High Availability and Faster Convergence Techniques in IP Networks

Er. Avtar Singh
Research Scholar (M.Tech)
Department Computer engineering
Punjabi University Patiala
Patiala, Punjab

Er. Harpreet Kaur
Assistant Professor
Department Computer engineering
Punjabi University Patiala
Patiala, Punjab

Abstract : This paper explains the importance of High Availability and Faster Convergence in today's Networks and the current set of high availability and faster convergence protocols and technologies used. Network Convergence and High Availability of IP networks are related with each other and both play a very important role in achieving the business continuity and to make IT works for any business. As Networks act as a base to transport the data from one part to other and most of the business units are dependent on IT for its functions to run, the paper includes those technologies and protocols that helps achieving best results to achieve around hundred percent business continuity with the help of networks.

Keywords : FHRP, LINK AGGREGATION, LACP, MPLS TE, Fast Reroute, LSA and SPF Pacing.

1.) INTRODUCTION

IT is changing and it is also bringing the revolutionary shift in the Businesses Worldwide. Almost every business is connected with IT and those who are not connected will use it in future as it has become a kind of a need to compete with competitors. Technologies like Cloud and Internet-of-Things are taking the Business innovations to the entire different level where you can achieve Machine to Machine connectivity and also can connect to your applications, data and machines from any place any time. One thing that is making these all things possible is Networks. Networks are acting as the beating heart of the business continuity as network is the thing that makes all the applications accessible to the clients or employees from any part of the world. Enterprises, Individual Users are moving their data to the private and public clouds and as more and more data is shifting towards the clouds ,like Amazon AWS or Rackspace for Enterprises or Apple iCloud, Dropbox, Google Drive etc for individuals, the importance of network availability has increased with the time. Stock Markets, Enterprises, SOHO Users etc, in order to always connect with their business applications requires different technologies and protocols related to Highly Available and Faster Convergence Networks. Faster Convergence and High Availability are two different terms, but are related to each other in a manner that both are needed to make business continuity work. Convergence in Networks means at what rate the data traffic shifts to the backup link, if primary link goes down. Better the convergence, higher availability can be achieved. Sub-Second convergence is what is needed in today's networks to make the business continuity work and to make the enterprise applications always available to the users. Convergence can be measured using the formula :. **Failure Detection + Event Propagation + Routing Process + FIB Update**

High Availability and Faster Convergence both work together in a way that faster the convergence higher the availability. A great example that one can take is of an e-commerce company like Amazon.com, if Amazon.com becomes unreachable for 5 minutes from his customers, how much bad impact(financial and reputation) does it make. Sub-Second convergence is what is needed when you are using VoIP in the network as VoIP uses User Datagram Protocol(UDP) for transporting Voice and Video traffic and delay can end the Voice or Video connection instantly.

On the other hand, High Availability is measured using the following formula :

Availability = (MTBF-MTTR)/MTBF

where**MTBF** is mean time between failure means "What, when, why and how does it fail ?"and**MTTR** is mean time to repair means "How long does it take to fix ?"

	Availability	DPM	Downtime Per Year (24x365)		
Reactive?	99.000%	10000	3 Days	15 Hours	36 Minutes
	99.500%	5000	1 Day	19 Hours	48 Minutes
Proactive?	99.900%	1000		8 Hours	46 Minutes
	99.950%	500		4 Hours	23 Minutes
Predictive?	99.990%	100			53 Minutes
	99.999%	10			5 Minutes
	99.9999%	1			30 Seconds

Figure 1.1 - High Availability Measurement Table

Above high availability measurement table shows availability of networks in terms of percentage and downtime per year and in today's network era, 99.999% and 99.9999% are termed as highly available networks.

An highly available or predictive network needs to have:

- There should not be single points of failures.
- Fault, performance and workflow process tools.
- Excellent consistency is needed with Hardware, Software, Configuration and design.
- Consistent processes for fault, security and performance.

Following are the faster convergence and high availability protocols that we have used in IP Networks:

A)First Hop Redundancy Protocols - FHRPs mainly comes in three variants i.e. HSRP, VRRP and GLBP. These three protocols are used for gateway redundancy and is used for high availability of networks. Hot Standby Router Protocol and Gateway Load Balancing Protocol are licensed under Cisco Systems Inc., while Virtual Router Redundancy Protocol is an Open Standard Gateway Redundancy Protocol and is defined in IETF RFC 3768. In both HSRP and VRRP, Load Balancing is enabled by default and in case user requires load balancing, he has to configure multiple groups, while GLBP does load balancing by default. All these protocol use the concept of virtual IP which is shared among different gateway routers and primary to backup link failure depends on the timers, which can be reduced to single second or in ms(in case of HSRP and GLBP). Below is the figure showing HSRP scenario:-

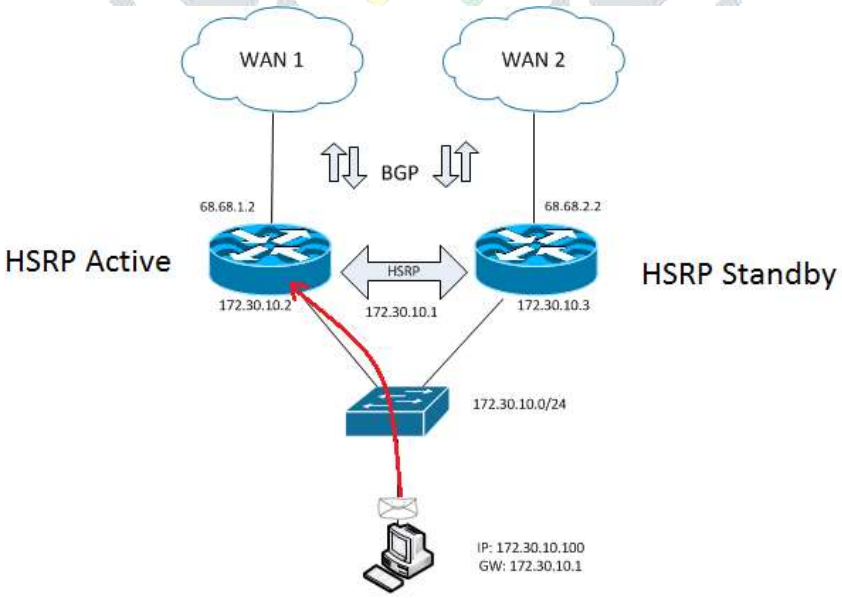

Figure 1.2 - HSRP Scenario - Source - https://thecciejourney.wordpress.com/2014/10/12/hsrp-no-longer-for-the-weak-of-heart/

B)OSPF Fast Hello Detection, LSA Group Pacing and Tuning SPF Timers :

We can detect a neighbor failure or link failure using Polling interval method by polling through fast hellos which are transmitted at Layer 2 and Layer 3. We can fasten the hello packet interval time in milliseconds, for example ospf can transmit 5 hello packets every 1 second and the dead timer is set to 1 second. If we need to configure 200 millisecond hellos with OSPF, then we can set 5 hello packets in a dead interval of 1 using the following command in Cisco IOS software :ipospf dead-interval minimal hello-multiplier [multiplier], where multiplier is the number of hello packets that we need to send in 1 second.

OSPF produce large bursts of LSA Flooding traffic every 30 minutes, while individual aging do fragmentary re-flooding. LSA group pacing feature would help in controlled bursting. For example if we change the group pacing timer to 10 minutes, a small batch of LSAs that are close to be aged-out are processed together. A figure below shows LSA group pacing effect :

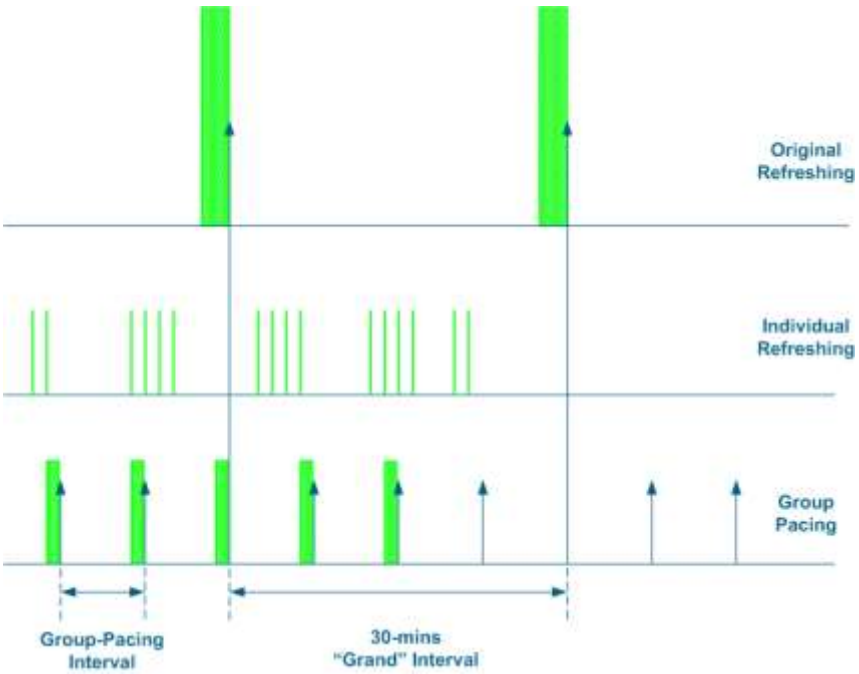

Figure 1.3 LSA Group Pacing in OSPF

Tuning SPF Timers comes under OSPF Exponential Backoff for the generation of Link State Advertisements. It is also known as LSA throttling. It includes three attributes :

- **spf-start -** It is the initial Shortest Path First schedule delay in milliseconds.
- **spf-hold**- It is the minimum hold time between two consecutive SPF calculations.
- **spf-max-wait** - It is the maximum wait time between two consecutive SPF calculations.

C)MPLS Fast Re-Route

Multiprotocol Label Switching is used in almost all the service providers in the world. It is the technology that is the heart of Internet Service Provider core networks. From one Provider edge to other provider edge, mostly there are two or more than two paths. For faster convergence and for link protection in case of link failure, MPLS Fast Reroute is used. MPLS FRR provides protection against link or node failures. The FRR mechanism provides sub-second convergence by having backup path pre-calculated which is used in case of primary link failure. It allows data flow to continue even when the headend router tries to create a new end-to-end Label Switch Path that is used to bypass the failure. Notification of primary link failure to headend router is sent by Interior Gateway Routing Protocol like OSPF and ISIS and through RSVP. Below figure shows the MPLS FRR process :

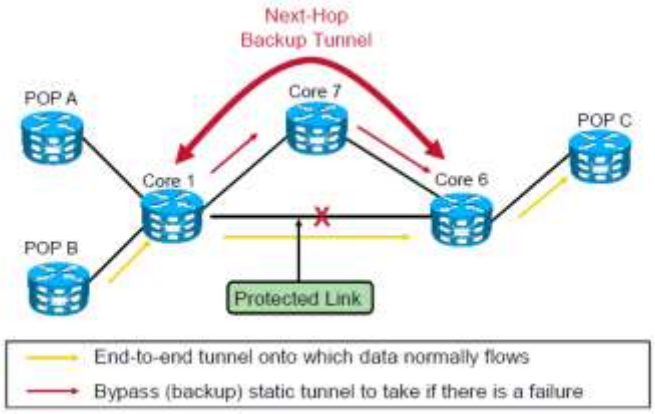

Figure 1.4 - MPLS FRR Sample Process

In the above MPLS FRR figure, it is shown that when primary link between Core 1 router of ISP and Core 6 goes down, then the backup link comes up immediately. Backup link come to use immediately within a second helps in almost zero loss of data. LSP paths are calculated at the Headend. Backup tunnels that bypasses next-hop nodes for LSP paths are known as next-next-hop backup tunnels and the reason is that they terminate at the node which is following the next-hop-node of the LSP path, therefore it bypasses the next-hop-node. Protection from failure of nodes is also made sure with MPLS FRR. Therefore if a node along the LSP path goes down, then the LSP is created over the backup with less than a second, and the convergence is very fast. Figure below shows the node failure and MPLS FRR.

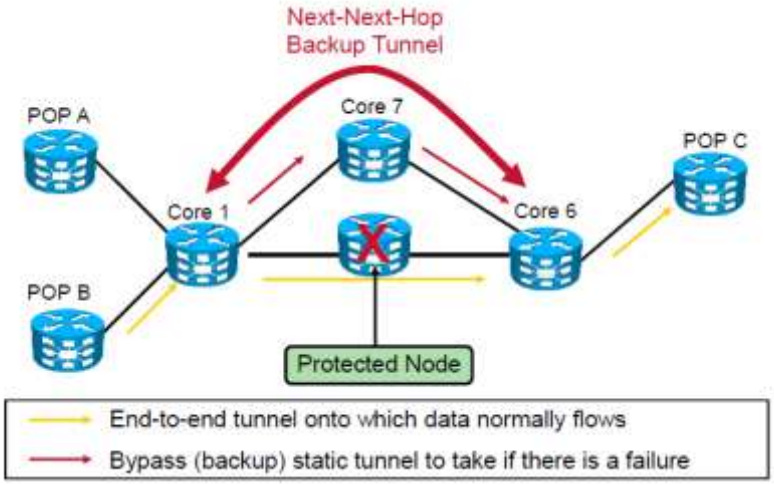

Figure 1.5 - MPLS Core Node Failure and how MPLS FRR does a sub second convergence

D)Link Aggregation - Also known as Ether Channels or NIC Teaming, it is used to aggregate multiple physical links to create a single logical link. Multiple 8 links can be in use. Different methods can be used to bundle multiple links like we can use static method by using On mode, or we can use dynamic method by using protocols to negotiate EtherChannel. Protocols that can be used are Port Aggregation Protocol and Link Aggregation Channel Protocol. In LACP, maximum 16 links can be bundled, but a maximum of 8 wil be in use at a time. So the capacity can reach up to 80 Gbps, if we bundle 8 links of 10Gbps each. Traffic between the links can be decided on the basis of Load Sharing Algorithms used with Ether Channels. If a single address based load sharing method is used, then the BITS method, while XOR is used in case where source and destination addresses are used in load sharing mechanism. Below is a simple scenario of LACP :

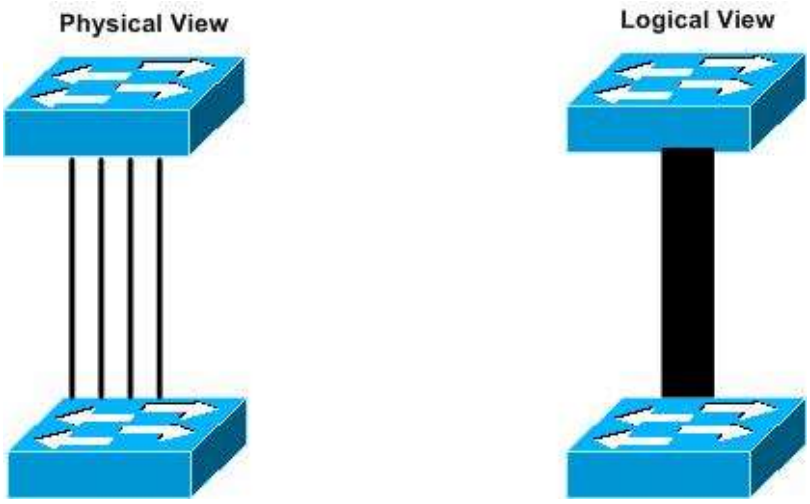

Figure 1.6 - Link Aggregation or EtherChannel, Source - http://ericleahy.com/index.php/etherchannels/

2.) LITERATURE SURVEY–

MPLS Traffic Engineering – Fast Reroute [1] by ShugufthaNaveed, S. Vinay Kumar of Vasavi College of Engineering(Osmania University), Hyderabad in May, 2014 under IJSR – ISSN: 2319-7064 draws a conclusion that in the event of link failure, traditional recovery technologies takes unacceptable time in case of VoIP and Video based critical solutions, while MPLS traffic engineering Fast Reroute meets the requirements of real-time applications with fast recovery that facilitate high availability to converge. The research shows that MPLS Fast Reroute method provides great performance in case of link failure as compared with traditional IP networks..

Fast Reroute Extensions to RSVP-TE for LSP Tunnels [3] by P. Pan, Ed. Of Hammerhead Systems, G. Swallow, Ed. Of Cisco Systems and A. Atlas, Ed. Of Avici Systems in IETF RFC 4090 defines RSVP-TE extensions to establish backup label-switched path(LSP) tunnels for local repair of LSP tunnels. These mechanisms enable the re-direction of traffic onto backup LSP tunnels in 10s of milliseconds, in th event of failure.

Survey on the RIP, OSPF, EIGRP Routing Protocols [4] by V. Vetriselvan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1058-1065, specifies a performance evaluation of various routing protocols with certain criteria's like Jitter, Convergence Time, end to end delay.

Bidirectional Forwarding Detection (BFD) [5] by D. Katz and D. Ward of Juniper Networks in IETF RFC 5880 describes a protocol intended to detect faults in the bidirectional path between two forwarding engines, including interfaces, data link(s), and to the extent possible the forwarding engines themselves, with potentially very low latency. It operates independently of media, data protocols, and routing protocols.

Bidirectional Forwarding Detection (BFD) on Link Aggregation Group (LAG) Interfaces [6]by M. Bhatia of Alcatel-Lucent, M. Chen of Huawei Technologies, S. Boutros, M. Binderberger of Cisco Systems. J. Haas of Juniper Networks in IETF RFC 7130 defines a mechanism to run Bidirectional Forwarding Detection(BFD) on Link Aggregation Group(LAG) interfaces.

Graceful OSPF Restart by J.Moy of Symacore Networks, P. Pillay-Esnault of Juniper Networks and A .Lindem of Redback Networks in IETF RFC 3623 [7]describes where an OSPF router can stay on the forwarding path even as its OSPF software is restarted. This process is called "graceful restart" or "non-stop forwarding". In this paper, operation of the restarting router, Graceful restart advantages in unplanned outages, its format is defined.

OSPFv3 Graceful Restart by P. Pillay-Esnault of Cisco Systems and A. Lindem of Redback Networks in IETF RFC 5187[10] describes the OSPFv3 graceful restart mechanism. It is pretty much identical to OSPFv2. There are very few differences which are specified in the document. This includes the format of the grace Link State Advertisements(LSAs).

Basic Specification of IP Fast Reroute for IP Fast Reroute: Loop-Free Alternatives by A. Atlas, Ed. Of British Telecom and A. Zinin, Ed. Of Alcatel-Lucent in IETF RFC 5286[11] describes the use of loop-free alternatives to provide the local protection for unicast traffic in pure IP and MPLS networks if a failure on a link occurs. The objective of MPLS Fast Reroute is to reduce the loss of packets that happens while the routers converge after the primary link failure. It has a rapid failure repair with the pre-calculated backup next-hops towards the destination.

Fast Reroute Extensions to RSVP-TE for LSP Tunnels by P. Pan, Ed. of Hammerhead Systems, G. Swallow, Ed. of Cisco Systems, A. Atlas, Ed. of Avici Systems in IETF RFC 4090[12] in May 2005 defines two methods in this paper. First is a one-to-one backup

method that creates detour LSPs for each protected LSP at each potential point of local repair. Other one is a facility backup method that creates a bypass tunnel to protect a potential failure point by using MPLS stacking, this bypass tunnel can protect the set of LSPs that have similar backup constraints. Both these methods are used in case of link or node failures.

3.) PROBLEM DEFINITION -

• High Availability in Networks is one of the major goals of every company. Large downtime of network in a company creates big loss in companies(data centers, ISPs, Enterprises, E-commerce Companies). Various High availability protocols can be used, but all of them are used according to a specific network design. A specific set of protocols are used for different layers, all of them have different requirements, both economically and infrastructure wise.

• Same is the case with Faster convergence as if not properly implemented, the results can be severe.

• So designing a highly available and faster converging network is a very difficult task with set of different protocols trying to achieve the same.

• As world is diving towards VoIP and Video solutions, that needs high bandwidth and low delay, faster convergence has become much more important.

4.) OBJECTIVES

a.) Comparative analysis of High Availability technologies and Faster Convergence Technologies such as MPLS Fast Reroute, Tuning SPF Timers , LSA Pacing Timers, OSPF Fast Hello Mechanism, and Dampening will be done.

b.) Selection of best High Availability and Faster Convergence method in Medium to Large Service Providers and Data Centers.

c.) Selection of best High availability and faster convergence methods that generates minimum delay for VoIP based networks.

5.) METHODOLOGY

- To study High Availability and Faster Convergence Standard and informational documents which are used by different vendors while implementing protocol on their devices.
- Design a network in GNS3 simulation environment and also on Real Cisco Devices.
- Implementation of High Availability and Faster Convergence Protocols in a simulated service provider and enterprise network environment.
- Implementation of Simple Network Management Protocol(SNMP) between Network Monitoring System(NMS) and Routers running in GNS3 and in real.
- Monitoring Tool that will be used is Paessler Router Traffic Graph(PRTG) to draw output graphs that will help us comparing different outputs.

6.) RESULTS AND DISCUSSIONS

Following section includes the results and topologies we have used in our work:

5.1 LSA Throttling/Tuning SPF Timers –

SPF and LSA throttling can be done to minimize the convergence in a network where Link State Routing Protocols are used. Below is the topology that we have used:

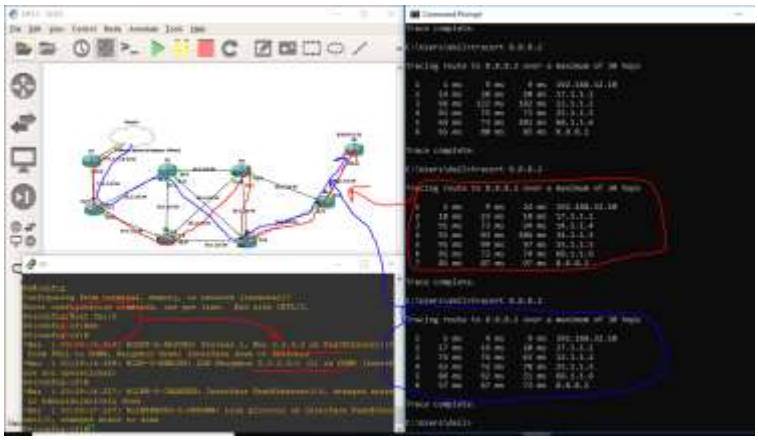

Figure 5.1 – SPF Topology Link Convergence

SPF timers can be throttled in order to fasten the protocol convergence at the Network Layer. Below is the result of tuning the SPF and LSA:

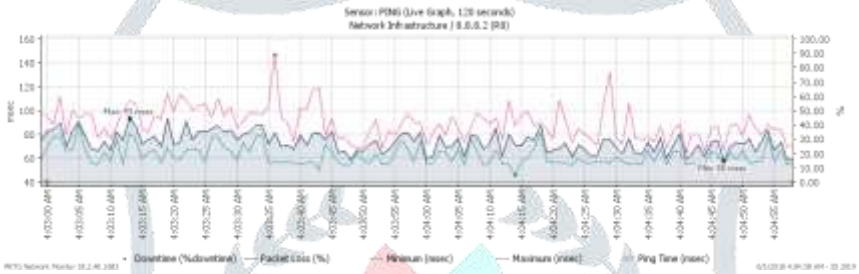

Figure 5.2– SPF/LSA Throttling convergence time

As one can see in the above figure, delay is not seen in the graph and if there is any delay that also is pretty meagre which is not recognizable at all.

5.2 MPLS Fast Reroute

MPLS FRR is mainly used in the MPLS Backbone networks to reduce the convergence time between the primary to backup link shift. Below is the topology used for MPLS FRR:

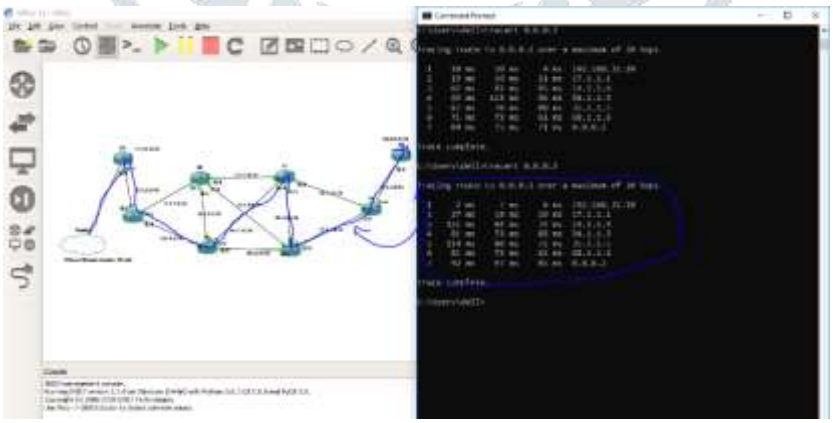

Figure 5.3– MPLS FRR Topology and primary path

As shown above in the GNS3 topology, CE-CE traffic goes via 17.1.1.1(R1)->14.1.1.4(R4)->34.1.1.3(R3)->35.1.1.5(R5)->68.1.1.6(R6), which is the primary path from PE1-PE2. In case of link failure, as we have multiple paths in the PE-PE path, traffic shifts to the next best path towards destination as shown below:

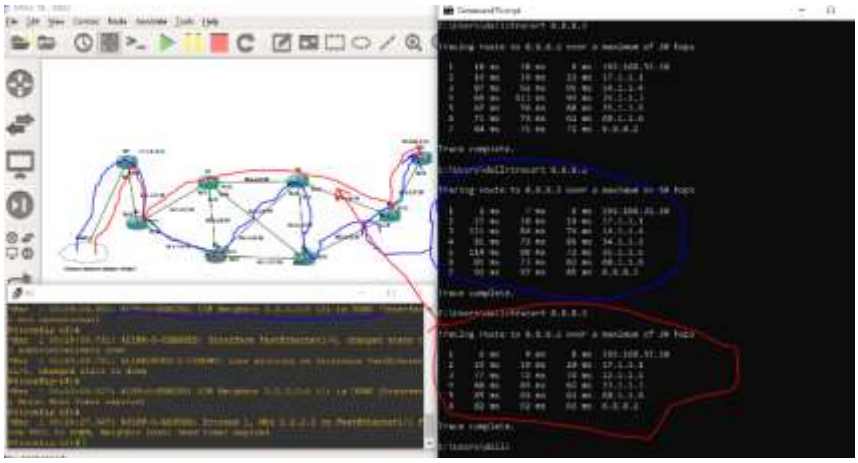

Figure 5.4 – MPLS Backbone Link Converged from primary to backup link

As shown above in the figure, link is converged from primary to backup path as the new path now appears to be 17.1.1.1(R1)->12.1.1.2(R2)->23.1.1.3(R3)->68.1.1.6(R6). Default MPLS Backbone does not provide fast convergence and high availability automatically. Belwo graph shows the convergence time it takes in MPLS Backbone when shifting the link from one path to another.

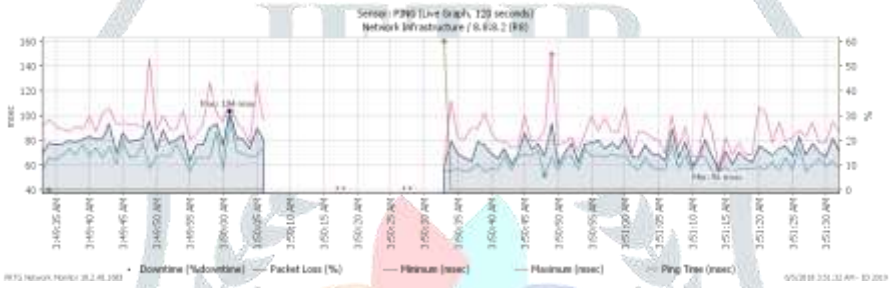

Figure 5.5 – Default MPLS Backbone Convergence Time without TE configured.

MPLS FRR is configured in the topology under MPLS Backbone and below is the graph in PRTG that shows the amount of time it took for the convergence:

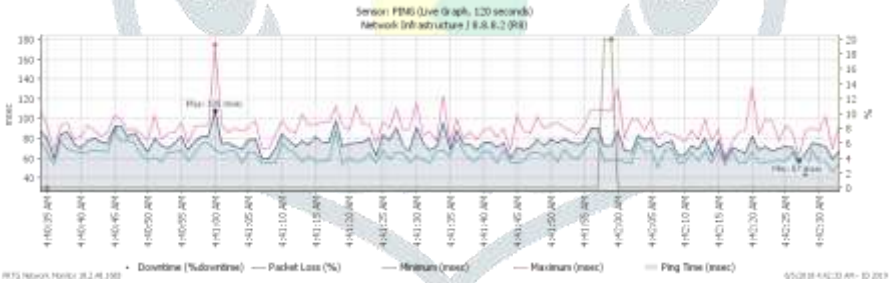

Figure5.6 - MPLS FRR Convergence Graph.

Above graph clearly shows that MPLS FRR provides with in a second convergence between the MPLS Backbone paths from one PE to other PE.

5.3 First Hop Redundancy Protocols

FHRP is the suite of gateway redundancy protocols with HSRP, VRRP and GLBP as their three major protocols. They are also used for faster convergence and higher availability at the access layer. We have used VRRP in our topology to provide faster convergence at the gateway level. Below is the topology used:

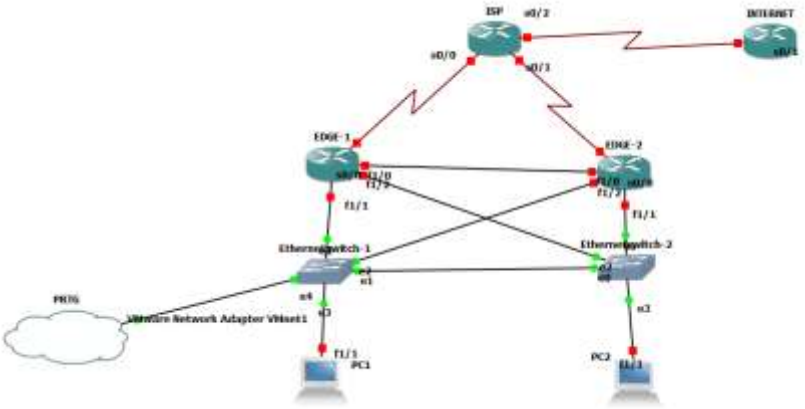

Figure 5.7 – VRRP Topology

As shown in above topology, two edge routers are used to connect with ISP for internet services. As there are multiple gateways for host machines running at access layer, there are multiple physical gateways present which helps at edge layer redundancy which means that in case the primary link of Edge-ISP goes down, backup link comes up, but now the problem that arises is that Host Machines can have only a single gateway, so if the primary to backup link is shifted, then one have to manually change the gateway of the host machines which results in higher amount of delay with large number of host machines. VRRP solves this problem by sharing a Virtual IP between the gateways and if primary link goes down, then there is no need to change the gateway. We can lower the time of primary of backup link convergence to 1 second which is way better than changing manually. Below is the graph depicting the Convergence from primary to backup link with normal timer parameter:

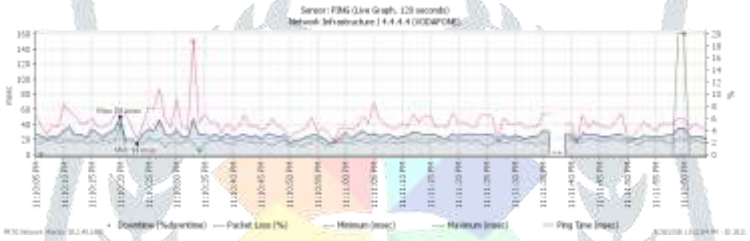

Figure 5.8 – VRRP Convergence with Normal Timer Parameter.

Above graph created in PRTG shows the delay to be around 3 seconds which can be tuned to reduced by tuning the timers to much lower value. Below graph shows the convergence time of VRRP primary to backup link failure with timers tuned:

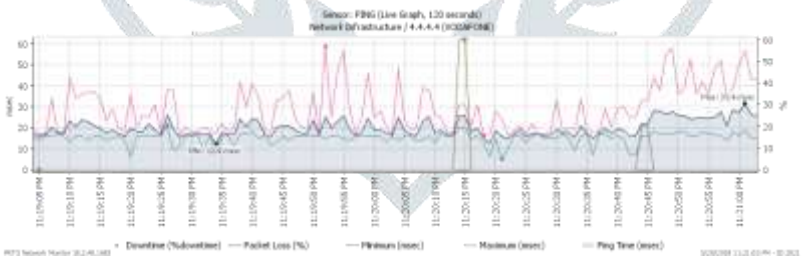

Figure 5.9- VRRP Convergence with Tuned Timer Parameter.

Above graph clearly shows a much better convergence when compared with normal timer parameter. With tuned timer parameter, convergence time is reduced to within one second.

5.4 Link Aggregation Channel Protocol–

LACP is used in Data Centers in order to improve the bandwidth and redundancy by binding multiple physical links to make a single logical channel. Max 8 physical links can be used to bind in to a single logical channel. When we have multiple links between the switches, Spanning Tree Protocol or STP makes utilization of only a single link possible and rest all the links are put in to the blocking state. Because of that, if a single primary link goes down, then the backup link goes through different states i.e. Blocking->Listening-> Learning->Forwarding, that takes around 30 seconds to make backup link in the data plane. Below is the graph that shows convergence time from primary to backup link:

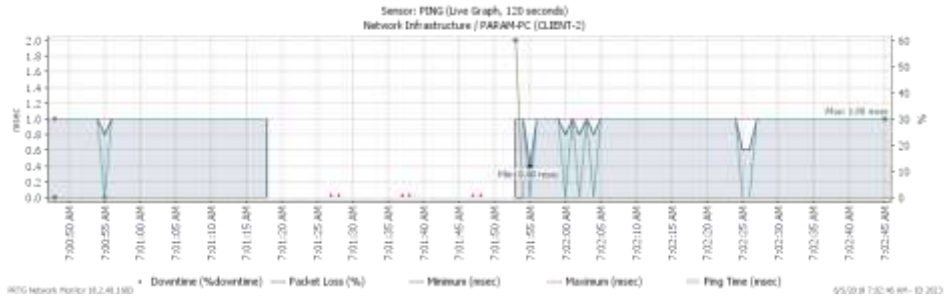

Figure 5.10 – Layer 2 Convergence without LACP

In case, LACP is used all the links between the two directly connected switches acts as they are a single logical link and there is no backup link as all the links up to 8 are in the forwarding state. Graph below shows the amount of convergence time used in case of a physical link failure during LACP configuration:

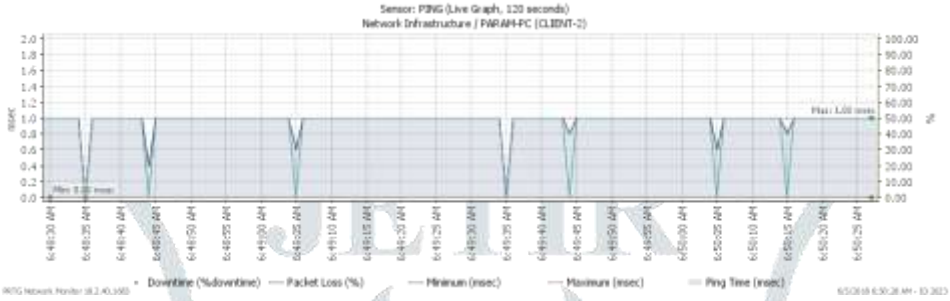

Figure 5.11 – LACP Convergence Time in case of physical link failure.

7.) CONCLUSION& FUTURE SCOPE

Networks are getting faster and faster over time and with new applications like self-driving cars, automation, IoT, cloud and VoIP services are increasing, higher availability of internet and faster convergence is the need of time to run these sort of applications where even a single second of a delay can bring the application to destroy the whole working. Data is stored in the data centers in case of cloud storage and users reach to the data using Internet from internet service providers, so there are different organizations that play an important role in defining the actual convergence and quality of the link. Different technology services can be used in order to better the convergence time that automatically brings high availability in the network. When talking about data centers servers and storage is connected with Layer 2 Network switches directly either using Top-of-Rack or End-of-Row topologies. Servers can be connected with Switches with multiple links connected and creating an Ether Channel by using any link aggregation technique like PAGP or LACP which automatically reduces time to with in a second in case of switch to switch direct multiple link connectivity or in case of server to switch connectivity. VRRP can be used at the edge of the data centers and enterprise networks in order to provide faster convergence which can be as fast as with in a second so that there is no need to change the default gateway of the servers or the host machines. SPF and LSA throttling or tuning can be made in the service providers where link state routing protocols like OSPF or ISIS are used almost every time for the interior gateway routing protocol. Tuning the SPF protocol to fasten the convergence brings the service provider PE-PE MPLS Backbone convergence with in 1 second. This can also be achieved in the enterprise networks where OSPF is used in medium-large scale enterprises. MPLS FR can be used when QoS and TE is in use with MPLS Backbone and that helps in reducing the convergence time to around 1 second.

As network speeds are increasing and technologies like Voice and Video are also increasing in the industry. New innovations in the industry and technologies related with automation needs internet services and convergence as fast as nano second, and Software Defined Networks(SDN) is the next big thing in the network and also an 5G enabler so the traditional network will obsolete over the time and faster convergence and high availability of the SDN controllers and controller->controller interaction and controller->switch interaction and convergence is what is needed to be ultra-fast. Future Work in SDN based faster convergence technologies is needed to be done as it is an emerging technology and one of the most researched technologies these days.

8.) REFERENCES

[1] MPLS Traffic Engineering – Fast Reroute by ShugufthaNaveed, S. Vinay Kumar of Vasavi College of Engineering(Osmania University), Hyderabad in May, 2014 under IJSR – ISSN: 2319-7064.

[2] Virtual Router Redundancy Protocol (VRRP) by R. Hinden, Ed. Of Nokia in Internet Engineering Task Force – RFC 3768

[3] Fast Reroute Extensions to RSVP-TE for LSP Tunnels by P. Pan, Ed. Of Hammerhead Systems, G. Swallow, Ed. Of Cisco Systems and A. Atlas, Ed. Of Avici Systems in IETF RFC 4090

[4] Survey on the RIP, OSPF, EIGRP Routing Protocols by V. Vetriselvan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1058-1065

[5] Bidirectional Forwarding Detection (BFD) by D. Katz and D. Ward of Juniper Networks in IETF RFC 5880.

[6] Bidirectional Forwarding Detection (BFD) on Link Aggregation Group (LAG) Interfaces by M. Bhatia of Alcatel-Lucent, M. Chen of Huawei Technologies, S. Boutros, M. Binderberger of Cisco Systems. J. Haas of Juniper Networks in IETF RFC 7130

[7] Graceful OSPF Restart by J.Moy of Symacore Networks, P. Pillay-Esnault of Juniper Networks and A .Lindem of Redback Networks in IETF RFC 3623

[8] Graceful Restart Mechanism for BGP by S. Sangli, E. Chen of Cisco Systems, R. Fernando, J. Scudder, Y. Rekhter of Juniper Networks in IETF RFC 4724

[9] Graceful Restart Mechanism for BGP with MPLS by Y. Rekhter and R. Aggarwal of Juniper networks in IETF RFC 4781

[10] OSPFv3 Graceful Restart by P. Pillay-Esnault of Cisco Systems and A. Lindem of Redback Networks in IETF RFC 5187

[11] Basic Specification of IP Fast Reroute for IP Fast Reroute: Loop-Free Alternatives by A. Atlas, Ed. Of British Telecom and A. Zinin, Ed. Of Alcatel-Lucent in IETF RFC 5286

[12] Fast Reroute Extensions to RSVP-TE for LSP Tunnels by P. Pan, Ed. of Hammerhead Systems, G. Swallow, Ed. of Cisco Systems, A. Atlas, Ed. of Avici Systems in IETF RFC 4090