# EANTC Independent Test Report

## NFWare Virtual CGNAT on Lenovo NFVi Intel Select Solution Performance Benchmarking

October 2019

## Introduction

Global Internet traffic is transported between Internet Protocol (IP) endpoints with public IP addresses. The majority of Internet content is reachable by Internet Protocol version 4 (IPv4). IPv4 has a limited capacity of the public addresses that are globally routable on the Internet (less than 4.3 billion).

That said, the total number of devices connecting to the Internet increases continually. The pool of unallocated IPv4 addresses have long been depleted, which led to what is known as IPv4 address exhaustion today. The Internet Engineering Task Force (IETF) addressed this problem by developing a newer version of the Internet protocol named Internet Protocol version 6 (IPv6). IPv6 supports a larger block of IP addresses to alleviate the insufficient capacity of the addresses for the foreseeable future.

Internet Service Providers (ISPs) consider IPv6 as a strategic enabler for the evolution of their networks and radical solution for IPv4 address exhaustion problem. The complete shift from IPv4 networks (devices, Infrastructure & content) to full IPv6 networks requires prolonged plans to achieve that because:

1. The majority of Internet web content is only reachable by IPv4 addresses. Referring to the Internet Society State of IPv6 Deployment 2018 Report, less than 30% of Alexa's Top 1,000 global websites are reachable via IPv6
2. IPv6 is not backward compatible with IPv4 natively. Due to this incompatibility, extra investments are required in the devices of the subscribers, network infrastructure, and Internet content

While implementing the rollout plans for IPv6 deployments, ISPs utilize the Carrier-Grade NAT (CGNAT) as an interim solution to relieve IPv4 exhaustion. In principle, CGNAT is an extension of the standard NAT for large scale service provider deployments by sharing a public IP address between many subscribers to get access to the Internet. The flexibility of CGNAT yields to many deployment schemes in the ISP environment that serve various purposes:

- NAT44: The IPv4 source address is translated to another private or public IPv4 source address. While NAT44 is a translation among two IPv4 address domains, the extended-term NAT444 describes the process of address translation among three IPv4 address domains
- NAT64: The IPv6 source address is translated to IPv4 source address to be routable within IPv4-only network
- NAT46: The IPv4 source address is translated to IPv6 source address to be routable within IPv6-only network

### Test Highlights

→ High-performance CGNAT solution optimized for full-server NFV environments

→ Up to 230 Gbps throughput performance (1520 Bytes frame size) using PCI Passthrough

→ Peak session establishment rate of 7 million sessions per second

→ Scalable sessions table up to 100 million concurrent sessions

In parallel with the network evolution of the Communication Service Providers (CSP) toward network virtualization and Telco-Cloud, CGNAT is one of the early-adopted network functions that can be extended in the virtual environment, thus providing flexibility and cost-efficiency for small and large deployments.

EANTC has been commissioned by Lenovo to verify the performance and scalability of the NFWare virtual Carrier-Grade Network Address Translation (vCGNAT) solution. In this test, the Lenovo ThinkSystem SR650 server hosted NFWare vCGNAT VM. EANTC tested the maximum throughput (Gbps), sessions setup rate (Sessions Per Second), and maximum concurrent sessions of NFWare vCGNAT.

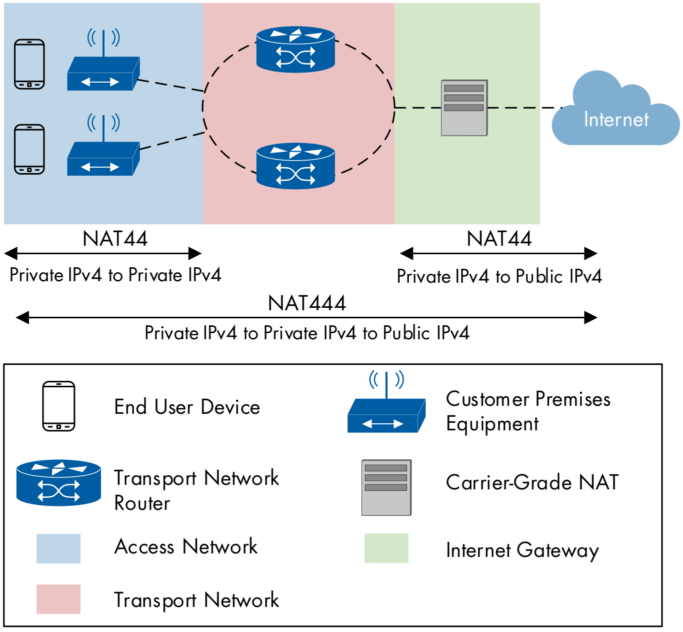


**Figure 1: IPv4 Address Translation in SP Network**

## Executive Summary

NFWare vCGNAT is a high-performance software-based solution for transparent address and protocol translation. According to NFWare, the solution enables service providers to extend their IPv4 networks quickly and to migrate to IPv6 smoothly. A Lenovo ThinkSystem SR650 server was used to host the NFWare vCGNAT Virtual Network function (VNF). The server was powered by dual Intel® Xeon® Gold 6252 second generation scalable processors (@ 2.10 GHz,80 24 cores). To verify the maximum traffic throughput that the NFWare vCGNAT can process in a single compute node, the server was equipped with five Intel dual-port 25 GbE Network Interface Cards (NICs). By request of the vendor, the ten 25GbE ports were attached to the NFWare vCGNAT VNF using PCI Passthrough. NFWare explained that the full potential of their VNF could be exploited only using PCI Passthrough mode. The Virtual Infrastructure Manager (VIM) of the testbed was supported by Red Hat OSP 13. Moreover, Lenovo leveraged its organic Lenovo Open Cloud automation toolsets for the rapid deployment of the OpenStack environment.

The NFWare vCGNAT VNF was configured to serve large-scale deployment of IPv4 to IPv4 address translation. This use case is typically found at Internet gateways (IGW) of communications service providers.

The combined solution of Lenovo ThinkSystem SR650 server and NFWare vCGNAT VNF showed superior performance and capabilities. EANTC measured a maximum throughput performance of 230 Gbps (1520 Bytes frame size), 72 Gbps (84 Byte frame size), accompanying around 100 million IPv4 flows. These results confirm that the solution is ready to be deployed by service providers in their regional Internet gateways.

## Test Bed Description

The test execution environment combined the Network Function Virtualization Infrastructure (NFVI), the Function Under Test (FUT) as well as the testing tools and the traffic generators. Lenovo provided the NFVI which is compliant with the "Base Configuration Reference Design of Intel Select Solution for NFVI v2". The NFVI included one compute node (Lenovo ThinkSystem SR650) to host the FUT as shown in Figure 2 and three controller nodes (Lenovo ThinkSystem SR630) to host the controller virtual machine for Red Hat OpenStack OSP 13 deployments. Table 1 lists the components of the NFVI used in the test.

| Item | Description | Qty |
|---|---|---|
| **Compute Node** | **Lenovo ThinkSystem SR650** | **1** |
| Processor | Intel® Xeon® Gold 6252 24C 2.1GHz processor | 2 |
| Memory | ThinkSystem 32GB TruDDR4 2667MHz RDIMM (total 768GB) | 24 |
| NIC (Integrated) | ThinkSystem 1Gb 4-port RJ45 LOM (uses Intel X722 1/10 GbE) | 1 |
| NIC | Intel XXV710-DA2 PCIe 25Gb 2-Port SFP28 Ethernet Adapter | 5 |
| Storage (NVMe) | ThinkSystem U.2 Intel P4500 2TB Entry NVMe PCIe 3.0 x4 Hot-Swap SSD | 2 |
| Boot Drive | ThinkSystem 2.5" Intel S4500 480GB Entry SATA 6Gb Hot-Swap SSD | 2 |

**Table 1: Lenovo supported Configuration for Intel Select Solution for NFVI Certification (Base)**

Lenovo ThinkSystem NE2572 RackSwitch as shown in Figure 3 was used for the data traffic connectivity between the compute node server and the traffic generator using 25GbE physical links. Separately, the physical connectivity to the management network was provided by Lenovo RackSwitch G8052.

**Figure 2: Lenovo ThinkSystem SR650 Server**

**Figure 3: Lenovo ThinkSystem NE2572 RackSwitch**

| Ingredient | SW | Version |
|---|---|---|
| Firmware | BIOS | 2.13 |
| | BMC | V2.12 |
| | Intel® Ethernet Controller XXV710 | 6.80 |
| Host OS | Red Hat Enterprise Linux Server | RHEL 7.7 (Kernel: Linux 3.10.0-1062.el7.x86_64) |
| Hypervisor | KVM/QEMU | 2.12.0 |
| Net Driver | i40e | 2.7.7.1 |
| VIM | Red Hat OpenStack Platform | Red Hat OpenStack 13 |
| | Lenovo Open Cloud Automation | v0.9 (pre-GA) |

**Table 2: Software & Firmware Stack**

Table 2 lists the software and firmware details.

NFWare defined the compute and networking resources of the Function Under Test (FUT); which is a single VM instance of NFWare vCGNAT. Table 3 lists the allocated resources for the FUT. Using Red Hat OSP13, the required NFVI resources for the FUT was allocated smoothly and straightforward. Furthermore, the VM was instantiated successfully.

| Configuration Parameter | Value |
|---|---|
| NFWare vCGNAT VNF Version | 3.3.0.4148 |
| Assigned vCPU (Threads) | 84 |
| Assigned Memory value | 280 GBytes<br>(140 GBytes per socket) |
| Assigned Physical Network Interface | 10x25 GbE with PCI Passthrough |
| Assigned storage space | 80 GBytes |

**Table 3: NFWare vCGNAT Hardware and Software Configurations**

## NFVI Optimization for High-Performance Data Plane

Lenovo team configured the NFVI environment based on some common best practices to maximize the data plane performance.

Besides enabling Intel® Hyper-Threading Technology and Intel® Turbo Boost Technology, the following technologies were intentionally configured, including:

- Non-Uniform Memory Access (NUMA) Topology Awareness: The compute node server integrates two processors with two NUMA nodes; each processor contains 24 core/48 threads. The PCI buses of 3 NICs were assigned to NUMA 0, and the other 2 NICs were assigned to NUMA 1.
- CPU Pinning: CPU pinning is the process of pairing a virtual CPU to a physical CPU to make sure processes that run on a certain vCPU will be run on the physical CPU that we have determined. Similarly, CPU isolation is the process of isolating the CPUs to make sure it handles only those interrupts that we have determined. Total 84 vCPUs; 42 vCPUs from each NUMA node are assigned to the VNF. Also, they are isolated to handle only the vCGNAT interrupts and pinned to dedicated cores in the host machine.

Additional configuration and optimization information for Lenovo's NFVI environment can be found at https://lenovopress.com/lp0913.pdf.

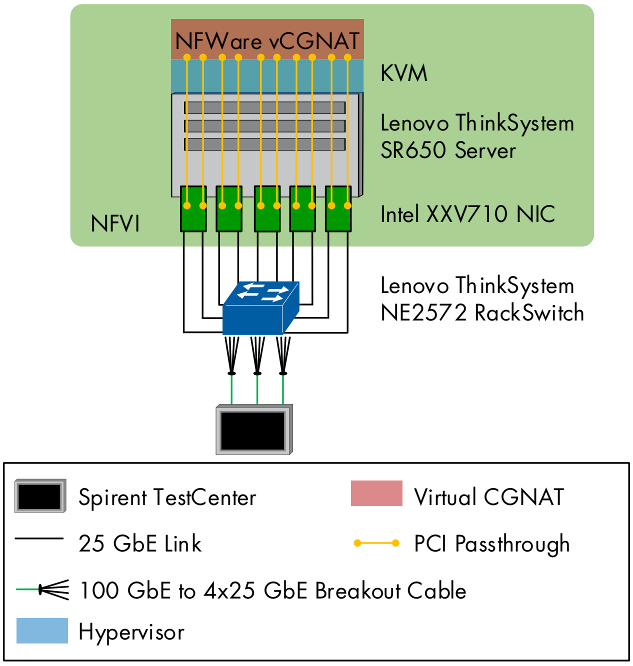


**Figure 4: Logical Test Bed**

## Test Equipment and FUT Configurations

The test cases were executed using Spirent Communications N4U chassis with a PX3 line card. PX3 module delivers the highest density high-speed Ethernet solution per module by integrating 12x100GbE ports.

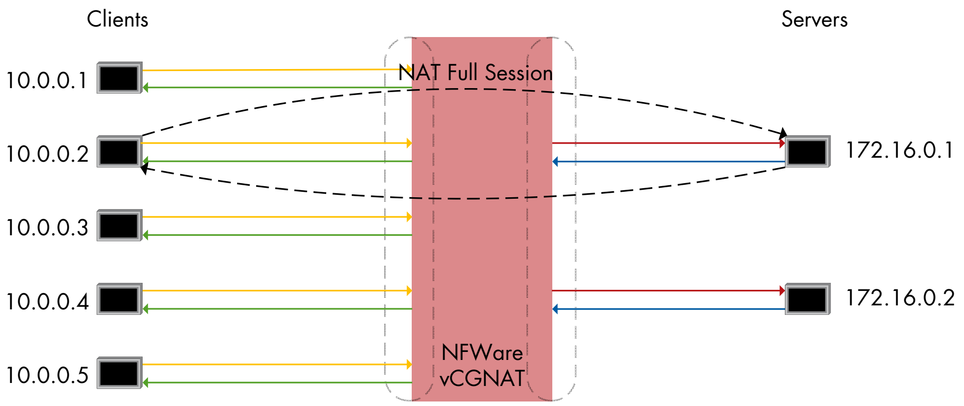
In this test, we utilized 3x100GbE ports and connected each port physically to data traffic switch (Lenovo ThinkSystem NE2572 RackSwitch) through 100GbE-to-4x25GbE breakout cables, as shown in Figure 4.

The client packets are expected to flow through the FUT. The FUT replaces the internal source IPv4 address with an external source IPv4 address and a specific UDP source port number and forwards the packet toward the destination server. Once the server replies to the client, it sets the destination IPv4 address and the UDP destination port number as the same values of the external source IPv4 address and the UDP source port number of the client packet. The traffic generator has limited features to send back the packet from the server to the client with the same destination port number. NFWare configured the FUT based on static-NAT rules. Each internal IPv4 address is mapped one-to-one with an external IPv4 address.

This allows us to experience a deterministic behavior during the test by enforcing the return traffic to hit the same external IPv4 address and to overcome the traffic generator limitation, as shown in Figure 5.

However, the FUT configurations don't have any impact on the achieved results (Throughput, Session Setup Rate and Concurrent Sessions). The FUT was ready to handle up to 160 million sessions based on the following configurations:

- 5 IPv4 address pools were used as IP internal clients, each pool consists of 4K IPv4 addresses, in a total of 20K
- 5 IPv4 address pools were used as servers IP addresses, each pool consists of 400 IPv4 addresses, in a total of 2K
- Each internal client IPv4 address used four source UDP ports to generate traffic toward each server IP address
- Static-NAT rules were configured to map each permutation of the internal client (IPv4 address, UDP port) to a static permutation of the external client (IPv4 address, UDP port)



| Internal Addresses | | | | External Addresses | | | |
|---|---|---|---|---|---|---|---|
| Src. IP | Dst. IP | Src. Port | Dst. Port | Src. IP | Dst. IP | Src. Port | Dst. Port |
| 10.0.0.1 | 172.16.0.1 | 1025 | 1025 | 10.8.0.1 | 172.16.0.1 | 1025 | 1025 |
| 10.0.0.2 | 172.16.0.1 | 1026 | 1025 | 10.8.0.2 | 172.16.0.1 | 1026 | 1025 |
| 10.0.0.3 | 172.16.0.2 | 1027 | 1025 | 10.8.0.3 | 172.16.0.2 | 1027 | 1025 |
| 10.0.0.4 | 172.16.0.2 | 1025 | 1025 | 10.8.0.4 | 172.16.0.2 | 1025 | 1025 |
| 10.0.0.5 | 172.16.0.2 | 1028 | 1025 | 10.8.0.5 | 172.16.0.2 | 1028 | 1025 |

| | | | |
|---|---|---|---|
| ■ | Spirent TestCenter spirent Promise. Assured. | ⇆ | NAT Session |
| → | Outbound Traffic Before NAT | → | Inbound Traffic Before NAT |
| → | Outbound Traffic After NAT | → | Inbound Traffic After NAT |

**Figure 5: Static NAT Sessions Table**

## Test Results

The targeted scope of this test was to verify the performance of NFWare vCGNAT when it is running on a specific NFVI Lenovo ThinkSystem SR650 server. EANTC set three Key Performance Indicators (KPIs) to evaluate the performance of the solution. Those three KPIs are:

1. Traffic Throughput (Gbps): To measure the maximum bidirectional throughput that the FUT can forward in Gigabits per second. The throughput is measured based on 100 million established sessions to emulate the typical use case for a large scale SP.
2. Sessions Establishment Rate (Sessions Per Second): To measure the maximum number of new NAT sessions that can be created by FUT in one second with zero frame loss.
3. Maximum Concurrent Sessions: To measure the maximum number of NAT sessions that the FUT can hold without service interruption or frame loss.

One of the guiding factors for any EANTC benchmark is reproducibility. We aim to share sufficient details to enable readers to reproduce our test setup and results independently.

The configuration of NFWare vCGNAT remained the same during the whole test execution. Figure 6 depicts the logical network topology used in the three test cases.

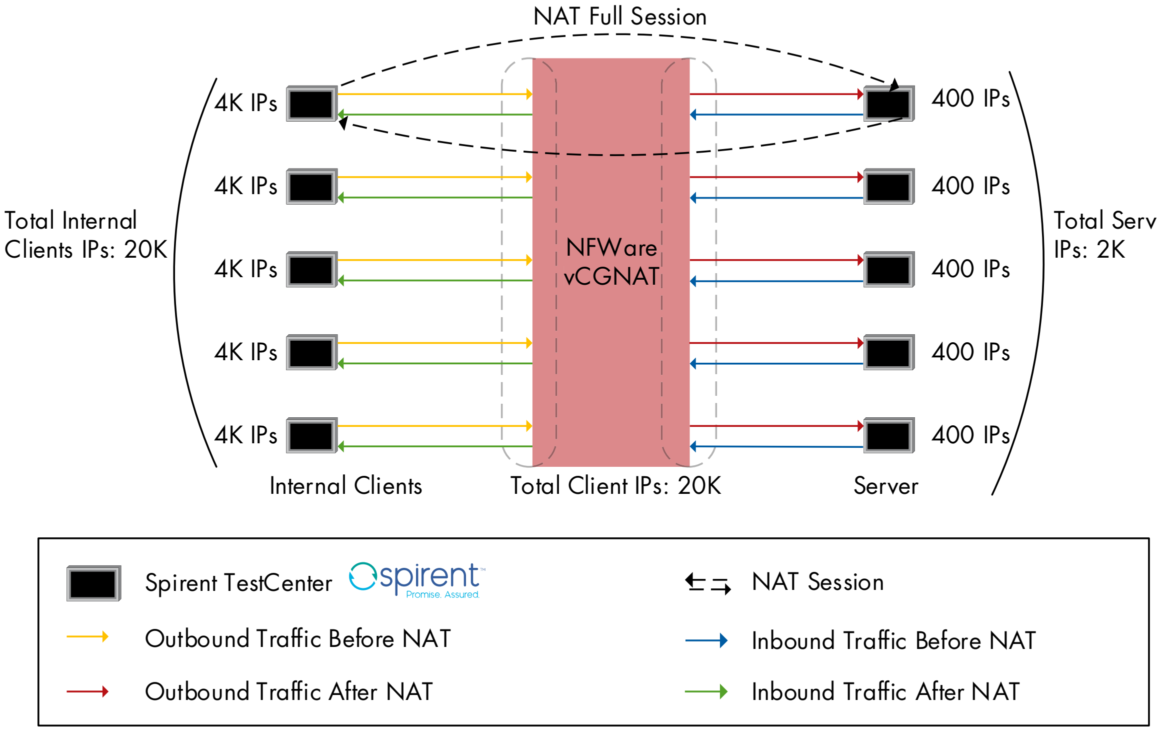


**Figure 6: Traffic Flow Diagram**

### Throughput Test

The first performance test was to measure the maximum throughput that can be handled by NFWare vCGNAT along with 99.6 million sessions.

We configured the traffic generator to emulate 20K internal client IP addresses and 1245 server IPs. We generated bidirectional traffic from the clients to the servers and vice versa.

We started the test with a low load initial traffic for 15 seconds to establish all the targeted number of sessions on the vCGNAT (99.6 million sessions). Then, we generated the traffic with the targeted rate while all the sessions were successfully established and within the timeout period.

Table 4 shows the IMIX distribution we used in the test. Table 5 lists the achieved results of the throughput test for different frame sizes.

| Frame Size (Bytes) | Weight | Percentage (%) |
|---|---|---|
| 82 | 1 | 3.7 |
| 118 | 11 | 40.74 |
| 391 | 3 | 11.11 |
| 588 | 2 | 7.4 |
| 1318 | 3 | 11.11 |
| 1536 | 7 | 25.92 |

**Table 4: IMIX Distribution**

Figure 7 shows the maximum throughput test results for different frame sizes measured with Millions of Frames per second and with Gigabits per second.

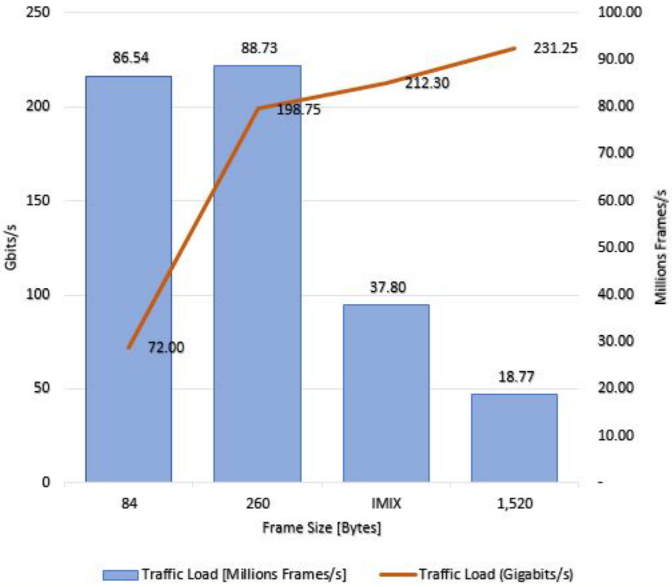


**Figure 7: Maximum Throughput**

| Number of Flows | Frame Size (Bytes) | Offered Load (Gbps) | Latency (µs) | | |
|---|---|---|---|---|---|
| | | | Min | Max | Avg |
| 99,600,000 | 84 | 72.00 | 8 | 743 | 83 |
| 99,600,000 | 260 | 198.75 | 10 | 555 | 110 |
| 99,600,000 | IMIX | 212.30 | 10 | 444 | 33 |
| 99,600,000 | 1,520 | 231.25 | 10 | 442 | 28 |

**Table 5: Maximum Throughput Results**

### Session Establishment Rate

The objective of the second test case was to measure the maximum connections per second the FUT can handle without frame loss.

For that, we configured the Spirent TestCenter to instantiate IP sessions at different rates. In each iteration, we observed the maximum sessions establishment rate on the FUT along with the frame loss. We kept the test running for 10 seconds to verify the sustainability of the achieved session establishment rate. In the end, the sessions table was flushed to ensure no sessions left for the next iteration.

The FUT was successfully capable of creating 7 million sessions per second without frame loss, resulting in establishing 70 million sessions in 10 seconds.

Figure 8 shows the session establishment rate achieved.

### Maximum Concurrent Connections

The third test case was to measure the maximum number of NAT sessions that the FUT can hold without service interruption or frame loss. To measure the maximum number of concurrent sessions, we configured the traffic generator to instantiate unidirectional sessions toward the FUT by a rate of 3.5 million sessions per second; which is 50% of the achieved maximum sessions per second. The FUT was able to achieve 100 million concurrent sessions in total without dropping any NAT session or traffic frames. Each 5 million sessions were stored in a different RAM memory block which is assigned to a different NUMA node. The maximum memory utilization for the sessions was 62.5%. Figure 9 shows the memory assignment and utilization from the Command Line Interface (CLI) of the FUT.

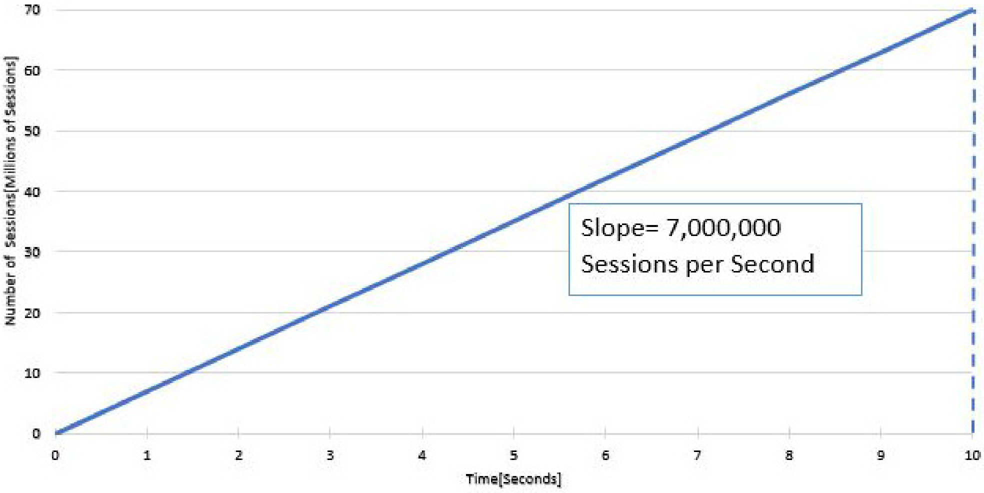
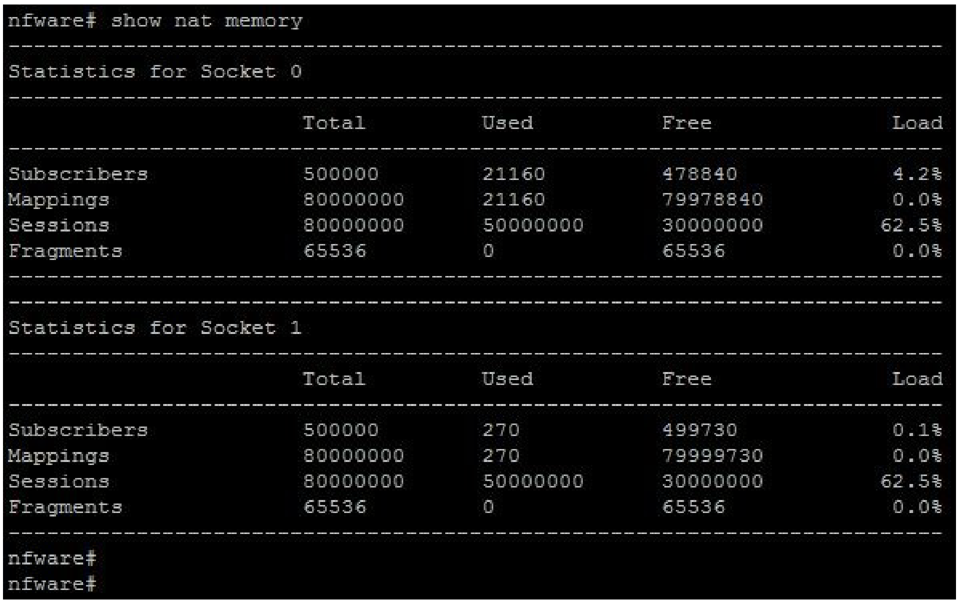


**Figure 8: Session Establishment Rate**



```
nfware# show nat memory
--------------------------------------------------------------------------
Statistics for Socket 0
--------------------------------------------------------------------------
                       Total         Used          Free              Load
--------------------------------------------------------------------------
Subscribers            500000        21160         478840            4.2%
Mappings               80000000      21160         79978840          0.0%
Sessions               80000000      50000000      30000000         62.5%
Fragments              65536         0             65536             0.0%
--------------------------------------------------------------------------
--------------------------------------------------------------------------
Statistics for Socket 1
--------------------------------------------------------------------------
                       Total         Used          Free              Load
--------------------------------------------------------------------------
Subscribers            500000        270           499730            0.1%
Mappings               80000000      270           79999730          0.0%
Sessions               80000000      50000000      30000000         62.5%
Fragments              65536         0             65536             0.0%
--------------------------------------------------------------------------
nfware#
nfware#
```

**Figure 9: Memory Assignment and Utilization for Maximum Concurrent Sessions**

## Conclusion

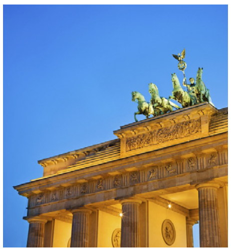
The NFWare vCGNAT VNF uses the PCI Passthrough mechanism, which connects hardware resources directly with the VNF, using these hardware resources exclusively and circumventing the abstraction layer of the virtualization environment. The current solution works great as an appliance exclusively utilizing the server's resources – which is often sufficient for a very large CGNAT scenario. The PCI passthrough architecture does not allow workload migration; high availability scenarios must be provided on the application layer. We did not test any failover functions.

In the future, we hope to verify another release of the vCGNAT VNF which might exploit the full potential of the NFV architecture by using virtual switching functions.

Taking the PCI Passthrough solution into account, we confirmed outstanding performance of NFWare vCGNAT NFV with 230 Gbps traffic throughput and 100 million concurrent sessions. Furthermore, we validated superior utilization of the allocated NFVI resources by the vCGNAT VNF, efficiently utilizing 92% of the compute node networking capacity (250 Gbps) and 87.5% of the compute node processing capacity (96 vCPUs).

The combination of Lenovo ThinkSystem SR650 server and NFWare vCGNAT VNF implements a fixed configuration appliance using a lot of aspects of the NFV architecture. It is a high-performance solution ready for large-scale service provider scenarios, able to process hundreds of Gbps traffic and ready for large-scale deployment scenarios.

## About EANTC



EANTC (European Advanced Networking Test Center) is internationally recognized as one of the world's leading independent test centers for telecommunication technologies.

Based in Berlin, the company offers vendor-neutral consultancy and realistic, reproducible high-quality testing services since 1991. Customers include leading network equipment manufacturers, tier 1 service providers, large enterprises and governments worldwide. EANTC's Proof of Concept, acceptance tests and network audits cover established and next-generation fixed and mobile network technologies.

While every reasonable effort has been made to ensure accuracy and completeness of this publication, the authors assume no responsibility for the use of any information contained herein. All brand names and logos mentioned here are registered trademarks of their respective companies in the United States and other countries.

EANTC AG
Salzufer 14, 10587 Berlin, Germany
info@eantc.com, http://www.eantc.com/
[v1.0 20191010]