A Survey of Cloud Network Overlay Protocols

Muhammad Aamir Nadeem , Taimur Karamat

Department of Computer Science, Virtual University of Pakistan
M.A. Jinnah Campus, Defence Road, Off Raiwind Road, Lahore, Pakistan.
ms140400023@vu.edu.pk;taimur.karamat@vu.edu.pk

Abstract — **Cloud computing is model for sharing of computing and storage resources over Internet. Virtualization is the key technique behind Cloud. It enables provisioning of various kinds of services to the Cloud users on one hand; on the other it increases utilization of computing resources and bandwidth. Virtualization of server and network puts pressure on existing hardware features and protocols by arousing issues of scalability, limitation of number of ports, under utilized resources and bandwidth. MAC address table of Network Interface Card has a limited size, VLANs are not enough scalable to cater demands of Cloud, and Spanning Tree Protocol results in idle or under utilization of resources. Cloud network overlay protocols are the intricacies devised to resolve these issues. It resolves the issues of scalability, mobility, better resources and bandwidth utilization and workload migration. This paper presents comparative overview of different overlay protocols namely Shortest Path Bridging (SPB), Transparent Interconnection of Lots of Link (TRILL), Locator/ID Separation Protocol (LISP), Virtual Extensible LANs (VXLANs), Network Virtualization using Generic Routing Encapsulation (NVGRE), and Stateless Transport Tunnelling (STT) followed by conclusions and future recommendations.**

***Keywords*— Cloud Computing; Cloud network overlay protocols; SPB; TRILL; LISP; VXLAN; NVGRE; STT;**

I. Introduction

Cloud Computing (Cloud) provides ubiquitous network access to a poll of resources which are shared and configurable. It is based on shared services and convergence of infrastructure. It emerged from evolution and adoption of many prevalent technologies. It inherits many of its characteristics from client-server model, grid computing, mainframe computing, utility computing, and peer-to-peer architecture. Virtualization is the key technique behind Cloud Computing. It adopts Service Oriented Architecture (SOA) which enables its clients to transform their requirements problems into services thus benefited by the solution provided by the Cloud. Key advantages of cloud include agility, reduced costs, device independence, location independence, easy maintenance, high performance, extremely scalable and flexible, increase productivity, privacy, and security.

Increased densities of processing and storage resources, with virtualization enabled Cloud computing to cater thousands of clients with varying nature of services. Virtualization of servers, network, and storage increased the resources utilization and processing efficiencies on one hand, on the other it aroused new complexities in processing and networking architecture. The existing network protocols are tied with physical networks and ports. Virtualization rapidly explodes the MAC address table, possible number of VLANs prove to be insufficient, and Spanning Tree Protocols (STP) causes networking resources be idle or underutilized.

It demands for more scalable, robust and efficient solutions that can resolve these issues and can be implemented in the existing infrastructure and protocol stack. Cloud computing uses overlay networks to resolve these issues.

This paper surveys the routing protocols used in Cloud network overlays. This paper is structured as follows; Section II briefly describes Cloud Computing. Section III gives a quick overview of some background concepts which provides foundation for Section IV in which overlay protocols are discussed. Section V concludes the discussion and gives direction for future work.

II. Cloud Computing

"*Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*" [1].

A general representation of Cloud architecture is shown in Figure 1 [22]. The key characteristics that clouds must have, the services it can provide and how it can be deployed are

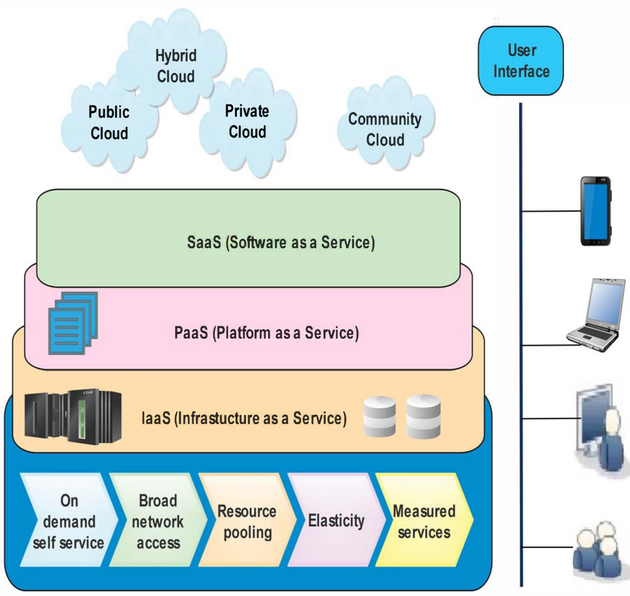

Fig. 1: Cloud Architecture

ISBN: 978-1-4673-9609-7 ©2016 IEEE

summarized in Table I [1].

III. Background Concepts

Cloud network overlay protocols are build on the foundation of some already existing algorithms and protocols. Here a brief overview of these is presented to have better insight while discussing overlay protocols.

A. *Distance Vector Routing Protocol (DVRP)*

This protocol is based on Bellman-Ford algorithm. In it every node periodically shares its routing table to its adjacent neighboring nodes. Routing table contains information of nodes that the sending node can reach and associated cost to reach them. Ultimately it determines the direction and cost to reach every node. In the even of change in topology, only neighboring nodes are informed. The protocol is used in Routing Information Protocol (RIP) and Interior Gateway Protocol (IGP). The original algorithm suffers from count-to-infinity problem as it can't prevent loops. Split horizon, poison reverse and maximum hops counter are used to remedy the problem [2].

B. *Link State Routing Protocol (LSRP)*

This protocol is based on Dijikstra algorithm. Here every node (only switching nodes means routers) periodically sends its links state to its adjacent neighboring nodes in the form of Link State Packets (LSP). LSP contains sending node identifier, identifiers of the nodes to which it is directly connected and associated cost. Having received LSPs, each node draws a graph of network topology and computes next best hop (shortest path) to reach some other node in the network. Quick convergence, rapid adoption of topology changes, and less traffic are the nice features of this protocol. Addition sequence number field in LSPs ensures reliable flooding [3]. The protocol is used in Open Shortest Path First (OSPF) protocol, Intermediate System-to-Intermediate System (IS-IS) protocol.

C. *Intermediate System-to-Intermediate System Protocol(IS-IS)*

It was defined in the Open System Interconnection (OSI) model of International Standards Organization (ISO) [4] and Internet Engineering Task Force (IETF) republished it in [5]. It is based on LSRP, and is de facto standard protocol for Internet Backbones. It is designed to be used within autonomous system (AS). It differs from the OSPF in the following aspects. It is layer2 protocol while OSPF is layer3 protocol. The other difference is notion of 'areas' is a bit different in IS-IS. Unlike the OSPF area demarcation at area border router (ABR), IS-IS routers are designated /configured as Level1 or Level2 router. Level2 routers are designated for inter-area routing; they exchange link state information with other neighboring level2 router. Level1 routers are designated for intra-area routing; they exchange link state information with other Level1 routers and also send to its designated Level2 router. Another difference is IS-IS has no Area 0 which is defined in OSPF as Backbone for inter-area routing, instead here level 2 routers form logical topology of Backbone. Being a layer2 protocol, it requires MAC addresses, and its LSPs contain area identifier, system identifier, and net-selector fields. Instead of Address Resolution Protocol (ARP),

Table I: Summary of Cloud Architecture [1]

Essential characteristics	On-demand self-service	• Provisioning of computing capabilities to the customer. • Automatic provisioning on demand. • e.g. network storage, server time etc.
	Broad network access	• Capabilities are accessible through standard networks, primarily Internet. • Devices that can access these capabilities include mobile phones, workstations, laptops etc.
	Resource pooling	• Multi-tenancy • Resources are shared by all customers. • Resources are location independent. • Resources may be physical or virtual.
	Rapid elasti city	• Elastically provisioning /releasing of resources. • Must be scalable and flexible enough to meet maximum demands.
	Measured service	• Measuring capabilities for the type of services provisioned. • Automatic control and optimization of resources usage. • Monitoring and controls. • Reports and accounting of utilized services for both customer and service provider.
Service Models	Software as a Service (SaaS)	• Customer can use only the provided application running on underlying cloud infrastructure. • Capabilities are accessible through Internet or APIs. • No control of underlying infrastructure.
	Platform as a Service (PaaS)	• Customer can deploy/configure it's or third party application, the only limitation is supported by underlying infrastructure. • No control of underlying infrastructure like operating system, servers, storage etc.
	Infrastructure as a Service (IaaS)	• Customer is provisioned computing resources of processing, network, storage. • Customer can run/deploy arbitrary software including operating system, applications. • Customer can only control components related to it. • No control of underlying infrastructure.
Deployment Models	Private cloud	• Cloud infrastructure is exclusively provisioned to single organization. • Ownership, management and operation may be of organization or third party or by both. • Commonly exist on premises of organization, may exist off premises.
	Community cloud	• Cloud services are provided to some specific community. • Community belongs to the organizations having shared concerns. • Ownership, management and operation may be of multiple organizations or third party or by both. • No restriction of on premises may exist off premises.
	Public cloud	• Services are provisioned to general public. • Ownership, management and operation may be of business/ government/ academic organization or third party or by both. • Commonly exist on premises of cloud owner.
	Hybrid cloud	• Combination of two or more above mentioned infrastructure types. • The infrastructures are only bound together by application portability and data. • Their distinctness remains preserved.

Internet Control Message Protocol (ICMP), Inter-domain Routing Protocol (IDRP); IS uses End System-to-Intermediate System (ES-IS) protocol. IS-IS is easily expandable and more scalable [6].

D. *Virtual Local Area Network (VLAN)*

Specifications of VLAN are defined in [7]. It enables creation of multiple LAN over a single physical Ethernet. The bridges/switches are configured to cater unicast and broadcast communication to each VLAN. To achieve these functionalities, VLAN tags are embedded into Ethernet frame. VLAN-aware part of network treats these frames according to fields specified in tag. VLAN tag contains 16 bit Tag Protocol Identifier (TPID), and 16 bit Tag control information (TCI) that includes 3bit Priority code point (PCP) single bit for Drop eligible indicator (DEI) and 12bit VLAN identifier (VID). Thus there are 4096 possible VLAN can be supported [8].

E. *Overlay Networks*

Overlay networks, in simple words, are logical networks on top of physical networks as depicted in Figure 2 [9]. Internet itself is an overlay build on top of telephone network. In overlays decision is application which is typically required in distributed application. Each node in overlay also exists in underlying physical network; however links between overlay nodes are implemented by tunneling.

Tunneling provides a mechanism of delivering packets from one end to the other. Overlay networks paves the way for experimenting new internet protocols, for extending the functionalities and features of Internet. Gnutella and Bit Torrent are the earlier examples of peer-to-peer networks build using overlays [9].

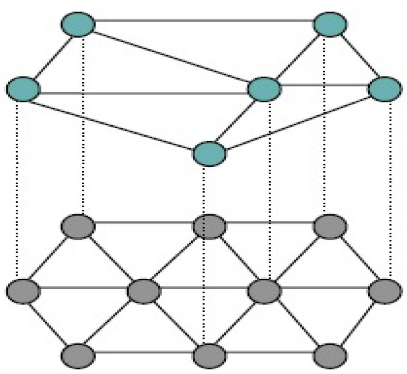

Fig. 2: Overlay network

IV. Overlay Routing Protocols

In Cloud networks overlays are used to resolves the problems mentioned in Section I. overlays are built on layer 2 or layer 3.

A. *Shortest Path Bridging (SPB)*

IEEE, to make network creation and configuration simpler with enabled multi-path routing, specified SPB in [10]. It is intended to replace Spanning Tree Protocols (STP) that blocks some paths to avoid loops. It results in some of routers are not used in forwarding packets thus resources are idle and bandwidth remains under-utilized. SPB keeps all paths active but with equal cost path, thus forming large number of layer2 topologies that enable to share load among different paths and quick adopt changes in topology. The result is better utilization of hardware resources and bandwidth. It is designed with plug-and-play characteristic, thus enable easier configuration of network and avoid human errors. SPB uses IS-IS in computing shortest paths. It can support thousands of switches having multi-terabyte capacity with just two tiers in non-blocking manner. It can be used in data centers networks by interconnecting few dozen switches to form a big Layer2 switch thus achieving non-blocking and efficient processing as shown in Figure 3 [23]. SPB frame format is depicted in Figure 4 [10] [24].

SPB combined with Virtual LAN (VLAN), termed as SPBV provides backward compatibility by supporting all STP based protocols. SPB with MAC termed as SPBM extends the functionalities. It supports VLAN, IPv4, and IPv6. In Winter Olympics 2014, a network fabric with SPB switches handled 54 Terabits per second of traffic [11].

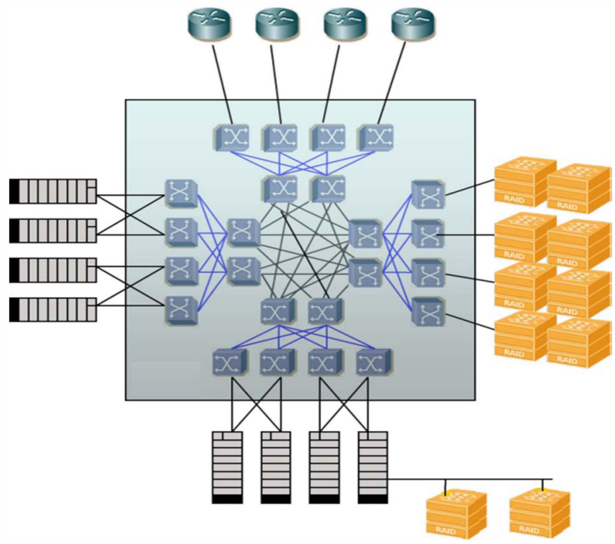

Fig. 3: Data centre network with SPB

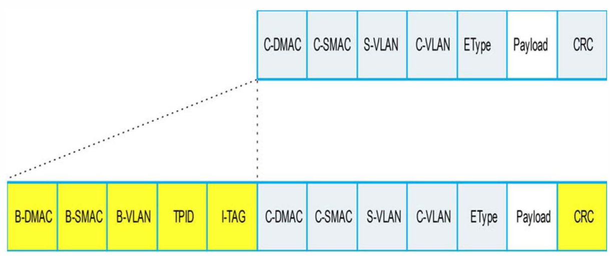

Fig. 4: SPB frame format

B. Transparent Interconnection of Lots of Links (TRILL)

TRILL is also an attempt to replace STPs from IETF [12, 13, 14, 15]. To eliminate the blocking characteristic of STP, TRILL provides a way of building a way of fully connected network of RBridges by configuration. The outcome is a mesh topology where all paths are active however it selects the optimum path as shown in Figure 5 [25]. RBridges are switches on which TRILL is implemented. They use IS-IS to acquire global routing information and maintain their link state database. RBridges are strictly implemented at Layer2, so TRILL is Layer2 protocol; hence it is independent of IP addressing. RBridges. Packet frames before transmitting are encapsulated in TRILL header at sending RBridge and encapsulated at destination RBridge. Address field in TRILL header is 16 bit wide enabling it to support around 64,000 RBridges. TRILL frame format is depicted in Figure 6 [15] [24]. It may employ routing techniques of Layer3; also it supports VLAN, IPv4, and IPv6. TRILL is most suitable for data center's network.

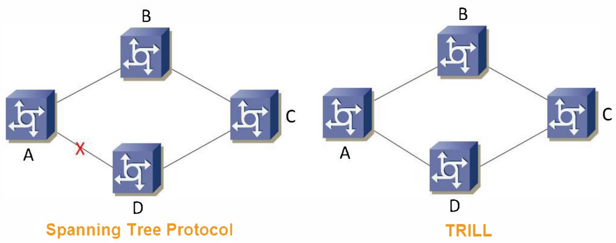

Fig. 5: Comparison of paths between Spanning Tree Protocols and TRILL

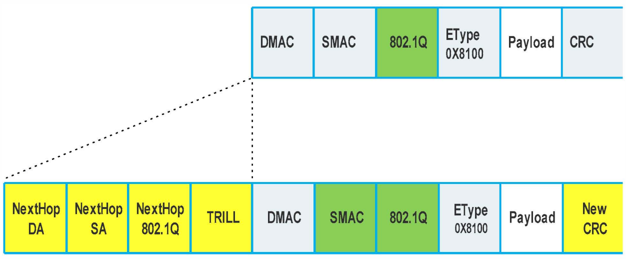

Fig. 6: TRILL frame format

C. Locator/ID Separation Protocol (LISP)

LISP separates the identity of host from its location. It defines a new architecture. It splits IP address into two fields; first End Point Identifier (EID), second Routing Locator (RLOC). IP addresses of IPv4/IPv6 can be used as EID. So it does not require any changes in existing 'core' of Internet architecture and protocol stack. However RLOC needs some mapping to be established. LISP provides support for multi-homing, mobility independent of location, scalability and traffic engineering [16, 17]. LISP frame format is depicted in Figure 7 [16] [24].

To implement LISP, it requires addition of following new components in the architecture. 1) Ingress Tunnel router (ITR) carry out mapping of source EID with RLOC of it, and mapping of destination EID with RLOC of Egress Tunnel Router to do so it sends map requests to map resolver. Then it encapsulates these source and destination EIDs and RLOCs in the packet header and forwards it according to route. 2) Egress Tunnel Router (ETR) periodically sends map-register request to map server. After receiving a map request, it responds back to ITR, its other functionality includes decapsulation of LISP packet. 3) Map server accepts registration requests, aggregate them using EID prefixes, also implement mapping database. 4) Map Resolver also connected with Alternative Logical Topology through partial mesh. It decapsulates ITR map requests, and then forwards them to ETR. 5) ALT router accepts EID prefixes and advertises them. 6) Proxy ITR: implements look up ITR mapping-database for non-LISP sites. 7) Proxy ETR: implements ETR function for non-LISP sites.

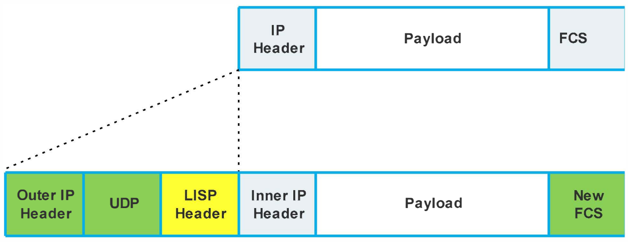

Fig. 7: LISP frame format

D. Virtual Extensible LAN (VXLAN)

VXLAN is one among the network virtualization technologies. It aims to resolve the issues of scalability. It enables encapsulation of layer2 frames which are MAC based into UDP datagram of layer4 and using port 4789 (IANA assigned port for UDP VXLAN). By this technique layer2 network can be extended over IP network. Server virtualization enabled several VM instances, each requiring different MAC address resulted huge number of MAC address tables in switched Ethernet to cater communication across hundreds of VMs. Implementation of VMs in the data center environment and their classification into groups require thousands of VLANs. It resolves the problems that may arise while serving multiple tenants and to cater MAC traffic over logical IP tunnels [18]. VXLAN frame format is depicted in Figure 8 [18] [24].

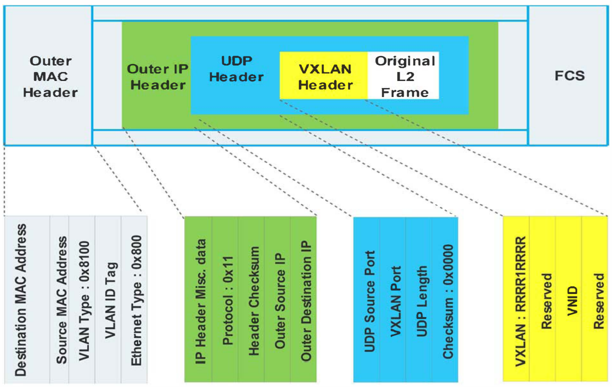

Fig. 8: VXLAN frame format

E. *Network Virtualization Using Generic Routing Encapsulation (NVGRE)*

GRE encapsulates any layer3 protocol to any other layer3 protocol. GRE packet contains delivery header which is any layer3 routing protocol, GRE header, and payload which is any Layer3 protocol packet [19, 20].

Network Virtualization technique creates virtual layer2 and layer3 topologies over any physical layer2 and layer3 network. Connectivity among the created topologies is provided using tunneling of Ethernet frames over physical IP. This enables a virtual layer2 network to span over multiple physical subnets. A 24bit Tenant Network Identifier is associated with every layer2 virtual network, enabling it create 16 million subnets. In Data Center environment, it enables assigning of workload of virtual layer2 to any server, and migration of workloads to any server. Thus increasing the resources utilization and processing efficiencies. NVGRE frame format is depicted in Figure 9 [20] [24].

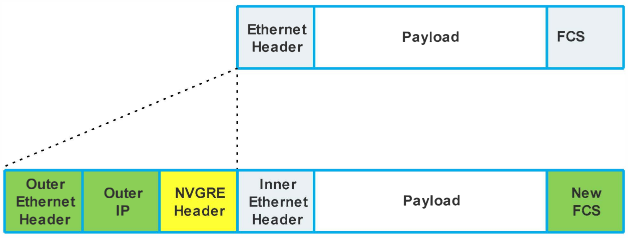

Fig. 9: NVGRE frame format

F. *Stateless Transport Tunneling (STT)*

STT builds overlay networks on top of virtual networks. Having the capabilities of exploiting network interface card (NIC) feature for performance gains, it is more suitable for the configurations where tunnel endpoint is in end-system. STT features include management of overlapping addresses, decoupling of virtual topology from the physical topology, VM mobility, extremely scalable for provisioning of virtual networks, and isolation between physical network and virtual network. STT header is TCP-like header, which is encapsulated in IP header. However unlike to TCP it is stateless. TCP-like header enables it to exploit NIC features. Primarily STT carries Ethernet frames from one end of tunnel to other end. For appropriate delivery of frames at the endpoint, it uses 64bit context ID to identify the entity at destination side [21].

V. CONCLUSION AND FUTURE WORK

SPB and TRILL uses multipath routing, they are more suitable for traffic engineering, load balancing. There band width utilization is extremely efficient. LISP provides more flexibility in terms of scalability, multi-homing, and traffic engineering. It makes device location independent thus decouples its address from physical location. It is more suitable where application, server, VM mobility are required which result in even better utilization of resources. VXLAN and NVGRE both resolves the issue of scalability, they can support up to 16 million VLANs thus suitable for multi-tenancy. NVGRE is useful in the environments where workload migration is required. STT is suitable where address overlapping is needed; it also exploits the features of NIC.

Aside from the benefits that these protocols offer, they involve overhead of encapsulation/tunneling which results in decrease of actual data transfer per packet. However this cost is acceptable against the returns. Suitability of these protocols for larger packet sizes which reduces per packet tax needs further study.

Leading vendors who provide networking devices that support these protocols include AVAYA, HP, CISCO, Alcatel-Lucent, Huawei etc. In the next step of this study, for detailed comparison of these protocols for various performance metrics, Proprietary softwares/toolkits of these vendors will be used to carry out simulations and to get statistical results.

REFERENCES

[1] P. M. Mell and T. Grance, The NIST Definition of Cloud Computing, National Institute of Standards and Technology: Special Publication (NIST SP) - 800-145, 2011.

[2] A. S. Tanenbaum and D. J. Wetherall, COMPUTER NETWORKS, 5 ed., Boston, Massachusetts: PRENTICE HALL, 2011.

[3] J. Kurose and K. Ross, Computer Networking A Top Down Approach, 5 ed., Boston: Addison Wesley, 2010.

[4] ISO, Information technology-Telecommunications and information exchange between systems -- Intermediate System to Intermediate System intra-domain routeing information exchange protocol, ISO/IEC 10589:2002, 2002.

[5] D. Oran, Ed., OSI IS-IS IntradomainRouting Protocol, RFC 1142: IETF, 1990.

[6] P. Göransson and C. Black, Software Defined Networks:A Comprehensive Approach, Waltham: Elsevier, 2014.

[7] IEEE, Virtual LANs, IEEE 802.1Q, 2005.

[8] B. A. Forouzan, Data Communications and Network, 5 ed., McGraw-Hill, 2013.

[9] L. L. Peterson and B. S. Davie, ComputerNetworks: a systems approach, 5 ed., Elsevier, 2012.

[10] D. Fedyk and M. Seaman, Eds., Shortest Path Bridging, IEEE

802.1aq, 2012.

[11] "The world's largest guest network," [Online]. Available: http://www.avaya.com/usa/documents/sochi-2014-olympic-winter-games.pdf. [Accessed 06 02 2016].

[12] D. Eastlake, R. Perlman, A. Ghanwani, H. Yang and V. Manral, Transparent Interconnection of Lots of Links (TRILL): Adjacency, IETF-RFC 7177, 2014.

[13] J. Touch and R. Perlman, Transparent Interconnection of Lots of Links (TRILL):Problem and Applicability Statement, IETF-RFC 5556, 2009.

[14] R. Perlman, D. Eastlake, D. Dutt, S. Gai and A. Ghanwani, Routing Bridges (RBridges): Base Protocol Specification, IETF-RFC 6325, 2011.

[15] D. Eastlake, A. Banerjee, R. Perlman and A. Ghanwani, Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS, IETF-RFC 6326, 2011.

[16] D. Farinacci, V. Fuller, D. Meyer and D. Lewis, The Locator/ID Separation Protocol (LISP), IETF-RFC 6830, 2013.

[17] L. Jakab, A. Cabellos-Aparicio, F. Coras, J. Domingo-Pascual and D. Lewis, Locator/Identifier Separation Protocol (LISP):Network Element Deployment Considerations, IETF-RFC 7215, 2014.

[18] M. Mahalingam, D. Dutt, K. Duda, P. Agarwal, L. Kreeger, T. Sridhar, M. Bursell and C. Wright, Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks, IETF-RFC 7348, 2014.

[19] P. Garg and Y. Wang, Eds., NVGRE: Network Virtualization Using Generic Routing Encapsulation, IETF-RFC 7637, 2015.

[20] D. Farinacci, T. Li, S. Hanks, D. Meyer and P. Traina, Generic Routing Encapsulation (GRE), IETF-RFC 2784, 2000.

[21] B. Davie and J. Gross, A Stateless Transport Tunneling Protocol for Network Virtualization (STT), IETF-Internet Draft, 2012.

[22] S. D. Choubey and M. K. Namdeo, "Study of Data Security and Privacy Preserving Solutions in Cloud Computing," in *International Conference on Green Computing and Internet of Things*, Delhi, India, 2015, pp. 1101-1106

[23] P. Ashwood-Smith, "Shortest Path Bridging IEEE 802.1aq Overview," in *Asia Pacific Regional Internet Conference on Operational Technologies*, Hong Kong, 2011. [Online]. Available: https://www.apricot.net/apricot2011/program/layer2bridges/

[24] "Data Center Overlay Technologies (White Paper)," CISCO, [Online]. Available: http://www.cisco.com/c/en/us/products/collateral/switches/nexus-9000-series-switches/white-paper-c11-730116.pdf. [Accessed 10 February 2016].

[25] "Transparent Interconnection of Lots of Links (TRILL)," [Online]. Available: http://www.ipinfusion.com/products/zebos/protocols/data-center-ethernet/TRILL. [Accessed 10 February 2016].