White Paper

5G Core
5G Performance

Samsung Exceeds 1 Tbps in 5G Core User Plane Performance with 4th Gen Intel® Xeon® Scalable Processors

Authors

Samsung
Ilgee Kang, Kwangseop Hwang, Sungyoon Ryu, Myungwon Bang, Giljung Kim, Namgyun Kim, Gyuil Choi, Wonsuk Song, Vijay Ram Inavolu , Ramkumar Balu

Intel
Steve Dinkins, Jianwei Ma, Jass Jang Henry Jeong, Chetan Hiremath

Table of Contents

Executive Summary

Growth in network traffic is non-stop, and the mix of data and services will only get more complex.

The industry response is to adopt cloud native architectures to support the migration to a standalone 5G core (5GC) network. The adoption of cloud native IT infrastructure in the 5GC supports fully automated network management and promises significant improvements in network agility, responsiveness and speed of innovation with much lower capital and operating expenses over time. Samsung 5G Core builds on the cloud native architecture to provide highly stable solutions that deliver the high capacity and reliability required for network applications.

While sustainability initiatives and complexity from cloud native deployments require operators to emphasize metrics focused on energy efficiency, control plane latency and security, user plane throughput remains the top key performance indicator. Samsung and Intel have collaborated on performance and latency optimizations of Samsung 5G Core, especially with user plane function (UPF) workloads.

This paper describes a recent test of UPF throughput using the latest Samsung 5G Core software running on Intel reference servers with 4th Gen Intel® Xeon® Scalable processors and Intel® Ethernet Network Adapters. Throughput more than doubled when compared to the previous generation of Intel Xeon Scalable processor, reaching 1 Tbps of throughput in the 5GC user plane.

Samsung is fully capable of meeting the exponentially increasing traffic demands from diverse subscribers and industries and is ready to provide its solutions for global network operators to evolve their 5G networks.

1. Introduction

To date, many network operators have chosen to deploy non-standalone 5G core networks, which support legacy 4G/LTE services and the delivery of enhanced mobile broadband (eMBB).

The move to 5G standalone core networks will support more advanced services, such as network slicing, mobile private networks, ultra-reliable low latency communications, Voice of New Radio, machine-to-machine communications, vehicle-to-everything and more. These new services represent significant monetization opportunities.

In its recent paper titled, "5G Core: The Key to Monetizing 5G Standalone Networks," Dell'Oro Group noted that China Unicom reported over $500 million in contract value in the first half of 2022 from industrial applications running on 5G standalone mobile private networks.

The wave of data traffic resulting from this expanded array of services will also require new levels of throughput performance in the cloud native 5G standalone core.

The UPF plays a critical role in 5G network performance because it is responsible for traffic forwarding and policy enforcement. We can increase performance by optimizing user plane pipelines, packet processing pipelines and software architectures, making it a key enabler for advanced 5G use cases and services.

2. Samsung 5G Cloud Native Core Overview

2.1. Introduction to Samsung 5G Cloud Native Core

Samsung's 5G cloud native core allows network operators to launch new services quickly and upgrade frequently according to their business needs while reducing operating expenses by providing higher operational efficiency.

Samsung 5G Core solutions leverage a micro-services architecture, end-to-end dynamic orchestration and automation, continuous integration and continuous delivery (CI/CD), open source platform services, telco-grade performance support and telco-oriented open sources to deliver an end-to-end solution.

Samsung 5G Core is also designed with critical features to ensure the stability and reliability of the network such as an overload control feature to counteract sudden traffic spikes as well as geo-redundancy support.

Samsung 5G Core automates end-to-end network operation by integrating RAN, core, and MEC centering on its automation platform to respond to any changes in 5G network such as operation, upgrades and monitoring to boost operational efficiency. It simplifies and automates the operation of Samsung 5G Network by minimizing operator intervention through its closed-loop control that constantly repeats data collection from the 5G network, followed by analysis, optimization, control and monitoring.

Telecom services, unlike IT services, demand strict performance and reliability characteristics. Based on its cloud native architecture, Samsung 5G Core uses open source technologies to provide highly stable solutions that deliver the high reliability required for network operator applications.

2.2. The Basic Principle of Samsung 5G Core Cloud Native Network Function (CNF)

Samsung core network functions (NFs) are deployed either as VNFs (virtualized network functions) or CNFs (containerized network functions). The 5G core CNF delivers the advantages of cloud native applications by using cloud native design principles and are deployable on any container platform that is aligned with Cloud Native Computing Foundation (CNCF) principles, such as containers and microservices and widely used cloud tools such as Kubernetes, Istio/Envoy, Prometheus and others.

- **Stateless & Microservices –** The 5G core CNF uses stateless and loosely coupled software for micro-services and Kubernetes container environments to provide efficient support for failover handling, scaling in/out and seamless upgrades. This approach also improves reusability by minimizing dependencies between microservices.
- **Open API –** The 5G core CNF uses REST-based, open APIs that are found in most web applications today. As a result, CNFs can access extensive cloud native capabilities and tools to enable more convenient monitoring, debugging and tracing, which in turn reduces development time. In addition, third parties can use open APIs to integrate their applications more easily into Samsung 5G Core CNF.
- **Platform-Agnostic –** The 5G core CNF remains agnostic to the tools provided by the platform service. CNFs frequently use well-known tools, such as Prometheus and EFK, but network operators may require different tools to support their 5G core CNFs. For example, Istio/Envoy is a commonly used service mesh tool, but an operator's environment may require a different service mesh tool.
- **Resiliency –** The 5G core CNF should always be reliable and available whether the base infrastructures are stable or not. In cloud native environments, infrastructures are not always reliable, and all applications should adapt to these environments. To help with the availability, the careful design of the 5G core CNF minimizes the size of its components to facilitate quick and easy recovery.

2.3. Introduction to Samsung 5G User Plane Network Functions

Samsung is developing, maintaining and evolving its 5G-UPF based on the Samsung 5G Core container principles listed above to deliver the following features:

- **DPDK/VPP –** The DPDK and VPP frameworks support high-performance packet forwarding in the 5G data plane. VPP is commonly used for processing multiple packets at a time and provides the improved performance and the low latency to meet the high and stable transfer quality requirements in 5G networks. The Data Plane Development Kit (DPDK) is a set of user-space libraries and drivers that accelerate packet-processing workloads running on all major CPU architectures. Samsung 5G-UPF uses DPDK to make packet processing faster by directly accessing the network interface card without the overhead of a kernel.
- **High-Performance and Diverse Interface –** Samsung 5G-UPF uses packet acceleration technologies, such as SR-IOV, for telco-grade I/O performance. This integration supports direct access to the network interface card to improve packet transfer performance and enables the use of Multus container network interface (CNI) to add additional network interfaces to pods.
- **Dynamic Device Personalization (DDP) –** Samsung 5G-UPF requires multiple dedicated cores (packet distribution/packet distribution) to steer the same flow packets to fixed packet processing cores and transmit them to the network adapter in order. A packet distribution core parses inner/outer IP headers of all incoming uplink (GTP-U)/downlink packets, assigns a unique hash value from five tuples of each IP header and delivers the packets to the dedicated cores. The distribution/transmission load requires more processing to improve performance, which makes it more challenging to avoid forwarding bottlenecks and maintain target latency at a high load. When using the

DDP capabilities of the Intel Ethernet Network Adapter E810, the NIC parses more in-depth layer protocol to get inner IP header fields of incoming GTP-U packet for RSS hash calculation. As a result, the UPF can offload packet distribution/transmission workloads to the network adapter and reuse packet distribution cores as packet processing cores, boosting UPF performance.

- **CG-NAT –** Samsung 5G-UPF supports rapid deployment because of fully integrated highly-optimized CG-NAT features that are processed in a single container, eliminating any additional investment to support east-west traffic.

3. Key Intel Technologies for UPF Optimization

The UPF throughput test described in this paper used Samsung 5G Core running on Intel reference servers with dual-socket, 60-core Intel Xeon Platinum 8490H processor and Intel Ethernet Network Adapters E810-2CQDA2 ports.

4th Generation Intel Xeon Scalable processors deliver industry-leading per-core performance, energy efficiency and the widest array of built-in accelerators, giving network operators a fast path to deploy new services without requiring expensive external accelerators to build custom systems. For 5G core workloads, built-in accelerators help increase throughput and decrease latency, while advances in power management enhance both the responsiveness and the efficiency of the platform.

Intel Xeon Platinum 8400 processors feature a new architecture with higher per-core and multi-socket performance than the previous generation with up to 60 cores per socket and up to eight sockets per system. The processors also deliver improvements in input/output, memory, storage and network technologies to balance these gen-to-gen core count increases. Because the 5GC UPF is a very intensive I/O bandwidth application that requires careful optimization of memory access and cache system, the following platform advances are important for these 5G core workloads:

- DDR5 memory provides up to 1.5x the bandwidth and speed of DDR4, for 4,800 megatransfers per second (MT/s).
- I/O improvements with 80 lanes of PCIe Gen5 per socket compared to 64 lanes of PCIe 4.0 per socket in the prior generation, maintains backward compatibility and provides foundational slots for CXL.
- CXL 1.1 supports high fabric bandwidth and attached accelerator efficiency.
- Increase shared last-level cache (LLC) (up to 100 MB LLC shared across all cores).

Intel® Advanced Vector Extensions 512 (Intel® AVX-512) supports up to two fused multiply-add (FMA) units and includes optimizations to accelerate performance for demanding computational tasks. Intel AVX for vRAN increases density up to 2x compared to the previous generation, with the same power envelope.[1] The AVX-512 instructions accelerate performance for 5GC UPF workloads by delivering improvements to the VPP infrastructure library. This enables execution of one instruction on multiple data sets simultaneously which is beneficial in Tx and Rx traffic operations.

Intel® QuickAssist Technology (Intel® QAT) accelerates encryption, decryption and data compression, offloading these tasks from the processor core to help reduce system resource consumption.

Intel® Crypto Acceleration reduces the penalty of implementing pervasive data encryption and increases the performance of encryption-sensitive workloads, such as for Secure Sockets Layer (SSL) web servers, 5G infrastructure and VPNs/firewalls.

Intel® Dynamic Load Balancer (Intel® DLB) provides efficient hardware-based load balancing by dynamically distributing network data across multiple CPU cores as the system load varies.

3.1. IA Platform Optimization

Intel Ethernet 800 Series is the next generation of Intel Ethernet Controllers and Network Adapters. The Intel Ethernet Network Adapter E810 is designed with an enhanced programmable pipeline, allowing deeper and more diverse protocol header processing. This on-chip capability is called Dynamic Device Personalization (DDP). This allows the consistent parsing and steering of traffic from a given user equipment (UE) or UE flow to a worker core.

A network controller packet pipeline is responsible for packet identification and reporting protocol information on the packet's receive (Rx) descriptor. This information is used by the filters and queue management on the controller and by upper layers of software.

To optimize UPF performance, it is important for the NIC to understand the protocols and tunnels that are received to allow for filtering and various stateless offloads to assist in the packet processing.

The Enhanced DDP Package for Telecommunications used for UPF supports GPRS Tunnelling Protocol (GTP). Metadata can be extracted from the GTP headers, then used in the subsequent steps of the packet processing engine of the network adapter including the switch, Receive Side Scaling (RSS) and Intel Ethernet Flow Director (Intel Ethernet FD)

4. System Test Environment

4.1. IA Platform Optimization

For this performance test, the Samsung's 5G-UPF data plane ran on Intel reference servers with Intel Xeon Platinum 8490H processors and Intel Ethernet Network Adapters E810-2CQDA2 using an Intel best known configuration.

The test utilized multiple UPF instances to support faster and easier application scaling, avoiding potential bottlenecks inherent in a single monolithic instance where multiple cores compete to access VFs.

To simulate the complete 5G network architecture, the team deployed multiple TeraVMs to emulate applications and simulate the whole 5G network architecture, including UE, gNodeB and DN servers and Samsung's AMF and SMF for control plane. See Figure 1.

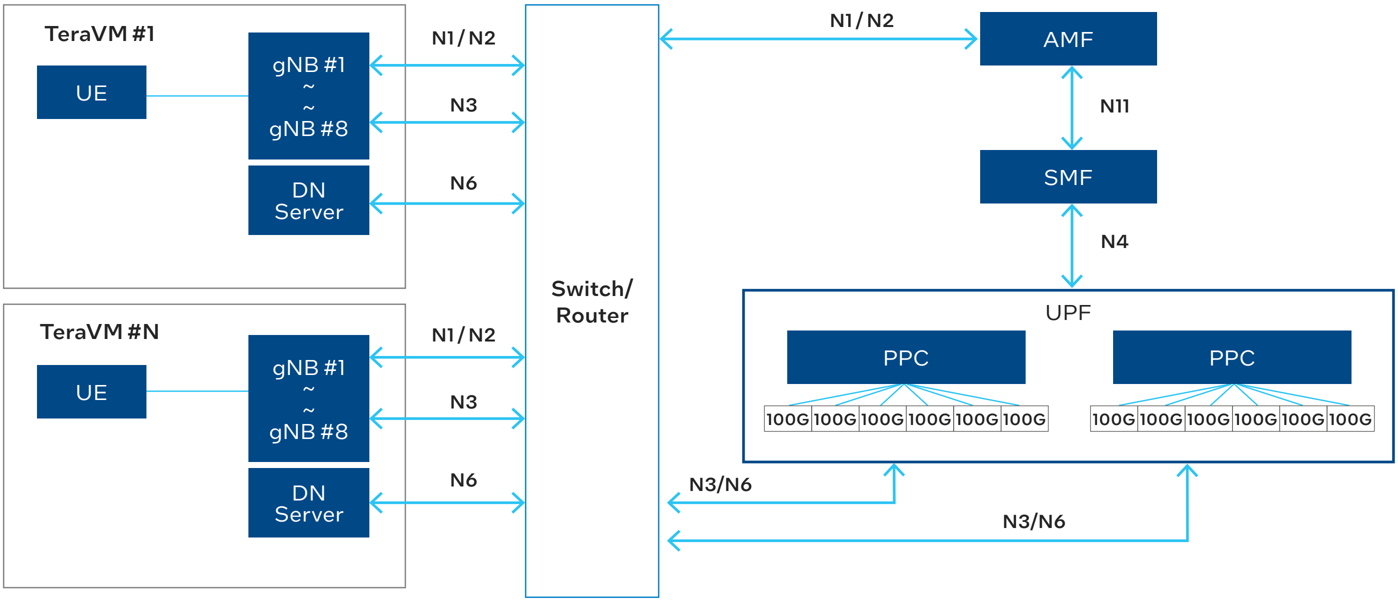

Figure 1. Network Test Call Setup

After full 5G-call setup, TeraVMs sent and received downlink/uplink high-volume data traffic, and switches/routers in Samsung Network Data Center forwarded them between TeraVM (DN server, UE) and 5G-UPF.

4.2. Hardware Configuration

The test used an Intel reference server. Two-socket Intel Xeon Platinum 8490H processors provided 60 physical cores with Intel® Hyper-Threading Technology enabled to get more packet forwarding cores. Each socket was connected to two Intel Ethernet Network Adapters E810-2CQDA2.

Samsung Electronics Labs conducted all tests on April 23, 2023.

CPU	Intel® Xeon® Platinum 8490H processor
Number of CPUs	2 (60 physical cores per CPU)
Number of vCPU for packet forwarding	114 x 2
Memory	512GB DDR4
Network Adapter	6 Intel® Ethernet Network Adapters E810-2CQDA2

4.3. Software Configuration

OS	Ubuntu 18.02
UPF(CNF)	SCR(Samsung CNF Release) 22D
DPDK	21.08
VPP	21.01
Cloud Management	Kubernetes 1.4

4.4. Test Traffic Model

Number of Subscribers	Total Subscribers: 600,000 Active Subscribers: 10,000
UPF(CNF)	Samsung CNF Release (SCR) 22D
PDR/QER/URR	1000/200/100
Traffic Volume Ratio	UDP 100%
Packet Size	920 bytes

4.5. Network Topology and UPF Resource Configuration

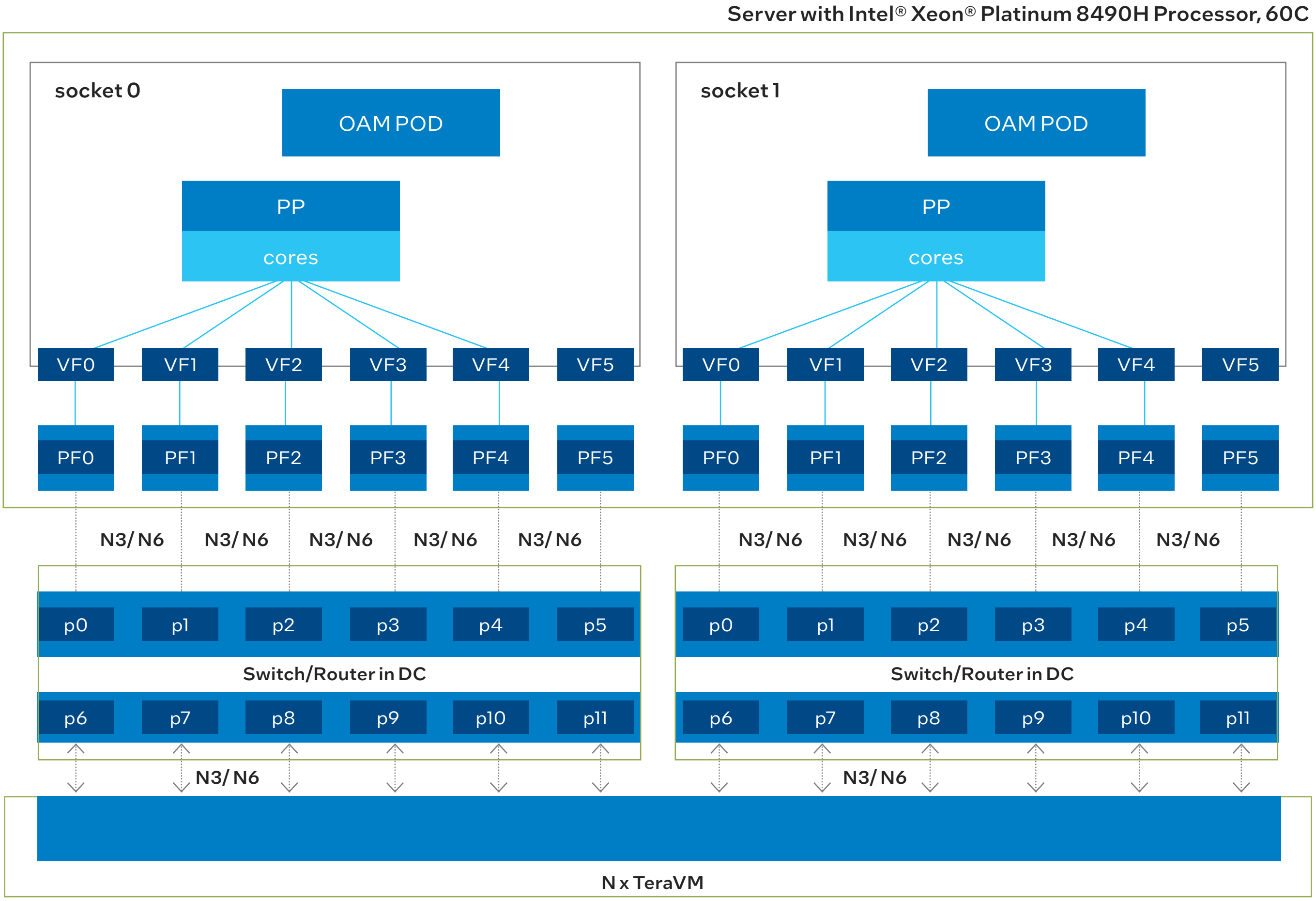

Figure 2. UPF Performance Test Network Topology

The performance test used two Intel reference servers with Intel Xeon Platinum 8490H processors, and Kubernetes for automating deployment, scaling and management of containerized applications.

Kubernetes master node creates and manages two pods on Kubernetes work node on each socket, based on YAML file configuration. Kubernetes schedules Packet Processing (PP) pod on each socket of the Intel 8490H processor server.

A single PP pod occupies 114 vCPUs. One vCPU is allocated to the timer and one for communication for external containers. The database for OAM monitoring and configuring the data plane occupies two vCPUs.

12 x 100Gbps E810 ports of 2 x 5G-UPF pod are prepared and connected to TeraVM ports through Samsung Data Center network. Each physical function (PF) on E810 is matched to a VF. The whole network topology for the performance test is shown in Figure 2.[2]

4.7. Performance Test Results

Basic forwarding performance of Samsung 5G-UPF was measured with URR/QER policy configuration enabled.

The measured performance was 988 Gbps with average core load of 90%. Prior tests measured 305 Gbps. Achieved 1Tbps without URR/QER.[2] See Figure 3.

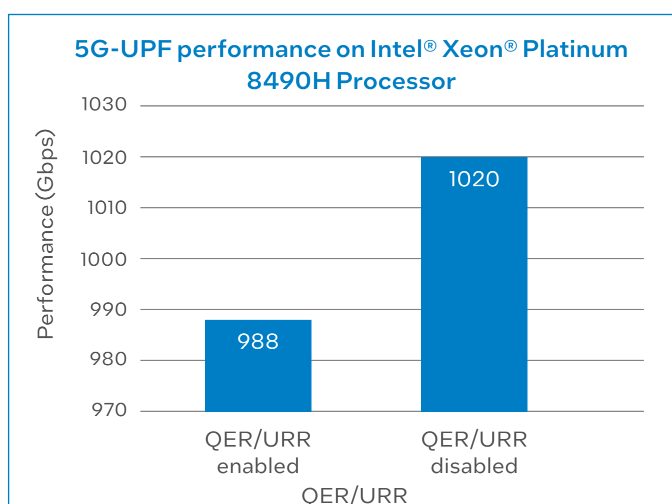

Figure 3. Forwarding performance with QER/URR policy enabled / disabled

5. Conclusion

Samsung achieved 1 Tbps of throughput in the 5GC user plane with the latest Samsung 5G Core software running on Intel reference servers with 4th Gen Intel Xeon Scalable processors and Intel Ethernet Network Adapters.

Samsung and Intel have collaborated on performance and latency optimizations of Samsung 5G Core, which builds on the cloud native architecture to deliver high capacity and reliability required for network applications. As a result, Samsung is fully capable of meeting the exponentially increasing traffic demands from diverse subscribers and industries and is ready to provide its solutions for global network operators to evolve their 5G networks.

Term	Description
5GC	5G Core
AMF	Access and Mobility Management Function
CG-NAT	Carrier-Grade Network Address Translation
CNFs	Containerized Network Functions
CNI	Container Network Interface
DDP	Dynamic Device Personalization
DL/UL	Downlink/Uplink
DN	Data Network
DPDK	Data Plane Development Kit
gNodeB	Next Generation Node Base Transceiver Station
GTP	GPRS Tunnelling Protocol
LLC	Last-Level Cache
OAM	Operation And Management
PP	Packet Processing
RSS	Receive Side Scaling
SMF	Session Management Function
SR-IOV	Single Root Input/Output Virtualization
UE	User Equipment
UPF	User Plane Function
URR/QER	Usage Reporting Rules/QoS Enforcement Rules
vCPU	Virtual Central Processing Unit
VNF	Virtualized Network Function
VPP	Vector Packet Processor

More Information

https://www.intel.com/content/www/us/en/products/sku/231747/intel-xeon-platinum-8490h-processor-112-5m-cache-1-90-ghz/specifications.html

https://download.intel.com/newsroom/2023/data-center-hpc/4th-Gen-Xeon-Scalable-Product-Brief.pdf

https://ark.intel.com/content/www/us/en/ark/products/192558/intel-ethernet-network-adapter-e810cqda2.html

Notices & Disclaimers

See backup for workloads and configurations. Results may vary.

Performance varies by use, configuration and other factors. Learn more at https://edc.intel.com/content/www/us/en/products/performance/benchmarks/overview/

[1] See [N9] at https://edc.intel.com/content/www/us/en/products/performance/benchmarks/4th-generation-intel-xeon-scalable-processors/. Results may vary.

[2] See prior whitepaper - https://builders.intel.com/docs/networkbuilders/samsung-achieves-305-gbps-on-5g-upf-core-utilizing-intel-architecture.pdf

This test result was conducted at optimized lab environment. Actual field result may vary under circumstances. 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 and Samsung technologies may require enabled hardware, software or service activation.

0823/BB/MESH/PDF 355878-001US