WHITE PAPER

Communication Service Providers
Core Network

User Plane Innovations Enable Profitable 5G Service and Mobile Growth

Ericsson's User Plane function delivers up to 193 Gbps User Plane throughput, offering high performance at low cost per gigabit

Authors

Terence Nally
Wireless Core Segment Lead, Intel

Daniel Seiler
Solution Line Packet Core, Ericsson

Table of Contents

Executive Summary

5G mobile deployments have begun around the world. 5G is the next evolution of connectivity, creating a paradigm shift that will enable incredible growth in subscriber traffic and peak rates over the next five years. This greatly increases demand on user plane performance in packet core networks.

Ericsson and Intel have collaborated over the years to deliver innovative solutions for telecommunications. A new effort continues this momentum with Ericsson's optimized software solutions for their Dual-Mode 5G Cloud Core on Intel® Xeon® Scalable processor-based platforms and Intel® network controllers. As in the past, Ericsson and Intel are pushing the envelope by further enhancing one of the key elements of the network—the user plane (UP)—to support future growth by delivering up to 193 Gbps user plane throughput on one dual-socket Intel® processor-based server. This level of performance and flexibility provides a foundation for lower total cost of ownership (TCO) through lower cost per gigabit and cost per watt.

This white paper describes the user plane challenges that Communications Service Providers (CoSPs) will face with increasing growth and data demands and the Ericsson software and Intel hardware solution that addresses those challenges.

With 5G, Mobile Data Will Be Among the Biggest Challenges

5G deployments are just getting started in 2019. The new technology brings higher peak speeds with lower latency and greater bandwidth. These features will be needed, considering the expected growth in subscribers and increased traffic from a wide variety of new use cases—many yet undiscovered and unimagined but enabled by 5G. This growth will put considerable pressure on User Plane performance.

> "No previous generation of mobile technology has had the potential to drive economic growth to the extent that 5G promises."
>
> —2019 ERICSSON MOBILITY REPORT

Mobile Data Traffic to Grow 4x by 2024[1]

The Ericsson Mobility Report predicts mobile traffic will grow more than 400 percent over the next five years, from 30 exabytes/month to 131 exabytes/month by 2024. Most of this traffic will come from increased smartphone subscriptions—from 5.1 billion devices to 7.2 billion in 2024. Also in 2024, 25% of all smartphone subscriptions will be 5G; these will generate 35 percent of all mobile traffic. The remaining traffic will come from 4G/LTE (65% of all smartphone subscriptions), 3G, and 2G services.

Cellular Internet of Things (IoT) connections are predicted to grow with 27% CAGR from 2018 to 2024 to reach an additional 4.1 Billion connections using 4G, 5G, NB-IoT, Cat-M, and 2G/3G technologies.

Growth expectations of this scale are beyond what today's deployed hardware and software technologies can profitably support.

Focus on New Services

With 5G's architecture for superior communications capability, an important market shift from previous mobile technologies—mainly optimized to support consumers—will take place. In addition to growing mobile access for consumer voice, gaming, streaming, and emerging applications, such as augmented and virtual reality, 5G will support engagement at a new level of commercial and industrial applications, such as:

- Fixed wireless access (FWA)
- Enhanced mobile broadband (eMBB)
- Large-scale Internet of Things (IoT) used in industry and healthcare (such as massive Machine Type Communications, mMTC)
- Autonomous vehicles (V2X)
- Critical infrastructure (traffic control, smart cities, safety, etc.)

5G Core Architecture for Cloud-Native Operations

3GPP R15 introduces 5G core (5GC) architecture, which is a paradigm shift. 5GC is architected for flexibility and agility, considering NFV and software defined infrastructure (SDI), focusing on cloud-type deployment and orchestration of services. The new network core architecture is not backwards compatible and will support 5G standalone (SA) deployments.

5GC facilitates network slicing and more granular local breakout, which will enable many new industrial and enterprise use cases. While evolving the network to address these segments with new use cases, Communication Service Providers (CoSPs) need to continue to serve consumer and other high volume and throughput use cases.

> "…in the first five years, 5G subscription uptake is expected to be significantly faster than that of LTE, following its launch back in 2009."
>
> **—2019 ERICSSON MOBILITY REPORT**

Modernizing Mobile Networks

Current evolved packet core (EPC) network architecture was developed before the availability of widespread computing innovations that are now deployed in IT and cloud data centers. These technologies include virtualization and containerization, distributed computing, management/orchestration, and other technologies that have enabled flexibility, fast deployment of innovative new services, and reduced total cost of ownership (TCO).

EPC is introducing some of these capabilities in the first steps of Network Functions Virtualization (NFV), but earlier choices in the architecture prevent a full adoption of modern mechanisms. For example, certain aspects of networking, choice of networking protocols, and the use of point-to-point communication do not comply with modern technology choices such as HTTP and K8s networking.

5G NR standardizes a high-throughput design for advanced radio access network (RAN) with faster speeds, greater bandwidth, and spectrum sharing. EPC has been extended to support non-standalone (NSA) operations of 5G NR, allowing operators to reuse their existing core network assets to simultaneously serve 4G/LTE and 5G subscribers on all 4G/LTE and 5G NSA.

Ericsson's Dual-Mode 5G Cloud Core allows operators to serve subscribers on all those technologies including 5G SA in one software system running on one infrastructure. The flexible allocation of infrastructure to container-orchestrated microservices allows shifts of connections between different access technologies to happen without over-provisioning. One key product in Ericsson Dual-Mode 5G Cloud Core is the Packet Core Gateway (PCG) which is a universal User Plane Function (UPF).

NFV Infrastructure on Intel® Architecture

NFV infrastructure (NFVI) forms the foundation that supports communications compute capacity, scalability, high throughput with low latency, flexible deployment of services, and energy efficiency. Standards-based software and commercial-off-the-shelf, high-performance hardware will help reduce TCO and enable reusability of applications and equipment, while providing a base for telecom equipment manufacturers (TEMs) to differentiate their offerings to stay competitive in their markets.

Intel® Architecture (IA) has long been a trusted platform of choice for reliable, capable, scalable, and flexible compute, networking, and storage for communications infrastructure. Intel's partnership with the mobile communications industry has helped build a strong and broad ecosystem of solutions, and its latest generation of processors and networking silicon are ideal for enabling the future growth of core networks.

The UP is under pressure due to increased peak rates and significant traffic growth. Intel and Ericsson are continuously working together to improve economics for joint customers through the combination of Intel's powerful and versatile silicon in combination with Ericsson's efficient and highly optimized user plane software. This collaboration will help deliver innovative deployments of services and capabilities through a modernized user plane function catering to EPC, and 5G NSA and SA traffic.

Enabling an Efficient User Plane with Ericsson Dual-Mode 5G Cloud Core on Intel® Technologies

The data growth expected in mobile communications, especially with deployments of new 5G services, will place significant demand on the UP. Ericsson's UPF software running on Intel processors will enable profitable 5G service through a number of enhancements that lead to the following operating characteristics:

- User plane throughput of up to 193 Gbps on one dual-socket Intel® Xeon® Scalable processor-based server with direct I/O.
- User plane throughput of up to 26 Gbps per socket for Intel Xeon Scalable processor-based server with virtio.
- 5G peak rates with high background traffic load.

With a high-performance, more efficient UPF distributed in a scalable manner across an orchestrated NFVI, CoSPs can continue to profitably serve customers.

193 Gbps User Plane Throughput Per Dual Socket Server With Direct I/O

A Dell* dual-socket Intel Xeon Scalable processor-based server with Ericsson's UPF software using PF pass through I/O virtualization can deliver up to 193 Gbps at a CPU CapEx of under USD 49/gigabit. This level of performance on commercial-off-the-shelf hardware means high throughput at low power (0.64 Gbps/CPU Watt), reducing the associated power and cooling (OpEx) costs. The UPF software generates on average up to 2.2 Gbps of throughput per provisioned vCPU, this higher density will reduce the number of servers needed for peak capacity expansions.

All of Ericsson's UPF enhancements are part of the evolution of their virtualized evolved packet core (vEPC) software; they are included in their Dual-Mode 5G Cloud Core.

Test configuration and result details for achieving this level of throughput at low cost and low power are listed in the Appendix (Tables A1 and A2). Figure 1 illustrates the testing configuration. Figure 2 charts the measured results from testing.

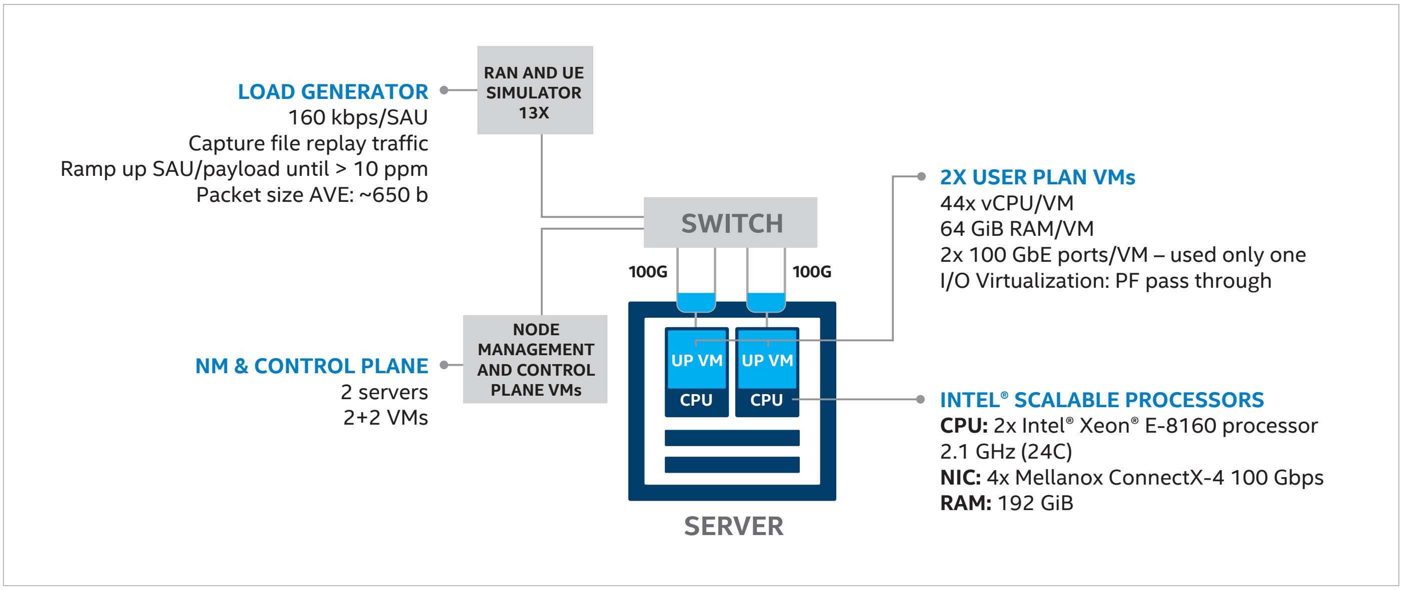

Figure 1. UP Testing Setup

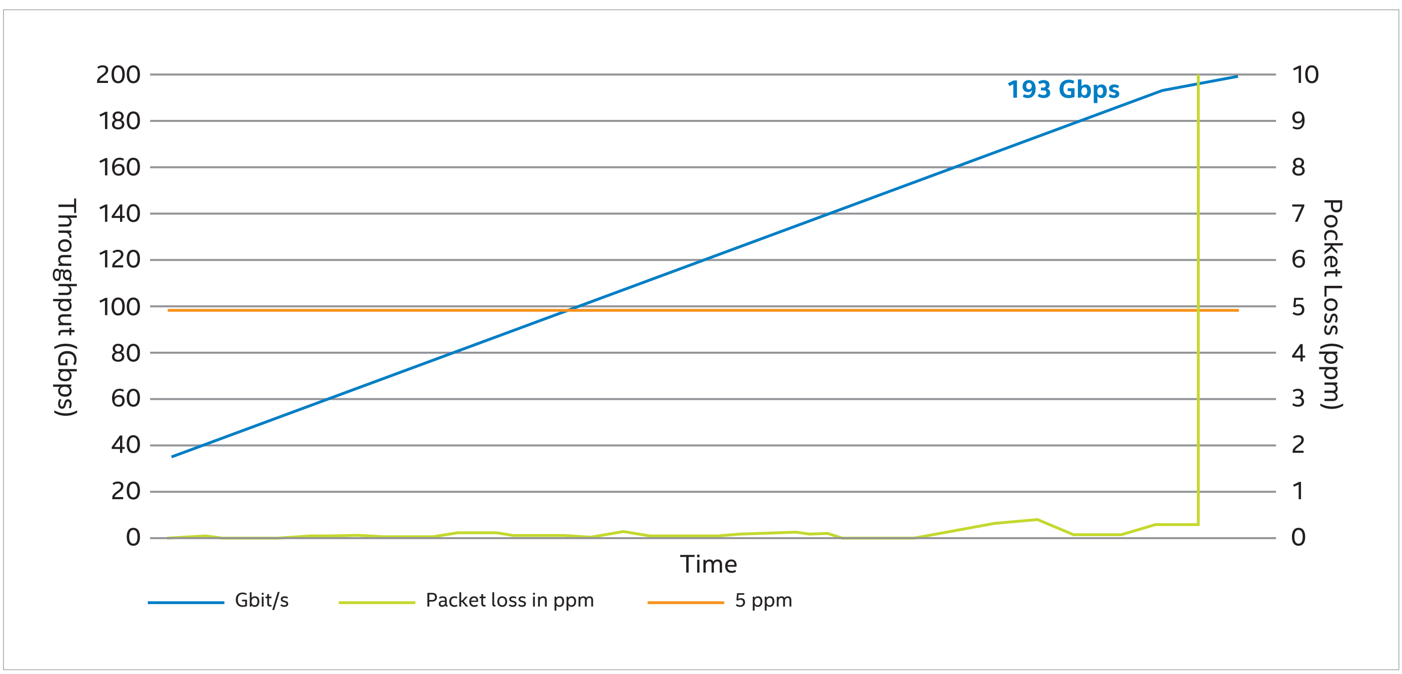

Figure 2. UP Testing Results

Virtio 26 Gbps per CPU socket

Ericsson re-architected networking and I/O handling in the user plane into Cloud-Native I/O (CNIO). One key feature of CNIO is to enable the Ericsson user plane function to leverage the multi-queue features of Open Virtual Switch (OVS).

Typical forwarding rates before using multi-queue on OVS are 5-10 Gbps total user plane throughput for a host. With Ericsson's user plane software a single Intel® Xeon® Scalable processor socket can forward up to 26 Gbps subscriber traffic, saturating one of the most deployed NICs in modern data clouds today—the 25 GE network adapter.

Test configuration and results details are listed in the Appendix (Tables A3 and A4). Figure 3 illustrates the configuration. Figures 4 and 5 chart the results. For testing, the latest Ericsson NFVi was utilized, which enables the use of multi-queue OVS.

The ability to perform well on an OVS is very important in NFV infrastructures that are built on OpenStack,* where OVS is the default networking mechanism. Many advanced OpenStack features, for example VM live migration, are relying on OVS.

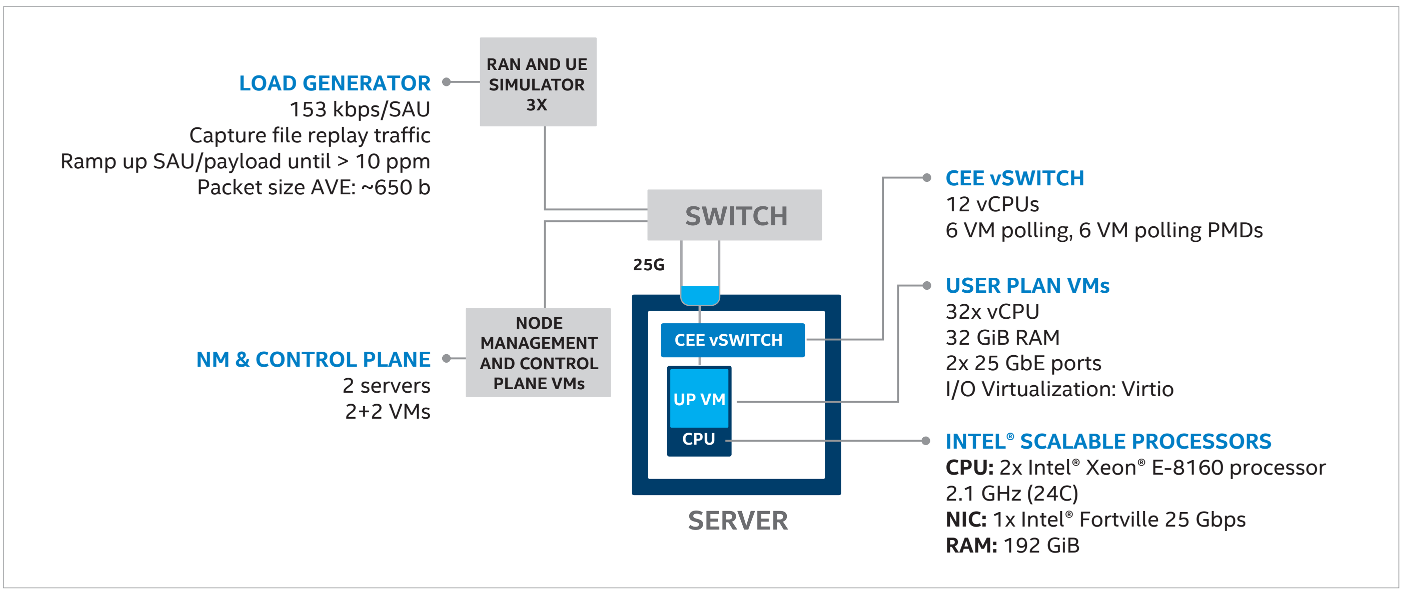

Figure 3. OVS Testing Configuration

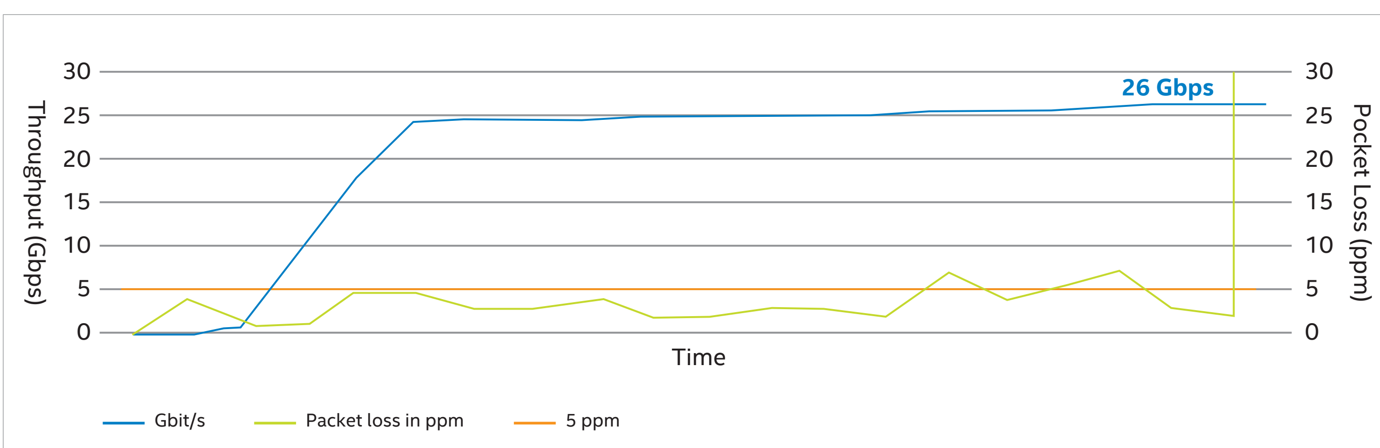

Figure 4. Virtio Throughput

Enabling Peak Speeds for 5G with Background Load

While increasing the aggregate capacity for a single server, Ericsson also enables the 5G experience for individual subscribers. 5G introduces new individual subscriber peak rates of 20 Gbps downlink and 10 Gbps uplink. Ericsson has enhanced the user plane function to dynamically deliver high subscriber peak rates, while continuing to handle a significant number of subscribers' base load in the system.

These enhancements allow for significantly better utilization of Intel processors, as less buffer overhead for peak speed is required.

In earlier user plane implementations, all traffic forwarding and service handling for individual subscribers were executed on one specific vCPU (Figure 6). But, one vCPU can process the traffic for many subscribers. Already in 4G, individual subscriber peak rates have increased through carrier aggregation. Thus, the likelihood of multiple subscribers in the same CPU requesting to satisfy peak demand with insufficient resources was growing.

With the newly standardized 20 Gbps downlink and 10 Gbps uplink maximum peak rates for 5G, it becomes imperative that user plane implementation evolves to address these peak rates and to allow efficient handling of subscribers across the pool of vCPUs.

Even with the earlier described configuration of running up to 193 Gbps across 88 vCPUs on two Intel® Xeon® processors it becomes obvious that a single vCPU alone will not be able to handle the maximum possible 5G peak rates of one individual subscriber.

To facilitate peak rates higher than the single vCPU maximum, Ericsson's UPF software distributes flows of individual subscribers across many vCPUs (Figure 7).

This facilitates a higher aggregate peak rate for individual subscribers, improving TCO in two distinct ways:

- Achieves higher subscriber peak rates by pooling a number of resources instead of requiring a more capable individual resource (like a higher frequency CPU).
- Allows operators to increase the load in the CPUs that are handling user plane processing. Distributing flows enables high individual subscriber peak rates without reserving headroom on each individual vCPU where a subscriber was anchored.

In one demonstration of this peak rate handling in Ericsson's user plane software executing on an older Intel CPU, a single peak rate subscriber achieved 34 Gbps throughput while 1.2 million subscribers were generating a background load of 30 Gbps across the system (see chart A5 in Appendix). This demonstration was executed on Ericsson Smart Services Routers (SSR).

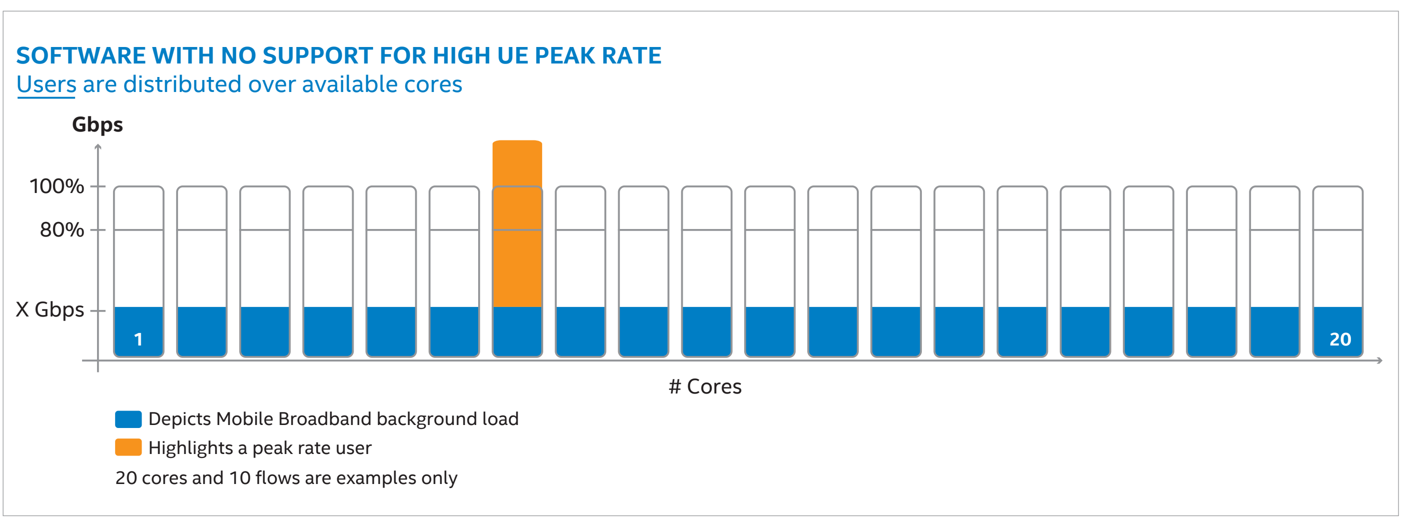

Figure 5. Inefficient Peak Rate Handling with Traditional UPF Implementation

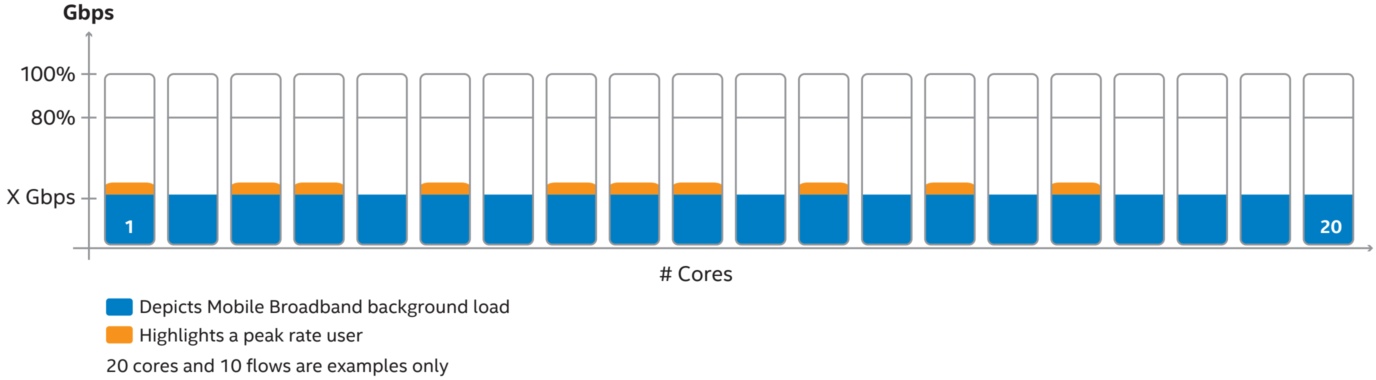

Figure 6. Ericsson's UPF Distributes User Traffic Across Multiple vCPUs for Better Resource Utilization and Improved User Experience

Summary

Cost effectively meeting user plane data demand will increase in importance for CoSPs over the next five years as 5G deployments grow along with the increase of 4G/LTE subscriptions and traffic. To support 5G while continuing to maintain and expand existing services, Ericsson has introduced their Dual-Mode 5G Cloud Core for mobile networks. Their new 5G core architecture includes re-architected user plane software that delivers up to 193 Gbps user plane throughput from a single dual-socket server with Intel® Xeon® Platinum 8160 processors. The combination of Ericsson's software on commercial-off-the-shelf, standards-based hardware platforms significantly improves the economics of running 5G user planes with lower cost-per-gigabit (CapEx) and cost-per-watt (OpEx) as illustrated in the testing in this paper.

The testing to achieve the above performance was completed on Intel Xeon Scalable processors, but Ericsson and Intel continue to optimize Ericsson's software for 2nd Generation Intel® Xeon® Scalable processors.

For more information about Ericsson's next-generation Dual-Mode 5G Cloud Core for cloud-native operations, visit **ericsson.com/5g-core**

For more information on 2nd Generation Intel Scalable Processors, visit **https://www.intel.com/content/www/us/en/products/processors/xeon/scalable.html**

Appendix

Testing Parameters/Results for Direct I/O

The following tables list the test configuration and results for the tests completed with Direct I/O.

Table A1.
PF Performance Test Configuration

SERVER CONFIGURATION	
Platform	Dual-socket, Dell* R740
Processor	2x Intel® Xeon® Platinum 8160 processor
Cores/Threads	24/48
Frequency	2.1 GHz
Memory	192 GB
Network Adapters	2x Mellanox* ConnectX-4 100 Gbps per socket; 4x total system
USER PLANE CONFIGURATION	
VMs	2 (1 User Plane VM per socket)
vCPUs/ User Plane VM	44
Memory/ User Plane VM	64 GB
Network Adapters/ User Plane VM	2x 100 GB Ethernet ports
I/O virtualization	PF Pass Through

Table A2.
PF Performance Test Parameters and Results

TESTING PARAMETERS	
Number of Sessions	1,200,000
Packet Loss (ppm)	<10
Throughput per Session (kbps)	160
TEST TRAFFIC INFORMATION (NO DPI)	
Average Packet Size (Bytes)	650
UL/DL Ratio	40%/60%
IP Protocol	UDP
Bearers per Session	1
Flows per Bearer	1
MEASURED PERFORMANCE	
Total Throughput (Gbps/Mpps)	193/36.67

Testing Parameters/Results for Virtio

The following tables list the test configuration and results for the tests completed with Virtio.

Table A3.
Virtio Test Configuration

SERVER CONFIGURATION	
Platform	Dual-socket, Dell* R740
Processor	2x Intel® Xeon® Platinum 8160 processor
Cores/Threads	24/48
Frequency	2.1 GHz
Memory	192 GB
Network Adapters	1X Intel® Ethernet Converged Network Adapter XXV710-DA2
vSWITCH CONFIGURATION	
vCPUs	12
Software	Ericsson Cloud Execution Environment* Release 6, vSwitch
UP CONFIGURATION	
VMs	1
vCPUs	32
Memory	32 GB
Network Adapters	2x 25 GB Ethernet ports
I/O virtualization	Virtio

Table A4.
Virtio Test Parameters and Results

TESTING PARAMETERS	
Number of Sessions (million)	169,800
Packet Loss (ppm)	<10
Throughput per Session (kbps)	153
TEST TRAFFIC INFORMATION (NO DPI)	
Average Packet Size (Bytes)	650
UL/DL Ratio	40%/60%
IP Protocol	UDP
Bearers per Session	1
Flows per Bearer	1
MEASURED PERFORMANCE	
Total Throughput (Gbps/Mpps)	26/5

Testing Parameters/Results for Peak Rate

The following table lists the test configuration and results for the tests on Ericsson SSR.

Table A5.
34 Gbps Individual Peak Test

BACKGROUND TEST TRAFFIC INFORMATION	
Aggregate Background Traffic (Gbps)	30
Average Packet Size (Byte)	650
UL/DL Ratio	40%/60%
IP Protocol	UDP
Sessions (millions)	1,200,000
Bearers per Session	1
Flows per Bearer	1
Throughput per Flow (kbps)	25
PEAK USER TEST TRAFFIC INFORMATION	
Aggregate Background Traffic (Gbps)	34
Average Packet Size (Byte)	650
UL/DL Ratio	40%/60%
IP Protocol	UDP
Sessions	1
Bearers per Session	1
Flows per Bearer	12
Throughput per Flow (Gbps)	~2.8

[1] https://www.ericsson.com/assets/local/mobility-report/documents/2019/ericsson-mobility-report-june-2019.pdf