Networking Protocols for Multipathing in Cloud

Siddhi Khamitkar

Computer Engineering Department

Mukesh Patel School of Technology Management and Engineering

NMIMS University

siddhi.khamitkar94@gmail.com

Abstract—Cloud computing is an ever-growing field in today‘s era. With the accumulation of data and the advancement of technology, a large amount of data is generated every day. Storage, availability and security of the data form major concerns in the field of cloud computing. With the beginning of virtualisation, data centre networking is reaching a high level of management. Interconnection in the data centres is not just based on practical knowledge but also on the traffic present in the particular network and the traffic engineering concerns. Scalability is a major concern for cloud service providers and data centre managers. Many researchers have proposed scalable ethernet as a solution for scalability but complex networking is a drawback for the scalable ethernet. Specific wiring topologies needed can create problems in deployment, network switch configuration. Network today, is the bottleneck which prevents businesses from realising the benefits of the information infrastructure. In this paper, various network protocols are discussed for multipathing and an insight is also given to view the improvisations from traditional data centres to the present day real time data centres. In this paper, we discuss about protocols such as LISP, MPTCP, TRILL, SPAIN, spanning trees and various other technologies, their proposals and the protocols used.

Keywords: Cloud computing, networking, protocols, multipathing

I. INTRODUCTION

Scalability is an important concern when it comes to cloud computing. High bandwidth, low latency is possible and advantageous in a scalable network. The cost of the equipment by a particular service provider is reduced only if the commodity is available off the shelf (COTS).Ethernet is highly used for connectivity in present times, specially over the fibre channel (FCOE).Self configuration, high bandwidth at low cost, high speed are all the benefits of the ethernet. The low scalability problem of the ethernet is solved by IP routers but which are available for a higher price. Layer 3 IP is added over the Layer 2 ethernet.

Several applications are hosted by cloud providers of different clients, thus providing a multi-tenant scenario in which the servers and the network is shared by tenants among each other. Without changing the tenants' systems, software's and protocols, services should be made available to the tenant by the cloud service provider. The requirement for high transmission rates, high bandwidth, less jitter are placed by the tenant to the provider. Multipathing is one of the ways in which these requirements can be met. Usually, redundancy is deployed at the data centre to avoid loss of information or data. Topologies are formed with multi rooted trees to offer multiple paths for fault tolerance. For easy deployment of services, it should be ensured that minimum modification is done at the infrastructure level.In this paper, an approach is discussed as well for efficient multipathing.This approach is formulated by using genetic algorithms.

In section II of the paper, we describe the issues related to the networking in traditional data centre and what is the ideal data centre network topology to overcome the drawbacks of the traditional data centre. In section III, we describe the two phase multipathing scheme evaluated by genetic algorithms. In section IV, a brief description is given about SPAIN COTS whereas in section V, a short brief about various protocols used for multipathing in networking is given.

II. COMPARISON BETWEEN LEGACY DATA CENTRES AND MODERN DATA CENTRES

A. Drawbacks of legacy data centres

To be concise, legacy data centres are highly location dependant and complex which greatly impacts their performance and has a negative effect on the user experience as well.

1) *High Complexity:* With the growth of the data centre, the number of devices also increase. It is necessary to manage the switch points and the switch interactions. The number of interactions between switches is given by the formula: n*(n-1)/2.Thus, according to the formula, 10 switches can generate 45 interactions while 100 can generate nearly 5000 interactions.

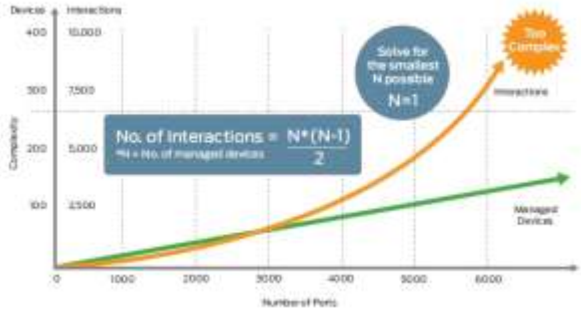

Fig 1: Growth in the number of interactions between switches when new devices are added to the network.[1]

To avoid this high a complexity, subnetworks are created within a network. If in such a scenario, spanning tree protocol is used, half the bandwidth will be unavailable as STP disables half the available links. Thus, in an organisation, the number of switches should be relative to the number of servers [1].

2) *High location dependency:* Application performance highly depends on the physical location of the servers ie where each server resides in the hierarchical structure. All switches are arranged in a hierarchical pattern in layers. This architecture forces each packet to transit switches up and down in the tree. Thus, the packets take divergent paths, thus low latency is difficult to maintain. Thus not fulfilling the SLA's as well.[1]

Engineering the networks to manage changing workloads is also difficult. The network traffic is engineered in such a way that it makes less number of hops, i.e. the source and destinations of the flows are close. This improves the performance of the application.

3) *Inequality of data centres:* Businesses often have to make trade-offs between performance, scalability, management complexity and cost as the requirements by each business differ. Some companies require a high performance for some interval of time and at regular times, require a normal performance. Instead of using individual equipment's provided by individual vendors, it is better to have an integrated product in the data centre network which addresses all their needs.

B. Ideal data centre characteristics

An ideal data centre offers simplicity and high performance.

1) *Simplicity:* The best way to induce simplicity in a network is to make the entire network behave as a single switch. Inside every switch is a flat fabric mesh and provides random connectivity between ports. There is no need for sharing protocols such as STP or control or management that is required. Server, storage and other devices can be connected seamlessly to the fabric offering high scalability and the performance does not degrade even if there are hundreds or thousands of ports [1].

2) *Performance:* With a flat fabric, the data path is all simplified. Latency is minimised and it is ensured that all the members of the fabric are via a single hop and each connection benefits from the bandwidth.

Eliminating tyres of switches has additional benefits. The fabric based data centre network requires few devices and interconnections which reduces the cost of building network infrastructures. Thus a fabric based architecture can pay heed to a lot of data centre requirements. By offering high scalability, performance and providing cost benefits, a fabric based network eliminates the need for IT to trade off one capability for the other [1].

III. TWO PHASE MULTIPATHING

A two phase multipathing scheme divides the problem in two phases: 1) Multipath configuration and 2) Multipath

Selection.

In multipath configuration, multiple paths are created using genetic algorithms with the objective of minimising path cost and maximising link usage diversity. The multipath share a common feature of the VLAN switches, ie they use VLAN tag to map multiple paths to multiple trees. Thus mapping problem is simplified by Genetic Algorithms which allows the creation of multiple VLAN trees with multiple objective functions such as the two mentioned above.

The two phase multipathing scheme explores path redundancy using algorithms. First, several independent trees are generated by the scheme to interconnect various servers of the data centres using genetic algorithms to minimise path sizes and maximise link usage diversity. Each tree is configured with a separate VLAN tag with the configuration performed offline, thereby not delaying the process of scheduling workflows. A network controller may take the advantage of SNMP (Simple Network Management Protocol) to gather topology and take VLAN advantage in switches. Only one path exists between each pair of servers. The use of several trees at a time spreads the traffic which increases the throughput and the load balancing. A single tree is selected to forward the flow and hence the packet follows only one path to avoid reordering at the receptor [4].Tree selection uses heuristics based on network usage.Smart Path Assignment in Networks (SPAIN) uses a similar technique to explore paths in a data Centre. The SPAIN approach is described in the next section.

IV. SPAIN COTS

SPAIN pre-computes the set of paths which utilises the redundancy in physical wiring. To provide high bandwidth and increased fault tolerance. These paths are then merged into a set of trees and each tree is mapped to a separate VLAN.These VLANs are installed on switches. Generally, a few number of switches are required to cover the network since a single VLAN ID can be re-used for multiple disjoint subtrees.

SPAIN highly improves performance as it uses COTS (Commodity off the Shelf) which provides the best TCO (Total Cost of Ownership).These COTS switches are used unmodified. SPAIN works with arbitrary topologies, supports incremental deployment and requires no centralised controllers. [6]

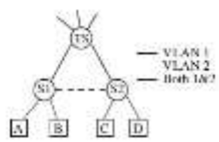

Fig 2: VLAN‘s used for multipathing.[6]

With SPAIN, a pair of end hosts can use different VLAN’s traversing potentially through different links at different times for different flows. Thus, high throughput and better fault tolerance can be achieved than the spanning tree.

In the figure shown above, VLAN 1 is reserved to include all the nodes.Thus, it is always available as a fallback path. Certain switch features required by SPAIN include:

- MAC address learning and VLAN support (already present in COTS switches). SPAIN can also exploit other switch features for improvised performance, high scalability and fault tolerance
- Switch is required to store multiple table entries. (One table per VLAN).

As shown in the above figure, Fig 2, there are two VLANs.VLAN 1 for normal spanning tree and VLAN 2 for the alternate link. Once the two VLANs are configured, VLAN 1 is used by endhost A for flows to C and VLAN 2 used by endhost B for flows to D, thereby doubling the bandwidth [6]. Offline configuration of the network includes

- Set of paths to be used.
- Mapping paths to VLANs.
- Handling unplanned topology changes.

V. NETWORKING PROTOCOLS

A. MPTCP-Multi Path TCP

MPTCP is an ongoing effort by engineers to enable the simultaneous use of IP addresses by a modification of the TCP which presents a regular interface to the applications, while in fact spreading data across various sub flows [9].

B. LISP-Locator/Identifier Separation Protocol

IP addresses, today, are assumed to have two functions:

1) Localisation 2) Identification of the owner.

When it comes to LISP, each endpoint (EID) is associated to one or many ip addresses of intermediate IP interfaces. named Routing LOCators (RLOC) which are typically supposed to be at the border routers of the endpoint network. Upon reception of the packet from the local network to an outer EID, the border acts as an Ingress Tunnel Router (ITR).It retrieved the EID to RLOC from the mapping system and assigns to it a LISP header, and an outer IP header with RLOC as destination IP address.The RLOC that is at the receiving end acts as the ETR (Egress Tunnel Router) which decapsulates the packet and then forwards it to the destination EID. If the traffic is not LISP compliant, the encapsulation or decapsulation process may be carried out by a Proxy ITR.The usage of RLOC priorities helps in load balancing [3].

C. TRILL-Transparent Interconnection of Lots of Links

TRILL(Transparent Interconnection of Lots of Links) runs a link state protocol between the routing bridges(Rbridge).It uses broadcast communication so that each Rbridge knows about all other Rbridges.Thus each RBridge has enough information to compute pairwise optimal path for unicast traffic. Distribution trees and an RBridge is used as a root for forwarding. Each node present in the network calculates the TRILL header and performs functions like the MAC address swapping.STP is not needed in TRILL.[2]

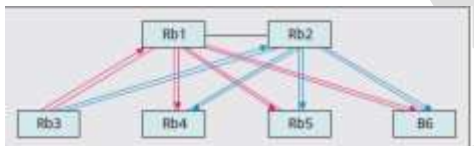

Fig 3: TRILL multicast domain with RBridges[2].

Benefits of TRILL:

- Introduces a loop free path
- Alleviates problems related to MAC address tables
- Handling unplanned topology changes.

D. SPB

Shortest Path Bridging(SPB) is a layer 2 protocol which was formerly introduced by IEEE as provider link state bridging(PLSB) which was developed for the telecommunication market.802.1 ah frame for separating MAC BOOK and VLAN ids thus enabling data centre virtualisation called ISID.The ISID abstracts service from the network.ISID provides mechanism for granular ID control as well.When a new device is attached to the SPB and wishes to communicate with an existing device, there is an exchange to identify the device and its immediate neighbours. This exchange is enabled by the ISIS protocol.The ISID is reachable to the end of the network where learning is .The bidirectional paths are then enabled from the requesting device to the destination using ECMP.The same kind of approach is used for Unicast as well as broadcast packets. Once the entire tree has been developed, the tree is pruned and traffic is assigned to a particular path. The endpoints reach each other by transmitting an address port. Thus the endpoints are well aware of the entire traffic path [2].

E. Virtual Cluster Switching

Brocade proposed virtual cluster switching.VCS is a layer 2 protocol based on the basic functionality of TRILL. It is compliant with TRILL and uses the basic TRILL framework format. The core of Brocade uses Shortest Fabric Path First (SFPF), which is the standard path selection protocol in Fibre Channel Storing area. According to Brocade, a single VCS fabric is synchronised with 32000 MAC addresses. Currently the Brocade switches can scale up to 600 physical ports in a single fabric.

F. QFabric

Juniper proposed QFabric using the ideal data centre strategy. At the heart of this strategy is the Juniper 3-2-1 Data Centre Network Architecture which eliminates the layers of switching to flatten them into two layers which is further reduced to a single layer by the Quantum Fabrics or the

QFabric. [2]

According to Juniper, QFabric allows multiple devices in the network to share a common management plane. This is just why the QFabric devices and switches are not referred to as edge core and the overall approach is that of a fabric than a network. Juniper can support upto 6144 physical ports with the oversubscription of 3:1 at the edge of the fabric. QFabric provides equal latency between any two ports in the fabric [2].

G. ODIN-Open Datacentre Interoperable Network

ODIN was released by IBM.ODIN helps in interoperability between the devices, thus facilitating multi-vendor interaction. It deals with various industry standards and the best of practices including TRILL, Lossless ethernet, Open flow, Wide Area Networking and ultra-low latency networks. It has been publicly endorsed by a number of companies, a few of the famous ones include Brocade, Juniper, Huawei, NEC, BigSwitch and Marist College. Some of these companies have developed individual protocols to support ODIN.Eg QFabric by Juniper, VCS by Brocade. Each of these protocols have different features such as scalability, latency, oversubscription and management. ODIN has not yet been implemented in the market but has high hopes associated with it. [2]

VI. CONCLUSION

In this paper, a detailed study of the various networking protocols for multipathing has been done. A comparison has been made between the legacy and present day data centres to show how the present day data centres have gone an improvisation since t6he past. The various protocols and their roles in the data centre management have been discussed.

VII. ACKNOWLEDGEMENT

I would like to thank my professors and mentor for their immense support and encouragement to guide me throughout my research in the area.

REFERENCES:

[1] Juniper white paper 2000380-003-EN (Dec. 2011), Revolutionizing Network Design: Flattening the Data Centre Network with the Qfabric Architecture,http://www.juniper.net/us/en/local/pdf/whitepapers/ 2000380-en.pdf, accessed 27-Dec-2011, see also independent Qfabric test results, available: http://newsroom. juniper.net/pressreleases/juniper-networks-qfabric- sets-new-standard-for-net-nyse-jnpr0859594, accessed 30 June 2012.

[2] Carolyn J.Sher Decusatis,Casimer M.Decusatis ,Aprico Carranza, ”Communication Within Clouds: Open Standards and Proprietary Protocols for Data Centre Networking”, Communications Magazine, IEEE, vol.50, no.9, pp26-33, September 2012.

[3] M.Coudron et al,”Boosting Cloud Communications through a Cross Layer Multipath Protocol Architecture”, Proc of IEEE, SDN4FNS’13, Italy, Nov, 2013.

[4] Lyno Henrique,G.Ferraz,Diogo Mennez Ferrazani Mattos ”A two phase Multipathing Scheme based on Genetic Algorithm for Data Centre Networking ”,IEEE-GLOBECOM 2014,December 2014.

[5] D.Bellabed,S.Secci,D.Medhi,G.Pujolle ”Impact of Ethernet Multipath Routing on Data Centre Network Consolidation”,Proc of the 4th int workshop on Data Centre Performance,ICDCS,IEEE,June 2014.

[6] J.Mudigonda,P.Yalagandula,M.Alfares,J.Mogul,”SPAIN:COTS Data centre ethernet for multipathing over arbitrary topologies ”,Proc NSDI,2010,San Jose,CA,USA,April 2010.

[7] S.Secci,San Murugesan, ”Cloud networks:Enhancing performance and resiliency”,IEEE,Computer 2014:47(10):82-85

[8] P. Mell and T. Grance, The NIST Definition of Cloud Computing, 07 Oct. 2009: http://csrc.nist.gov/groups/ SNS/cloud-computing/, accessed: 29-Jan-2011.

[9] www.Multipath TCP.org