Layer 3 Cross-site VM Live Migration Scheme

draft-zeng-nvo3-l3-vm-migration-00.txt

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on December 12, 2014.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.
Abstract

This document utilizes software-defined network (SDN) to propose a cross-site live VM migration scheme in layer 3 environment. The architecture and main components are described in detail. The components include central controller, local controller, gateway and virtual switch, which are in charge of cross-site communication management, in-site network configuration, inter-site connection and VM address resolution respectively. This scheme can implement layer 3 cross-site VM migration without using complex mechanism.

Table of Contents

1. Introduction ................................................ 2
2. Conventions used in this document............................ 3
   2.1. Requirements Language................................... 3
   2.2. Definitions ............................................ 3
3. Scheme Architecture ........................................ 4
4. Component Detail ............................................ 5
   4.1. Central SDN Controller.................................. 5
   4.2. Local SDN Controller.................................... 6
   4.3. Gateway ................................................ 7
   4.4. Virtual Switch ........................................ 7
5. Migration Process ............................................ 7
6. Security Considerations...................................... 11
7. IANA Considerations ........................................ 11
8. Conclusions ................................................ 11
9. References ................................................ 11
   9.1. Normative References................................... 11
10. Acknowledgments ........................................... 11

1. Introduction

The support of Virtual Machine (VM) mobility is an important feature of nvo3. This document describes a VM live migration scheme which addresses several network related issues of this problem. The scheme mainly concerns VM live migration between different data centers, especially the ones with different subnets. As for live migration of VM, it is defined as the mobility of VM while retaining the VM’s addresses and communication status.

The scheme in this document utilizes SDN to control and configure data center network. SDN’s global network view and flexible forwarding control make it possible to quickly change routing and provide peer to peer forwarding which enables cross-subnet connection. These features help simplify the scheme.
2. Conventions used in this document

2.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

2.2. Definitions

live migration: VM can keep communication while migration.

flow: forwarding rules configured in switches.

host: physical machine which hosts VMs and virtual switch.
3. Scheme Architecture
The scheme architecture is shown in figure 1. VMs are interconnected by intra data center network, which is controlled by local SDN controller. All SDN controllers in different data center are connected with central SDN controller, which is in charge of cross-site VM migration. Cross-site migration process is invoked by central SDN controller. Central SDN controller stores VMs’ location and communication information, using this information, it can calculate which data centers are involved in the migration process and the job types each data center needs to handle, e.g. setting up migration path, changing routing. After calculation, central controller notices the involved local controllers to configure each intra data center network to complete the migration process.

4. Component Detail

4.1. Central SDN Controller

Central SDN controller is in charge of cross-site communication. It needs to store information like VMs’ location and communication pairs, and to have the ability to calculate and dispatch the configuration jobs.

Central SDN controller do not need to configure the network devices by itself, this work is done by local controllers. The central controller only needs to notice the involved local controller its job type and configuration parameters.

Job types include setting up migration path and changing routing. Setting up migration path refers to connect the source host and the destination host before copying the VM’s disk and memory, it involves intra and inter data center connection configuration. Changing routing means to calculate and configure the changed routing path caused by VM migration. This work can be conducted during the migration process to minimize the service down time.

Setting up migration path needs parameters like source host, source data center, VM IP and MAC addresses, destination host and destination data center. Central controller divides the connection configuration into three parts, source host to source data center gateway, source data center gateway to destination data center gateway and destination data center gateway to destination host and respectively notices the source and destination data centers’ local controllers to configure the switches and gateways.
Changing routing needs parameters like VM IP and MAC addresses and virtual network (VN) ID. Central controller needs to search its database to find VMs of the same VN belongs to which data centers and then notices the data centers’ local controller to calculate and modify the flows related to the migrating VM.

4.2. Local SDN Controller

Local SDN controller is in charge of the network configuration issues. Each local controller controls the physical network (include physical switches, virtual switches, gateway, etc.) of the data center it belongs to. Local controller should have at least two functions: network configuration and communication with central controller. Network configuration function mainly manages the forwarding logic of the data center, more specifically, it manages network virtualization, communication within VNs and communication between different data centers.

Network virtualization can be realized by setting VLAN, classifying MAC address, etc. The way to virtualize the network is not specified in this document, different data centers can even implement network virtualization in different ways as long as VMs in different data centers, but of the same VN can communicate with each other.

Communication within VNs function can be provided by classical forwarding methods, like MAC-learning, OSPF, etc. These methods should be implemented as controller modules which can automatically set up the route. Cross-site forwarding is like intra-site forwarding in most aspects, if the local controller identifies a certain communication as cross-site communication, it should forward the packets to gateway and notices the central controller to configure the relative data center’s local controller to set up the remaining path. After receiving the notice from central controller, the other local controller needs to forward the packets from gateway to its destination VM.

Communication between different data centers is implemented by inserting NAT flows to the gateways of the involved data centers. Gateways should have one or more public IPs, if a flow is identified as a cross-site flow, central controller assigns a source public IP, destination public IP and TCP port tuple as the flow’s ID. After that, the central controller notices the local controllers to insert NAT flows to the gateways. At the source gateway, the NAT flow replaces the packets’ source IP, MAC addresses with the gateway’s addresses, replaces the TCP port with the assigned port number, and replaces the packets’ destination IP, MAC addresses with the destination gateway’s addresses. At the destination gateway, the NAT flow does the reverse.
process and recovers the packets’ headers. If VM’s status is changed or the flow is no longer needed, central controller should notice the local controllers to delete the flows and release the TCP port.

4.3. Gateway

Gateway’s main function is to support inter data center communication and user to VM’s communication. Inter data center communication’s process is explained in 4.2. User to VM’s communication’s process is almost the same but only involves one gateway.

One data center can deploy one or more gateways to balance the load, or even one gateway per VN per data center to simplify the VN implementation. To support NAT flows, gateway should support packet header modification action of at least MAC, IP, and TCP fields.

4.4. Virtual Switch

Virtual switches are usually deployed at the host server in which the VMs nest. As the last hop of the whole network, virtual switches provide not only the basic forwarding function but also the address resolution function. After migration, VM is usually in a different subnet, which can impede the normal address resolution process. To avoid this, virtual switches need to identify all ARP packets and upload them to the local controller to process them. With the ARP packets, local controller can recognize the destination VM and set up the forwarding path accordingly. After that, it can reply the source VM with the destination VM’s MAC address and the source VM can start communication.

5. Migration Process

This section describes the timeline of the migration process. The whole migration process is divided into three parts: pre-migration, migration and post-migration.
Figure 2 Pre-migration Phase
Pre-migration Phase

Pre-migration phase has one main object, which is to set up the routing path for image copy. This object is fulfilled by three steps: 
First, central SDN controller receives the migration request, which contains the VM’s IP address and the destination host name, then the central controller queries the VM’s MAC address, source and destination data centers’ local controllers’ addresses. After the query, the central controller deploys the forwarding job to the involved local controllers. Second, the local controllers set up the path from host to gateway. And last, the local controllers configure the gateway to connect the two data centers. The NAT parameters are provided by the central controller within the job request.

![Diagram of Migration Phase]

Migration Phase

In migration phase, hypervisors conduct the VM migration process. During this time, the SDN controllers can calculate and configure the route changes caused by migration to minimize the VM’s service down...
time. In order not to disturb the ongoing communication, in this phase, only the flows which do not impact the service are added or updated. Other flows are updated after migration is completed.

---

**Figure 4 Post-migration Phase**

---

Zeng  Expires December 12, 2014  [Page 10]
This phase mainly does some clean work, including deleting disk copy flows, release NAT resources, etc. Besides these, some route changes need to be configured after the VM migration to avoid unnecessary service down (like ingress and egress flows).

6. Security Considerations

This requirements document does not raise in itself any specific security issues.

7. IANA Considerations

IANA does not need to take any action for this draft.

8. Conclusions

This document proposes a scheme for layer 3 cross-site VM live migration and describes its architecture, main components and process phases.

9. References

9.1. Normative References


10. Acknowledgments

This work is supported by Chinese National Major Scientific and Technological Specialized Project (No.-2013ZX03002001), National Basic Research Program of China (973 Program Grant No.-2013CB329105), China’s Next Generation Internet (No.-CNGI-12-03-007), and ZTE Corporation.

This document was prepared using 2-Word-v2.0.template.dot.
Authors’ Addresses

Lieguang Zeng
Department of Electronic Engineering, Tsinghua University
Department of Electronic Engineering, Tsinghua University, Beijing, China

Email: zenglg@mail.tsinghua.edu.cn

Xu Yang
Department of Electronic Engineering, Tsinghua University
Department of Electronic Engineering, Tsinghua University, Beijing, China

Email: yxpoke@sina.com

Mao Yang
Department of Electronic Engineering, Tsinghua University
Department of Electronic Engineering, Tsinghua University, Beijing, China

Email: yangmao210@163.com

Yong Li
Department of Electronic Engineering, Tsinghua University
Department of Electronic Engineering, Tsinghua University, Beijing, China

Email: liyong07@tsinghua.edu.cn

Depeng Jin
Department of Electronic Engineering, Tsinghua University
Department of Electronic Engineering, Tsinghua University, Beijing, China

Email: jindp@mail.tsinghua.edu.cn

Li Su
Department of Electronic Engineering, Tsinghua University
Department of Electronic Engineering, Tsinghua University, Beijing, China

Email: lisu@tsinghua.edu.cn