



Supporting Evolved Packet Core for One Million Mobile Subscribers with Four Intel® Xeon® Processor-Based Servers

The telecommunications industry continues to work toward capturing the full benefits of Network Functions Virtualization (NFV)—in terms of cost savings, flexibility, and ease of deployment—in large-scale operational deployments. This development work seeks to meet the performance challenges associated with complex interdependencies, state, and scale in mobile telecommunications networks, without the use of costly special-purpose hardware. Similar to the trend in general data centers, decoupling software functions from the underlying hardware through advances in NFV and Software-Defined Networking (SDN) enhances flexibility, compared to using equipment that is performance-optimized for a specific network function.

As demand for telecom services continues its dramatic expansion, infrastructure capacity challenges become more prevalent among Communication Service Providers (CSPs). Transforming carrier networks to take advantage of general-purpose hardware using approaches such as NFV has become a competitive and strategic imperative. At the same time, however, workload characterization and performance measurement for mobile network infrastructure has remained primarily the province of the large manufacturers of special-purpose equipment. Proprietary Key Performance Indicators (KPIs) specific to a single vendor's equipment have been the industry norm, leaving a gap in terms of identifying workloads and techniques to specify requirements and measure performance on virtualized mobile network infrastructure.

This paper represents a step toward developing test workloads and processor, memory, and I/O capacity requirements for transformed mobile carrier networks based on NFV and general-purpose, standards-based Intel® architecture by testing an Evolved Packet Core (EPC) running on a scalable server cluster.

The work reported on here was conducted by Intel in collaboration with Sprint and other industry experts on the topics of call modeling and conformance to 3GPP* specifications. It involved the analysis of wireless infrastructure topologies and markets, as well as the development of workloads to represent real-world operational networks. In addition to design guidelines for ensuring robust operation on Intel architecture, modeling and analysis tools were created to study the requirements and outcomes of the infrastructure's operation under both typical and stress traffic loads.



Call models and traffic models were established to represent usage by one million subscribers in the United States (US). Test workloads were then created to stress the system at far higher levels than those real-world scenarios, to ensure that the system under test had ample headroom to handle large-scale usage spikes. A control-plane load of 12,000 events per second was processed while simultaneously processing approximately 5.6 Gbps uplink and 33 Gbps downlink user-plane traffic. These results support the conclusion that a four-node cluster of servers based on Intel® Xeon® processors, configured as shown in Table 1, can conservatively support one million typical US mobile subscribers, with sufficient unused cores for quality of service, policy, and deep packet inspection, which will be studied in future testing. Finally, further tests prove that the server cluster can scale by adding more server nodes.

Table 1. Configuration of test servers used in study.

| Component | Description |
|-------------|--|
| Processors | Two Intel® Xeon® processors E5-2697 v2 (2.70 GHz, 30 MB cache) |
| RAM | 64 GB RAM |
| Network I/O | Intel® 82599 10 Gigabit Ethernet Controller |

1 Background Overview of Wireless Infrastructure Planning and Operation

A high-level representation of the overall topology of a mobile network is illustrated in Figure 1, divided primarily into the traffic network and the transport network. The traffic network fans out services from the carrier’s core infrastructure to the far reaches of the geographic area of service; it comprises three main parts:

- **User Equipment (UE)** such as mobile phones, tablets, and Wireless Wide-Area Network (WWAN) cards
- **Access edge**, made up primarily of transceivers, antennas, and cell towers
- **Backhaul**, which connects the access edge to the carrier’s core infrastructure

The transport network includes the core infrastructure (which is built over and tied together by the CSP’s backbone) and the Border Gateway (BGW) routers that connect the core infrastructure to the Internet at peering points. Voice and data networking for 4G LTE* traffic are converged into the EPC architecture, the main components of which are discussed later in this section.

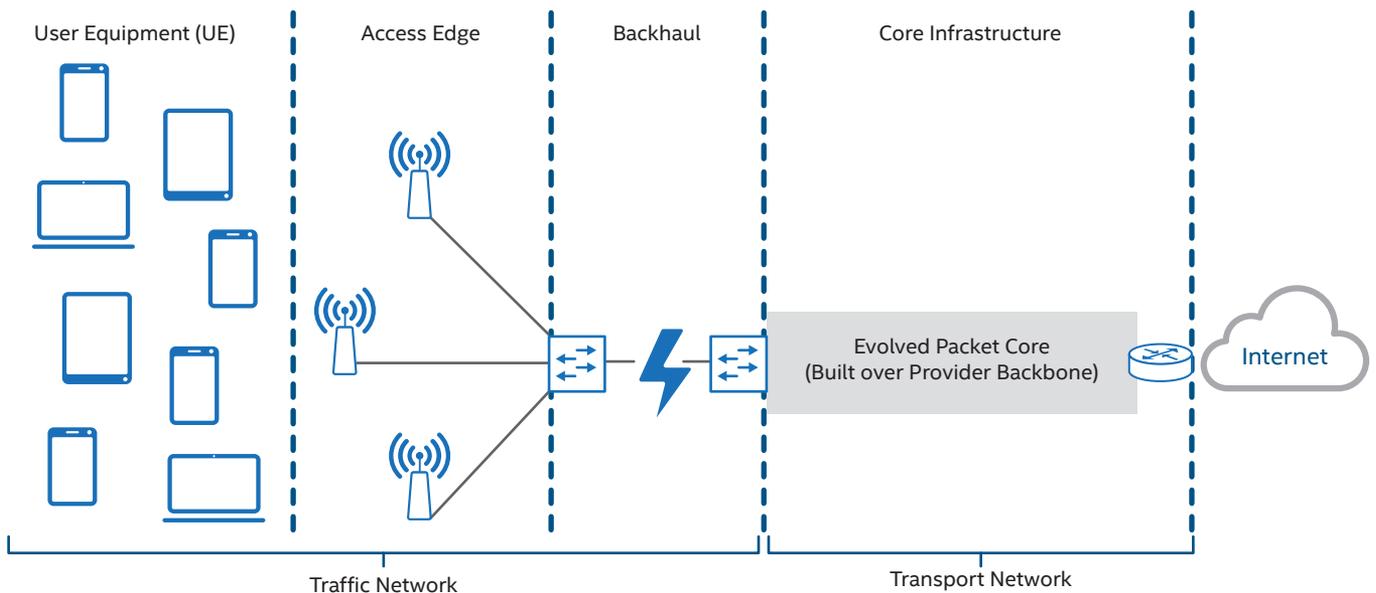


Figure 1. High-level mobile network topology.

Capacity Planning and Network Engineering

Sizing of the elements throughout the mobile network is based on the data volume (sometimes referred to as “tonnage”) that the infrastructure must be able to handle in a given period of time. Historical tonnage information is combined with growth projections for the network to specify the capacity requirements. An accurate understanding of the performance to be expected from a given network design is critical; the capital-intensive nature of deployments makes over-provisioning financially untenable, while under-provisioning will compromise service quality and potentially result in lost customers or a damaged reputation for the CSP.

Network load is commonly quantified as capacity needed to support the busy hour, which corresponds to the maximum amount of traffic a single network element (such as a router, switch, or gateway) must handle over any one-hour period. The average capacity per second is defined as busy hour capacity divided by 3600 (the number of seconds in an hour). Note that busy hour traffic load increases as one moves from the access edge toward the network core and traffic aggregates into larger flows. Likewise, variations in capacity demand per busy hour decrease as progress is made toward the core infrastructure. Thus, traffic load increases from access to core, but with less variation over time.

Average capacity demand cannot be used to determine network capacity requirements, especially in the access edge and backhaul portions of the network, because capacity requirements at some points in time will necessarily exceed that average. Instead, CSPs base capacity dimensioning on a calculated value called Effective Bandwidth (EBW), which accounts for variations in demand, as well as average and peak capacities. Network elements are typically over-provisioned by amounts approaching 30 percent at the access edge, decreasing progressively with the variations in capacity demand per BH closer to the network core.

Typical EPC Architecture and Configuration

The EPC, which has a typical structure similar to that shown in Figure 2, is the voice- and data-processing center of critical operator infrastructure assets. It connects to the Radio Access Network (RAN)—not shown in the diagram—which is made up of the access edge, coverage cells, and connection bearers bundled into licensed bands. The signaling or control plane is terminated at the Mobility Management Entity (MME), which processes all UE signaling events. The Non-Access Stratum (NAS) resides between the UE and the MME in the control plane and is responsible for call-processing session-management functions. The user plane that carries application traffic is terminated at the Serving Gateway (SGW).

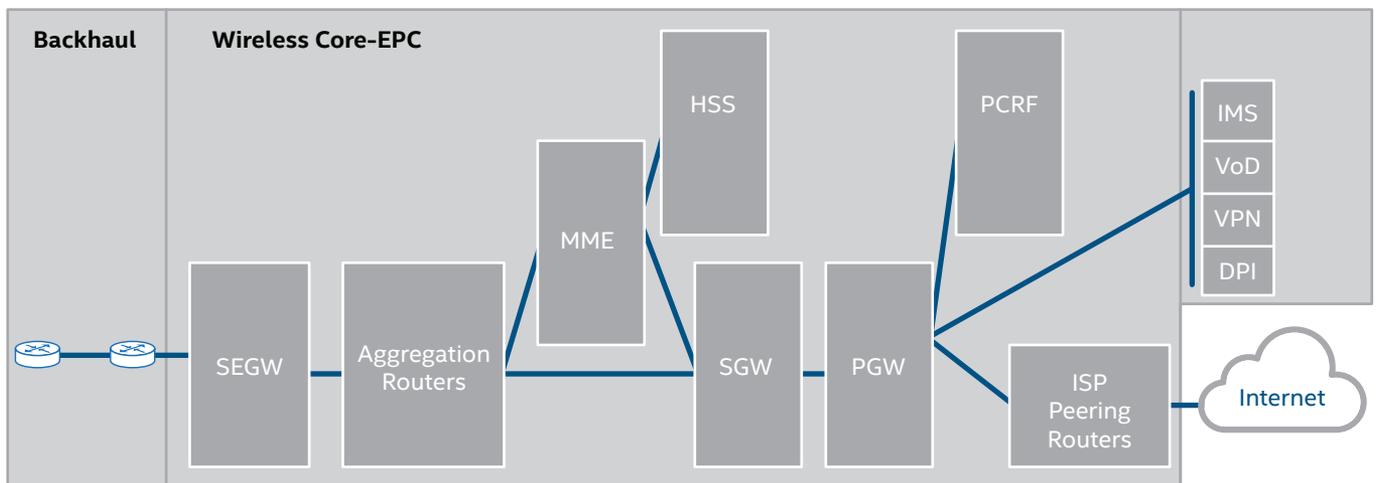


Figure 2. Typical wireless core infrastructure (EPC) configuration.

Application traffic is negotiated through the Packet Gateway (PGW) onto the Internet through Internet Service Provider (ISP) Peering Routers. The CSP may also offer additional, operator-hosted services that are not bound to the EPC; application traffic flows belonging to these operator hosted services are terminated at the SGI-LAN. The SGI-LAN also presents an area for soft landing of new services before they can be moved further south into the EPC core.

The Home Subscription Server (HSS) is the central database that maintains a list of subscribed services for UEs registered with the network. The MME interrogates the HSS for authorization of the UE bearer establishment requests. The PGW retrieves rules to be applied to the UE application flows from the Policy Control and Rules Function (PCRF). The UE application flows are then passed via forwarding routers to the Internet or the SGI-LAN appliances.

Wireless network infrastructure planning involves three main factors:

- **Market design** is driven by the physical geography over which the wireless provider is licensed to radiate at given frequency bands. The number and placement of evolved NodeBs (eNodeBs) is based on the potential number and distribution of mobile subscribers. Workloads for this study are based on a market size of one million subscribers, which roughly reflects a mid-sized city.
- **Call model** is a calculation of the expected number of NAS signaling events per subscriber in a busy hour for a specific market design. This information helps quantify the dimensioning requirements for signaling equipment, especially the MME, with the key design criterion being the aggregate rate of signaling events per unit of time, rather than throughput (which is typically in the order of a few Mbps, even in large markets).
- **Traffic model** is based on maximum usage (during the busy hour) for each application type planned within the market design. The traffic model calculates the busy hour data demand per subscriber based on projected usage patterns and data rates for each application type. Based on that information, capacity requirements associated with each application type are calculated.

2 Test Workload Design

The separation of the NAS control plane from the user plane is an important architectural characteristic of mobile networks. The NAS control plane signaling procedures (e.g., call establishment, UE state transitions, UE hand-over across towers) must be successfully processed within specified response times to establish and then maintain a call session. Each user-plane application service flow is associated to a physical bearer that has been successfully negotiated by the control plane. UE application service flows can be transported over multiple bearers, with each bearer encapsulated inside a GTP-U (GPRS Tunneling Protocol-User Plane) tunnel terminating at the SGW.

Equipment implementing EPC functions is generally specified in terms of the maximum number of control-plane messages per second that can be processed for sustainable user-plane throughput. The control plane and user plane represent the two dimensions along which network load is applied, and EPC capacity must be able to scale independently in these two dimensions.

Stream Control Transmission Protocol (SCTP) allows for multiple transmission streams to occur simultaneously between two endpoints. The number of SCTP terminations on the MME and the event arrival rate in terms of NAS procedures per second are the two stress vectors on the control plane. The number of SCTP terminations the system is capable of handling determines the maximum number of transceivers in the access edge (eNodeBs), which typically correspond to cell towers) that can connect to it. The NAS event arrival rate capacity of the system defines the aggregate UE and application signaling activity that can be sustained over the connected eNodeBs.

Similarly, on the user plane, the number of GTP-U tunnels the system can handle determines the number of bearers that can attach to the system. (The number of bearers equals the number of UEs unless there is more than one bearer per UE.) The aggregate packet arrival rate in terms of packets per second that can be handled by the EPC defines the sustainable system throughput for a specified application traffic mix over the established number of bearers.

In general, workloads are designed to characterize system performance under various stress conditions. Given that EPC system capacities are typically proprietary information, the key objectives of the workload design for this series of tests were the following:

- **Test EPC system performance** on Intel architecture under various network load conditions.
- **Parameterize system resources** required for different network loads.
- **Gauge the impact of virtualization** on system resources for parameterized workloads (future work).

It was a key requirement of the test design that the stress applied by the workloads along the control-plane and user-plane dimensions were several times greater than would be the case in a similarly sized real-world scenario. Further, to ensure a true measure of performance, the EPC responses to NAS event messages were forced to be completely compliant to the 3GPP protocol specifications, with zero retransmissions or failures. The overall design approach was first to obtain baseline EPC hardware requirements and then to incrementally increase the applied loads along individual dimensions to measure the associated impacts on resource requirements.

Details of the call-model workload used for this study are shown in Table 2. For each event type, the maximum value without packet loss was targeted. As characterized in the rightmost column, the event rates used are substantially higher than would be expected for one million US subscribers. Note that only NAS events that presented large-scale loads on the EPC are included in the workload.

Table 2. Call model used in study.

| Call Model Non-Access Stratum (NAS) Events | Busy Hour Event Rate: One Million US Subscribers | Test Workload NAS Event Rate | X Times Higher than Typical US Market |
|---|--|------------------------------|---------------------------------------|
| User Equipment (UE) Attach Attempts | 556 | 7,400 | 13 |
| UE-Initiated Service Requests | 5,556 | 12,000 | 2 |
| UE Detach Attempts | 139 | 7,400 | 53 |
| S1 Handovers Between Mobility Management Entities with Serving Gateway (SGW) Relocation | 56 | 12,000 | 216 |

The traffic model used in this study is shown in Table 3, including the types of traffic used as well as the proportions of uplink versus downlink traffic. This traffic model indicates the target uplink and downlink packet sizes (673 bytes and 172 bytes respectively), as well as the 6.4 Gbps downlink and 0.64 Gbps uplink target traffic rates expected for the one million subscriber sample.

Table 3. Traffic model used in study.

| Traffic Model Parameters | Typical US Data | | Workload Values | | X Times Higher than Typical US Market |
|---|----------------------|--------------------|----------------------|--------------------|--|
| | Downlink Traffic Mix | Uplink Traffic Mix | Downlink Traffic Mix | Uplink Traffic Mix | |
| Voice over LTE* | 5% (estimate) | 32% (estimate) | 5% | 321% | same |
| Web/HTTP | 52% | 43% | 52% | 43% | same |
| Streaming Video | 29% | 12% | 29% | 12% | same |
| Smartphone Apps | 7% | 7% | 7% | 7% | same |
| Email and Other | 7% | 6% | 7% | 6% | same |
| Average Packet Size (bytes) | 673 | 172 | 345 | 172 | 0.5 D/L 1.0 U/L (Smaller = more demanding) |
| 2012 S1U TPUT/1M subs (Gbps) [§] | 6.4 | .64 | 33 | 5.6 | 5 |
| Packet Rate (Mpps) | 1.07 | .32 | 11.5 | 3.6 | 11 |

[§]Overhead = 74 bytes (GPU + Ethernet + MAC)

Maximum uplink and downlink packet rates with negligible data loss were established. Additionally, to stress the system beyond the levels that would be expected in real-world usage, the test download-packet size was reduced by approximately half, from the expected 673 bytes to 345 bytes. This reduced packet size is a reflection of the fact that smaller packet sizes consume higher levels of CPU and other system resources for the same given throughput. Accordingly, packet rates are a more critical metric than data size, and the test workload targeted rates 11 times the expected levels for one million subscribers. Sustainable uplink and downlink rates well above target levels were generated while control plane messages were processed within 3GPP protocol specifications, without packet loss or retransmission.

3 System Test Setup

After several phased tests, the four-node integrated Control Plane Engine (CPE), Data Plane Engine (DPE) cluster shown in Figure 3 was established. A server cluster was set up using four servers based on Intel Xeon processors E5-2697 v2 (see Table 1) in a mesh configuration. A Switch-Route-Forward (SRF) module combines multiple compute units that can be connected either using top-of-rack switches or a mesh topology into a single logical entity. A mesh topology was used in this set of tests.

The NG40-RAN* packet generator provided by the German manufacturer ng4T provides 2G, 2.5G, 3G, and 4G emulations. Used by many operators and equipment manufacturers, the packet generator is designed to run on standards-based, commercial-off-the-shelf server hardware as well as in virtual machines. It was used to inject packets on the uplink and downlink interfaces of the servers at the highest rate sustainable.

The EPC stack from Connectem, a company headquartered in Silicon Valley, provides a streamlined EPC architecture that supports collapsed internode communication and independent dimensioning of signaling capacity and bearer capacity.

As shown in Figure 3, each of four clustered server nodes hosts an instance of the Connectem DPE (which implements user-plane processing functions of the typical serving gateway and packet gateway in the EPC), with a minimal CPE (which aids in signaling functions of the EPC). In addition, NODE1 hosts the complete MME function, as well as an IP Manager (IPM), Authentication Center (AuC), and related functions of the EPC. This node also interfaces with the other nodes by means of an internal GbE management interface (not shown).

Each node in Figure 3 has two processor sockets with 12 physical cores each; Intel® Hyper-Threading Technology brings the total to 24 logical cores (L-cores) per socket, represented in Figure 3 by the grid pattern. Note that a significant number of L-cores are not utilized, especially in NODE2, NODE3, and NODE4. This approach was taken in the testing to allow processing overhead for other functions, as well as to verify the viability of implementing a multi-node cluster in a real-world scenario that would scale well.

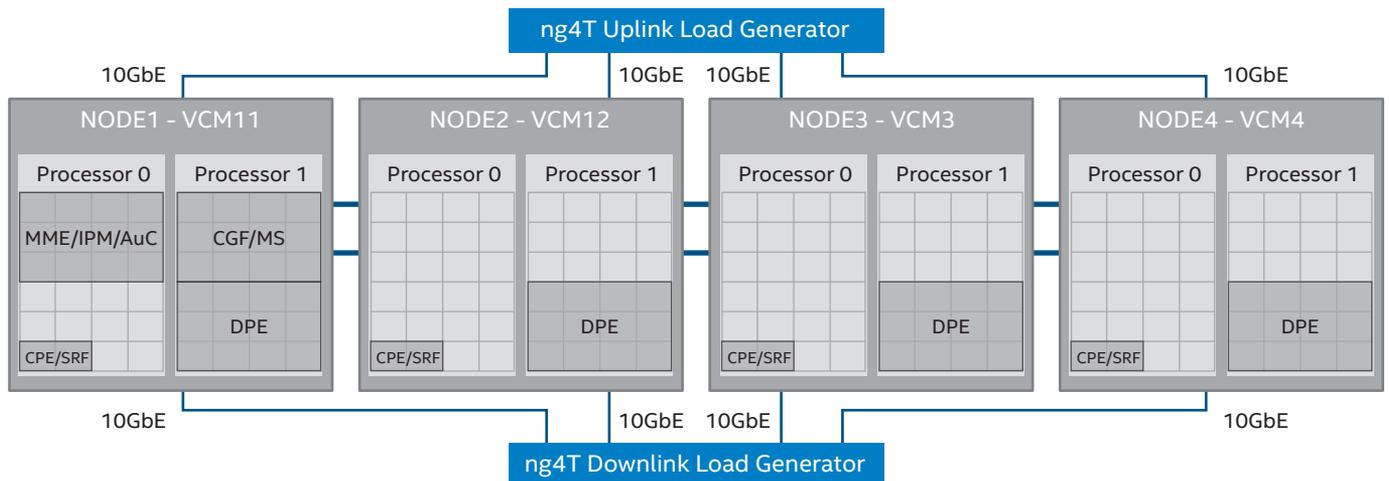


Figure 3. High-level view of the four-node integrated control plane engine/data plane engine cluster.

4 Results

Testing showed that the four-server cluster configuration shown in Figure 3 could process 12,000 control plane events per second while simultaneously processing 5.6 Gbps uplink traffic at a 172-byte packet size and 33 Gbps downlink traffic at a 345-byte packet size. Control messages and user-plane packets were processed within 3GPP protocol specifications, with zero retransmissions or failures. During these tests, 84 logical processing cores were used out of a total of 192 deployed across the cluster, leaving 108 available, which could be used in a real-world implementation for other required processing, to support functionality such as Quality of Service (QoS) or Deep Packet Inspection (DPI).

Since the test call model and test traffic model represent factors of traffic greater than expected for one million US subscribers, this study strongly suggests that the four-server cluster tested could serve one million US mobile LTE subscribers (or by extension, one million subscribers with similar or less demanding usage patterns elsewhere in the world).

Additionally, independent linear scaling of the user plane was demonstrated by the following method.

1. Traffic load (as shown in Table 2) equivalent to 600,000 subscribers (with 600,000 GTP-U tunnels) was applied to a two-node cluster with 7,200 control-plane events per second. The first node was configured like NODE1 in Figure 3 with the MME running on 12 logical cores as shown. The second node was configured like NODE2 in Figure 3 – a typical data-plane node. It is important to note that control plane traffic on NODE1 was not at maximum in the two-node cluster. This approach allows an increase later in testing to prove that control-plane load does not impact user-plane throughput.
2. User plane load was then applied to the two-node cluster, and maximum processing capacity was measured at ~1.4 million packets per second (mpps) per logical core at a typical data plane (i.e., DPE) node with negligible packet loss.
3. Next, the cluster size was increased to four server nodes by adding two more data-plane nodes to create the entire cluster, as shown in Figure 3.
4. Without changing the configuration of NODE1, control-plane load was then brought to a capacity of one million subscribers (with one million GTP-U tunnels) and 12,000 control-plane events per second.
5. User plane processing capacity was measured again and found to remain at ~1.4 mpps per logical core, showing that the user plane can scale by adding servers to the cluster and that user-plane performance is not significantly impacted by control-plane load.

As a result of this testing, a reusable EPC test bed is now available for further inquiry, including the use of unused CPU cores for QoS, DPI, and other higher-layer processing. Additional testing will investigate higher levels of scaling, as well as server-platform requirements for new call models driven out of new services such as machine-to-machine communication.

5 System Optimization

Pre-integrated “EPC in a box” solutions typically combine pre-validated combinations of elements while retaining the underlying architecture of a similar solution produced from separate components. Thus, such a solution instantiated in a virtualized cloud deployment tends to lack architectural flexibility for on-demand capacity provisioning, thereby failing to achieve a fundamental design goal of NFV.

As an alternative, the test design for this study approaches the NFV-friendly EPC implementation by seeking to take advantage of the following factors associated with Intel architecture:

- The homogenous nature of underlying systems
- Packet-processing optimizations for processing, memory, and I/O features
- The ecosystem of pre-tuned, open-standards software

The streamlined architecture of the Connectem EPC stack eliminates a substantial amount of internode EPC overhead. The Connectem EPC maintains 3GPP compliance at all external interfaces, while combining several internal EPC interfaces into a Control Plane Engine (CPE), Data Plane Engine (DPE), and session database. The SGW and PGW are combined into the DPE, and the CPE and DPE communicate asynchronously through the session database, which also implements the HSS functionality. This architecture not only eliminates a great deal of internode communication overhead as stated, but also decouples the control plane and user plane, allowing independent scaling of EPC functions.

Pipeline Architecture

The user-plane stack was designed to support scalable, performance-optimized EPC functionality, largely by due consideration of the following requirements:

- **Avoid serialization** in the packet-processing pipeline, including serializing events such as locks, special instructions such as CLFLUSH, and large critical sections.
- **Access data from the cache where possible** by making use of prefetch instructions and observing best practices in design of the software pipeline.
- **Design data structures to be cache-aligned** and avoid occurrences of data being spread across two cache lines, partial writes, and contention between write and read operations.
- **Maintain affinity between software threads and hardware threads**, as well as isolate software threads from one another with regard to scheduling relative to hardware threads.
- **Break down user-plane functionality** so that it can be implemented with a combination of RTC (Run to Completion) and pipeline methods.

Switch-Route-Forward (SRF) Module Design

The SRF module is designed either to coexist with the EPC application or to be deployed separately as part of the Open vSwitch* layer. To optimize scalability, the SRF module combines multiple compute units that can be connected either using top-of-rack switches or a mesh topology into a single logical entity. The mesh topology used in this set of tests consists of two fundamental blocks:

- **SRFd – Switch Route Forward Daemon** mainly performs the control portion of the bridging functionality.
- **SRF – Switch Route Forward Data module** mainly handles the packets coming from external entities.

This architecture enables fine-grained, flow-level load balancing while maintaining the coarse-level view to external entities.

6 Conclusion

NFV and open-standards servers based on Intel Xeon processors represent a compelling alternative to the higher-cost, purpose-built hardware that is widely used in mobile networks. Intel architecture offers greater flexibility than older, single-purpose equipment, as well as outstanding scalability. As the industry moves to adopt NFV, carriers should be aware that Intel architecture—especially when combined with software such as the Connectem EPC that has been architected and optimized for homogeneous x86 servers—can provide outstanding performance and scalability for next-generation telecommunications infrastructure, with four servers easily supporting the equivalent of one million subscribers in this testing. Ongoing work with this test bed will include investigation of areas such as virtualization, QoS, and DPI, leading to even greater capabilities that will power the future of mobile carrier networks.

Learn more about Intel® Technology in Communications: www.intel.com/communications

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information. The products described in this document 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. Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order. Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or by visiting Intel's Web Site www.intel.com.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to www.intel.com/performance.

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families: [Go to Learn About Intel® Processor Numbers](#).

Intel Hyper-Threading Technology (Intel® HT Technology). Available on select Intel® Core™ processors. Requires an Intel® HT technology-enabled system. Consult your PC manufacturer. Performance will vary depending on the specific hardware and software used. For more information including details on which processors support HT Technology, visit www.intel.com/content/www/us/en/architecture-and-technology/hyper-threading/hyper-threading-technology.html.

Copyright © 2015 Intel Corporation. All rights reserved. Intel, the Intel logo, Intel Core, and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

