Traffic Load Balancing Techniques

Dharamjeet Brar

University of Alberta

Juned Noonari

Table of Contents

PROBLEM **5**

DISCUSSION **7**

LAG 7

LOOP PREVENTION MECHANISMS 9

G.8032. ..... *11*

FIRST HOP REDUNDANCY PROTOCOLS 16

INTERIOR GATEWAY PROTOCOLS 19

BORDER GATEWAY PROTOCOL 22

BGP Ingress load sharing. *22*

BGP Egress load sharing. *28*

LAB DEMO **32**

CONFIGURING LAG 34

CONFIGURING G.8032 35

CONFIGURING CORE ROUTERS 37

CONFIGURING VRRP 39

CONFIGURING BGP LOAD SHARING 40

BEST-CASE SCENARIO 45

WORST-CASE SCENARIO 46

CONCLUSION **47**

FUTURE WORK 48

Table of Figures

FIGURE 1: LAG EXAMPLE 7
FIGURE 2: MC-LAG EXAMPLE 8
FIGURE 3: MSTP EXAMPLE 10
FIGURE 4: G.8032 EXAMPLE (HOLNESS, 2013) 13
FIGURE 5: VRRP EXAMPLE 17
FIGURE 6: VRRP EXAMPLE 2 18
FIGURE 7: ECMP EXAMPLE 19
FIGURE 8: ECMP EXAMPLE 2 20
FIGURE 9: AS-PATH PREPENDING DEMONSTRATION 25
FIGURE 10: AS-PATH PREPENDING EXAMPLE 26
FIGURE 11: DUAL-HOMING EXAMPLE 30
FIGURE 12: LOAD SHARING SCENARIO 33
FIGURE 13: LAG CONFIGURATION 34
FIGURE 14: G.8032 CONFIGURATION 36
FIGURE 15: CORE CONFIGURATION 1 37
FIGURE 16: CORE CONFIGURATION 2 38
FIGURE 17: VRRP CONFIGURATION 39
FIGURE 18: BGP LOAD-SHARING 1 40
FIGURE 19: BGP LOAD-SHARING 2 41
FIGURE 20: BGP LOAD-SHARING 3 42
FIGURE 21: BGP LOAD-SHARING 4 43
FIGURE 22: BGP BEST CASE 45
FIGURE 23: BGP WORST CASE 46
FIGURE 24: VENN DIAGRAM SCREENSHOT (EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE, 2012) 50

Traffic Load Balancing Techniques

Increasing number of users and data hungry applications in networks today demand more network throughput than they did years ago. Global IP Internet traffic increased 34% from 10,942 Petabyte per month in 2009 to 47,176 Petabyte per month in 2014, (Cisco Systems, 2010, p. 7). While network speeds do not double every two years like number of transistors on a chip (Moore's Law observation), keeping up with throughput requirement demands efficient network design. Load balancing provides efficient distribution of network traffic to increase throughput in a network. Careful network planning needs to be in place to implement load-balancing techniques as well. Middle boxes like hardware load balancers are another solution to provide redundancy and network load balancing. They are also known as layer 4-7 routers (Caforilo, 2007). These hardware devices prove to be fast and reliable. But they have limited use cases: mainly in hosted environments.

Redundancy in a network is essential to keep the network up and running in worst scenarios possible. For load sharing, sometimes redundancy is necessary. Redundancy introduces few complicated forwarding paths in the networks, which, can both be a blessing and a nightmare depending on how effectively the network is designed. Redundancy could be:

- Node level
- Link level

Node level redundancy provides node failure backup and link level redundancy provides link level backup. Examples of node level redundancy would be MC-LAG, FHRPs, etc. Examples for link level redundancy would be LAG, or layer 2 looped topology.

Problem

Load balancing is a concept for distributing traffic over available links. It is a one-way process; similar action should be taken at the other side to leverage the benefits associated with traffic load balancing. In modern networks it is technically possible but practically un-achievable at large scale. Load balancing is only achievable in very specific scenarios.

Sometimes load-sharing is mistaken as load balancing. Load balancing refers to the even distribution of network traffic disregarding the link weights. While load sharing is a process inherent to router to share forwarding of traffic. Load-sharing techniques help distribute traffic on available links fairly equally. It still has the possibility of unbalanced forwarding, in case unequal cost paths exist. From a network point of view, load-sharing techniques can't load balance traffic all over the network evenly in most possible scenarios. Simply putting, few parts of the network are left underutilized.

Network devices inherently have load-balancing capabilities. Layer two devices and layer three devices are capable of bundling links in LAGs (Link Aggregation Groups). Traffic then can be split over the available links in a few desired fashions (depending on the vendor implementation) like: source, destination or both IP, MAC, and port. Some vendors provide more variations to select traffic distribution criteria. Layer three devices have a feature called ECMP (equal cost multipath). Multiple paths to the same destination having the same cost can be used to load balance traffic over the available links. ECMP is also vendor dependent. Cisco Systems Inc. has a few options for packet switching including but not limited to: process switching (per packet), fast switching (per flow), Cisco Express Forwarding (per flow, per packet). A fact to consider is that LAG/ECMP are unidirectional and they require careful

implementation on the other side of path as well.

On the other hand, load-balancing techniques are limited in certain ways. First example would be unidirectional property of load balancing techniques. Distributing traffic evenly on the way back is sometimes easier said than done. Second example: In a redundant network, to load-balance traffic over possibly all links if case the redundancy level increases, it gets hard to mange network load balancing. Another simple example would be the following figure where ECMP is not an option in layer-two domain and LAG is limited in use here. Such scenarios provide some level of redundancy. Along with redundancy, a few complicated forwarding paths are introduced. These forwarding paths have to be managed carefully for mitigate forwarding loops. Fortunately, redundancy can also help load share traffic over available links. Different networking protocols can be used to resolve switching loops depending on the type of layer two technologies used. While most of the loop prevention technologies can provide load sharing, but it is the network planner's responsibility to assess the traffic and distribute among available paths to make certain the path utilization is fairly even.

To use load balancing and redundancy to another advantage, FHRPs (First Hop Redundancy Protocols) can also implemented along with looped topologies. FHRPs provide gateway redundancy and can help load share traffic to both gateways in conjunction with looped infrastructure. Gateways in case of Interior Gateway Routing are fairly in favor of load balancing. But issue gets complicated when load sharing traffic on different link while multi-homed to two or more Service Providers. BGP inherently selects only one best path to get to a destination, eliminating the use of ECMP for load balancing.

These are the few issues that may impact network performance essentially requiring effective load-balancing strategies.

Discussion

LAG

Link Aggregation Groups are both a layer two and layer three features. Multiple links can be bundled together and treated as a single logical link. Traffic can then be load balanced on all available links. An example of LAG is below.

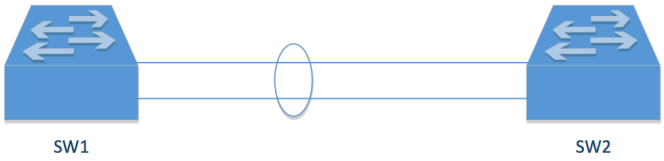

Figure 1: LAG Example

Link Aggregation can be configured statically or dynamically. Dynamic Link Aggregation Groups have a few implementations based on different vendors like EtherChannel from Cisco, Port Trunking from HP, Channel-Group from Dell, and LACP (IEEE 802.1AX-2008). These protocols provide link redundancy and link aggregation between two adjacent nodes.

In some scenarios, node redundancy is also required. Multi-Chassis LAG is the solution in such scenarios. Its implementation varies from vendor to vendor. Cisco offers Virtual Port Channel, and Dell offers Virtual Line Trunking. MC-LAG provides both link level and node level redundancy and required in some critical applications. Following diagram illustrates MC-LAG:

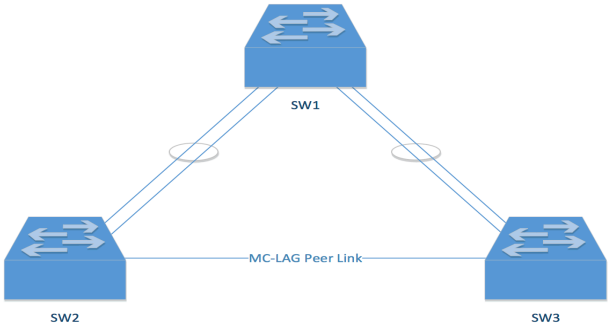

Figure 2: MC-LAG Example

Loop Prevention Mechanisms

MSTP. Multiple Spanning-Tree-Protocol, an extension to Rapid Spanning Tree Protocol (802.1w) and later incorporated into 802.1Q-2005 is a VLAN aware Spanning Tree Protocol, unlike other Spanning-Tree-Protocols. MSTP creates separate spanning-tree instances for each group of VLANs. This allows creating different spanning trees with blocked ports on different physical ports for each instance. MSTP is also backwards compatible with RSTP, so it can interpret RSTP BPDU messages. The protocol has following timers and parameters as specified in IEEE Standard for Local and metropolitan area networks — Media Access Control (MAC) Bridges and Virtual Bridge Local Area Networks (Institute of Electrical and Electronics Engineers., 2011):

Parameter	Default
Bridge Hello Time	2 sec
Bridge Max Age	20 sec
Bridge Forward Delay	15 sec
Transmit Hold Count	6 sec
Max Hops	20

While vendor specific extensions may exist for other Spanning-Tree-Protocols, the standard 802.1D does not have support for VLANs. In Provider or Customer infrastructure, VLANs are an essential part, as they support isolation of forwarding paths by creating Virtual LANs. Following is an example of a Layer two network running separate MST instances. MST 1 forwards traffic for one set of VLANs, and MST 2 forwards traffic for another set of VLANs. This enables traffic distribution over all available links essentially providing load-sharing functionality. While network performance may vary according to the usage of VLANs in different MST instances. For example, if VLANs in MST 1 are bandwidth hungry and MST 2

VLANs barely utilize more than 50% of the available bandwidth, this leaves with underutilized links in one part of the network. Such scenarios require careful network design.

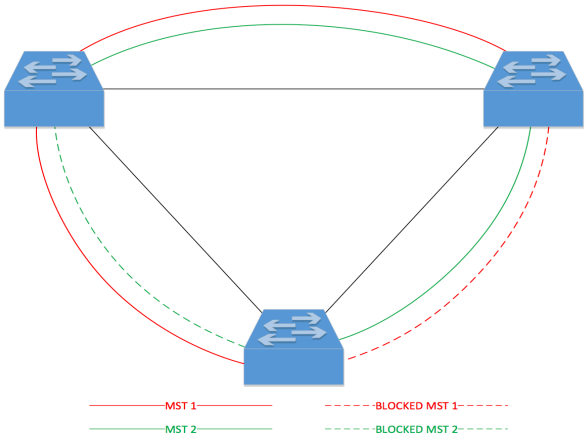

Figure 3: MSTP Example

G.8032.

G.8032 recommendation specifies a protocol and protection switching mechanisms for Ethernet networks. Each node is connected to two adjacent nodes participating in the same Ethernet ring. The principals, as stated in G.8032/Y.1344: Ethernet ring protection switching (International Telecommunication Union Telecommunication Standardization Sector, 2012), are as follows:

- Loop avoidance
- Utilization of learning, filtering and forwarding database (FDB) mechanisms defined in the Ethernet flow forwarding functions (ETH_FF).

There is a central node called Ring Protection Link (RPL) owner, which blocks one of the available ports in Ethernet ring to prevent loops from forming. It uses Ring-Automatic Protection Switching (R-APS) messages to coordinate switching activities. Any failure in the ring triggers R-APS signal fail (SF) message on the neighboring nodes of the failed link along both directions. Consequently, RPL unblocks the blocked RPL link. This switchover occurs in less than 50ms in most cases. Which, in contrast to MSTP, is order of magnitudes faster.

Operational differences as compared to MSTP include:

- Continuity Check Messages (CCM) on G.8032 are sent every 3.33ms to detect failure in 10ms, which makes 300 messages per second. On the other hand, MSTP sends hello messages every 2 seconds.
- Each logical ring has its own RPL owner. But MSTP is not primed for ring topologies; however, it works in any possible implementation adhering to some restrictions. So, in MSTP, a logical ring can have more than one root bridges (for better load sharing, and to utilize the unused link).

G.8032 has two versions, namely G.8032v1, and G.8032v2. G.8032v2 has all the features from G.8032v1 and some additional modification. Both are explained below:

G.8032v1.

- Ethernet protection switching using Ethernet OAM Connectivity Check Messages (CCMs).
- Service oriented protection switching based upon VLAN tags compared to traditional link based Ethernet resiliency mechanisms (e.g. Spanning Tree Protocol (STP), or Link Aggregation (LAG).
- Sub 50ms protection switching performance for rings less than 1000km and/or 16 nodes.

G.8032v2.

- Support for both Revertive/ Nonrevertive operational modes for controlling ring behavior after the event that caused the protection switch is cleared.
- Additional administrative commands include a Forced Switch (FS) command that enables the operator to force traffic on the protect path (useful for maintenance or upgrades) and Manual Switch (MS) for blocking a particular ring port.
- The Flush FDB (Filtering database) Logic significantly reduces amount of flush FDB operations in the ring and improves performance during a protection switch event.
- Support of multiple ERP instances on a single ring. Future support for hierarchical ladder rings.

G.8032 in Access and Mesh networks. Generally, Access Layer has a few implementations: Layer-2 or Layer-3. Fully routed access layer designs are not widely deployed. If you are considering an Ethernet based ring topology in your high performance network, consider spanning tree limitations and choose more resilient technology like G.8032 (Ethernet Ring Protection Switching aka ERPS). Many vendors have their own variation for similar implementation, not necessarily identical: Resilient Ethernet Protocol (REP) by Cisco, Ethernet Automatic Protection Switching (EAPS) by Extreme Networks, Ethernet Protection Switching Ring (EPSR) by Allied Telesis, Rapid Ring Protection Switching (RRPS) by Huawei/H3C, and eRSTP by RuggedCom.

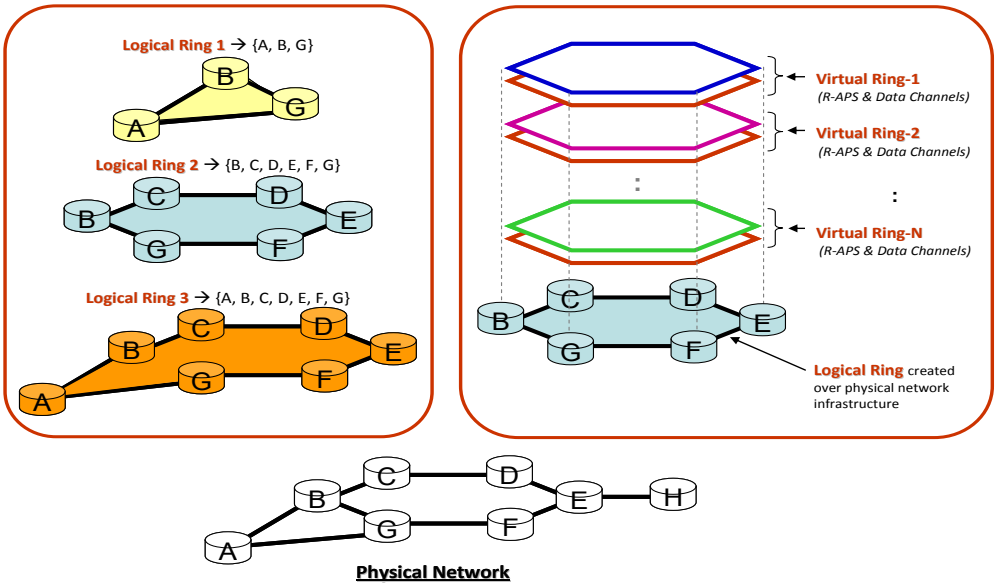

Figure 4: G.8032 Example (Holness, 2013)

Guidelines for end-to-end network/service *resiliency* involving G.8032 are still under study. Inherently G.8032 is designed to serve core-switching purpose. End-to-end resilience is not yet widely supported by G.8032. To provide end-to-end resilience, G.8031 or MPLS-TP can be used on top of G.8032. Connecting sub-rings for extending ERPS support to access layer is also possible.

Implementing G.8032 in mesh networks may also benefit by increasing network availability. It can be implemented in any desired configuration in mesh networks since it is physical layer independent. Logical rings can be configured with interconnected/overlapping rings. Moreover, a device supporting G.8032 can run multiple instances of G.8032. It provides load-balancing capability and ability to interconnect multiple rings at overlapping nodes.

Using G.8032 in simple ring. ERPS doesn't provide marginal benefits if implemented in non latency-sensitive ring-topology, e.g. a ring of switches in a small office network. It provides maximum availability, which may not be required in such simple rings. MSTP would take care of such topology without any noticeable packet loss. However, MSTP falls short on failover time when implemented in large core networks, or interconnected rings (in which case, its convergence time may increase exponentially).

Applications.

- Site-to-site Layer 2 VPNs
- Ethernet Private Line
- Wholesale Ethernet access
- 3G/4G cell site Mobile Backhaul
- Ethernet access to IP services
- Delivery of Private Cloud services
- Residential services

G.8032 vs. MSTP. Spanning Tree Protocol by comparison to Ethernet Ring Protection Switching is complicated to understand. G.8032 is primed for inherently simple ring (logical or physical) topologies. There is support for multi-ring/ladder topology as well; still core principles stay the same, protecting each ring/sub-ring individually.

On the other hand, STP designed for any possible looped topology. It treats all devices (for IEEE 802.1D and IEEE 802.1w) connected together in a single protection domain. There are some proprietary variations that extend the functionality of standard 802.1D and 802.1w protocols. While even VLAN aware 802.1s (ultimately 802.1Q) may help load balance traffic and create different protection domains, it still suffers from the state problem. Because Spanning Tree BPDU messages are sent from/to all devices on the segment; bigger the segment, more time it takes to converge on failure.

G.8032 on the other hand has to manage each logical ring individually. A single ring failure may or may not affect other rings (depending on the physical connectivity relation between two rings).

Maximum number of nodes supported by G.8032 is from 16 to 255. For a less than 16 nodes in an Ethernet ring and maximum of 1200 Kilometer of ring fiber circumference, the failover time shall be less than 50ms. While in Spanning Tree Protocol, the default timer dictate 7 hops diameter (approximately 15 hops in a ring). Default maximum number of hops for both Cisco and Juniper are 20, after which BPDU messages are discarded. The failover may take <1sec in best cases and may go over 5 seconds.

First Hop Redundancy Protocols

First Hop Redundancy Protocols provide gateway redundancy. There are a few FHRP protocols like: HSRP and GLBP from Cisco, NSRP from Juniper, R-SMLT from Avaya, ESRP from Extreme Networks, and VRRP Open Standard Protocol. In the event if a gateway fails, FHRP provides fast switchover to the available gateway. For the sake of support, we are going to concentrate this work on open standard protocol: VRRP (Virtual Router Redundancy Protocol).

Gateway protection is very important for network infrastructure. FHRP provides gateway redundancy by creating virtual routers. An IP address is assigned to a virtual router group and then the virtual router acts as a gateway. Multiple routers can participate in a virtual router group. Virtual router priority decides which router takes over the master role in case the master virtual router gateway fails. VRRP supports multiple virtual router groups; it helps assigning virtual router group master role to different routers for different groups. The figure below shows how VRRP enables gateway redundancy. Both routers participate in the VRRP virtual router group and one of them acts as a master (active gateway) and the other one stays idle until a switchover is required.

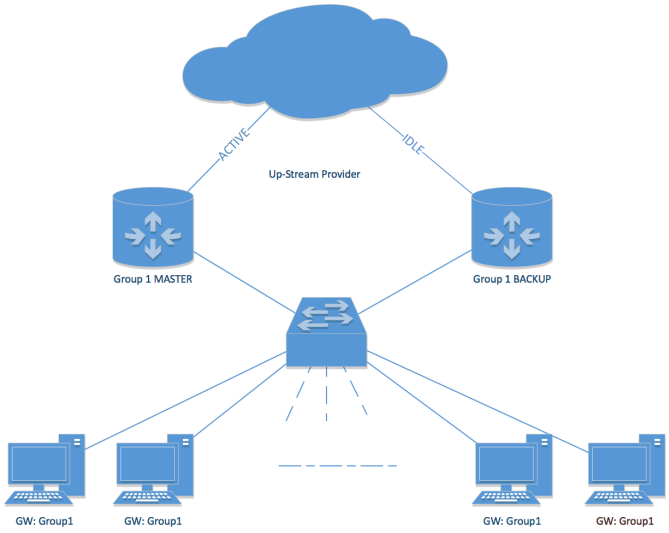

Figure 5: VRRP Example

VRRP group feature can also be leveraged to implement load sharing where one gateway forwards traffic for one group and other gateway forwards traffic for the other group. This enables the utilization of available hardware without compromising redundancy. If any of the available gateways fails, the switchover would occur instantly. Load balancing scenario is depicted in the following figure.

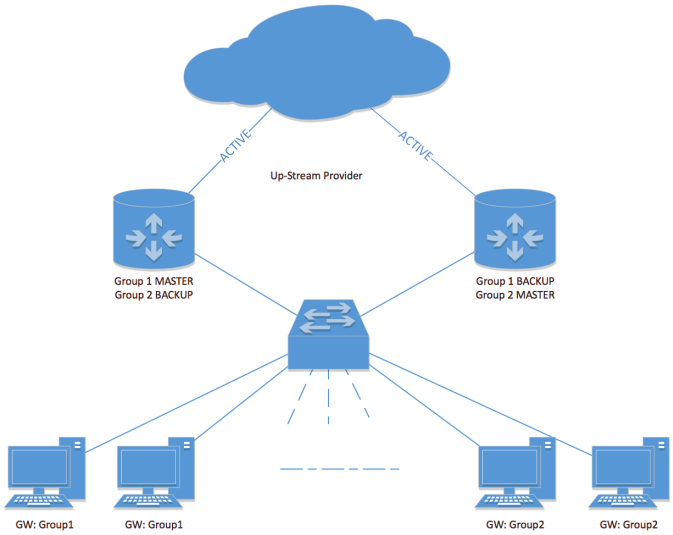

Figure 6: VRRP Example 2

Multiple Virtual Router group master advertises its state at a preconfigured advertisement interval. Default advertisement interval is 1 second. If a backup router doesn't hear an advertisement for 3*advertisement intervals + skew time, it takes over the master role. The second example is more common VRRP deployment. Gateways don't fail in most cases, so 3 seconds switchover is a fair to wait for VRRP advertisement. Few vendor deployments facilitate advertisement interval to be configured in milliseconds, bringing the 3+ seconds switchover time down to a few milliseconds.

Interior Gateway Protocols

ECMP. Equal Cost Multipath is the feature of Router's forwarding process. If an IGP injects multiple paths into the Forwarding Information Base, the equal cost paths can then be used to load balance traffic. ECMP does not take into account any differences in the bandwidths of the outgoing interfaces, but it has been a common practice to use a link metric that is proportional to the inverse of the link's bandwidth. Normally ECMP is implemented to link-state routing protocols, because they need quite small modification to their path calculation, but ECMP implementations to DV protocols have been published as well, (Lappetelainen, A., 2011). For example, the Interior Gateway Routing Protocol (IGRP) and Enhanced Interior Gateway Routing Protocol (EIGRP) support ECMP. Following illustration helps explain ECMP behavior:

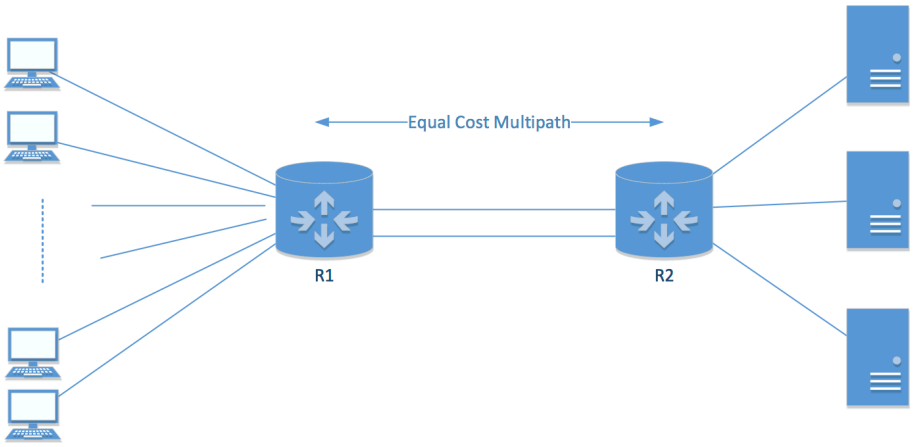

Figure 7: ECMP Example

In this scenario, router R1 has multiple paths to get to the network(s) connected to R2. It applies to router R2 as well. Now both devices can forward traffic destined to each end over both available paths depending on the available packet switching technique used. Forwarding could take place in round robin fashion, or per flow.

Another example to demonstrate ECMP is as follows:

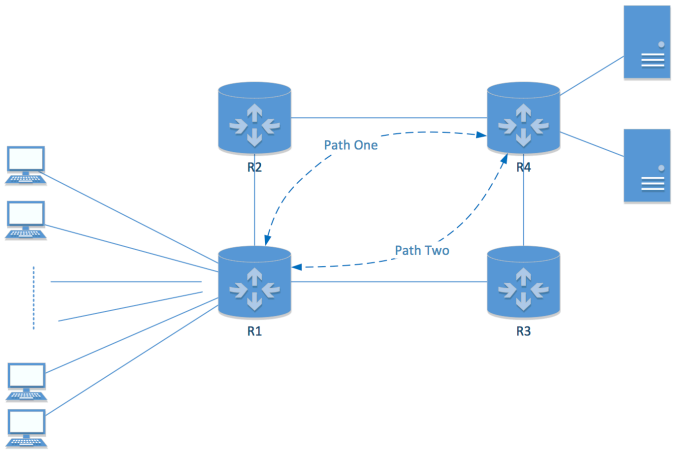

Figure 8: ECMP Example 2

This scenario demonstrates the benefits of equal cost paths more clearly. Traffic originating from both ends can be distributed over the entire network fairly evenly, depending on the forwarding scheme used. Packets may be subjected to reordering if per-packet forwarding is used. Reordering of packets often results in degraded TCP performance and underutilization of available links. To prevent undesired packet reordering, per-flow forwarding must be used. Traffic can be forwarded based on source-destination-IP to better utilize available links. But per-flow forwarding does not guarantee even distribution of traffic. Fortunately, this uneven distribution of traffic is un-noticeable often.

Unequal Cost Multipath (UCMP) is another option to leverage multipath forwarding. Protocols like EIGRP and IGRP offer this feature to install routes in the Forwarding Information Base even if the cost is unequal. It enables forwarding over multiple available paths without compromising network performance. Traffic is distributed inversely proportional to the link speed. Commonly deployed routing protocols like OSPF, and IS-IS do not support ECMP. RIP also does not support ECMP.

There is another protocol that supports UCMP or UCLB (Unequal Cost Load Balancing) as some may call it. BGP can load share traffic in many ways depending on its configuration. There are quite a few ways do load sharing in BGP. Only a few of them leverage ECMP or UCMP functionality.

Border Gateway Protocol

When designing effective load sharing scheme, type of services that are being accessed/provided over the network do matter. Some services like VoIP calls may experience drastic performance issues due to Asymmetric Inter-AS routing. Implemented state in the network, such as stateful firewalls, may also require symmetry in the inspected traffic.

Only a little number of applications requires network symmetry. Instead, networks (if designed well) may benefit from asymmetric routing as the resource utilization goes up by dividing traffic evenly.

Taking delay sensitive customer traffic and stateful networks into account, we may round down our selection of BGP load-sharing strategies to the ones promoting symmetry. Mainly because delay/jitter in distributed autonomous-systems isn't predictable. At the edge, it may be better to implement symmetrical routing rather than spreading inbound and outbound traffic over different parts of the Internet (where performance for delay sensitive traffic may suffer).

BGP Ingress load sharing.

BGP Multihop. This feature is widely used in case of dual homing to same provider over single router (to single router) with loopback addresses. It is only for eBGP not for iBGP. In some cases, BGP multi-hop may be used to provide connectivity to not directly connected neighbors only. But in load balancing scenarios, BGP multi-hop is used to trigger implicit ECMP load balancing (the IGP's switching architecture). In such scenarios, the neighbor is connected through more than one physical links, and the neighbor-ship is not based on the physical interface addresses, rather it is based on loopback interfaces. The router will have more than one path to the neighboring router and it will do ECMP load balancing over all available links

(provided the cost is equal). It provides efficient load balancing but limited to the multiple physical links to neighboring router.

BGP Multipath. This feature selects multiple paths to be installed in the IP routing table for load sharing. Multiple paths are used to distribute traffic from router's perspective, not to advertise multiple BGP routes to its neighbors. Generally works for dual-homing scenarios where the AS is connected to same service provider with multiple links. Traffic can be shared on a per-packet or per-destination bases. Few vendor specific variations of multipath may enable load sharing over multiple providers as long as the AS Path length is the same. Drawbacks include increased routing table size because multiple paths to the same destination are being installed, more memory/CPU utilization. May not be a best bet if used with multiple providers on a low-memory/CPU router.

MED. MED is an optional non-transitive attribute. It works by sending a metric to the neighboring Autonomous System about its own routes. It provides a dynamic way to instruct neighboring AS on how to reach the advertised prefixes. When the receiving AS has multiple direct paths to sending AS, the MED value can be compared and best path chosen accordingly. Lower MED value is preferred over higher MED value to choose a preferred path. A BGP speaker will only propagate the received MED attribute to iBGP neighbors. So, the received MED value for a prefix never leaks out of an AS. In the case of a provider, multiple routes can be compared for same prefix even if received from different Autonomous Systems. This feature is called always-compare-med. Another noticeable feature is to group multiple routes from same AS together and then decide the best, called deterministic MED. These two features paired together provide an efficient routing decision. Metric is certainly a useful attribute for dual-homed customers to a single provider and requires less reconfiguration.

Communities. BGP community attribute is very useful. Service providers support this option to offload some responsibilities to customers. While it may not be a viable option to ask the service provider to change certain attributes every now and then, with BGP community attribute, a customer can tag routes with BGP community numbers and have the corresponding action executed at the service provider site. For example, if the service provider has provided you with a list of BGP community attribute numbers corresponding to the actions they trigger, you may use specific communities to do different tasks (adjusting local preference, limiting advertisements within the AS, not advertising to the internet, distinguishing different class of traffic) in the provider network. This list is often called community definition. Prefixes can be tagged with community number and advertised to the provider to leverage the features provided. Community list may vary from provider to provider, but it offers granular control and solves many issues without contacting service provider.

AS-path prepending. AS-Path Prepending is quite diverse and provides higher degree of control over ingress load sharing. This attribute may not be necessary in dual-homed environments to single service provider, but it does provide great benefits in multi-homing scenarios. To influence ingress traffic in multi-homing scenarios, this option is widely used in practice. The following scenario helps demonstrate its operation clearly:

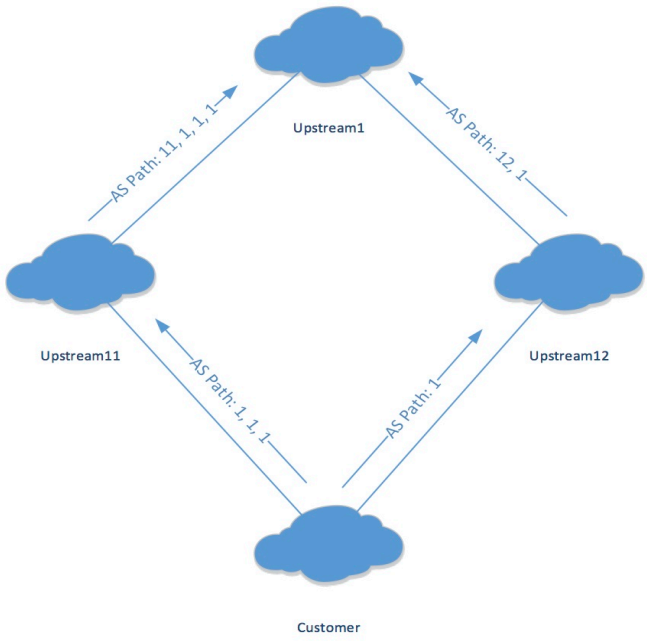

Figure 9: AS-path prepending demonstration

In the figure above, a Customer is connected to two upstream providers Upstream11 and Upstream12. In this multi-homing scenario, Customer uses AS path prepending and prepends its own AS number two times, essentially making the AS path length 3 as it is advertised. The Upstream11 provider sees path to Customer as 3 AS away, and Upstream12 provider sees Customer as 1 AS away. Now both providers advertise these prefixes to their neighboring AS Upstream1. Upstream1 sees Customer as 4 AS away on Upstream11-Upstream1 link, and 2 AS away on Upstream12-Upstream1 link. So it prefers the link to Upstream12 to reach Customer. AS path prepending is an effective attribute that helps load share traffic with least amount of reconfiguration in multi-homing scenarios. And its effects are seen throughout the WAN.

It has very high degree of success in sharing traffic in multi-homing scenarios, but as always, there are drawbacks in a few scenarios. The following scenario helps demonstrate such a case:

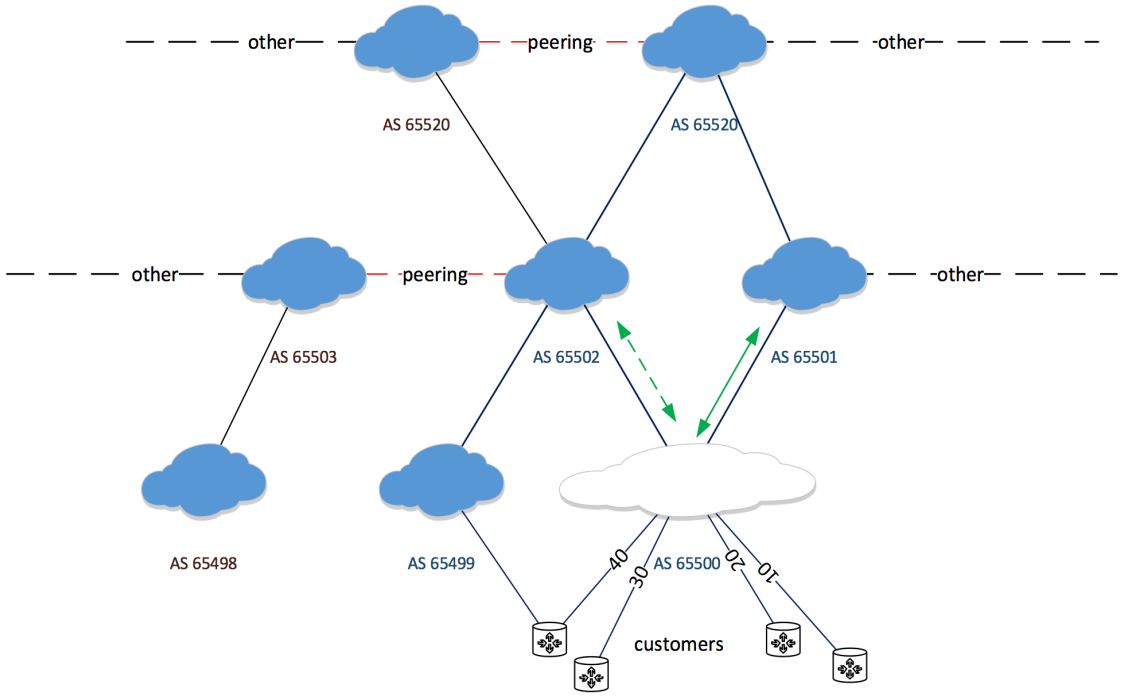

Figure 10: AS-Path Prepending Example

In this scenario, let us assume AS65500 is we. We provide Internet access to four customers: Customer 10, 20, 30, and 40. Customer 40 is has Internet access from AS 65499 as well. When we select path to AS 65501 as our primary traffic path for customer 30 and 40, and perform AS Path prepending two or three times on the link to AS 65502, AS 65502 should not serve us ingress traffic for those prefixes.

In this case, path from AS 65498, AS 65503, AS 65499 and AS 65502 is same to AS 65500 if the path was prepended three times. Even less if it was two times. Such cases do exist, and there is one way around: perform more AS Path prepending, this way AS Path gets longer and longer. While routing updates with longer AS paths can be filtered out, but it is not necessary most of the time as the Internet is so closely connected the AS Path length barely goes over 5-10 AS's. There was one noticed scenario in February 2009 where a Czech provider prepended the AS Path in an unusual manner that it went over 255 in AS length when it reached

its upstream providers. This caused a flood of routing updates in the Internet backbone essentially resulting from router meltdowns that couldn't handle this update. Those meltdowns triggered the routing updates flood. This flap that was heard around the world, impacting more than 1000 prefixes in the Internet, by a single AS Path prepend.

Longer Prefix announcement. Routing 101: Longest prefix length over shorter prefix length. Using this property requires that the customer owns its IP space. If the IP addressing space is owned by its provider and/or is being aggregated by its up-stream provider, this technique may not be beneficial. In longer prefix announcement, the network is split into mostly two subnets (or more if multi-homed to more than two providers). Different longer prefixes are announced on different upstream links, to different up-stream providers. This technique is used in multi-homing scenarios. Single/Dual-homed customers to a single service provider have more options to choose from, as described above.

Summary address is also advertised along with the longer prefix to make sure both service providers can reach the customer in case one of the service providers experiences any problems. This is also one of the best ways to load share traffic among different service providers in BGP. According to Beijnum (2002), "Announcing more specifics is also useful when someone else announces your address block (by mistake, or by your request but no longer needed) and you don't want to wait for them to fix this."

A drawback associated with this technique: black-holing. In case a chunk of IP addresses in the customer's network goes unreachable and the BGP speakers are still advertising the routes to up-stream providers, uplinks will keep bringing traffic destined for unreachable devices. This is bad, both from network throughput perspective and financial standpoint where the customer would be paying upstream provider every time the connectivity to a smaller subnet goes down. It

is a possible scenario where smaller subnets are provided further to users/customers of the "Customer" in this case. Black-holing of course leaves customers unhappy, and prompts more phone calls to the providers.

Longer Prefix announcement (without summary, conditional). Announce the prefix only if certain conditions are met. IP SLA can provide extensive set of parameters to match up on. Selection is including but not limited to: Reachability, Link Performance, and External events. In this way, only advertise certain networks if they are reachable. Different providers may have different ways of implementing tracking. Cisco does that by using Track statement and IP SLAs. Y.1731 is another option to check the connectivity.

The triggered announcements make sure the prefixes being advertised are, for sure, reachable. Carefully designed conditional advertisements can handle most of the failure scenarios automatically, without any technical intervention. Of course, it falls back to how well the conditions configured and tested before deploying them to take care of BGP advertisements.

BGP Egress load sharing.

- Full BGP routing table
- Partial BGP routing table
- Default route

While BGP Multipath and Multihop can also be used for load sharing in the same manner as they can be used to influence ingress traffic if you have equal cost paths over multiple links. These techniques are mainly used in single/dual-homing scenarios with the same service provider. If you are getting full BGP routing table, it is not wise to use ECMP to load balance traffic as the number of routes increase by order of magnitude as new links are introduced. It is useful if you have a router that can handle 100,000+ routes per link. Sometimes it is not feasible.

There are few more techniques available to influence egress traffic as follows:

Weight. Using inbound route filtering, different prefixes can be assigned weights over different links accordingly. For example, routing updates coming from link-1 can have higher weight than the same routing updates coming from link-2 for half the routing updates. It is useful in full and partial BGP routing table updates where you can choose what routes to reach over what link. But, asymmetric routing may be introduced if ingress traffic doesn't take the same route as the egress traffic. Also, in case of default route, there is no incoming routing updates to set weight attribute on. Moreover, weight attribute is local to the router, so it may not be useful in more than one boundary routers.

Local Preference. Unlike the weight attribute, local preference is local to the AS. Hence, it can be used in scenario with multiple border routers. Inbound route filtering capability can be used to set local preference to favor certain routes over others. To effectively do egress load sharing, depending on the number of provider links, the inbound prefixes can be split. Asymmetric routing may occur if local preference is not set in conformance with the ingress traffic policy. Similar to the weight attribute, local preference is used with full or partial BGP routing updates. Setting local preference on default routes does not help load balance traffic over multiple links. Instead it just makes one default route more preferred over the other. It still may be useful in certain scenarios. The following illustration shows multiple border routers with single upstream providers:

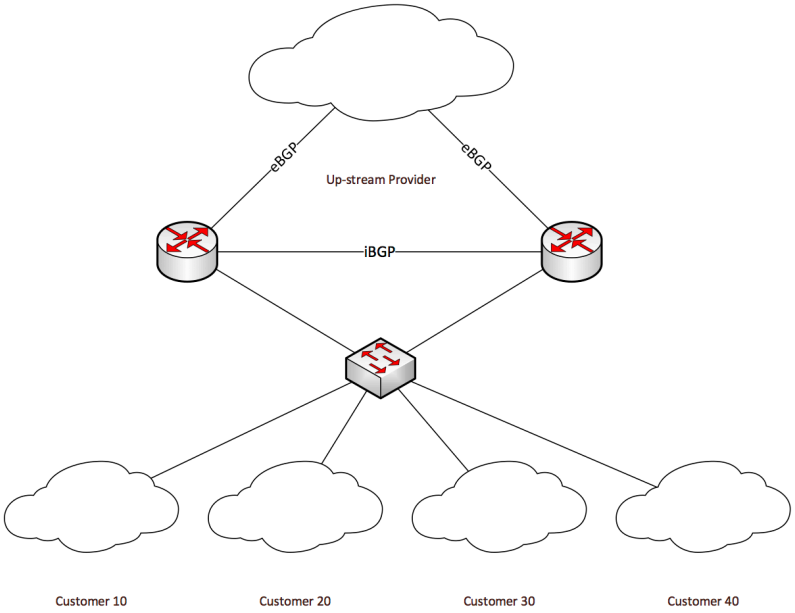

Figure 11: Dual-Homing Example

Unlike the Weight attribute, Local Preference attribute can be propagated within the Autonomous System. It helps in route selection with multiple boundary routers. Both routers know where to forward traffic for designated prefixes.

Communities. Similar to the ingress load sharing case, BGP Community attribute can be used for egress load sharing as well. Community options can be leveraged to help make decisions over available routes. Matching routes based on different criteria like: originating AS, traversed AS, specific prefixes, and already set community attributes with the help of route filtering. These capabilities enable BGP community attribute to be used with full/half BGP routing updates and default routes depending on the implementation.

Route-filtering. Route filtering feature is available in most of the routers in market. This capability helps filtering inbound or outbound routes based on certain criteria. The criteria can be specified on which you want the route filtering to happen, while some restriction may apply. Route filtering can help filtering routes based on many options and it can also help manipulate inbound updates.

Filtering inbound routing updates to allow partial set of prefixes in can help load share egress traffic on different links. This feature is most usable in multi-homed scenarios with multiple border routers. This way symmetric routing can be guaranteed.

Policy Based Routing (Conditional). Conditional routing based on certain conditions. Conditions could be matched on predefined criteria like match-groups or route-maps for example. Policy based routing can help accomplish some amazing conditional routing. For example, you can set next-hop addresses based on reachability of certain router, is just one of many things you can do with policy based routing.

Lab Demo

Following is taken into consideration for network topology

- Independent metro region with ring connectivity.
- Core region with multiple routers connected to metro region.
- Multi-homed to multiple providers.

Reasons for choosing ring as metro region backbone

- Easy to implement and extend.
- Provide redundancy inherently.
- Modern ring protection switching techniques offer SONET like failover speed.

While Mesh/Tree topologies are inherently complex to implement and introduce a lot of reconfiguration during scaling, ring topology is simple and requires less reconfiguration for scaling up. New loop prevention techniques like G.8032 and many more vendor specific techniques provide fast switchover on failure (<50ms) to provide resilient backbone.

Available Hardware

- Alcatel Lucent 7750SR-c12 (3 Units for metro ring)
- Cisco 2921 Routers (4 Units; 2 border routers, 2 for each neighbor AS)

Network diagram for the lab scenario

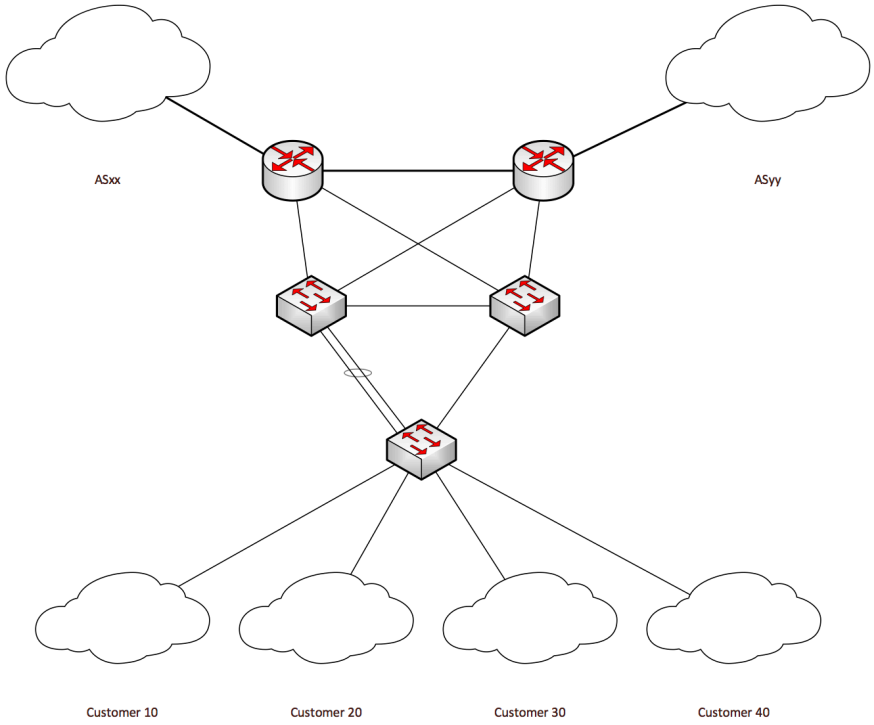

Figure 12: Load Sharing Scenario

Network Configuration steps

- Configuring LAG
- Configuring G.8032
- Configuring Core Routers
- Configuring VRRP
- Configuring BGP Load Sharing

Configuring LAG

Requirements for configuring LAG.

- "Autonegotiate" feature should be turned off or limited on the physical ports.
- Port configuration should be identical.
- At least one port should be operationally up for bringing the LAG up.

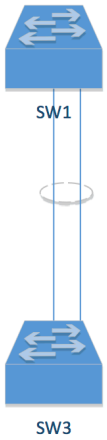

Figure 13: LAG Configuration

In Alcatel Lucent 7750 Service Router, VPLS known unicast traffic is hashed based on the IP source and destination addresses for IP traffic. We are implementing PBB-VPLS, so the traffic will be sprayed across available links based on source and destination IP addresses.

Configuring G.8032

G.8032 rings in Alcatel Lucent Service Router require Ethernet Connectivity Fault Management (IEEE 802.1ag) running to trigger switchover on failure. Then two logical rings on top of physical topology are created with “configure eth-ring” command. Each ring is configured in revertive mode with 60 seconds of time before switching back to the failed link that just came up. This prevents link flapping. Each ring requires an RPL (Ring Protection Link) to be configured manually. Manual blocking helps in load sharing such that it enables network designer to select which link should not forward traffic in normal operation.

Each ring is configured with a VLAN tag that is used as a control ring. Control ring does not forward user traffic; it is dedicated to G.8032 protection domain. VPLS instance are configured in the backbone with Service Access Points (SAPs). These SAPs are assigned individual control ring tags. This marks the control plane of G.8032.

To configure data rings, PBB-VPLS instances are created for individual rings. PBB (IEEE 802.1ah-2008) has 22 bytes of overhead, so it needs larger service MTU size than the customer MTU. Here the backbone rings are up and running. Following is an illustration of configured Ethernet rings:

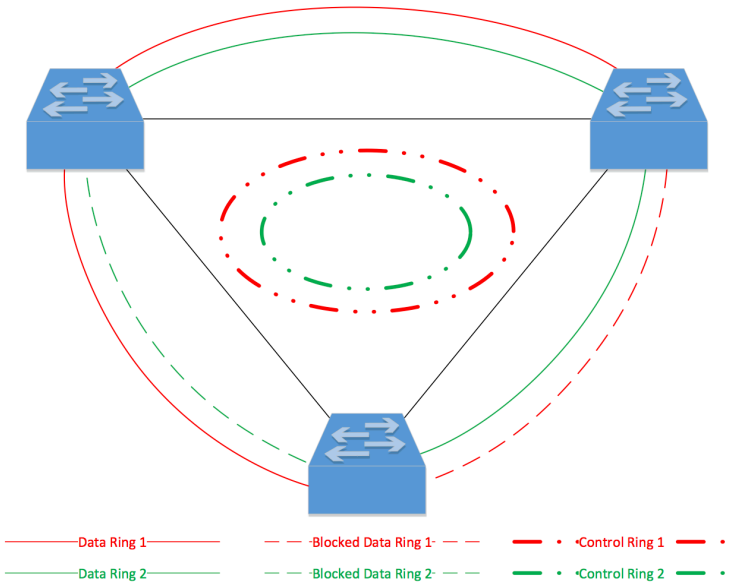

Figure 14: G.8032 Configuration

Internal dotted rings represent control rings. Note that there is two control rings, one for each data ring. We configured each ring in opposite direction to help utilize all the links to load share traffic throughout the ring. In case one link fails, one of the rings should switch over and the other one operate normally. Physical links are omitted for clarity in this figure; only logical links are represented by black links (LAG or single physical link.

Outer rings represent data rings. Each ring has an inactive link to prevent loop from occurring. PBB enables encapsulation of customer VLANs inside backbone VLANs. Customer VLAN can either be Null, IEEE 802.1Q encapsulated, or IEEE 802.1ad encapsulated. In this scenario, we create dot1Q encapsulated VLANs for each customer. Each customer site has its own VLAN ID. These dot1Q VLANs are configured as I-VPLS SAPs on all switches and encapsulated in PBB-VPLS to be switched over the backbone. Entire backbone is PBB-VPLS switched, without any MPLS or routed VPLS configuration. Customer traffic enters and leaves the metro ring at each designated switch.

Configuring Core Routers

Core Cisco 2921 routers are configured with the following

- OSPF Area 0
- BGP AS 65500
- EHWIC Configuration
- VRRP Virtual Router Groups
- IP SLA
- Object Tracking Statements

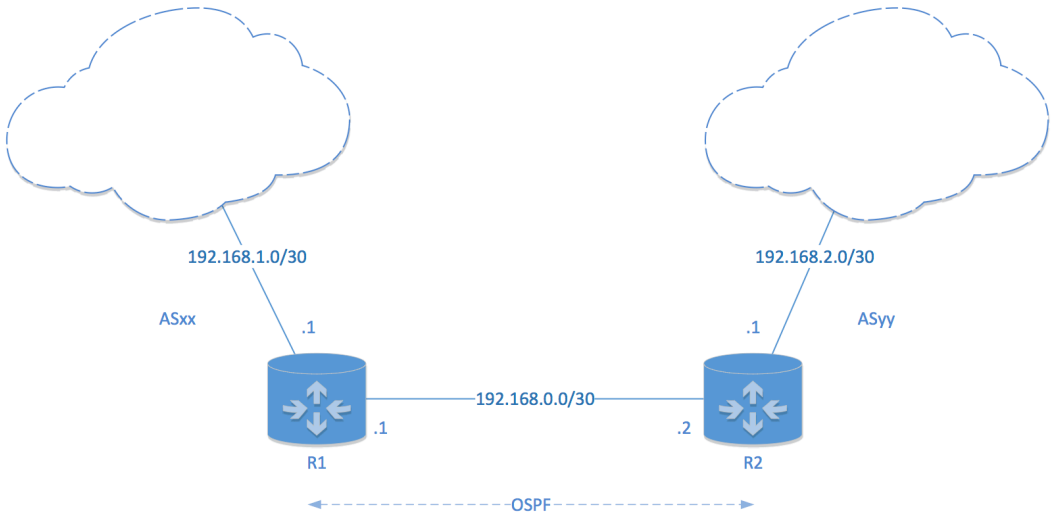

Figure 15: Core Configuration 1

The OSPF protocol is configured on the links between both R1 and R2 to advertise routing table and connected routes (not the static routes). IP subnets are specified in the figure above.

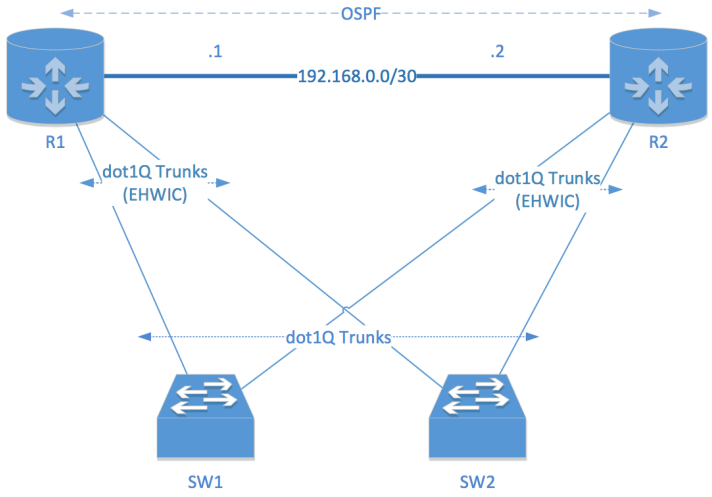

Figure 16: Core Configuration 2

Each router has an EHWIC card with four Gigabit Ethernet ports. We use two of those ports on each router to create triangle-looped topology for redundancy and load sharing. EHWIC on Cisco routers do not support RSTP (IEEE 802.1w) or MSTP (IEEE 802.1s or essentially 802.1Q-2005). They only support Cisco proprietary PVST (Per VLAN Spanning-Tree). Shutting down spanning tree on core routers will introduce layer two switching loops since metro area switches are running G.8032 only. For loop prevention, we leave PVST up and running.

Please note that all EHWIC ports in use are IEEE 802.1Q trunks. Routed VLAN interfaces are created on core routers for each customer VLAN (10, 20, 30, and 40).

Configuring VRRP

VRRP Configuration

- Router R1 is VRRP group master for Customer 10 and 20.
- Router R2 is VRRP group master for Customer 30 and 40.

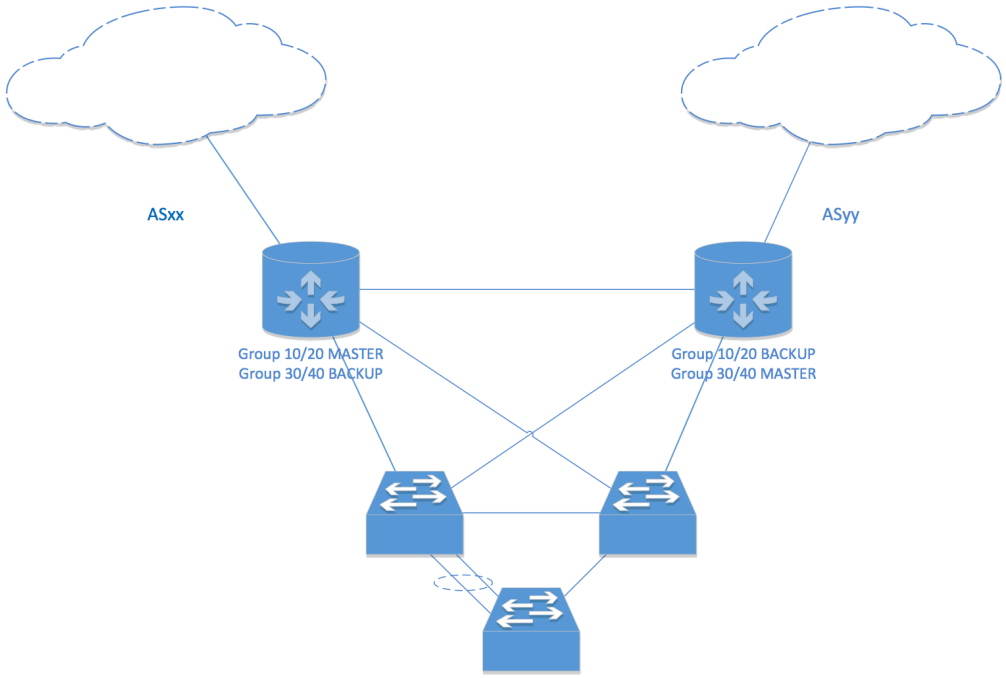

Figure 17: VRRP Configuration

Virtual Router Redundancy Protocol is configured on each Customer VLAN on both core routers and there is no physical IP address owner. VRRP default advertisement interval is 1 second on both routers. Following are the gateway addresses for each customer:

Customer ID	Gateway Address	VLAN-ID
Customer 10	192.168.10.1	10
Customer 20	192.168.20.1	20
Customer 30	192.168.30.1	30
Customer 40	192.168.40.1	40

Configuring BGP Load Sharing

Now putting the pieces together, only BGP configuration is pending. Few considerations before we begin with BGP scenario:

- Upstream providers originate default routes.
- We own the IP address space, so it is unlikely that the upstream provider will aggregate the prefixes before announcing BGP routes to any other Autonomous Systems.
- We are going to perform conditional prefix advertisement for ingress load balancing.
- Egress load balancing is influence by PBR.

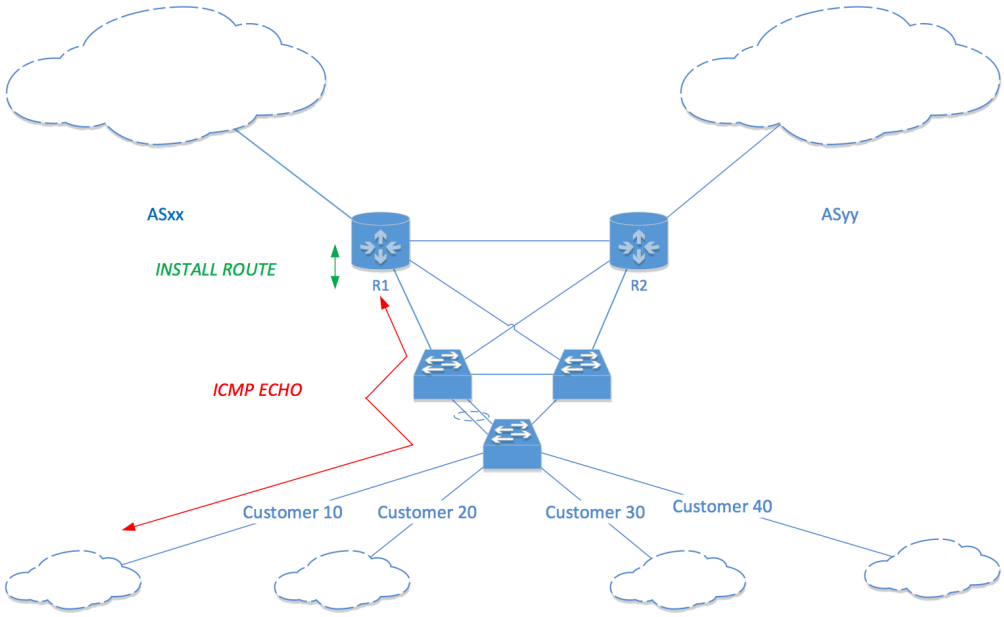

Figure 18: BGP Load-Sharing 1

The figure above demonstrates how IP SLA influences BGP load sharing. IP SLA statements are configured on both core routers to check connectivity to all its customers with ICMP-ECHO request/response messages. Track statements are used track IP SLA objects. Multiple track statements are combined to create track-lists, and these lists help us verify other

conditions like: Is the Customer 10 reachable and Customer 20 not? To withdraw routes and advertise the prefixes that are actively reachable. In this case, Router R1 is designated to forward traffic for customer 10 and 20 as specified in VRRP configuration earlier. Similarly Router R2 is designated to forward traffic for customer 30 and 40 in normal conditions. As long as core routers can reach the locally attached subnet at customer sites from the core side, the customer's internal network routes are installed in core router's Forwarding Information Base. Following are the subnets assigned to each customer:

Customer ID	Gateway	Assigned Subnet	Outside Interface IP
Customer 10	192.168.10.1	172.16.0.0/24	192.168.10.2
Customer 20	192.168.20.1	172.16.1.0/24	192.168.20.1
Customer 30	192.168.30.1	172.16.2.0/24	192.168.30.1
Customer 40	192.168.40.1	172.16.3.0/24	192.168.40.1

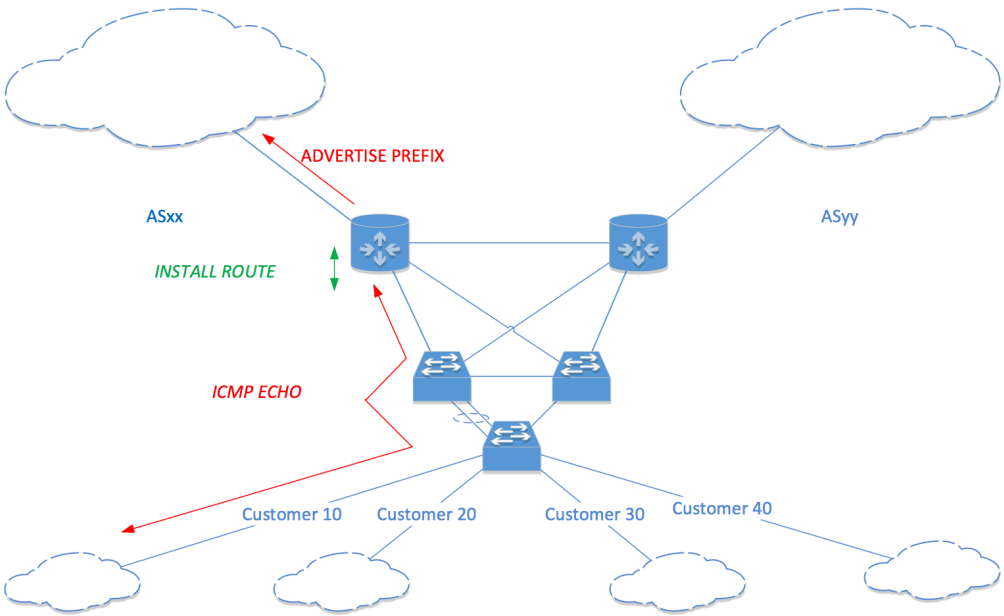

Figure 19: BGP Load-Sharing 2

After verifying customer reachability, route is installed and BGP advertisement is triggered. In normally operating network, router R1 will advertise customer 10 and customer 20's subnets by summarizing it to 172.16.0.0/23. Similarly router R2 will advertise customer 30 and 40's subnets by summarizing it to 172.16.2.0/23. It is normal operation mode.

If, for some reason, router R2 can't reach out to customer 30 and 40 through its EHWIC interface, and it can still reach them through router R1 (because R1 has connectivity to for customer 30 and 40, and OSPF is advertising connected subnets), router R2 can still continue advertising customer 30 and 40's prefixes to ASyy.

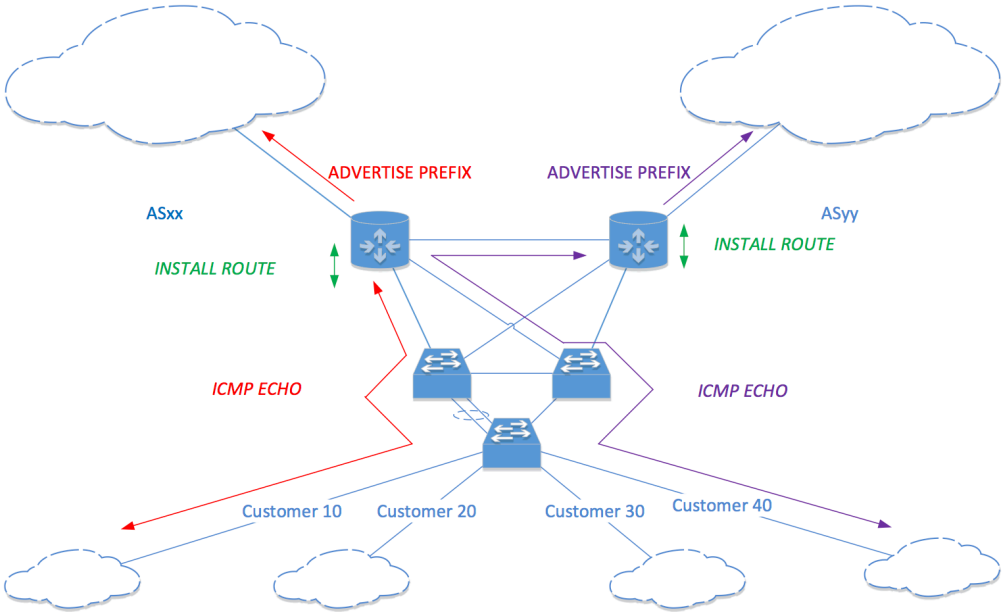

Figure 20: BGP Load-Sharing 3

Checking reachability through adjacent core router helps load share traffic even if both of the EHWIC interfaces, for some reason, become unreachable to the customers. There may be some unfortunate scenarios where connectivity to one service provider goes down. In such cases, the default route from the provider that just went down will not exist. So we can track default route from the service provider to make sure all customers can reach the Internet. To track the default route, matching is based on the following:

- Default route
- Originating AS number

Router R1 is configured to track ASyy's (connected to router R2) default route and router R2 is configured to track ASxx's (connected to router R1) default route. If the default route disappears but the other router's gateway IP addresses are still reachable through the LAN segment, routers won't advertise the customer subnets "NOT" designated to them.

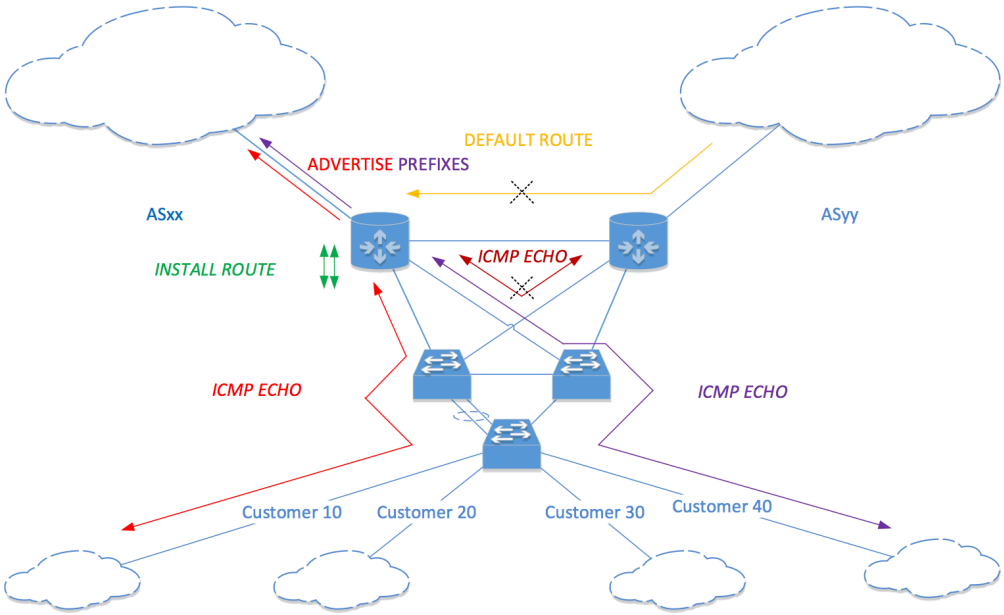

Figure 21: BGP Load-Sharing 4

Router R1 will not advertise customer 30 and 40's subnets as long as there is a default route from ASyy. In case the default route disappears "AND" router R1 can't reach gateway address on router R2 on the LAN segment, only then router R1 advertises those subnets. This helps make sure router R2 is completely unreachable on the LAN segment.

Case 1: If router R2's gateway IP addresses become reachable again, the prefix is withdrawn by router R1.

Case 2: if default route to ASyy becomes available again, router R1 withdraws the prefix

from ASxx.

In the first case, it is router R2's responsibility to forward the customer traffic. In the second case, router R1 acts as transit (not the BGP transit AS) to router R2. This traffic redirection is done with policy based routing. Gateway IP addresses are not checked to forward customer 30 and 40's traffic to router R2. It is considered that if the default route to ASyy is available, the only reason router R1 is receiving customer 30 and 40's traffic is because VRRP group 30 and 40 is acting as virtual router master, so router R2 should be unable to receive customer 30 and 40's traffic.

Best-case scenario

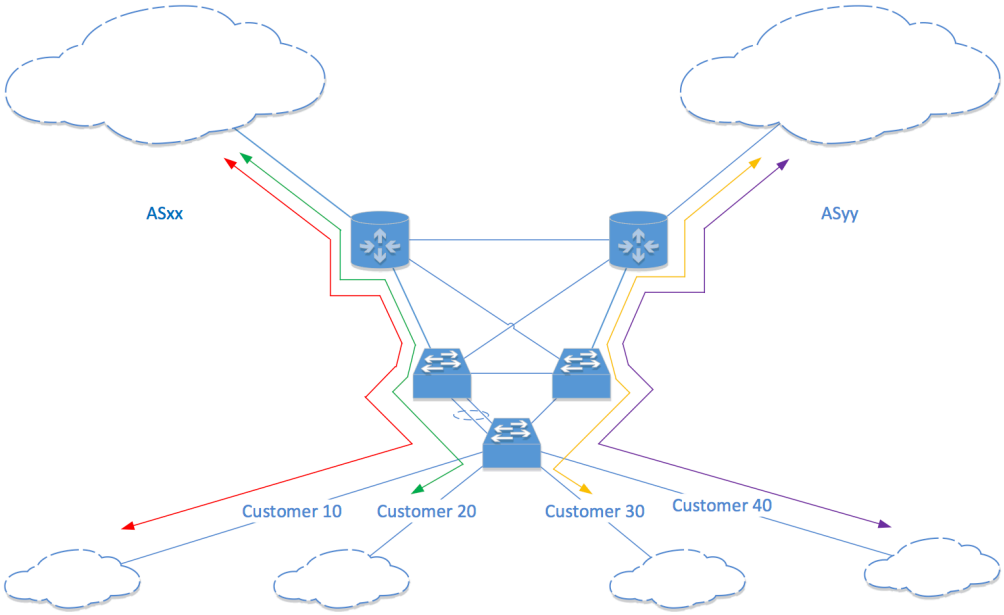

Figure 22: BGP Best Case

In the best-case load-sharing scenario, both core routers will forward traffic for two customers each. The egress load sharing is done based on where the prefixes are being advertised (to ASxx or ASyy). Customer 10 and 20 will use router R1 to exit the network, and customer 30 and 40 will use router R2 to exit the network under normal mode of operation. If router R1 receives traffic from customer 30 and/or 40 while default route to ASyy is active, traffic is redirected with policy based routing (PBR). If there is no default route to ASyy, no match in PBR route-map occurs and traffic is forwarded to the neighboring ASxx (in this case the prefix for customer 30 and/or 40 will be advertised to ASxx as well).

Another case would be, if one of the metro region switches fail, G.8032 will switch over to the available link and traffic should be forwarded with minimum degree of service disruption. Suppose switch SW2 in metro region fails, customer 30 and 40 can still reach router R2 through switch SW1.

Worst-case scenario

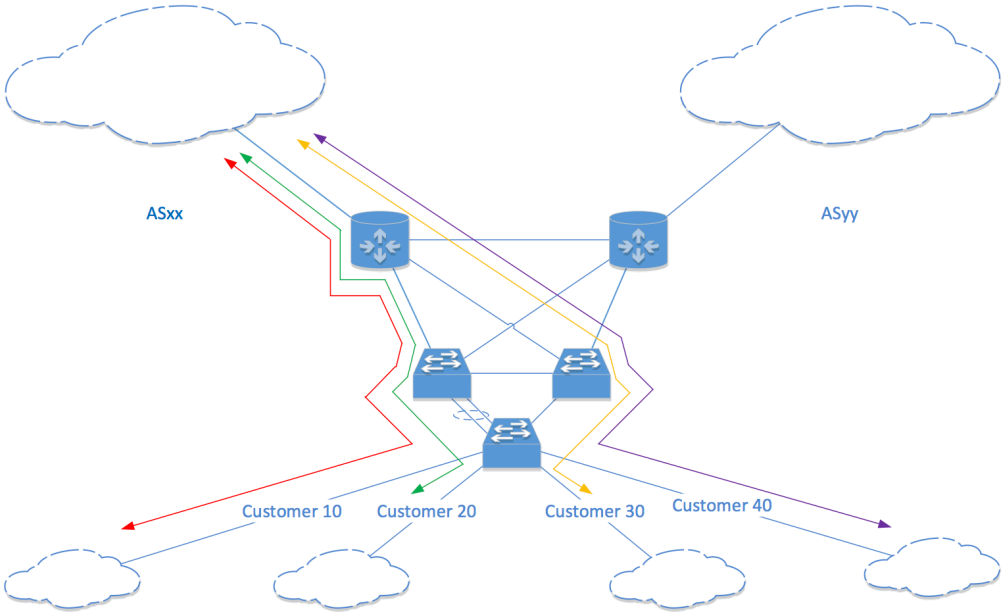

Figure 23: BGP Worst Case

In the worst-case load-sharing scenario, some devices may fail. For example, if the router R2 fails, the traffic can still be forwarded through router R1, considering ASxx is reachable. Considering another failure, switch SW1 or SW2 fails in the metro region. Traffic from all the customers can still get to router R1. But if router R1 also fails or ASxx goes unreachable, there is no way out.

Even worst case would be, if the MTU (switch SW3) goes down, no customer can reach the network even if ASxx and ASyy are reachable along with both router R1 and R2 running perfectly fine.

Conclusion

The Internet traffic is growing at an increasing speed. Keeping up with network throughput demand whether you are a service provider or an educational institution providing network services to users, is essential. An alternate solution to network throughput other than fatter pipes is load sharing as discussed throughout this paper. While it largely depends on network design, and also on the type of traffic the network is carrying, delay sensitive or not. We discussed some techniques promoting symmetric routing. Non-delay sensitive data may benefit from asymmetric routing.

In conclusion, load sharing largely depends on network design and its implementation. Steps can also be taken to improve load sharing in networks experiencing performance degradation. In hosted environments, solutions like Elastic Load Balancing (for EC2) can prove beneficial. But, for load balancing in underlying network infrastructure, ECMP/LAG or other load sharing techniques discussed in this paper can be used. Few other techniques worth mentioning that are not discussed in this paper are MPLS Traffic Engineering (Raszuk, 1999), Shortest Path Bridging (Kerravala, 2014), and TRILL (Eastlake 3rd, 2009).

Future work

BGP load sharing is different in a way that it requires ingress and egress traffic control to be implemented for deterministic control over traffic. The border routers can influence external BGP load sharing as they receive multiple routes from the neighboring AS, because they mostly receive multiple routes to same destination in dual-homed or multi-homed scenarios. Only best available prefixes are forwarded to internal BGP neighbors from the border routers, because of the BGP's default behavior. For some time, there has been an effort to improve this single best path problem in the topology mainly due to link flapping that triggers updates across the topology. One of the major goals of such efforts is to fast link restoration, by introducing multiple paths instead of one best path. Another goal achieved through introducing multiple paths is mitigation of MED oscillation.

BGP Add-Paths (Walton, D., Retana, A., Chen, E., & Scudder, J., 2013) provides such a capability by modifying the BGP protocol. BGP Add-Paths' introduction requires major software and possibly hardware changes in the network to accommodate modified BGP protocol. A BGP speaker that is willing to receive multiple paths from its peer, or send multiple paths, should advertise this capability to the peer. The speaker should use extended NLRI encoding to advertise multiple paths to it peer as specified in the document. As specified in the goals, its applications include: Eliminate persistent route oscillations, optimal routing and routing convergence. Standard BGP behavior dictates if a BGP speaker receives an advertisement of an NLRI and path from a peer that also subsequently advertises the same NLRI with different information like next-hop or BGP attributes, the new path effectively overwrites the existing path. But in BGP add-paths function, this behavior is overridden so that different attributes for the same prefix can be kept in the routing base. If the path identifier (as described in the document), is same as the NLRI information in the routing base, only then the route is replace or

withdrawn. In terms of scalability, add-paths consume more memory as it introduces multiple paths per prefix. There are many modes specified in the document to select from, to best suit the customer needs, but still the router memory and CPU consumption takes a toll over less powerful devices. Keeping in mind the memory requirements and the upgrade requirements, BGP add-paths proves to be a promising modification to mitigate some of the major issues associated with BGP routing updates today.

To (temporarily) address the implications of BGP add-paths, and interim solution introduced until BGP add-paths is standardized, is Diverse BGP Path distribution. It achieves multiple route distribution by using multi plane route reflection, where the main route reflection plane advertises/distributes the best path, and the next route reflector plane distribute second best path, and so on. Since the installation of one or more devices with multiple route reflector planes is easier than infrastructure wide upgrade, the implementation overhead is less. Benefits of Diverse BGP Path include: No BGP protocol modification, safe and gradual deployment on route reflector cluster basis, and does not require infrastructure wide changes. While Diverse BGP path buys more time for BGP add-paths standardization, it in no way, is intended to compete with add-paths, as stated in the document (Raszuk, E. R., Fernando, R., Patel, K., McPherson, D., & Kumaki, K., 2012).

Network Functions Virtualisation (NFV) is concept of taking network functions and handling them on industry standard virtual servers. With NFV, network functions will not require dedicated/proprietary hardware devices for different functions; instead they will leverage current traditional server virtualization techniques. Software Defined Networking (SDN), on the other hand, differs from NFV in a way that it focuses on separation of control plane of communication networks from the data plane. SDN focuses on separation of complex functions

on network communication devices and offloading them to device(s) connected through an out-of-band network. Out-of band networks have been used in the industry for decades, since the advent of AT&T SS7. NFV and SDN are often thought to be identical, confusingly. Following screenshot from the NFV white paper paints a picture of logical differentiation:

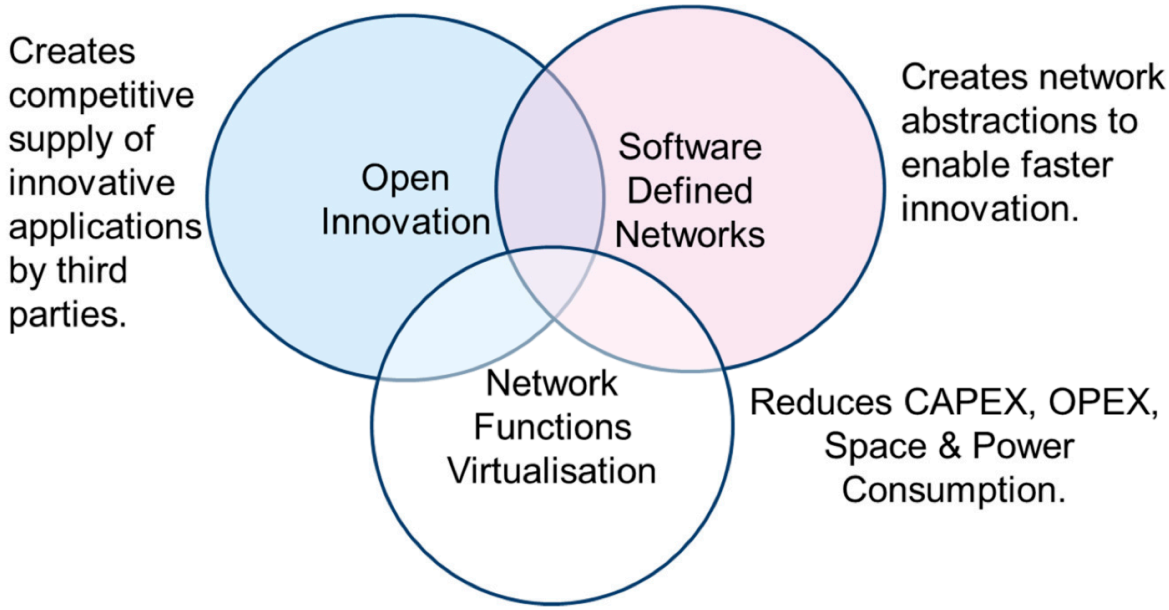

Figure 2: Network Functions Virtualisation Relationship with SDN

Figure 24: Venn diagram screenshot (European Telecommunications Standards Institute, 2012)

According to a case study by Google, their largest largest production network running on SDN and OpenFlow operating at 95% network utilization as mentioned in Google's Inter-Datacenter WAN Using SDN and OpenFlow Case Study, (Open Networking Foundation, 2012). As mentioned in the case study: "Centralized traffic engineering provides a global view of the supply and demand of network resources. Managing end-to-end paths with this global view results in high utilization of the links." Implementing SDN at a massive scale becomes problematic for small-to-medium organizations. Even Google had to make their own hardware devices as there was no OpenFlow support for hardware devices when they started this effort in

2010, as mentioned in the case study. Centralized traffic engineering does provide a great option to effective load balancing, but current market support is still limited. There is still more research to be done so that the industry can leverage this function feasibly.

References

Alcatel-Lucent. (2012). G.8032 Ethernet Ring Protection Single Ring Topology. Retrieved from https://infoproducts.alcatel-lucent.com/html/0_add-h-f/93-0267-HTML/7X50_Advanced_Configuration_Guide/G_8032.html

APNIC Training (2013, March 6). SP-Multihoming [Video file]. Retrieved from http://www.youtube.com/watch?v=_IDCVFKASic

Beijnum, I. V. (2002) BGP. Sebastopol, CA: O'Reilly & Associates, Inc.

Bollapragada, V., Murphy, C., & White, R. (2000) Inside Cisco IOS Software Architecture (CCIE Professional Development). Indianapolis, IN: Cisco Press.

Caforio, J. R. (2007, April). Hardware load-balancing device (HLD). Retrieved from http://searchnetworking.techtarget.com/definition/hardware-load-balancing-device

Cisco Systems. (2013). Threshold metric through track timer. Retrieved from http://www.cisco.com/c/en/us/td/docs/ios-xml/ios/ipapp/command/iap-cr-book/iap-t1.html

Cisco Systems. (2010, June 2). Cisco Visual Networking Index: Forecast and Methodology, 2009–2014. Retrieved from http://large.stanford.edu/courses/2010/ph240/abdul-kafi1/docs/white_paper_c11-481360.pdf

Cisco Systems. (2005, August 23). *Load Sharing with BGP in Single and Multihomed Environments: Sample Configurations* Retrieved from http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13762-40.pdf

Davis, D. (2009, January 8). How to use the Cisco IOS Policy-Based Routing Features. Retrieved from http://www.petri.co.il/how-to-use-cisco-ios-policy-based-routing-

features.htm

Eastlake 3rd, D. E. (2009, December). RBridges and the IETF TRILL Protocol. Retrieved from http://www.nanog.org/meetings/nanog48/presentations/Monday/Eastlake_RBridge_N48.pdf

European Telecommunications Standards Institute. (2012, October 22-24). Network Functions Virtualisation: An Introduction, Benefits, Enablers, Challenges & Call for Action. Retrieved from http://portal.etsi.org/NFV/NFV_White_Paper.pdf

Hinden, E. R. (2004, April). Virtual Router Redundancy Protocol (VRRP). Retrieved from https://tools.ietf.org/html/rfc3768

Hogg, S. (2009, June 29). Ethernet on a Ring [Web log message]. Retrieved from http://www.networkworld.com/community/node/43116

Holness, M. (2012, July 13). *ITU-T G-Series Supplement 52 Overview -- G.8032 Usage and Operational Considerations*. Paper presented at Joint IEEE-SA and ITU Workshop on Ethernet, Geneva, Switzerland. Retrieved from http://www.itu.int/en/ITU-T/Workshops-and-Seminars/ethernet/201307/Documents/S1P4_Marc_Holness.ppt

How does Hardware load balancing work?. (n.d.). Retrieved from http://www.hardwareloadbalancer.com/#how_does_load_balancing_work?

Institute of Electrical and Electronics Engineers. (2004, June 9). 802.1D : IEEE Standard for Local and metropolitan area networks Media Access Control (MAC) Bridges.

Institute of Electrical and Electronics Engineers. (2011, August 31). IEEE Standard for Local and metropolitan area networks — Media Access Control (MAC) Bridges and Virtual Bridge Local Area Networks.

International Telecommunication Union Telecommunication Standardization Sector. (2012,

February). G.8032/Y.1344: Ethernet ring protection switching. Retrieved from http://www.itu.int/rec/T-REC-G.8032-201202-I/en

Kerravala, Z. (2014, February 10). Inside the network powering the Sochi Olympics. Retrieved from http://www.networkworld.com/community/blog/inside-network-powering-sochi-olympics

Lappetelainen, A. (2011, March 17). Equal Cost Multipath Routing in IP Networks. Retrieved from http://lib.tkk.fi/Dipl/2011/urn100416.pdf

Milivojevic, M. (2011, January 24). Old CCIE Myths: Spanning-Tree Diameter. Retrieved from http://blog.ipexpert.com/2011/01/24/old-ccie-myths-spanning-tree-diameter/

Open Networking Foundation. (2012). Inter-Datacenter WAN with centralized TE using SDN and OpenFlow. Retrieved from https://www.opennetworking.org/images/stories/downloads/sdn-resources/customer-case-studies/cs-googlesdn.pdf

Raszuk, E. R., Fernando, R., Patel, K., McPherson, D., & Kumaki, K. (2012, November). Distribution of Diverse BGP Paths. Retrieved from http://tools.ietf.org/html/rfc6774

Raszuk, R. (1999). MPLS Traffic Engineering NANOG18. Retrieved from https://www.nanog.org/meetings/nanog18/presentations/raszuk.ppt

Rekhter, E. Y., Li, E. T., & Hares, E. S. (2006, January). A Border Gateway Protocol 4 (BGP-4). Retrieved from http://www.ietf.org/rfc/rfc4271.txt

Ruhann. (2009, June 3). Load-sharing vs Load-balancing [Web log message]. Retrieved from http://routing-bits.com/2009/06/03/load-sharing-vs-load-balancing/

Walton, D., Retana, A., Chen, E., & Scudder, J. (2013, October 16). Advertisement of Multiple Paths in BGP. Retrieved from http://tools.ietf.org/html/draft-ietf-idr-add-paths

Xu, Z. (2009) Designing and Implementing IP/MPLS-Based Ethernet Layer 2 VPN Services: An Advanced Guide for VPLS and VLL. Indianapolis, IN: Wiley Publishing, Inc.

Zhang, R., & Bartell, M. (2003) BGP Design and Implementation. Indianapolis, IN: Cisco Press.