IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 2, 2017, accepted October 2, 2017, date of publication October 12, 2017,
date of current version November 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2762362

SDN Based VxLAN Optimization in Cloud Computing Networks

ZHIFENG ZHAO, FENG HONG, AND RONGPENG LI
College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

Corresponding author: Rongpeng Li (lirongpeng@zju.edu.cn)

This work was supported in part by the Program for Zhejiang Leading Team of Science and Technology Innovation under Grant 2013TD20, in part by the National Natural Science Foundation of China under Grant 61731002 and Grant 61701439, in part by the National Postdoctoral Program for Innovative Talents of China under Grant BX201600133, and in part by the Project funded by China Postdoctoral Science Foundation under Grant 2017M610369.

ABSTRACT Nowadays, cloud computing networks significantly benefit the deployment of new services and have become one infrastructure provider in the digital society. The virtual extensible local area network (VxLAN), which belongs to the network schemes to overlay Layer 2 over Layer 3, is one of the most popular methods to realize cloud computing networks. Though VxLAN partially overcomes the capacity limitation of its predecessor virtual local area network, it still faces several challenges like annoying signaling overhead during the multicast period and interrupted communication with a migrating virtual machine (VM). In this paper, we propose a new software defined network (SDN) based VxLAN network architecture. Based on this architecture, we address how we can deploy an intelligent center to enhance the multicast capability and facilitate the VM migration. Besides, we discuss the means to update the global information and guarantee the communication continuum. Finally, we use Mininet to emulate this SDN based VxLAN architecture and demonstrate effective load balancing results. In a word, this proposed architecture could provide a blueprint for future cloud computing networks.

INDEX TERMS Cloud computing networks, software defined networks (SDN), virtual extensible local access networks (VxLAN).

I. INTRODUCTION

“*Clouds come floating into my life from other days,*
no longer to shed rain or usher storm,
but to add color to my sunset sky.”

Every time we read the poem by the Indian poet Rabindranath Tagore, we seam to see the blossoming clouds embellished scene. Similar to the real life where clouds float here and there flexibly, the digital life is empowered by cloud computing techniques, in which shared computer processing resources and data are provided to their tenants on demand with broad network access and rapid elasticity [1]. For a tenant, network elements are invisible, as if obscured by a cloud. Apparently, cloud computing networks should possess dynamic infrastructure and flexible resource management capability, so as to deliver the content fast. In order to obtain such a goal, cloud computing networks usually rely on the mechanism of overlay networks (e.g., virtual local access networks (VLAN) and its successor virtual extensible LAN (VxLAN)) [2]. In other words, the cloud computing network is built on top of another network and the nodes in this cloud network are connected by virtual or logical links, which might correspond to many physical links in the underlying network. For example, VLAN and VxLAN overlay the Layer 2 in the open system interconnection (OSI) reference model over Layer 3 networks, and it is thus possible to conveniently assign one unique virtual ID to a tenant of the cloud computing networks. However, as VLAN merely stores around 4000 IDs in its tag field, it suffers in the cloud environment with a large number of tenants. Therefore, the Internet engineering task force (IETF) advocates VxLAN to replace VLAN. In particular, VxLAN realizes the tunneling by encapsulating original frames with additional headers (e.g., outer Internet protocol (IP) address, outer user datagram protocol (UDP) address and VxLAN headers). Usually, the VxLAN header includes 24 bits of VNI (VxLAN Network Identifier) and VxLAN supports more than 16 million tenants to coexist within the same physical infrastructure of cloud computing networks. However, as clouds in the real life might be murky to obscure the sun, the VxLAN based cloud computing network is also imperfect and still faces several challenges.

- In order to extend the overlay capacity, VxLAN deploys IP multicasting techniques [3] by encapsulating the original message with headers at one VxLAN tunnel end point (VTEP) and sending the duplications out to other VTEPs to jointly determine the destination host. Inevitably, during this multicasting procedure, it costs a significant amount of extra signaling traffic.
- In cloud computing networks, it is common to migrate VMs to better providing services to tenants. Currently, the VxLAN network has to manually migrate the VMs and lacks an intelligent mechanism to automatically supervise the VM traffic and balance the loads. Besides, the migrating VMs will have to interrupt the communication. Consequently, it fails to gain the required scalability to cope with millions of tenants.

SDN (Software Defined Networks) based VxLAN is a concept of introducing SDN as the control plane [4]–[6], meanwhile employing the VxLAN network structure as the data plane. In this way, the notion of separating control plane from data plane is introduced, and by taking advantage of global view and programmable features, it is able to make a centralized control for VxLAN networks and prominently improve the performance of the network, and optimize the VxLAN mechanism. In the remainder of this paper, we envision an SDN-based VxLAN network architecture for realizing a large-scale cloud network, by making full use of the light weight architecture of VxLAN overlay network and the high control flexibility of SDN switches for building up a cost-efficient large-scale network with superior performance. Later, we use several examples (e.g., IP multicasting, VM migration, and load balancing) to demonstrate the advantages of this architecture. We boldly argue that the SDN-based VxLAN architectures perfectly meet the challenges in future cloud computing networks.

II. THE SDN BASED VxLAN NETWORK

Usually, as depicted in the shaded part of Fig. 1, the VxLAN based cloud computing network depends on the leaf-spine structure, in which gateways and switches serve as spines and VTEPs, respectively. In particular, the spine allows VMs to connect with different subnets and routing the traffic from/to the servers in the distant network. The VxLAN leverages the tunneling technology to transmit network data, so that VMs could pay no attention to potential network variants. Meanwhile, the distributed spine architecture avoids the potential bottleneck in a centralized scheme. To realize such a goal, the VxLAN uses edge switches (or VTEP in the exact terminology) in the access layer to encapsulate and decapsulate VxLAN messages. In other words, the VxLAN tunnel is a logical or virtual channel between two VTEPs and responsible for transmitting the encapsulated VxLAN messages. In addition, VMs with different VNIs under the same VTEP no longer need to go through the spine to complete the connection, since the VTEP itself could complete the work. In the traditional VxLan network, both spines and VTEPs possess limited controlling capability, and is in independent charge of traffic routing based on their own local information.

In this paper, we put forward a three-tier SDN based VxLAN network architecture in Fig. 1. Compared to the traditional VxLAN network, though spines and VTEPs still appear in the core layer and access layer, the control functionalities have been decoupled from them, and converged into the brand-new control layer. Besides centralizing most of control functionalities into the SDN controllers and leaving few in the hypervisor of the access layer, the control layer also include an intelligent center [7]. The intelligent center embodies typical artificial intelligence algorithms and empower the cloud computing network by interacting with other types of applications. For example, the intelligent center can run some classification algorithms (e.g., gradient boosting decision tree or replicator neural networks [8]) to find flow patterns and sense the VM resource utilization, after being fed the latest network information by the database application like the IP address, MAC address, VNI (VXLAN Network Identifier), VTEP ID (VXLAN Tunnel End Point ID) of the VM. The VM migration application can rely on the intelligent center to smartly determine when and where to migrate a certain VM by supervising the traffic load of the VTEPs. It is also possible to sense the requirements of migration from the tenant and start a migration procedure proactively or passively. Furthermore, based on the reasoning algorithms in the intelligent center (e.g., actor-critic method, Q-learning method), the load balancer is possible to analyze the traffic load information of the VTEP and Spine ports, so as to accurately predict the load variations and timely balance the traffic load among the VTEPS and Spines.

In summary, our proposed SDN-based VxLAN network realizes a flexible large-scale cloud network. As listed in Table 1, the design aims to make full use of the light weight VxLAN overlay network and the high control flexibility of SDN to build a novel data center network with significant improvement on multi-tenancy management, higher server capacity, enhanced stability, centralized yet flexible control.

III. APPLICATIONS OF SDN BASED VxLAN NETWORK

In this part, we highlight some typical examples to manifest the advantages of SDN based VxLAN network and demonstrate the contributions of the intelligent center.

A. IP MULTICASTING

VxLAN realizes network multicast technology by emulating a resolution process similar to the address resolution protocol (ARP). Traditionally, after an encapsulated flow is sent from the source VTEP, its duplications will also be distributed to other VTEPs. At each VTEP, there exists a mapping table to indicate the relationship between the VNIs and IP multicast addresses [9]. If the mapping table indicates that all the VMs attached to one VTEP doest not belong

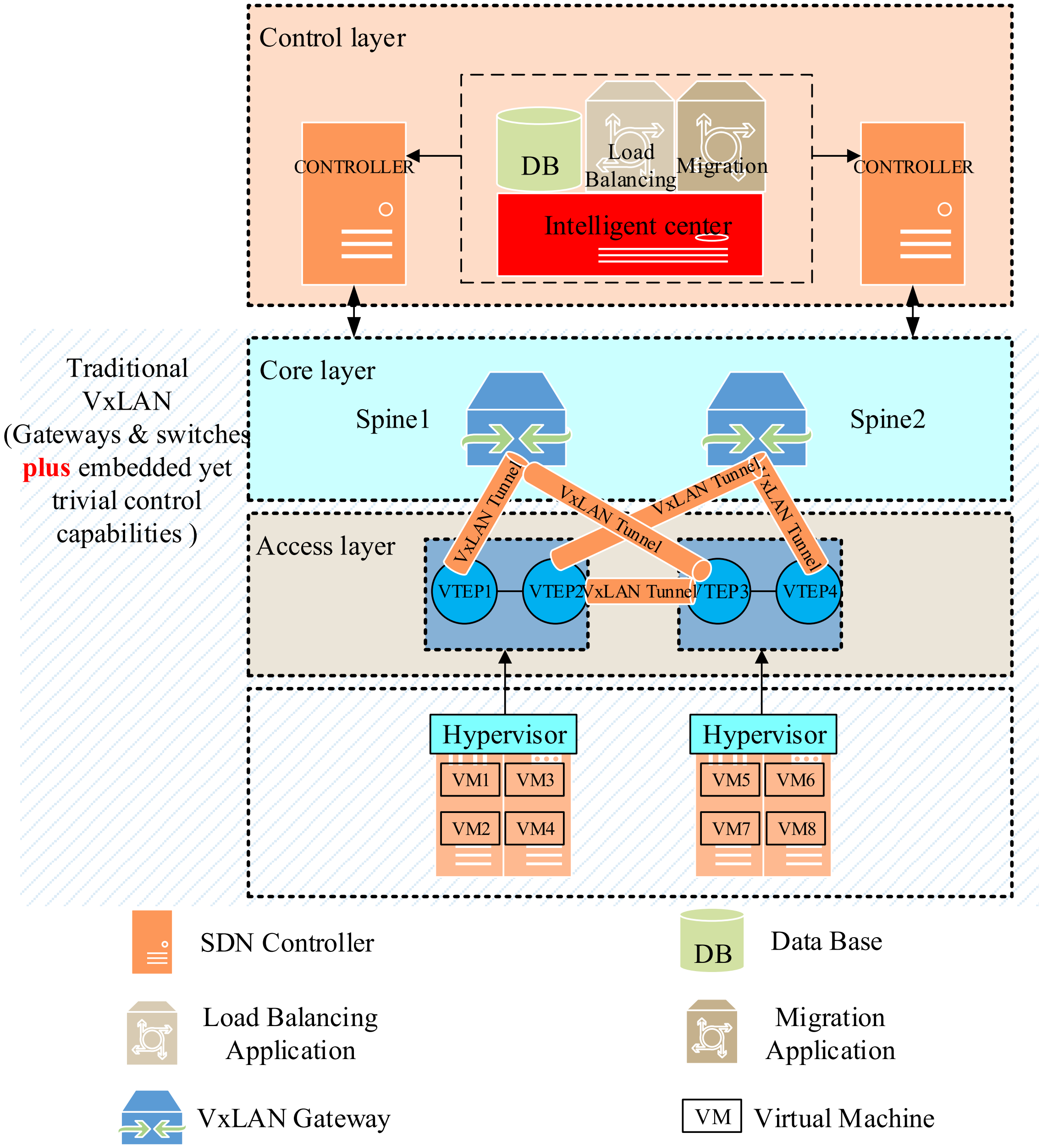


FIGURE 1. The illustration of the SDN based VxLAN network architecture.

TABLE 1. The functionalities of SDN based VxLAN and traditional VxLAN.

Functionalities	SDN Based VxLAN	Traditional VxLAN
Tenant Management	Real-time supervision	Manual Configuration
VM Dilatation	Automatical configuration	Manual Configuration
Fault Detection	Enhanced capability by global information	Difficult to locate the fault
New Service	Flexible deployment in the intelligent center	Inevitable hardware configuration modifications
VM Migration	Proactive or Paasive	Manual Configuration
Load Balancing	Automatical detection	Some simple mechanism.

to the multicast groups, the VTEP will discard the related message. Otherwise, the VTEP will forward to the related VMs. Besides, when the mapping table changes, VTEPs will change their routing policy consistently [3]. Obviously, during the ARP procedure, it requires a great amount of resources to store the mapping table and generates some useless messages. In this part, we show how SDN based VxLAN network with the intelligent center could contribute

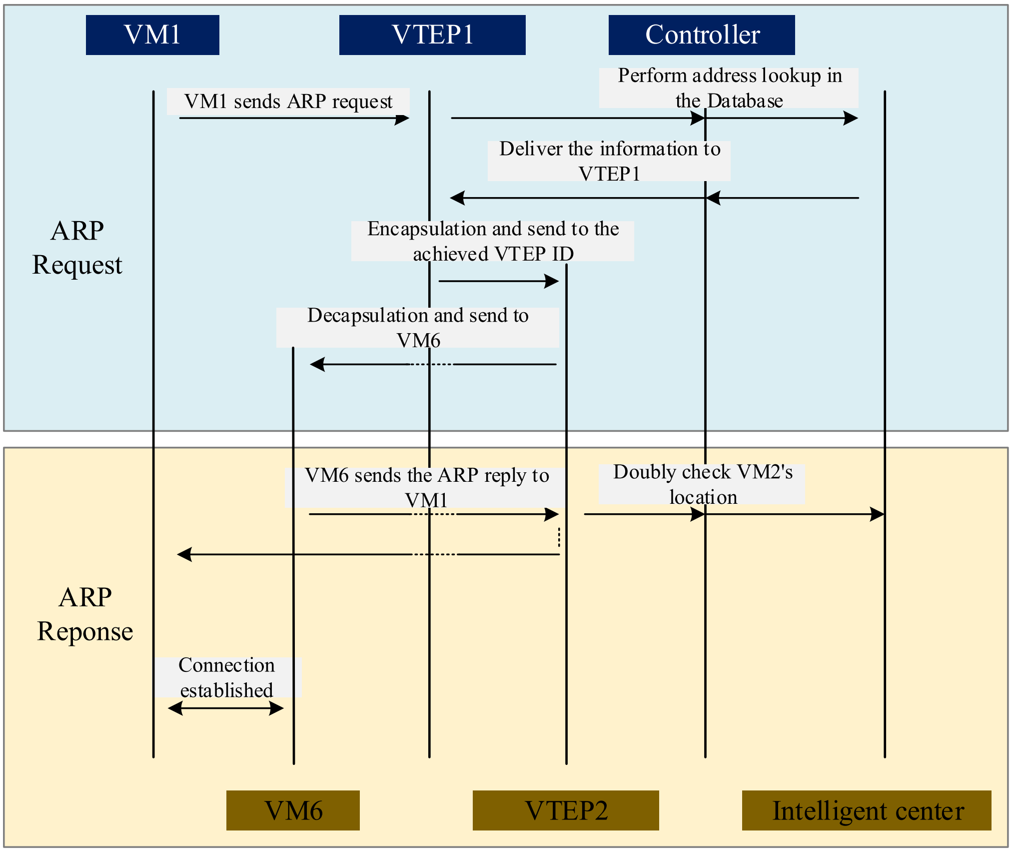


FIGURE 2. The multicast diagram in the SDN based VxLAN network.

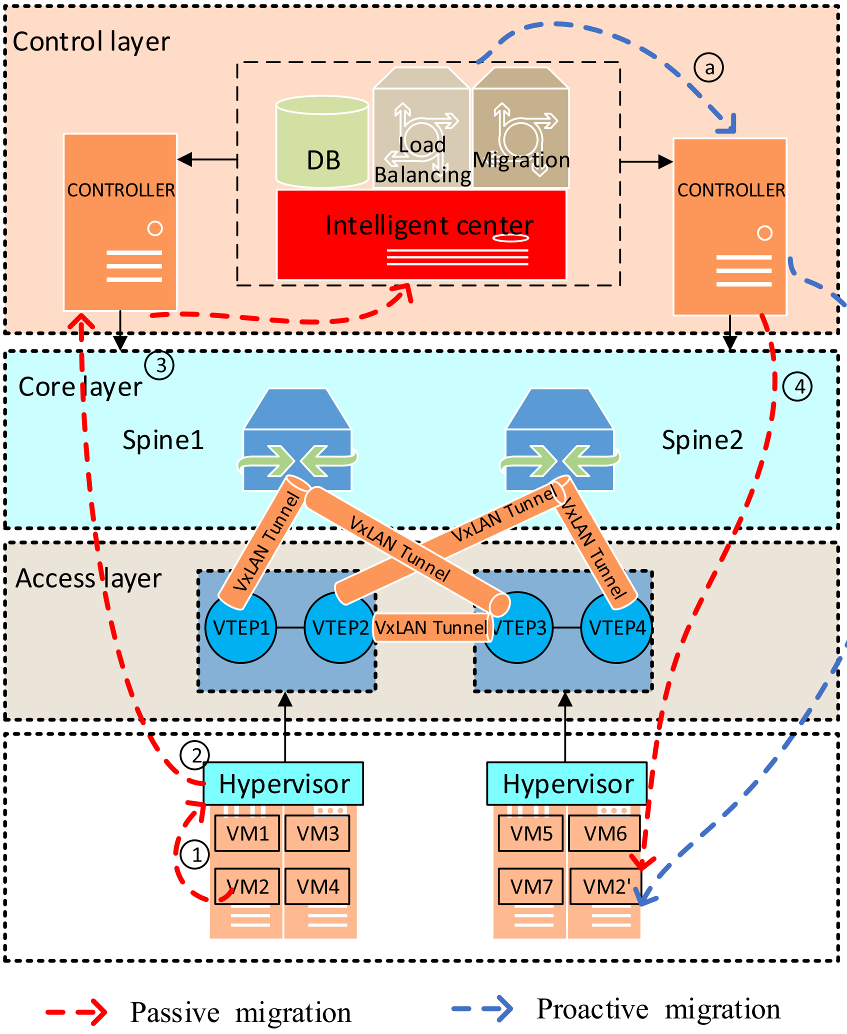
to solving the aforementioned issues. As depicted in Fig. 2, the related ARP procedure here is significantly distinct from that in the traditional method. Whenever a source VM plans to start a multicasting procedure, it first queries the intelligent center for the MAC address and VTEP ID, instead of initiating an ARP request. The intelligent center performs address lookup in the database and delivers the information to the VTEP which the destination VMs belong to. Afterwards, the messages can be easily forwarded to the destination VM, and the corresponding response packet generated by the destination VM can also be used by the intelligent center to doubly check the location of the receiver. If the response is not received in time by the SDN controller, the controller could ask the monitor application to track the actual location of the receiver and update the database. In this way, the network is able to implement the multicast method with less link resource cost and reduced negative impact on the network itself.

B. VM MIGRATION

In the conventional VxLAN network, the VM migration mainly relies on manual configuration, and it is unable to establish a new connection during the configuration process until the end of the migration, thus lasting long time for a VM to re-start the communication [10]. In the SDN based VxLAN network, the system could benefit from the centralized control and the intelligent center to complete the VM migration in a either proactive or passive manner. Moreover, the migrating VM can set up smooth connection during the process of migration [11].

As depicted in Fig. 3, the passive migration manner implies that the system could initiate a VM migration process, by following the tenant command. Specifically, the tenant sends out a request to migrate a VM, which explicitly indicates the destination of migration. The request will be achieved by the hypervisor, and will be sent to the corresponding SDN controller to kick off the migration process. Comparatively, the proactive migration manner means that the system will conduct a migration process automatically if the intelligent center finds the system status meets the pre-defined requirements. For example, for the passive manner [12], the intelligent center will have a real-time supervision of the VTEP ports to gather the load information in the database and periodically feed them to the monitor application. Conceptually, when the load of a VTEP exceeds the threshold, the system will perform the VM migration to offload some VM from one VTEP with heaviest throughput to another one with lighter one.

On the other hand, the SDN based VxLAN scheme could facilitate the VM migration process. Without loss of



① The *tenant* launches a migration request, and explicitly indicates the destination location;

② The *hypervisor* receives the request;

③ The *hypervisor* sends the request to the corresponding SDN controller.

④ The *migration* module starts to operate, and the migration process initiates.

ⓐ The *monitor* module detects that the load of a VTEP exceeds the threshold.

ⓑ The *migration* module determines the migrating target and the destination location based on the load information. Afterwards, it starts to operate.

FIGURE 3. **The comparison between the proactive and passive VM migration in the initialization part.**

generality, Fig. 4 illustrates the procedure by migrating VM 2 from Spine 1 to Spine 2. In the first place, the SDN controller will construct a mirror VM 2' and randomly generates the corresponding IP and MAC address and VNI. Afterwards, the intelligent center will collect the related information by informing the monitor application and updating the database. Next, the SDN controller establishes a link between the VM 2 and VM 2', so that VM 2 could transfer the data and

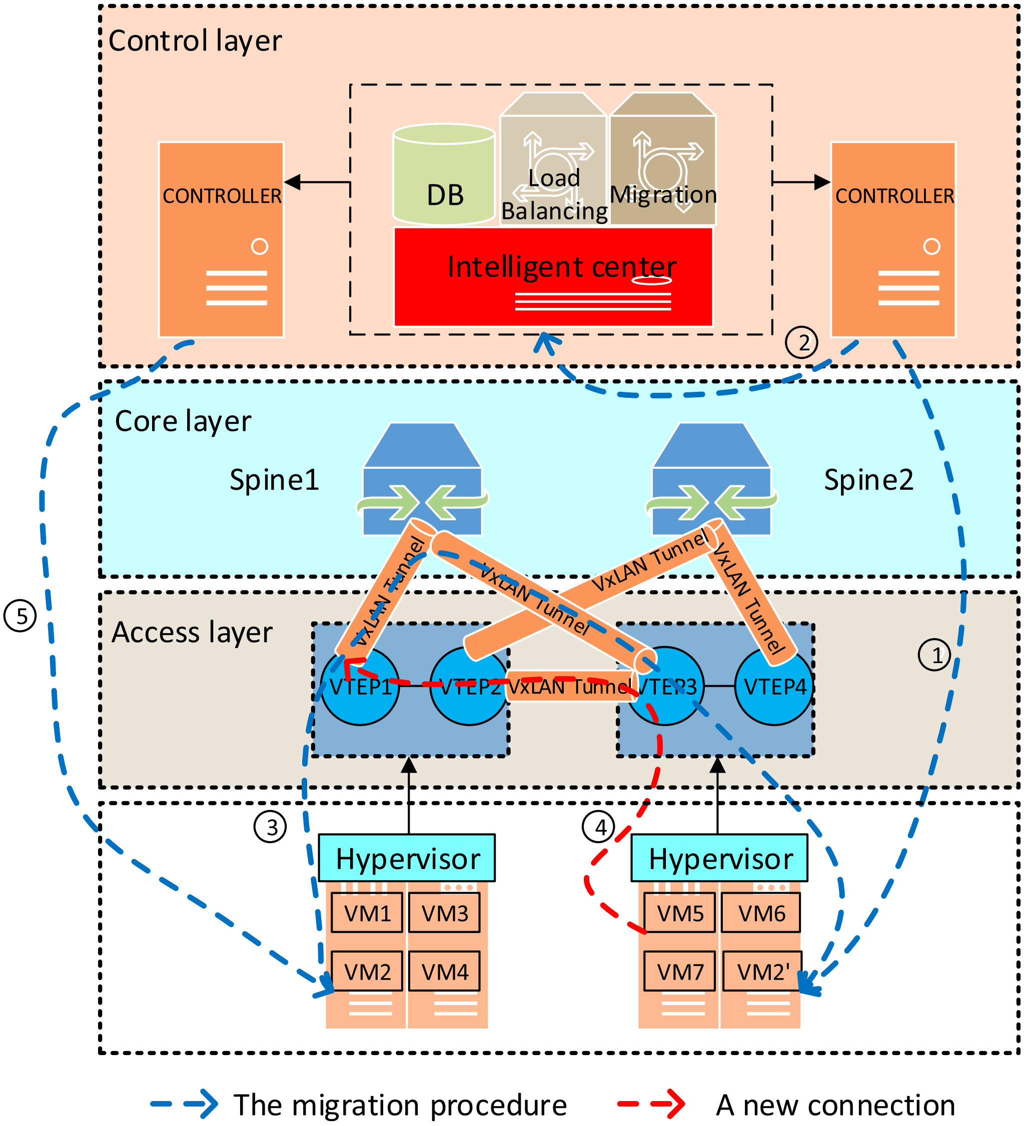


FIGURE 4. The migration process in the SDN based VxLAN process.

configuration to VM 2'. During the migration process, if another VM (i.e., VM 5) wants to communicate with VM 2, VM 5 sends the ARP request to VTEP 1. VTEP 1 begins to search the database and finds that VM 2 is migrating to VM 2'. Hence, VTEP 1 forwards the request from VM 5 to VM 2' and a connection between the VM5 and VM 2' is established. Once the migration process is completed, the IP and MAC address of VM2' will be modified to the same content of VM2. Meanwhile, the corresponding table entries in the database will be updated. Therefore, the communication with a migrating VM is guaranteed.

C. LOAD BALANCING

In this part, we provide a specific example to show how the SDN based VxLAN network could contribute to the load balancing, by taking full advantage of the centralized control and the intelligent center. Specifically, we investigate the effect of proactive migration threshold on the network performance, since we argue that the intelligent center could sense the network performance and adaptively optimize the threshold.

Our test scenario follows the illustration of Fig. 5(a) and consists of one Layer 3 gateway (i.e., Spine) and four VTEPs, each of them having four VMs. Besides, we adopt the Mininet [13] to create such a realistic virtual network, in which we select RYU OpenFlow controllers [14] and Openvswitch as SDN controllers and VTEPs and Spine [15] respectively and generate the topology of the IP network. Notably, we do not specify the topology environment and all switching devices are managed by the RYU controller. We also emulate the traffic from each VM to its random destination by *hping3*, with the load and inter-arrvial time of

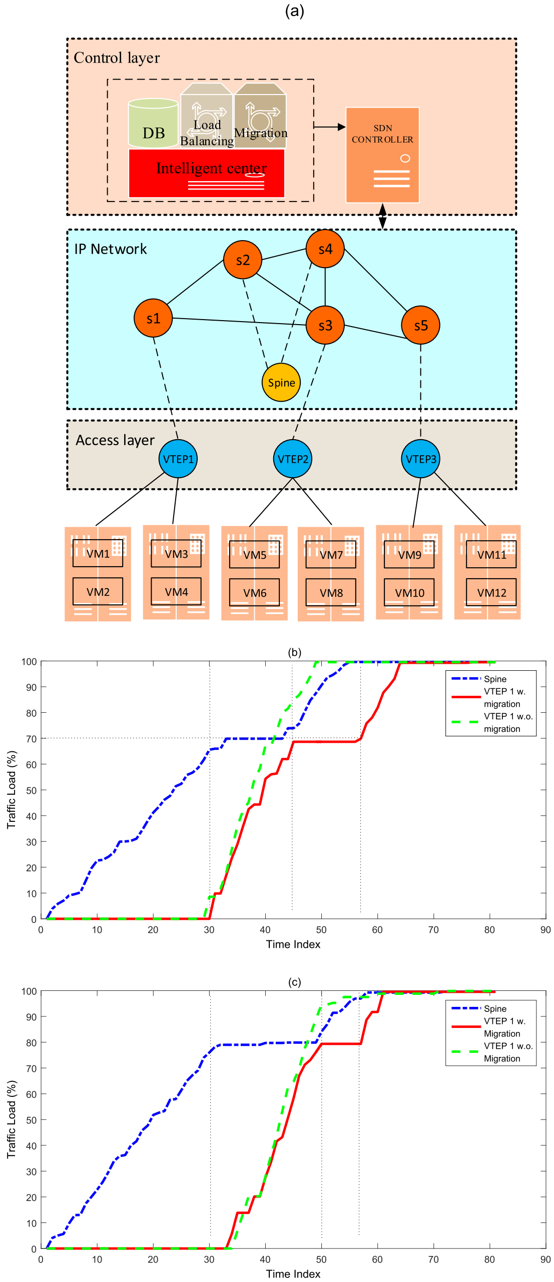


FIGURE 5. **(a) The Mininet-based simulation scenario; (b)∼(c) The load of Spine and VTEPs during a load balancing procedure when the threshold differs.**

flows satisfying exponential distribution. Since spines usually exhibit stronger performance than VTEPs, the spines are selected as preferential load balancing devices. The maximum load, which spines and VTEPs could support, is assumed to be as 20GB/s and 10GB/s, respectively. Fig. 5(b)∼(c) depicts the performance comparison when the migration threshold of the spine is set as 70% and 80%. In other words, with the increase of each flow load, the network starts to shift the traffic and VMs and different tenants start to communicate under different anchoring points.

During the load balancing procedure, the intelligent center will intuitively migrate the VM with the heaviest throughput to other Layer 3 equipment (e.g., other VTEPs). But more advanced reinforcement learning based migration policies [8] are still applicable in this case. As depicted in Fig. 5(a), along with the increase of traffic flow, the load of the Spine first arrives at the threshold (i.e., 70%), since the spine is chosen as the preferential device and most of the flows go through the Spine. Then, the traffic begins to be offloaded to VTEPs. When the load of one VTEP (i.e., VTEP 1) also gets to 70%, the intelligent center will conduct proactive migration to reduce the load of VTEP 1. Consistent with our intuition, both the Spine and VTEPs finally reach 100% utilization. Fig. 5(b) provides the related results when the threshold changes to 80%. From Fig. 5(b), most of the trends remain but the load balancing procedure is postponed, which implies some spines or VTEPs will operate under higher utilization. In other words, the intelligent center could further leverage some intelligent algorithms to adapt the threshold, so as to better provision the VMs.

IV. CONCLUSION

In this paper, we have envisioned a three-tier SDN based VxLAN architecture including control layer, core layer, and access layer. In particular, we have assessed the advantages of such an architecture like reduced resource cost for IP multicasting, enhanced communication continuum during the VM migration, and advanced and agile configuration for load balancing. We boldly argue that SDN based VxLAN architecture can meet the requirements for realizing a highly reliable large-scale multi-tenant cloud computing networks and adding more sunshine to the digital life.

REFERENCES

[1] P. Mell and T. Grance, “The NIST definition of cloud computing,” Rec. Nat. Inst. Standards Technol., Gaithersburg, MD, USA, NIST Special Publication 800-145, Sep. 2011.

[2] M. Mahalingam *et al.*, *Virtual Extensible Local Area Network (VxLAN): A Framework for Overlaying Virtualized Layer 2 Networks Over Layer 3 Networks*, document IETF RFC 7348, Aug. 2014.

[3] Y. Nakagawa, K. Hyoudou, and T. Shimizu, “A management method of IP multicast in overlay networks using openflow,” in *Proc. ACM HotSDN*, Helsinki, Finland, Aug. 2012, pp. 91–96.

[4] S. S. W. Lee *et al.*, “Design of SDN based large multi-tenant data center networks,” in *Proc. IEEE CloudNet*, Niagara Falls, ON, Canada, Oct. 2015, pp. 44–50.

[5] F. Hu, Q. Hao, and K. Bao, “A survey on software-defined network and OpenFlow: From concept to implementation,” *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2181–2206, 4th Quart. 2014.

[6] H. Kim and N. Feamster, “Improving network management with software defined networking,” *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 114–119, Feb. 2013.

[7] Y. Zhou, Z. Zhao, and H. Zhang, “Towards 5G: Hetegogeneous cellular network architecture design based on intelligent SDN paradigm,” *J. Telecommun. Sci.*, no. 6, pp. 28–35, Jun. 2016.

[8] R. Li *et al.*, “Intelligent 5G: When cellular networks meet artificial intelligence,” *IEEE Commun. Mag.*, 2017, to be published.

[9] A. Craig, B. Nandy, I. Lambadaris, and P. Ashwood-Smith, “Load balancing for multicast traffic in SDN using real-time link cost modification,” in *Proc. IEEE ICC*, London, U.K., Jun. 2015, pp. 5789–5795.

[10] N. F. Mir, J. N. Kotte, and G. A. Pokuri, “Implementation of virtualization in software defined networking (SDN) for data center networks,” in *Proc. ICN*, Jun. 2016, p. 136.

[11] A. Mayoral, R. Vilalta, R. Muñoz, R. Casellas, and R. Martínez, “Experimental seamless virtual machine migration using an integrated SDN IT and network orchestrator,” in *Proc. OFC Conf.*, Anaheim, CA, USA, Mar. 2015, pp. 1–3.

[12] J. Liu, Y. Li, and D. Jin, “SDN-based live VM migration across datacenters,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 583–584, Apr. 2015.

[13] K. Kaur, J. Singh, and N. S. Ghumman, “Mininet as software defined networking testing platform,” in *Proc. ICCCS*, Chennai, India, Feb. 2014, pp. 1–4.

[14] R. Khondoker, A. Zaalouk, R. Marx, and K. Bayarou, “Feature-based comparison and selection of software defined networking (SDN) controllers,” in *Proc. IEEE WCCAIS*, Hammamet, Tunisia, Jan. 2014, pp. 1–7.

[15] B. Pfaff *et al.*, “The design and implementation of open vswitch,” in *Proc. ACM/IEEE NSDI*, Santa Clara, CA, USA, Mar. 2015, pp. 1–14.

ZHIFENG ZHAO received the bachelor’s degree in computer science, master’s degree in communication and information System, and Ph.D. degree in communication and Information system from the PLA University of Science and Technology Nanjing, China, in 1996, 1999, and 2002, respectively. From 2002 to 2004, he was a Post-Doctoral Researcher with Zhejiang University, where his research was focused on multimedia next-generation networks and soft-switch technology for energy efficiency. He is currently an Associate Professor with the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. From 2005 to 2006, he was a Senior Researcher with the PLA University of Science and Technology, where he performed research and development on advanced energy-efficient wireless routers, ad hoc network simulators and cognitive mesh networking test-beds. His research interests include cognitive radio, wireless multi-hop networks (ad hoc, mesh, WSN), wireless multimedia networks and green communications. Dr. Zhao was the Symposium Co-Chair of ChinaCom 2009 and 2010. He is the Technical Program Committee Co-Chair of the 10th IEEE International Symposium on Communication and Information Technology.

FENG HONG received the B.E. and M.E. degrees Xidian University, Xi’an, China and Zhejiang University, Hangzhou, China in 2013 and 2017, respectively. His research interests are in network optimization and artificial intelligence.

RONGPENG LI received the B.E. degree from Xidian University, Xi’an, China, in 2015, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2010. He is currently a Post-Doctoral Researcher with the College of Computer Science and Technologies and the College of Information Science and Electronic Engineering, Zhejiang University. From 2015 to 2016, he was a Researcher with the Wireless Communication Laboratory, Huawei Technologies Co. Ltd., Shanghai, China. He was a Visiting Doctoral Student with Supélec, Rennes, France in 2013, and an Intern Researcher with the China Mobile Research Institute, Beijing, China, in 2014. He has authored or co-authored several papers in related fields and his research interests currently focus on resource allocation of cellular networks (especially full duplex networks), applications of reinforcement learning, and analysis of cellular network data. He was granted by the National Postdoctoral Program for Innovative Talents, which has a grant ratio of 13% in 2016. He serves as an Editor for *China Communications*.

• • •