2024 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) | 979-8-3503-6498-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/WECONF61770.2024.10564620

A Way to Implement Network Services in a Data Center

I. G. Buzhin
Moscow Technical University of Communications and Informatics
Moscow, Russia
i.g.buzhin@mtuci.ru

V. M. Antonova
Moscow Technical University of Communications and Informatics
Moscow, Russia
v.m.antonova@mtuci.ru

V. S. Gnezdilov
Moscow Technical University of Communications and Informatics
Moscow, Russia
v.s.gnezdilov@mtuci.ru

Yu. B. Mironov
Moscow Technical University of Communications and Informatics
Moscow, Russia
i.b.mironov@mtuci.ru

E. A. Gayfutdinov
Moscow Technical University of Communications and Informatics
Moscow, Russia
e.a.gaifutdinov@mtuci.ru

***Abstract*— In recent years, IP-fabric has often been used in data center networks, which is a set of network equipment (usually switches) interacting via control protocols, which provides unified addressing, security, etc. services, as well as data delivery along the optimal route. With such an organization, the implementation of some services is transferred to the virtualization of network functions. This article suggests an effective way to implement virtualization of network functions and deliver traffic to it in the IP fabric of a data center. The method is based on VPP (Vector Packet Processing platform) and FRR (Free Range Routing), VxLAN (Virtual extensible local area network), EVPN (Ethernet VPN) and ICMP (Internet Control Message Protocol) technologies. The results of test measurements of the processing of various traffic on the data plane are shown. The results of experimental tests show that on Xeon Cascade Lake core, VPP technology produces a performance of about 20 Mpps with IPv4 traffic. Combinations of IPv4 lookup, VxLAN, Encap/Decap, and ACL (Access Control List) lookup processing show lower performance.**

Keywords*— *IP fabric; FRR; Network Function Virtualization (NFV); VPP; VxLAN, EVPN.

I. INTRODUCTION

Modern data centers use IP fabrics, which are a consolidated system consisting of highly interconnected network devices within the data center and deployed over two or three layers. This approach simplifies the network architecture by minimizing the distance between end nodes, reducing latency, and improving device efficiency. Many of the procedures required for network interaction are moved to the overlay network, including segmentation, virtualization, and other interconnected services. These procedures are shifted to the overlay network, which runs on top of the IP fabric. Standard dynamic routing protocols are supported in this case.

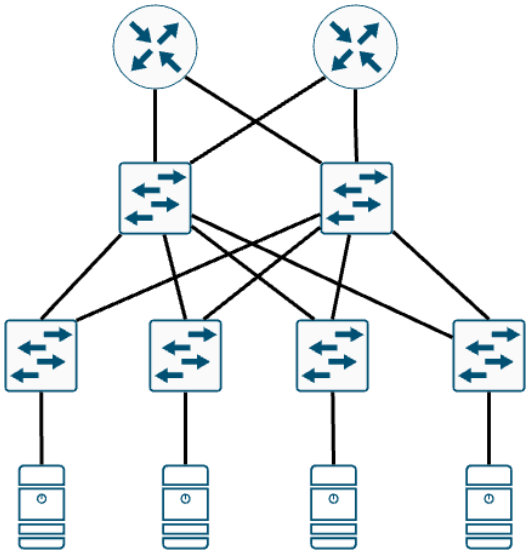

Fig. 1. IP fabric architecture

The IP fabric is a network architecture that provides a high level of connectivity, high redundancy, guaranteed bandwidth, and QoS (Quality of Service) for peer-to-peer networks. The IP fabric provides transport services for various overlay IP networks. This topology relies on a fat-tree and spine-leaf structure. The latter is also known as a Clos network and represents a multicast switching network. In many modern data centers, the Clos architecture is implemented in the form of a

979-8-3503-6498-9/24/$31.00 ©2024 IEEE

«leaf-spine» scheme, where the «spine» layer represents switches at the backbone level, and the «leaf» layer represents edge switches at the input and output. This scheme is also sometimes called the fat tree topology. The purpose of this topology is to reduce the total number of switches and ports required. The architecture of the IP fabric (Fig. 1) includes many high-speed direct links that prevent bottleneck-induced slowdowns in the network and ensure high forwarding efficiency and low latency. Using such an architecture in a data center provides a number of advantages compared to the classic two-tier architecture with a "collapsed" network core:

- fault tolerance increases. There can only be two independent cores in a collapsed core network - this is a limitation of MLAG (Multi-Chassis Link Aggregation) aggregation technologies, which allow interfaces to be grouped into groups between a maximum of two switches. There can be many more Spine switches. It makes it easier to increase bandwidth by simply adding the required number of Spine switches.
- failures affect fewer devices – the switches are completely independent, fault tolerance is achieved without the use of stacking or MLAG.
- the delay becomes predictable – there is only one Spine switch in the path of any internal traffic flow.
- the level of utilization of network interfaces increases – traffic between Leaf and Spine is transparently and efficiently distributed using ECMP (Equal Cost Multipath) mechanisms.

However, the implementation of the IP fabric architecture utilizes network equipment of data processing centers. This equipment may not allow for additional traffic processing services, such as NAT (address translation) functions, special L3/L4 traffic filtering, intrusion detection and prevention systems (IPS/IDS), or asymmetric traffic return from the service (Direct Server Return), when necessary. Simultaneously, this implementation of the IP fabric does not allow for any modifications to the network stack.

Therefore, providing additional traffic processing services on the IP fabric requires solving several problems. These include selecting the appropriate method of traffic processing, determining how to deliver traffic to the processing point, meeting network architecture requirements (with no possibility to modify the network stack), and ensuring scalability and redundancy.

II. Methods for Handling Traffic on a Public Server

The Network Function Virtualization (NFV) approach is used to implement network traffic processing on public servers.

In accordance with the NFV reference architecture [8], a system consisting of hardware and software for the implementation of NFV can be represented from the following elements (Fig. 2):

- Virtual Network Functions (VNF(s));
- Network Function Virtualization Infrastructure (NFVI);
- NFV Management and Orchestration Software.

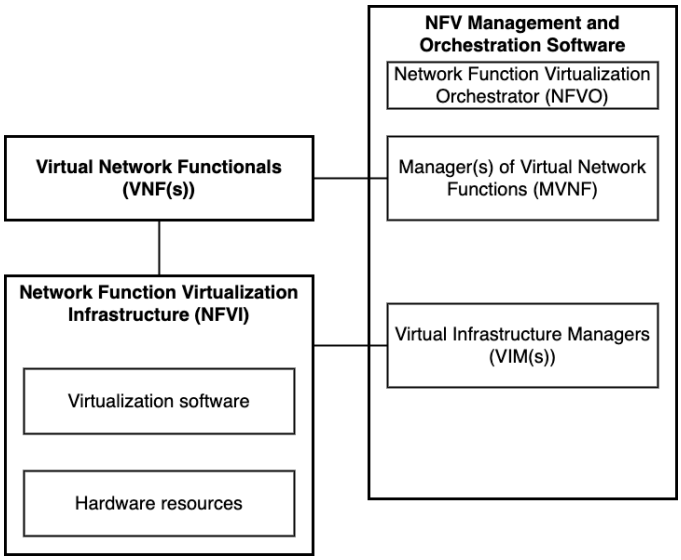

Fig. 2. Simplified Network Function Virtualization Reference Architecture

Virtual network functions implement the functionality of communication hardware (switches, routers, etc.) Examples of such functionality can be routing protocols, network address translation, implementation of access lists and firewall, implementation of QoS mechanisms, etc. [3]. The NFV infrastructure is an abstraction of the hardware of communication networks, in fact, it is a set of software tools for implementing virtualization of hardware resources, including memory and computing resources. The NFVO orchestrator is used for the proper and coordinated functioning of the NFV infrastructure and functions. Its main functions are connecting and creating instances of VNF and network services, and changing infrastructure resources. The orchestrator constantly interacts with the VNF manager, who monitors and ensures the lifecycle of all network functions. The VIM manager manages and controls the infrastructure. Its key functions are the identification of infrastructure facilities (hardware and software), the redistribution of resources between virtual entities (virtual machines or containers), monitoring the parameters of the functioning of virtual entities and their interactions.

Thus, the NFV concept allows replacing the functions of BNG/BRAS (Broadband Remote Access Server) devices – gateway servers for connecting subscribers to the global network. The network functions of BRAS are the protocols for establishing PPPoE (Point-to-point protocol over Ethernet) sessions, DHCP/IPoE, the implementation of the RADIUS server for managing subscriber sessions, the automatic creation of VLANs for each individual service (S-VLAN), the implementation of the GRE protocol for the HTTP Redirect server, the implementation of QoS at the VLAN level. In the classic hardware version, BRAS have a high cost and proprietary implementations of the above functions, which are tied to the hardware. NFV allows you to implement such functionality on a group of public servers with the ability to scale the functionality.

Thus, the NFV concept makes it possible to increase network performance, optimize the resource utilization of hardware devices, reduce the cost of purchasing and maintaining specific hardware and hardware-software telecommunications devices, reduce network power consumption, and unify the implementation of network functions on hardware.

In order to implement this approach, it is necessary to select an implementation for the data plane and for the control plane. The network function server is shown in Fig. 3.

To achieve the necessary level of performance for the data plane, a DPDK (Data Plane Development Kit) [10] based implementation of VPP [9] was chosen. VPP [9] is a cutting-edge platform for high-speed packet processing. It operates within the user space and utilizes kernel bypass methods to access the hardware. Its design is independent of the hardware, kernel and deployment. Unlike older packet-by-packet processing methods, VPP can handle a vector of packets at once. Thus, the processing is done 'vector by vector'. VPP utilizes DPDK services [10] to bypass the kernel and access the hardware. DPDK implements network I/O, while VPP handles packet processing. VPP uses DPDK to retrieve a vector of packets from the network card, and then it handles all the packets in the vector. A vector consists of a group of packets that are processed simultaneously, with the current vector size being 256.

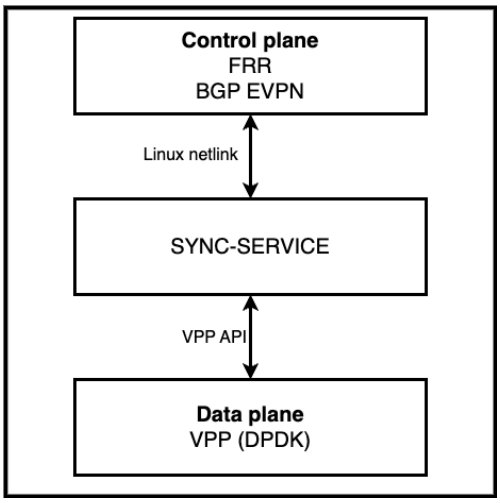

Fig. 3. Network function server

When the packet buffer (one vector) becomes available, VPP starts processing all the packets within the vector. VPP processing is executed as a forwarding graph. Each node represents a network function. There are 4 types of nodes: 1) Pre-input 2) Input 3) Internal 4) Process. Pre-input nodes are functions that handle packet sampling from the hardware and make the index buffer available to VPP. An input node can be in one of two states: 1-interrupting or 2-polling. The interrupting state of input nodes occurs when the number of packets available in the buffer is less than 5. Conversely, the node switches to the polling state when more than 10 packets are available. In the polling state, the node continuously checks for available packets to fill the vector and to process it. Process nodes are individual nodes that are not connected to the graph and are executed separately as flows. Each node predicts the following node after processing the vector and sends packets to different following nodes based on its decision. This process is known as branching prediction.

Test results (Fig. 4) demonstrate that VPP operating on a single Xeon Cascade Lake core achieves a performance of approximately 20 Mpps for IPv4 traffic. The processing of traffic involving IPv4-lookup, VxLAN, Encap/Decap, and ACL-lookup combinations demonstrates lower performance. Moreover, the test results indicate that the traffic processing performance increases non-linearly with the addition of CPU cores. When processing IPv4-VxLAN-ACL traffic, a server with an Intel Xeon 6238 2.2GHz processor, an Intel x710 network card and 512 GB of RAM achieved a performance of 20 Mpps on 16 cores.

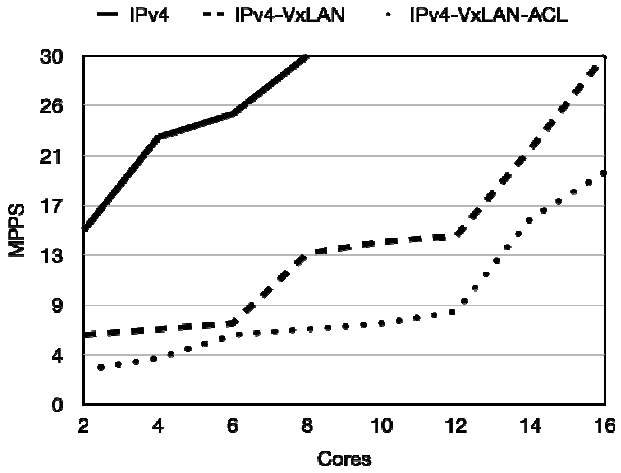

Fig. 4. VPP capacity

FRR [11], an open-source implementation of BGP [20], was selected as the control plane. FRR provides IP routing services. In the network stack, it is responsible for exchanging routing information with other routers, making routing and policy decisions as well as informing other layers about these decisions. In the most common scenario, FRR installs routing decisions into the operating system kernel allowing the kernel's network stack to make appropriate forwarding decisions. In addition to dynamic routing, FRR supports a full range of L3 configurations, including static routes, addresses, router advertisements, etc. It has some lightweight L2 features, but their presence is dependent on the specific platform. It also supports EVPN VxLAN and can be easily integrated with the selected data plane. To implement FR, the FRRouting software package is used [11], which allows you to implement, in addition to BGP (Border Gateway Protocol), such protocols as OSPF, IS-IS, EIGRP, VRRP, etc. Cisco-like commands are used for configuration. FRRouting can be installed and used on virtual entities with NIX systems.

III. Traffic Delivery to the IP Fabric Network Functions

In [1], the problem of the growing demand for virtualization of resources in data centers is analyzed. Hardware virtualization provides advantages such as higher hardware utilization (due to the implementation of several

virtual entities on the hardware), reduced user waiting time, and efficient power consumption. However, the data center data network must also support the increased demand for virtualization. According to [1], this growing demand can be overcome through the use of NFV, using traffic isolation, address space isolation (using the same address space in isolated virtual networks).

VxLAN [12] is used to deliver traffic to the IP fabric network functions. This technology is designed to address VLAN scalability issues. With its 24-bit segment identifier, VxLAN can support up to 16 million VLANs. VxLAN encapsulates Ethernet frames into UDP packets allowing the creation of virtualized L2 subnets of L3 layer. This establishes a VxLAN tunnel directly from the IP fabric switch to the Network Functions Server. In this scenario, there is no need to redirect traffic to On-Demand network functions (i.e. to establish connections only when a certain type of traffic is active), the services that require specific traffic handling are predefined. A group of services is defined by a single VRF with access to the global table via ACL. Thus, VxLAN allows you to do L3 isolation. Also, for L2 security, the Switched Virtual Interface (SVI) is used [13] to hide the MAC address. SVI works as the default gateway. Also, for security, VxLAN can use a separate VRF table to isolate the address space, which also allows the same IP addresses to be used by different clients in different address spaces.

EVPN VXLAN Type5 IP-PREFIX-ROUTE (RFC9136 [14]), IP-VRF-to-IP-VRF model, without ESI/GW Overlay Index, was proposed for the overlay construction.

[2] describes EVPN technology for the NFV management layer when used in multi-tenant data centers. One of the functions of the NFV management layer is to manage traffic flows between NFV endpoints. To implement traffic data flows, a data plane is needed that implements traffic forwarding (routing) in the network based on the management layer. [2] considers the possibility of a EVPN to isolate network traffic for each client, expand connectivity on L2 through MAC mobility and the ability to scale the management layer.

EVPN is an extension of the MultiProtocol-BGP (MP-BGP) community [15, 16], which is a new framework for calculating the reachability of the BGP Network Layer (NLRI). NLRI EVPN uses an MP-BGP extension called Address Family Identifier (AFI) and Secondary Address Family Identifier (SAFI). For EVPN the AFI is 25 (i.e. L2 VPN) and the SAFI is 70 (i.e. VPN) [4]. [5] describes the types of routes for EVPNs, which are presented in Table I.

TABLE I. EVPN ROUTE TYPES

Type	Description
RT-1	Ethernet Auto-discovery Route
RT-2	MAC/IP Advertisement Route
RT-3	Inclusive Multicast Ethernet Tag Route
RT-4	Ethernet Segment Route
RT-5	IP Prefix Route

ECMP (Equal-Cost Multi-Path routing) is used to balance traffic between network function servers (solving the problem of scaling and redundancy) [6]. ECMP determines the cost (weight) of paths to the delivery point, allows for the same paths by weight, and can load-balance traffic between paths with the same weight. The use of ECMP for load balancing between routes of the same cost increases the efficiency of network resources, as well as network fault tolerance. ECMP is recommended to be used in conjunction with a EVPN [17].

Since there is a large amount of north-south traffic in data centers, it can be serviced (transferred to its destination) based on a comparison of the client's VRF table and the VRF routing table on border leaf switches, which are connected directly to the border routers. Border leaf switches put route information in a global routing table for interaction with the outside world. Placing route information from VRF clients in the global routing table is possible by using VRF on border leaf switches [7].

EVPN also allows multitasking of Ethernet (ES) segments (a device or network of devices connected to one or more communication channels), unlike VxLAN [18]. A fully active multi-link connection allows you to distribute the traffic load across streams for efficient channel utilization. However, this mode of operation has a drawback – cycles may occur. To avoid such situations, EVPN uses the horizon splitting approach, which does not allow the return of a packet that came over a multicast connection. The MPLS label uses EVPN for filtering based on horizon splitting [19]. When the device receives a multicast frame and tries to forward it, it checks the label, and if the outgoing and incoming interfaces match, the frame is not forwarded. In addition, EVPN allows you to select a forwarder so that multicast traffic can be sent to only one channel. This prevents the transmission of duplicate multicast traffic and the formation of avalanche traffic.

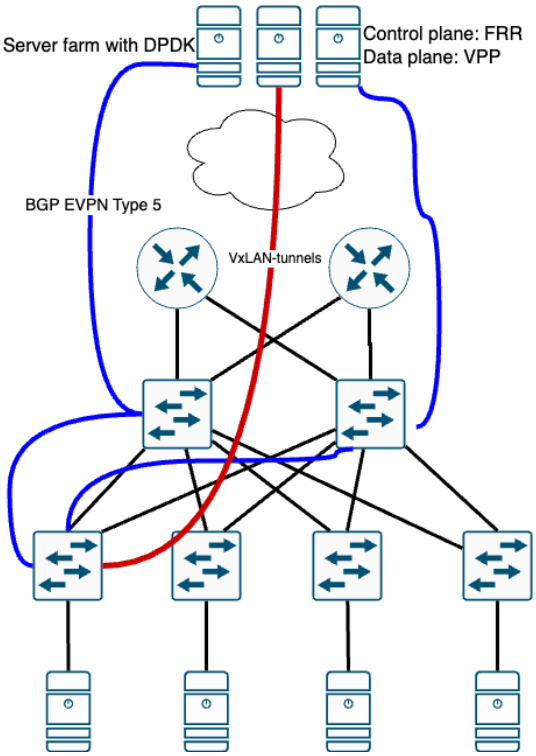

Fig. 4. Architecture of network functions implementation in the IP fabric

The deployment of EVPN-VxLAN [21] provides advantages such as the availability of an open standards-based architecture, efficient L2/L3 connectivity based on the control

plane, segmentation of the internal network and address space, MAC address mobility allows you to effectively scale and deploy the network.

Thus, Fig. 5 represents the implementation of network functions in a data centre IP fabric.

IV. Conclusion

This article searches for an effective way to implement virtualization of network functions in the IP factory of a data center. The method is based on VPP and FRR technologies. VPP uses DPDK services to bypass the kernel and access the hardware. Experimental measurements of the processing of various network traffic using the proposed NFV implementation method have been carried out. The results of experimental tests show that on 1 core of Xeon Cascade Lake, VPP technology produces a performance of about 20 Mpps with IPv4 traffic. Combinations of IPv4 lookup, VxLAN, Encap/Decap, and ACL lookup processing show lower performance. Thus, the validity of this approach has been proved experimentally. An architecture for delivering various packet traffic to the network functions of an IP factory based on VxLAN is also proposed. To build an overlay, it is proposed to use EVPN VXLAN Type5 IP-PREFIX-ROUTE, the IP-VRF-to-IP-VRF model, without the ESI/GW Overlay Index. The proposed architecture allows efficient processing of north-south traffic in the data center, isolating address spaces and load balancing along routes with the same weight.

References

[1] "IEEE Draft Standard for Local and Metropolitan Area Networks - Bridges and Bridged Networks Amendment: Virtual Station Interface (VSI) Discovery and Configuration Protocol (VDP) Extension to Support Network Virtualization Overlays Over Layer 3 (NVO3)," IEEE P802.1Qcy/D2.5, July 2018 (Draft Amendment to IEEE Std 802.1Q-2018) , vol., no., pp.1-35, 30 July 2018.

[2] G. Salazar-Chacon and L. Marrone, "Open Networking for Modern Data Centers Infrastructures: VXLAN Proof-of-Concept Emulation using LNV and EVPN under Cumulus Linux," *2022 IEEE Sixth Ecuador Technical Chapters Meeting (ETCM)*, Quito, Ecuador, 2022, pp. 1-6, doi: 10.1109/ETCM56276.2022.9935681.

[3] E. F. Naranjo and G. D. Salazar Ch, "Underlay and overlay networks: The approach to solve addressing and segmentation problems in the new networking era: VXLAN encapsulation with Cisco and open source networks," *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, Salinas, Ecuador, 2017, pp. 1-6, doi: 10.1109/ETCM.2017.8247505.

[4] S. T. Radhakrishnan and S. R. Mohanty, "Egress Engineering over BGP Label Unicast in MPLS-based Networks," *2021 IEEE International Conference on Networking, Architecture and Storage (NAS)*, Riverside, CA, USA, 2021, pp. 1-4, doi: 10.1109/NAS51552.2021.9605412.

[5] T. Singh, V. Jain and G. S. Babu, "VXLAN and EVPN for data center network transformation," *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2017, pp. 1-6, doi: 10.1109/ICCCNT.2017.8203947.

[6] M. Batayneh, D. Schupke, M. Hoffmann, A. Kirstaedter and B. Mukherjee, "Reliable Multi-Bit-Rate VPN Provisioning for Multipoint Carrier-Grade Ethernet Services Over Mixed-Line-Rate WDM Optical Networks," *Journal of Optical Communications and Networking*, vol. 3, no. 1, pp. 66-76, January 2011, doi: 10.1364/JOCN.3.000066.

[7] S. Georgiev and K. Nikolova, "Implementation of an Agile SDLC CI/CD pipeline for managing a SDN VXLAN-EVPN fabric," *2023 31st National Conference with International Participation (TELECOM)*, Sofia, Bulgaria, 2023, pp. 1-4, doi: 10.1109/TELECOM59629.2023.10409668.

[8] G. Portaluri, D. Adami, S. Giordano and M. Pagano, "A novel allocation strategy for virtual machines in software defined data center," *2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, Berlin, Germany, 2017, pp. 204-209, doi: 10.1109/NFV-SDN.2017.8169873.

[9] M. Zhu, W. Gong, F. Peng and H. Qin, "OpenStack Oriented Networking-VPP Network Optimization Method," *2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, Shenyang, China, 2021, pp. 187-191, doi: 10.1109/ACCTCS52002.2021.00045.

[10] M. -A. Kourtis et al., "Enhancing VNF performance by exploiting SR-IOV and DPDK packet processing acceleration," *2015 IEEE Conference on Network Function Virtualization and Software Defined Network (NFV-SDN)*, San Francisco, CA, USA, 2015, pp. 74-78, doi: 10.1109/NFV-SDN.2015.7387409.

[11] A. Putina et al., "Unsupervised real-time detection of BGP anomalies leveraging high-rate and fine-grained telemetry data," *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Honolulu, HI, USA, 2018, pp. 1-2, doi: 10.1109/INFCOMW.2018.8406838.

[12] M. Elmadani and S. O. Sati, "MTU Analyzing for Data Centers Interconnected Using VxLAN," *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, Manama, Bahrain, 2024, pp. 1825-1829, doi: 10.1109/ICETSIS61505.2024.10459403.

[13] J. Xu, G. Tong, Q. Chen and M. Wu, "A New Evaluation Method of The Fault Recovery Scheme for Medium-low Voltage DC Distribution Network," *2020 5th Asia Conference on Power and Electrical Engineering (ACPEE)*, Chengdu, China, 2020, pp. 1730-1735, doi: 10.1109/ACPEE48638.2020.9136433.

[14] K. Razazian and M. C. Bali, "Evaluating Various Machine Learning Techniques in Selecting Different Modulations in G3-PLC Protocol," *2023 IEEE International Symposium on Power Line Communications and its Applications (ISPLC)*, Manchester, United Kingdom, 2023, pp. 55-60, doi: 10.1109/ISPLC57122.2023.10104174.

[15] J. Mai and J. Du, "BGP performance analysis for large scale VPN," *2013 IEEE Third International Conference on Information Science and Technology (ICIST)*, Yangzhou, China, 2013, pp. 722-725, doi: 10.1109/ICIST.2013.6747647.

[16] L. Hiryanto, S. Soh, K. -W. Chin, D. -S. Pham and M. M. Lazarescu, "Multi-Path Routing in Green Multi-Stage Upgrade for Bundled-Links SDN/OSPF-ECMP Networks," *IEEE Access*, vol. 9, pp. 99073-99091, 2021, doi: 10.1109/ACCESS.2021.3093899.

[17] C. Trăistaru, "VXLAN - A practical approach to cloud computing scalability," *2023 22nd RoEduNet Conference: Networking in Education and Research (RoEduNet)*, Craiova, Romania, 2023, pp. 1-4, doi: 10.1109/RoEduNet60162.2023.10274934.

[18] G. Salazar-Chacon and L. Marrone, "Open Networking for Modern Data Centers Infrastructures: VXLAN Proof-of-Concept Emulation using LNV and EVPN under Cumulus Linux," *2022 IEEE Sixth Ecuador Technical Chapters Meeting (ETCM)*, Quito, Ecuador, 2022, pp. 1-6, doi: 10.1109/ETCM56276.2022.9935681.

[19] S. Santhanamahalingam, S. Alagarsamy and K. Subramanian, "A study of cloud-based VPN establishment using network function virtualization technique," *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2022, pp. 627-631, doi: 10.1109/ICOSEC54921.2022.9951894.

[20] R. B. da Silva and E. Souza Mota, "A Survey on Approaches to Reduce BGP Interdomain Routing Convergence Delay on the Internet," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2949-2984, Fourthquarter 2017, doi: 10.1109/COMST.2017.2722380.

[21] O. Komolafe, "IP multicast in virtualized data centers: Challenges and opportunities," *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, Lisbon, Portugal, 2017, pp. 407-413, doi: 10.23919/INM.2017.7987305.