A Multicast Scheduling Method Based on EVPN-VXLAN Extension in Data Center Networks

Nan Wu
China Telecom Corporation Limited Research Institute, Beijing, 102209, China
wun3@chinatelecom.cn

Guozhen Dong
China Telecom Corporation Limited Research Institute, Beijing, 102209, China
donggz@chinatelecom.cn

Wenjuan Xing
China Telecom Corporation Limited Research Institute, Beijing, 102209, China
xingwj@chinatelecom.cn

Shizhong Nie
China Telecom Corporation Limited Research Institute, Beijing, 102209, China
nieshzh@chinatelecom.cn

Yunpeng Xie
China Telecom Corporation Limited Research Institute, Beijing, 102209, China
xieyp@chinatelecom.cn

ABSTRACT

Data center networks need to provide various cloud services to a large number of users, which requires efficient and flexible management of network resources. However, the existing EVPN-VXLAN (Ethernet Virtual Private Network - Virtual Extensible Local Area Network) technology, which is widely used for network virtualization, cannot support multicast resource scheduling in the overlay network. Multicast service can save network bandwidth and reduce latency for high-demand applications, but it faces challenges of end-to-end delivery, path selection, and resource allocation in the data center EVPN-VXLAN network. This paper proposes a novel multicast scheduling method based on EVPN-VXLAN, which defines a new EVPN routing type to deliver multicast resource information and uses it for multicast routing selection and forwarding. The proposed method can improve the performance and efficiency of multicast service in the data center EVPN-VXLAN network at the overlay layer, and is compatible with the existing EVPN-VXLAN technology.

CCS CONCEPTS

• **Networks** → Networks protocols; Networks layer protocols.

KEYWORDS

Datacenter networks, EVPN-VXLAN, multicast scheduling, EVPN routing type

ACM Reference Format:
Nan Wu, Guozhen Dong, Wenjuan Xing, Shizhong Nie, and Yunpeng Xie. 2023. A Multicast Scheduling Method Based on EVPN-VXLAN Extension in Data Center Networks. In *2023 International Conference on Communication Network and Machine Learning (CNML 2023), October 27, 28, 2023, Zhengzhou, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3640912.3640978

CNML 2023, October 27, 28, 2023, Zhengzhou, China

ACM ISBN 979-8-4007-1668-3/23/10
https://doi.org/10.1145/3640912.3640978

1 INTRODUCTION

The need for scale in data centers continues to grow with the development of the Internet, video content, e-commerce, analytics applications, and customized cloud services. Data centers can provide a variety of cloud services, including applications, platforms, computing and storage infrastructure services, which can be accessed via the Internet or virtual private networks. Different services also have different requirements for compute, storage, connectivity, and bandwidth. There are also differences in the types of data centers, services, and how they are provided. For example, a private data center may serve multiple users within an enterprise, who may be connected via an enterprise private or virtual private network. Typically, this service is deployed inside the enterprise, but it can also be partially connected to the Internet through a secure gateway. Public data centers provide services to users through the Internet, including multi-user services and commercial public cloud services. In order to provide virtual services for a large number of users, data centers need to allocate computing resources and storage resources to users. It is very difficult to manage and arrange resources with traditional VLAN (Virtual Local Area Network) and IP routing [2]. A key step in network virtualization is to separate the virtual network from the physical network, so that the virtual network can be more easily managed, automated, and orchestrated.

At present, the data center network basically uses EVPN-VXLAN technology [4]. EVPN (Ethernet Virtual Private Network) is a kind of VPN technology for layer-2 network interconnection. EVPN technology adopts a mechanism similar to BGP/MPLS IP VPN, and defines a new NLRI (Network Layer Reachability Information) on the basis of BGP protocol, namely EVPN NLRI. The EVPN NLRI defines several new BGP EVPN route types for learning and publishing MAC addresses between different sites in a layer-2 network. There is no control plane in the original VXLAN implementation scheme, and the VTEP (Virtual Tunnel End Point) discovery and host information (including IP address, MAC address, VNI, gateway VTEP IP address) are learned by traffic flooding in the data plane, which leads to a lot of flooding traffic in the data center network. In order to solve this problem, VXLAN introduces EVPN as the control

plane, and realizes the automatic discovery of VTEPs and mutual notification of host information by exchanging BGP EVPN routes between VTEPs, thus avoiding unnecessary data traffic flooding.

However, the way that EVPN-VXLAN technology uses BGP EVPN routing information to schedule and manage services makes it impossible to realize the multicast resource scheduling of the overlay. Multicast service in data center network can improve the efficiency of high bandwidth demand network traffic by saving network traffic, and reduce the total delay of multicast traffic by reducing the operation of sending multiple copies of multicast flow to different destination servers. Most of the existing multicast services are based on the traditional three-layer network for resource scheduling, which is not suitable for the data center scenario of basic virtualization technology. Although some multicast network schemes use controllers to schedule each node, the controller scheduling scheme is subject to the controller performance and cannot automatically synchronize and schedule information based on network information, which cannot be applied to large data center scenarios and cannot meet the multicast resource scheduling of data center networks based on EVPN-VXLAN.

In this paper, a new EVPN routing type is defined to deliver multicast resource information by the overlay network, and the multicast scheduling method based on EVPN-VXLAN is implemented. The scheduling can be performed according to the multicast resource information table delivered by this type of routing, which meets the requirements of multicast scheduling in EVPN-VXLAN-based data center networks.

2 RELATED WORK

Multicast traffic scheduling is a key technology to enhance the efficiency and resource utilization of data center networks. However, existing multicast scheduling algorithms are either based on Internet protocols that are unsuitable for data center networks or depend on centralized controllers, which cause scalability and complexity problems. Moreover, these algorithms often fail to achieve load balancing and prevent network congestion, which degrades network performance and reliability. Therefore, in the academic community, various multicast scheduling algorithms have been devised for data center networks, especially fat-tree topologies. For instance, reference [1] introduces a reliable multicast protocol to support Internet of Things (IoT) applications in remote areas. The authors in [3] suggest a blocking cost-driven multicast scheduling algorithm, which applies optimization theory to minimize the blocking probability of the multicast subnetwork. Reference [5] presents a distributed dynamic multicast solution that enables hosts to join or leave multicast without interrupting real-time traffic. These studies aim to address the challenges and requirements of multicast traffic scheduling in data center networks in terms of scalability, efficiency, reliability, and load balancing. However, all of these researches focus on the theoretical analysis of the algorithm and do not mention how to achieve the virtualized resource multicast scheduling process through the popular EVPN-VXLAN technology.

Another research direction is to leverage Virtual Extensible Local Area Network for multicast traffic scheduling in data center networks. VXLAN enables creating a logical network that spans across different physical networks. It supports multicast communication by encapsulating and forwarding multicast packets through VXLAN tunnels. However, VXLAN also poses some challenges and limitations for multicast traffic scheduling, such as multicast table synchronization, multicast packet loss, and core device load. To address these issues, some studies have devised methods and devices for multicasting VXLAN through virtual tunnel end point devices, which are responsible for encapsulating and unencapsulating VXLAN packets. For instance, reference [6] presents methods and devices to synchronize multicast table entries on VTEP devices, which prevent multicast data flooding in VXLAN and ensure data delivery to multicast group members. Reference [7] introduces a method and device to multicast VXLAN through VTEP device to avoid broadcast, unknown unicast, and multicast (BUM) message loss, reduce core device load and enhance forwarding performance. The authors in [8] propose an application layer multicast system and method for data center networks, which achieve data transmission by multicast and instruct the multicast source node to send data to multiple destination nodes based on the network topology information. Reference [9] suggests a scheduling method for multicast flows in data center networks, which determines the reconfigurable crossbar switch (RCS) configuration graph and the matching RCS switch of a subset of inseparable multicast flows in the set of multicast flows to be transmitted and outputs the schedulable flow set, period and RCS configuration graph as the scheduling scheme. Reference [10] describes a method for VXLAN multicast communication using network devices, which register network devices as VTEP of VXLAN and use underlay network to transmit multicast traffic to multicast receivers without encapsulating multicast traffic in VXLAN packets. The author in [11] outlines a method and a device for forwarding VXLAN multicast data message from an IP core device in a VXLAN. The method involves searching matched multicast entries in a core replication multicast table, copying and transmitting the received VXLAN multicast data message to multiple multicast tunnel outlet ports and a loopback port, removing the outer layer VXLAN encapsulation of the loopback VXLAN multicast data message and performing local forwarding operation on inner layer data message. These solutions consider the operational mechanism and traffic scheduling of multicast services, but do not involve how to further support resource multicast scheduling by extending EVPN routing announcement.

The Underlay layer is the infrastructure layer network that represents the traditional IP network dedicated to carrying user traffic. It can provide packet forwarding service for the business. The Overlay layer is a virtual network built on top of the Underlay layer using tunneling technology. It is the service layer network or the user layer network that can transmit data based on the Underlay layer. In the related technologies, multicast service packets are forwarded on the Underlay layer. When the multicast service packets travel between two virtual tunnel end point devices, the intermediate nodes forward them hop-by-hop by looking up the multicast routing table entries. This requires each intermediate node to maintain multicast routing table entries, which increases the storage resource consumption of intermediate nodes. Moreover, the hop-by-hop forwarding of service packets also results in low processing efficiency of service packets.

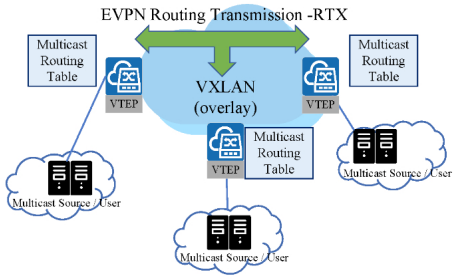

Figure 1: Illustration of multicast information announcement.

3 METHOD

In this paper, we propose a new route type (RT) in Ethernet VPN protocol to achieve the multicast routing announcement method, which synchronizes the multicast information of virtual tunnel end point nodes in the domain and enables multicast scheduling. The new RT in EVPN delivers multicast resources, while the original RT remains unchanged, and the multicast resources can be announced among VTEP nodes in the local domain. This way, the overlay-based multicast forwarding is achieved. Figure 1 shows the illustration of multicast information announcements.

There are five commonly used types of EVPN routing, and we name the new EVPN RT routing type RT6, which announces the multicast resource information in the EVPN VXLAN network, realizes the end-to-end multicast delivery, and does not need to maintain the multicast route forwarding representation on each node, saving equipment resources. The description of the message advertised by the route is as follows:

1. Ethernet A-D Route (Type1): Ethernet automatic discovery.

2. MAC/IP Advertisement Route (Type2): MAC address advertisement, used to announce the host MAC address, host ARP and host routing information.

3. Inclusive Multicast Route (Type3): Integrated multicast for automatic VTEP discovery and dynamic VXLAN tunnel establishment.

4. Ethernet Segment Route (Type4): ES members are automatically discovered.

5. IP Prefix Route (Type5): IP Prefix route (type5) is used to advertise incoming external routes as well as route information to the host.

6. Computing Power Route (Type6): Computing power route (type6) is used to advertise computing power route information.

As shown in Figure 2, the self-defined MGP-BGP multicast route transmits TLV-RT-6(VXLAN) packets in the following format:

Business process flow (as shown in Figure 3):

1. When the service node is online, VTEP1 node converts the multicast route according to the multicast information and announces it through RT6 routing.

2. VTEP2 receives the RT6 route to establish the EVPN VXLAN tunnel.

3. When client VM2 goes online, VTEP2 generates normal RT2/RT5 routes according to client IP for delivery.

RD (8 Bytes)	**The RD of IP-VRF**
ESI (10 Bytes)	**Iterate the clues, with Gw IP at least one of them being 0**
Ethernet Tag (4 Bytes)	**Broadcast domain identifiers**
Source Multicast Group Address (4 or 16 Bytes)	
Multicast Source Status (1 Bytes)	**0 is revocation and 1 is activation**
Multicast Address Requested (4 or 16 Bytes)	
Multicast User Status (1 Bytes)	**0 is revocation and 1 is activation**
IP Prefix Length (1 Bytes)	**0~32 for IPv4, 0~128 for IPv6**
IP Prefix (4 or 16 Bytes)	**Consistent with GW, IPv4/IPv6**
GW IP Address (4 or 16 Bytes)	
MPLS Labels 1 (3 Bytes)	**IP-VRF tag, VXLAN tunnel for P2P VNI tag**
MPLS Labels 2 (0 or 3 Bytes)	

Figure 2: The MGP-BGP multicast route transports the TLV-RT-6(VXLAN) packet format.

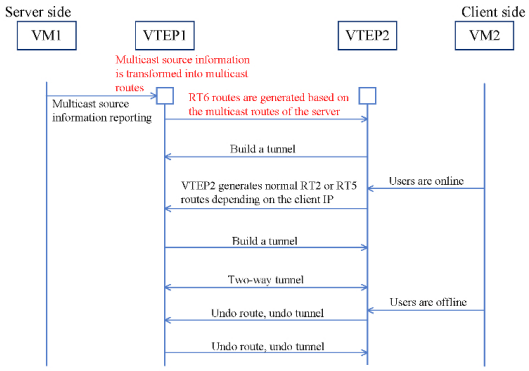

Figure 3: The business process of announcing multicast routing RT6.

4. VTEP1 establishes the EVPN VXLAN tunnel after receiving the RT2/RT5 route.

5. Bidirectional tunnel is established to realize bidirectional communication.

6. When the user leaves, the route is revoked, and with it the tunnel.

7. If the multicast source is out of service, the route will be canceled, and the tunnel will be canceled.

8. The multicast source update process is the same as (1) the service node on-line process.

4 ADVANTAGES

The scheme advertises the multicast route through the self-defined route type 6 (RT6), which enables the end-to-end delivery of multicast routes in overlay scenarios. In the data center scenario, the scheme adds multicast routing delivery rules, which do not consume underlay multicast routing resources and multicast protocols

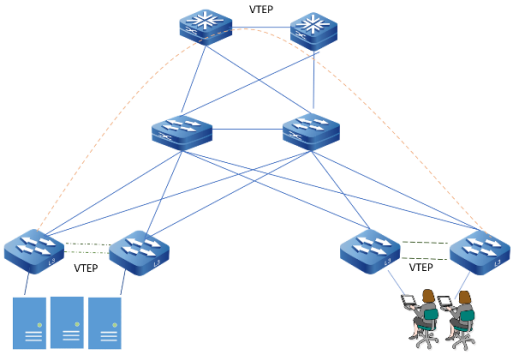

Figure 4: The network diagram of multicast forwarding process based on EVPN-VXLAN.

and improve forwarding efficiency. Moreover, in the data center EVPN-VXLAN network, the scheme delivers the multicast routing end-to-end and makes the intermediate link unaware, which effectively saves network resources. The scheme forms an overlay multicast transmission tunnel between the tunnel endpoints to connect servers and users, without the traditional multicast protocol hop-by-hop transmission and maintenance of a large number of multicast routing entries.

In the actual application process, we find that this mechanism can be widely used in data center EVPN-VXLAN scenarios. VTEP nodes use multicast routing as one of the forwarding rules to realize the rational application of resources. It can solve the end-to-end delivery problem of multicast routing and the path selection problem of multicast routing in the current data center EVPN-VXLAN network. On the basis of not changing the original RT routing type, the new selection of multicast routing can meet the requirements of multicast routing scheduling in the data center EVPN-VXLAN network.

5 EXPERIMENTAL SIMULATION

In this paper, the multicast forwarding process based on EVPN-VXLAN is verified through network simulation. As shown in Figure 4, the traditional three-layer CLOS architecture of the data center is used, and the multicast source and multicast users are connected to the VTEP respectively. The multicast traffic passes through the leaf to the spine and then to the DCGW, and then passes through the spine to the leaf. When the user goes online, the multicast join message is encapsulated in a VXLAN message and delivered to the DCGW. The DCGW node then transmits the message to the VTEP on the multicast source side, which completes the announcement of the entire multicast join message.

As shown in figure 5, a multicast forwarding table is generated on each VTEP node. Assuming that the user requests a multicast channel of 225.0.0.1, the corresponding multicast forwarding table will be generated. The table still contains the multicast source address and multicast exit. The intermediate spine node does not need to generate a multicast forwarding table, which avoids the hop-by-hop generation of multicast forwarding tables by underlay devices in traditional multicast forwarding, and directly achieves end-to-end transmission of the overlay.

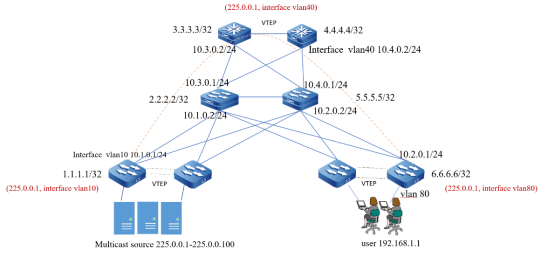

Figure 5: The diagram of the generation of a multicast forwarding table.

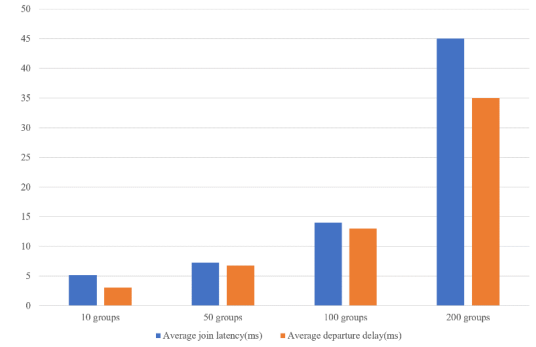

Figure 6: The diagram of experimental results.

After the verification test, taking 100 groups as an example, the multicast join/leave delay based on EVPN-VXLAN is shown in Figure 6. The results are basically in line with expectations and the delay changes greatly when the number of groups is large. The experimental results show that the multicast scheduling method based on EVPN-VXLAN is feasible, which can not only optimize the generation of multicast tables, but also meet service requirements for forwarding delay.

After the verification test, taking 100 groups as an example, the multicast join/leave delay based on EVPN-VXLAN is shown in Figure 6. The results are basically in line with expectations and the delay changes greatly when the number of groups is large. The experimental results show that the multicast scheduling method based on EVPN-VXLAN is feasible, which can not only optimize the generation of multicast tables, but also meet service requirements for forwarding delay.

6 CONCLUSION AND FUTURE DIRECTION

This paper proposes a multicast scheduling method based on EVPN-VXLAN for data center networks, which can improve the efficiency and performance of multicast services in the overlay network. The method defines a new EVPN routing type to deliver multicast resource information, and uses it as a basis for multicast routing selection and forwarding. The method can solve the problems of end-to-end delivery, path selection, and resource allocation of multicast routing in the data center EVPN-VXLAN network. This method can provide a way to solve the problems of end-to-end delivery, path selection and multicast routing resource allocation in the data

center EVPN-VXLAN network on the overlay layer. In the future, we plan to conduct more experiments to evaluate the performance of the proposed method in different network conditions and compare it with other multicast scheduling methods. We also intend to explore other possible applications of the new EVPN routing type in data center networks, such as load balancing, security, and fault tolerance.

REFERENCES

[1] Wong, K.-S. and Wan, T.-C. 2019. Current State of Multicast Routing Protocols for Disruption Tolerant Networks: Survey and Open Issues. Electronics, 8 (2): 162-2019. https://www.mdpi.com/2079-9292/8/2/162.

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

[3] Li, G., Guo, S. and Yang, Y. 2017. Blocking cost-driven multicast scheduling in fat-tree data center networks. Concurrency and Computation: Practice and Experience, 29 (19): e4166. doi: https://doi.org/10.1002/cpe.4166.

[4] Goralski, W. 2017. Chapter 21 - EVPN and VXLAN. In: The Illustrated Network (Second Edition), W. Goralski Ed. Boston: Morgan Kaufmann, San Francisco. pp. 535-560. doi:10.1016/B978-0-12-811027-0.00021-7.

[5] Fan, F., Hu, B. and Yeung, K. L. 2016. Distributed and dynamic multicast scheduling in fat-tree data center networks. In: 2016 IEEE International Conference on Communications (ICC). Cheng Du. pp. 1-6. doi: 10.1109/ICC.2016.7511250.

[6] Yang, B. 2017. Multicast table item synchronizing method, involves encapsulating multicast data message, and transmitting virtual extensible local area network encapsulated multicast data packet through virtual extensible local area network tunnel port. Patent CN107612809-A; CN107612809-B Patent Appl. CN107612809-A CN11045010 31 Oct 2017 CN107612809-B CN11045010 31 Oct 2017, [Online]. Available: <Go to ISI>://DIIDW:2018080263.

[7] Wang, Y. and Wang, F. 2016. Method for multicasting virtual extensible local area network by virtual tunnel end point device, involves sending big bum message to virtual tunnel end point device based on copy mode in multicast group. Patent CN106656722-A Patent Appl. CN106656722-A CN11141814 12 Dec 2016, [Online]. Available: <Go to ISI>://DIIDW:2017318830.

[8] Tian, C., Wang, Y. and Wang, B. 2019. Data center network application layer multi-cast system, has multi-cast scheduling module for performing scheduling process based on network topology information until multiple destination nodes receives data. Patent CN110213063-A; CN110213063-B Patent Appl. CN110213063-A CN10499361 10 Jun 2019 CN110213063-B CN10499361 10 Jun 2019, [Online]. Available: <Go to ISI>://DIIDW:2019789974.

[9] Luo, L., Yu, H., Sun, G. and Jin, Q. 2019. Method for scheduling multicast stream in data center network, involves removing non-split multicast stream subset from multicast streams, and outputting scheduleable flow set, cycle duration and configuration diagram as scheduling scheme. Patent CN111131064-A; CN111131064-B Patent Appl. CN111131064-A CN11416495 31 Dec 2019 CN111131064-B CN11416495 31 Dec 2019, [Online]. Available: <Go to ISI>://DIIDW:202043447W.

[10] mmidi, K., Kumar, S., Nidumolu, K. C., Srinivas, P. and Yue, G. 2018. Method for communicating multicast traffic in virtual extensible local area network, involves receiving multicast traffic from multicast source, and transmitting multicast traffic to multicast receivers using underlay network. Patent US2019207779-A1; US10880112-B2 Patent Appl. US2019207779-A1 US234415 27 Dec 2018 US10880112-B2 US234415 27 Dec 2018, [Online]. Available: <Go to ISI>://DIIDW:2019575415.

[11] Cheng, J. 2019. Method for forwarding multicast data message of internet protocol core device in virtual extensible local area network, involves receiving loopback multicast data message, and performing local forwarding operation on inner layer message. Patent CN110445702-A; CN110445702-B Patent Appl. CN110445702-A CN10615237 09 Jul 2019 CN110445702-B CN10615237 09 Jul 2019, [Online]. Available: <Go to ISI>://DIIDW:2019971611.