Technology Guide

Intel® Data Streaming Accelerator (Intel® DSA) - Calico VPP Multinet with Intel® DSA on 4th Gen Intel® Xeon® Scalable Processor

Authors

Yunjiang Zhang

Mrittika Ganguli

Xiaobing Qian

Sunny Wang

1 Introduction

Calico VPP is the integration of Vector Packet Processing (VPP) as a Calico data plane. It focuses on improving network performance for applications and also on extending network features. Multinet is one of the important features that allows pods to attach to multiple networks.

In some Kubernetes use cases, complex connectivity features are needed. Some applications require isolated networks for handling different types of traffic. Others may need high-performance interfaces, but only for specific classes of traffic. Calico VPP Multinet implementation is an attempt at fulfilling these requirements. Essentially, it allows creation of several isolated Kubernetes networks to which pods can attach, and it exposes the usual Kubernetes abstractions (services and policies) in each of them.

A shared memory packet interface (memif) provides high performance packet transmit and receive using memory copy between user application and VPP or multiple user applications. To address the industry challenge, the 4th Gen Intel® Xeon® Scalable processor has embedded a DMA engine to accelerate memory copy in the memif.

This document takes the Calico VPP Multinet using memif as the example to demonstrate the memory copy acceleration benefit based on the latest 4th Gen Intel® Xeon® Scalable processor with Intel® Data Streaming Accelerator (Intel® DSA).

This document is part of the Network & Edge Platform Experience Kits.

Table of Contents

Figures

Tables

Document Revision History

Revision	Date	Description
001	September 2023	Initial release.
002	January 2024	Update legend for figure 5 and 6.

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
IPAM	IP Address Management
Memif	Shared Memory Packet Interface
NTB	Non-Transparent Bridge
PCI	Peripheral Component Interconnect
PPS	Packets Per Second
QUIC	Quick UDP Internet Connections
SVM	Shared Virtual Memory
SW	Software
TCP	Transmission Control Protocol
TUN	A virtual IP point-to-point interface
UDP	User Datagram Protocol
VPP	Vector Packet Processing

1.2 Reference Documentation

Table 2. Reference Documents

Reference	Source
Intel® I/O Acceleration Technology	https://www.intel.com/content/www/us/en/wireless-network/acceltechnology.html
Introducing the 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
VXLAN encapsulation/decapsulation	https://en.wikipedia.org/wiki/Virtual_Extensible_LAN
Calico	https://docs.tigera.io/calico/latest/about
Multinet	https://github.com/projectcalico/vpp-dataplane/blob/5d6a3fb6166b3157f6f368f16b781b7eb46762d5/docs/multinet.md
Multus	https://github.com/k8snetworkplumbingwg/multus-cni/blob/master/README.md
Whereabouts	https://github.com/k8snetworkplumbingwg/whereabouts/blob/master/README.md

2 Calico VPP Multinet

2.1 Calico VPP Data Plane

Calico is the industry standard, free and open-source networking and network security solution for containers, virtual machines, and native host-based workloads. Calico supports a broad range of platforms including Kubernetes, OpenShift, Mirantis Kubernetes Engine (MKE), OpenStack, and bare metal services. Calico VPP is the integration of VPP as a Calico data plane.

The Calico VPP data plane brings the performance, flexibility, and observability of VPP to Kubernetes networking. The VPP data plane does not have any additional requirements compared to regular Calico. All the networking configuration, including traffic encapsulation, IP Pools, BGP Configuration, and so on, is done through regular Calico means. VPP-enabled nodes are entirely compatible with regular Calico nodes - meaning that it is possible to have a cluster with some VPP-enabled nodes and some regular nodes. This makes it easy to migrate a cluster from Linux or eBPF Calico networking to VPP-accelerated networking.

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

- There is very fast container traffic encryption with IPSec.
- SRv6 is supported for node-to-node transport.
- The VPP HostStack can be consumed by network intensive endpoint applications (using TCP, TLS, UDP, QUIC, ...) with the VPP Client Library VCL.
- Memif packet interfaces can also be required by network intensive applications for optimized user-space networking.
- Multinet can be leveraged by containerized network functions that require multiple high speed or isolated interfaces.

In this document, we refer to the Calico VPP that does not enable the Multinet feature as Singlenet. Singlenet does not implement the same network isolation function as Multinet using VXLAN encapsulation method, which results in a certain performance loss. We also test the performance of Singlenet and compare it with the performance data of Multinet to analyze the impact of enabling Multinet feature on Calico VPP (detailed in Section 6.3).

2.2 Memif

The memif feature in Calico VPP, which provides pod connectivity through memif, is a method for deploying high-performance packet processing applications or CNFs (Cloud Native Network Functions). Great performance aside, 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.

Memifs are straightforward packet interfaces that can be used to exchange packets between two processes and are based on a shared memory segment. A single thread can transmit or receive up to 15 Mpps on a memif. Memifs support both L2 and L3 modes.

The memif has two identities, primary and secondary, and both transmit control messages through sockets. There is a one-to-one relationship between primary and secondary. A primary can only connect to one secondary, and both have the same id. The primary is responsible for creating sockets and monitoring secondary connection requests. The secondary is responsible for creating shared memory files. Initializing shared memory actively initiates socket connection requests. After the connection is established, user application can receive or transmit packets over memif.

2.3 Multinet

Multinet is a feature of Calico VPP that allows pods to connect to multiple isolated Kubernetes networks and expose Kubernetes abstractions (services and policies) in each of them.

Multinet introduces a new Kubernetes resource (CRD), named Network. A Network is defined by a VNI, (Virtual Network Identifier) that allows to identify the network in the data plane and contains a CIDR range that defines the IP addresses to assign to that network's pods. When a Network is created, the NetWatcher component in the Calico VPP agent creates a dedicated VRF (Virtual Routing and Forwarding) in VPP to isolate the routes of the network.

To associate pods to the network, Multinet leverages Multus. Multus is a CNI plugin that enables pods to have multiple network interfaces. It performs multiplexing by calling the Calico CNI once per attachment. To make Multus aware of the existing networks, a network attachment definition object must be created for each network with a matching name. In this object, you can specify the IPAM (IP Address Management) and the `dataplane_options`, which include the `network_name` parameter to indicate which network to connect to. The IPAM that is used to assign addresses to pods in secondary networks is Whereabouts.

If we create two "blue" and "red" networks, a pod that has an interface in the blue network can communicate with other pods attached to that network. However, interfaces of different networks cannot communicate between each other, that is, a pod cannot ping a red interface from a blue one.

On the data plane front:

- All blue interfaces belong to the same blue VRF, which is isolated from the red VRF.
- When packets go from node to node, Multinet needs to carry the VRF color along with them. In the current implementation, Multinet has chosen to create a VXLAN tunnel between each pair of nodes, carrying the network VNI as the VXLAN VNI.

Multinet supports memif, which can scale to much higher PPS than kernel interfaces. It is ideal for packet processing workloads.

2.4 VPP Graph Node of Multinet

VPP adopts the packet processing graph as its core design. The graph nodes are organized as tree shaped graphs in VPP.

We use "trace" command of "calivppctl" to analyze the packet vectors flow from memif-input all the way to "HundredGigabitEthernet-tx" and "dpdk-input" all the way to memif-output.

The "dpdk-input" tracks the process of the VPP container receiving the packet from the virtual 100 G network adapter to transmitting packets to memif, which represents how the Multinet handles the received packet. "memif-input" tracks the process of the VPP container receiving the packet from the memif to the virtual 100 G network adapter sending the packet, which represents the process of Multinet for sending packets. The graph node of Multinet sending and receiving packet process is shown in Figure 1.

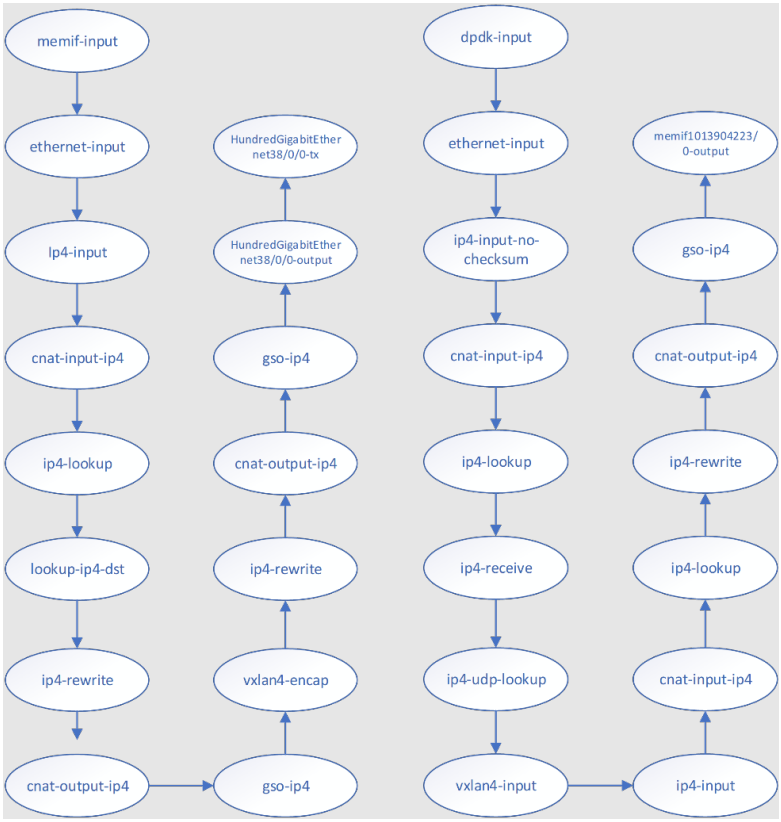


Figure 1. VPP graph node of Multinet sending and receiving packet process

3 Intel® Data Streaming Accelerator (Intel® DSA)

Intel® Data Streaming Accelerator is a high-performance data copy and transformation accelerator integrated in 4th Gen Xeon Scalable processors and future Intel® processors. It aims to provide higher overall system performance for data mover and transformation operations, while freeing up CPU cycles for higher level functions.

Intel DSA hardware supports high-performance data mover capability 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.

In addition to performing basic data mover operations, Intel® DSA can also perform some higher-level transformation operations on memory. For example, 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.

Intel DSA also supports Shared Virtual Memory (SVM) operation, which allows it to work directly in the application's virtual address space without requiring a pinned memory. Moreover, it can support memory overlap for the source and destination addresses. There is no restriction for memory alignment, which makes it possible to accelerate any type of memory copy.

4 Optimize Calico VPP Multinet Memif

Calico VPP Multinet uses the TUN interface by default, and a packet goes through the kernel as it travels between VPP and pods. Calico VPP Multinet supports memif to improve packet transmission performance between memif primary and secondary. We call this kind of memif software (SW) memif since it uses CPU to copy memory data between memif primary and secondary. If Intel DSA is enabled on memif, we call it DSA memif because it uses Intel DSA instead of the CPU to copy memory data. Compared with SW memif, DSA memif has a faster memory copy speed, which enables better packet transmission performance between applications and VPP. The comparison of TUN interface, SW memif, and DSA memif is shown in Figure 2.

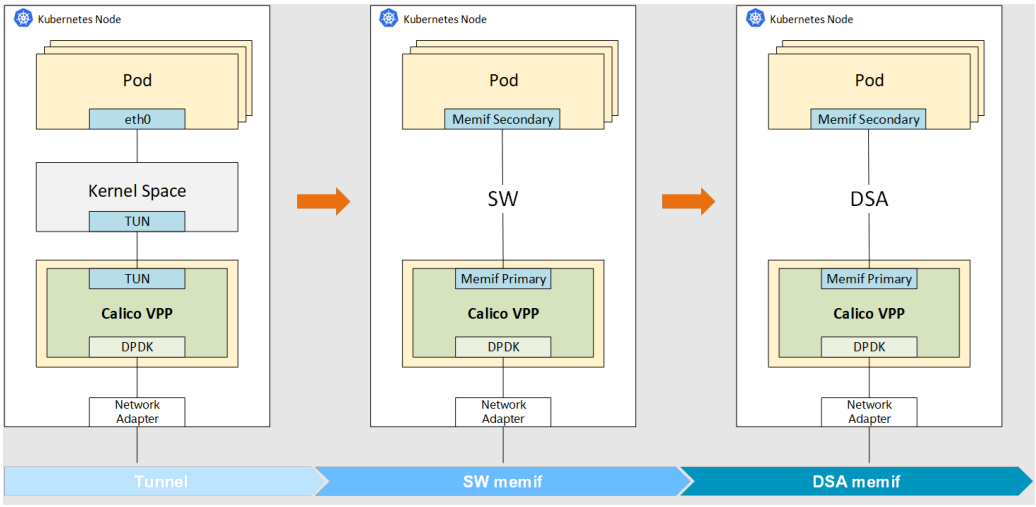


Figure 2. Comparison of TUN interface, SW memif. and DSA memif

5 Performance Benchmark Test

5.1 Overview

We ran a performance benchmark test to demonstrate the memory copy acceleration benefit for Calico VPP Multinet memif based on the 4th Gen Intel Xeon Scalable processor with Intel DSA. We analyzed performance and identified memory copy as a common bottleneck between user application and Vector Packet Processing.

In this benchmark test, TRex is used as the traffic generation tool and VPP-L3FWD is the traffic forwarding tool. TRex is a fast realistic open-source traffic generation tool, running on standard Intel processors based on DPDK. It supports both stateful and stateless traffic generation modes. VPP-L3FWD is a sample application that demonstrates how to use the VPP library to implement a high-performance routing function. It forwards packets using the Longest Prefix Match algorithm based on the forwarding table.

TRex is set to send 100 Gbps (line rate of network adapter) traffic to ensure that the sending side is not the bottleneck of this benchmark test. We used TRex to monitor L2 RX throughput and convert it to L1 RX throughput, which represents the forwarding capability of VPP-L3FWD's memif. We demonstrate the memory copy acceleration benefit of Intel DSA by comparing forwarding capability of VPP-L3FWD's SW memif with that of VPP-LFWD's DSA memif.

5.2 Workload Architecture

We used TRex with blue network and red network SW memifs to send packets to VPP-L3FWD with blue network and red network SW memifs or blue network and red network DSA memifs through Calico VPP. Both the blue network interface and the red network interface are equipped to test the network isolation function of Multinet. In the performance test, only the throughput of a single color is tested.

TRex pod and VPP-L3FWD pod are deployed on two Kubernetes nodes. On the TRex side, a total of 19 cores are utilized. Among them, VPP container uses one main core and eight worker cores, TRex 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 is not the bottleneck of the benchmark test. On the VPP-L3FWD side, the VPP container and the VPP-L3FWD pod both use one main core, and they use the same number of worker cores, which depends on core scaling. The workload architecture is shown in Figure 3.

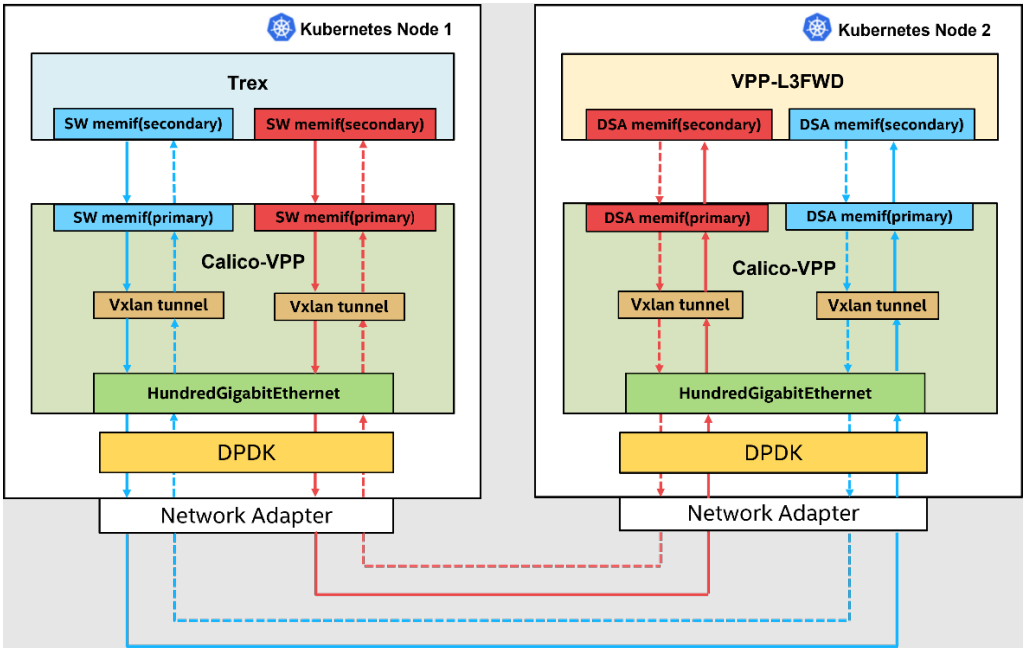


Figure 3. Workload architecture

5.3 KPI

This benchmark result focuses on one KPI: Mean Throughput (Gbps).

Multinet uses a method like VXLAN encapsulation to achieve network isolation, but this encapsulation is invisible to TRex. Therefore, the L1 RX Throughput obtained by TRex is inaccurate.

The components of the package sent by TRex, the package encapsulated by Multinet, and the package transmitted in Ethernet (L1 Layer) are shown in Figure 4.

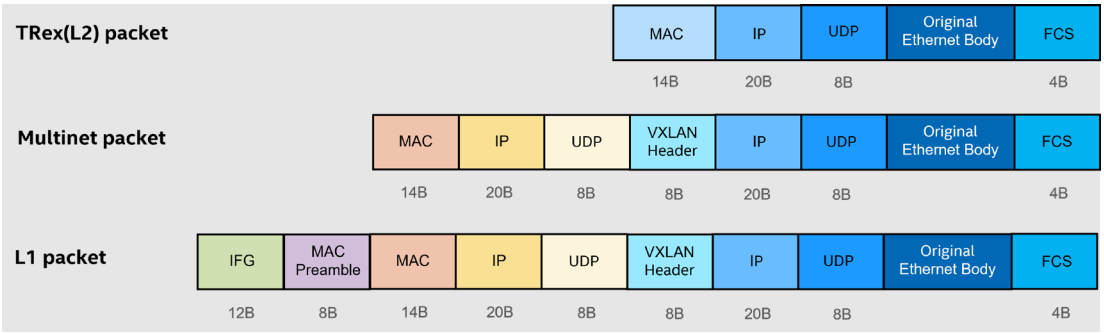


Figure 4. Multinet packet composition

We use TRex to obtain L2 RX throughput and use the following formula to convert and get L1 RX throughput:

$$L1\ RX = \frac{L2\ RX * (Trex\ packet\ size - 14 + 8 + 8 + 20 + 14 + 8 + 12)}{TRex\ packet\ size} = L2\ RX * (1 + \frac{56}{TRex\ packet\ size})$$

5.4 System Configuration

For the performance benchmark test on the 4th Gen Intel Xeon Scalable processors, refer to the system configuration of DUT as detailed in Table 3.

Table 3. System configurations

	TRex node	VPP-L3FWD node
Manufacturer	Intel Corporation	Intel Corporation
Product Name	Intel Reference Platform	Intel Reference Platform
BIOS Version	EGSDCRB1.86B.9409.P15.2301131123	EGSDCRB1.86B.9409.P15.2301131123
OS	Ubuntu 22.04.1 LTS	Ubuntu 22.04.1 LTS
Kernel	5.15.0-43-generic	5.15.0-25-generic
Microcode	0xab000190	0xab000190
IRQ Balance	Enabled	Enabled
CPU Model	Intel® Xeon® Platinum 8487CL CPU	Intel® Xeon® Platinum 8472C CPU
Base Frequency	1.9 GHz	2.5 GHz
Maximum Frequency	3.8 GHz	3.8 GHz
All-core Maximum Frequency	2.8 GHz	3.0 GHz
CPU(s)	224	208
Thread(s) per Core	2	2
Core(s) per Socket	56	52
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
PPIN(s)	270c94cc50d17a48,272242cb54bb3069	28c780bc79d8303e,28c782cc57997f74
Power & Perf Policy	Performance	Performance
TDP	350 watts	350 watts
Frequency Driver	intel_pstate	intel_pstate
Frequency Governor	performance	performance
Max C-State	9	9
Installed Memory	512 GB (16x32GB DDR5 4800 MT/s [4800 MT/s])	512 GB (16x32GB DDR5 4800 MT/s [4800 MT/s])
Huge Pages Size	1048576 kB	1048576 kB
Transparent Huge Pages	madvise	madvise
Automatic NUMA Balancing	Enabled	Enabled
Network Adapter Summary	1x Intel® Ethernet Network Adapter I225-T1, 2x Intel® Ethernet Network Adapter E810-CQDA2	1x Intel® Ethernet Network Adapter I225-T1, 2x Intel® Ethernet Network Adapter E810-CQDA2
Drive Summary	1x 894.3G INTEL SSDSC2KB96, 1x 0B 2309 PRAM, 8x 3.5T INTEL SSDPF2KX038TZ	1x 29.9G Flash Drive FIT, 4x 3.5T INTEL SSDPF2KX038TZ, 1x 894.3G INTEL SSDSC2KB96

5.5 Software Configuration

To perform the performance benchmark test, refer to the software configuration as detailed in Table 4.

Table 4. Software configurations

	TRex node	VPP-L3FWD node
Calico VPP Version	3.24.0	3.24.0
Calico Version	3.24.0	3.24.0
VPP Version	22.02	22.02
Compiler	gcc version 11.3.0 (Ubuntu 11.3.0-1ubuntu1~22.04)	gcc version 11.3.0 (Ubuntu 11.3.0-1ubuntu1~22.04)
DPDK Version	21.11.0	21.11.0
Containerd Version	1.6.18	1.6.18
Kubernetes Version	1.24.4	1.24.4
Isolcpus	None	None
Calico VPP Core Number	1 (main core), 2-9 (worker)	1 (main core), 2-7 (worker, depends on core scaling)
VPP L3FWD Core Number	N/A	28,29-34 (depends on core scaling)
Traffic Generator Type	TRex	N/A
Traffic Generator Version	2.91	N/A
TRex Server Core Number	15 (master thread), 16 (latency), 17-24 (worker)	N/A
TRex DPDK Version	21.02.0-rcl	N/A
Network Adapter Model	Intel® Ethernet Network Adapter E810-CQDA2	Intel® Ethernet Network Adapter E810-CQDA2
Network Adapter Firmware	4.10 0x8001518e 1.3310.0	2.15 0x800049c3 1.2789.0
Network Adapter Driver	1.10.1.2.2	5.15.0-25-generic

6 Calico VPP Multinet Performance

6.1 Packet Size Scaling

For the packet size scaling performance test[1], both SW memif and DSA memif scenarios comply with the following configurations:

- MTU: 1500
- VPP-L3FWD Worker Core Number: 1
- VPP Worker Core Number (L3FWD): 1
- DPDK PCI TX/RX Queue (L3FWD): 1
- TRex Worker Core Number: 8
- VPP Worker Core Number (TRex): 8
- DPDK PCI TX/RX Queue (TRex): 8
- Calico VPP RX Mode: Polling
- Internal Packet Format: UDP
- VPP Worker Core Op Freq (TRex/VPP-L3FWD): 3.8 GHz
- VPP Worker Core utilization (TRex/VPP-L3FWD): 100%
- DSA: 1 instance, 4 engines, 4 work queues

[1] See backup for workloads and configurations. Results may vary.

We tested the throughput of different packet sizes: 128 B, 256 B, 512 B, 1024 B, and 1500 B. The results are shown in Figure 5.

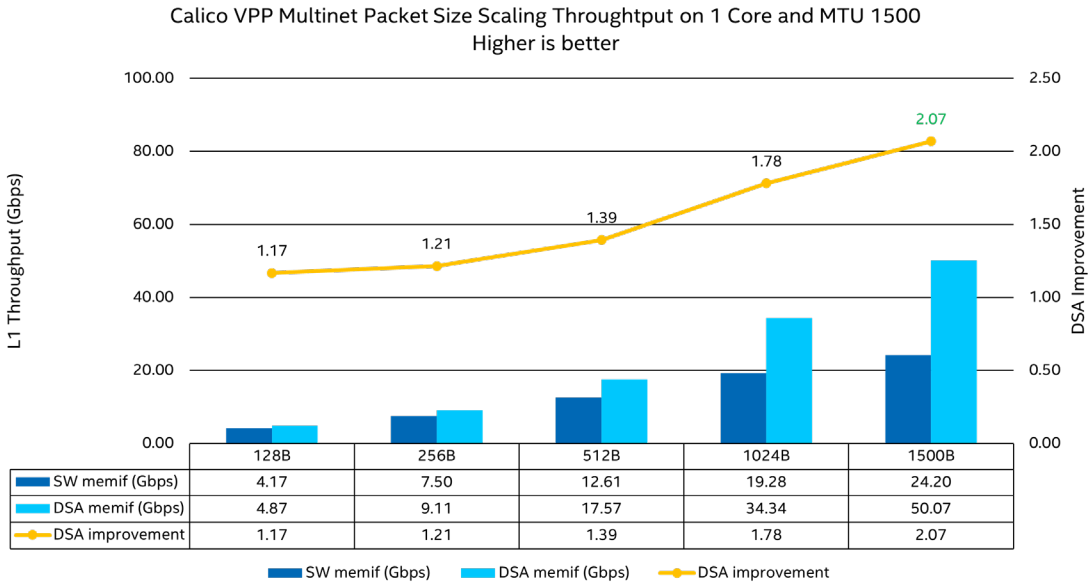


Figure 5. Calico VPP Multinet packet size scaling throughput on one core and MTU 1500

From the performance data of packet size scaling[2], we can see that as the packet size increases, the throughput of Multinet with SW and DSA memif improves. When packet size is increased to 1500 bytes, Multinet performance with DSA memif is up to ~2.07x of SW memif.

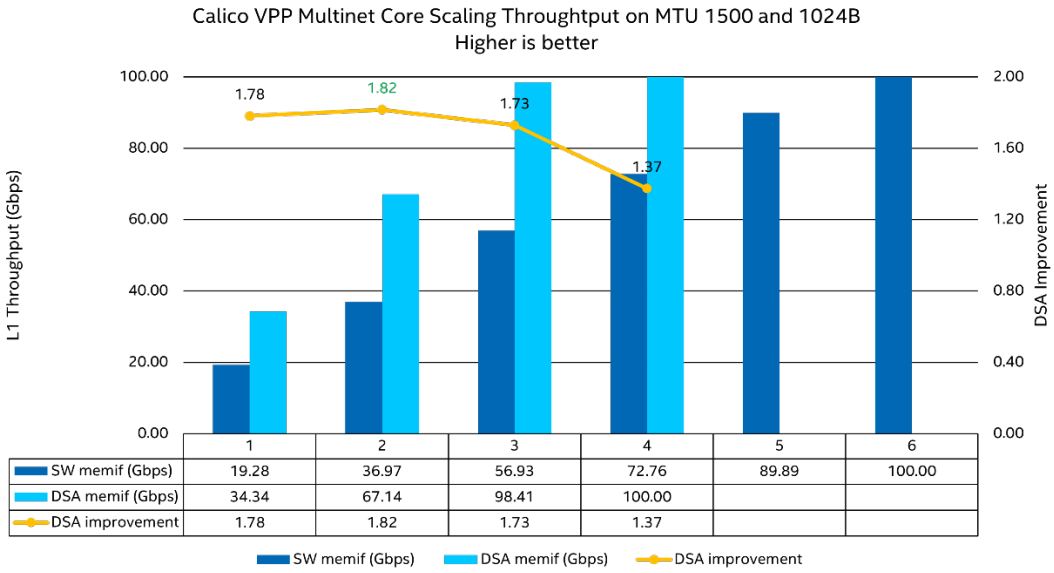
6.2 Core Scaling

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

- MTU: 1500
- TRex Worker Core Number: 8
- VPP Worker Core Number (TRex): 8
- DPDK PCI TX/RX Queue (TRex): 8
- Calico VPP RX Mode: Polling
- Internal Packet Format: UDP
- VPP Worker Core Op Freq (TRex/VPP-L3FWD): 3.8 GHz
- VPP Worker Core utilization (TRex/VPP-L3FWD): 100%
- DSA: 1 instance, 4 engines, 4 work queues

We test the throughput of different VPP worker core numbers of VPP-L3FWD nodes: 1/2/3/4/5/6. The DPDK PCI TX/RX Queue Number on the VPP-L3FWD node and the VPP-L3FWD pod Worker Core Number are the same as the VPP worker core number of the VPP-L3FWD node. The results are shown in Figure 6.

[2] See backup for workloads and configurations. Results may vary.



	1	2	3	4	5	6
SW memif (Gbps)	19.28	36.97	56.93	72.76	89.89	100.00
DSA memif (Gbps)	34.34	67.14	98.41	100.00		
DSA improvement	1.78	1.82	1.73	1.37		

Figure 6. Calico VPP Multinet core scaling throughput on MTU 1500 and packet size 1024 B

When the CPU cores number is four, we can see that Multinet performance with DSA memif achieves 100 Gbps[3]. This throughput is only 37% higher than that of SW memif and is not scaled because the line rate of the network adapter has been reached. From the performance data of core scaling, we conclude that Multinet performance with DSA memif is up to ~1.82x of SW memif when packet size is 1024 B. DSA memif needs ~3 CPU cores to achieve the max throughput, and SW memif needs six cores, so DSA memif can save ~3 CPU cores.

6.3 Comparison of Multinet and Singlenet

We compare the performance data of packet size scaling and core scaling of Multinet and Singlenet to analyze the impact of VXLAN encapsulation and decapsulation in Multinet on performance. The results of packet size scaling and core scaling of Multinet and Singlenet are compared in the following tables.

Table 5. Comparison of packet size scaling results of Multinet and Singlenet with SW memif

Architecture	128 B	256 B	512 B	1024 B	1500 B
Singlenet (Gbps)	5.60	9.87	15.86	24.03	28.83
Multinet (Gbps)	4.17	7.50	12.61	19.28	24.20
Comparison	0.74	0.76	0.80	0.80	0.84

Table 6. Comparison of packet size scaling results of Multinet and Singlenet with DSA memif

Architecture	128 B	256 B	512 B	1024 B	1500 B
Singlenet (Gbps)	6.59	12.35	23.80	46.23	68.82
Multinet (Gbps)	4.87	9.11	17.57	34.34	50.07
Comparison	0.74	0.74	0.74	0.74	0.73

From the results shown in Table 5 and Table 6, we can see that the performance of Multinet with SW memif is ~0.74-0.84x of Singlenet, and the performance of Multinet with DSA memif is ~0.73-0.74x of Singlenet on one VPP core of VPP-L3FWD for different packet sizes. The VPP container of Multinet needs to perform VXLAN encapsulation and decapsulation after receiving data from memif-input and dpdk-input, which brings 16%-26% performance loss to SW memif and 26%-27% performance loss to DSA memif on the one core scenario.

[3] See backup for workloads and configurations. Results may vary.

Table 7. Comparison of core scaling results of Multinet and Singlenet with SW memif

Architecture	1	2	3	4	5	6
Singlenet (Gbps)	24.03	43.46	64.76	78.45	92.49	98.98
Multinet (Gbps)	19.28	36.96	56.93	72.76	89.89	100.00
Comparison	0.80	0.85	0.88	0.93	0.97	1.01

Table 8. Comparison of core scaling results of Multinet and Singlenet with DSA memif

Architecture	1	2	3
Singlenet (Gbps)	46.24	83.17	99.51
Multinet (Gbps)	34.34	67.14	98.41
Comparison	0.74	0.81	0.99

From the results shown in Table 7 and Table 8, the performance[4] of Multinet with SW memif is ~0.8-1x of Singlenet, and the performance of Multinet with DSA memif is ~0.74-0.99x of Singlenet with 1024 B for different VPP core number. As the number of VPP cores increases, the impact of VXLAN encapsulation and decapsulation gradually decreases. Multinet with SW memif achieves the performance close to Singlenet at six cores, and Multinet with DSA memif achieves the performance close to Singlenet at three cores.

6.4 Memif Memory Copy Overhead Analysis

Perf can be used to analyze Multinet with SW memif and DSA memif performance on one CPU core scenario. The Perf data of VPP worker core on VPP-L3FWD node is shown in Table 9.

Table 9. VPP core Perf analysis of Multinet with memif

object	symbol	object_SW	symbol_SW	object_DSA	symbol_DSA
memif_plugin.so	memif_device_class_tx_fn_icl	**38.51%**	27.04%	**9.43%**	3.98%
	memif_input_node_fn_icl		11.28%		2.91%
	mspace_usable_size_with_delta@plt		0.20%		0.10%
	memif_dma_completion_cb_ma_icl				0.78%
	memif_tx_dma_completion_cb_ma_icl				1.65%
libvnet.so.22.06	ip4_input_node_fn_icl	23.15%	3.51%	45.99%	6.71%
	ip4_lookup_node_fn_icl		3.23%		6.42%
	gso_ip4_node_fn		2.54%		8.84%
	ip4_rewrite_node_fn_icl		2.25%		4.69%
	ip4_local_inline		1.66%		2.61%
	vnet_per_buffer_interface_output_node_fn_icl		1.57%		1.74%
	ethernet_input_node_fn_icl		1.27%		2.24%
	lookup_ip4_dst_node_fn		1.27%		2.14%
	vxlan4_input_node_fn_icl		1.27%		2.33%
	vxlan4_encap_node_fn_icl		1.07%		2.91%
	ip4_input_no_checksum_node_fn_icl		0.98%		1.46%
	udp4_local_node_fn_icl		0.78%		1.26%
	vnet_l2_compute_flow_hash		0.68%		1.17%
	vnet_interface_output_node_fn_icl		0.49%		0.49%

[4] See backup for workloads and configurations. Results may vary.

object	symbol	object_SW	symbol_SW	object_DSA	symbol_DSA
	virtio_pre_input		0.39%		0.49%
	ip4_tcp_udp_validate_checksum		0.20%		0.49%
cnat_plugin.so	cnat_input_feature_ip4_node_fn	15.32%	8.19%	18.88%	7.89%
	cnat_output_feature_ip4_node_fn		5.27%		8.08%
	cnat_snat_policy_k8s		1.46%		2.63%
	_hash_get@plt		0.39%		0.29%
dpdk_plugin.so	ice_xmit_pkts	9.59%	5.07%	10.01%	5.64%
	ice_recv_scattered_burst_vec_avx512_offload		2.08%		1.85%
	dpdk_input_node_fn_icl		1.36%		0.39%
	dpdk_device_class_tx_fn_icl		1.07%		2.03%
	__tls_get_addr@plt				0.10%
libvppinfra.so.22.06	lookup	11.03%	8.69%	11.01%	7.29%
	get_indirect		1.36%		1.41%
	_hash_get		0.49%		1.53%
	mspace_usable_size_with_delta		0.49%		0.78%
libvlib.so.22.06	vlib_worker_loop	1.95%	0.88%	3.30%	1.07%
	vlib_get_next_frame_internal		0.39%		1.16%
	vlib_buffer_enqueue_to_next_fn_icl		0.39%		0.29%
	vlib_put_next_frame		0.10%		0.39%
	mspace_usable_size_with_delta@plt		0.10%		0.10%
	vlib_frame_alloc_to_node		0.10%		0.29%
dsa_intel_plugin.so	intel_dsa_batch_submit			0.71%	0.51%
	intel_dsa_batch_new				0.10%
	intel_dsa_node_fn				0.10%

From the results shown in Table 9, we can see that the CPU usage of `libvnet_plugin` is significantly increased[5], and Intel DSA saves up to ~29.08% CPU computing capacity. The libvnet.so provides implementation of the network protocol stack in VPP. It implements various layers in the protocol stack, including link layer, network layer, transport layer, and implements various protocols, such as Ethernet, IPv4, IPv6, TCP, and UDP. It also provides functions such as routing, forwarding, and filtering of data packets, as well as support for various network devices. So, this Perf data means that the use of Intel DSA enables the 4th Gen Intel Xeon Scalable processor to have more capacity to deal with the work of the network protocol stack. Therefore, Intel DSA improves throughput.

6.5 EMON Analysis

We use EMON to analyze Multinet with SW memif and DSA memif performance on one CPU core scenario. The EMON data of Multinet on VPP-L3FWD node is shown in Table 10.

[5] See backup for workloads and configurations. Results may vary.

Table 10. EMON data of Multinet on VPP-L3FWD node

	SW memif	DSA memif	Ratio
metric_CPU operating frequency (GHz)	3.7	3.7	
metric_CPU utilization %	100	100	
metric_CPU utilization % in kernel mode	0.15%	0.14%	
metric_CPI (cycle per instruction)	**0.41**	**0.25**	**0.61**
IPC (=1/metric_CPI) (instruction per cycle)	**2.45**	**4.07**	**1.66**
metric_kernel_CPI (cycle per instruction)	1.92	1.8	
metric_TMA_Frontend_Bound (%)	4.69	5.49	
metric_TMA_..Fetch_Bandwidth (%)	2.68	3.94	
metric_TMA_Bad_Speculation (%)	0.79	1.58	
metric_TMA_Backend_Bound (%)	54.94	26.79	
metric_TMA_..Memory_Bound (%)	**44.2**	**11.35**	**0.26**
metric_TMA_....L1_Bound (%)	4.69	3.05	
metric_TMA_....L2_Bound (%)	1.3	1.13	
metric_TMA_....L3_Bound (%)	**32.52**	**3.66**	**0.11**
metric_TMA_Retiring (%)	**39.57**	**66.13**	**1.67**

From the results shown in Table 10, we can see that the memory-bound ratio of DSA memif is only 0.26x of SW memif; L3-bound ratio of DSA memif is only 0.11x of SW memif; Retiring ratio of DSA memif is 1.67x of SW memif; IPC of DSA memif is 1.66x of SW memif. These metrics show the advantages of DSA memif over SW memif.

7 Summary

This document introduces Calico VPP, Memif, Multinet, Intel DSA technology, the optimized Calico VPP Multinet memif, and workload architecture. It also introduces in detail the VPP graph node of Multinet, the impact of VXLAN encapsulation on the packet, the calculation method of KPI, and the configuration of the system and software. The performance data of packet size scaling and core scaling are tested to obtain the performance gain of Intel DSA for memif. We also compare the performance of Singlenet and Multinet to see the impact of VXLAN encapsulation on Multinet performance. Finally, we use Perf to analyze the reasons for the performance gain brought by Intel DSA and EMON to show the advantages of DSA memif over SW memif.

This guide demonstrates how Intel DSA can increase Calico VPP Multinet with memif performance[6] by accelerating memory copy, giving the CPU more cycles to process the network protocol stack. DSA memif performance is up to ~2.07x of SW memif when packet size is 1500 B. DSA memif needs ~3 CPU cores to achieve the max throughput (100 Gbps), and SW memif needs six cores when packet size is 1024 B, so DSA memif can save ~3 CPU cores.

[6] See backup for workloads and configurations. Results may vary.

0124/DN/WIT/PDF 787129-002US