Network Working Group   Vic Liu
Internet Draft   Dapeng Liu
Intended status: Informational   China Mobile
                                    Bob Mandeville
        Iometrix   Brooks Hickman
                                            Spirent Communications
                                                     Guang Zhang
                                        IXIA

Benchmarking Methodology for Virtualization Network Performance

                draft-liu-bmwg-virtual-network-benchmark-00.txt

Status of this Memo

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

<Lastname>   Expires January 4, 2015   [Page 1]

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 January4,2009.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. 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.

Abstract As the virtual network have been wide establish in IDC. The performance of virtual network has become a consideration to the IDC managers. This draft introduce a benchmarking methodology for virtualization network performance.

Table of Contents

1. Introduction   As the virtual network have been wide establish in IDC. The   performance of virtual network has become a consideration to the IDC   managers. This draft introduce a benchmarking methodology for   virtualization network performance.

2. Peculiar issues   In a conventional test setup with real test ports, it is quite   legitimate to assume test ports provide the golden standard and in   measuring the performance metrics. If and when test results are sub   optimal, it is automatically assumed it's the Device-Under-Test

(DUT) that is at fault. For example, when testing the max no-drop throughput at a given frame size, if the test result shows less than 100% throughput, we can safely conclude that it's the DUT that can't deliver line rate forwarding at that frame size(s). We never doubt that tester will be an issue. In a virtual test environment where both the DUT as well as the test tool itself are VM based, it's quite a different story. Just like the DUT, tester in VM shape will have its own performance peak under various conditions. Even worse, conditions are multitude and in many forms. Tester's calibration without DUT is essential in benchmarking testing in a virtual environment. Furthermore, to reduce the enormous combination of various conditions, tester must be calibrated with the exact same combination and parameter set the   user wants to measure against a real DUT. A slight variation of   conditions and parameter value will cause inaccurate measurements of   the DUT.   While the exact combination and parameter set are hard to list,   below table attempts to give a most common example how to calibrate a tester before testing a real DUT under the same condition.   Sample calibration permutation

----

Type	Hypervisor Speed	VM VNIC CPU Allocation	VM Memory Size	Packet Throughput	No Drop
				ESXi	1G/10G
512M/1Core	64				
128					256
				512	
			1024		Vic Liu

                                    | 1518 | | --------------------
            --   Key points are as following: a)
The hypervisor type is of ultimate importance to the test results.   Tester V
M must be installed on the same hypervisor type as the   real DUT. When feasi
ble, tester VM software should be installed on   a separate, but identical ty
pe of, hypervisor.   b)The VNIC speed will have impact on test results. Tester m
ust   calibrate against all VNIC speeds to be tested against a real DUT.   c)
VM allocation of CPU resources and memory will affect test results   d)Packet si
zes will affect test results dramatically due to the   nature of virtual mach
ine   e)Other possible extensions of above table: The number of VMs to be   c
reated, latency and/or jitter readings, one VNIC per VM vs.   multiple VM sha
ring one VNIC, and uni-directional traffic vs. bi-   directional traffic.   I
t's important to have a test environment for tester's calibration   as close as
possible to when a real DUT will be involved for the   benchmark test. Above tes
t setup illustrate below key points: 1.One or more tester's VM need to be crea
ted for both traffic   generation and analysis   2.vSwitch is need to accommo
date performance penalty due to extra VM   addition   3.VNIC and its type is
needed in the test setup to once again   accommodate any performance penalty
when real DUT VM is created   4.ToR switch is needed to accommodate delays intro
duced by the   external device   In summary, calibration should be done in su
ch an environment that   other than the DUT VM, all possible factors that may ne
gatively   impact test results should be accommodated.Vic Liu   Ex
pires January4,2015   [Page 5]

3. Key Performance Index   We listed numbers of key performance index for virtual network as   follows: a) No drop throughput under various frame sizes: Forwarding performance under various frame sizes is a key performance index   of interest. Once this performance number is obtained, vendors can   always allocate more CPU and memory for mission critical   applications where line rate performance is expected.   b) DUT consumption of CPU and memory: when adding one or more VM.   With addition of each VM, DUT will consume more CPU and memory.   c) Latency readings: Some applications are highly sensitive on   latency.It's important to get the latency reading with   respective to various conditions.   VxLAN performance can be looked at from two perspectives. First,   addition of VxLAN on an existing VM will consume extra CPU resources   and memory. This can be easily included in the benchmark table.   Tester VM are strictly traffic generator and analyzer. No   calibration is needed when adding VxLAN to DUT VM.   Once basic performance metric is obtained with respective to single   VxLAN, we need to look at performance metrics with many VM and   VxLAN. The idea is verify how many VM/VxLAN can be created and what   their forwarding performance (no drop throughput), latency, and   CPU/memory consumptions are.

4. Test Setup The test bed is constituted by two physical server with 10GE NIC, a   test center, a 10GE TOR switch for test traffic and a 1GE TOR switch for management.
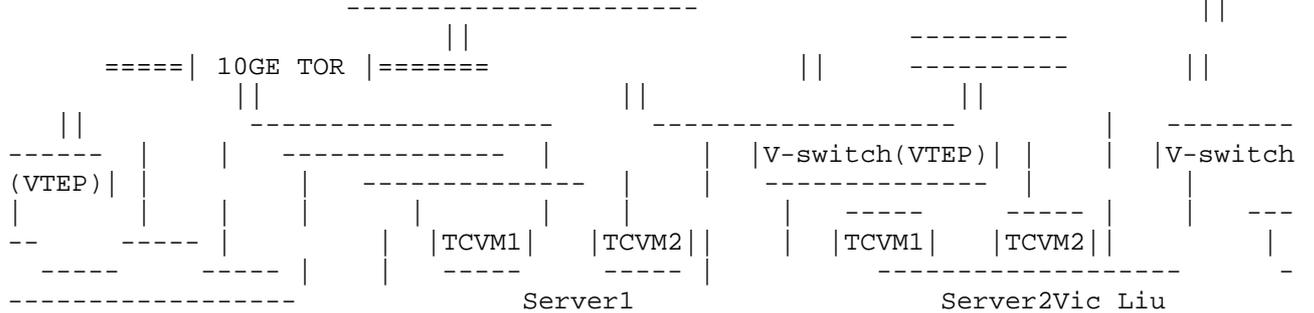
```
                     ----------------------              |Test Center PHY 10GE*2|
                          ----------------------               ||
                              ||                  ----------
      =====| 10GE TOR |======              ||   ----------        ||
              ||                       ||             ||
      ||      ------------------    ------------------      |  --------
------  |   |  ------------- |     |  |V-switch(VTEP)| |   | |V-switch
(VTEP)| |   |  ------------- |     |  ------------- |     |   |
|    |  |   | |   |  |    |   |  |TCVM2||  |  ----- ---  |   |  ---
--   ----- |   | |TCVM1|   |TCVM2||  |  |TCVM1|   |TCVM2||   |
 -----   ----- |   | ----- ----- |   ------------------    -
------------------            Server1                 Server2Vic Liu
```

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.

```
   ---------------------                       |Test Center PHY 10GE*2|   ||
              ---------------------                        ------------------
             ||                |  --------------  |              |  |V-
- switch(VTEP)| |                       | --------------  |          ----- ----- |
      |      |      |    |    |  |TCVM1|   |TCVM2||          |  -----
   ----- |                         ------------------       |  -----
       Server1    In traffic model B: We use the benchmark to test the performance of   VxLAN.
                                          ----------            =====| 10GE T
OR |======              ||     ----------        ||             ||      ||
             ||               ||              ||           ||        ----
---------------    ------------------        | --------------  |    | --
----------- |       |  |V-switch(VTEP)| |    |  |V-switch(VTEP)| |    |      |
  --------------  |       | --------------  |        |    |    |
     |      |      |    |  -----  ----- |    | -----  ----- |
  | |TCVM1|   |TCVM2||   |  |TCVM1|   |TCVM2||   |  -----  ----- |
  |  -----  ----- |      ------------------      ------------------
   -----        Server1           Server2
```

Vic Liu   Expires January4,2015   [Page 8]
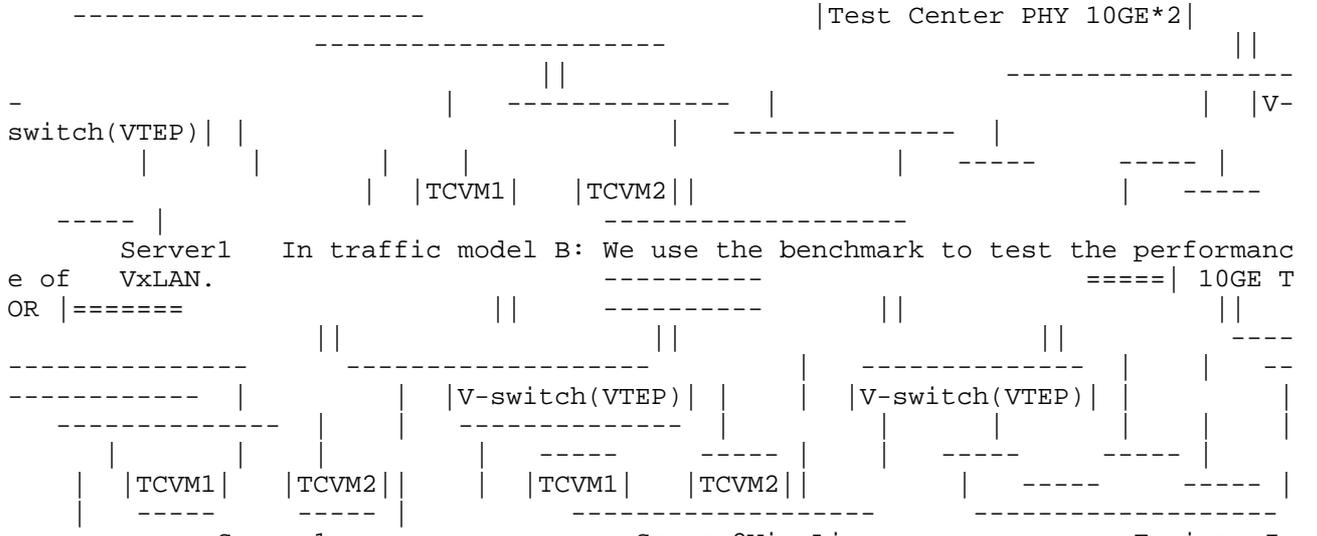
5. Proposed Benchmark Tests

5.1. Throughput

Unlike the traditional test cases which the DUT and the tester are separated, it has brought unparalleled challenges to virtual network test. In that case, the tester and the DUT (visual switches) are in one server (physically converged), so the CPU and MEM share the same resources. Theoretically, the tester's operation may has some influences on the DUT's performances. However, for the specialty of virtualization, this method is the only way to assess the truth of assessment method. Under the background of existing technology, when we mean to test the virtual switch's throughput, the concept of traditional physical switch will not be applicable. The traditional throughput indicates the switches' largest transmit capability, for certain selected bytes and selected cycle under zero-packet-lose conditions. But in virtual environment, the fluctuant of performance on virtual network will be much greater than dedicated physical devices. In the same time, because the DUT and the tester cannot be separated, which only proved the DUT realize same network performances under certain circumstances, it also means the DUT may achieve higher capability. Therefore, we change the throughout in virtual environment to actual throughput, hoping in future, as the improvement of technique, the actual throughput will approach the theoretical throughput gradually. Of course, under actual condition, this throughout have certain referential meanings. In most cases, common throughput application cannot compare with professional tester, so for virtual application and data center's deployment, the actual throughout already have great refinance value.

5.1.1. Objectives

Under the condition of certain hardware configuration, test the DUT(virtual switch) can support maximum throughput.

5.1.2. Configuration parameters

Network parameters should be define as follows:

a) the number of virtual tester (VMs)   b) the number of vNIC of virtual tester   c) the CPU type of the server   d) vCPU allocated for virtual tester (VMs)   e) memory allocated for virtual tester (VMs)   f) the number and rate of server NIC

5.1.3. The test parameters   a) test repeated times   b) test packet length

5.1.4. Testing process   1. Configure the virtual tester to output traffic through V-Switch.   2. Increase the number of vCPU in the tester until the traffic has no packet loss.   3. Record the max throughput on VSwitch   4. Change the packet length and repeat step1 and record test results.

5.1.5. Test results formats

Byte	Throughput (GE)
0	0
128	0.46
256	0.84
512	1.56
1024	2.88
1518	4.00

5.2. CPU consumption   The operation of DUT (VSwitch) can increase the CPU load of host server. Different V-Switches have different CPU occupation. This can

Vic Liu

be an i
mportant indicator in benchmark the Virtual network   performance.5.2.1. Objecti
ves   The objectives of this test is verified the CPU consumption caused   by th
e DUT (VSwitch).5.2.2. Configuration parameters   Network parameters should be d
efine as follows: a)The number of virtual tester (VMs)   b)The number of vNIC
of virtual tester   c)The CPU type of the server   d)vCPU allocated for virtual
tester (VMs)   e)Memory allocated for virtual tester (VMs)   f)The number and ra
te of server NIC5.2.3. The test parameters: a)test repeated times   b)test pac
ket length5.2.4. Testing process   1.Record CPU load value of server according t
o the steps of 5.1.3.   2.Under the same throughput, Shut down or bypass the DUT
 (VSwitch)   record the CPU load value of server.   3.Calculate the increase
of the CPU load value due to establish the   DUT (VSwitch).5.2.5. Test result
s formats   ---

| Byte| Throughput(GE)| Server CPU MHZ | VM CPU | --------------------

----------------------------- | 0 | 0 | 515

 | 3042 | ---

  | 128 | 0.46 | 6395 | 3040 | --------------

----------------------------------- | 256 | 0.84 | 6517

      | 3042 | ---

| 512 | 1.56 | 6668 | 3041 | ------------------
------------------------------ | 1024| 2.88 | 6280
   | 3043 | ---
       | 1450| 4.00 | 6233 | 3045 | ------------
-------------------------------------   5.3. Memory consumption   The operati
on of DUT (VSwitch) can increase the CPU load of host   server. Different V-Swit
ches have different memory occupation. This   can be an important indicator in b
enchmark the Virtual network   performance.5.3.1. ObjectivesThe objective of thi
s test is to verify the memory consumption by the DUT (VSwitch) on the Host ser
ver.5.3.2. Configuration parameters   Network parameters should be define as fol
lows: a) The number of virtual tester (VMs)   b) The number of vNIC of virtual
 tester   c) The CPU type of the server   d) vCPU allocated for virtual tester (
VMs)   e) Memory allocated for virtual tester (VMs)   f) The number and rate of
server NIC5.3.3. The test parameters: a) test repeated times   b) test packet
length5.3.4. Testing process   1. Record memory consumption value of server acco
rding to the steps   of 5.1.3.   2. Under the same throughput, Shut down or b
ypass the DUT (VSwitch)   record the memory consumption value of server.Vic L

3. Calc
ulate the increase of the memory consumption value due to   establish the DUT
 (VSwitch).5.3.5. Test results formats   -------------------------------
---------------- | Byte| Throughput(GE)| Host Memory | VM Memory | |
       --- | 0 |
   0 | 3042 | 696 | -------------------------------
---------------- | 128 | 0.46 | 3040 | 696 |
        --- | 256 | 0.
84 | 3042 | 696 | -------------------------------
-------------- | 512 | 1.56 | 3041 | 696 |
      --- | 1024| 2.88
  | 3043 | 696 | -------------------------------
------------- | 1450| 4.00 | 3045 | 696 |
     ---   5.4. Latency Physical
tester's time reference from its own clock or other time   source, such as GPS,
which can achieve the accuracy of 10ns. In   virtual network circumstances, the
virtual tester gets its reference   time from Linux systems. But the clock on Li
nux of different server   or VM can't synchronized accuracy due to current metho
d. Although VM   of some higher versions of CentOS or Fedora can achieve the acc
uracy   of 1ms, if the network can provide better NTP connections, the   result
will be better.   In the future, we may consider some other ways to have a bette
r   synchronization of the time to improve the accuracy of the test.   5.4.1.
Objectives The objective of this test is to verify the DUT (VSwitch) for latenc
y of the flow. This can be an important indicator in benchmark the Virtual net
work performance.   5.4.2. Configuration parameters   Network parameters should
 be define as follows:Vic Liu   Expires January4,2015
   [Page 13]

a) The number of virtual tester (VMs)   b) The number of vNIC of virtual tester   c) The CPU type of the server   d) vCPU allocated for virtual tester (VMs)   e) Memory allocated for virtual tester (VMs)   f) The number and rate of server NIC5.4.3. The test parameters: a) test repeated times   b) test packet length5.4.4. Testing process   1. Record latency value of server according to the steps of 5.1.3.   2. Under the same throughput, Shut down or bypass the DUT (VSwitch)   record the latency value of server.   3. Calculate the increase of the latency value due to establish the   DUT (VSwitch).5.4.5. Test results formats   TBD.6. Formal Syntax   The following syntax specification uses the augmented Backus-Naur Form (BNF) as described in RFC-2234[RFC2234].   Vic Liu   Expires January4,2015   [Page 14]

7. Security Considerations

8. IANA Considerations

9. Conclusions

10. References

   10.1. 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.

   10.2. 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.

11. Acknowledgments

   Thanks for Paul To, alick luo of sprient communication and
   Jianjun Lu of VMware Beijing Office supporting this draft.

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

Authors' Addresses

    Vic Liu
    China Mobile
    32 Xuanwumen West Ave, Beijing, China
    Email: liuzhiheng@chinamobile.com

    Dapeng Liu
    China Mobile
    32 Xuanwumen West Ave, Beijing, China
    Email: liudapeng@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

    Guang Zhang
    IXIA
    Email: GZhang@ixiacom.com