Network Working Group
Internet Draft
Intended status: Informational

Vic Liu
China Mobile
Bob Mandeville
Iometrix
Brooks Hickman
Spirent Communications
Weiguo Hao
Huawei Technologies
Zu Qiang
Ericsson

Expires: January 2015

July 3, 2014

Problem Statement for VxLAN Performance Test
draft-liu-nvo3-ps-vxlan-perfomance-00.txt

Status of this Memo

This Internet-Draft will expire on January 3,2015.

Copyright Notice

Abstract

As the number of data center tenant increased, 4K VLANs, mobility, broadcasting issues have become the network bottleneck. VxLAN has being take into consideration in China Mobile IDC.

There are two implementation solutions for VXLAN. The first one is that NVE resides in TOR (top of rack switch), another one is that NVE resides in V-Switch established in hypervisor of physical server. For virtualized network, it's much better to implement NVE in V-switch because it's directly connect with Virtual Machines and easier for business to carry on. As we research in VxLAN solution, some problems are take into our considerations. How much resources will be consumed by VxLAN in a virtualized network environment. This draft introduces the problem for VxLAN performance by test.

There is no methodology which can effectively evaluate VxLAN forwarding performance. This draft also attempts to address this issue, give a VXLAN performance evaluation method, especially when VxLAN resides in the virtual switch.

Table of Contents

1. Introduction 3
2. Consideration on VxLAN performanc 4
2.1. Test methodology for VxLAN performance in virtual network 4
2.2. Large-scale VxLAN test issues 4
2.3. Key index in VxLAN performance 5
2.4. Test Bed Setup 5
2.5. Benchmark test on virtualized network 8
3. Problem statement on VxLAN performance 9
3.1. VxLAN performance on test bed 9
3.2. VxLAN Scalable test issues 11
4. Security Considerations 11
5. IANA Considerations 11
6. References 11
6.1. Normative References 11
6.2. Informative References 11
7. Acknowledgments 12

1. Introduction

As the number of data center tenant increased, 4K VLANs, mobility, broadcasting issues have become the network bottleneck. VxLAN has being take into consideration in China Mobile IDC.

There are two implementation solutions for VXLAN. The first one is that NVE resides in TOR (top of rack switch), another one is that NVE resides in V-Switch established in hypervisor of physical server. For virtualized network, it's much better to implement NVE in V-switch because it's directly connect with Virtual Machines and

easier for business to carry on. As we research in VxLAN solution, some problems are take into our considerations. How much resources will be consumed by VxLAN in a virtualized network environment. This draft introduces the problem for VxLAN performance by test.

There is no methodology which can effectively evaluate VxLAN forwarding performance. This draft also attempts to address this issue, give a VXLAN performance evaluation method, especially when VxLAN resides in the virtual switch.

2. Consideration on VxLAN performance

While we testing performance on virtualized network, some issues and key index should be considered clearly.

2.1. Test methodology for VxLAN performance in virtual network

It's different from test for physical switch. Because firstly in virtual network, the DUT (VxLAN on V-switch), hypervisor and virtual test center (it's a VM) is all in one physical server. Secondly, it's not like RFC 2544 that the test center generate line rate traffic(usually 1G or 10G) and test the physical server's performance. As we generate traffic from one server to another (model A below), it has a fold point during traffic increase from 1G to 10G because the vCPU is overloading. For example, server A generate 1G traffic and server B can receive 100%, but server A generate 10G traffic and server B can only receive 530Mb traffic.

So in this test, the test process is designed as follows:

a) Firstly use the server to connect with a physical test center.

b) Make a traffic benchmark of 128, 256, 512, 1024, 1518bytes.

c) Setup the test bed this the benchmark to get performance without VxLAN.

d) Setup VxLAN and running the same performance test.

2.2. Large-scale VxLAN test issues

When test the scale of VLANs, it can be simulate 4K VLAN on the test center. But, in virtual network, the virtual center is a virtual machine. And virtual machine can only establish one VxLAN with VSwitch. So it can't test the large scale VxLAN performance.

2.3. Key index in VxLAN performance

a) CPU: CPU utilization is very important for VxLAN. However, vCPU can be allocated for VM. But it cannot allocated for hypervisor and VSwitch. We use the test methodology to evaluate the CPU performance for VxLAN.

b) Memory: Memory is not sensitive from the test result. But we still think it should be listed as one VxLAN performance index.

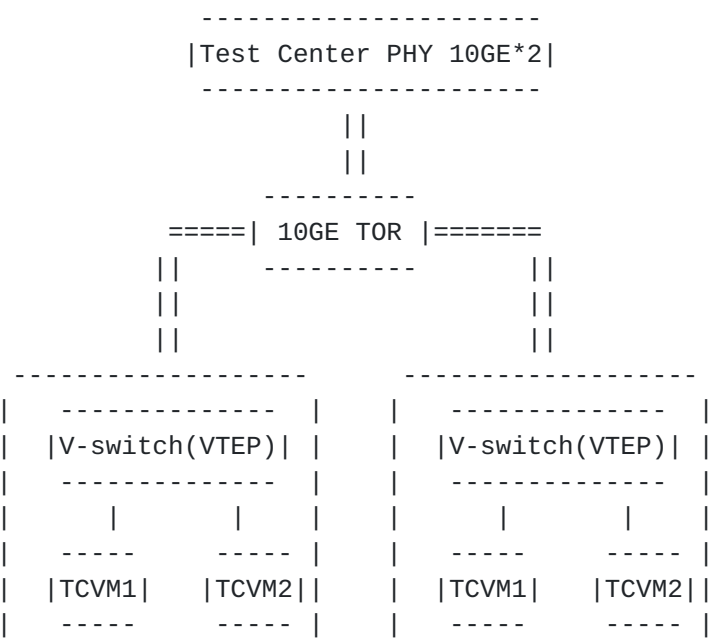
c) Latency: When traffic is forwarded between VM to VM across two different physical server. Latency should be an index.

d) Throughput: We use the benchmark as the traffic throughput.

e) Packet-lost: Virtual network may have few packet-lost because of unstable of vCPU. Less than 2% of packet-lost is acceptable.

2.4. Test Bed Setup

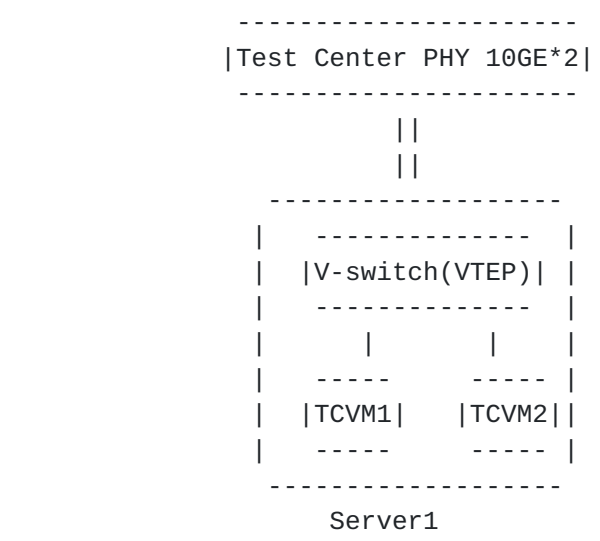
The test bed includes two physical server with 10GE NIC, a test center, a 10GE TOR switch for test traffic and a 1GE TOR switch for management.



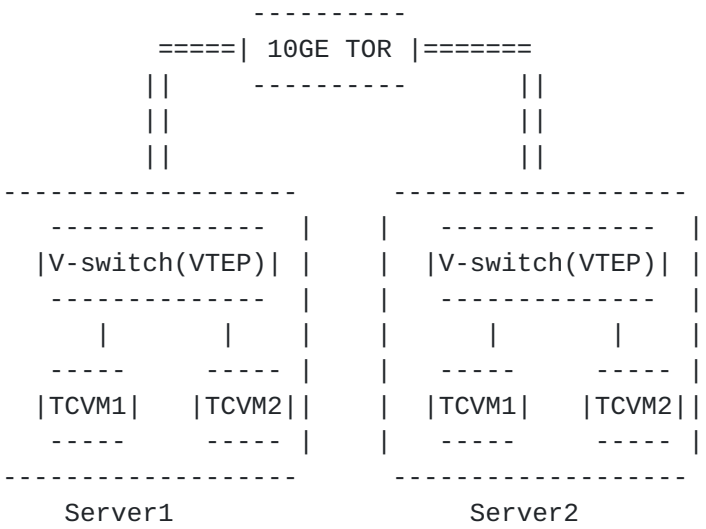

Server1
Server2

Two Dell server are R710XD (CPU: E5-2460) and R710 (CPU: E5-2430) with a pair of 10GE NIC. And in the server we allocate 2 vCPU and 8G memory to each Test Center Virtual Machine (TCVM).

In traffic model A: We use a physical test center connect to each server to verify the benchmark of each server.



In traffic model B: We use the traffic model A benchmark to test the performance of VxLAN.



2.5. Benchmark test on virtualized network

The reason we need a benchmark test is we realized that the virtualized network is different from physical network device. We cannot use test methodology like RFC 2544. The performance is not linear growth with traffic we generate. It has an inflection point.

To get the benchmark, we use traffic model A and get the result table below:

Server 1: CPU E5-2430

Byte	Rate(GE)	Server CPU MHZ	Server Mem	VM CPU	VM Mem
0	0	505	3022	372	695
128	0.46	6085	3021	5836	695
256	0.84	6365	3021	6143	696
512	1.56	6330	3021	6099	696
1024	2.88	5922	3021	5726	696
1518	4.00	5713	3023	5441	696

Server 2: CPU E5-2620

Byte	Rate(GE)	Server CPU MHZ	Server Mem	VM CPU	VM Mem
0	0	505	2900	239	698
128	0.61	5631	2900	5117	698
256	0.94	5726	2896	5157	698
512	2.02	5786	2901	5217	698
1024	4.02	5884	2901	5097	698
1518	5.61	5856	2901	5197	698

Byte	Rate(GE)	Server CPU MHZ	Server Mem	VM CPU	VM Mem
0	0	515	3042	374	696
128	0.46	6395	3040	5748	696
256	0.84	6517	3042	5923	696
512	1.56	6668	3041	5857	696
1024	2.88	6280	3043	5506	696
1450	4.00	6233	3045	5309	696

As we get the benchmark, we use the lower benchmark (server1) in traffic model B and test VxLAN performance.

3. Problem statement on VxLAN performance

3.1. VxLAN performance on test bed

We use the lower benchmark (server 1) to generate traffic from server 1 to server 2 with VxLAN encapsulation and get the performance result of the two servers. And because of VxLAN encapsulation increases the packet length, to avoid MTU problem we use 1450 instead 1518 as original packet length.

Server 1 with VxLAN: CPU E5-2430

Byte	Rate(GE)	Server CPU MHZ	Server Mem	VM CPU	VM Mem
0	0	515	3042	374	696
128	0.46	6395	3040	5748	696
256	0.84	6517	3042	5923	696
512	1.56	6668	3041	5857	696
1024	2.88	6280	3043	5506	696
1450	4.00	6233	3045	5309	696

Server 2: CPU E5-2620

Byte	Rate(GE)	Server CPU MHZ	Server Mem	VM CPU	VM Mem
0	0	450	2905	239	698
128	0.46	6203	2905	5897	698
256	0.84	5937	2906	5797	698
512	1.56	5993	2909	5737	698
1024	2.88	5710	2912	5697	698

1450	4.00	5863	2902	5697	698	

By analyzing the testing result, we have conclusion as follows:

a) CPU: VxLAN function resided in VSwitch increases physical CPU usage. The table below shows the increasing percentage of CPU usage after using one VxLAN ID. The average increase is 6.51% in server 1 and 4.07% in server 2. This increase is cost by one VxLAN. We will still evaluate increase in large scale VxLAN scenario.

Byte	Server 1	Server 2
0	4.04%	24.65%
128	7.57%	7.26%
256	1.15%	7.59%
512	8.66%	2.41%
1024	4.49%	0.14%
1450	11.61%	1.84%
*AVG	6.51%	4.07%

b) Memory: Memory of both physical server and virtual machine are not sensitive during VxLAN test.

c) Packet-loss: Because virtual network is based on X86 architecture. When vCPU utilizing rate reaches over 90%, there will be about 2% packet-loss. It is different with VXLAN on physical switch that no pack-lost forwarding is a necessary requirement.

d) Line-rate forwarding: It is well known to us, in virtual network, traffic become unstable as CPU goes overload. Whatever we try, we can't reach line rate using any packet length. Finally, we reach 10Gb using 1518 byte without VxLAN between two server by add 3 pair of TCVM and each TCVM allocated 2 vCPU and 8G memory. While we add VxLAN and decrease 1518 to 1450(in case of fragment), on same network, we can only get 5.6 Gb unstable throughput.

3.2. VxLAN Scalable test issues

All the tests above are based on one VN. As we considering Multi-VN scenario. One problem we can't overlook is, the VSwitch can only recognize (or study) the VN ID from VNIC of VM (TCVM). As we generate thousands of VxLAN by the TCVM, none can be studied by VSwitch except the VxLAN to VNIC. We calculate, one VM can provide 10 VNIC (MAX) which allocate 10 VxLAN, and one physical server install 20 VM. If we make a 5000VxLAN scale performance test, there will be at least 25 server.

4. Security Considerations

5. IANA Considerations

6. References

6.1. Normative References

[1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[2] Crocker, D. and Overell, P.(Editors), "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, Internet Mail Consortium and Demon Internet Ltd., November 1997.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2234] Crocker, D. and Overell, P.(Editors), "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, Internet Mail Consortium and Demon Internet Ltd., November 1997.

6.2. Informative References

[3] Faber, T., Touch, J. and W. Yue, "The TIME-WAIT state in TCP and Its Effect on Busy Servers", Proc. Infocom 1999 pp. 1573-1583.

[Fab1999] Faber, T., Touch, J. and W. Yue, "The TIME-WAIT state in TCP and Its Effect on Busy Servers", Proc. Infocom 1999 pp. 1573-1583.

7. Acknowledgments

<Add any acknowledgements>

Authors' Addresses

Vic Liu
China Mobile
32 Xuanwumen West Ave, Beijing, China

Email: liuzhiheng@chinamobile.com

Bob Mandeville
Iometrix
3600 Fillmore Street
Suite 409
San Francisco, CA 94123
USA
bob@iometrix.com

Brooks Hickman
Spirent Communications
1325 Borregas Ave
Sunnyvale, CA 94089
USA
Brooks.Hickman@spirent.com

Weiguo Hao
Huawei Technologies
101 Software Avenue, Nanjing 210012, China

Email: haoweiguo@huawei.com

Zu Qiang
Ericsson
8400, boul. Decarie
Ville Mont-Royal, QC,
Canada

Email: Zu.Qiang@Ericsson.com