IEICE TRANS. COMMUN., VOL.E95–B, NO.8 AUGUST 2012

INVITED PAPER *Special Section on Networking Technologies for Cloud Services*

Network Virtualization Technology to Support Cloud Services

Hideo KITAZUME[†a)], ***Member***, **Takaaki KOYAMA**[†], ***Nonmember***, **Toshiharu KISHI**[†], ***and*** **Tomoko INOUE**[†], ***Members***

SUMMARY Recently, server virtualization technology, which is one of the key technologies to support cloud computing, has been making progress and gaining in maturity, resulting in an increase in the provision of cloud-based services and the integration of servers in enterprise networks. However, the progress in network virtualization technology, which is needed for the efficient and effective construction and operation of clouds, is lagging behind. It is only recently that all the required technical areas have started to be covered. This paper identifies network-related issues in cloud environments, describes the needs for network virtualization, and presents the recent trends in, and application fields of, network virtualization technology.

key words: *cloud, open source, network virtualization, OpenFlow*

1. Introduction

The recent progress in server virtualization technology has brought about the following in the cloud service environment: (1) improved efficiency in the use of physical servers, (2) sharing of servers by multiple tenants, and (3) the needs for virtual machine migration. With regard to the network in a data center, there is a growing concern over the explosion of the number of MAC addresses and the number of VLANs, over the construction of L2 networks that traverse server sites, and over the migration across networks.

The attempted solutions to these issues can be broadly classified into two approaches: (i) the architecture that extends the existing technologies, and (ii) network virtualization. The former approach uses high-speed, high-capacity L2 switches, typically forming a 40-Gb to 100-Gb Ethernet. Several L2 switches are interconnected with only a single hop on a flat, non-hierarchical, basis, and are operated and managed as a huge, logically single L2 switch. Although this technology is expected to become widespread as a next-generation technology for large data center networks, it poses some problems, such as a heavy dependence on switch vendors, inadequacy in coordinated operation with other network devices such as firewalls and load balancers, and the inability to reduce network construction costs.

In the network virtualization approach, issues of physical networks shch as MAC address and VLAN-ID explosion are solved by making the logical network independent of the physical network, and network functions such as firewalls and load balancers are logically integrated by swich software which construct logical networks.

OpenFlow [1], which is being actively standardized to develop programmable switch, are attracting attention because they facilitate the design of multi-tenant networks and allow flexible virtual machine (VM) migration. LISP, which is being introduced to proprietary products, and VXLAN [2], which is planned to be introduced to these products, are also attracting attention. They are now considered to be promising network virtualization technologies for building software-designed networks in a cloud environment.

This paper discusses the needs, technologies, and application areas of network virtualization.

2. Needs for Network Virtualization

As was mentioned in the previous section, cloud services are required to isolate the network of an individual user in order to make multi-tenant servers possible. Also, they are required to allow networks to be constructed and their structure to be changed flexibly and quickly to achieve merger or migration of servers. The methods based on existing VLANs and their products cannot satisfy these demands easily.

2.1 Difficulty in Designing and Managing Networks within a Data Center

In data centers, tag VLANs are usually used to isolate the network dedicated to an individual user. However, this method poses two problems. One problem is the limit in the number of packets that a tag VLAN can handle. In a tag VLAN, one of 4094 VLAN-IDs is attached to each packet. Therefore, a tag VLAN can support only up to 4094 networks. Since each network in a data center needs to have a unique VLAN-ID, a tag VLAN cannot serve more than 4094 users simultaneously. The other problem is that data center products are based on vendor-specific specifications. There are many vendors that adopt proprietary specifications for the design and configuration of their data center products. For example, with regard to network design, one vendor provides a virtual chassis that makes multiple switches look like a single switch, while another vendor uses a tag VLAN for communication between server products. With regard to configuration, there are several vendor-specific control protocols. This situation makes it necessary to adjust VLAN-

Manuscript received December 22, 2011.
Manuscript revised April 4, 2012.

[†]The authors are with NTT Information Sharing Platform Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
a) E-mail: kitazume.hideo@lab.ntt.co.jp

DOI: 10.1587/transcom.E95.B.2530

IDs and control protocols to each vendor's specification. In addition, the design and management of networks in a data center has become increasingly difficult because the number of servers and switches in a data center has grown to several hundred, and the owners of data centers tend to buy multi-vendor devices in order to avoid lock-in with vendors and so to reduce costs.

2.2 Quick and Automatic Change in Network Configuration in Parallel with Virtual Machine (VM) Migration

To provide a cloud service, it is necessary to move virtual machines between data centers on an hourly basis in order to reduce the number of power-consuming sites at night, when the demand for services is reduced, or to reduce the effect of a serious disaster should it occur. Specifically, it is necessary to change the network configuration in an extremely short period of time using a protocol such as TCP/UDP in parallel with virtual machine migration without causing the suspension of the services being provided on the virtual machines.

As a solution to these two issues, cloud service providers are seeking a new standard network virtualization technology that does not rely on VLANs, and is not dependent on vendor-specific specifications.

3. OpenFlow and ONF

OpenFlow was started by Stanford University as a network control technology for a scientific purpose in 2008, and is now being studied at the OpenFlow Switching Consortium. Key features are as follows.

3.1 CD Separation

The basic concept of OpenFlow is that the controller that provides central control distributes programs to switches that comply with the OpenFlow specifications, and that these switches operate under the control of these programs (Fig. 1). In sharp contrast to the conventional Internet, where each network device autonomously exchanges routing information with other devices and autonomously selects its routing, the controller has a centralized controlling authority in OpenFlow, and switches operate under its control. This mechanism is referred to as CD (controller-data) separation.

3.2 Control Mechanism

Control in OpenFlow is specified by combinations of rules and actions. A rule specifies the packets to be processed. For example, it is possible to specify the criteria for the contents of the L2 to L4 headers with the incoming physical interface, such as specifying the packets whose TCP port number is 80. An action defines the action applied to the packets that match the associated rule. Typically, it can specify transfer

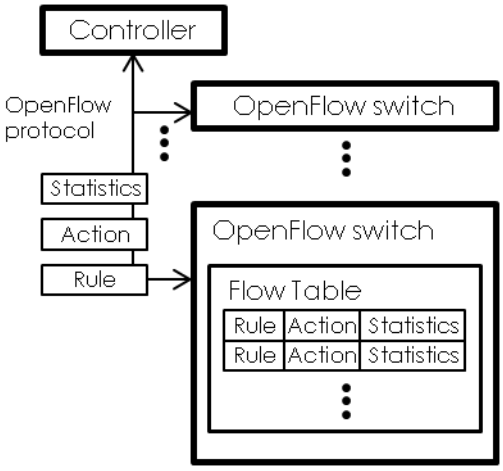

Fig. 1 Switch control scheme in the OpenFlow technology.

of the packets to another port, rewriting of the packet headers, discarding of the packets, etc. For example, it is possible to specify that those packets that arrive at a specific port number are discarded. In other words, simply by distributing an extremely simple program, the controller can make a switch function as a router, a firewall, or a load balancer depending on the intended purpose.

It is fair to say that the flexibility achieved by specifying the type of control with a program is one of the reasons why OpenFlow is attracting attention.

3.3 ONF (Open Network Foundation)

ONF is an OpenFlow promotion body launched by Google, Facebook, Yahoo, etc. in March 2011. NTT is its member.

An interesting characteristic of the ONF is that, rather than network vendors, what may be called network user enterprises, such as those that operate large data centers, serve as its board members.

The launch of ONF indicates that OpenFlow, which had been considered to be a scientific endeavor, will be applied for commercial purposes, which is the reason why OpenFlow is now attracting significant attention.

4. LISP and VXLAN

4.1 LISP

LISP is an IP-in-IP tunneling method proposed by Cisco. And we use router function instead of switch software because LISP can transport not L2 frame but IP packet. It is used globally, mainly for achieving IP mobility of VMs, multi-homing, and traversal of IPv4 by IPv6. Key features are as follows.

4.1.1 CD Separation

CD separation is achieved by a mechanism in which the Map Server, which manages network information, informs

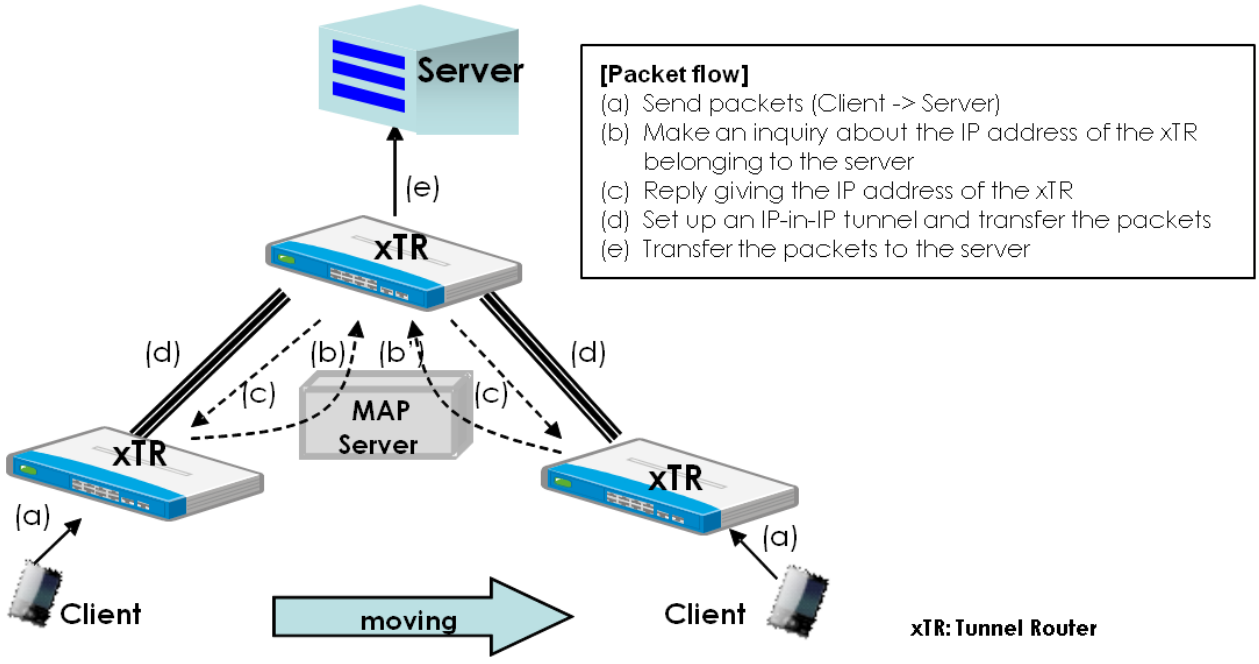

Fig. 2 Overview of the operation of LISP.

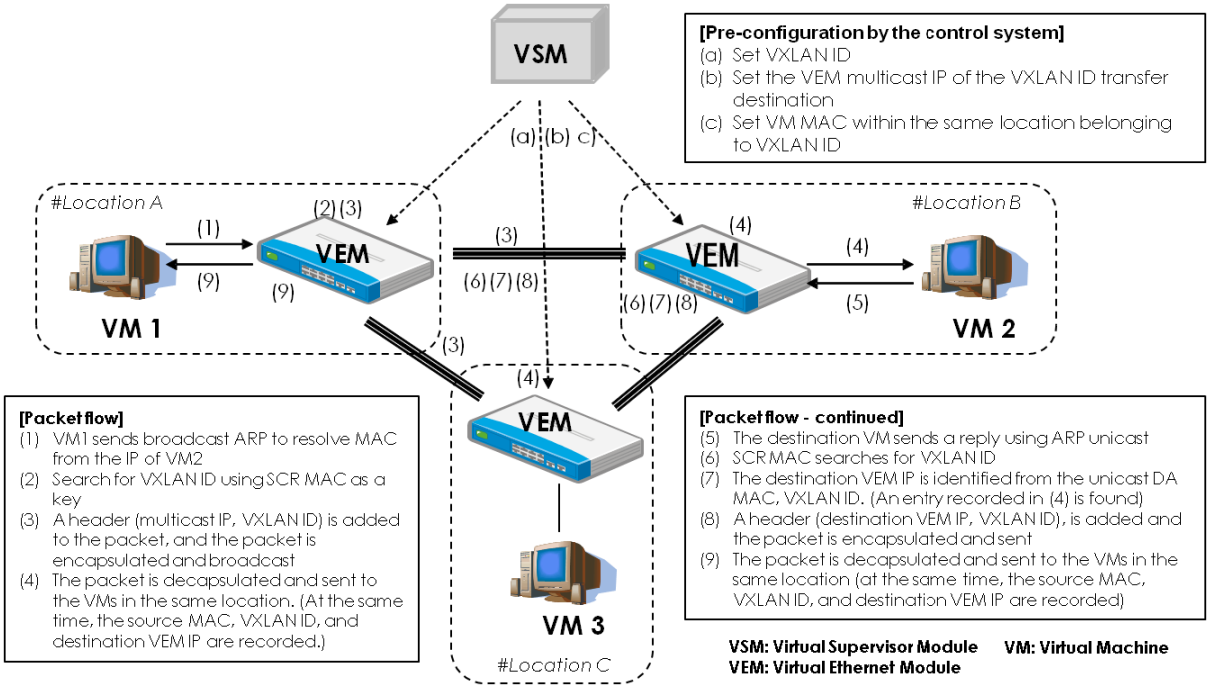

Fig. 3 Overview of the operation of VXLAN.

a tunnel router function (Tunnel Router: xTR) of the destination xTR (see Fig. 2). Since a xTR transports packets to another xTR by sending an inquiry to the Map Server, the Map Server has centralized authority over the control of the entire network, and xTRs operate under its control.

4.1.2 Mechanisms of L3-in-L3 Transfer and Control

Packets are transferred as follows. Packets are L3-in-L3 encapsulated by xTR that has a tunnel termination function, transferred to the destination xTR, decapsulated, and delivered to the end user. The Map Server manages the identifier (EID) and the destination (Locator). As shown in the

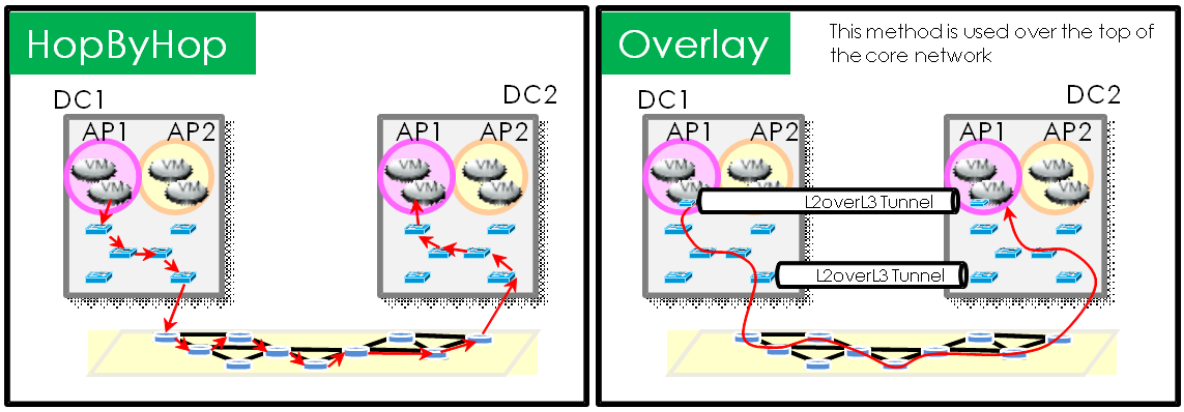

Fig. 4 Hop-By-Hop and Overlay methods.

Fig. 2, upon receiving a packet from a VM, the xTR sends an inquiry about its destination to the Map Server, and receives the packet transfer information from it. Since this mechanism does not have the ability to specify the required action, as does OpenFlow, it does not have as much flexibility as OpenFlow. For example, it cannot make a xTR as a load balancer. This L3-in-L3 encapsulation is suitable for achieving IP mobility of individual VMs, rather than achieving the accommodation of IP segments or the extension of IP segments to other sites.

4.2 VXLAN

VXLAN is an inter-switch software tunneling technology for network virtualization. A draft of VXLAN was submitted to IETF by Cisco, VMware, Citrix, Redhat, and Broadcom in August 2011. Key features are as follows.

4.2.1 CD Separation

Like Openflow, Cisco's virtual switch software called Nexus1000 V, for example, consists of two parts: a control system and multiple virtual switch software that have a tunnel termination function (Fig. 3). CD separation is achieved by the mechanism in which the control system (Virtual Supervisor Module: VSM) sends the transfer destination information to a virtual switch software (Virtual Ethernet Module: VEM).

4.2.2 L2-over-L3 Transfer and the Control Mechanism

The virtual switch software encapsulates an user L2 packet in a L3 packet, and sends them to the appropriate egress virtual switch software for the tenant network and for the destination MAC address (Fig. 3). Since a virtual switch software change the destination virtual switch software for each tenant network using a tenant network ID, the virtual network can be configured in a variety of ways, such as accommodating a virtual network for each tenant network, or extending a virtual network among datacenters. Unlike OpenFlow, no action can be specified.

5. Methods of Using OpenFlow

From the perspective of network virtualization, OpenFlow can be applied for routing in two ways: hop-by-hop and overlay. (Fig. 4).

5.1 Hop-by-Hop Method

The routing of the hop-by-hop method is such that the controller knows of all the switches and designs appropriate route for each service, and each switch passes packets to another switch under the instructions of the controller in a bucket-brigade manner until the packets reach the destination.

Although this method can fully exploit the features of OpenFlow, it has been pointed out that the method has a problem in scalability because each switch needs to hold all routing information. While the hop-by-hop method is suitable for building a small network, some ingenious measure, such as grouping routes, is needed if it is to be applied to a large network.

5.2 Overlay Method

In the overlay method, the controller does not control all routes. Instead, the communication ends cooperate with each other to control their route using a tunneling technology, which will be described later. As such, this method is referred to as edge networking.

The controller and the switches need only to know the source and destination of packets, and can reuse the conventional routing mechanism without any modification.

Therefore, the volume of routing information that needs to be managed remains within a practical range even with a large network. Unlike the hop-by-hop method, the overlay method can be applied to real services rapidly.

The overlay-type virtual network relies on L2-over-L3 tunneling, in which L2 frames for the user are encapsulated with L3 packets (Fig. 5). Its three technical features are described below.

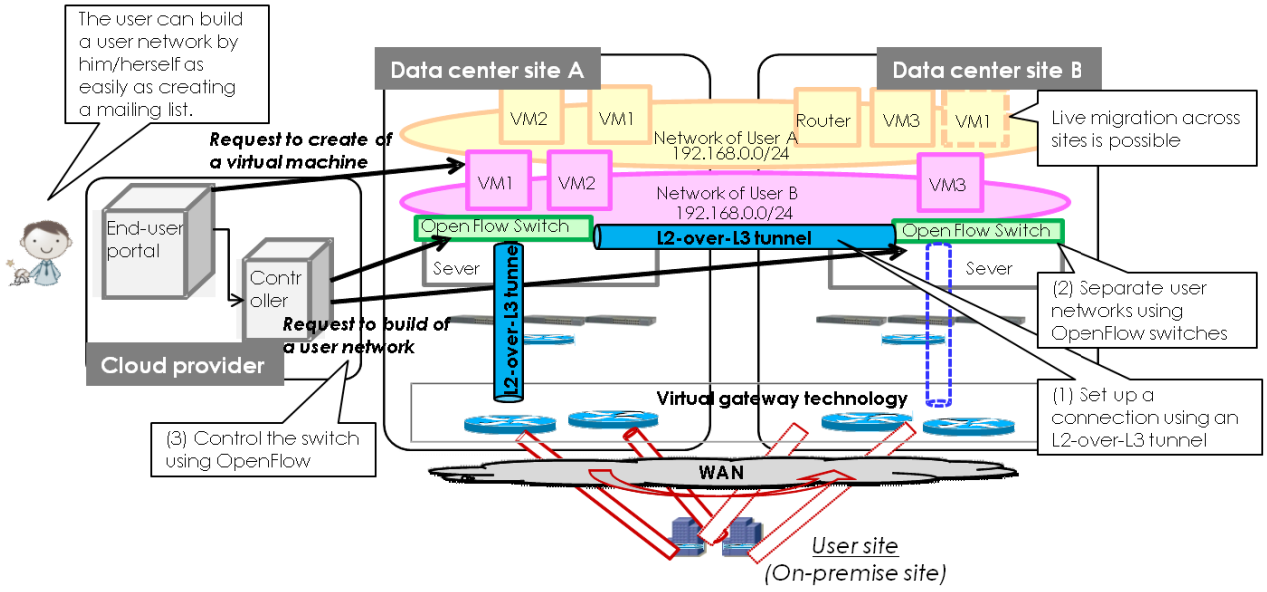

Fig. 5 Overview of the overlay-type virtual network.

5.2.1 L2-over-L3 Tunneling

In this technology, the OpenFlow switch within a hypervisor has an L2-over-L3 tunnel termination function, and hypervisors are interconnected with full-mesh tunnels. For on-premise environments, machines in user sites and hypervisors in datacenters are also interconnected by a virtual gateway, which is based on an OpenFlow switch having a tunnel termination function. Since this technology does not use any VLAN, it is free from VLAN-ID depletion and vendor-specific VLAN-ID design.

5.2.2 Separation of Virtual Networks for Individual Users

Separation of virtual networks for individual users is achieved by user identifiers. These identifiers are assigned by the tunnel termination function of the OpenFlow switches, and user packets are encapsulated to tunneling protocol with their identifiers. Physical switches and routers in the datacenter are only needed to keep IP reachabilty among hypervisors in order to bring encapsulated packets.

5.2.3 Standardized Switch Control Protocol

Since OpenFlow is used as a switch control protocol, a controller can be developed as an operation device capable of controlling multi-vendor products. The use of multi-vendor servers and switches can reduce their cost and the development of the operation device can reduce the operation and maintenance cost.

6. Interworking between an Overlay Virtual Network and an MPLS/VPLS Network

In the overlay-type network described in the previous section, a virtual network is built between virtual machines by interconnecting hypervisors with L2-over-L3 tunnels. This means that a network that interconnects the hypervisors can use any L3 networks. Therefore, as shown in Fig. 5, in a network configuration in which two data centers are interconnected by an arbitrary L3 wide-area network, it is possible to build a virtual L2 network that spans the two data centers, and to allow live migration of virtual machines between the data centers.

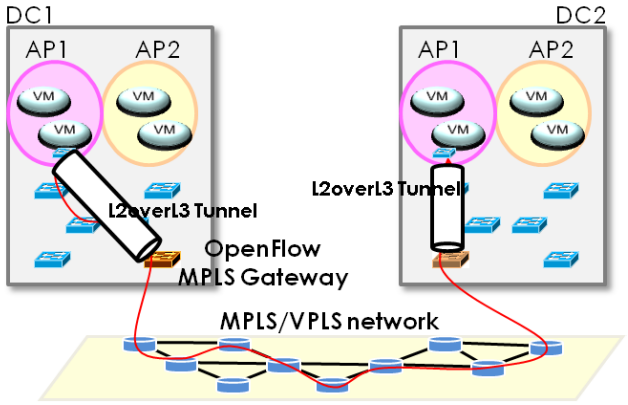

Fig. 6 Overlay with MPLS/VPLS.

From the perspective of the wide-area network, it is difficult to indentify individual packet flows because packets flowing between virtual machines are encapsulated by the OpenFlow switches located within the hypervisors, which are the edges of the network. This would be sufficient for services that simply require connectivity, but is insufficient for services that need to guarantee end-to-end SLA requirements or QoS between virtual machines.

So it is important to be able to build an overlay-type virtual network in each data center, and to make the gate-

way that connects a data center to the wide-area network take over the SLA and QoS information defined for each flow. For example, consider an MPLS VPN or a wide-area Ethernet (VPLS), which are often used to build enterprise networks, as shown in Fig. 6. The gateway terminates the L2-over-L3 tunnel within the data center, and replaces the MPLS label when it exchanges packets with the MPLS/VPLS network. By embedding QoS information in the MPLS label at this stage, it is possible to control QoS within the wide-area network for each flow.

7. Application Areas

This section introduces several application areas for overlay-type virtual networks.

7.1 DR/BCP

Since the East Japan Earthquake Disaster, there has been growing interest in disaster recovery (DR), and the business continuity plan (BCP) based on the construction of a backup office.

To achieve DR, it is necessary to copy data between remote sites, and to transfer virtual machines between these sites.

Conventionally, it has taken months to achieve the migration of virtual machines because it has been necessary to interconnect the two sites concerned with dedicated devices, and reconfigure all the networks between these sites.

The use of virtual networks has enabled an end user to achieve live migration of virtual machines to a remote site by him/herself within minutes. This makes for smooth disaster recovery.

NTT Laboratories have built a software-based L2-over-L3 logical network between the cloud environments in NTT Musashino R&D Center and in NTT Atsugi R&D Center using virtual network control technology, and have successfully achieved remote live migration, thereby confirming the feasibility of smooth disaster recovery.

7.2 Reduction in Power Consumption

Since the East Japan Earthquake Disaster, cloud services are attracting attention as means of reducing power consumption. The use of cloud services makes it possible to reduce the number of virtual servers and hence the number of physical servers, thereby reducing power consumption.

However, since the usage of virtual machines varies over time, the allocation of virtual servers to physical servers may not always be optimal if a virtual server is kept running on the same physical server.

By concentrating virtual machines onto fewer physical servers, depending on the usages of these virtual machines, it is possible to optimize the allocation of the virtual machines, and by suspending the operation of the vacant physical servers, to reduce power consumption. In addition, the ability to make virtual machines migrate to a remote site using a virtual network makes virtual machines operate flexibly. For example, it is possible to move virtual machines to the physical servers located in the area where there is spare power supply potential.

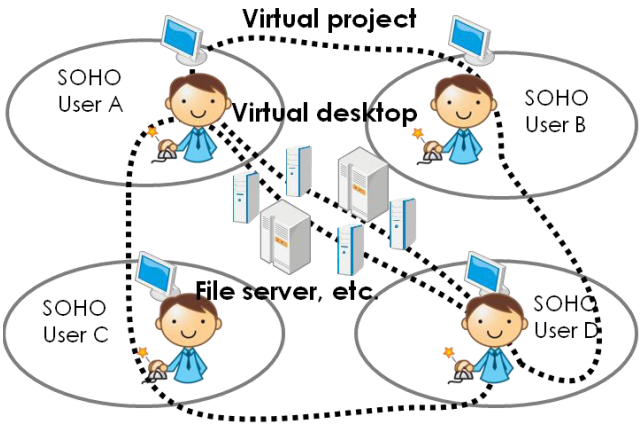

Fig. 7 Virtual desktop service.

7.3 DaaS

DaaS (Desktop as a Service) keeps the customer's desktop environment in a cloud, and thereby enables the customer to perform, on an inexpensive PC or smartphone, sophisticated operations that he or she is accustomed to. NTT Laboratories is studying a new service that combines the above feature with a virtual network.

For example, it will be possible not only to keep the desktop environment of each employee of group companies in a cloud but also to set up a logical network between the desktops of any of these employees on demand. When a project that spans employees of several organizations or group companies is launched, a shared space can be created easily for the project by installing a shared server or a chat server on such a logical network (Fig. 7).

Conventionally, it has been necessary to set up a virtual desktop and a VPN for each project, and end users have had to access the virtual desktop that has been specifically set up for a particular project. In contrast, when DaaS is provided in combination with a virtual network, all that is required is to set up virtual desktops for each user, and the end user can switch between the virtual desktops of different projects on demand.

8. Future Issues and Expected Futuer Development

If the current overlay-type network virtualization technology is to be introduced in an actual business scenario, it is necessary to address the following issues.

8.1 Management of the Correspondence between the Physical and Logical Networks

Since overlay-type network virtualization technology cre-

ates a logical network that is independent of physical networks, network management can become complicated. In particular, a mechanism to manage the correspondence between physical and logical networks will become necessary because the maintenance staff need to be able to determine quickly the effect on logical networks of a failure in the physical network.

8.2 Storage Migration

When a virtual machine is to be made to migrate to a remote data center to implement DR or reduce power consumption as mentioned in Sect. 8, it is necessary to pay attention to the performance of the new storage to which the storage used by this virtual machine migrates. In particular, when the transfer delay is big, as in the case where the migration takes place between very remote data centers, or when the packet loss of the link used is very large, it takes a long time for the migration to complete. Therefore, it is necessary to study the use of high-speed WAN technology, etc.

8.3 Coordinated Operation of Network Controllers

When virtual network technology based on CD separation is adopted, servers and switches within a data center may be divided into groups, or servers and switches may be grouped according to data center sites. Since there may be many virtual switches, or virtual switches may be distributed geographically, different controllers may manage them. In such cases, virtual networks are built on a mix of servers and switches in different groups, and thus the controllers need to operate in coordination.

8.4 Expected Future Development

The above sections have introduced network virtualization technology, which is necessary in a cloud environment. In particular, it has been mentioned that the overlay-type network virtualization can solve many network-related issues in the next-generation data center, and make it possible to provide network services in a data center to multiple tenant users quickly.

The following trends are noteworthy in considering the future development of the technology.

(1) Virtualization of network appliances

Besides switches and routers, so-called network appliances, such as load balancers and firewalls, are often used in data centers.

Creating virtual functions of these appliances and providing them in combination with virtual networks will make it possible to provide required network services in a data center flexibly and quickly. It is expected that the provision of simple L2 connectivity using the overlay-type virtual network will be followed by the commercial implementation of the technologies mentioned above.

(2) Coordinated operation with the cloud management system

The virtual machines required by users and the user-specific L2 networks required for connecting them can be provided flexibly and quickly by coordinating the operation of the cloud management system, which manages the creation, deletion and migration of virtual machines, with the operation of the network controllers, which control the virtual networks. It will become possible for the user to provide virtual machines on his or her own initiative by incorporating appropriate user portal functions.

One of the activities to bring about this coordination is OpenStack [3]. This is an open source community proposed by RackSpace, NASA, etc. It aims to make it possible to build an IaaS environment with typical PCs and network devices, and as part of this endeavor, has begun an activity to separate the network control function from the main environment and to extend this function. In particular, one of its projects, called Quantum, aims at standardizing the control and management of virtual L2 services using the overlay-type virtual network technology, and making this control and management possible through common APIs. Quantum adopts a plug-in structure to allow the user to incorporate any virtual network technologies freely. Such an arrangement will enable the user to select the preferred network virtualization method, and to control L2 networks using a common method.

9. Conclusions

This paper has described the current trends and issues in network virtualization technology, which is needed in a cloud environment. In particular, it has mentioned that the overlay-type virtual network technology is actively being implemented and has reached the level at which it can provide L2 connectivity between virtual machines. The paper has also provided some examples of expected applications.

In the future the authors plan to focus on how to secure connectivity between data centers.

References

[1] http://www.openflow.org/
[2] IETF draft, “VXLAN: A framework for overlaying virtualized layer 2 networks over layer 3 networks,” draft-mahalingam-dutt-dcops-vxlan-00.txt, Aug. 2011.
[3] http://www.openstack.org/

Hideo Kitazume Senior Research Engineer, Supervisor, Network Security Project, NTT Information Sharing Platform Laboratories. Hideo Kitazume received B.E. and M.E. degrees in computer science from Gunma University in 1987 and 1989, respectively. He joined NTT in 1989 and engaged in R&D of an ATM-LAN system, ATM traffic control studies, and the development of a global networking service platform. From 1998 to 2010, he worked in NTT EAST and engaged in the development, design, and operation of IP-VPN services. He is currently working on R&D of virtual networking technologies for cloud systems. He is a member of the Operations Research Society of Japan.

Takaaki Koyama Senior Research Engineer, Network Security Project, NTT Information Sharing Platform Laboratories. Takaaki Koyama received B.A. and M.M.G. degrees in media and governance from Keio University in 1994 and 1996, respectively. He joined NTT Software Laboratory in 1996 and has been studying software CALS. Since 1999, he has been studying GMN-CL, which is a kind of IP-VPN technology and developing some network equipment. Recently, his research interests have extended to enterprise cloud network systems. He is a member of the Information Processing Society of Japan.

Toshiharu Kishi Researcher, Secure Networking System Group, Network Security Project, NTT Information Sharing Platform Laboratories. Toshiharu Kishi received the B.E. and M.E. degrees in medical electronics from Chiba University in 2007 and 2009, respectively. He joined NTT Information Sharing Platform Laboratories in 2009 and worked on threat analysis of web applications. Since 2011, he has worked on enterprise cloud network systems and has been studying the architecture and construction of virtual networks in a cloud environment.

Tomoko Inoue Researcher, Network Security Project, NTT Information Sharing Platform Laboratories. Tomoko Inoue received a B.A. degree in literature from Ritsumeikan University, Kyoto, in 2003 and an M.A. degree in informatics from Kyoto University in 2005. She joined NTT WEST in 2005 and moved to NTT Information Sharing Platform Laboratories in 2011. Since 2011, she has been working on enterprise cloud network systems and is studying the architecture and construction of virtual networks in a cloud environment.