Optimizing Networking and Security Performance Using VMware vSphere and NVIDIA BlueField DPU with BWI

Whitepaper, Proof of Concept

WP-11504-001 | August 2023

Document History

WP-11504-001

Version	Date	Authors	Description of Change
01	August 28, 2023	> Martin Schulz > Marcus Schoent > Joerg Roesch > Karthik Ganesan > John F. Kim > Jose Castanos	Initial release

Table of Contents

Executive Summary

In 2022, VMware launched the capability to run vSphere on DPUs, which allows vSphere to offload networking and NSX services to a DPU like NVIDIA® BlueField®. In this joint testing between BWI GmbH and VMware, we study how offloading security features such as distributed firewalling to the DPU improves network performance for highly secure environments requiring hundreds of firewall rules. Our results show that we can achieve close to line rate network throughput when using an NVIDIA BlueField-2 DPU with vSphere to accelerate networking and firewall functions, even in the presence of thousands of firewall rules. We also show how the fully-accelerated DPU mode almost completely frees up the CPU cores on the host from network and firewall tasks. This enables these cores to process applications in isolation for maximum performance without much interference from sharing the CPU resources with other tasks.

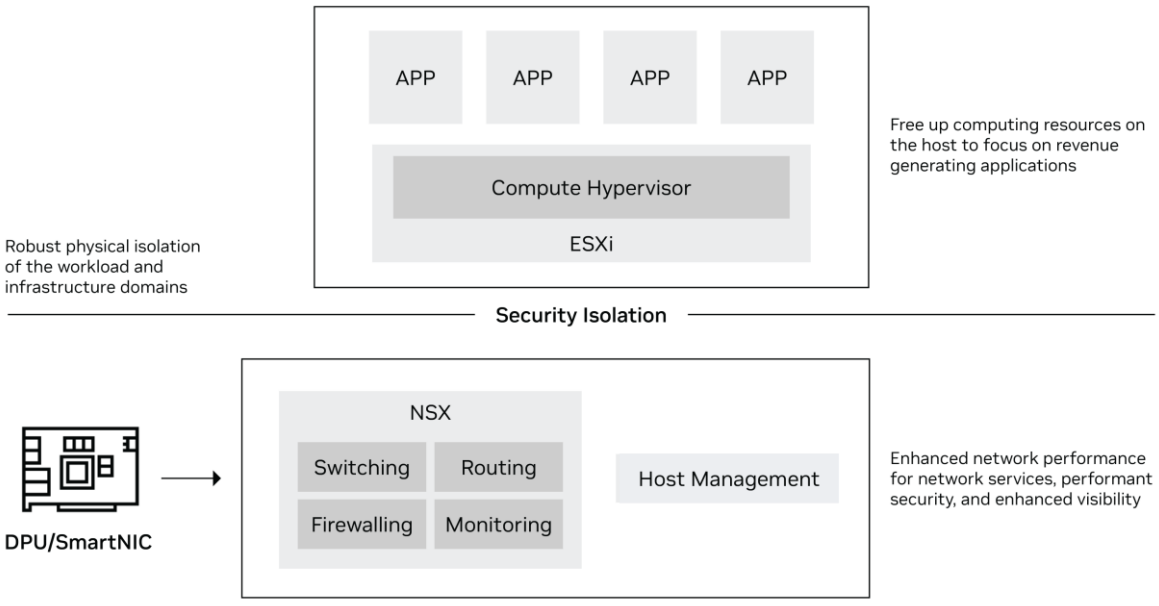
Introduction

With vSphere 8 and NSX 4, VMware has introduced support for SmartNICs or Data Processing Units (DPUs). The DPU implementation in vSphere is called vSphere Distributed Service Engine. This is the capability formerly known as “Project Monterey”.

DPUs (SmartNICs) are network cards with built-in intelligence that can perform a variety of network functions directly on the adapter through their own programmable processors. In addition to the networking accelerators, DPUs like NVIDIA BlueField also have general-purpose Arm processor cores that can run a full ESXi general system.

With the DPU technology, NSX services like routing, switching, firewall and monitoring are offloaded to the DPU from the host hypervisor. With these capabilities it is possible to improve performance, free up resources on the host and isolate workload and infrastructure domains (see Figure 1).

Figure 1. DPU Architecture



This whitepaper provides information on how DPUs (SmartNICs) improve performance for highly secure environments that have stringent regulatory requirements. Specifically, the focus is on the security function of the distributed firewall. The test scenarios include NSX Distributed Firewall (Microsegmentation) tests with “Any-Any-Allow-Rule", with from 1000 to 4000 rules activated. The test scenarios include traffic between source and destination VMs placed on different ESXi hosts, and on the same “overlay” network.

BWI Use-case and NSX Distributed Firewall

BWI GmbH (abbreviated BWI in this document) is the IT system house of the German armed forces. As a digitization partner, BWI is accompanying the German armed forces’ biggest transformation to digitization. A major focus here is the design and implementation of cloud computing for mapping administrative tasks through to military operations, considering the high requirements for information security and secrecy protection.

For this reasons BWI is developing a German armed forces private cloud (named pCloudBw), with a focus on cloud-native provisioning of functions based on Kubernetes. As a long-standing customer and partner of VMware, BWI uses a variety of products from VMware's portfolio at the core of the pCloudBw architecture, including VMware Cloud Foundation®, VMware NSX®, VMware Tanzu®, etc. The top design goals of the development are standardization and automation.

The biggest challenge in the architecture of the private cloud pCloudBw is the conception of a security architecture and the implementation of different tenant groups. Within BWI, tenants are defined by the art of the data to be processed. To be able to identify these tenants, a control matrix was designed to classify data.
In Figure 2 the different security and information domains are visible.

- **Information Space** – Defines the national sovereignty of the system and the basic regulations to be applied.
- **Security Domains** – Defines the requirements for classification of data.
- **Information Domains** – Combined requirements of the information space and security domain and defines the border of physical separation.

Figure 2. BWI Information and Security Domains

Control Matrix

Security Domain	Information Space		
	A	B	Mission
Top Secret	A - Top Secret	B - Top Secret	Mission - Top Secret
Secret	A - Secret	B - Secret	Mission - Secret
Confidential	A - Confidential	B - Confidential	Mission - Confidential
Restricted	A - Restricted	B - Restricted	Mission - Restricted
Unclassified	A - Unclassified	B - Unclassified	Mission - Unclassified
Public	A - Public	B - Public	Mission - Public

Defines the border of physical separation

- **Information Space:** Defines the national sovereignty of the system and the basic regulations to be applied
- **Security Domain:** Defines the requirement for the Classification of data
- **Information Domain:** Combined requirements of Information Space and Security Domain, defines the border of physical separation

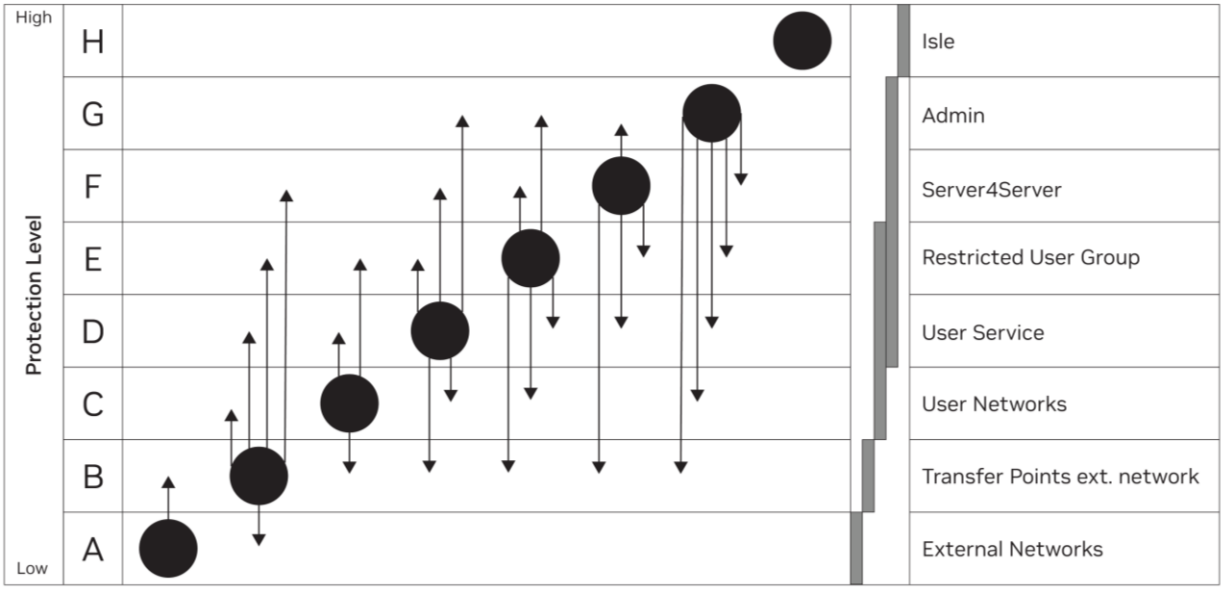
The tenant classifications resulting from the control matrix are to be implemented in the form of dedicated infrastructures. This means using dedicated compute hardware and dedicated management systems for each type of tenant. VMware Cloud Foundation is used with the Software-Defined-Networking solution NSX to separate the network layer with overlay technologies.

Further, the networks within a security and information domain are segmented. This segmentation is based on the creation of security zones. The zones are separated with NSX-T Overlay. The following zones are defined (see Figure 3):

- **Zone A** - All networks that are not under the administrative control of BWI
- **Zone B** - The network between Zone A and the BWI-controlled Zones C – G, the equivalent of a network DMZ (Demilitarized Zone).
- **Zone C** - In this network zone all end devices (workstations as desktop or laptop, printers, etc.) are placed
- **Zone D** - This network zone is classically assigned to the data center. It is used to place frontend systems for server applications in 2- or 3-tier web applications.
- **Zone E** - This network zone is classically assigned to the data center. It is used to place application servers/backend systems for server applications in 2- or 3-tier web applications.
- **Zone F** - This network zone is classically assigned to the data center. It is used to place database systems or other persistent data storing systems for server applications in 2- or 3-tier web applications.
- **Zone G** - This network zone contains the workstations of the administrators for the operation and administration of all systems in the German armed forces IT system.

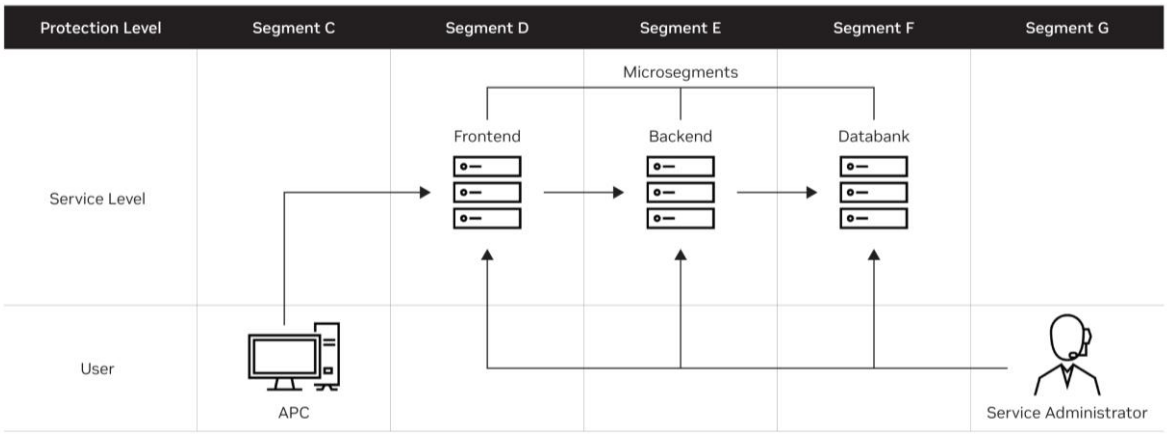
The zone model also defines the direction in which connections may initially be established.

Figure 3. BWI Networking Zoning Model



Microsegmentation using the NSX Distributed Firewall is used as the final stage of security segmentation. The networks that are assigned to a network zone are broken down into many small security groups. The advantage from NSX Distributed Firewall of having no dependencies on IP ranges helps to realize an automated, scalable and operable Microsegmentation design. Microsegments are created as needed when applications are implemented. Each application has its own security group in each security zone (see Figure 4). By default, the individual microsegments assigned to the different applications cannot communicate with each other, even though they might be in the same network zone and same physical network.

Figure 4. Microsegmentation Design



In BWI, these regulations for detailed Microsegmentation need to be broken down for like overlay network encapsulation and decapsulation offload. Every single workload, like VMs or containers, requires a large firewall ruleset. These must be implemented accordingly in NSX as policies. With more than 3000 active policies per VM, the firewall rule sets in BWI are very extensive and incur significant resource costs affecting network throughput, latencies, and compute resources in production. In addition to the use of firewall functionalities, IDPS (intrusion detection and prevention systems) functions are also used in the network architecture. The acceleration of this security function represents an enormous added value for BWI when using DPUs.

For these reasons, BWI tested vSphere on NVIDIA DPUs with the NSX Distributed Firewall to verify if the performance and the scaling limits fit their requirements.

vSphere on NVIDIA BlueField-2 DPU

NVIDIA BlueField-2 DPU

A Data Processing Unit, or DPU, is a specialized piece of silicon optimized for running data center infrastructure tasks such as networking, storage, security, and management.

BlueField-2 is the second generation of NVIDIA's DPU infrastructure-on-a-chip designed to optimize enterprise, AI, and high-performance computing workloads. It provides a variety of software-defined networking, storage, security, and isolation services that can be offloaded to specialized engines with no server CPU overhead.

The BlueField-2 chip includes an NVIDIA ConnectX® 6-Dx network adapter combined with an array of Arm cores, DRAM controller, PCIe switch and various acceleration engines. The DPU runs its own system image independent of the main system image, providing management and security isolation. In the case of VMware, the system runs a light version of ESXi ported to the Arm cores. The ConnectX portion of BlueField provides a highly programmable embedded eSwitch. Networking pipelines defined between virtual or physical ports in this switch enable hardware-accelerated processing and manipulating of network traffic at line speed without intervention of the server's CPU. The role of the Arm processors is to program and monitor these accelerated pipelines, therefore offloading the control-plane (in addition to offloading the data-plane) to the DPU.

In addition to the eSwitch, the BlueField-2 incorporates a variety of acceleration engines: IPSec/TLS data data-in-motion encryption, data at-rest storage encryption, public-key acceleration, NVMe-oF storage acceleration and virtualization, data hashing and deduplication and deep packet inspection (for intrusion detection and stateful L3 firewall rules).

To support high performance firewalls with stateful rules, BlueField-2 supports stateful connection tracking in hardware at line rate. This is the main BlueField-2 feature we are exploring with VMware vSphere in this whitepaper, in addition to standard SDN services like overlay network encapsulation and decapsulation offload.

vSphere on DPUs: Running on NVIDIA BlueField

VMware is the leading provider of multi-cloud services for all apps, enabling digital innovation with enterprise control. vSphere is the enterprise workload platform that brings the benefits of cloud to on-premises workloads, supercharges the performance through DPUs and GPUs, and accelerates innovation with an enterprise-ready integrated Kubernetes runtime.

While using traditional virtualization with any hypervisor, including VMware, as VMs become more powerful and as network speeds increase, an increasing percentage of CPU power is devoted to managing data movement and infrastructure. Tasks associated with network routing, network overlays (for Geneve or VXLAN), security (such as a distributed firewall), telemetry, storage, and remote management can consume more than 30% of CPU resources on each virtualized server.

One way to resolve this overhead is to offload networking functions and security tasks to a DPU that contains purpose-built “engines” for offloading these specific workloads. The launch of vSphere 8 ushered in a new era of heterogeneous computing by introducing Data Processing Units to enterprises through vSphere Distributed Services Engine (DSE). vSphere Distributed Services Engine is the next step in the evolution of cloud infrastructure for modern applications, in which the stewardship for running infrastructure services is distributed between the CPU and the DPU.

Starting with vSphere 8, VMware introduced Uniform Passthrough version 2 (UPTv2) or VM-direct compatibility, which is called the fully accelerated mode and allows a VMXNET Generation 3 (VMXNET3) adapter to be configured to use the capabilities of the DPU in passthrough mode. Typically, a device that is connected in passthrough mode–using SR-IOV (Single Root IO Virtualization) to bypass the hypervisor and provide high performance–must sacrifice specific virtualization features like VMware vSphere® vMotion®, vSphere HA, etc., because networking is bypassing the hypervisor. But the VM-direct mode brings the best of both worlds by enabling one to achieve SR-IOV like high networking performance without losing the workload services that vSphere enables.

Some of this high performance can be attributed to the fact that the BlueField DPU in the fully accelerated mode accelerates network packet processing by doing most of this in hardware, in comparison to an emulated regular NIC that entails running some slower packet processing tasks in software on the CPU.

One of the major pain points solved by offloading networking to DPUs is the cache pollution caused from running hypervisor network processing alongside application logic on the same host cores. In this case, it is common for the frequently used networking and application information to push each other out of the shared cache, reducing the cache hit rate for both. By offloading and isolating network processing to the DPU, the application logic running on the host cores can enjoy better cache locality while the networking processes use a separate cache on the DPU, often resulting in significant application performance boosts in terms of latency and throughput.

Note: To be able to use the VM-direct mode, one should ensure specific VMXNET3 driver versions, full memory reservation for the VMs, and availability of DPU Virtual Functions (VF) on the host.

If one does not want to adhere to the restrictions needed for the fully accelerated (VM-direct) mode, the DPU can also be used with EDP in the default mode. The default mode also provides DPU-based offload and acceleration for network processing, but also incurs some CPU overhead on the host for tagging packets. This mode should also free up some CPU cores in comparison to using a non-DPU standard NIC, but it will not free up as many cores as in the fully accelerated mode. In fact, using vSphere vMotion for the VM-direct mode is achieved by automatically switching a VM to default mode temporarily during vMotion, and then back to VM-direct mode on the destination host, provided the destination also has a SmartNIC VF available; this enables access to vMotion and related operations in VM-direct mode in a way that appears seamless to the application owner. Note that using EDP with either VM-direct or default mode requires the NSX Manager to configure networking. Moreover, using VM-direct may have additional licensing requirements. Contact a VMware representative for more details.

vSphere Distributed Services Engine can be deployed on a flat network (all nodes on the network segment can see each other) or with an overlay network tunnelling protocol such as Geneve and VXLAN. Overlay networking allows the creation and management of virtual Layer 2 networks that can span different Layer 3 subnets. They allow a set of nodes to act as if they had their own dedicated network that only they can access, and many such virtual networks can co-exist on one physical network. As such, overlay networking is very useful for tenant separation and enhanced security in private or hybrid cloud deployments, but not all vSphere deployments require overlay networking.

Testing Goals

BWI and VMware set up joint benchmark testing to explore and demonstrate the DPU performance on vSphere with NSX Distributed Firewall. One of the main goals is to ensure that the DPU enables running network throughput close to line rate even with thousands of NSX Distributed Firewall functions activated. We also include some simulated load on host to mimic other workloads and application logic that can run alongside this network test and showcase how that does not degrade the performance of NSX functions which are running on the DPU. The NSX Distributed Firewall policies were tested from Layer 4 rules up to Layer 7 application rules. We compare all the performance results using the DPU in default and accelerated modes with that of a comparable regular NIC using standard datapath and enhanced datapath (EDP) standard modes on the host.

Experiment Setup

Testbed Configuration

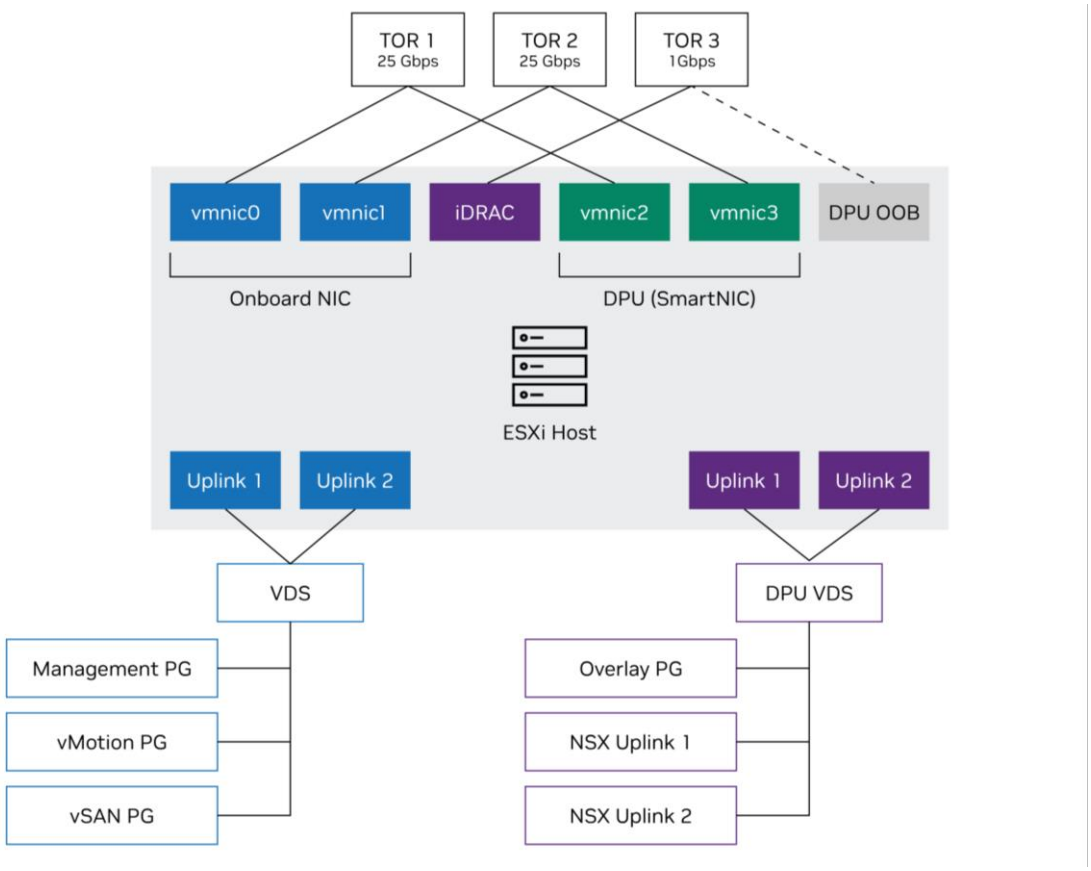
We used the first generation of “Monterey-ready” servers from Dell: PowerEdge R750 (see Figure 5) with one Intel® Xeon® Silver 4316 CPU running 40 physical cores at 2.30 GHz and 256 GB of main memory. With hyperthreading enabled, ESXi has 80 logical processor cores (l-cores). These servers have a standard NIC for management and vSAN traffic, but for this test, we only focus on the NSX distributed vSwitch (DVS) running on top of the 25 GbE BlueField-2 DPU. The servers each have a third NIC (NVIDIA ConnectX-5, also running at 25 GbE) that we use as a baseline to compare a traditional vSphere networking scenario against one with NVIDIA DPU offloads.

The equipment was hosted in an isolated environment with a dedicated vCenter and NSX Manager.

Figure 5. Testing Hardware: 3 x DELL Power Edge R750 S

Every ESXi Host (see Figure 6) is configured with a DPU (2 Ports at 25 Gbps) and an onboard standard NIC (2 Ports with 25 Gbps). An onboard 1 Gbps port for the iDRAC and a 1 Gbps port for the DPU management are also available. The iDRAC stands for "Integrated Dell Remote Access Controller" and is necessary for out-of-band management. In the vSphere 8 versions studied in this paper, only one DPU is supported per ESXi Host.

Figure 6. ESXi Network Architecture



One VDS (VMware vSphere® Distributed Switch™) is used per network card. All administration traffic—like management, vMotion and vSAN—is running over the traditional NIC card (see Figure 6 marked in light blue). All production traffic with NSX uses the DPU card, overlay and uplink VLANs for north south communication.

Workload Used for Testing

The open source iPerf Tool was used for measuring network bandwidth and performance. It allows users to test the speed, throughput, and quality of a network connection by sending and receiving data packets between two endpoints. iPerf operates in a client-server model, where one computer acts as the server, and another computer (or multiple computers) acts as the client(s) that initiate the network tests. iPerf version 3 was used with the default TCP mode.

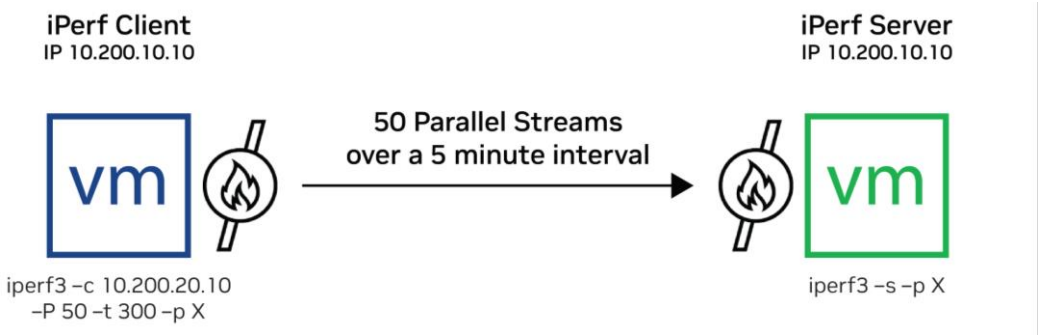
Testing Methodology

The following four modes are used for testing:

- **Regular NIC (Standard DP)** – A comparable regular NIC is used in the standard datapath mode.
- **Regular NIC (EDP)** – A comparable regular NIC is used in enhanced datapath (EDP) standard mode.
- **DPU (Default mode)** – DPU used in default mode. Significant portion of network/security processing tasks offloaded to DPU and allows workload mobility using vMotion.
- **DPU (Full Acceleration Mode)** - DPU is used in VMDirectPath mode – Passthrough like performance with workload mobility using vMotion capabilities and most of the network/security processing tasks offloaded to DPU.

We used the iPerf3 tool to drive network load between two VMs running a Linux operating system. The green VM in - shows the iPerf server VM, where the server process is started with the command iperf3 -s -p X, X stands for a randomly chosen port. The blue VM shows the iPerf client VM. The command iperf3 -c 10.200.20.10 -P 50 -t 300 -p X in the client establishes a communication over TCP with 50 parallel streams over a 5-minute (300 seconds) interval. While running the experiments, we recorded the maximum network bandwidth that we were able to achieve between the client and the server. We also recorded the CPU utilization on the source and destination hosts for network processing and other diagnostics data to understand the value-add from using a DPU. We scaled the number of iperf3 processes to four.

Figure 7. IPerf3 Server-Client Architecture



Benchmark Results

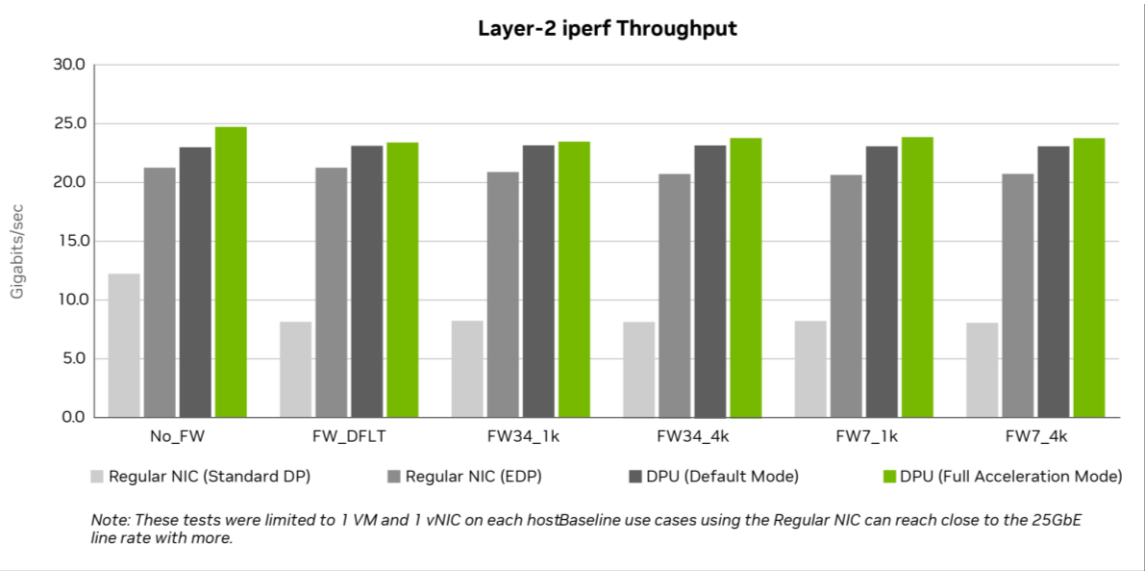
Iperf Throughput

For the IPerf throughput tests, we keep the focus on VMs placed on the same L2 overlay network segment as the overall trends are similar for other cases like when the VMs are placed on separate overlay segments. We focus on the cases with default rule, 1000 and 4000 firewall rules targeting both Layer 3/4 and Layer 7 to implement them as shown in Table 1. Figure 8 shows the network bandwidth achieved (higher is better) between the client and the server VMs for each of the cases from Table 1.

Table 1. Layer-2 iPerf Throughput Tests With Varying Number of Firewall Rules Applied at L3/4 and L7 Layers

Test Name	Number of Firewall Rules	Firewall Layer in the OSI Stack
No Firewall	0	N/A
FFW_DFLT	1	Layer 3/4
FW34_1k	1000	Layer 3/4
FW34_4k	4000	Layer 3/4
FW7_1k	1000	Layer 7
FW7_4k	4000	Layer 7

Figure 8. Layer-2 iPerf Throughput with Varying Number of Firewall Rules Applied at L3/4 and L7 Layers

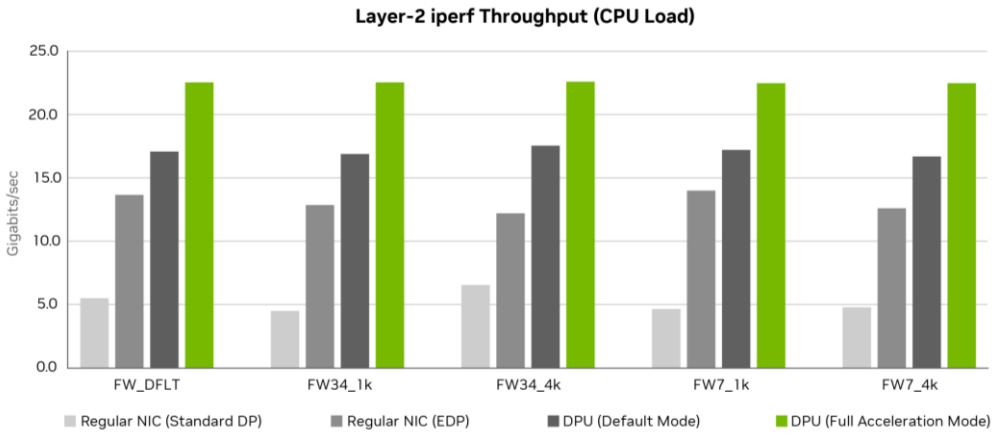


One can notice how the DPU outperforms both the baseline cases using a comparable regular NIC configured in standard datapath mode and a regular NIC with enhanced datapath (EDP) standard modes, by achieving higher bandwidth close to line rate. Also, even as we increase the number of firewall rules, the DPU in both default and full acceleration modes continues to provide maximum performance, close to line rate. Since the distributed firewall system works by building a rule lookup tree, even with an increasing number of firewall rules, the number of lookups in the tree may not increase a whole lot. This results in only modest variations in performance with an increasing number of rules, both in the case of the regular NIC and with the DPU. To keep the testing simple, we use a single VM on each host with default settings for the vNICs and hence don't get close to line rate for the Regular NIC case in standard mode. If we had used higher values for VMX settings such as ethernetX.pNicFeatures on the VM or scaled out the number of VMs or number of vNICs on each VM, we should have been able to effectively scale up regular NIC throughput to fully utilize a 25G card but at the cost of using more CPUs on the host.

One of the most important advantages of using a DPU to offload and accelerate network and security services is that it frees up the host CPUs to provide better performance for the applications in isolation from infrastructure workloads. Application processing, when isolated from other hypervisor tasks like networking and security, often enjoys a better quality of service from the CPU cores by eliminating interference factors like cache pollution. In addition, running the network and security tasks on the DPU also means they are not affected by any application processing CPU load on the host.

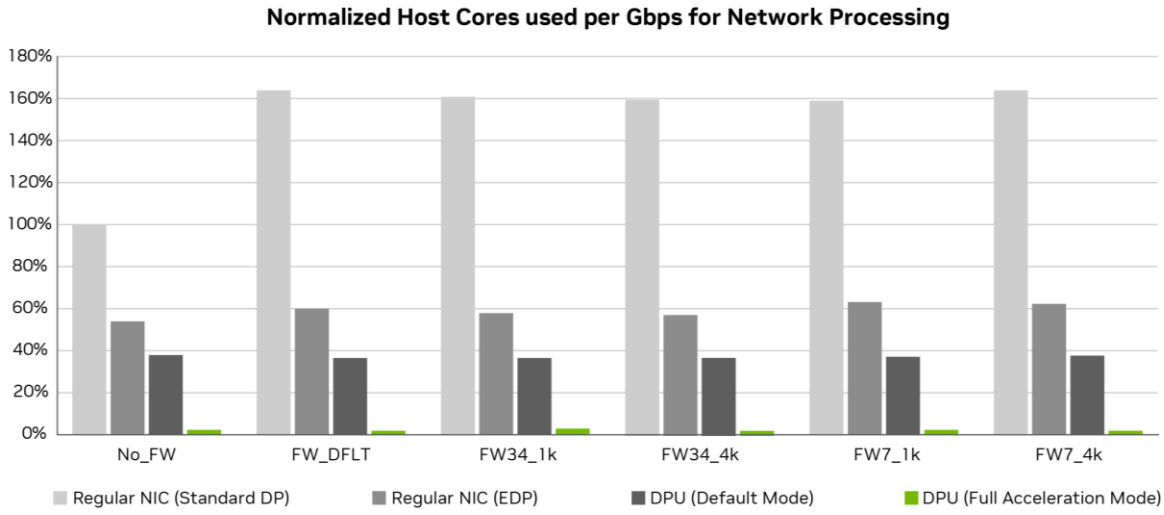
To showcase this better, we simulate an application load on the host to drive host core utilization to 100% and showcase how network processing gets slowed down for the firewalled cases listed in Table 1. Figure 9 shows these results where one can notice how the CPU load on the host slows down network processing for each of the cases studied. When compared to the bandwidth achieved as shown in Figure 8, the DPU in the accelerated mode continues to achieve the same maximum performance close to line rate even when there is high CPU load on the host. Whereas the cases with the regular NIC in standard and enhanced datapath modes have much larger performance impacts since the network tasks have to contend with the application on the host for CPU cycles. The DPU in default mode offloads a good part of the host side network and security processing to the DPU and hence fares better than the regular NIC cases, though it experiences a larger drop in performance than when using the DPU in full acceleration mode. At high CPU load, the regular NIC standard DP case with 4k L3/L4 rules shows higher performance than the same with 1k L3/L4 rules. This could be due to the baseline case being sensitive to any variations from CPU scheduling decisions on the host at high load.

Figure 9. Layer-2 iPerf Throughput Alongside CPU Load with Varying Number of Firewall Rules Applied at L3/4 and L7 Layers



While achieving better performance for network and security tasks is certainly an advantage in case of the DPU, the CPUs saved by offloading to the DPU can be used to process a bigger application workload scale and better application performance (like lower transaction latencies) using the same hardware on the host due to the lack of interference from network and security tasks. Figure 10 shows the amount of host CPU used for every Gigabit of network processing for each of the cases studied. One can notice that the CPU overheads increase for the regular NIC cases when we go from No_FW to the firewalled cases on the host. On the other hand, the DPU completely hides the overhead from firewall processing from the host CPUs. Overall, the DPU in default mode uses less CPU than the regular NIC cases and the DPU full acceleration mode (green bar) has close to zero host CPU used. This can result in better workload consolidation in the datacenter saving hardware, software licensing, energy expenditures and lower TCO over time.

Figure 10. Normalized host Used for Network Processing per Gbps of Data Transmitted an L2 Network (CPU utilization with regular NIC and no firewall normalized as 100%)



Summary

The tests show that a high load on the x86 platform does not influence the NSX services which are running on the DPU. The DPU is able to achieve close to line rate performance and sustains the same even when thousands of firewall policies are included in the environment.

For BWI the performance tests show an impressive result in accelerating network performance and freeing up CPU resources on the x86 ESXi hosts. The figures show how using the DPU in accelerated mode help us achieve much better performance in comparison to using a regular NIC. BWI assumes that further tests for database systems, storage traffic, or applications will also deliver corresponding acceleration.

From BWI business perspective the DPU with the network acceleration and the additional CPU resources on the x86 ESXi host brings a significant advantage. BWI could fulfil the separation of their information and security domain with DPU and NSX security.

References

To learn more about vSphere on DPUs, visit https://www.vmware.com/products/vsphere/distributed-services-engine.html

To learn more about the NVIDIA BlueField DPU, visit https://www.nvidia.com/en-us/networking/products/data-processing-unit/

About the Authors

Martin Schulz – Strategic Technology Advisor, BWI

Marcus Schoen – BWI Cloud Architect, BWI

Joerg Roesch – Lead Solution Engineer Network and Security, VMware

Karthik Ganesan – Lead Performance Engineer, VMware

John F. Kim – Director of Storage Networking, NVIDIA

Jose Castano – System Software Principal Engineer, NVIDIA

Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation (“NVIDIA”) makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer (“Terms of Sale”). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer’s product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA accepts no liability related to any default, damage, costs, or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, “MATERIALS”) ARE BEING PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, ConnectX,and BlueField are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

VESA DisplayPort

DisplayPort and DisplayPort Compliance Logo, DisplayPort Compliance Logo for Dual-mode Sources, and DisplayPort Compliance Logo for Active Cables are trademarks owned by the Video Electronics Standards Association in the United States and other countries.

HDMI

HDMI, the HDMI logo, and High-Definition Multimedia Interface are trademarks or registered trademarks of HDMI Licensing LLC.

Arm

Arm, AMBA, and Arm Powered are registered trademarks of Arm Limited. Cortex, MPCore, and Mali are trademarks of Arm Limited. All other brands or product names are the property of their respective holders. ”Arm” is used to represent Arm Holdings plc; its operating company Arm Limited; and the regional subsidiaries Arm Inc.; Arm KK; Arm Korea Limited.; Arm Taiwan Limited; Arm France SAS; Arm Consulting (Shanghai) Co. Ltd.; Arm Germany GmbH; Arm Embedded Technologies Pvt. Ltd.; Arm Norway, AS, and Arm Sweden AB.

OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com

Copyright