TECHNOLOGY GUIDE

Intel Corporation

Open vSwitch – Optimized Deployment Benchmark

Authors

Ian Stokes

Marko Kovacevic

Harry van Haaren

Cian Ferriter

Abhishek Khade

1 Introduction

This benchmark report presents the throughput performance improvements achievable for various deployments and configurations of Open vSwitch (OVS) with the Data Plane Development Kit (DPDK) for a given use case.

The configuration and deployments demonstrate some of the latest Intel® architecture technology advancements, such as migration between 2nd and 3rd Gen Intel® Xeon® Scalable processors, which include additional CPU capabilities such as Intel® Speed Select Technology – Base Frequency (Intel® SST-BF), Intel® Advanced Vector Extensions 512 (Intel® AVX-512) software optimizations within both OVS and DPDK, as well as external software optimizations that impact OVS throughput performance, such as Virtual I/O Device (VIRTIO) version 1.1. In addition to the features mentioned, further upcoming Intel® AVX-512 optimizations are also demonstrated.

The performance improvements for these features are presented both individually as well as collectively. An outline of future performance enhancements in both hardware and software that are not yet merged as part of the OVS codebase are also discussed.

This guide is written for network administrators who want to use OVS DPDK with Intel® Architecture-based deployments, specifically to demonstrate the performance gains expected with the features highlighted. It is not intended as a user configuration guide.

This document is part of the Network Transformation Experience Kit, which is available at https://networkbuilders.intel.com/network-technologies/network-transformation-exp-kits.

NOTE: The general information contained within this document applies to both the 3rd Generation Intel® Xeon® Scalable processor and the Intel® Xeon® D processor. However, please note that the performance, configurations, and feature set in this document apply specifically to the 3rd Gen Intel® Xeon® Scalable processor and may vary for the Intel® Xeon® D processor.

Table of Contents

Figures

Tables

Document Revision History

REVISION	DATE	DESCRIPTION
001	April 2021	Initial release.
002	February 2022	Added a note regarding Intel® Xeon® D processor in the Introduction section.

1.1 Terminology

Table 1. Terminology

ABBREVIATION	DESCRIPTION
DPDK	Data Plane Development Kit (DPDK)
DPIF	DataPath InterFace, OVS datapath component that represents the software datapath as a whole
DPCLS	Datapath Classifier, OVS software datapath component that performs wildcard packet matching
EMC	Exact Match Cache, OVS software datapath component that performs exact packet matching
IA	Intel Architecture
IP	Internet Protocol
MFEX	Miniflow Extract
NIC	Network Interface Card
OVS	Open vSwitch
PMD	Poll Mode Driver
SMC	Signature Match Cache, OVS software datapath component that performs wildcard packet matching
SST	Intel® Speed Select Technology
SST-BF	Intel® Speed Select Technology – Base Frequency (Intel® SST-BF)
TEP	Tunnel Endpoint
VXLAN	Virtual Extensible LAN

1.2 Reference Documentation

Table 2. Reference Documents

REFERENCE	SOURCE
Applying SIMD Optimizations to the OVS Datapath Classifier (OVS Conference, 2018)	https://youtu.be/5-MDlpUIOBE
Next steps for higher performance of the software data plane in OVS (OVS Conference, 2019)	https://youtu.be/x0bOpojnpmU
Next steps for even higher performance of the SW Datapath in OVS (OVS Conference, 2020)	https://youtu.be/5dWyPxiXEhg
Intel® Xeon® Gold 6252N Processor	https://ark.intel.com/content/www/us/en/ark/products/193951/intel-xeon-gold-6252n-processor-35-75m-cache-2-30-ghz.html
3rd Generation Intel® Xeon® Scalable Processors	https://ark.intel.com/content/www/us/en/ark/products/series/204098/3rd-generation-intel-xeon-scalable-processors.html
Intel® Speed Select Technology – Base Frequency - Enhancing Performance Application Note	https://builders.intel.com/docs/networkbuilders/intel-speed-select-technology-base-frequency-enhancing-performance.pdf
Intel® Speed Select Technology – Base Frequency Priority CPU Management for Open vSwitch (OVS) Application Note	https://builders.intel.com/docs/networkbuilders/intel-speed-select-technology-base-frequency-priority-cpu-management-for-open-vswitch.pdf
What's new in VIRTIO 1.1	https://www.dpdk.org/wp-content/uploads/sites/35/2018/09/virtio-1.1_v4.pdf
Virtual I/O Device (VIRTIO) Version 1.1	https://docs.oasis-open.org/virtio/virtio/v1.1/csprd01/virtio-v1.1-csprd01.html
OVS Datapath Classifier Performance	https://docs.openvswitch.org/en/latest/topics/dpdk/bridge/#datapath-classifier-performance
OVS Enabling AVX-512	https://docs.openvswitch.org/en/latest/intro/install/dpdk/#possible-issues-when-enabling-avx512
ICE Vector PMD	https://doc.dpdk.org/guides-20.11/nics/ice.html#vector-pmd
Introducing the Intel® Data Streaming Accelerator (Intel® DSA)	https://01.org/blogs/2019/introducing-intel-data-streaming-accelerator
OVS-DPDK Datapath Classifier	https://software.intel.com/content/www/us/en/develop/articles/ovs-dpdk-datapath-classifier.html

REFERENCE	SOURCE
DPIF Framework + Optimizations	https://patchwork.ozlabs.org/project/openvswitch/cover/20210212171718.2189798-1-harry.van.haaren@intel.com/
Intel® AVX-512 – Packet Processing with Intel® AVX-512 Instruction Set Solution Brief	https://networkbuilders.intel.com/solutionslibrary/intel-avx-512-packet-processing-with-intel-avx-512-instruction-set-solution-brief
Intel® AVX-512 – Instruction Set for Packet Processing Technology Guide	https://networkbuilders.intel.com/solutionslibrary/intel-avx-512-instruction-set-for-packet-processing-technology-guide
Intel® AVX-512 – Writing Packet Processing Software with Intel® AVX-512 Instruction Set Technology Guide	https://networkbuilders.intel.com/solutionslibrary/intel-avx-512-writing-packet-processing-software-with-intel-avx-512-instruction-set-technology-guide

2 Overview

Open vSwitch (OVS) is a production quality, multilayer virtual switch licensed under the open source Apache 2.0 license. OVS can operate as a soft switch running within the hypervisor and as the control stack for switching silicon. It has been ported to multiple virtualization platforms and switching chipsets.

The Data Plane Development Kit (DPDK) consists of libraries to accelerate packet processing workloads running on a wide variety of CPU architectures. Designed to run on x86, POWER, and ARM processors, it runs mostly in Linux userspace, with a FreeBSD port available for a subset of DPDK features. DPDK is licensed under the Open Source BSD License.

Official support and integration for OVS with DPDK has been available since OVS 2.7 with the view of provisioning a high performance I/O alternative to standard Linux network devices.

Over several releases, much work has been completed within OVS and DPDK to improve switching performance and throughput by taking advantage of the latest hardware and software features available.

By default, these optimizations and features may not be enabled for vanilla deployments of OVS with DPDK due to requirements such as specific Intel hardware and, as such, it may not be clear to users the advantages they provide.

This benchmark report discusses a typical complex OVS with DPDK deployment that consists of north/south network traffic via a network interface card (NIC), traffic isolation via logical port tunnel types to provide encapsulation/decapsulation of VXLAN, as well as virtual back end interfaces to receive/transmit traffic back out of the OVS deployment.

As part of the benchmark, multiple configurations of the deployments are benchmarked demonstrating:

- Performance gain between 2nd and 3rd Gen Intel® Xeon® Scalable processors
- Throughput improvement using Intel® Speed Select Technology – Base Frequency (Intel® SST-BF)
- Intel® Advanced Vector Extensions 512 (Intel® AVX-512) for software optimization for increased workload

This report highlights where this technology is already in place and available for use by users. A high level summary of each technology is provided and highlighted in each deployment configuration.

The paper also provides performance numbers for proposed Intel® AVX-512 optimizations currently being considered by the OVS community but not yet merged as part of the OVS codebase.

Finally, consideration is given to future IA technology that will help drive higher performance, such as the Intel® Data Streaming Accelerator (Intel® DSA) and further Intel® AVX-512 software optimizations in OVS itself.

3 Topology

The test topology has interfaces representing 32 VMs with one million flows. The test runs on four physical cores for OVS and 10 hyper-threaded cores for TestPMD and works in a north-south workflow.

The Ixia traffic generator was configured to have an acceptable loss rate of 0.1% for received traffic. Traffic is generated and sent from the Ixia traffic generator and is received by the Intel® Ethernet Network Adapter E810-C 100G. The profile of the traffic as received at the NIC is as follows: **Ether()/IP()/UDP()/VXLAN()/Ether()/IP()**

After the packet is received, it traverses the physical bridge into the first Virtual Extensible LAN (VXLAN) Tunnel Endpoint (TEP), at which point it is decapsulated. After the packet has been decapsulated, the inner packet is then transmitted to the appropriate Vhost port. The DPDK TestPMD sample application is used to emulate the VIRTIO back end (in this case representing a guest). After this VIRTIO port receives a packet, it transmits the packet back to the Vhost port in OVS. The same path is followed by the packet but in this case southbound. The packet is encapsulated on the VXLAN port before being transmitted out of the virtual switch via the NIC. This operation is completed for each of the 32 VIRTIO user ports created.

This test uses two types of traffic profiles, burst and scatter.

- BURST:
  On the outer IP, the source address remains the same for 32 instances before changing by 0.1 for 1024 times, while the destination remains the same. For the inner IP, both destination and source IPs are incremented by 0.1 for a total of 1,048,576, which creates the one million flows for the test.
- SCATTER:
  On the outer IP, the source address changes incrementally for 1-32, unlike the burst profile that uses the same IP for the 32 instances. The destination address remains the same on the outer IP. For the inner IP, the source IP remains the same while the destination address increments for a total of 1,048,576 times, which creates the one million flows for the test.

Important: All throughput has been recorded as received traffic at the Ixia traffic generator, i.e., traffic that egresses from the vSwitch itself. Note that the vSwitch is both transmitting and receiving this amount of traffic at any one moment during the tests. As such, at any one time the vSwitch is dealing with double the amount of data per second than what is being presented in the benchmarks in Section 6. If the vSwitch is transmitting 50,000 Mbps, it has a total switch load of 100,000 Mbps for both rx/tx operations.

Both exact match cache (EMC) and signature match cache (SMC) were disabled in OVS for the deployments. Due to the number of flows used in the deployments, the EMC and SMC can become saturated quickly. The cost of accessing/updating/maintaining these becomes overhead in itself. To avoid this overhead, it is simpler to disable their usage in the deployment and use the more flexible datapath classifier (DPCLS) that supports wildcarding.

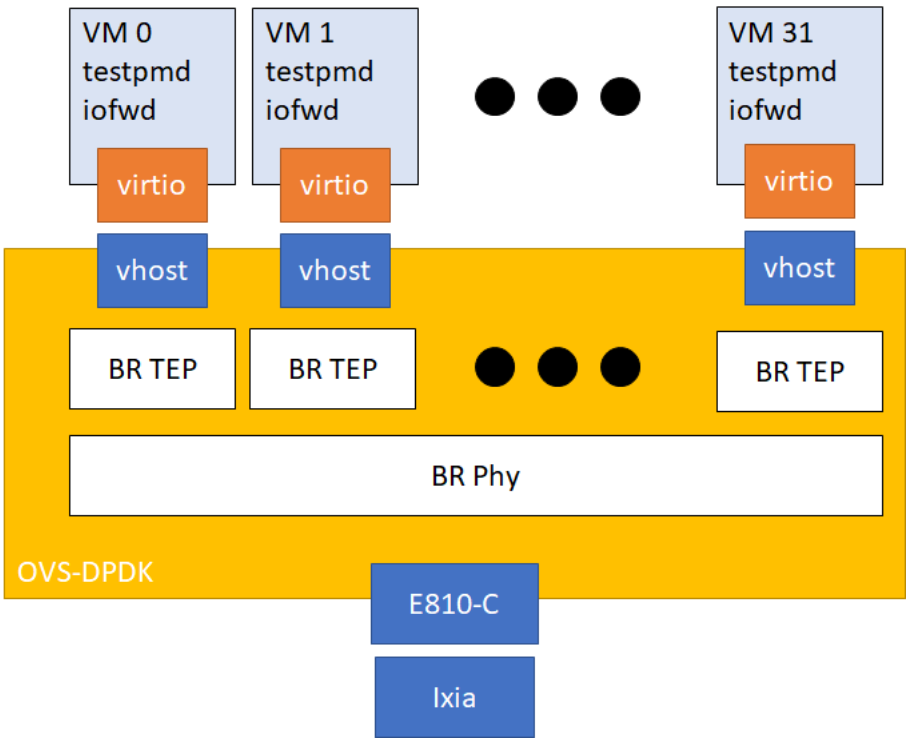

Figure 1. Topology Diagram

4 Ingredients

This section defines the hardware and software used in the benchmarks. At the time of writing some software has not been upstreamed to the Open vSwitch community as it is under review. This has been flagged with a link to the patch and review for the reader.

4.1 Hardware Bill of Materials

Table 3. Hardware Bill of Materials

NAME	CONFIG 1 (2ND GEN INTEL XEON SCALABLE PROCESSOR)	CONFIG 2 (2ND GEN INTEL XEON SCALABLE PROCESSOR WITH SST-BF)	CONFIG 3 (3RD GEN INTEL XEON SCALABLE PROCESSOR)	CONFIG 4 (3RD GEN INTEL XEON SCALABLE PROCESSOR WITH SST-BF)
Test By	Intel	Intel	Intel	Intel
Test Date	December 18, 2020	December 18, 2020	April 2021	April 2021
Platform	Supermicro X11DPG-QT	Supermicro X11DPG-QT	Intel Corporation Reference Platform*	Intel Corporation Reference Platform*
# Nodes	1	1	1	1
# Sockets	1	1	2 (Tested on Socket 1 only)	2 (Tested on Socket 1 only)
CPU	6252N	6252N	6338N	6338N
Cores/socket, Threads/socket	24, 48	24, 48	32, 64	32, 64
Microcode	0x5002f01	0x5002f01	0x8d0001f0	0x8d0001f0
Hyper-Threads	On	On	On	On
Turbo	Off	Off	Off	On
Power Management (disabled/enabled)	Disabled	Enabled (SST-BF)	Disabled	Enabled (SST-BF)
BIOS Version	American Megatrends Inc, Version: 3.3 dated February 21, 2020	American Megatrends Inc, Version: 3.3 dated February 21, 2020	SE5C6200.86B.0020.P22 .2103231313	SE5C6200.86B.0020.P22. 2103231313
System DDR Mem Config: slots / cap / speed	6 slots / 16 GB / 2933 (run at 2933)	6 slots / 16 GB / 2933 (run at 2933)	16 slots / 16 GB / 3200 (run at 2666)	16 slots / 16 GB / 3200 (run at 2666)
Total Memory/Node (DDR, DCPMM)	96, 0	96, 0	256, 0	256, 0
Storage - Boot	1x Intel 240 GB SSDOS Drive	1x Intel 240 GB SSD OS Drive	1x Intel 240 GB SSD OS Drive	1x Intel 240 GB SSD OS Drive
NIC	E810-C	E810-C	E810-C	E810-C
PCH	C620	C620	C621	C621
NIC Firmware	1.3.13.0	1.3.13.0	1.3.13.0	1.3.13.0
DDP Profile	2.15	2.15	2.15	2.15

*Intel® Reference Platform (RP) for 3rd Gen Intel® Xeon® Scalable Processor

4.2 Software Bill of Materials

Table 4. Software Bill of Materials

NAME	CONFIG 1 (2ND GEN INTEL XEON SCALABLE PROCESSOR)	CONFIG 2 (2ND GEN INTEL XEON SCALABLE PROCESSOR WITH SST-BF)	CONFIG 3 (3RD GEN INTEL XEON SCALABLE PROCESSOR)	CONFIG 4 (3RD GEN INTEL XEON SCALABLE PROCESSOR WITH SST-BF)
OS	Ubuntu 20.10	Ubuntu 20.10	Ubuntu 20.10	Ubuntu 20.10
Kernel	5.8.0-25-generic	5.8.0-25-generic	5.8.0-25-generic	5.8.0-25-generic
Workload & Version	OVS-2.15.0	OVS-2.15.0	OVS-2.15.0	OVS-2.15.0
Compiler	GCC 10.2.0	GCC 10.2.0	GCC 10.2.0	GCC 10.2.0
Other SW	DPDK-20.11.0	DPDK-20.11.0	DPDK-20.11.0	DPDK-20.11.0
Other SW	qemu-5.1.0	qemu-5.1.0	qemu-5.1.0	qemu-5.1.0
Operating Frequency	2.3 GHz	1.9 GHz, 2.8 GHz	2.2 GHz	1.9 GHz, 2.4 GHz
CPU Utilization	100% (active DPDK cores 4)	100% (active DPDK cores 4)	100% (active DPDK cores 4)	100% (active DPDK cores 4)

5 Deployment Details

This benchmark report uses multiple configurations of the deployment use case outlined in Section 3. Each section below outlines the specifics of a given configuration with a high-level overview of the technology involved in each case.

5.1 2nd Generation Intel® Xeon® Scalable Processor

This configuration deploys OVS with DPDK as outlined in Section 3 on a platform using a 2nd Generation Intel® Xeon® Scalable processor, specifically the Intel® Xeon® Gold 6252N processor. No extra configuration options or specific Intel technologies are enabled. The case presents a baseline of what is achievable with the previous generation of Intel server processors. For more information regarding 2nd Gen Intel® Xeon® Scalable processors, see https://ark.intel.com/content/www/us/en/ark/products/193951/intel-xeon-gold-6252n-processor-35-75m-cache-2-30-ghz.html.

5.2 3rd Gen Intel® Xeon® Scalable Processor

This configuration deploys OVS with DPDK as outlined in Section 3 on a platform using the 3rd Gen Intel Xeon Scalable processor, which at the time of writing represents the latest generation Intel® Xeon® 6338N Scalable processor. No extra configuration options or specific Intel technologies are enabled. The case presents a baseline of what is achievable with the latest generation of Intel server processors for as-delivered deployments. For more information regarding 3rd Generation Intel® Xeon® Scalable processors, see https://ark.intel.com/content/www/us/en/ark/products/series/204098/3rd-generation-intel-xeon-scalable-processors.html.

5.3 2nd Gen Intel® Xeon® Scalable Processor + Intel® SST-BF

The configuration described in this setup uses a CPU capability called Intel® Speed Select Technology – Base Frequency (Intel® SST-BF), which is available on select SKUs of 2nd Gen Intel® Xeon® Scalable processors (5218N, 6230N, and 6252N). The placement of key workloads on higher frequency Intel® SST-BF enabled cores can result in an overall system workload increase and potential overall energy savings when compared to deploying the CPU with symmetric core frequencies. In this case the PMD coremask of OVS is set so that each PMD thread resides on a CPU with Intel® SST-BF enabled. For more information regarding the use of Intel® SST-BF or it's use with OVS DPDK, see the following application notes: https://builders.intel.com/docs/networkbuilders/intel-speed-select-technology-base-frequency-enhancing-performance.pdf and https://builders.intel.com/docs/networkbuilders/intel-speed-select-technology-base-frequency-priority-cpu-management-for-open-vswitch.pdf.

5.4 3rd Gen Intel® Xeon® Scalable Processor + Intel® SST-BF

The configuration described in this setup is similar to that described in Section 5.3 in that it uses a CPU capability called Intel® Speed Select Technology – Base Frequency (Intel® SST-BF) but it is instead deployed on a 3rd Gen Intel® Xeon® Scalable processor (6338N). With this setup we can easily compare the performance gain from 2nd Gen to 3rd Gen Intel Xeon cores.

5.5 3rd Gen Intel® Xeon® Scalable Processor with Intel® SST-BF + VIRTIO 1.1

This benchmark configuration builds upon on the previous configuration of the 3rd Gen Intel® Xeon® Scalable processor with SST-BF enabled and VIRTIO 1.1. VIRTIO is a mechanism to provide virtual devices to guests. In this case it is a net device that is being exposed to the TestPMD application (which takes the role of the guest). The original VIRTIO 1.0 specification has been improved

upon in VIRTIO 1.1 in several ways including but not limited to providing packed virtqueues and in-order device descriptor completion. This results in improved performance from the virtual back end for an OVS DPDK deployment as bottlenecks specific to the guest interface are reduced. For more information on VIRTIO 1.1, see https://www.dpdk.org/wp-content/uploads/sites/35/2018/09/virtio-1.1_v4.pdf and https://docs.oasis-open.org/virtio/virtio/v1.1/csprd01/virtio-v1.1-csprd01.html.

5.6 3rd Gen Intel® Xeon® Scalable Processor with Intel® SST-BF + DPCLS Intel® AVX-512

This benchmarking configuration builds on the previous configuration of 3rd Gen Intel® Xeon® Scalable processor N SKU with SST-BF enabled, enabling optimizations in the OVS wildcard matching engine known as DPCLS. The datapath classifier (DPCLS, https://software.intel.com/content/www/us/en/develop/articles/ovs-dpdk-datapath-classifier.html) is the component in the OVS datapath that matches incoming packets with the rules OVS has programmed. The complexity of the rules programmed in the DPCLS wildcard engine depend on the exact configuration; however, extensive compute is required in order to identify packets matching or missing on wildcard rules.

The enabled optimization in this deployment uses Intel® AVX-512 instructions to perform the required wildcard matching. This results in higher performance as fewer CPU cycles are required for rule matching, leaving more cycles for processing more packets. For further information regarding the use of Intel® AVX-512 instructions with OVS DPDK, see https://docs.openvswitch.org/en/latest/topics/dpdk/bridge/#datapath-classifier-performance and https://docs.openvswitch.org/en/latest/intro/install/dpdk/#possible-issues-when-enabling-avx512. For further information about Intel® AVX-512, see https://networkbuilders.intel.com/solutionslibrary/intel-avx-512-packet-processing-with-intel-avx-512-instruction-set-solution-brief, https://networkbuilders.intel.com/solutionslibrary/intel-avx-512-instruction-set-for-packet-processing-technology-guide, and https://networkbuilders.intel.com/solutionslibrary/intel-avx-512-writing-packet-processing-software-with-intel-avx-512-instruction-set-technology-guide.

5.7 3rd Gen Intel® Xeon® Scalable Processor with Intel® SST-BF + VIRTIO 1.1 + DPCLS Intel® AVX-512 + Intel® AVX-512 Vector ICE PMD

This benchmarking configuration builds on the previous, but also includes the Intel® AVX-512 optimized Vector PMD for ICE PMDs used by the Intel® Ethernet Network Adapter E810 in DPDK. By using Intel® AVX-512 in the PMD to handle NIC descriptor translation into the DPDK mbuf format it is possible to gain higher performance, as the Intel® AVX-512 instruction set provides wider registers to handle more packets in parallel in a single SIMD register. For more information about Vector ICE PMD, see https://doc.dpdk.org/guides-20.11/nics/ice.html#vector-pmd

5.8 3rd Gen Intel® Xeon® Scalable Processor with Intel® SST-BF + VIRTIO 1.1 + DPCLS Intel® AVX-512 + Intel® AVX-512 Vector ICE PMD + DPIF/MFEX Patches

This benchmark configuration builds on the previous configuration by including the DPIF and MFEX optimized patches using Intel® AVX-512. This means that the datapath interface (DPIF, or datapath glue code) is accelerated using Intel® AVX-512 to perform tasks like batching of packets. The MFEX (miniflow extract, or packet parsing code) is optimized using a “mask and match” approach, where specific traffic patterns are optimized. By combining the width of the Intel® AVX-512 SIMD registers and incoming packet headers, it is possible to probe for a single packet type in parallel. This avoids the branchy protocol-by-protocol parsing that is implemented in scalar code, and significantly reduces the instruction count to parse the specific packet type.

Note that the optimizations included in the benchmark configuration are not yet upstream in the OVS project; however, the patches used for this benchmark are publicly available here: https://patchwork.ozlabs.org/project/openvswitch/cover/20210212171718.2189798-1-harry.van.haaren@intel.com/.

6 Results

This section presents the benchmarks results and comparisons for different configurations and deployments as detailed in Section 5. Throughput is presented in megabits per second (Mbps) for four packet sizes, which are 112, 256, 512, and 1518 bytes respectively. In each section, short summary figures are provided as well as technical explanations for gains and differences. The aim of this section is to show performance gains of individual optimization features as well as to provide an overview of the performance when all features are enabled. The purpose here is to identify that large performance gains are not attributable to a single feature but are an amalgamation of many features, each providing a gain to give an overall optimized deployment.

6.1 2nd Gen Intel® Xeon® Scalable Processor

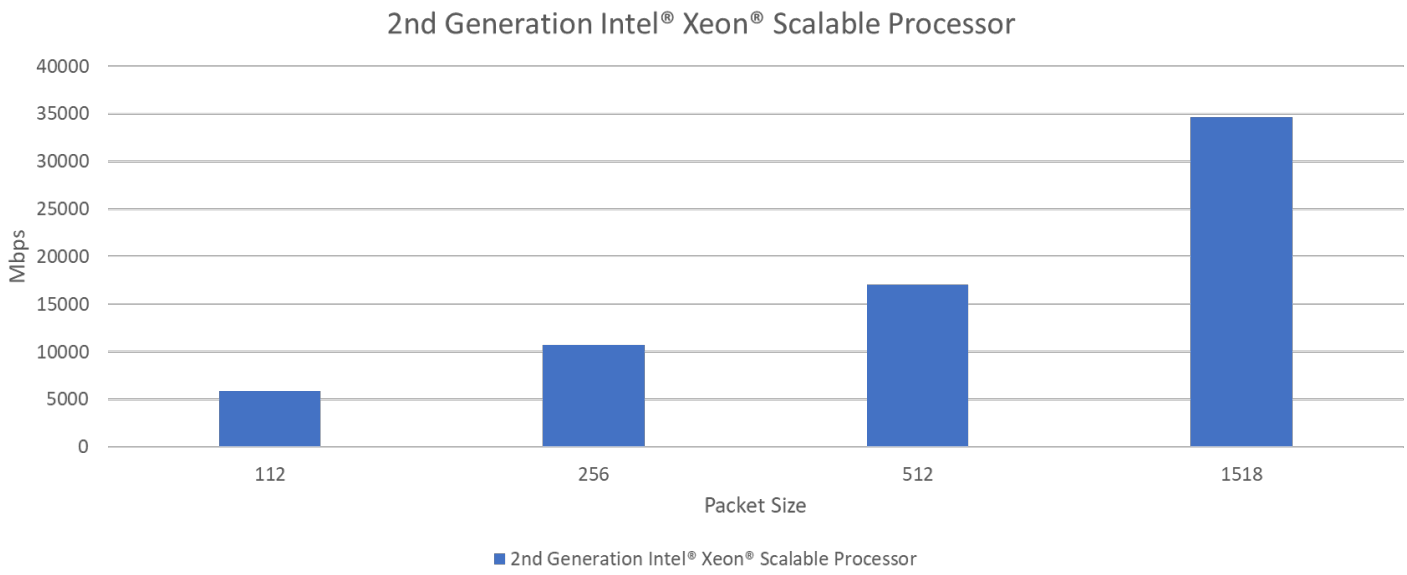

Figure 2 Throughput in Mbps Achievable for 2nd Gen Intel Xeon Scalable Processor

Figure 2 above represents the baseline achievable throughput with a 2nd Gen Intel® Xeon® Gold 6252N Scalable processor. Key data to note is that minimum achievable throughput is 5874 Mbps with 112-byte packets and maximum achievable throughput is 34,609 Mbps with packet size of 1518 bytes.

6.2 3rd Generation Intel® Xeon® Scalable Processor

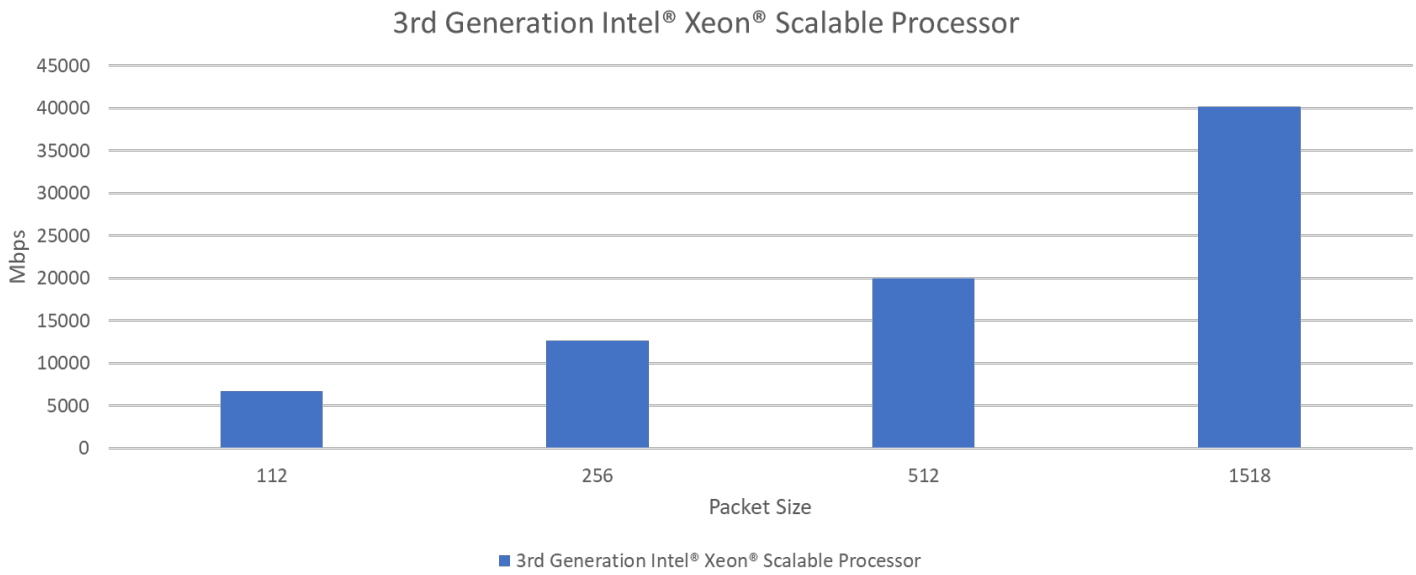

Figure 3 Throughput in Mbps Achievable for 3rd Gen Intel Xeon Scalable Processor

Figure 3 above represents the baseline achievable throughput with a 3rd Gen Intel® Xeon® 6338N Scalable processor. Key data to note is that minimum achievable throughput is 6704 Mbps with 112-byte packets and maximum achievable throughput is 40,155 Mbps with packet size of 1518 bytes.

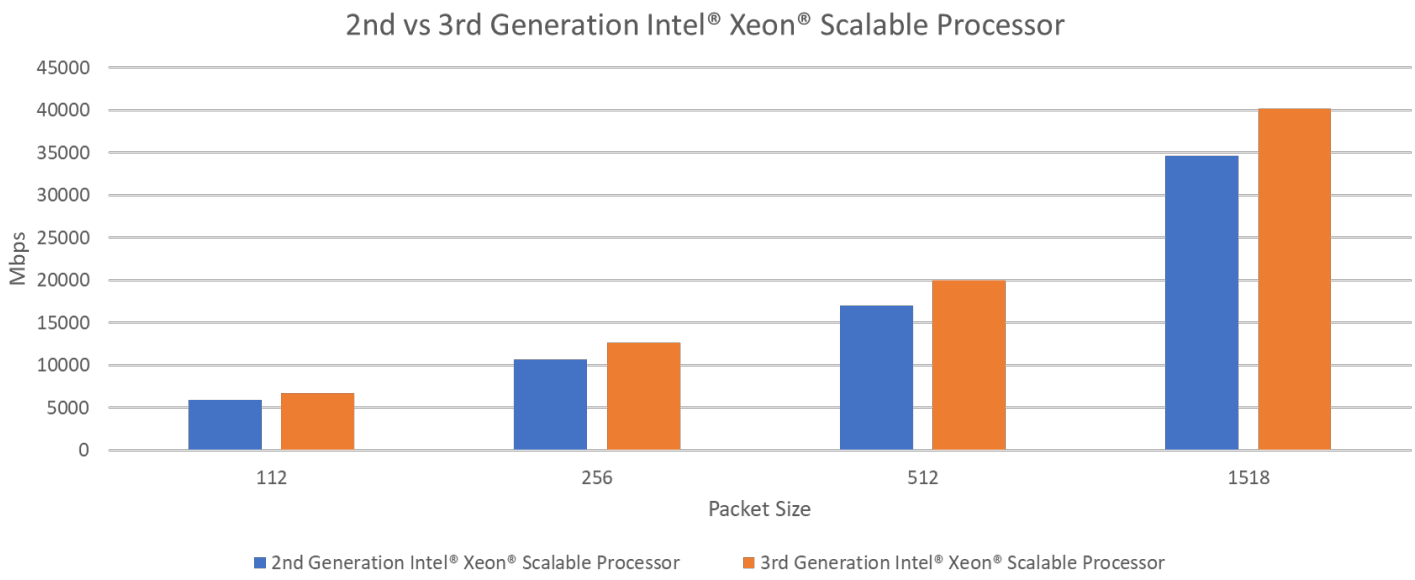

Figure 4 Comparison of Throughputs for 2nd Gen and 3rd Gen Intel Xeon Scalable Processors

Figure 4 above represents a comparison between of the baseline achievable throughput of 2nd and 3rd Generation Intel® Xeon® Scalable processors. Key data is represented in Table 5 below for easy comparison.

Table 5. Comparison of 2nd Gen Intel Xeon Scalable Processor and 3rd Gen Intel Xeon Scalable Processor Throughput and Performance Gain from Gen on Gen Migration

PACKET SIZE	2ND GEN INTEL XEON SCALABLE PROCESSOR MBPS	3RD GEN INTEL XEON SCALABLE PROCESSOR MBPS	PERFORMANCE GAIN
112	5874.95	6704.08	14%
256	10699.63	12601.33	17%
512	17031.28	19949.23	17%
1518	34609.43	40155.3	16%

From Table 5 above we can see a minimum performance gain of 14% for packet sizes of 112, 17% for packets of size 256 & 512, and finally 16% for 1518 byte packet size. This demonstrates on average 16% performance gain when migrating from the 2nd Gen Intel® Xeon® Gold 6252N Scalable processor to the 3rd Generation Intel® Xeon® 6338N Scalable processor.

6.3 2nd Gen Intel® Xeon® Scalable Processor + Intel® SST-BF

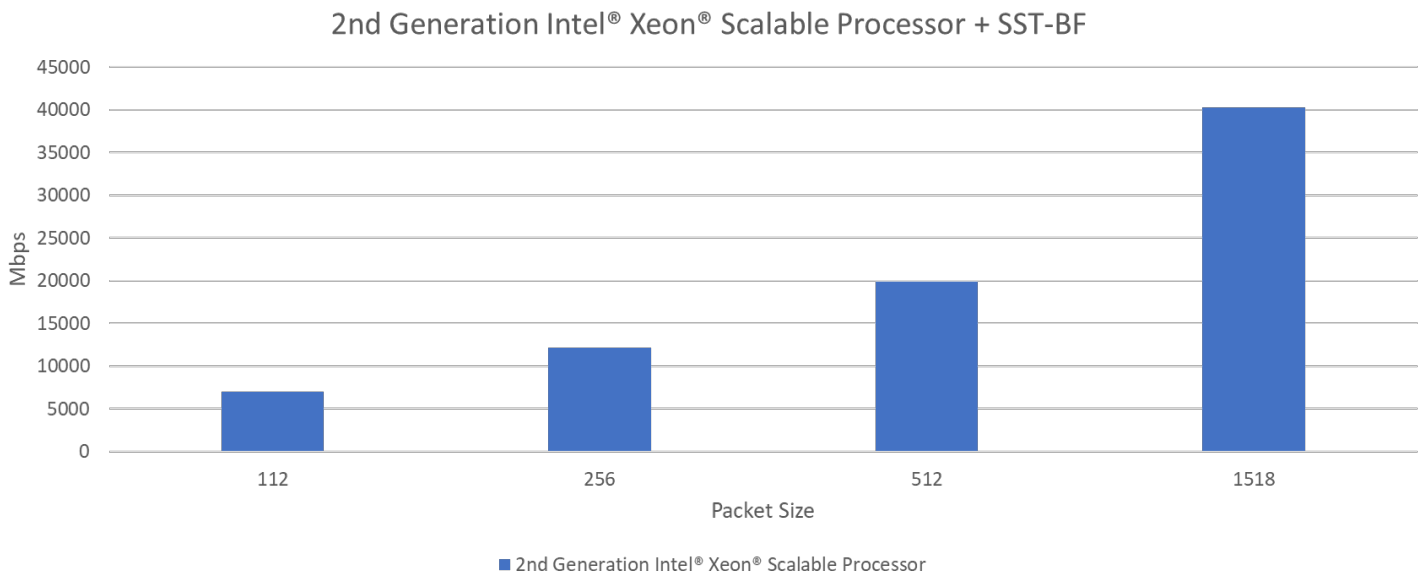

Figure 5 Throughput in Mbps achievable for 2nd Gen Intel Xeon Scalable Processor + Intel® SST-BF

Figure 5 above represents the throughput with a 2nd Gen Intel® Xeon® Gold 6252N Scalable processor and Intel® SST-BF enabled. Key data to note is that minimum throughput achievable with Intel® SST-BF enabled is 6977 Mbps with 112-byte packets and maximum achievable throughput is 40,234 Mbps with packet size of 1518 bytes. This roughly equals an average 16% gain in performance throughput on each packet size.

6.4 3rd Gen Intel® Xeon® Scalable Processor + Intel® SST-BF

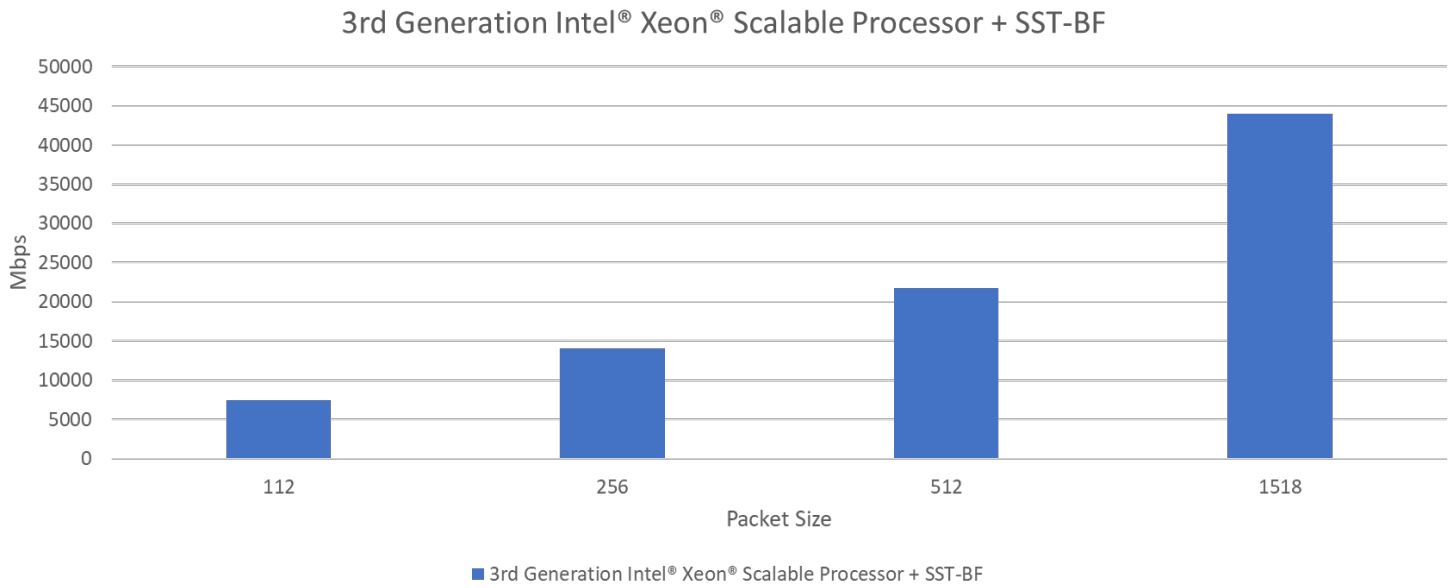

Figure 6 Throughput in Mbps achievable for 3rd Gen Intel Xeon Scalable Processor + Intel® SST-BF

Figure 6 above represents the throughput with a 3rd Gen Intel® Xeon® 6338N Scalable processor and Intel® SST-BF enabled. Key data to note is that minimum throughput achievable with Intel® SST-BF enabled is 7477 Mbps with 112-byte packets and maximum achievable throughput is 43925 Mbps with packet size of 1518 bytes. This roughly equals an average 10% gain in performance throughput on each packet size.

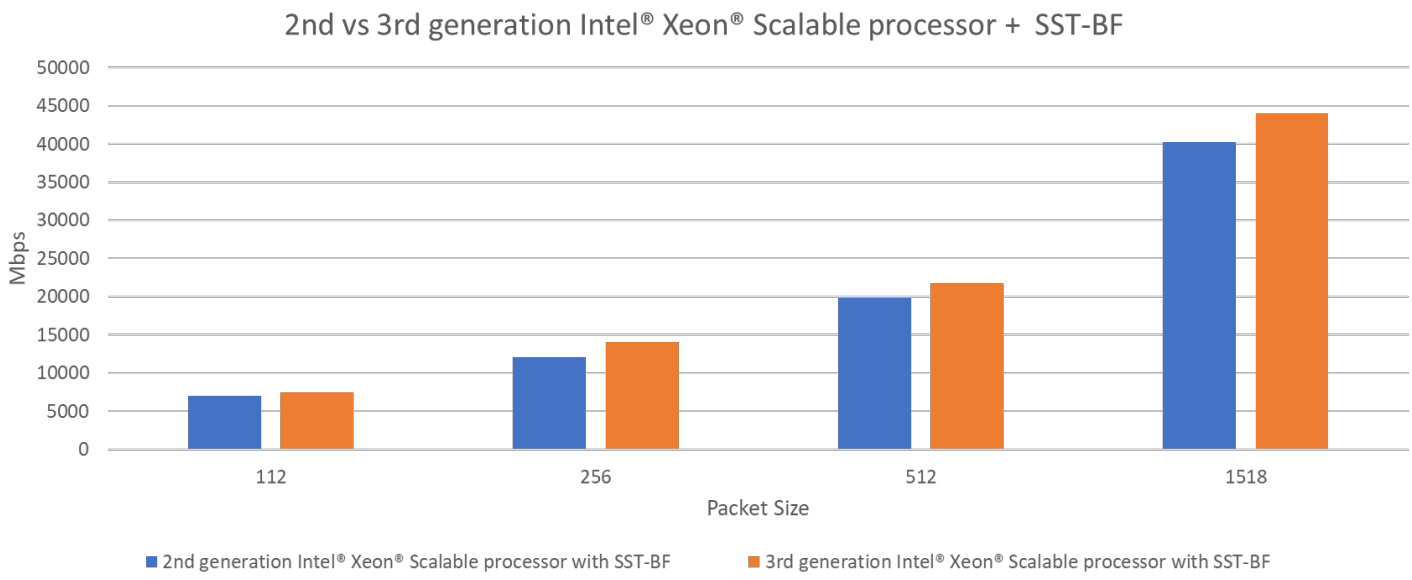

Figure 7 Comparison of 2nd and 3rd Gen Intel Xeon Scalable Processors, both + Intel® SST-BF

Figure 7 above represents a comparison between achievable throughput of both a 2nd and a 3rd Gen Intel® Xeon® Scalable processor with SST-BF enabled. Key data is represented in Table 6 below for easy comparison.

Table 6. Comparison of 2nd Gen Intel Xeon Scalable Processor and 3rd Gen Intel Xeon Scalable Processor Throughput when Using Intel SST-BF

PACKET SIZE	2ND GEN INTEL XEON SCALABLE PROCESSOR + SST-BF MBPS	3RD GEN INTEL XEON SCALABLE PROCESSOR + SST-BF MBPS	PERFORMANCE GAIN GEN ON GEN
112	6977.8	7477.52	7%
256	12109.39	14044.69	15%
512	19843.77	21786.13	9%
1518	40234.37	43925.78	9%

From Table 6 above we see that enabling SST-BF provides performance gains for both 2nd and 3rd Gen processors. The performance gain between the two processors narrows with SST-BF enabled. For the 2nd Gen Intel Xeon Scalable processor it averages about 16% throughput gain while with the 3rd Gen Intel Xeon Scalable processor it averages 10% throughput gain. It should be noted that the difference in throughput gain can be explained by the difference in CPU speed. The 2nd Gen Intel Xeon Scalable processor CPU benchmarked here has a base speed of 2.4 GHz and transitions to 2.8 GHz when enabled with SST-BF. The 3rd Gen Intel Xeon Scalable processor benchmarked in this case, however, has a base frequency of 2.2 GHz and transitions to 2.4 GHz with SST-BF enabled. It should be noted though that even with the higher clock speed for the 2nd Gen Scalable processor, the 3rd Gen Intel Xeon Scalable processor still provides higher throughput and achieves performance gains over the 2nd Gen Intel Xeon Scalable processor in each packet size, as highlighted above.

6.5 3rd Gen Intel® Xeon® Scalable Processor with Intel® SST-BF + VIRTIO 1.1

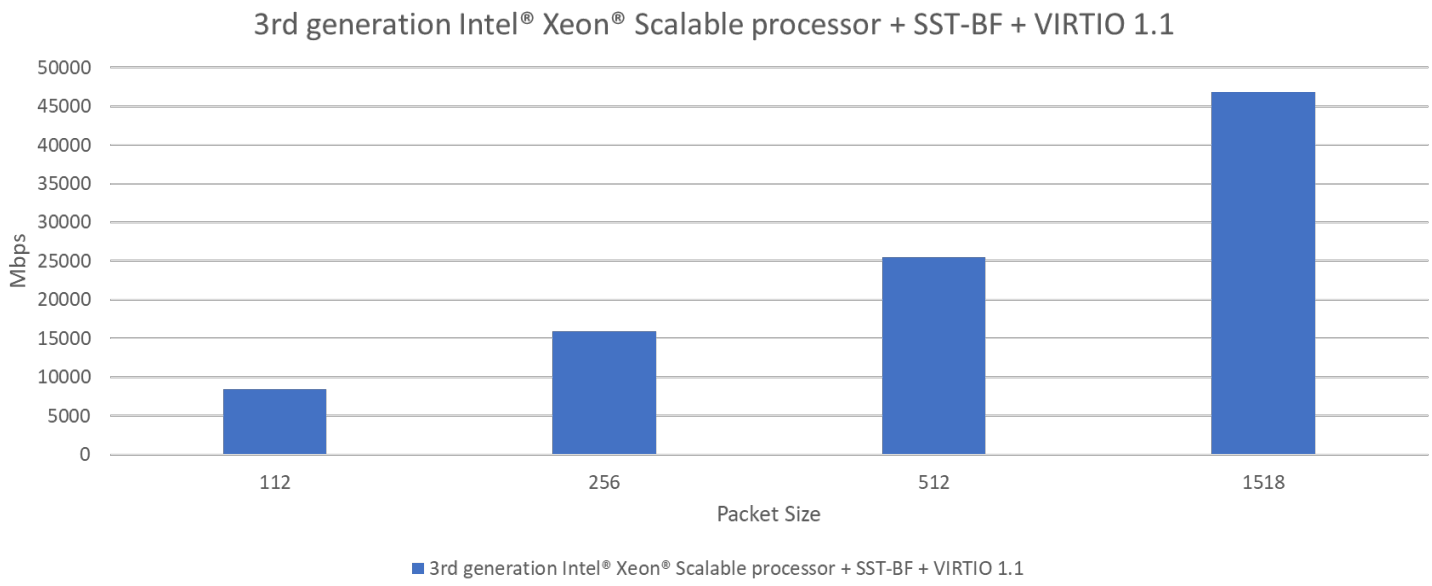

Figure 8 Throughput in Mbps Achievable for 3rd Gen Intel Xeon Scalable Processor + Intel® SST-BF + VIRTIO 1.1

Figure 8 above represents the throughput with a 3rd Gen Intel® Xeon® 6338N Scalable processor, Intel® SST-BF enabled, and VIRTIO 1.1 enabled for the VIRTIO back end. Key data to note is that minimum throughput achievable with both Intel® SST-BF and VIRTIO 1.1 enabled is 8,437 Mbps with 112-byte packets and maximum achievable throughput is 46,818 Mbps with packet size of 1518 bytes.

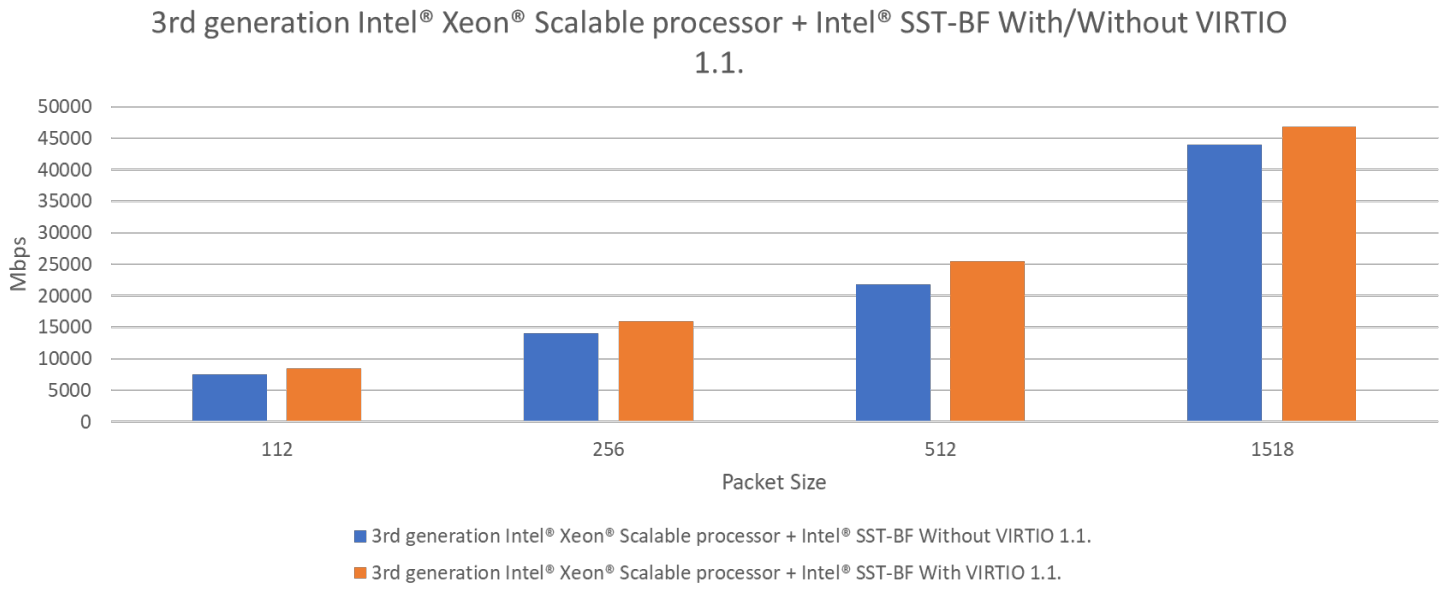

Figure 9 Comparison of 3rd Gen Intel Xeon Scalable Processor Throughput When Using Intel SST-BF With and Without VIRTIO 1.1.

Figure 9 above represents the throughput with a 3rd Gen Intel® Xeon® 6338N Scalable processor and Intel® SST-BF enabled **both with and without** VIRTIO 1.1 enabled for the VIRTIO back end. Key data is represented in Table 7 below for easy comparison.

Table 7. Comparison of 3rd Gen Intel Xeon Scalable Processor Throughput When Using Intel SST-BF With and Without VIRTIO 1.1.

PACKET SIZE	3RD GEN INTEL XEON SCALABLE PROCESSOR + SST-BF WITHOUT VIRTIO 1.1 MBPS	3RD GEN INTEL XEON SCALABLE PROCESSOR + SST-BF WITH VIRTIO 1.1 MBPS	PERFORMANCE GAIN
112	7477.52	8437.11	13%
256	14044.69	15888.68	13%
512	21786.13629	25459.4	17%
1518	43925.78	46818.15	9%

Performance gain from use of VIRTIO 1.1 can span from 13% to 17% for smaller packets. Larger packets do not see as large a performance gain but still see an improvement of 9%. A key item to identify here is that the VIRTIO 1.1 back end is not specific to OVS itself. It can itself become a bottleneck on the back end to the VM. Enabling the use of VIRTIO 1.1 can provide a boost across all packet sizes as a pure software optimization.

6.6 3rd Gen Intel® Xeon® Scalable Processor with Intel® SST-BF + VIRTIO 1.1 + DPCLS Intel® AVX-512 + Intel® AVX-512 Vector ICE PMD (Upstream Features)

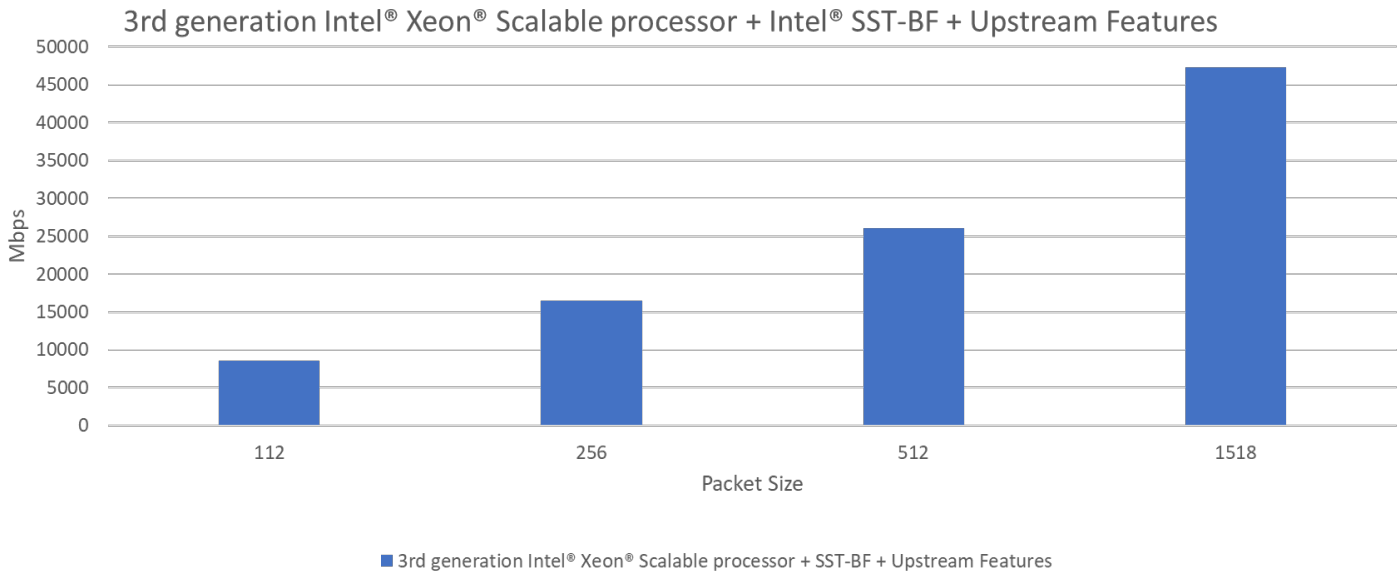

Figure 10 Throughput in Mbps Achievable for 3rd Gen Intel Xeon Scalable Processor + Intel® SST-BF + Available Upstream Features.

Figure 10 above represents the throughput with all available optimizations outlined so far as well as two extra optimizations in DPCLS Intel® AVX-512 and Intel® AVX-512 Vector ICE PMD E800 CVL PMD in DPDK. It should be noted, although there are only marginal gains in performance from what has been outlined previously, the DPCLS Intel® AVX-512 is a prerequisite for enabling future Intel® AVX-512 optimizations benchmarked later in the paper. As such, these results are provided as a baseline of what throughput performance can be expected with all available hardware and up-streamed software features for each project enabled. These figures are used in later comparisons.

6.7 3rd Gen Intel® Xeon® Scalable Processor with Intel® SST-BF + VIRTIO 1.1 + DPCLS Intel® AVX-512 + Intel® AVX-512 Vector ICE PMD + DPIF/MFEX Patches (Burst)

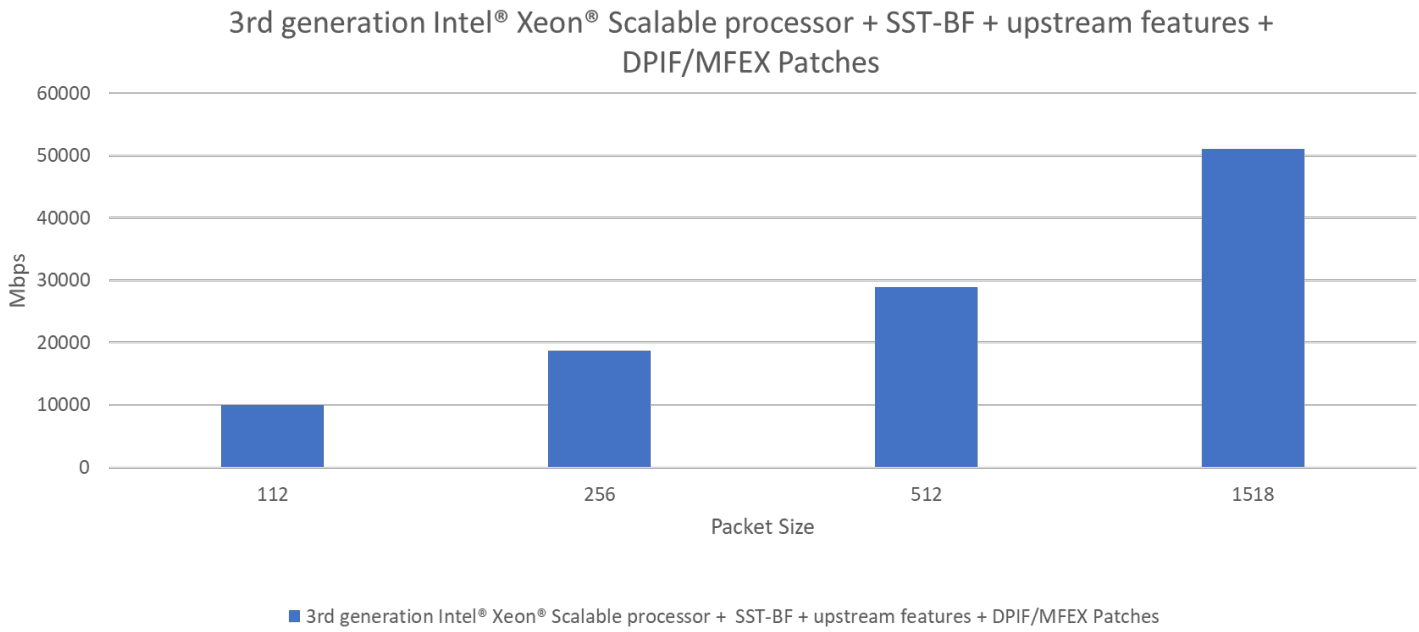

Figure 11 Throughput in Mbps Achievable for 3rd Gen Intel Xeon Scalable Processor + Intel® SST-BF + Available Upstream Features + DPIF/MFEX Patches

Figure 11 above represents the throughput with all available optimizations outlined so far as well two Intel® AVX-512 optimizations for OVS. The Intel® AVX-512 patches are applied to the datapath interface (DPIF) and miniflow extract mechanism of OVS. Note these patches have not been up streamed to OVS but, at the time of writing, are publicly available and under discussion with the OVS community. Key data to note is that the minimum throughput achievable with these patches applied on top of the existing outlined optimizations enabled is 9,991 Mbps with 112-byte packets and the maximum achievable throughput is 51,080 Mbps with packet size of 1518 bytes.

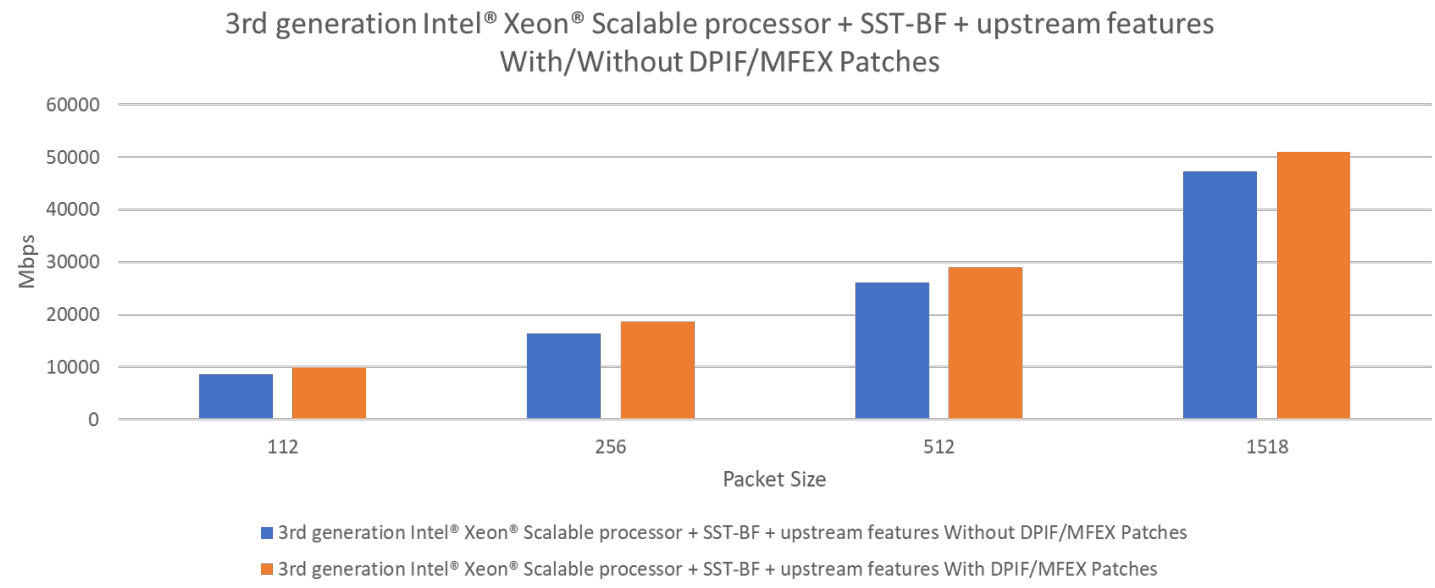

Figure 12 Comparison of 3rd Gen Intel Xeon Scalable Processor + SST-BF + Upstream Features with and without DPIX/MFEX Patches

Figure 12 above represents the throughput with a 3rd Gen Intel® Xeon® 6338N Scalable processor with all available upstream optimizations enabled both with and without the DPIF/MFEX patches enabled. Key data is represented in Table 8 below for easy comparison.

Table 8. Comparison of 3rd Gen Intel Xeon Scalable Processor With Intel SST-BF + Available Upstream Features with and without DPIF/MFEX patches

PACKET SIZE	3RD GEN INTEL XEON SCALABLE PROCESSOR + SST-BF + UPSTREAM FEATURES MBPS	3RD GEN INTEL XEON SCALABLE PROCESSOR + SST-BF + UPSTREAM FEATURES + DPIF/MFEX PATCHES MBPS	PERFORMANCE GAIN
112	8540.99	9991.08	16.98%
256	16456.37	18692.4	13.59%
512	26030.87	28938.9	11.17%
1518	47309.59	51080.11	7.97%

Performance gain from use of the Intel® AVX-512 optimization for DPIF/MFEX is significant. Performance gains for smaller packet sizes benefit on average 16.98% with larger packets benefiting to 7.97%. For larger packets, however, we see a total throughput of 51,080 Mbps.

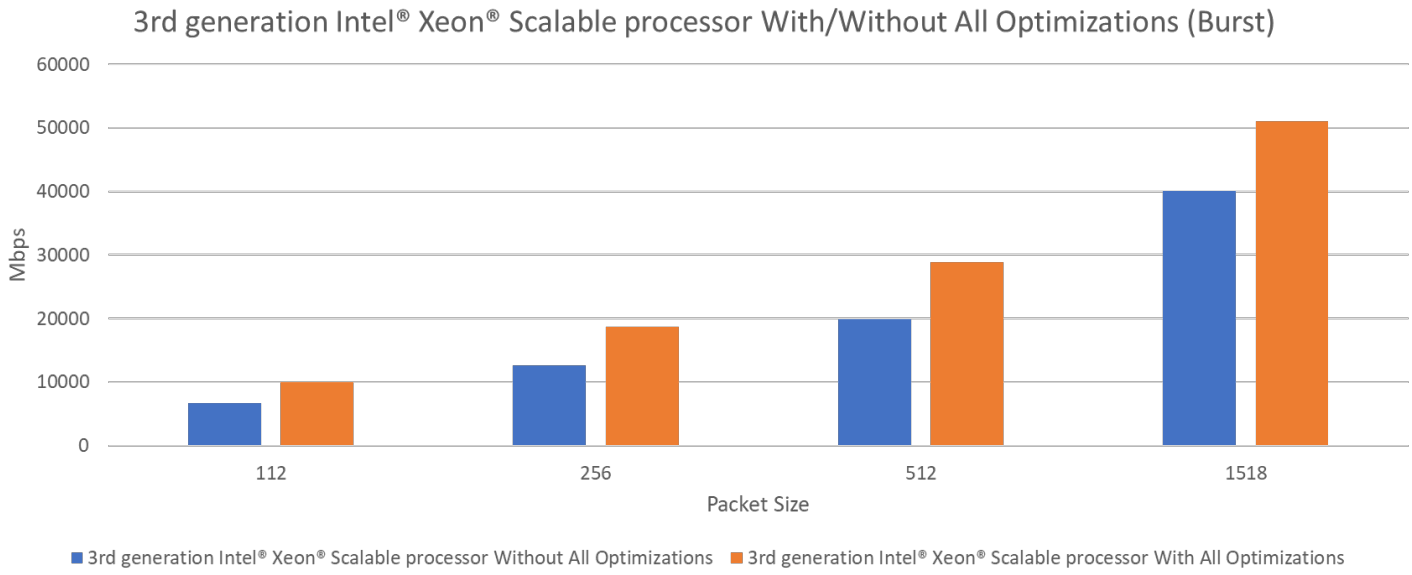

Figure 13 Comparison of Throughput of 3rd Gen Intel Xeon Scalable Processor With and Without All Optimizations

Figure 13 above represents the throughput of a 3rd Gen Intel® Xeon® 6338N Scalable processor with and without all the optimizations outlined in this paper so far to demonstrate the overall gain achieved from enabling all optimizations. Key data is represented in Table 9 below for easy comparison.

Table 9. Comparison of 3rd Gen Intel Xeon Scalable Processor With and Without All Optimizations

PACKET SIZE	3RD GEN INTEL XEON SCALABLE PROCESSOR WITHOUT OPTIMIZATIONS	3RD GEN INTEL XEON SCALABLE PROCESSOR WITH OPTIMIZATIONS	PERFORMANCE GAIN
112	6704.07	9991.08	49%
256	12601.33	18692.39	48%
512	19949.23	28938.9	45%
1518	40155.29	51080.1	27%

It is clear from the table above that the performance gains can be quite visible. Smaller size packets can befit from a range of 49% to 45%. Larger packets benefit to a gain of 27% to reach over 51,000 Mbps in a deployment, a 10,000 Mbps increase over the baseline for the same packet size.

6.8 3rd Gen Intel® Xeon® Scalable Processor with Intel® SST-BF + VIRTIO 1.1 + DPCLS Intel® AVX-512 + Intel® AVX-512 Vector ICE PMD + DPIF/MFEX Patches (Scatter vs. Burst)

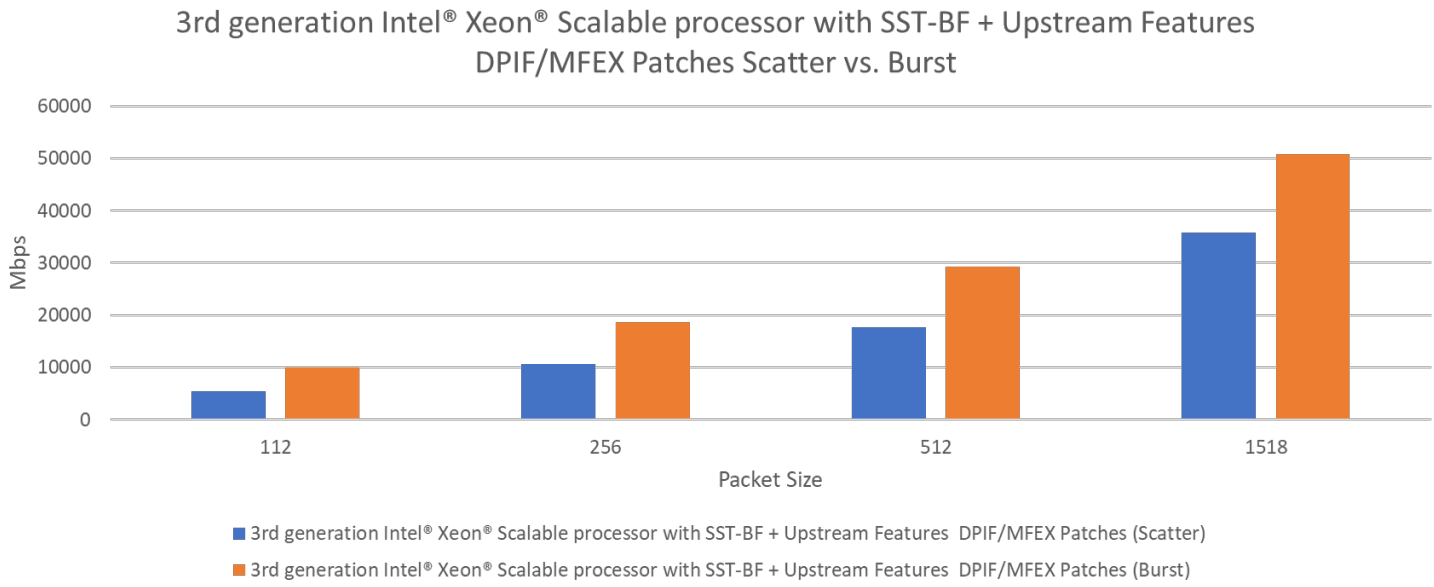

Figure 14 Comparison of 3rd Gen Intel Xeon Scalable Processor with All Optimizations Using Scatter and Burst Traffic Patterns

Figure 14 above compares the throughput of a 3rd Gen Intel® Xeon® 6338N Scalable processor with all optimizations outlined in the paper enabled but with the test traffic profile changed to scatter (worst case) and burst (best case). For details see the Topology section. Key data is represented in Table 10 below for easy comparison.

Table 10. Comparison of 3rd Gen Intel Xeon Scalable Processor With Optimizations Using Scatter and Burst Benchmark Traffic Configuration.

PACKET SIZE	3RD GEN INTEL XEON SCALABLE PROCESSOR + OPTIMIZATIONS + SCATTER	3RD GEN INTEL XEON SCALABLE PROCESSOR + OPTIMIZATIONS + BURST	PERFORMANCE DELTA
112	5833.99	9991.08	42%
256	11538.09	18692.4	38%
512	19272.16	28938.9	33%
1518	38125.01	51080.11	25%

The performance deltas between scatter and burst can be large with smaller packet sizes suffering. Max achievable throughput is 38,125 Mbps with packet size of 1518. Performance is still higher in all packet sizes than the 2nd Gen base burst benchmarks. Scatter is considered worst case scenario and although not typical in deployment has been presented here for full disclosure of performance.

Final note: All throughput has been recorded as received traffic at the traffic generator, i.e., traffic that egresses from the vSwitch itself. It should be noted though that the vSwitch is both transmitting and receiving this amount of traffic at any one moment during the tests. As such, at any one time the vSwitch is dealing with double the amount of data per second than what is being presented in benchmarks above. If the vSwitch is transmitting 50,000 Mbps, it has a total switch load of 100,000 Mbps for both rx/tx operations.

7 Future Work

The results presented above represent features that already are available to users either in hardware or in software. There are, however, several upcoming features in both Intel hardware and OVS with DPDK software that will further improve OVS switching performance. Below is a short summary of these items.

7.1 Intel® AVX-512 DPIF for Inner

In the Deployment Details section, it was mentioned that Intel® AVX-512 was used to accelerate the DPIF (datapath interface, or glue code of OVS Datapath). This optimization is used for most packets passing through OVS in the test topology in this report. It is

not used for all packets, however. Northbound packets (ingress packets, or packets going from NIC to VMs) in this scenario pass through the DPIF twice. This is because they are VXLAN packets and have outer and inner headers that OVS processes one at a time. The OVS term for this is recirculation. The DPIF optimization mentioned in this report is valid for all packets that enter the DPIF for the first time. The recirculated packets use the unoptimized scalar DPIF. Future work is planned around writing an Intel® AVX-512 accelerated DPIF for recirculated packets.

7.2 Intel® AVX-512 MFEX for Inner

As explained in the Deployment Details section, the MFEX is accelerated using Intel® AVX-512 and a mask-and-match approach where specific traffic patterns are optimized. Not all of the traffic patterns in this scenario are optimized. Currently, the southbound packets (egress packets, or packets going from VMs to NIC) are accelerated, since they have a more straightforward ETH/IPv4/UDP traffic pattern for which an MFEX Intel® AVX-512 implementation has been written. An Intel® AVX-512 MFEX implementation needs to be written for the recirculated packets on the northbound path. For these recirculated packets, MFEX needs to add tunnel metadata to the miniflow.

7.3 Intel® AVX-512 Action Implementations

The actions that OVS carries out in the test scenario in this report are output (where a packet should be sent on a port) and VXLAN pop (where the VXLAN header and outer header are removed from the packet headers). The VXLAN pop action takes a non-trivial amount of processing time, especially as other parts of the OVS datapath are being optimized to take less time. This action does some parsing of the packet, similar to how the scalar MFEX would, as well as moving some data from the outer header to tunnel metadata structures in OVS. Intel® AVX-512 acceleration of this action is worth investigating.

7.4 Intel® AVX-512 EMC/SMC

In OVS, there are three caches, or lookup table methods, use to find a particular rule for a packet. They are arranged in a hierarchy of EMC->SMC->DPCLS. EMC and SMC can be enabled/disabled; DPCLS is always on. Whether EMC and SMC should be used is very use-case dependent. As mentioned above in the report, there is an Intel® AVX-512 implementation of the DPCLS that accelerates lookup times. In this report's test scenario, the EMC and SMC are not used. Intel® AVX-512 implementations of these can be written and used. Whether this would show benefits depends on the use case. It's safe to say that for the test scenario in this report, EMC makes performance worse if turned on, since a high number of traffic flows are being used.

7.5 Intel® DSA Integration with OVS

There is ongoing investigation into integrating Intel® DSA into OVS. Intel® DSA is a high-performance data copy and transformation accelerator that will be integrated in future Intel® processors, targeted for optimizing streaming data movement and transformation operations. In the test scenario described in this document, a high number of transmission and reception operations to and from VHOST devices occurs. The ongoing investigation is looking at accelerating the TX to guest.

This will become more important as the parsing and lookup stages of the OVS pipeline become more optimized, at which time the bottleneck will become apparent in the VM back end. When this happens, VHOST IO functions will take a higher percentage of total processing time in OVS. Intel® DSA is a promising candidate to alleviate this type of situation. For more information, see https://01.org/blogs/2019/introducing-intel-data-streaming-accelerator.

8 Summary

In this paper we have outlined a complex OVS use case running on four CPU cores. We have benchmarked the baseline of achievable throughput of this use case with both 2nd and 3rd Gen Intel® Xeon® Scalable processors. Building upon the performance of the 3rd Gen processor, we have demonstrated individually the performance gain brought to the use case by enabling the latest Intel® architecture (IA) technology advancements, such as Intel® SST-BF and Intel® AVX-512. We have also demonstrated the importance of removing bottlenecks by enabling technology outside the scope of OVS, such as VIRTIO 1.1. We also provided benchmark figures demonstrating the complete performance gain when all the technologies are used in unison.

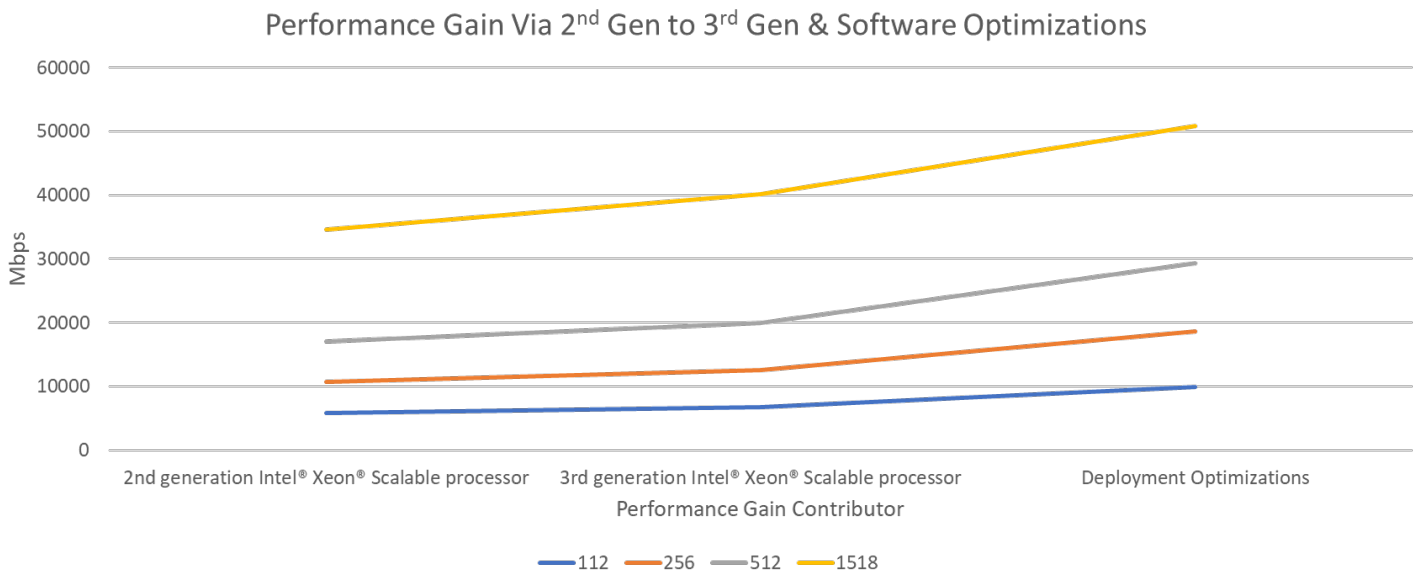

Figure 15 Performance Gain Achievable via 2nd Gen to 3rd Gen and Software Optimizations.

Figure 15 above shows a high-level summary for each packet size as both hardware optimizations and software optimizations are enabled. Migration from a 2nd Generation Intel Xeon Scalable processor to a 3rd Generation with all optimizations enabled can provide gains of up to 70% to 75% for smaller packet sizes (112, 256, 512) and up to 48% with larger packet size of 1518. This can result in vSwitch deployments that can process over 50,000 Mbps using four CPU cores on IA.

It should be noted, as flagged in the Topology section of this paper, the Mbps throughput reported relates to the rate of traffic received at the Ixia traffic generator. It does not convey the full workload of the vSwitch itself, i.e., does not account for both tx and rx operations. If we were to account for the full workload then the vSwitch would be achieving over 50 Gbps on four CPU cores with a packet size of 512 bytes.

Finally, we've outlined roadmap of future optimizations both in hardware and software that will further reduce bottlenecks at either the vSwitch level or the VIRTIO back end. With these in place, it is expected that further improved throughput performance will be gained for a vSwitch deployment running on IA.

0222/DN/WIPRO/PDF 636787-002US