Technology Guide

Intel® Data Streaming Accelerator (DSA) - Calico VPP with Intel® DSA on 5th Gen Intel® Xeon® Scalable Processor

Authors

Xianying Hu

Xiaobing Qian

Mrittika Ganguli

Sunny Wang

Guanxu Huang

1 Introduction

Calico VPP, as a Container Network Interface (CNI), integrates Vector Packet Processing (VPP) as a data plane for Calico. It focuses on improving network performance for applications, and on extending network features.

A shared memory packet interface (Memif) provides high-performance packet transmission and receive using memory copy between the user application and VPP. Calico VPP provides memif user space packet interfaces to the Kubernetes* pods. Memif can be enabled through pod annotation and the pod interface is selected as memif or tun for ingress traffic based on ports. Memif boosts user plane applications as it completely bypasses the kernel. By default, VPP is the controller for the memif interface.

To address the throughput limitation introduced by software implemented memif, the 5th Gen Intel® Xeon® Scalable processor has an embedded Direct Memory Access (DMA) engine to accelerate memory copy operations via the memif interface.

This document takes the Calico VPP instance with memif as a sample workload in order to demonstrate the benefits of accelerated memory copies with the latest 5th Gen Intel® Xeon® Scalable processor with Intel® Data Streaming Accelerator (Intel® DSA).

This document is part of the Network Transformation Experience Kits.

Table of Contents

Figures

Tables

Document Revision History

Revision	Date	Description
001	March 2024	Initial release.

1.1 Terminology

Table 1. Terminology

Abbreviation	Description
CIDR	Classless Inter-Domain Routing
DIF	Data Integrity Field
DMA	Direct Memory Access
DPDK	Data Plane Development Kit
DUT	Device Under Test
Intel® DSA	Intel® Data Streaming Accelerator
QUIC	Quick UDP Internet Connections
SVM	Shared Virtual Memory
SW	Software
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VPP	Vector Packet Processing

1.2 Reference Documentation

Table 2. Reference Documents

Reference	Source
Intel® Data Streaming Accelerator (Intel® DSA)	https://www.intel.com/content/www/us/en/content-details/759709/intel-datastreaming-accelerator-user-guide.html?wapkw=DSA
VPP Wiki	https://wiki.fd.io/view/VPP
Calico	https://docs.tigera.io/calico/latest/about
Calico VPP	https://github.com/projectcalico/vpp-dataplane

2 Calico VPP and Memif

2.1 Calico VPP Data Plane

Calico is an open-source networking and network security solution that is widely used in cloud-native and containerized environments. It is designed to provide seamless connectivity and robust security for workloads, making it an essential component of modern data centers and cloud platforms.

Calico VPP combines Calico and VPP (Vector Packet Processing) to provide a high-performance and scalable network infrastructure in cloud-native and containerized environments. By integrating Calico's advanced networking and security features with the speed and efficiency of VPP, Calico VPP delivers a robust platform for container.

VPP-enabled nodes are fully compatible with regular Calico nodes, allowing for mixed clusters with both VPP-enabled and standard nodes. This simplifies the transition of a cluster from Linux* or eBPF-based Calico networking to VPP-accelerated networking.

In addition, the VPP data plane provides some additional features that are not available in Calico. For instance:

- High performance traffic encryption with support for IPSec
- SRv6 support for node-to-node transport
- VPP HostStack can be consumed by network intensive endpoint applications (using TCP, TLS, UDP, QUIC, ...) with the VPP Client Library VCL
- Memif packet interface support for network-intensive applications for optimized user-space networking

2.2 Memif

Calico VPP's memif is a key feature that enhances the connectivity and networking capabilities within a Calico VPP-enabled environment. Memif is a high-performance, low-latency, and lightweight mechanism for creating direct memory-to-memory connections between different workloads, such as containers or virtual machines, running on VPP-enabled nodes. Memif does

not require special hardware and can be used on any cluster, and the Kubernetes* network policies and service load balancing features are implemented on the traffic path.

Memif provides simple packet interfaces that enable packet exchange between two processes by utilizing a shared memory segment. A single thread can achieve a packet transmission or reception rate of up to 15 million packets per second (Mpps) using memif, which supports both L2 and L3 modes.

Memif operates with two distinct identities, known as primary and secondary, and both transmit control messages via sockets. A one-to-one relationship exists between the primary and secondary identities, where a primary can connect to a secondary, and they share the same identification (ID). The primary takes on the role of creating sockets and monitoring incoming secondary connections. Conversely, the secondary is responsible for creating shared memory files. Upon Initializing shared memory, it proactively initiates socket connection requests. Once the connection is established, user applications gain the capability to send and receive packets over the memif interface.

3 Intel® Data Streaming Accelerator (Intel® DSA)

Intel® Data Streaming Accelerator is a high-performance data copy and transformation accelerator integrated in the 5th Gen Intel Xeon Scalable processors and future Intel® processors. It aims to provide higher overall system performance for data transfer and transformation operations, while freeing up CPU cycles for higher level functions.

Intel DSA hardware supports high-performance data transfer capabilities to/from volatile memory, persistent memory, memory-mapped I/O, and through a Non-Transparent Bridge (NTB) in the SoC to/from remote volatile and persistent memory on another node in a cluster. It offers a PCI Express compatible programming interface to the operating system and can be controlled through a device driver.

Besides performing basic data transfer operations, Intel DSA can also perform some higher-level transformation operations on memory. For instance, it can generate and test CRC checksum or Data Integrity Field (DIF) on the memory region to support usages typical with storage and networking applications. It can compare two memory regions for equality, generate a delta record, and apply a delta record to a buffer. These are compared and the delta generate/merge functions may be used by applications such as VM migration, VM fast check-pointing, and software managed memory deduplication.

4 Optimized Calico VPP

Calico VPP offers support for memif to enhance packet transmission performance between memif primary and secondary connections. We categorize this type of memif as "Software (SW) Memif " because it relies on the CPU to copy memory data between the memif primary and secondary identities. When Intel® DSA is enabled on Memif, it is referred to as a "DSA Memif " because it utilizes DSA instead of the CPU for memory data copying. In comparison to SW Memif, DSA Memif exhibits significantly faster memory copy speeds, resulting in superior packet transmission performance between applications and VPP. The comparison of SW Memif and DSA Memif is shown in Figure 1.

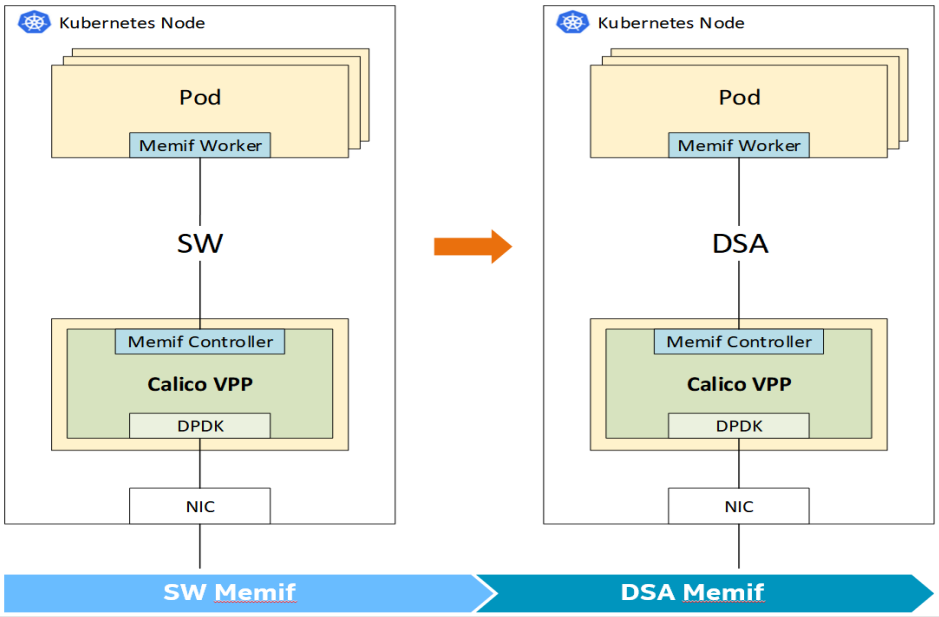


Figure 1. Comparison of SW Memif and Intel DSA Memif

5 Performance Benchmark Test

5.1 Overview

We conducted a performance benchmark test to showcase the advantage of memory copy acceleration in Calico VPP's memif, based on the 5th Gen Intel Xeon Scalable processor with Intel DSA. In this benchmark test, we deployed TRex as the traffic generation tool and VPP-L3FWD as the traffic forwarding tool. TRex is a high-speed, realistic open-source traffic generation tool designed to run on standard Intel processors utilizing DPDK. It supports both stateful and stateless traffic generation modes. VPP-L3FWD, on the other hand, is a sample application showcasing the use of VPP library to implement a high-performance routing function.

TRex is set to send 100 Gbps traffic, which matches the line rate of the network adapter. This setup ensured that sending side did not introduce any bottlenecks in our benchmark test. TRex was employed to monitor L2 RX throughput and convert it to L1 RX throughput, which represents VPP-L3FWD's memif forwarding capability. To illustrate the advantage of Intel DSA's memory copy acceleration, we conducted a comparative analysis of VPP-L3FWD's SW memif and DSA memif in terms of their forwarding capabilities.

5.2 Workload Architecture

TRex pod and VPP-L3FWD pod are deployed on two Kubernetes nodes. On the VPP-L3FWD side, core utilization is determined by the core scaling, while Calico VPP pod employs one main core with several worker cores, the quantity of which is determined by the core scaling configuration. Calico VPP container and VPP-L3FWD pod both use one main core, and they use the same number of worker cores, which depends on core scaling.

On TRex side, pod uses one core for master thread, one core for latency thread and eight cores for worker thread. Both the VPP container and TRex use eight worker cores to make the packet sending rate reach the line rate of the network adapter, so that the TRex side will not be the bottleneck of the benchmark test.

The workload architecture is shown in Figure 2.

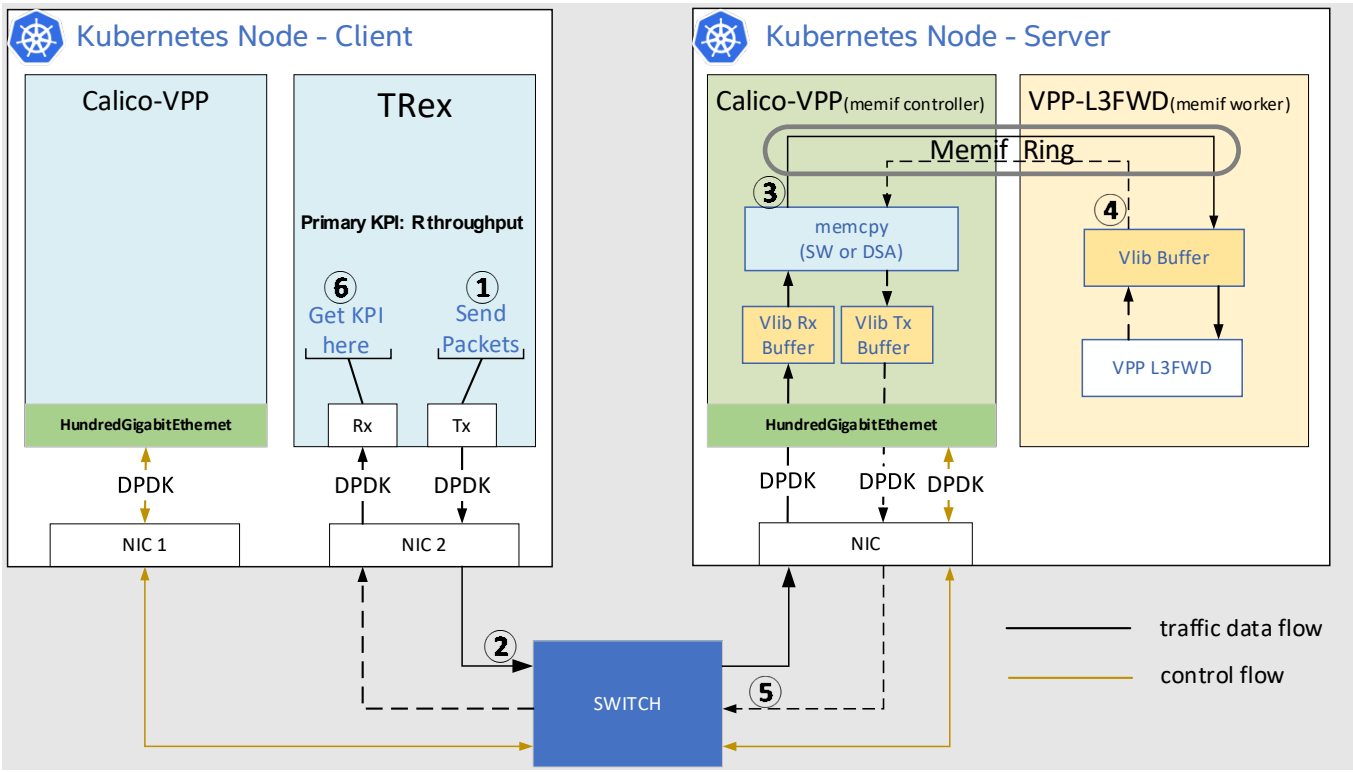


Figure 2. Workload architecture

5.3 System Configuration

For the performance benchmark test on the 5th Gen Intel Xeon Scalable processors, which has deployed VPP-L3FWD, refer to the system configuration of DUT as detailed in Table 3.

Table 3. System configurations

	VPP-L3FWD node (Server)	TRex node (Client)
Manufacturer	Intel Corporation	Intel Corporation
Product Name	Intel Reference Platform	Intel Reference Platform
BIOS Version	EGSDCRB1.86B.0105.D48.2308140023	SE5C7411.86B.9525.D18.2303071632
OS	Ubuntu 22.04 LTS	Ubuntu 22.04 LTS
Kernel	5.15.0-73-generic	5.15.0-73-generic
Microcode	0xa1000161	0x2b0004b1
CPU Model	Intel® Xeon® Platinum 8592+	Intel® Xeon® Platinum 8480+
Base Frequency	1.9GHZ	2.0GHz
Maximum Frequency	3.9GHz	3.8GHz
All-core Maximum Frequency	2.9GHz	3.0GHz
CPU(s)	256	224
Thread(s) per Core	2	2
Core(s) per Socket	64	56
Socket(s)	2	2
NUMA Node(s)	2	2
Prefetchers	L2 HW, L2 Adj., DCU HW, DCU IP	L2 HW, L2 Adj., DCU HW, DCU IP
Turbo	Enabled	Enabled
Power & Perf Policy	Performance	Performance
TDP	350 watts	350 watts
Frequency Driver	intel_pstate	intel_pstate
Frequency Governer	Performance	Performance
Max C-State	9	9
Installed Memory	512GB (16x32GB DDR5 4800 MT/s [4800 MT/s])	512GB (16x32GB DDR5 4800 MT/s [4800 MT/s])
Huge Pages Size	1048576kB	1048576kB
Transparent Huge Pages	madvise	madvise
Automatic NUMA Balancing	Enabled	Enabled
Network Adapter Summary	1x Intel® Ethernet Controller I225-LM, 2x Intel® Ethernet Controller E810-C for QSFP	2x I350 Gigabit Network Connection, 4x Intel® Ethernet Controller E810-C for QSFP
Drive Summary	1x 465.8G Hitachi HTS2505	1x 894.3G Micron_5300_MTFD

5.4 Software Configuration

To perform the performance benchmark test, refer to the software configuration as detailed in Table 4. Tested by Intel as of 10/9/2023.

Table 4. Software configurations

	VPP-L3FWD node (Server)	TRex node (Client)
Calico VPP Version	3.23.0	3.23.0
Calico Version	3.23.0	3.23.0
VPP Version	22.02	22.02

Compiler	gcc version 11.4.0 (Ubuntu 11.4.0-1ubuntu1~22.04)	gcc version 11.4.0 (Ubuntu 11.4.0-1ubuntu1~22.04)
DPDK Version	21.11	21.11
Containerd Version	1.6.18	1.6.18
Kubernetes Version	1.24.4	1.24.4
Isolcpus	64-95	56-83
Calico VPP Core Number	64 (main core), 65-70 (worker)	56 (main core), 57-62 (worker)
VPP L3FWD Core Number	74,75-80 (depends on core scaling)	70,71-76 (depends on core scaling)
Traffic Generator Type	N/A	TRex
Traffic Generator Version	N/A	v3.00
TRex Server Core Number	N/A	10,11,12-19
TRex DPDK Version	N/A	22.03
Network Adapter Model	Intel® Ethernet Controller E810-C for QSFP	Intel® Ethernet Controller E810-C for QSFP
Network Adapter Firmware	4.20 0x80017785 1.3346.0	4.20 0x80017785 1.3346.0
Network Adapter Driver	ice 5.15.0-73-generic	ice 5.15.0-73-generic

6 Calico VPP Performance

6.1 Frame Size Scaling

For the Frame size scaling performance test, both SW memif and DSA memif scenarios comply with the following configurations:

- MTU: 1500/9000
- Calico VPP / VPP-L3FWD Worker Core Number: 1
- Procotol: TCP
- Calico VPP Worker Core Op Freq (TRex/ VPP-L3FWD): 2.0 GHz
- Calico VPP Worker Core utilization (TRex/VPP-L3FWD): 100%
- DSA: 1 instance, 4 engines, 4 work queues

We tested the throughput of different MTU scenario: 1500 and 9000. MTU 1500 results are shown in Figure 3.

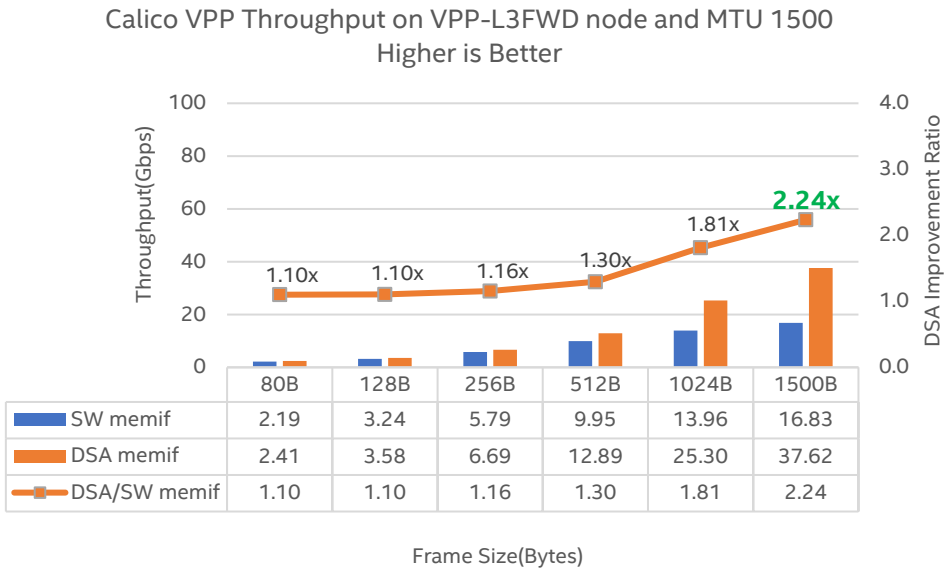
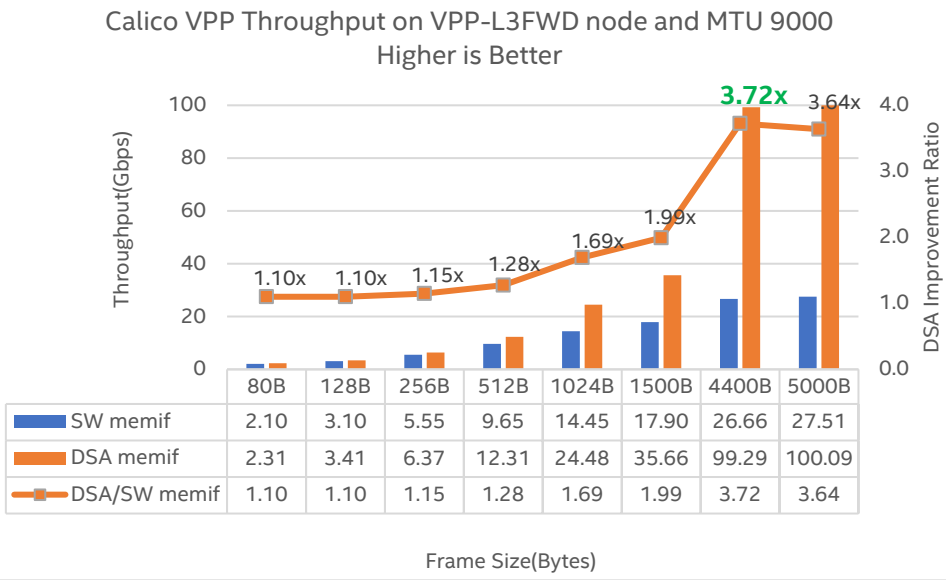


Figure 3. Calico VPP frame size scaling throughput on one core and MTU 1500

From the chart above, the throughput shows a linear growth with the increase of frame size. For frame size 1500 B, the DSA improvement reaches to the maximum. Calico VPP performance with DSA memif is ~2.24 times of SW memif.



	80B	128B	256B	512B	1024B	1500B	4400B	5000B
SW memif	2.10	3.10	5.55	9.65	14.45	17.90	26.66	27.51
DSA memif	2.31	3.41	6.37	12.31	24.48	35.66	99.29	100.09
DSA/SW memif	1.10	1.10	1.15	1.28	1.69	1.99	3.72	3.64

Figure 4. Calico VPP frame size scaling throughput on one core and MTU 9000

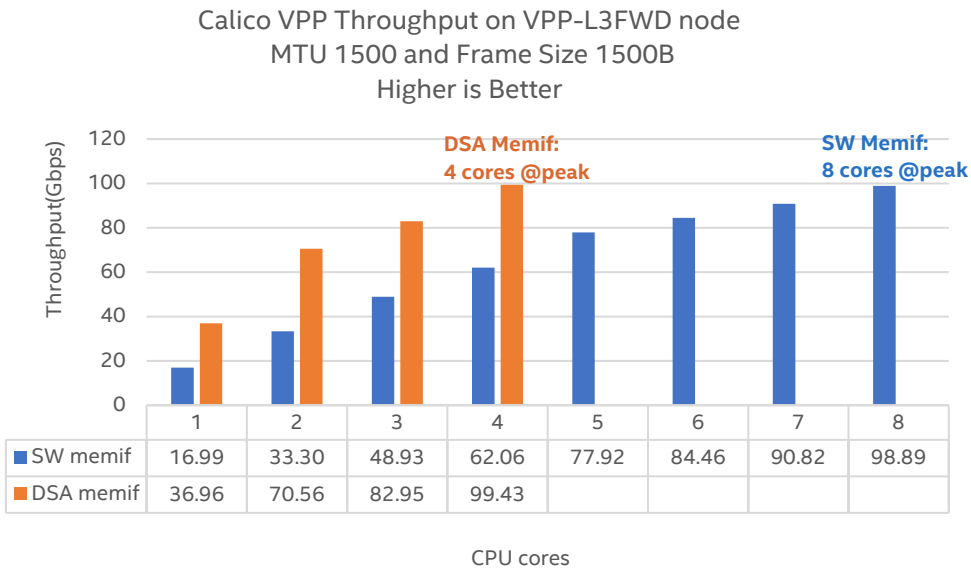
From the performance data of frame size scaling on MTU9000, as the frame size increases, the DSA improves higher. When the frame size increases to 4400 bytes, throughput with DSA Memif is up to ~3.72 times of SW Memif. The throughput has reached the line rate of 100 Gbps from 4400 B packet size, hence the DSA memif throughput is about the same in the 5000 B case.

6.2 Core Scaling

For the core scaling performance test, both SW Memif and DSA Memif scenarios comply with the following configurations:

- MTU: 1500/9000
- Calico VPP / VPP-L3FWD Worker Core Number: 1/2/3/4/5/6/7/8
- Procotol: TCP
- Calico VPP Worker Core Op Freq (TRex/ VPP-L3FWD): 2.0 GHz
- Calico VPP Worker Core utilization (TRex/VPP-L3FWD): 100%
- DSA: 1 instance, 4 engines, 4 work queues

We tested the throughput of different MTU scenario and VPP-L3FWD Worker Core Number are same with Calico VPP worker core number. On MTU 1500, the results are shown in Figure 5.



	1	2	3	4	5	6	7	8
SW memif	16.99	33.30	48.93	62.06	77.92	84.46	90.82	98.89
DSA memif	36.96	70.56	82.95	99.43				

Figure 5. Calico VPP core scaling throughput on MTU 1500 and frame size 1500B

We can see from the above chart that throughput with DSA memif achieves 100 Gbps when CPU cores are 4, throughput is ~1.60 times of SW memif. DSA Memif needs ~4 CPU cores to achieve the max throughput and SW memif needs ~8 cores, so DSA memif can save ~4 CPU cores.

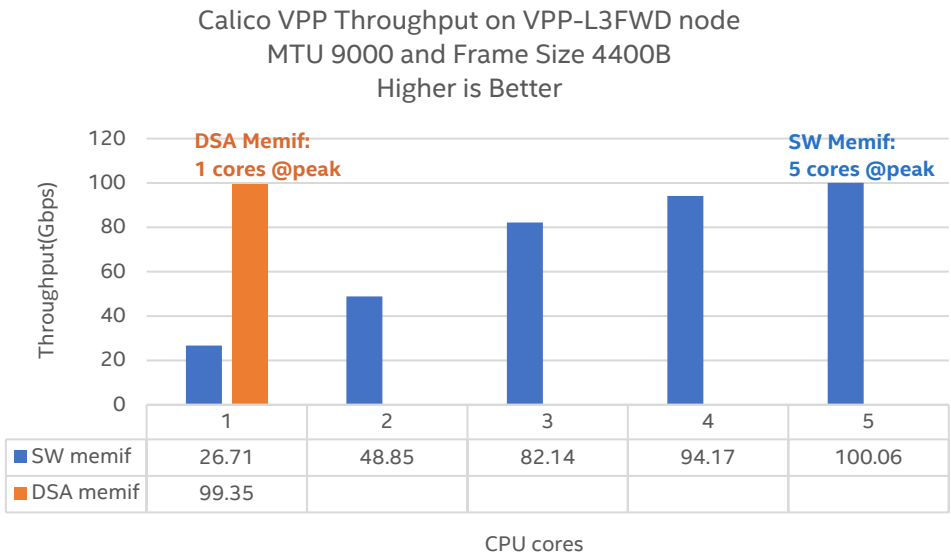


Figure 6. Calico VPP core scaling throughput on MTU 9000 and frame size 4400B

We can see from the above chart that DSA memif needs ~1 CPU cores to achieve the max throughput and SW memif needs ~ 5 CPU cores, so DSA memif can save ~4 CPU cores.

6.3 Memif Memory Copy Overhead Analysis

We used Perf to analyze Calico VPP with SW and DSA Memif performance on one CPU core scenario and frame size is 512 B, throughput data is shown in Table 5.

Both SW memif and DSA memif scenarios comply with the following configurations:

- MTU: 1500
- Calico VPP / VPP-L3FWD Worker Core Number: 1
- Procotol: TCP
- VPP Worker Core Op Freq (TRex/ VPP-L3FWD): 2.0 GHz
- VPP Worker Core utilization (TRex/VPP-L3FWD): 100%
- DSA: 1 instance, 4 engines, 4 work queues

Table 5. Calico VPP throughput on MTU 1500 and frame size 512B

	SW Memif	DSA Memif	DSA/SW Memif
Throughput (Gbps)	**9.59**	**12.78**	**1.33**

The Perf data of Calico VPP worker core on VPP-L3FWD node is shown in Table 6.

Table 6. VPP core Perf analysis of Calico VPP with Memif

object	symbol	object_SW	symbol_SW	object_DSA	symbol_DSA
memif_plugin.so	memif_device_class_tx_fn_icl	31.91	19.01	10.04	4.57
	memif_input_node_fn_icl		12.81		3.94
	memif_dma_completion_cb_ma_icl				0.45
	memif_tx_dma_completion_cb_ma_icl				0.99
libvnet.so.22.0	ip4_input_node_fn_icl	25.32	2.26	29.14	2.98
	ip4_lookup_node_fn_icl		3.98		3.76

	gso_ip4_node_fn		2.07		1.97
	ip4_rewrite_node_fn_icl		3.95		5.38
	vnet_per_buffer_interface_output_node_fn_icl		1.69		2.15
	ethernet_input_node_fn_icl		2.63		3.05
	lookup_ip4_dst_node_fn		4.14		4.40
	ip4_input_no_checksum_node_fn_icl		1.97		2.42
	vnet_interface_output_node_fn_icl		0.28		1.08
	virtio_pre_input		0.28		1.67
cnat_plugin.so	cnat_input_feature_ip4_node_fn	18.53	5.36	25.58	7.44
	cnat_output_feature_ip4_node_fn		10.89		15.99
	cnat_snat_policy_k8s		2.18		1.97
	_hash_get@plt		0.09		0.18
dpdk_plugin.so	ice_xmit_pkts	9.42	4.79	8.56	4.21
	ice_recv_scattered_burst_vec_avx512_offload		1.41		1.97
	dpdk_input_node_fn_icl		1.98		0.63
	dpdk_device_class_tx_fn_icl		1.13		1.64
libvppinfra.so.22.06	lookup	6.28	3.21	12.75	8.52
	get_indirect		1.34		1.61
	_hash_get		0.56		1.61
	mspace_usable_size_with_delta		0.85		0.81

From the results shown in Table 6, we can see that the CPU usage of Memif_plugin is significantly decreased, and Intel DSA saves up to ~21.87% CPU computing capacity. The Memif_plugin.so is a dynamic link library or shared object file that serves as a plugin within the context of network and communication software systems. The name " Memif " is derived from "Memory Interface," which indicates its role in facilitating high-speed memory-to-memory communication between processes or components. This Perf data indicates that enabling DSA provides the 5th Gen Intel® Xeon® Scalable processor with greater capacity to handle network protocol stack work. Therefore, Intel DSA can improve Calico VPP throughput while also reducing CPU usage.

6.4 EMON Analysis

We use EMON to analyze Calico VPP with SW and DSA memif performance on one CPU core scenario. The EMON data of Calico VPP on VPP-L3FWD node is shown in Table 7.

Table 7. EMON data of Calico VPP on VPP-L3FWD node

	SW Memif	DSA Memif	Ratio
metric_CPU operating frequency (in GHz)	2.00	2.00	1.00
metric_CPU utilization %	100.00	100.00	1.00
metric_CPU utilization% in kernel mode	0.11	0.10	0.91
metric_CPI	0.32	0.25	0.78
IPC (=1/metric_CPI)	3.12	4.00	1.28
metric_kernel_CPI	1.45	1.35	0.93
metric_TMA_Frontend_Bound(%)	3.15	4.30	1.37
metric_TMA_..Fetch_Bandwidth(%)	1.99	3.14	1.58
metric_TMA_Bad_Speculation(%)	1.19	1.59	1.34
metric_TMA_Backend_Bound(%)	46.55	29.09	0.62

metric_TMA_..Memory_Bound(%)	31.61	12.08	0.38
metric_TMA_....L1_Bound(%)	5.26	4.82	0.92
metric_TMA_....L2_Bound(%)	1.62	1.48	0.91
metric_TMA_....L3_Bound(%)	18.91	1.16	0.06
metric_TMA_Retiring(%)	49.10	65.03	1.32
metric_IO_bandwidth_disk_or_network_writes (MB/sec)	1261.19	5183.84	4.11
metric_IO_bandwidth_disk_or_network_reads (MB/sec)	1233.53	4746.88	3.85
metric_memory bandwidth read (MB/sec)	129.85	116.74	0.90
metric_memory bandwidth write (MB/sec)	107.38	98.92	0.92
metric_memory bandwidth total (MB/sec)	237.23	215.65	0.91

From the results of Table 7, we can see that memory bound ratio of DSA Memif is only 0.38x of SW Memif; L3 bound ratio of DSA Memif is only 0.06x of SW Memif; Retiring ratio of DSA Memif is 1.32x of SW Memif; IPC of DSA Memif is 1.28x of SW Memif. These metrics show the advantages of DSA Memif over SW Memif.

7 Summary

This document provides an overview of Calico VPP, Memif, Intel DSA technology, and their role in optimizing Calico VPP and workload architectures. The document includes performance testing data for packet size scaling and core scaling to assess the performance improvements achieved by Intel DSA. Additionally, we employ Perf for performance analysis to understand the factors contributing to the performance gains enabled by DSA. EMON is used to highlight the advantages of DSA Memif over SW Memif.

This guide demonstrates how Intel DSA can increase Calico VPP performance by accelerating memory copy, giving the CPU more cycles to process the network protocol stack. On MTU 1500, DSA Memif performance is up to ~2.24x of SW Memif when the packet size is 1500 bytes. DSA Memif can save ~4 CPU cores. For an MTU of 9000 B, DSA Memif performance is up to ~3.72x of SW Memif when packet size is 4400 bytes. DSA Memif can save ~4 CPU cores.

0324/DN/WIT/PDF 816471-001US