

## ZTE's High Performance 5G Core Network UPF Implementation Based on 3rd Generation Intel® Xeon® Scalable Processors

Using the latest 3rd Gen Intel® Xeon® Scalable processors and Intel® Ethernet 800 Series Network Adapters, performance tests achieved 462Gbps.

### Authors

Intel  
Yong Li  
Jerry Zhang  
Guoliang Gu  
Richard Li  
Jianwei Ma  
Gauss Wu  
Daniel Ying  
ZTE  
Peng Cheng  
Fangliang Lou  
Jiangying Ma  
Yang Xu  
Zhou Xu  
Chenghu Yao  
Haoming Zhang  
Yong Zhang  
Jinchun Zou

### Table of Contents

1. ZTE's 5G UPF Solution .....	1
1.1. ZTE 5G Common Core Introduction ..	1
1.2. ZTE 5G UPF Introduction .....	2
2. 3rd Generation Intel® Xeon® Scalable Processor .....	2
3. Intel® Ethernet 800 Series Network Adaptors .....	3
4. System Test Environment .....	4
4.1. A Telecom Carrier's Traffic Test Model .....	4
4.2. Test Instruments (Ixia) .....	4
4.3. Hardware Configuration .....	4
4.4. Software Configuration .....	4
4.5. Network Topology .....	5
4.6. BIOS Configuration for a Compute Node .....	5
4.7. Virtual Machine Configuration for UPF .....	5
5. Performance Test Results .....	6
5.1. Overall Forwarding Performance ..	6
5.2. Average Forwarding Latency <sup>1</sup> .....	7
5.3. Analysis of the System's Total Cost of Ownership (TCO) .....	7
6. Conclusions .....	8

In a continuously evolving digital world, disruptive and emerging technology trends are increasingly impacting use cases and experiences, and this continues to fuel the need for flexible computing, networking and storage. The exponential growth in data generation and usage, the rapid expansion of cloud-scale computing and 5G networks, and the convergence of high-performance computing (HPC) and artificial intelligence (AI) are demanding that today's data centers and networks transform. These demands require that network infrastructure must evolve quickly and in a scalable way.

The Intel® Xeon® Scalable processor provides the agility and scale that operators require. 3rd Gen Intel® Xeon® Scalable processors are designed to support a diverse set of network environments and provide optimization for many workloads and performance levels with a wide range of cores, frequency, features and power. Enterprise, cloud and communications service providers can now accelerate their more powerful digital solutions with a feature-rich and highly flexible platform.

This paper describes ZTE's 5G core network UPF solution based on the latest 3rd Gen Intel Xeon Scalable processors and Intel Ethernet 800 Series Network Adapters with Dynamic Device Personalization (DDP) capabilities. Test results show that the maximum overall performance can reach 287 Gbps on the Intel® Xeon® Gold 6330N processor and 462 Gbps on the Intel® Xeon® Platinum 8380 processor under an operator's standard traffic test model with 690 byte packet length and without any additional hardware acceleration.

This test was completed by ZTE on March 30th, 2021. See the System Test Environment section for specific test configurations.

## 1. ZTE 5G Virtualized and Cloud Native UPF Solution

### 1.1. ZTE 5G Common Core Introduction

5G capabilities will help enable continuous wide area coverage, low power consumption, low latency, and high reliability connectivity. It will meet the diversified service requirements of eMBB, mMTC, uRLLC and other application scenarios.

5G makes it possible to deliver a full-service, multi-technology converged network, which will meet the rapid development needs of various services including a wide range of data and connectivity use cases. Using innovative ICT (Information and Communications Technologies) will help meet the user centric network requirements. The ZTE 5G Common core utilizes many of these innovations and introduces a new architectural design in four areas: service-oriented architecture, network slicing, control/forwarding separation, and stateless design.

#### Service-oriented architecture

Functionality is componentized, services/microservices are decoupled from each other, and each service/microservice can be upgraded and deployed independently to achieve rapid business innovation.

#### Network slicing

Automated network slicing according to the needs of different networks to meet diverse business requirements.

### Control/forwarding separation

Centralized deployment on the control plane and distributed deployment on the forwarding plane enable centralized control and nearby forwarding to avoid user-plane routing detours, shorten transmission paths, and offer flexible deployment for user-plane egress on demand.

### Stateless design

User data is stored independently, and application and data are separated.

The 5G Common Core created by ZTE not only brings an upgraded 5G service experience to individual consumers, but also fully meets the needs of industry customers in different scenarios, providing strong support for digital transformation across all industries.

For personal, home, and general industry applications, ZTE's fully converged 5G core network, known as Common Core, and converged edge computing, known as Common Edge, allow for reusable resources and greatly reduce network upgrade and construction investments. For general industry applications using this solution, slicing technology can be used to achieve on-demand networking and private use of the public network. Multi-access edge computing (MEC) all-in-one machines are used on the edge to differentiate deployment at the network edge and user edge.

For specific industry applications, especially scenarios such as high security and safety assurance and exclusive use of network resources, network slicing may require additional security measures. 3rd Generation Intel Xeon Scalable processors with Intel® Software Guard Extensions (Intel® SGX) provide significant protection against security threats and data protection—moving beyond encrypted data to encrypted computing.

Network operators adopting a 5G fully converged core network solution can readily add or subtract functions and components according to industry needs. Furthermore, a Common Core solution supports flexible on-demand customization, lowers docking complexity, reduces operation and maintenance difficulties, and accelerates network construction of industry self-service.

### 1.2. ZTE 5G UPF Introduction

ZTE's 5G converged subscriber-plane function includes the UPF function of the 5G core network and the subscriber-plane function of the GGSN SAE-GW in 2/3/4G networks. ZTE's 5G converged user-plane function is based on a fully virtualized cloud-native architecture and supports cross-data center deployment based on slicing requirements.

The cloud native, microservice architecture is highly automated from blueprint design and provides resource scheduling and lifecycle management, application status monitoring, control policy updates, etc. The effective connection between each link provides a closed-loop feedback mechanism enabling one-click deployment and installation, full autonomy and efficient management of services.

Compared with virtualization, container technology has the advantages of fast pop-up, lightweight and high performance, which is rapidly developing and widely used in the IT industry. Cloud-native applications and underlying virtualization technologies are decoupled and can be deployed in container technology to achieve improved resource utilization, as well as rapid delivery and agile maintenance of services. ZTE's 5G converged core network supports both virtual machine-based and container-based deployments with the following benefits:

- Optimization based on virtual machines and containers meets the requirements of 5G core network deployment resources.
- Performance optimization includes open-source virtual switches with open architecture and interfaces.
- Optimization means forwarding performance improvements meet carrier-class network performance requirements.

## 2. 3rd Generation Intel® Xeon® Scalable Processor

3rd Generation Intel Xeon Scalable processors, based on Intel's 10nm process, are designed for the modern data center to deliver performance, total cost of ownership (TCO) improvements and user productivity by driving operational efficiency.

Systems built on 3rd Gen Intel Xeon Scalable processors deliver agile services with enhanced performance and breakthrough capabilities. The latest Intel Xeon Scalable processors support 8-40 powerful processor cores of varied frequencies to balance increased throughput and energy efficiency. Features include 8 channels of DDR4-3200 MT/s DIMMs, up to six Intel® Ultra Path Interconnect (Intel® UPI) channels for increased platform scalability, and improved cross-socket bandwidth for I/O-intensive workloads. It delivers an average of 62% performance gains over previous generation platforms across a range of widely deployed network and 5G workloads. Optimized for mainstream data center, multi-cloud, networking, and storage applications, 3rd Gen Intel Xeon Scalable processors support a wide range of XaaS environments.

3rd Gen Intel Xeon Scalable processors are built specifically for the flexibility to run complex AI workloads on the same hardware as existing workloads. Built-in AI performance with enhanced Intel® Deep Learning Boost, now incorporates the industry's first support of Brain Floating Point 16-bit (bfloat16) numeric format and Vector Neural Network Instructions (VNNI), which improves AI inference and training performance. Compared to the previous generation, AI training improves by 1.93X and image classification by 1.87x. 3rd Gen Intel Xeon Scalable processors help deliver AI readiness across the data center to the edge and back.

3rd Generation Intel Xeon Scalable processors have enhanced hardware security features that help stop malicious attacks while maintaining workload integrity. Intel® Software Guard Extensions (Intel® SGX) help protect data and application code from the edge to the data center and multi-tenant usage, allowing enhanced collaboration

without compromising privacy. Intel® Platform Firmware Resilience (Intel® PFR) is an Intel® FPGA-based solution that protects platform firmware, detecting corruption and restoring firmware to a normal state. In addition, built-in crypto acceleration reduces the performance impact of pervasive encryption.

Intel® Speed Select Technology (Intel® SST), supported by the 3rd Gen Intel Xeon Scalable processor allows for more and finer control over processor performance to optimize total cost of ownership (TCO). Intel® SST-CP technology can maintain a higher base frequency on a subset of processor cores and a lower base frequency on the remaining processor cores. Intel® SST-TF can maintain a higher RWD frequency on a subset of processor cores and a lower RWD frequency on the rest of the processor cores.

### 3. Intel® Ethernet 800 Series Network Adapters

The latest generation of Intel® Ethernet Network Adapters, Intel® Ethernet 800 Series, optimizes high-performance server workloads by innovatively improving application efficiency and network performance.

The Intel® Ethernet 800 Series Network Adapters support PCIe Gen4.0 and 3.0x16 host interfaces, as well as throughputs up to 100Gb/s. Multiple rates of 100/50/25/10/1GbE/100M are supported on a single port and can be flexibly configured and modified directly through the EPCT tool. The combination of more available ports and speeds simplifies validation and deployment to better meet the workload demands required by users.

Offering rich Dynamic Device Personalization (DDP) capabilities, Intel Ethernet 800 Series Network Adapters feature an enhanced DDP profile, loaded during driver initialization, as well as many protocols for specific workloads, resulting in greater flexibility. The default generic DDP profile supports common protocols, including TCP/UDP/IP/VLAN/MAC/ETYPE/SCTP/ICMP, and GRE/GENEVE/VXLAN/ARP/MPLS/NVGRE/LLDP. The Comms DDP profile, designed specifically for the telecom field, not only supports the protocols contained in the generic DDP, but also adds support for GTP/PPPoE/IPSEC/L2TPv3/PFCP/MPLS and other protocols. As new network services evolve and new protocols emerge, changes can

easily be updated by upgrading the DDP profile. This level of flexibility reduces latency, improves efficiency of message processing, and reduces processor load.

Intel Ethernet 800 Series Network Adapters support increased throughput and reduced latency through Remote Direct Memory Access (RDMA). RDMA provides high throughput and low latency performance for modern high-speed ethernet by eliminating network resource overhead, including TCP/IP stack processes, memory copies, and application context switching. Intel Ethernet 800 Series Network Adapters support all major storage transport protocols, including iWARP, RoCE v2 and NVMe over TCP. Customers can choose flexibly according to their needs or combine applications for easy networking and deployment.

Intel Ethernet 800 Series Network Adapters introduce the new Application Device Queues (ADQ) feature, which uses an optimized application thread-to-device data path to enable application-specific data control, transmission, and rate limiting. This dedicated queuing and the ability to adjust network traffic not only improves application performance, but also reduces latency and increases throughput. Resilient and predictable application-level performance becomes a key challenge as the modern data center scales horizontally. ADQ technology dramatically reduces performance jitter by creating dedicated queues for critical loads, significantly improving application scalability and predictability.

Support for both IEEE 1588 PTP v1 and v2 provide nanosecond-level time accuracy to precisely report the time each packet is received. This level of time accuracy helps ensure reliable synchronization for network deployments in areas ranging from 5G RAN to financial services, industrial automation, and energy monitoring.

The Intel Ethernet 800 Series family of network adapters provides resilient protection for platforms through three security mechanisms: protection, detection and recovery, and the Hardware Root of Trust. Built-in fault detection protects firmware and critical device settings and performs automatic device recovery to ensure devices return to their original programming state.

## 4. System Test Environment

### 4.1. A Telecom Carrier's Traffic Test Model

Number of users <sup>1</sup>	Access User: 600000 Data User: 6000
Content billing configuration rules (DPI)	L7 Networking: 40000 L3 Networking: 10000
PCC strategy	Static: 45; Dynamic: 5
Volume ratio	HTTP: 85% UDP: 15%
Average packet length <sup>2</sup>	690 byte

### 4.2. Test Instruments (Ixia)

The Ixia IxNetworks-XGS2 was used in the test, conducted by ZTE on March 30, 2021, to simulate the control and data planes when 5G mobile users access them through the telecom carrier's base stations.

### 4.3. Hardware Configuration

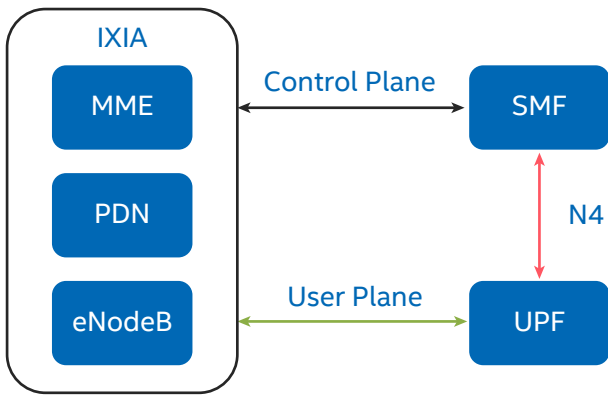


Figure 1. ZTE UPF test instruments

The servers used were ZTE's self-developed 5300-G4X, powered by 2-socket 3rd Gen Intel Xeon Scalable processors with each socket being connected to two or three Intel® Ethernet Network Adapters E810-CQDA2.

Server	ZTE 5300G4X
Ucode	0d000270
CPU	Intel® Xeon® Gold 6330N Intel® Xeon® Platinum 8380
Number of CPU	2
Memory	512G DDR4@2666MHZ
Network Adapter	4x Dual-port 100Gb Intel® Ethernet Network Adapter E810-CQDA2 on 6330N server 6x Dual-port 100Gb Intel® Ethernet Network Adapter
TOR	ZTE 5960-4M

### 4.4. Software Configuration

OS	ZTE CGSL 4.18.0-147.8.1.el8_1.x86_64
OpenStack Platform	ZTE TECS 7.2 (Openstack train)
QEMU	QEMU 4.1.0
UPF VNF	ZXUN-xGW(GUL)V7.20.41
i40e Driver	Driver Version: 1.3.2 Firmware-version: 1.40 0x80003ab8 1.2735.0 DDP Profile:ice_comms-1.3.22.0
DPDK	20.11

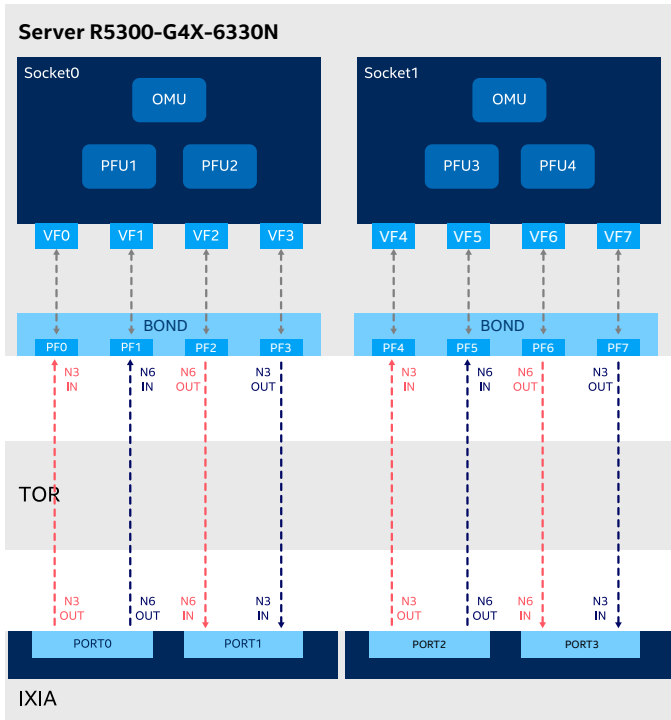


Figure 2. Test topology for the Intel® Xeon® Gold 6330N processor-based server

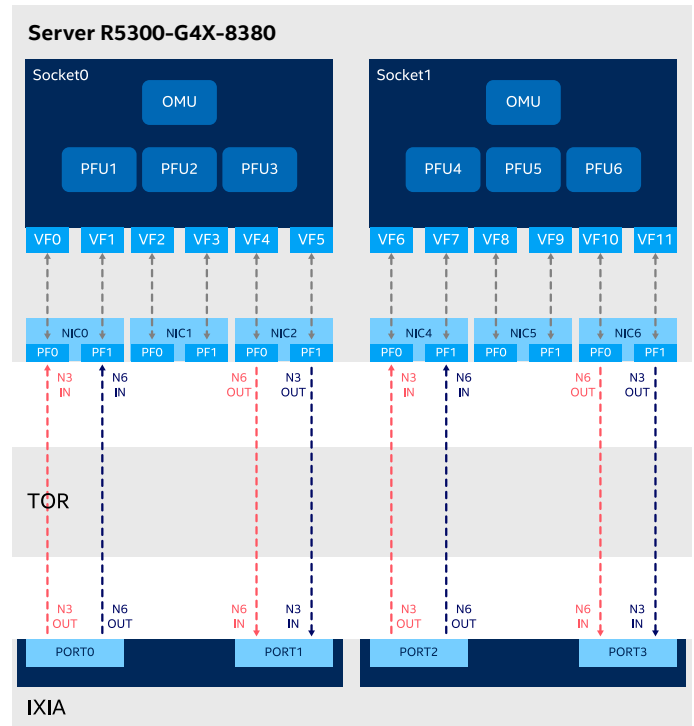


Figure 3. Test topology for the Intel® Xeon® Platinum 8380 processor-based server

4.5. Network Topology

The cloud management platform used in the test was ZTE TECS OpenStack. Two Intel processor-based servers were used as the control node and the compute node respectively. The UPF was deployed on the compute node while the UPF network element was connected to the Ixia tester with a 5960-4M switch. The test topologies for the Intel Xeon Gold 6330N processor-based server and the Intel Xeon Platinum 8380HL processor-based server are shown in Figure 2 and Figure 3.

4.6. BIOS Configuration for a Compute Node

The BIOS configuration used for the server as a compute node is shown in the following table.

4.7. Virtual Machine Configuration for UPF

One OMU virtual machine (VM) and two PFU VMs were deployed on each socket of the Intel® Xeon® Gold 6330N processor-based server. The OMU occupied 4 vCPUs and each PFU occupied 16 vCPUs (12 vCPUs ran worker forwarding threads). Intel SST-CP technology was used to set the core frequency of worker threads to 2.8 GHz and the remaining core frequencies to 1.5 GHz.

	OS		OMU		PFU1										PFU2													
6330N	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
	28	29	30	31	32	33	34	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	55

Figure 4. ZTE UPF virtual machine deployment on Intel® Xeon® Gold 6330N

One OMU VM and three PFU VMs were deployed on each socket of the Intel® Xeon® Platinum 8380 processor-based server. The OMU occupied 4 vCPUs and each PFU occupied 16 vCPUs (14 vCPUs ran worker forwarding threads).

	OS		OMU		PFU1										PFU2										PFU3															
8380N	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
	28	29	30	31	32	33	34	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	55	56	57	58	59	60	61	62	63	64	65	66	67

Figure 5. ZTE UPF virtual machine deployment on Intel® Xeon® Platium 8380

MENU	PATH TO BIOS SETTING	BIOS SETTING	REQUIRED SETTINGS
CPU CONFIGURATION	ADVANCED -> PROCESSOR CONFIGURATION	INTEL® HYPER THREADING TECHNOLOGY	ENABLED
		INTEL® VIRTUALIZATION TECHNOLOGY	ENABLED
POWER CONFIGURATION	ADVANCED -> POWER & PERFORMANCE	CPU POWER & PERFORMANCE POLCY	PERFORMANCE
	ADVANCED -> POWER & PERFORMANCE -> CPU P STATE CONTROL	ENHANCED INTEL SPEEDSTEP TECH	ENABLED
		INTEL® TURBO BOOST TECHNOLOGY	ENABLED
	ADVANCED -> POWER & PERFORMANCE -> HARDWARE P STATES	HARDWARE P-STATES	DISABLED
	ADVANCED -> POWER & PERFORMANCE -> CPU C STATE CONTROL	PACKAGE C-STATE	C0/C1 STATE
		CIE	DISABLED
PROCESSOR C6		DISABLED	
IO CONFIGURATION	ADVANCED -> INTEGRATED IO CONFIGURATION	INTEL VT FOR DIRECTED I/O	ENABLED

## 5. Performance Test Results

### 5.1. Overall Forwarding Performance

The bladed (dual processor) performance tests verified the basic forwarding capabilities of the UPF solution and the forwarding performance of the UPF solution with business processing capabilities, including billing and deep packet inspection (DPI).

#### Basic forwarding performance of the UPF solution based on dual Intel Xeon Gold 6330N processors:

- When offline billing and DPI were disabled, the forwarding performance reached 287 Gbps (51.3 MPPS). Worker forwarding threads' average core utilization was 84%.
- When offline billing and DPI were enabled, the forwarding performance reached 177 Gbps (31.6 MPPS). Worker forwarding threads' average core utilization was 85%.

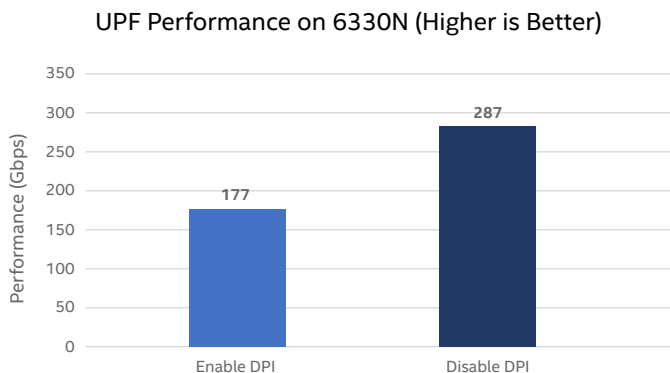


Figure 6. ZTE UPF performance result on Intel® Xeon® Gold 6330N processor

#### Basic overall forwarding performance of the UPF solution based on dual Intel® Xeon® Platinum 8380 processors:

- When offline billing and DPI were disabled, the forwarding performance reached 462 Gbps (81.4 MPPS). Worker forwarding threads' average core utilization was 83%.
- When offline billing and DPI were enabled, the forwarding performance reached 280 Gbps (49.4 MPPS). Worker forwarding threads' average core utilization was 84%.

Compared with Intel Xeon Gold 6230N and Intel Platinum 8280 processor, the forwarding performance of the UPF solutions are largely improved in the test.

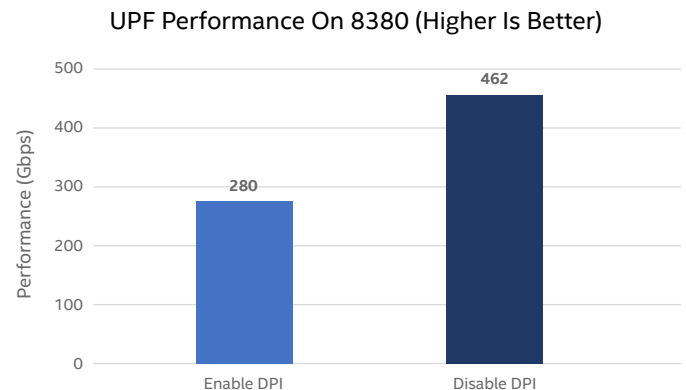


Figure 7. ZTE UPF performance result on Intel® Xeon® Platinum 8380 processor



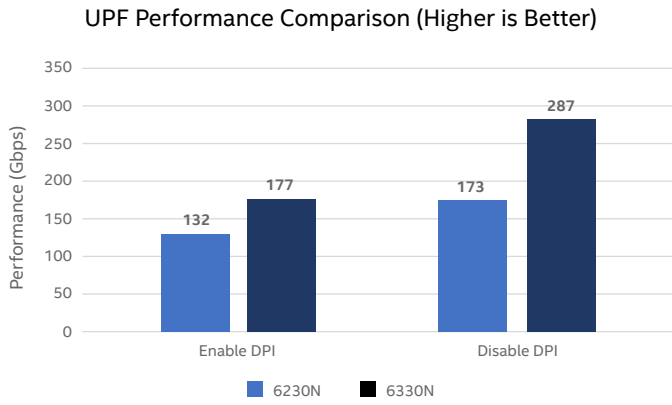


Figure 8. ZTE UPF performance comparison between 6230N and 6330N

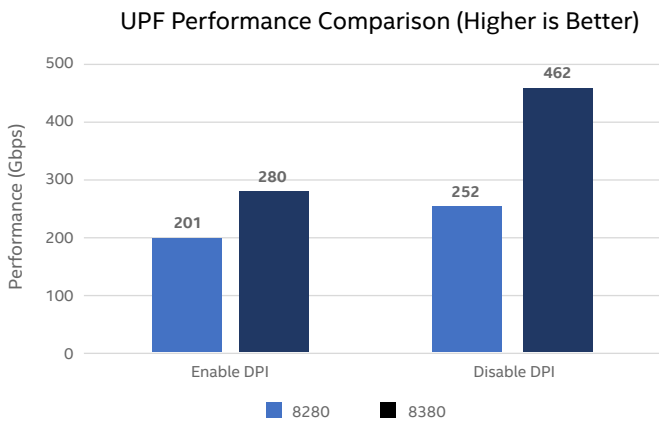


Figure 9. ZTE UPF performance comparison between 8280 and 8380

5.2. Average Forwarding Latency

Dynamic Device Personalization (DDP) acceleration technology does not require software to distribute messages among cores, greatly reducing the forwarding latency. The test results show that the one-way average forwarding latency of UPF messages was reduced from 150 μs to 74 μs.

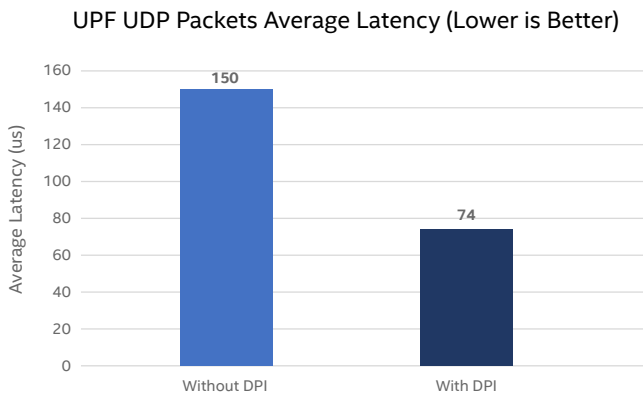


Figure 10. ZTE UPF UDP packets average latency

5.3. Analysis of the System's Total Cost of Ownership (TCO)

Intel SST and DDP technologies bring more than just performance improvement and latency reduction to UPF. Figure 11 estimates the device costs and ten-year electricity expenditures for three different solutions with different performance capabilities. Given the same processing requirements, the UPF solutions based on Intel Xeon Gold 6330N processors and Intel Xeon Platinum 8380HL processors, both of which are equipped with Intel SST and DDP technologies, deliver substantially lower total cost of ownership. The UPF solution based on Intel Xeon Gold 6138 processor, is not equipped with Intel SST and DDP technologies, resulting in a higher overall total cost of ownership. Over time, ZTE found that the TCO of Intel-based UPF solutions becomes even more advantageous.<sup>3</sup>

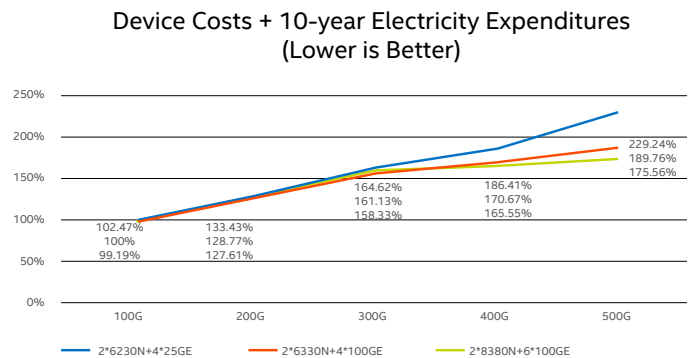


Figure 11. Device costs and ten years' electricity expenditures

Another TCO consideration for a telecom carrier's UPF element is the number of servers required for their performance requirements. The number of servers (excluding N+1 redundant servers and control plane servers) can be reduced if the total system performance is improved, as seen in UPF solutions based on Intel SST and DDP technologies.

Server Performance (Gbps)	UPF element reqs (Gbps)		
	100	200	300
132 (Intel® Xeon® Gold 6230N CPU)	2 servers	3 servers	4 servers
177 (Intel® Xeon® Gold 6330N CPU)	2 servers	2 servers	3 servers
280 (Intel® Xeon® Platinum 8380 CPU)	1 servers	2 servers	2 servers

## 6. Conclusions

ZTE's 5G Core Network UPF solution based on the latest 3rd Gen Intel Xeon Scalable processors and Intel Ethernet 800 Series has achieved over 400 Gbps forwarding capability in the telecom carriers' real-world traffic test models. The solution delivers the outstanding performance that telecom carriers are seeking for their 5G UPF deployments. In addition to performance improvement, the solution greatly reduces the forwarding latency and fulfills the end-end low-latency requirements needed for many 5G applications.<sup>4</sup> Moreover, the solution is also highly advantageous in terms of the system's total cost of ownership.

This testing also demonstrated that without any additional hardware acceleration devices, it is possible to achieve extraordinary performance with an Intel technology-based platform by using the latest general-purpose Intel processors, low-power Intel Ethernet 800 Series network adapters, and related technologies such as the Data Plane Development Kit (DPDK). By using these technologies together, the processing capabilities of VNFs can be improved, which helps the telecom carriers' and equipment manufacturers' in network function virtualization scenarios.

## 7. Abbreviations

eNB	eNodeB
FDIR	Fault Detection, Isolation, and Recovery
GPRS	General Packet Radio Service
GTP	GPRC Tunneling Protocol
IMIX	Internet Mix
IP	Internet Protocol
L2TP	Layer Two Tunneling Protocol
OEMs	Original Equipment Manufacturers
OMU	Operations Manager (Unix)
PDN	Packet Data Network
PFU	Packet Forwarding Unit
PPPOE	Point-to-Point Protocol Over Ethernet
RSS	Receive Side Scaling
SR-IOV	Single Root I/O Virtualization
TCP	Transmission Control Protocol
UE	User Equipment



<sup>1</sup> The Access Users number is the maximum number of users supported by the UPF system. The Data Users number is the actual number of users for the test.

<sup>2</sup> IMIX traffic.

<sup>3</sup> Based on ZTE internal analysis and data.

<sup>4</sup> See 3GPP 23.501 chapter 5.7.4

### Notices & Disclaimers

Performance varies by use, configuration and other factors. Learn more at [www.Intel.com/PerformanceIndex](http://www.Intel.com/PerformanceIndex).

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a nonexclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries.

\* Other names and brands may be claimed as the property of others.

© 2021 Intel Corporation Printed in USA

346628-001US Please Recycle

0421/DL/ICKK/PDF