MOBILE PACKET CORE PERFORMANCE INCREASES ON 2ND GENERATION INTEL® XEON® SCALABLE PROCESSORS

Contents

1.0 Executive Summary

This white paper presents performance data for several virtualized evolved packet core (EPC) and 5G Core Network (5GCN) functions running on three SKUs of 2nd Gen Intel® Xeon® Scalable processors, formally codenamed Cascade Lake. Intel engineers tested against typical enhanced Mobile Broadband (eMBB) and fixed wireless access (FWA) traffic models and pipelines. The testing was focused on user plane performance for both control and user plane separation (CUPS)-based EPC implementations and the 5G user plane functions (UPFs). The results demonstrate performance continues to improve steadily in this domain, with mid-range or Intel® Xeon® Gold processors delivering throughput in the range of 100 gigabits-per-second (Gbps) per socket for typical eMBB workloads and in excess of 250 Gbps per socket for FWA workloads. The forwarding throughput and virtual machine (VM) utilization rate for five use cases are summarized in Table 1.

Intel® Processor	Cores/CPU	Cores Used/CPU	Forwarding Throughput	Virtual Machine Utilization Rate	Use Case
Intel® Xeon® Gold 6230 Processor	20	17.5	212 Gbps on 2 CPU	78%	eMBB vEPC
Intel® Xeon® Gold 6230N Processor	20	17.5	211 Gbps on 2 CPU	76%	eMBB vEPC
Intel Xeon Gold 6230N Processor	20	19	210 Gbps on 2 CPU	69%	eMBB vUPF
Intel Xeon Gold 6230N Processor	20	19	222 Gbps on 1 CPU	63%	vFWA
Intel® Xeon® Platinum 8280 Processor	28	24	253 Gbps on 1 CPU	73%	vFWA

Table 1. Throughput Measurements for Three Intel® Xeon® Scalable Processors Running Various Packet Forwarding Workloads

This high performance is achieved entirely with software, meaning no hardware acceleration or offload of the packet core pipeline was employed. The packets per second and aggregate throughput performance have a very high level of determinism for all use cases, including both EPC and 5G core network (5GCN) pipelines. Moreover, the forwarding function throughput is linear with core counts (i.e., between 2 and 20 cores). In the future, Intel will publish more packet core benchmarks with the next generation of Intel® Xeon® processors and additional use case pipelines.

In addition, the forwarding throughput of 2nd Gen Intel Xeon Scalable processors (Cascade Lake) was approximately 25% higher than those measured on prior generation Intel Xeon Scalable processors (Skylake) for equivalent core count, as described in section 4.0.

2.0 Test Environment

ASTRI (Applied Science and Telecoms Research Institute - www.astri.org), based in Hong Kong, supplied the 4G EPC and 5GCN stacks used to test the five use cases. ASTRI gave Intel permission to use and modify their source code, which was used to demonstrate typical pipeline performance (via compiled binaries) and share best-known-methods with the communications service providers (CoSPs) and telecom vendor communities.

The test environment for characterizing the Intel Xeon processor-based platforms consisted of standard test equipment (Spirent Landslide). It tested the user plane performance for various call models and use cases, and not an entire EPC or 5GCN.

2.1 Core Network Overview

Figure 1 shows the 4G non-roaming architecture from the European Telecommunications Standards Institute (ETSI) in technical specification (TS) 23.401. The goal of this study was to maximize the performance of the serving gateway (SGW) and packet data network gateway elements (PGW). All other elements interfaced to the device under test (DUT) were either simulated or placed on external servers.

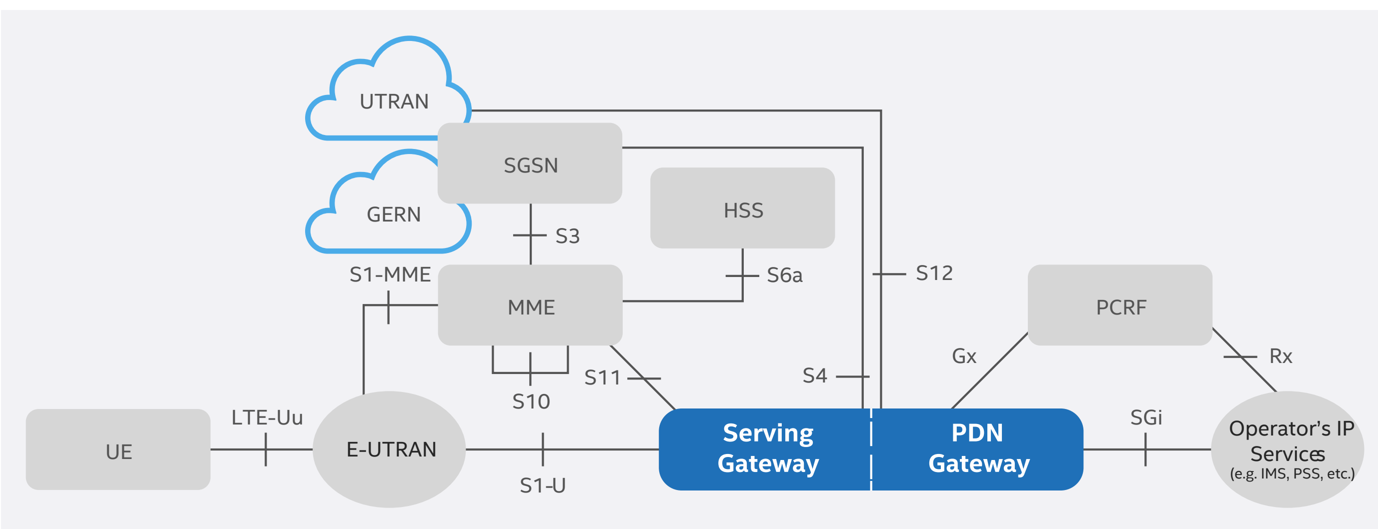


Figure 1. Non-Roaming 4G Architecture

The stack under test was Release 14 CUPS-enabled, as per TS 23.214 (Figure 2). More specifically, the DUT ran the user plane workloads, and the control plane workloads were external to the system.

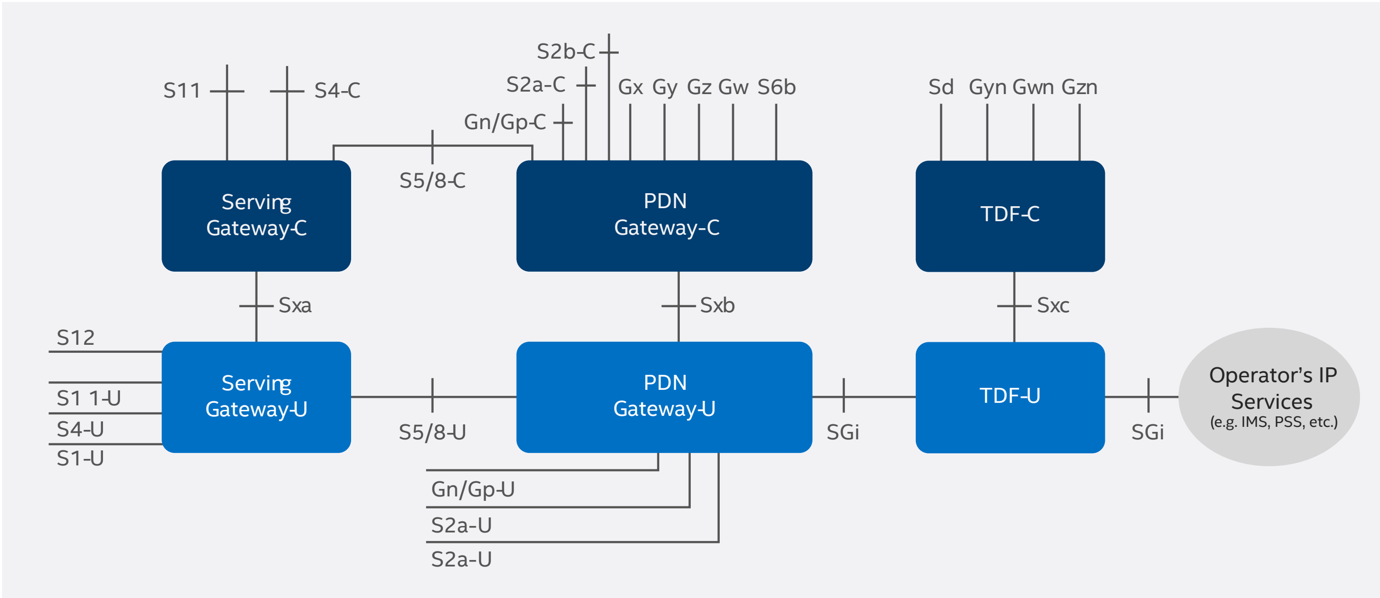


Figure 2. CUPS EPC Architecture

Figure 3 shows the 5G non-roaming reference model from TS 23.501. User plane processing is shown along the bottom of the figure: from the user entity (UE) to the radio access network (RAN), to the N3 interface, to the UPF and N6 IP encapsulation, and to the external data network (DN). All other functions are control plane orientated, such as for managing authentication, mobility, charging, slicing, and policy.

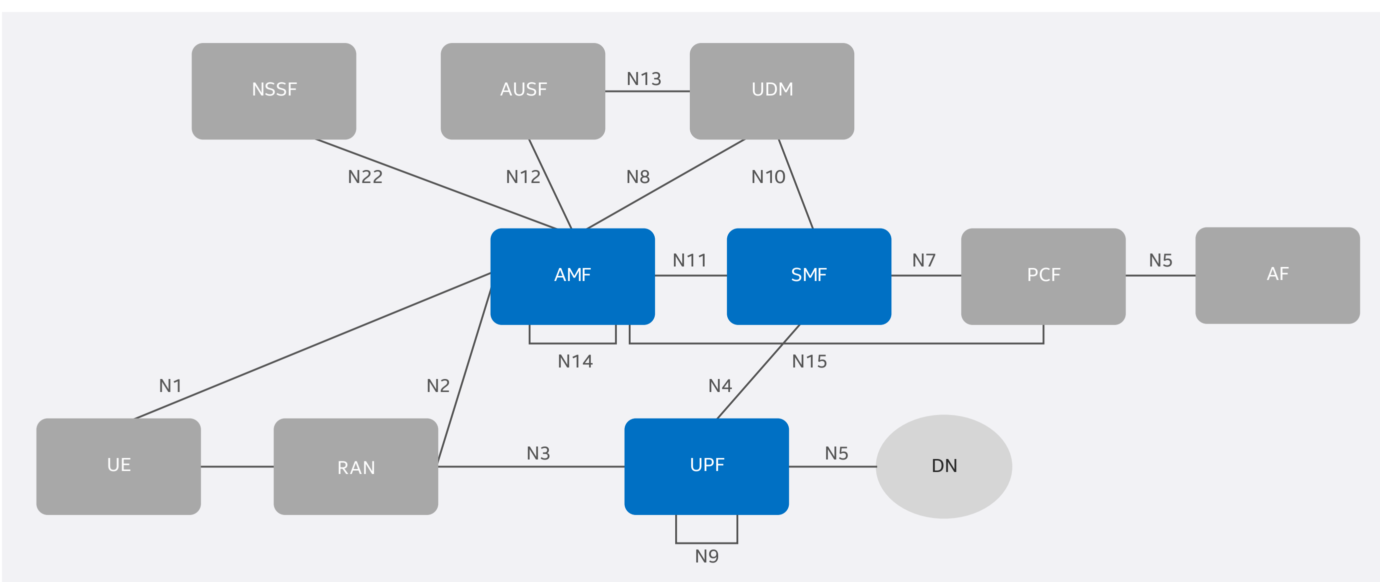


Figure 3. Non-Roaming 5G Architecture

The testing focused on the performance of the UPF, as this is the main forwarding element in the core architecture. Other elements, such as the UE, RAN, and DN, were simulated by testers. The Authentication Management Function (AMF) and Session Management Function (SMF) were used to test the user plane as it established, deactivated, and moved session in response to mobility events.

2.2 Test Infrastructure

2.2.1 4G Harness

The 4G testbed is shown in Figure 4. At the bottom left of the figure, the Spirent Landslide generates uplink traffic. This traffic transits the top-of-rack (ToR) switch and crosses the EPC user plane DUT. The Gi traffic is then terminated in external servers on the bottom right. These servers terminate the Gi uplink and reflect the downlink reply to the bearers; they also sink the downlink after it has traversed the DUT (sGi to S1). This harness is used to stress the system for different uplink/downlink ratios, depending on the desired CoSP use case. The Mobility Management Entity (MME), Serving Gateway Control Plane Function (SGW-C), PDN Gateway Control Plane Function (PGW-C), and Traffic Detection Control Plane Function (TDF-C) are hosted on a separate control plane server, shown in the top right.

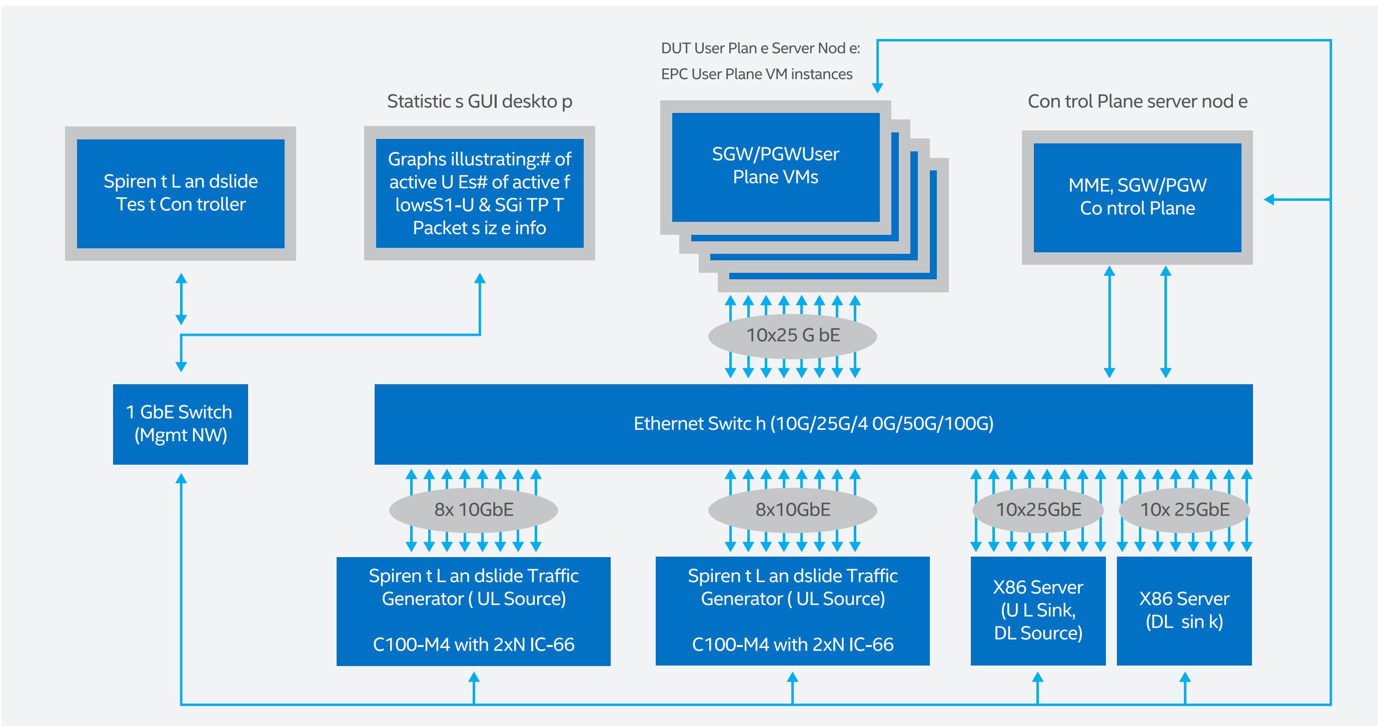


Figure 4. 4G Test Harness

For additional clarity, the traffic flows are shown in Figure 5. The uplink traffic (in red) is generated by the Spirent Landslide, and the downlink traffic (in blue) is generated by the reflection server.

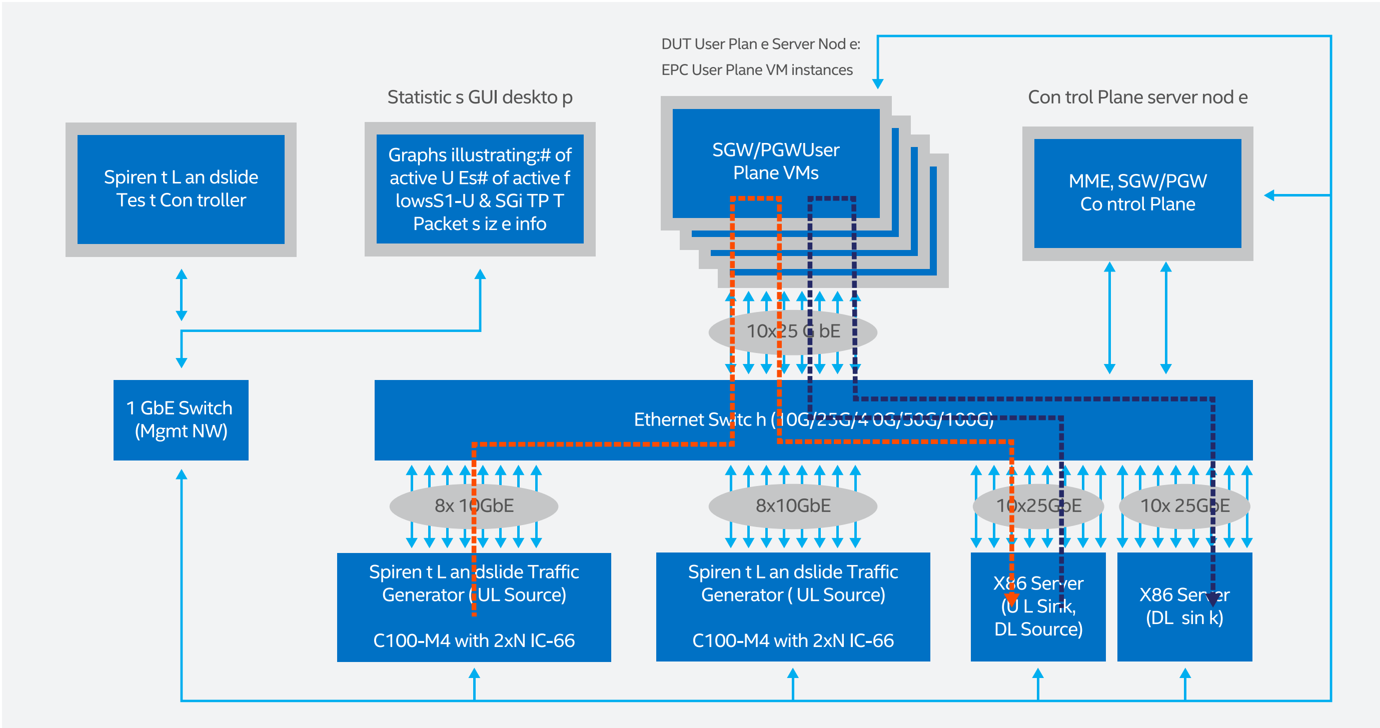


Figure 5. 4G Test Harness Traffic Flows

2.2.2 5G Harness

The 5G system uses a similar infrastructure model, as shown in Figure 6. Here, the DUT hosts the UPF, and the AMF and SMF functions are hosted by the server shown in the upper right box.

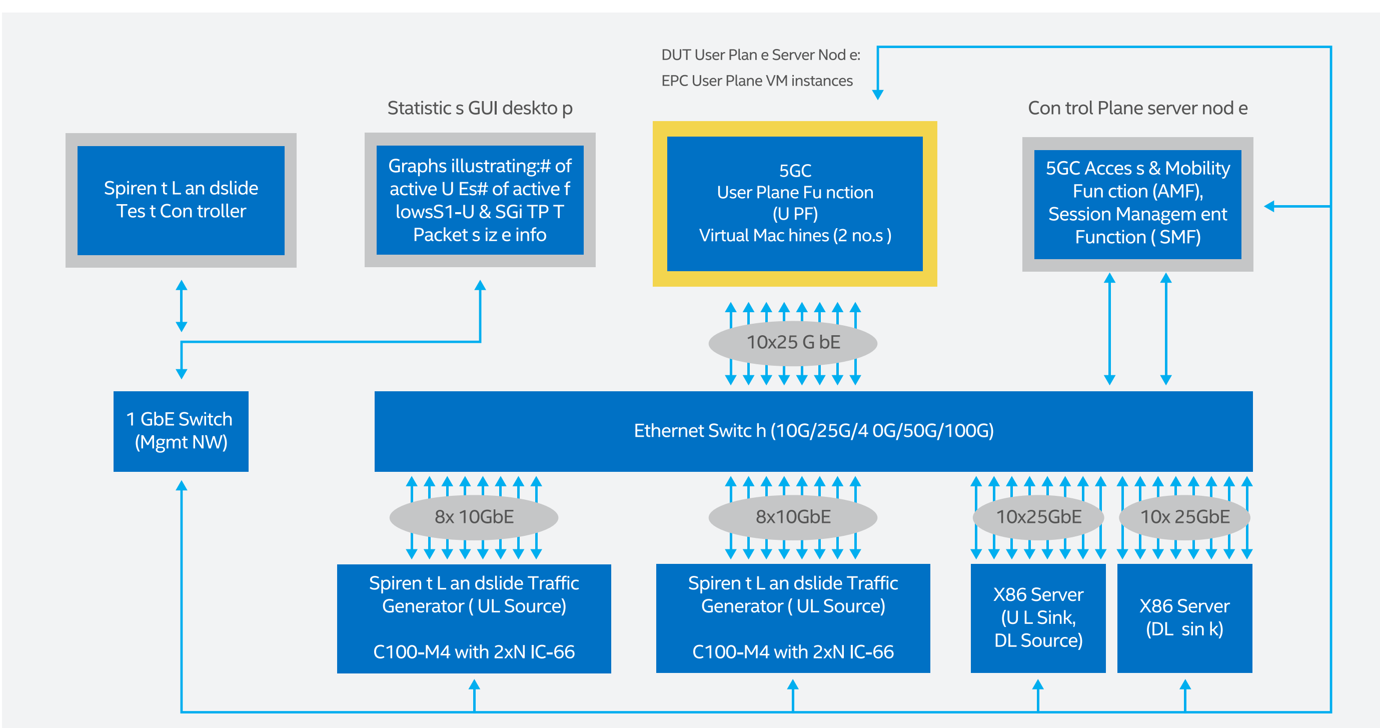


Figure 6. 5G Test Harness

2.3 Test Methodology

Two use cases using different Intel processors were run across the 4G and 5G harnesses: Enhanced Mobile Broadband (eMBB) and fixed wireless access (FWA). The two pipelines differ as follows:

1. eMBB: All user plane functions switched on for vEPC and 5GCN, as shown in Figures 7 and 8.
2. FWA: No charging, deep packet inspection (DPI), quality of service (QoS), access control lists (ACLs), or mobility.

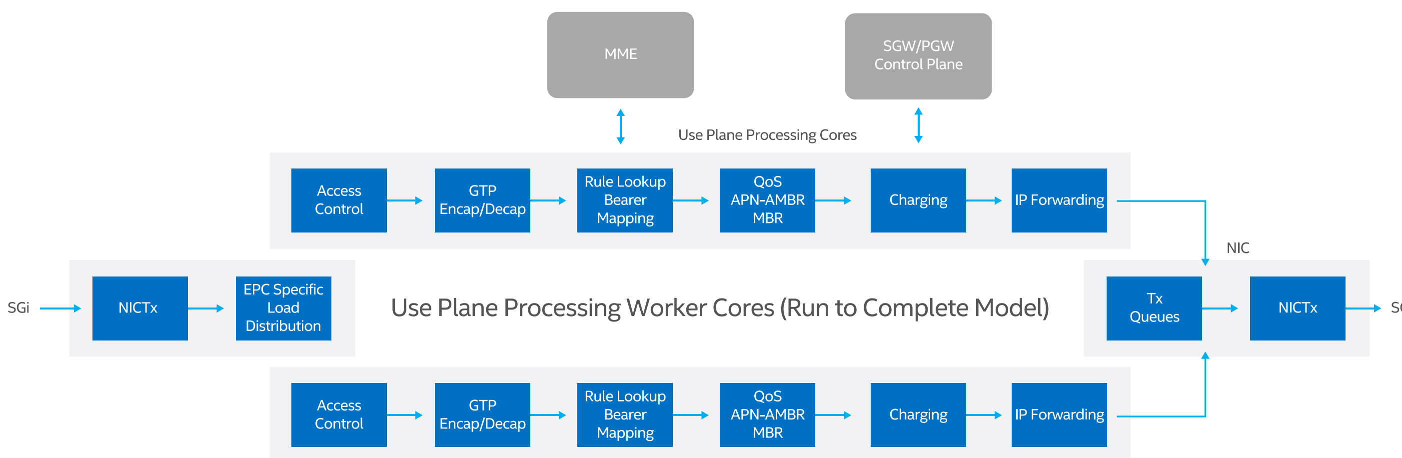


Figure 7: 4G eMBB Pipeline

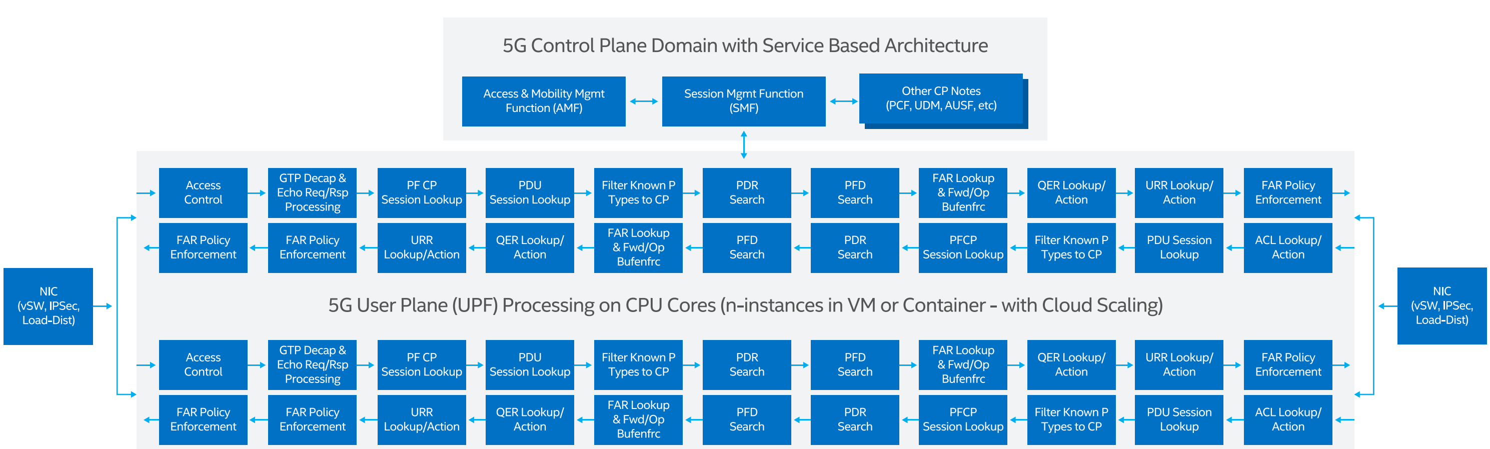


Figure 8: 5GCN eMBB Pipeline

The traffic model for the base setup is shown in Table 2. Changes to the base setup for a given test are indicated in the results.

UE Count	100KUEs, 2Mbits per UE in connected state on user plane
Bearers	2 bearers per UE:1 default,1 dedicated
Flows	4 flows per bearer, 800Kflows per DUT
IMIX	Packet size combination of 64B, 128B, 256B, 512B, 1024B, 1400B
Traffic Profiles	5:5 –50% uplink, 50% downlink 2.5:7.5 –25% uplink, 75% downlink 4:6 –40% uplink, 60% downlink 1:9 –10% uplink, 90% downlink
Pipeline Supports	Mobility, Downlink Data notifications, fragmentation/reassembly, network address translation (NAT), control plane-based charging data record (CDR) generation, but not included specifically in test scenario.
Pipeline Does Not Support	IPSec or DPI –These functions are being integrated into future testing

Table 2. Base Setup for the Traffic Model

2.4 Device under Test (DUT)

During testing that was conducted and concluded in March of 2019, the EPC and 5GCN forwarding functionality ran on an Intel reference board (Figure 9), featuring balanced I/O system with 2x16 PCI Express Gen3 lanes and 96 GB of DRAM (6x16GB) per socket.

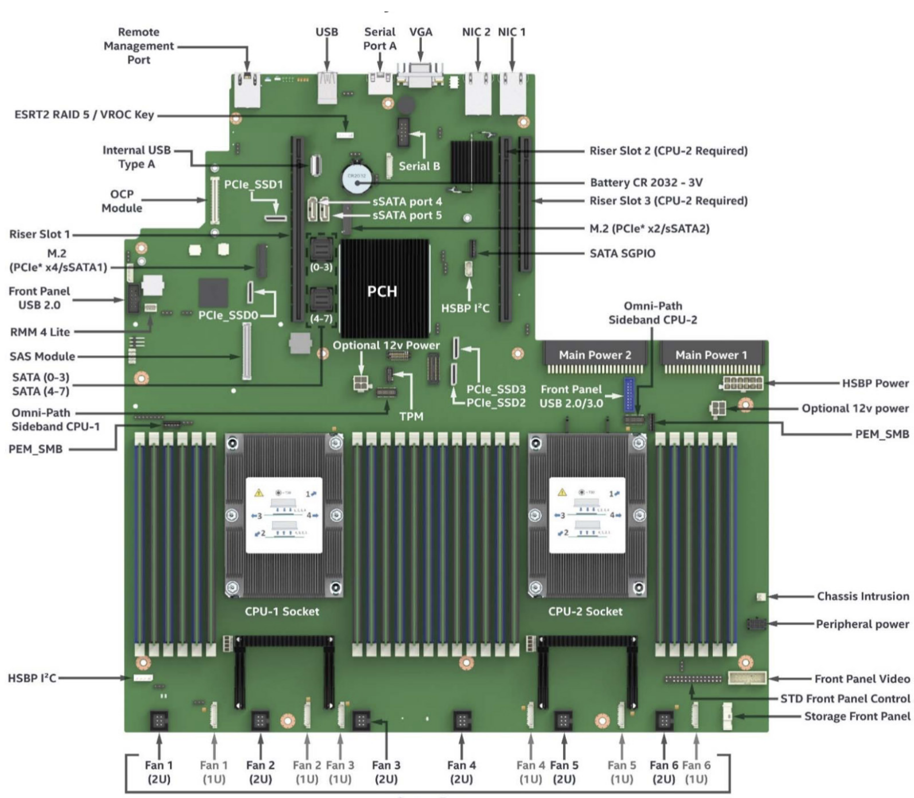


Figure 9: Intel Reference Board

3.0 Results

The following descriptions provide a high-level summary of testing of the five use cases. The platform details for the five use cases are shown in Appendix 1.

3.1 Intel® Xeon® Gold 6230 Processor: vEPC and eMBB

The VM mapping for the DUT is shown in Figure 10.

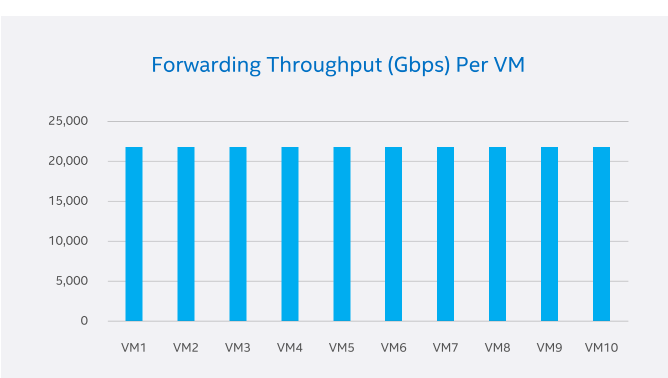
0		1		2		3		4		8		9		10		11		12		16		17		18		19		20		24		25		26		27		28	
0	40	1	41	2	42	3	43	4	44	5	45	6	46	7	47	8	48	9	49	10	50	11	51	12	52	13	53	14	54	15	55	16	56	17	57	18	58	19	59
Host OS	2	VM1							VM2							VM3							VM4							VM5						1	3	4	5

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

0		1		2		3		4		8		9		10		11		12		16		17		18		19		20		24		25		26		27		28	
20	60	21	61	22	62	23	63	24	64	25	65	26	66	27	67	28	68	29	69	30	70	31	71	32	72	33	73	34	74	35	75	36	76	27	77	28	78	29	79
Host OS	7	VM6							VM7							VM8							VM9							VM10						6	8	9	10

Figure 10: VM Mapping for the Dual Intel® Xeon® Gold 6230 Processors

Figure 11 shows the Intel reference board delivered consistent performance per VM, with an average forwarding throughput of 21.2 Gbps per VM and 212 Gbps (as measured on S1 aggregate UL + DL) for the entire board.



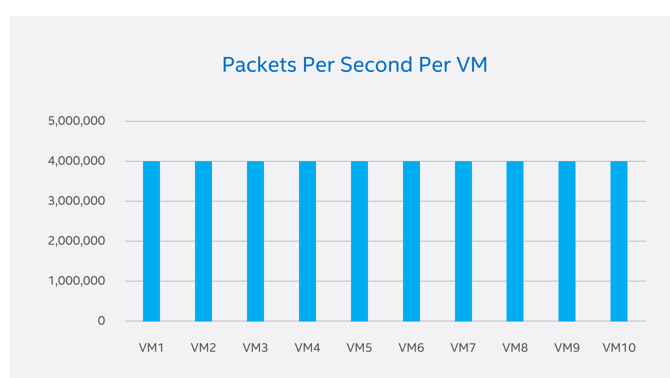


Figure 11: Dual Intel® Xeon® Gold 6230 Processor Performance: Forwarding Throughput and Packets per Second per Virtual Machine (VM)

The packet per second performance was also consistent among VMs, and the CPU core utilization rate varied between 74 and 78% across all VMs.

3.2 Intel® Xeon® Gold 6230N Processor: vEPC and eMBB

The VM mapping for the DUT is shown in Figure 12.

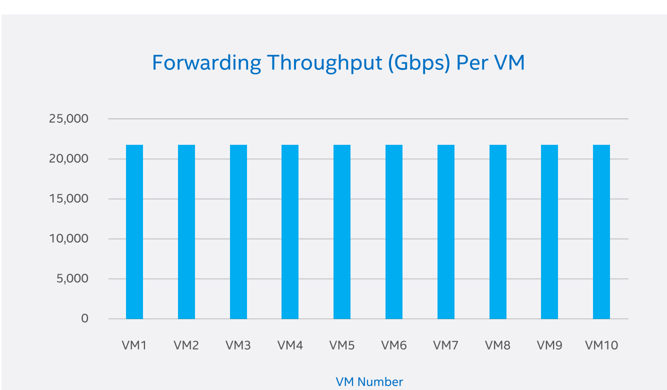
0		1		2		3		4		8		9		10		11		12		16		17		18		19		20		24		25		26		27		28	
0	40	1	41	2	42	3	43	4	44	5	45	6	46	7	47	8	48	9	49	10	50	11	51	12	52	13	53	14	54	15	55	16	56	17	57	18	58	19	59
Host OS	2	VM1							VM2							VM3							VM4							VM5						1	3	4	5

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

0		1		2		3		4		8		9		10		11		12		16		17		18		19		20		24		25		26		27		28	
20	60	21	61	22	62	23	63	24	64	25	65	26	66	27	67	28	68	29	69	30	70	31	71	32	72	33	73	34	74	35	75	36	76	27	77	28	78	29	79
Host OS	7	VM6							VM7							VM8							VM9							VM10						6	8	9	10

Figure 12: VM Mapping for the Dual Intel® Xeon® Gold 6230N Processors

Figure 13 shows the Intel reference board delivered consistent performance per VM, with an average forwarding throughput of 21.1 Gbps per VM and 211 Gbps (as measured on S1 aggregate UL + DL) for the entire board.



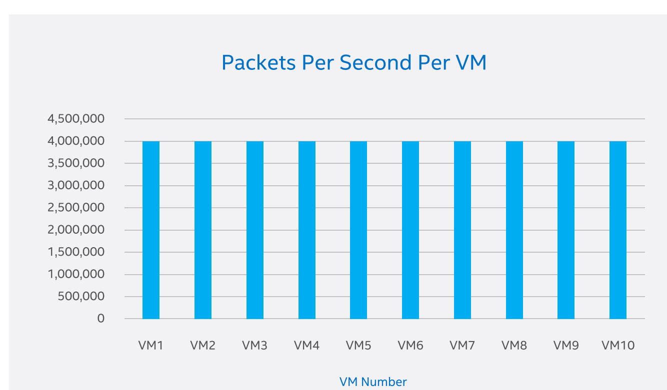


Figure 13: Dual Intel® Xeon® Gold 6230N Processor Performance: Forwarding Throughput and Packets per Second per Virtual Machine (VM)

Typically, we see about 10% performance increase from 6230 to 6230N as there is a frequency increase from 2.1 to 2.3GHz. It was not possible to show this performance increase in this scenario as the DUT was IO bound.

3.3 Intel® Xeon® Gold 6230N Processor: vUPF and eMBB

The VM mapping for the DUT is shown in Figure 14. In this test case, an entire CPU was assigned to a given UPF, and no performance degradation occurred. In other words, there was no penalty on Intel® architecture when scaling up VMs or container sizes from a low number to a high number of cores. The system consistently exhibited small deviation from minimum, average to maximum core utilization rate as traffic was scaled across the infrastructure.

SOCKET 0																																							
0		1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		19	
0	40	1	41	2	42	3	43	4	44	5	45	6	46	7	47	8	48	9	49	10	50	11	51	12	52	13	53	14	54	15	55	16	56	17	57	18	58	19	59
	0	1	20	2	21	3	22	4	23	5	24	6	25	7	26	8	27	9	28	10	29	11	30	12	31	13	32	14	33	15	34	16	35	17	36	18	37	19	38
Host OS House Keeping	VM1 House	VM1																																					

SOCKET 1																																							
20		21		22		23		24		25		26		27		28		29		30		31		32		33		34		35		36		37		38		39	
20	60	21	61	22	62	23	63	24	64	25	65	26	66	27	67	28	68	29	69	30	70	31	71	32	72	33	73	34	74	35	75	36	76	37	77	38	78	39	79
	0	1	20	2	21	3	22	4	23	5	24	6	25	7	26	8	27	9	28	10	29	11	30	12	31	13	32	14	33	15	34	16	35	17	36	18	37	19	38
Host OS House Keeping	VM2 House	VM2																																					

Figure 14: VM Mapping for the Dual Intel® Xeon® Gold 6230N Processors

Figure 15 shows the Intel reference board delivered consistent performance per VM, with an average forwarding throughput of 105.2 Gbps on VM1 and 210.4 Gbps (as measured on S1 aggregate UL + DL) for the entire board.

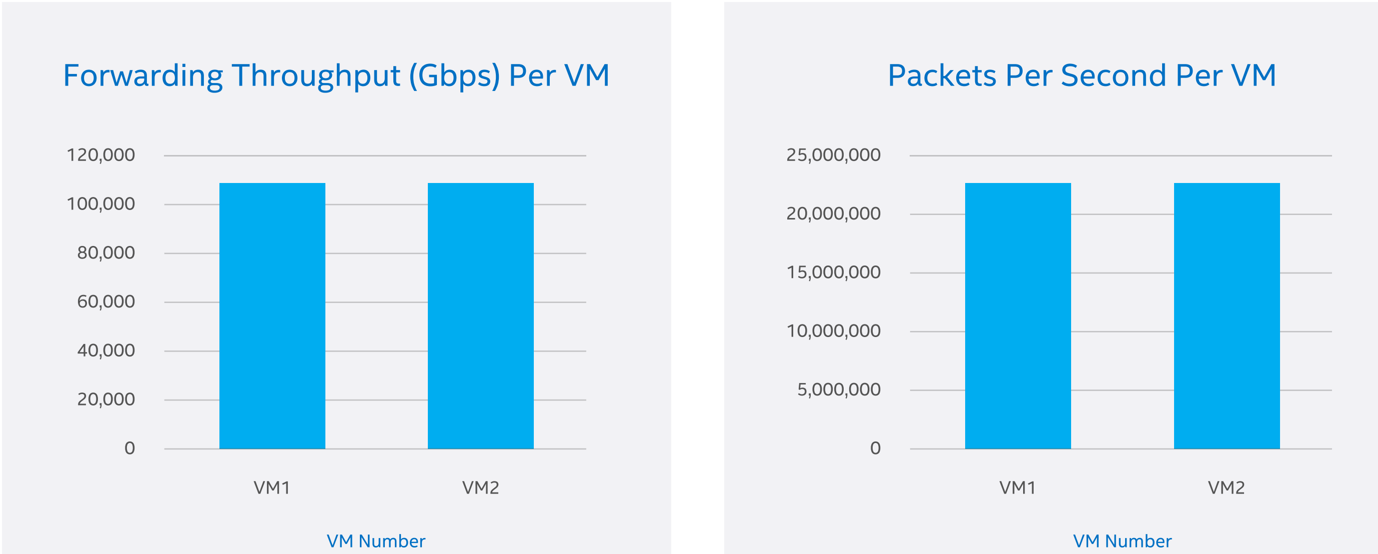


Figure 15: Dual Intel® Xeon® Gold 6230N Processor Performance: Forwarding Throughput and Packets per Second per Virtual Machine (VM)

The packet per second performance was similar for both VMs, more than 21 million packets per second and CPU core utilization rates averaging at 89% for both VMs.

3.4 Intel® Xeon® Gold 6230N Processor: vFWA

The VM mapping for the DUT is shown in Figure 16.

0		1		2		3		4		8		9		10		11		12		16		17		18		19		20		24		25		26		27		28	
0	40	1	41	2	42	3	43	4	44	5	45	6	46	7	47	8	48	9	49	10	50	11	51	12	52	13	53	14	54	15	55	16	56	17	57	18	58	19	59
Host OS	Unused																																						

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

0		1		2		3		4		8		9		10		11		12		16		17		18		19		20		24		25		26		27		28	
20	60	21	61	22	62	23	63	24	64	25	65	26	66	27	67	28	68	29	69	30	70	31	71	32	72	33	73	34	74	35	75	36	76	37	77	38	78	39	79
VM1			VM2			VM3			VM4			VM5			VM6			VM7			VM8			VM9			VM10			VM11			VM12			1	4	7	10
																																				2	5	8	11
																																				3	6	9	12

NIC-1	NIC-2	NIC-3	NIC-4	NIC-5	NIC-6 From CPU-0
86	88	af	b1	da	18

Figure 16: VM Mapping for the Dual Intel® Xeon® Gold 6230N Processors

Figure 17 shows the Intel reference board delivered consistent performance for all 12 VMs, with an average forwarding throughput of 18.5 Gbps and 222 Gbps (as measured on S1 aggregate UL + DL) for the entire board.

Figure 17: Dual Intel® Xeon® Gold 6230N Processor Performance: Forwarding Throughput and Packets per Second per Virtual Machine (VM)

The packet per second performance was consistent across the VMs in the system, and CPU core utilization rates averaged 63% across all VMs with very little variance.

3.5 Intel® Xeon® Platinum 8280 Processor: vFWA

The VM mapping for the DUT is shown in Figure 18. As mentioned previously, the vFWA was a simplified pipeline. For this test, 120K user entities (UEs) were run (with just one bearer per UE with two flows) in order to achieve higher throughput. The number of EPC user plane VMs was increased to 12 on Socket1, and NICs were added to drive more capacity.

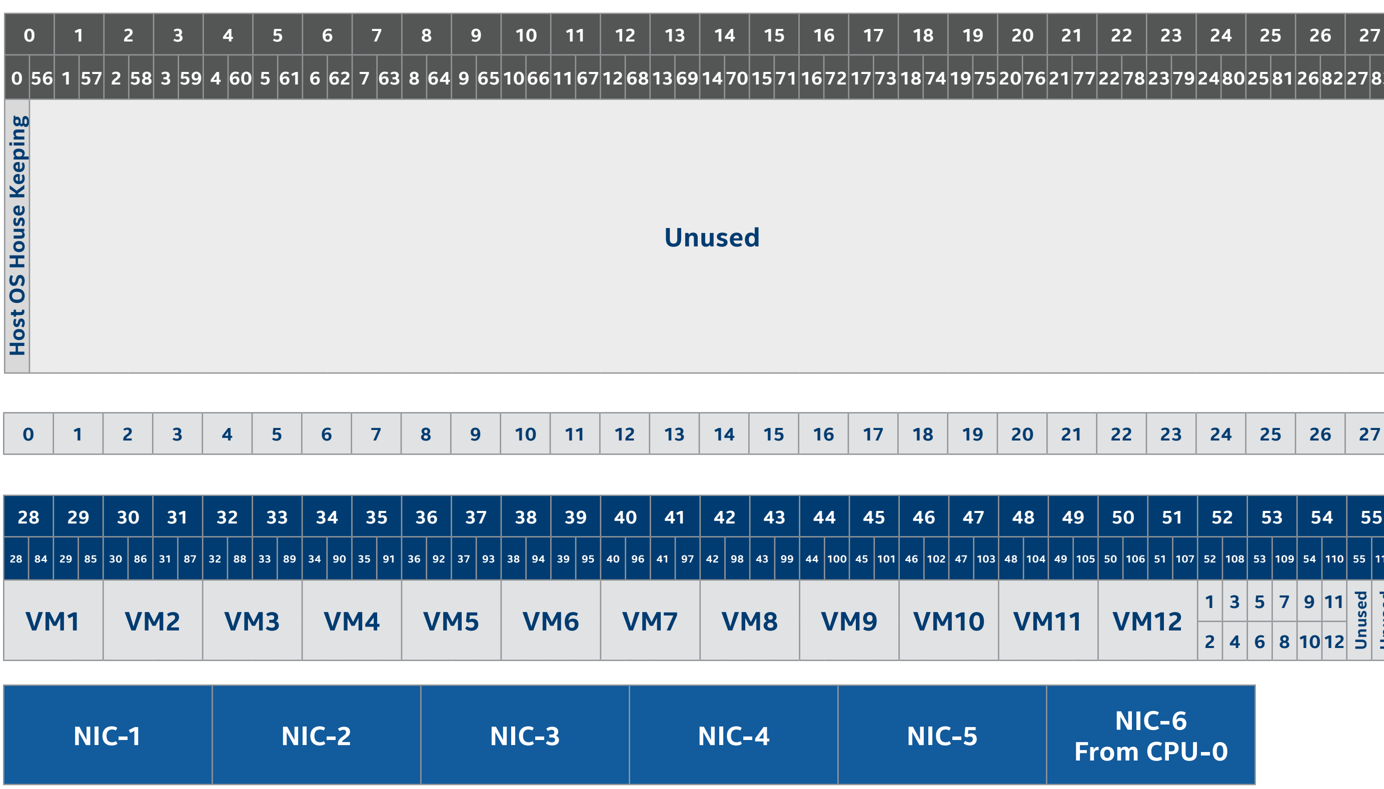


Figure 18: VM Mapping for the Dual Intel® Xeon® Platinum 8280 Processors

Figure 19 shows the performance on single CPU socket.

- Packet Rate: 31.7 MPPS (8.9 MPPS uplink, 22.8MPPS downlink)
- Aggregate TPT: 253 Gbps (26.7 Gbps uplink, 226 Gbps downlink)
- CPU utilization: 73%

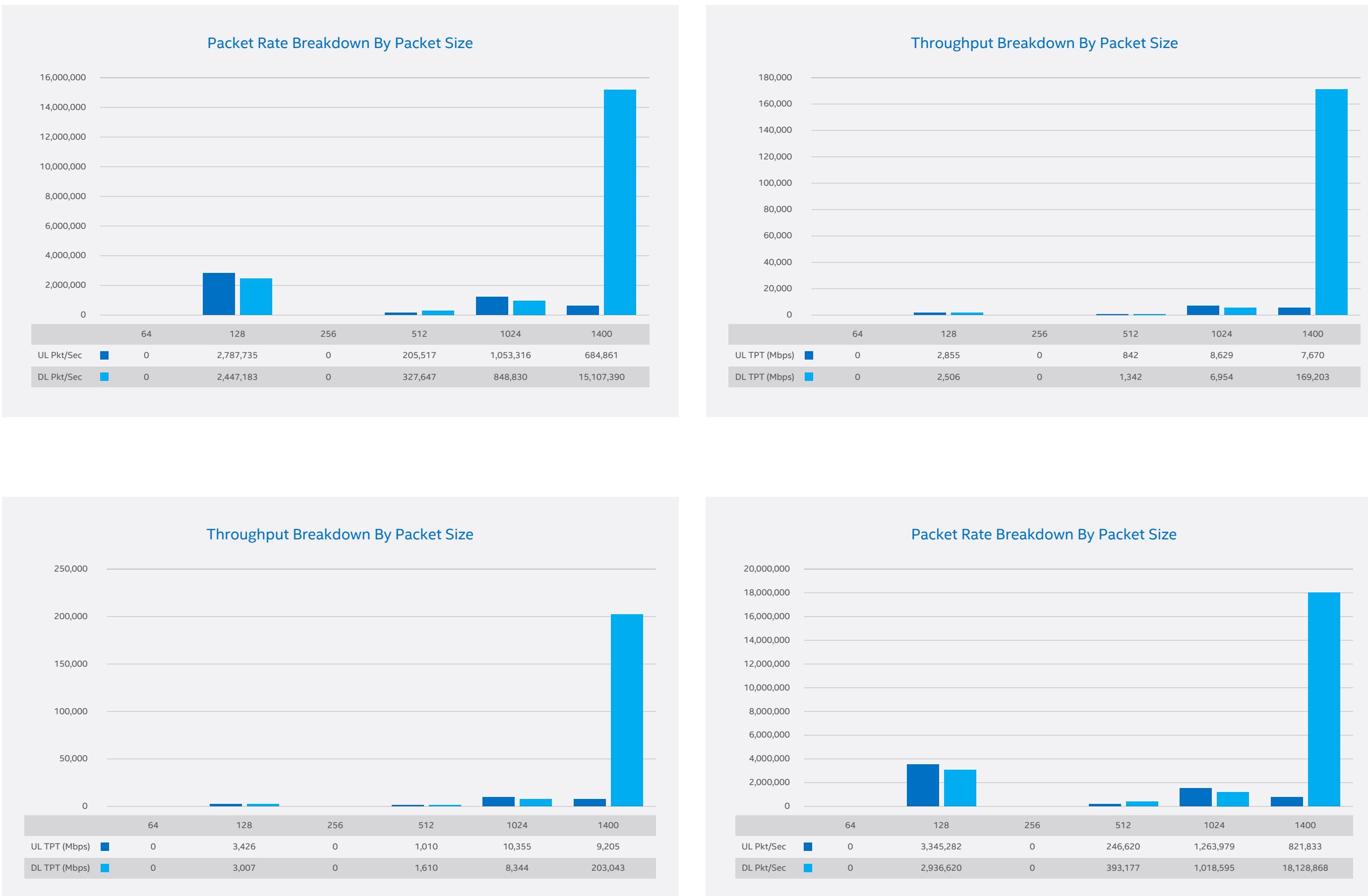


Figure 19: Dual Intel® Xeon® Platinum 8280 Processor Performance: Forwarding Throughput and Packets per Second per Virtual Machine (VM)

4.0 Conclusion

Figure 20 shows throughput performance increased up to 25% from the Intel® Xeon® Gold 6130 processor (from the Skylake family) to the 2nd Generation Intel Xeon Gold 6230N processor (from the Cascade Lake family). Testing of the vEPC pipeline showed the Intel Xeon Gold 6230N processor delivered 25% more performance than the prior generation Intel Xeon Gold 6130 at a 7% lower CPU utilization rate. The Intel Xeon Gold 6230N processor also achieved similar performance as the Intel® Xeon® Platinum 8280 processor (also from the Cascade Lake family) at a slightly lower utilization rate.

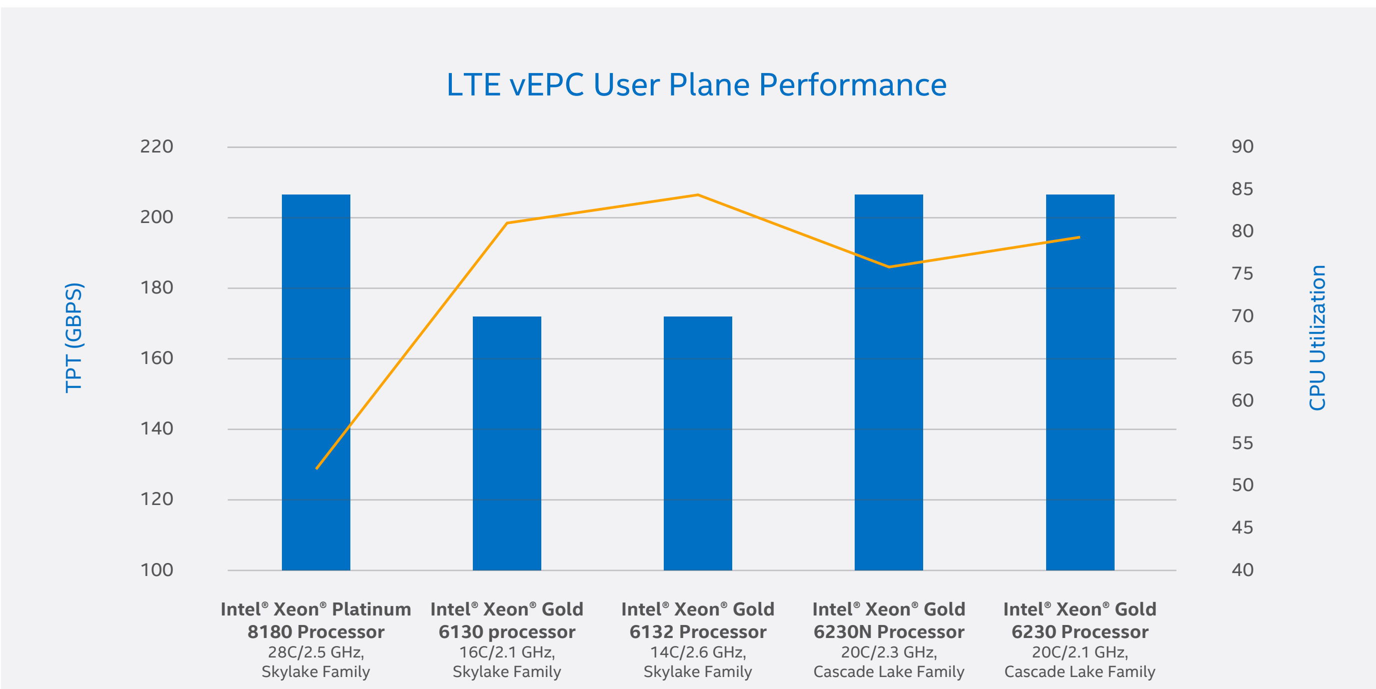


Figure 20: vEPC Performance for Various Intel® Xeon® Processors

The results demonstrated the tested Intel Xeon processors achieved a very high level of determinism across EPC and 5GCN UPF workloads, and this performance scaled linearly with the number of cores as they were scaled up from a few to many cores. The variation in CPU core utilization was also low. This performance consistency can help CoSPs and telecom equipment manufacturers (TEMs) better predict VNF performance and plan for capacity requirements.

5.0 Conclusion

This white paper shows 2nd Gen Intel Scalable processors offer around 25% more performance than predecessor generation processors without any additional platform tuning. The Intel reference board exhibited deterministic performance for 4G and 5G pipelines, even load distribution, and consistent throughput and latency. These results further verify the competitiveness and capability of running packet core workloads in software on Intel architecture without any hardware acceleration and associated orchestration.

6.0 Appendix: Platform Details

Use Cases								
Workload	vEPC, eMBB	vWEPC, eMBB	vEPC, eMBB	vEPC, eMBB	vEPC, eMBB	vUPF, eMBB	vFWA	vFWA
CPU	Intel® Xeon® Platinum 8180 Processor	Intel® Xeon® Gold 6132 Processor	Intel® Xeon® Gold 6130 Processor	Intel® Xeon® Gold 6230 Processor	Intel® Xeon® Gold 6230N Processor	Intel Xeon Gold 6230N Processor	Intel Xeon Gold 6230N Processor	Intel® Xeon® Platinum 8280 Processor
Frequency	2.5 GHz	2.6 GHz	2.1 GHz	2.1 GHz	2.3 GHz	2.3 GHz	2.3 GHz	2.7 GHz
Cores	28	14	16	20	20	20	20	28
Sockets	2	2	2	2	2	2	2, 1 under test	2
NICs	Intel® Ethernet Converged Network Adapters XXV710, 2x25G ports							
Memory	192 GB: 6x16 GB per CPU							
BIOS	SE5C620.86B.0D.01.0286.011120190816							

7.0 Glossary

Term	Description
CoSP	Communication Service Provider
CUPS	Control User Plane Separation
DPI	Deep Packet Inspection
EPC	Evolved Packet Core
FWA	Fixed Wireless Access
QoS	Quality of Service
NIC	Network Interface Controller
RAN	Radio Access Network
SP	Service Providers
SR-IOV	Single Root I/O Virtualization
TEM	Telecom Equipment Manufacturers
VM	Virtual Machine