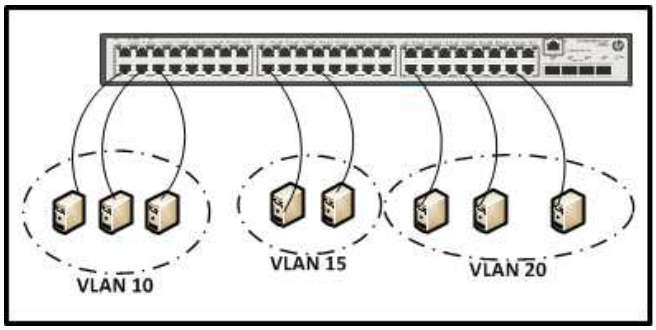

Diagram 4: VLANs

Trunk Port, VLAN Tagging

VLANs can span across switches, to create logical workgroups that are physically dispersed. To do this, the inter-switch link must allow the traffic from different VLANs to flow between the switches.

An identifier is required for each packet that traverses the VLAN trunk, so that the switch can correctly determine which ports to forward each packet to. "VLAN Tagging" is the mechanism used for this, IEEE 802.1q is the applicable standard. A "tag" is inserted into the frame header of each packet. The tag contains information about the specific VLAN that the frame and packet belong to.

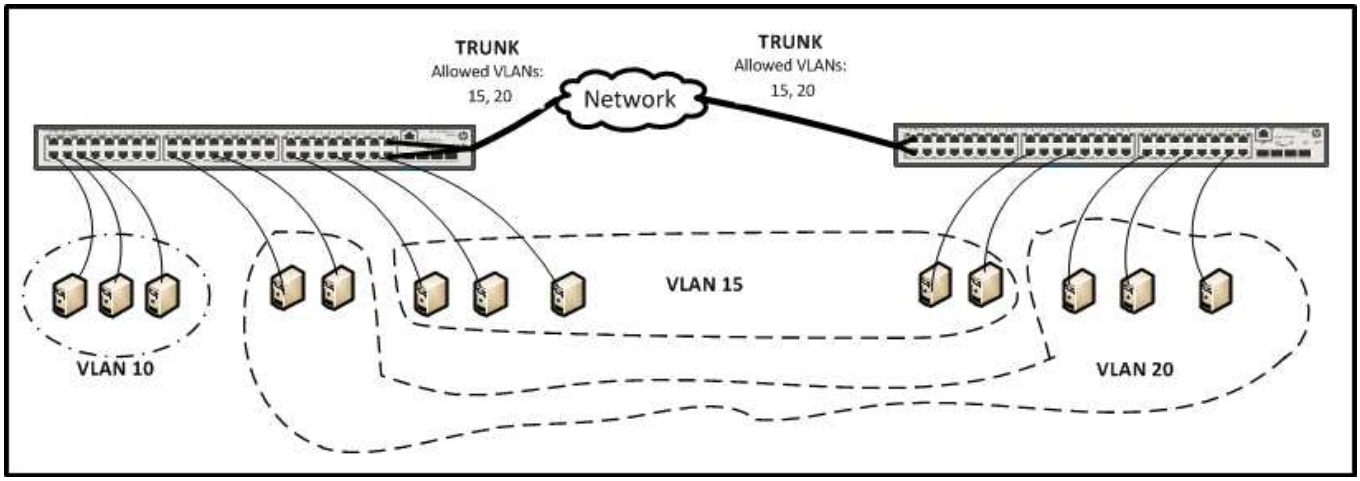

Diagram5: VLAN Trunk

Thus, a port configured as mode "trunk"

- carries traffic of one or more VLANs
- is configured with a list of "allowed" VLANs
- all traffic carried on the trunk are "tagged" (except for Native VLAN packets, discussed below)

Native VLAN

Layer 2 devices need to have a "default VLAN" to assign to their ports, including their management port(s). Certain Layer 2 protocols such as CDP, PAgP, and VTP needed to be sent on a specific VLAN on trunk links. VLAN 1 was chosen for these needs. By default, all Ethernet interfaces on Cisco switches are on VLAN 1, unless configured as a member of a user VLAN.

A trunk link will tag all frames from all VLANs with the 802.1q (or ISL) tag, except for frames belonging to VLAN 1. By default, frames from VLAN 1 belong to "native VLAN", and are carried across the trunk untagged.

The IEEE 802.1q committee decided to to support the so-called native VLAN, a VLAN that is not associated explicitly to any tag on an 802.1Q link. This VLAN is implicitly used for all the untagged traffic received on an 802.1Q capable port, and will be carried untagged across the trunk. On trunk links, for security reasons, it is best-practice to set the "native VLAN" explicitly to something other than VLAN1, and ensure that this "native VLAN" is not a user-VLAN anywhere in the network. When the "native VLAN" is changed to be other than VLAN1, control traffic like CDP, PAgP, DTP etc will still be carried on VLAN1, but as tagged traffic.

Access Port

An "access" port can carry traffic for only one VLAN, and has only one VLAN configured for the interface. If the port receives any packets that have a VLAN tag that is not the access port's VLAN value, the packet is dropped.

In Diagram 5 above, all ports except the trunk are access ports.

Specific Customer-Networking Scenarios

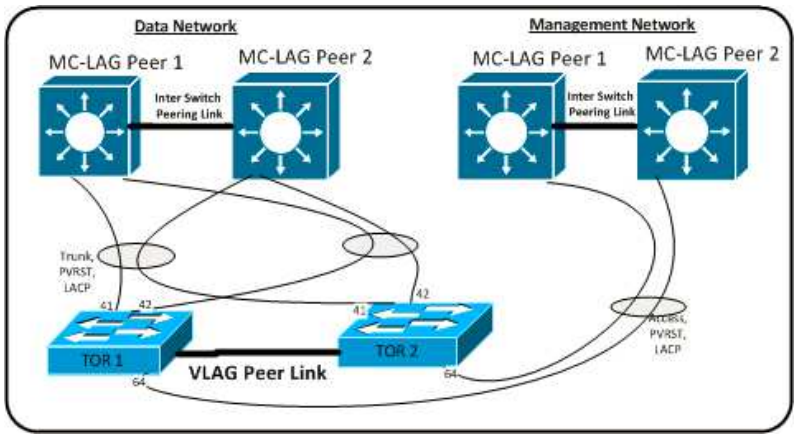

Diagram 6: Recommended Network Connection

The recommended connection to customer's network is shown in Diagram 6 above. This is not always available, and there are occasions where we have had to do manual configuration changes to the TOR switches after SGEN has completed its automated run. A few of these are detailed below.

HSRP

HSRP (Hot Standby Router Protocol) is a Cisco-proprietary protocol that provides router redundancy in the event of the primary router becomes unavailable. As the name states, it is a "hot standby" arrangement with only one router of the HSRP pair being "active" for a given route.

IPAS v1.1.0.0+cFix2 is the minimum IPAS version required to support a customer who is wants to connect PureApp rack to her/his HSRP pair.

Specific configuration as documented in **HSRP Configuration** is required. For later versions of IPAS, check with IPAS networking experts on any special configuration.