Reason for Outage
November 13th, 2019

Services Impacted:

The peering fabric was unstable and experienced a network loop on November 13th, 2019 between 11:32 to 12:51 (local Pacific time). The looping traffic caused broadcast *(IPv4 ARP)* & multicast *(IPv6 ND*) protocols to increase to an abnormal PPS level *(in excess of 500Mbps)*. This spike in control plane traffic impacted a portion of participant attached routers, in some cases triggering vendor default DDoS protection mechanisms. These protection features *(in their default unconfigured state)* caused participant control traffic to be dampened, in some cases disabling internal IGP and/or transit routing.

Background Information:

SFMIX has been migrating from a traditional layer 2 STP design to a modern layer 3 VXLAN architecture. The vast majority of IXP's which are multi-site utilize VXLAN or MPLS technology. The intention of the VXLAN protocol *(within the IXP context)* is to provide a loop-free switch fabric, facilitate additional transport between sites in a dynamic fashion, and provide more granular traffic engineering *(such as adjusting an IGP cost to "steer" traffic in a particular path).*

These migrations have been phased over this year, based on both financial and volunteer labor availability. This Monday *(November 11, 2019)* marked the final site receiving an upgrade to the new VXLAN-capable 100Gbps Arista switches - 365 Main St. The remaining locations were upgraded earlier in the year: 48233 Warm Springs Blvd & 200 Paul Ave *(March 30, 2019)*, along with 11 Great Oaks Blvd *(July 30, 2019)*.

Where possible, SFMIX strives for a consistent per-site deployment model. This includes multiple transport circuits to other SFMIX deployment locations, an Arista 7280SR-48C6-F 100Gbps VXLAN-capable switch, out-of-band management network, remotely manageable PDU, etc. The network loop event occurred while bringing the 365 Main location up to this parity, by enabling a secondary transport circuit back to 200 Paul. All of the other sites have already had multiple transport circuits present for several years.

Current Operational Status:

The peering fabric has been restored to “normal” operating state *(as of 12:51 Nov 13th)*.

Root Cause Analysis:

The problem manifested during the turn-up of an additional transport link between 365 Main and 200 Paul. Previously the configuration between these locations was a single 10Gbps link, operated in L2 STP trunk mode *(to facilitate connectivity to a non-VXLAN capable switch).* When the new VXLAN capable switch was installed, the L2 STP configuration was preserved to minimize participant traffic impact. An additional 10Gbps transport circuit was turned up in L3 VXLAN mode. The VXLAN VNI and VLAN 1q identifiers were the same, thus a network loop was inadvertently triggered. This was due to the same L2 path being made available with competing forwarding mechanisms.

The intention of the secondary transport turn-up was to facilitate a migration with no impact to participants. This approach of migrating VXLAN circuits from one interface to another had been done successfully in Fremont without impacting participant traffic *(this was done to free up SFP+ ports for a QSFP+ breakout cable, as that switch is rapidly becoming populated).* However, in this case, one of the interfaces was in a non-VXLAN mode and the configuration deployed was not fully understood. The loop was created inadvertently.

Reports came very rapidly via email, phone, Slack, and SMS - the event initially appeared to be more widespread than just SFMIX. For example, participants noted that their transit connectivity went down from ports not attached to SFMIX. During the triage process, it was discovered that SFMIX participants who had routers with DDoS protection mechanisms enabled *(default for at least Juniper MX and likely other platforms)* would automatically dampen their non-SFMIX control plane traffic *(such as BGP, OSPF, etc)* due to the loop. This would impact downstream customers and/or internal IGP traffic.

An initial assumption was made that a physical layer issue may have caused the issue *(due to the widespread non-SFMIX transit ports losing BGP sessions)*, however none of the SFMIX transport ports showed any link status change - nor did any provider confirm a physical layer issue. Additionally, participants consistently reported a single participant originating ~500Mbps of control plane traffic *(ARP and ND packets)* and it was initially assumed this was a “bad actor” router. It’s unclear why this particular participant was originating such high traffic levels - it’s now believed this is a false positive due to the network loop.

Log messages on the 200 Paul and 365 Main switches indicated that a loop was present between the newly configured and existing transport links between these sites. The original L2 STP link was converted quickly to the newer routed L3 VXLAN configuration to resolve the issue.

Event Timeline:

Please note that all times listed in the timeline below are in 24-hour clock format, and refer to Pacific Daylight Time - on November 13, 2019.

11:32	Secondary transport link between 365 Main and 200 Paul enabled, beginning of outage event.
11:35	Participant shared logs indicate that their Juniper MX router (attached to SFMIX) enabled DDoS protection at this time, killing BGP sessions to both their peers/transits along with IGP.
11:58	Voicemail received from first member noting that all SFMIX peering sessions were lost. Initial focus on physical layer and originated high pps control plane traffic, assumption was on physical and/or bad acting routing - not clear yet it was loop issue.
12:51	Peering fabric stabilized, both 365 Main transport links running in L3 VXLAN mode - no further `%ETH-4-HOST_FLAPPING` log messages.
12:56	Began interactive triage with participants present in Slack channel
13:49	Larger participant audience (entire members mailing list) sent Slack invitation link to collaborate with peers in real-time.

Corrective / Preventive Actions:

We will be adjusting the existing storm-control settings *(both broadcast & multicast traffic)* from 5% to 0.1% on all participant interfaces *(the PPS syntax is not supported on our switch platform)*. This adjustment will be done in an advertised maintenance window. One participant is already operating in this configuration *(at their request to trial the settings)* without reported issues.

SFMIX will develop a more rapid “heads up” mechanism to quickly confirm correct operation of the SFMIX fabric. This would include link state of all the transport links, alarming for participant ports in high PPS counters for non-unicast traffic, abnormal syslog entries, etc. Currently triage is done using LibreNMS *(at least 5mins lagged due to sampling interval)* and CLI commands, and neither approach is real-time or an aggregate of all the systems that need to be verified in a quick fashion.

Participate in industry forums (such as Euro-IX) to develop a “Precautions and protection mechanisms when connecting to an IXP” BCP document. Such examples might include utilizing a dedicated peering router, purposely setting DDoS protection settings (example syntax), implementing policiers/rate limiters facing IXP ports, etc.

- BGP Best Practice at IXPs
- [j-nsp] Syslog getting spammed by DDOS_PROTOCOL_VIOLATION_SET
- https://www.reddit.com/r/Juniper/comments/b1i08j/ddos_protection/

Migrate to a different mailing list platform. The current members mailing list is a Google Group *(in the interest of less software management overhead).* Unfortunately the Groups mailing list doesn't allow adding other Google Apps Groups, which several members utilize for their own distribution list management. Additionally, an audit needs to be performed to confirm all active participants have valid email addresses subscribed on the mailing list.