Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 3, 2012

Y. Gu
J. Yang
Huawei
C. Li
H. Xu
China Mobile
K. Li
Y. Fan
China Telecom
Z. Zhuo
M. Liu
Ruijie Network

October 31, 2011

State Migration
draft-gu-opsawg-policies-migration-01

Abstract

While Virtual Machine (VM) lively migrate around, not only the OS, memory, and the states on Hypervisor need to be migrated with VM, but also the states on the network side, e.g. on Firewall. Otherwise, the running services on the migrated VM could be disrupted, even stopped, In this draft, we describe the background and use cases of this proposal. We also raise a clear scope for the proposal.

Table of Contents

1. Introduction

VM live Migration enable us to migrate a VM from one place to another place without significant interruption to the running service on the VM. VMware lists some benefits that VMotion, VMware's VM live migration solution, can provide:

vMotion allows you to[VMotion]:

- Perform live migrations with zero downtime, undetectable to the user.
- Continuously and automatically optimize virtual machines within resource pools.
- Perform hardware maintenance without scheduling downtime and disrupting business operations.
- Proactively move virtual machines away from failing or underperforming servers.

VM Live Migration is a wonderful function to have for DC operators. However, some preconditions must be satisfied in order to make a successful live migration. One of the preconditions is that the flow-coupled state on network must be kept after VM migrates. A very obvious example of flow-coupled state is session table on Firewall. Assume that a VM migrates to a new place, which is under different Firewall from the original Firewall. If the session table, which records the existing connections to the VM, is lost, the following packets belonging to the existing connections will be dropped by the new Firewall, and the connections will finally be disconnected.

In the following sections, we will give more detail description of the problem with flow-coupled state in VM live migration. And we will conclude a feasible scope for further effort in IETF.

2. Terminologies and concepts

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 [RFC2119].

Source Network Device, Source switch, or Source device: the network device/switch/device from where the VM migrates. I.E. VM is originally located under the source network device/switch/device.

Destination Network Device, Destination switch, or Destination

device: the network device/switch/device to where the VM migrates. I.E. VM is relocated to the destination network device/switch/device.

Virtual Machine (VM), A completely isolated operation system which is installed by software on a normal operation system. An normal operation system can be virtualized into several VM.

Firewall (FW), A policy based security device, typically used for restricting access to/from specific devices and applications.

3. States On Firewalls

There are two kinds of physical Firewall deployment in DCs.

One is to place a pair of centralized powerful Firewalls at WAN connect point. In this case, any traffic, even the traffic between VMs within the same LAN, need to pass the Firewall.

The third way is distributed deployed Firewall. In stead of place a powerful centralized Firewall at the WAN connect point, Firewall is distributedly deployed at aggregation switches, even lower on access switches. The goal of this kind of distributed deployed Firewall is not to separate different security zones, but to off load the huge workload on centralized Firewall. This case is especially reasonable for large layer 2 network with tens even hundreds of thousands of Virtual Machines. To rely a centralized pair of Firewall to deal with traffic from such volume of VMs are not reliable and Firewall could be the bottleneck.

The following states are dynamically generated on Firewall.

3.1. Session Table

Firewall will establish session state for each connection to host within the DC. The host could a physical server or a VM. The session state includes most related information of the connection.

Item	Interpretation
Src IP	Source IP Address of the connection
Dst IP	Destination IP Address of the connection
Src Port	Srouce Port Number used to establish the session
Dst Port	Destination Port Number used to establish the session
Protocol	Protocol type
VLAN	VLAN ID

Expiration Time	The time that the session will be broken if no packet passes before that.

3.2. Cumulative Data

In order to protect DC from attacks, Firewall will cumulate various kinds of data. Assuming a use case, where there are both individual clients and enterprise servers in the DC. An untrust client might attack the servers in the same DC. One example of attacks is SYN Flooding. The client keeps sending SYN message to a specific server, which will be a DOS attack to the server, or to any server, which will become a DOS attack to the Firewall. Firewall cumulate the SYN message from a client. If the frequency of SYN message exceed a pre-defined rate, the IP address of this client will be drawn into a black list. Same situation to DNS Flooding attack.

4. Scenarios for Migration of States on Firewall

The following are scenarios that we need to migrate Firewall states with VM when proceed VM live migration.

4.1. VM Migration between different DCs

China Telecom deploys several separated DCs in one province in West China. These DCs have been built for several years and been upgraded during these years. But any single DC is limited in scale because most of the DCs are built in downtown. When facing with the requirements from large Service Provider, none of any single DC can accommodate the huge requirements for racks of servers by itself. So China Telecom has to split SP's requirements into multiple DCs. Interconnection between DCs must be provided to simulate a single DC, which is in order to enable inter-communication among the SP's VMs and to enable VM live migration.

Here we provide an example architecture of above situation. A DC provider has two DCs on different locations. One is at City A and the other is at City B, which is 30 kilometers away form City A. We assume that the physical distance and network bandwidth between City A and B satisfy the requirements of VM live migration. Two DCs are interconnected by VPLS to make them in the same LAN. Each DC has a pair of Firewalls on Core Switch. VRRP(Virtual Router Redundancy Protocol) [VRRP]is deployed on GW1 and GW1'.

At the very beginning, VMs are evenly created on Pod1 and Pod2. With time past, Pod1 and Pod2 becomes overloaded. In order to guarantee SLA, and to accommodate more service, Pod3 is created and some of the

VMs on Pod1 and Pod2 are migrated to Pod3, and the running service must be kept during the migration.

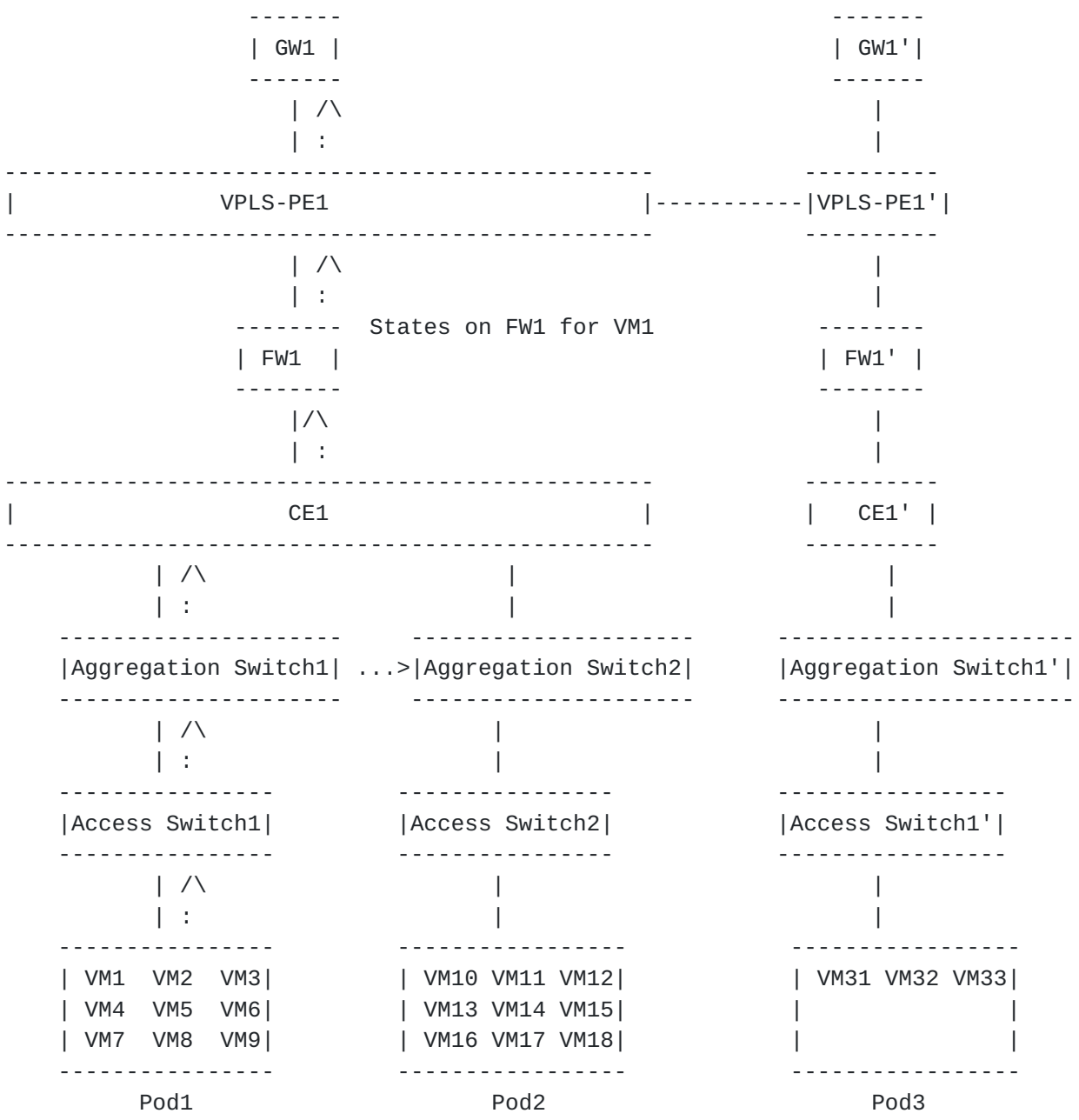
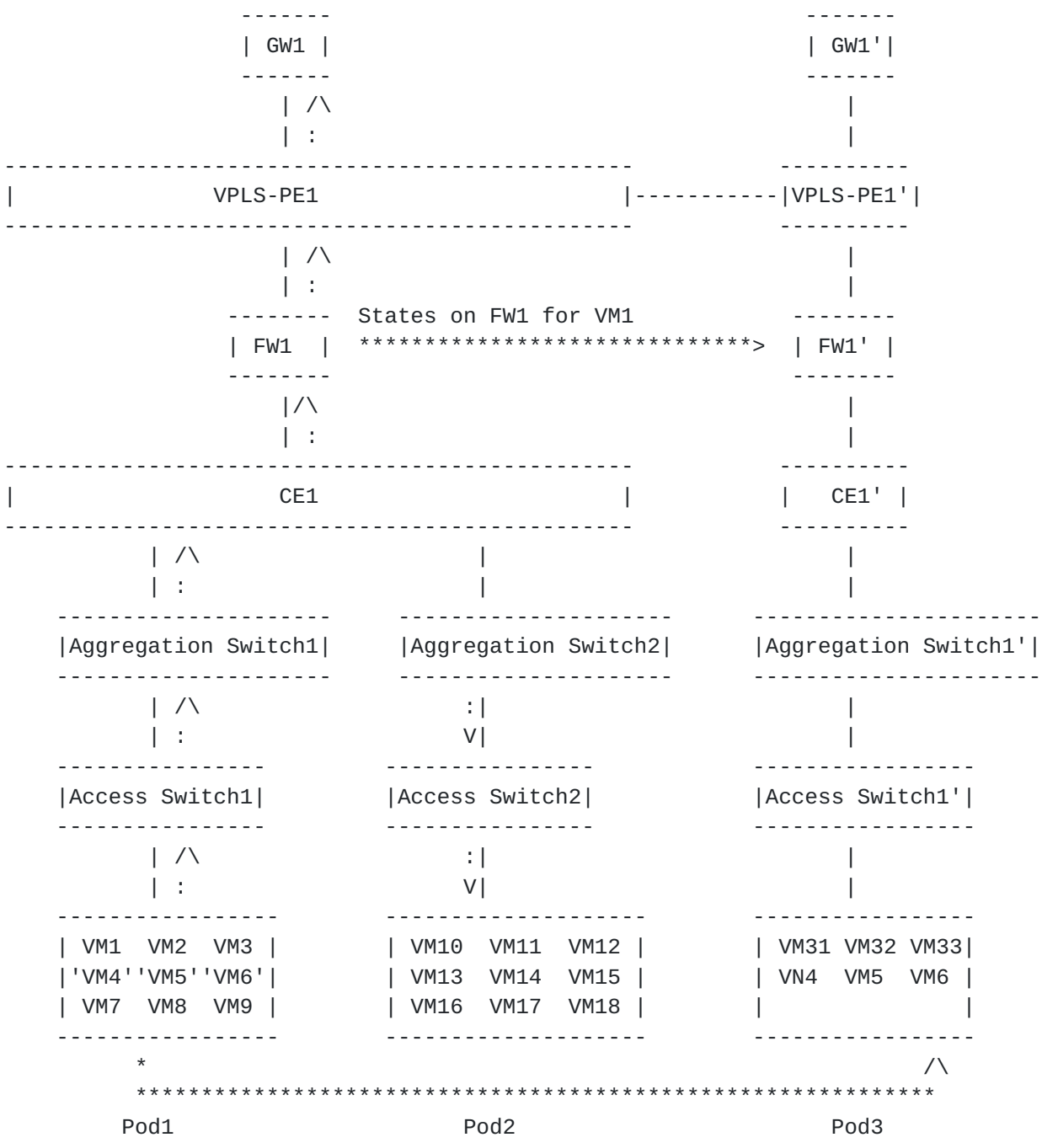


Figure 1: Example architecture

At payment day, a burst of access requests come to Finance Zone, the volume exceeds Server capability at Finance Zone 1. VM13 and some

other VM are migrated to Finance Zone 2 to utilize the idle resources in Finance Zone 2. The existing service on VM13 should be kept without disruption. So that the states on Firewall-2 that is related to VM13 should be migrated to Firewall-2'.



******** VM or States Migration
.... VM Traffic

Figure 2: VM and State Migration stage

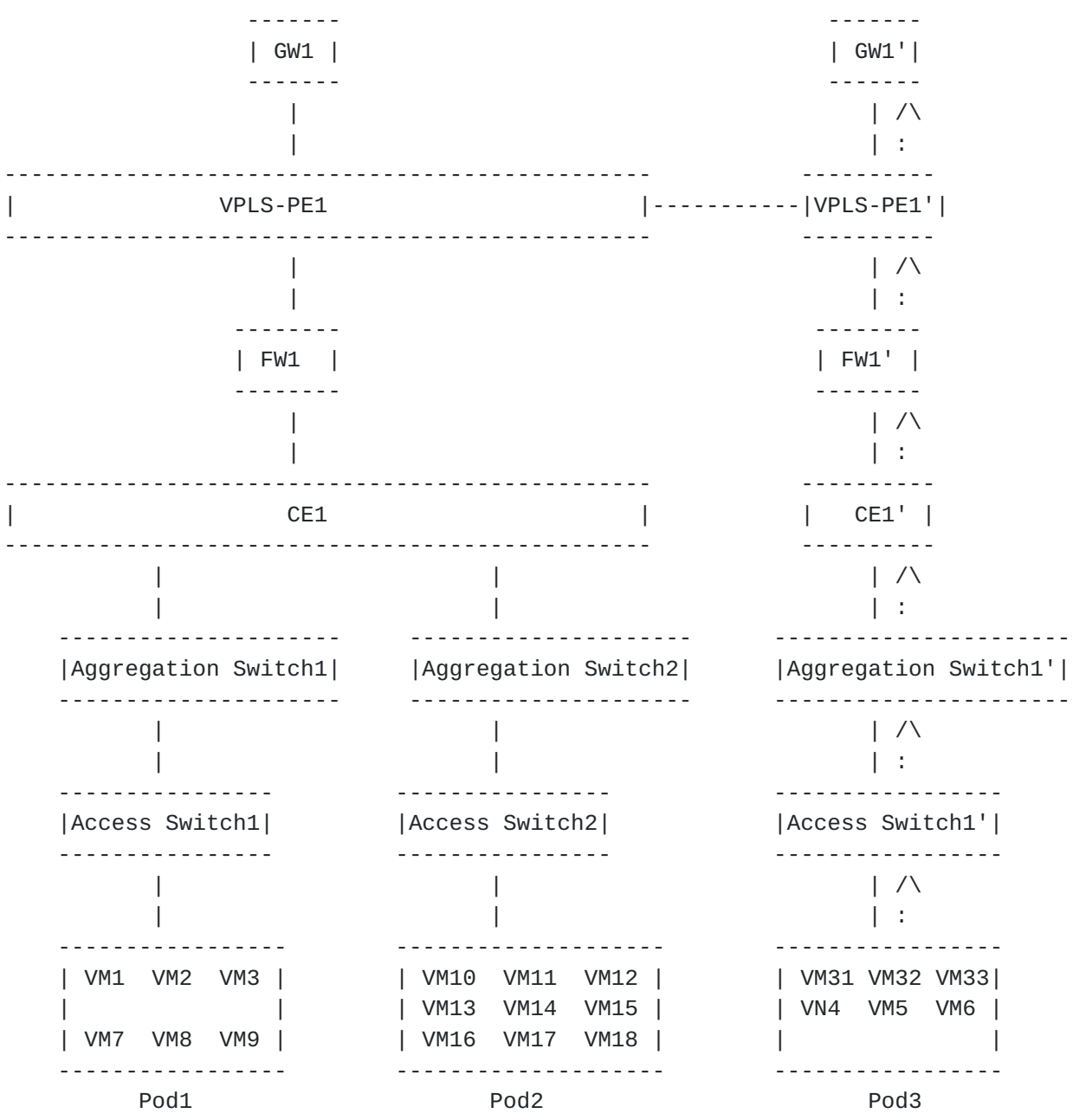


Figure 3: VM Migration Completion

4.2. VM Migration under Distributed Deployed Firewalls

In a DC with distributed deployed Firewalls on Aggregation Switches, an enterprise customer lease hundreds of physical servers, and each physical server carries 10 plus Virtual Machines (VM). The VMs provide VDI service to employees. At day time, the VMs are evenly deployed on each Pod3.

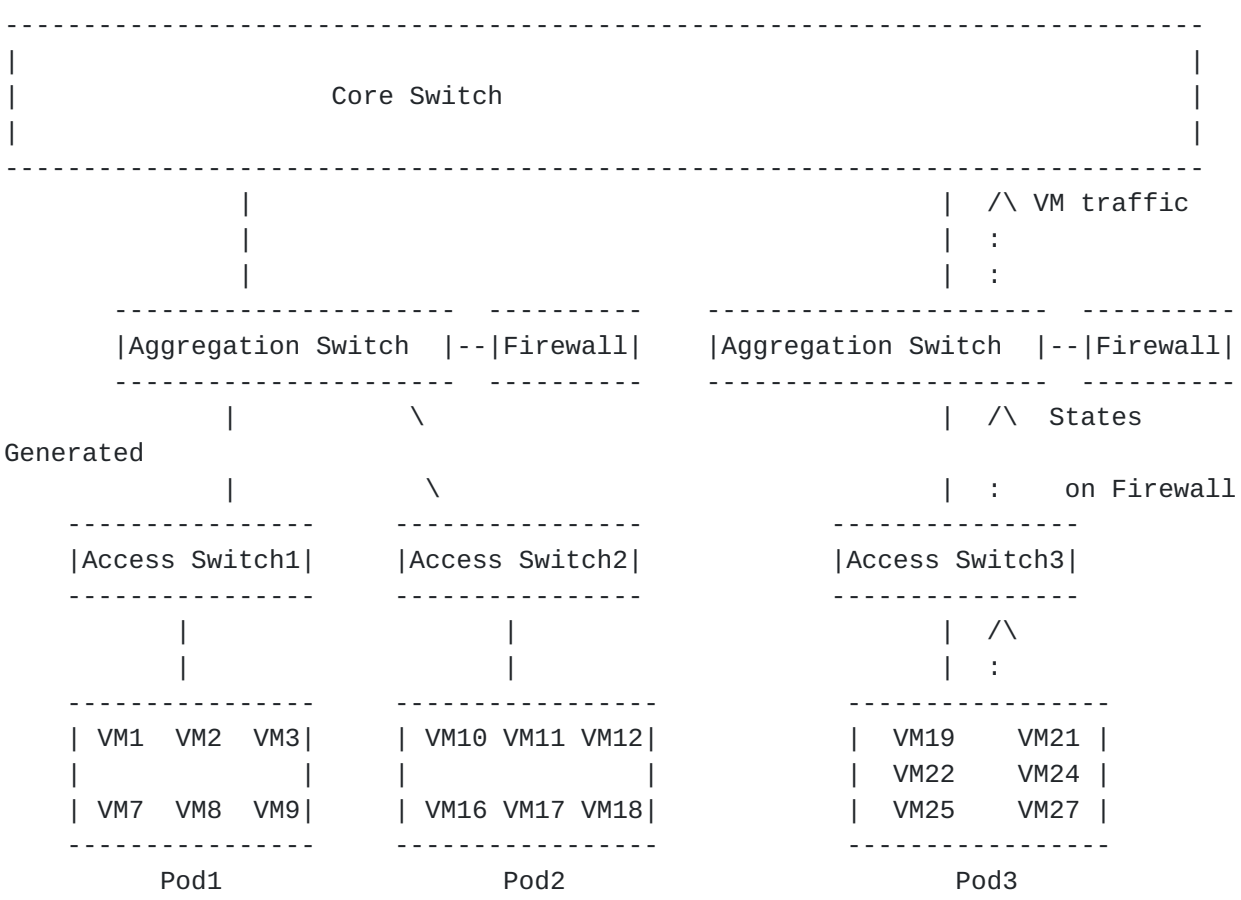


Figure 4: VDI service in DC

While at night, most of the VMs are shut down. Only a few VMs still working. In order to save energy, the active VMs are migrated to a few physical servers and the source physical servers, on which the migrated VM used to run, are shut down. The states on FW1' need to be migrated to FW1, otherwise the running service on migrating VM will be disrupted.

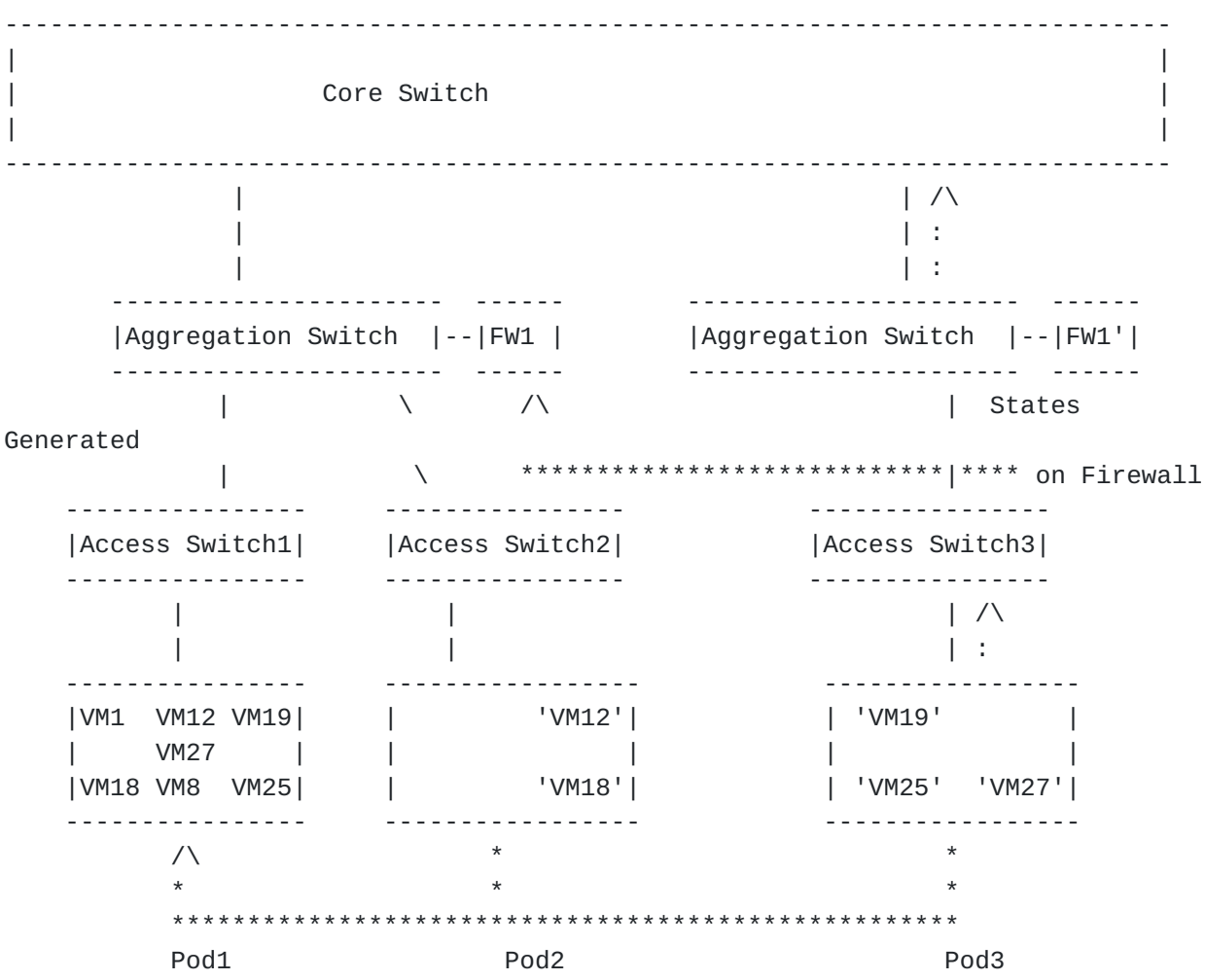


Figure 5: VM and State migration

5. Scope

SAMI (StAte MIgration) only considers the scenarios in which network conditions can satisfy the requirements raised by VM live migration. No matter VM is migrated within or between DCs, the scenario is in scope, as long as the network requirements for VM live migration can be satisfied. VM migration between L3 subnet, for now, is not in the scope. The solutions we develop in SAMI should enable both state migration within DC and between DCs, which is logically in the same Layer 2 network.

For the first stage, we only migrate Session tables on Firewall. But

the solution should be extensible to enable migration of other states

we may find that is necessary to be migrated during VM live migration.

We should always try to reuse existing IETF work to resolve SAMI problem. Only when there is no existing IETF work can use, with suitable extension, to achieve State migration, shall we develop a new mechanism to do this.

6. Security Considerations

The states described above are all about security. Besides, we need to be careful to avoid poisoned states from untrusted source. That means no matter how the states are migrated, authentication and verification are required.

7. Acknowledgments

The authors would like to thank the following people for contributing to this draft: Ning Zong, David harrington, Linda dunbar, Susan Hares, Serge manning, Barry Leiba, Jiang xingfeng, Song Wei, Robert Sultan.

8. References

8.1. Normative Reference

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", March 1997.

[RFC3303] Srisuresh, P., Kuthan, J., Rosenberg, J., Molitor, A., and A. Rayhan, "Middlebox communication architecture and framework", August 2002.

[RFC4761] Kompella, K. and Y. Rekhter, "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", Jan. 2007.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", Jan. 2007.

[RFC4861] "Neighbor Discovery for IP version 6 (IPv6)", Sep. 2007.

8.2. Informative Reference

[Data_Center_Fundamentals]
"Data Center Fundamentals", 2003.

[I-D.wang-opsawg-policy-migration-gap-analysis]
Wang, D. and Y. Gu, "I-D.wang-opsawg-policy-migration-gap-analysis", 2011.

[Vmotion_between_DCs]
VMware, "VMotion between Data Centers--a VMware and Cisco Proof of Concept, (http://http://blogs.vmware.com/networking/2009/06/vmotion-between-data-centersa-vmware-and-cisco-proof-of-concept.html)", June 2009.

[OTV] Grover, H., Rao, D., and D. Farinacci, "Overlay Transport Virtualization", July 2011.

[Amazon_VPC_User_Guide]
"http://docs.amazonwebservices.com/AmazonVPC/2011-07-15/UserGuide".

[VMotion] "http://www.vmware.com/products/vmotion/overview.html".

[LISP] "Location/ID separation protocol, http://tools.ietf.org/wg/lisp/".

[VRRP] "http://en.wikipedia.org/wiki/Virtual_Router_Redundancy_Protocol".

Authors' Addresses

Gu Yingjie
Huawei
No. 101 Software Avenue
Nanjing, Jiangsu Province 210001
P.R.China

Email: guyingjie@huawei.com

Yang Jingtao
Huawei
No. 101 Software Avenue
Nanjing, Jiangsu Province 210001
P.R.China

Email: yangjingtao@huawei.com

Li Chen
China Mobile

Email: lichenyj@chinamobile.com

Xu Huiyang
China Mobile

Email: xuhuiyang@chinamobile.com

Li Kai
China Telecom

Email: leekai@ctbri.com.cn

Fan Yongbing
China Telecom
No. 109 Zhongshan Road West
Guangzhou, Guangdong Province
P.R.China

Phone: 86-20-38639121
Fax: 86-20-38639487
Email: fanyb@gsta.com

Zhuo Zhiqiang
Ruijie Network

Email: zhuozq@ruijie.com.cn

Liu Ming
Ruijie Network

Email: lium@ruijie.com.cn