Solution Brief
Packet Processing on Intel® Architecture

Impressive Packet Processing Performance Enables Greater Workload Consolidation

Single Intel® Xeon® processor platform achieves Over 80 Mpps throughput[1]

With Intel® processors, it's possible to transition from using discrete architectures per major workload (application, control plane, packet and signal processing) to a single architecture that consolidates the workloads into a more scalable and simplified solution. This software-based approach, depicted in Figure 1, yields further benefits with Intel's 4:1 workload consolidation strategy. The hardware platform – based on general-purpose server technology – has been optimized using the best practices of the communications industry to ensure system robustness with core, memory and I/O scalability and performance improvements, to meet network operator's low to high-end system requirements.

This framework is made possible by the high performance level of Intel processors plus the Intel® Data Plane Development Kit (Intel® DPDK), which greatly boosts packet processing performance and throughput, allowing more time for data plane applications. As a result, telecom and network equipment manufacturers (TEMs/NEMs) can take advantage of lower development costs, fewer tools and support teams, and faster time-to-market with a solution benefiting from the economies of scale of the server industry. Such a notable level of consolidation is achievable through industry-leading performance, I/O throughput and performance per watt density – all on a standards-based platform. Solution providers listed in this paper are ready to help equipment architects jumpstart their next designs.

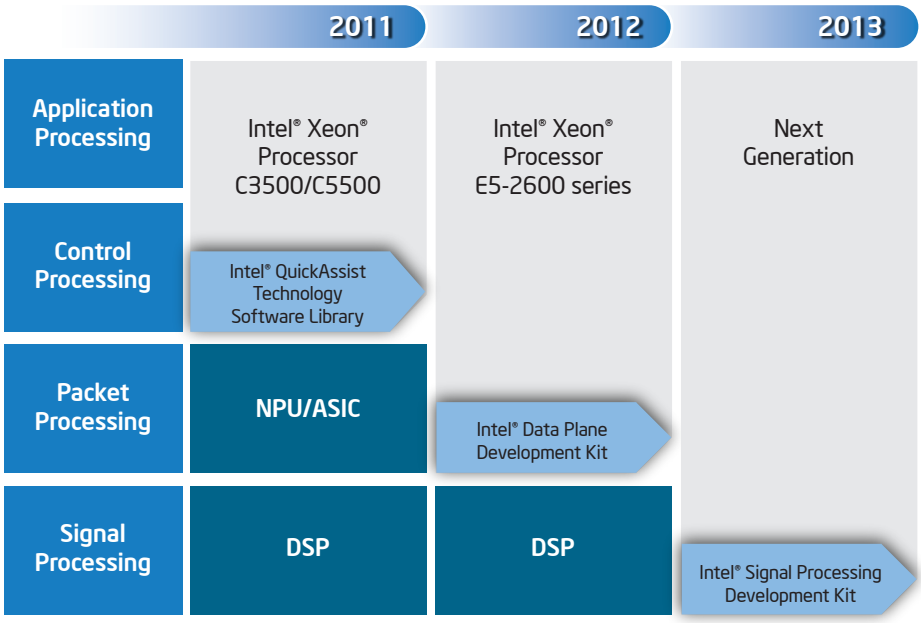

Figure 1. Intel's 4:1 Workload Consolidation Strategy

This approach also reinforces the synergies resulting from the convergence of telecom and data networks by blending the best of communications and computing competencies. What do both industries bring to the party in terms of technology and capability?

The communications industry excels in hardware density, high speed pipes, low latency transmission, robustness, exceptional quality of service (QoS), determinism and real-time I/O switching. On the computing side, there are standards-based technologies that enable dynamic resource sharing, workload migration, security, open APIs, developer communities, virtualization and power management.

What's Changed?

The Intel® Xeon® processor E5-2600 series, with an integrated DDR3 memory controller and an integrated PCI Express* controller, was a game-changer for delivering the small packet throughput required for next generation networks. This resulted in lower memory latency, decreasing from around 120 nanoseconds (ns) to about 70ns, on par with the arrival rate of 64 byte packets. Still, there's far more to do than just store packets to memory, and that's where the Intel DPDK makes a tremendous difference. The development kit reduces a significant amount of overhead when using an out-of-the-box, standard Linux* operating system. Significant time is saved by using core affinity, disabling interrupts generated by packet I/O, using cache alignment, implementing huge pages to reduce translation lookaside buffer (TLB) misses, prefetching, new instructions and many other concepts. The Intel DPDK also runs in user space, thus removing the high overhead associated with kernel operations and with copying data between kernel and user memory space.

Taking advantage of the breakthroughs enabled by the Intel DPDK, today's Intel® architecture-based platforms based on a single Intel Xeon processor E5-2600 series are achieving over 80 million packets per second (Mpps) of L3 forwarding throughput for 64 byte packets (Figure 2).[1]

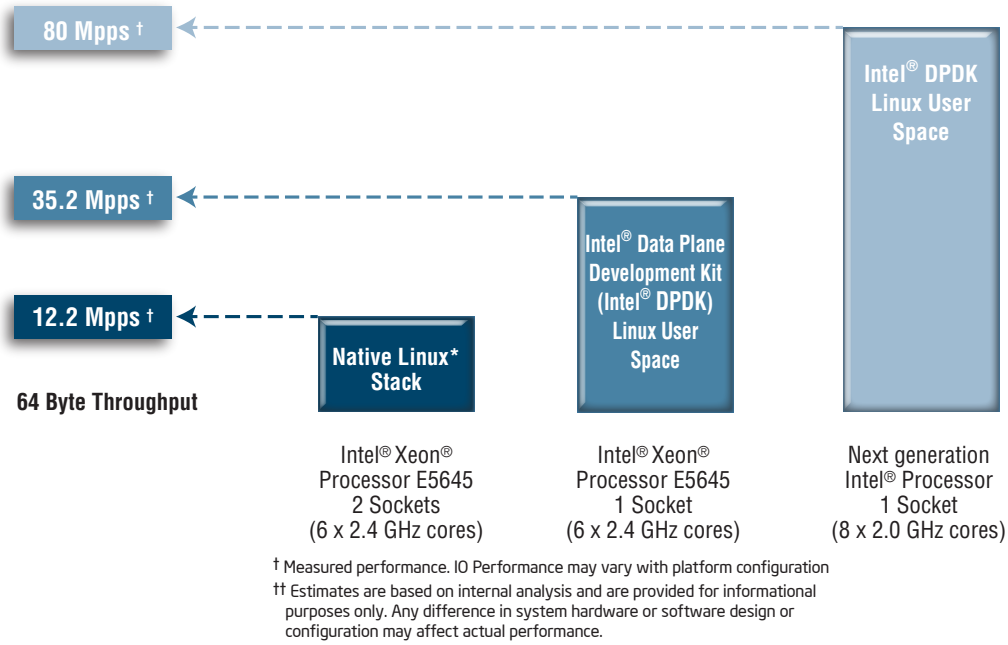

Figure 2. Breakthrough Data Plane Performance with Intel® Data Plane Development Kit (Intel® DPDK) L3 Packet Forwarding

Higher Packet Processing Performance

Intel architecture has a proven track record of delivering high performance and innovation over its long history of meeting requirements of various embedded communication markets. By integrating a memory controller and increasing memory bandwidth, the first generation Intel® Core™ i7 processor family achieved breakthrough performance when executing control and data plane workloads concurrently. This was possible, in large part, to the Intel DPDK, which greatly improved packet processing on Intel architecture.

The Intel DPDK, consisting of a set of libraries designed for high speed data packet networking, is based on simple embedded system concepts and allows users to build outstanding small packet (64 byte) high performance applications. It offers a simple software programming model that scales from Intel® Atom™ processors to the latest Intel® Xeon® processors, providing flexible system configurations to meet any customer requirements for performance and scalable I/O. While the Intel DPDK may be used in various environments, the Linux SMP user space environment is the most common among engineers due to the ease of development and testing of customer applications, along with maintaining leading edge data packet performance that easily outperforms native Linux, as shown in Figure 2. The L3 forwarding for the 2.40 GHz Intel® Xeon® processor E5645 on a native Linux stack is about 1 Mpps per core, compared to 6 Mpps (64 byte packets) per core using the Intel DPDK.

After testing how many 10 gigabyte pipes could be fully utilized using the Intel® Data Plane Development Kit (Intel® DPDK), John Basco, Distinguished Engineer at Verisign, couldn't believe his performance measurements. So, he went to some of his trusted colleagues and said, "You're going to think I'm crazy, but I'm getting these performance results. There must be a decimal point in the wrong place."

"They did their own evaluations, and it turned out the numbers were right. The Intel DPDK is really what I would consider a disruptive technology."

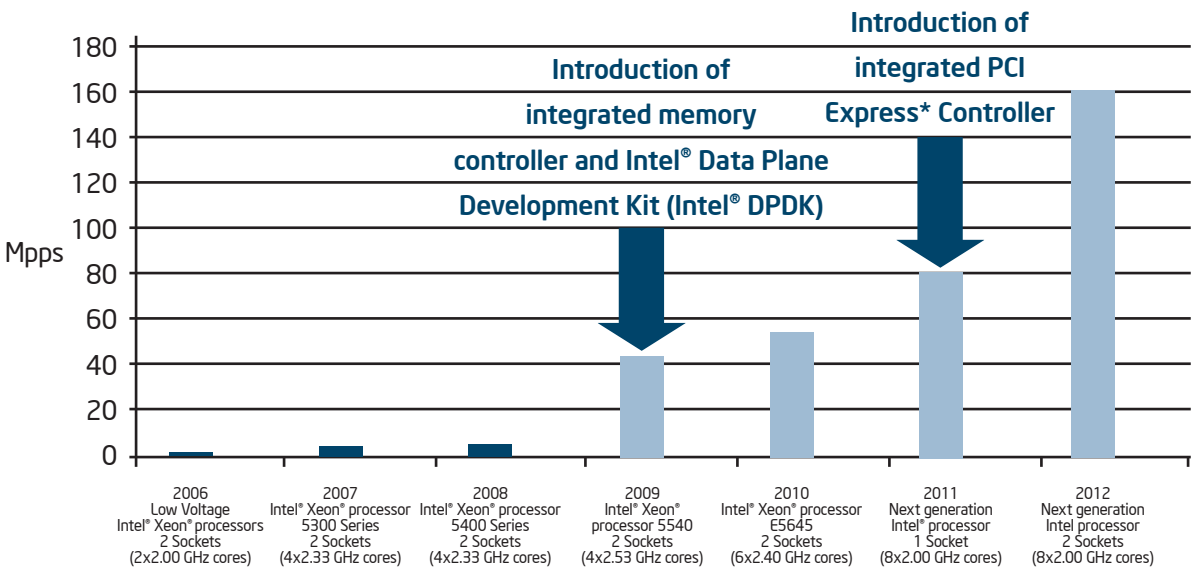

Figure 3. IPv4 Layer 3 Forwarding Performance for Various Generations of Intel Architecture Processor-based Platforms

With the addition of integrated PCI Express controllers, more cores and architecture enhancements, the 2nd Generation Intel® Core™ Processor Family provides even greater scalability and flexibility to embedded communication market segments. The Intel DPDK demonstrates the impressive small packet performance achievable using the latest Intel architecture in Figures 2 and 3. See the Appendix for platform information.

It should be noted in the case of IPv4 Layer 3 forwarding, the system throughput is I/O limited, constrained by the current PCI Express Generation 2 network interface cards (NICs); in other words, the processor's computing capacity was not fully utilized. Figure 4 shows the packet performance for various packet sizes and port usage cases for a single eight core 2.0 GHz Intel Xeon processor E5-2600 series with 20MB L3 cache and with Intel® Hyper-Threading Technology (Intel® HT Technology) disabled.

This impressive performance makes Intel architecture well suited to support a wide range of software-based networking components that perform routing, firewall, VPN and IPS, just to name a few services. Such a platform is highly adaptable and capable of running nearly any workload, allowing TEMs/NEMs to easily incorporate new features, services and applications. Furthermore, developers can differentiate their product offerings by leveraging complementary technologies from the server industry, like virtualization,

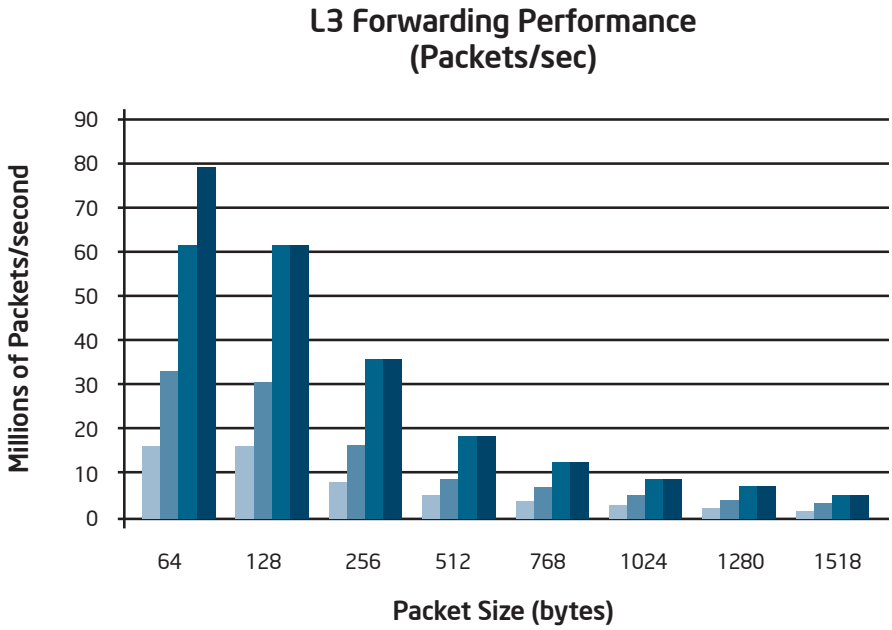

Figure 4. Eight Core Intel® Xeon® Processor E5-2600, L3 Forwarding Performance

power management and security. It is now possible to consolidate packet processing and network services onto a single processor architecture, which saves hardware and operating cost, eases application development and improves agility.

Development Kit Overview

The Intel DPDK provides Intel architecture-optimized libraries that allow developers to focus on their application. The Intel DPDK provides non-GPL source code libraries to support exceptional data plane performance and ease software development, while minimizing development time. This allows the developers to make additions and modifications to the Intel DPDK, as

"Telecom equipment vendors are trying to do more with less R&D spending, and it's important for them to move towards a common platform that runs multiple applications. The dramatic packet processing boost from the Intel DPDK enables our customers to converge on a single processor architecture, which helps them speed up their development cycles across all our platforms," said Paul Stevens, Telecom Sector Marketing Director at Advantech.

required, to meet their individual system needs.

Designed to accelerate packet processing performance, the Intel DPDK contains a growing number of libraries (Figure 5), whose source code is available for developers to use and/or modify in a production network element, with examples showing usage cases and performance. Developers can build applications with the libraries using "run-to-completion" and/or "pipeline" models that enable the equipment manufacturer's application to maintain complete control. The following provides a brief description of key software components:

- **The Environment Abstraction Layer** (EAL) provides access to low-level resources (hardware, memory space, logical cores, etc.) through a generic interface that hides the environment specifics from the applications and libraries.
- **The Memory Pool Manager** allocates NUMA-aware pools of objects in memory. The pools are created in huge-page memory space to increase performance by reducing translation lookaside buffer (TLB) misses, and a ring is used to store free objects. It also provides an alignment helper to ensure objects are distributed evenly across all DRAM channels, thus balancing memory bandwidth utilization across the channels.
- **The Buffer Manager** significantly reduces the amount of time the system spends allocating and de-allocating buffers. The Intel DPDK

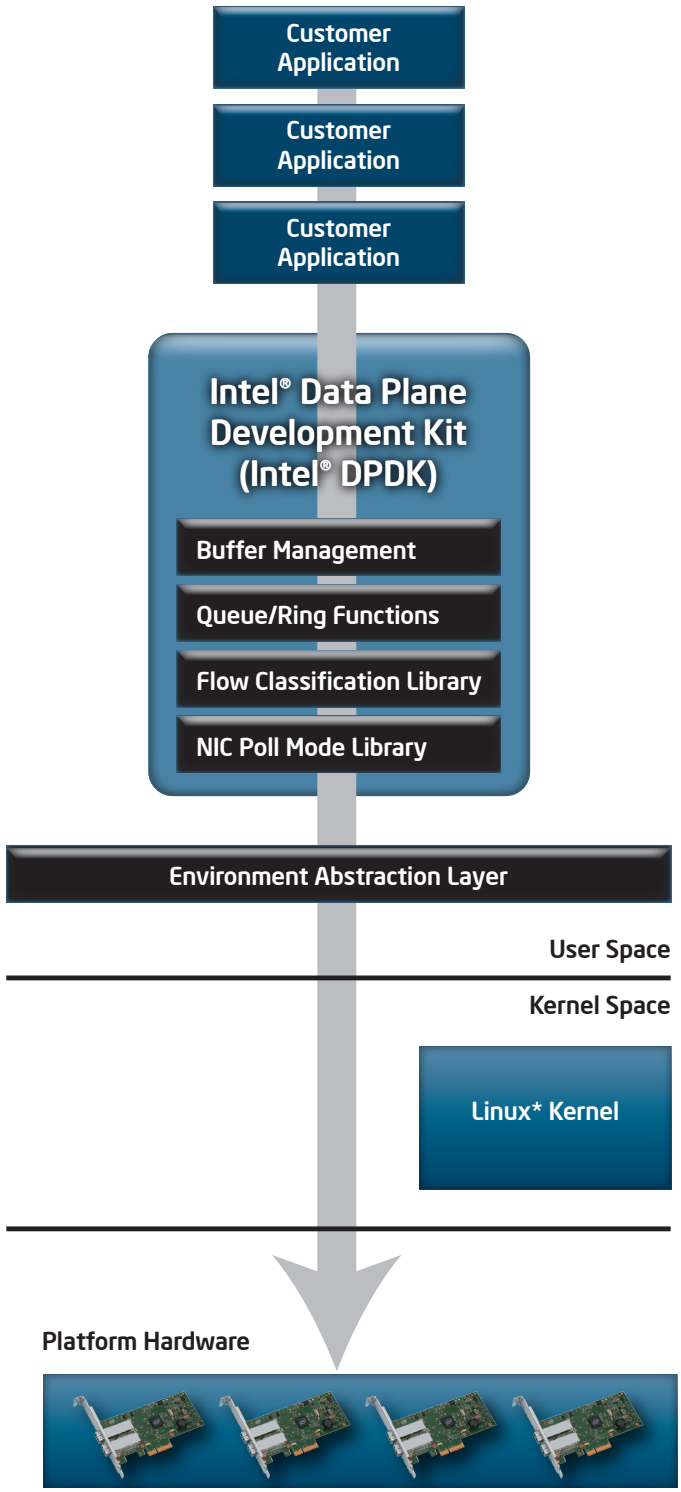

Figure 5. Major Development Kit Components

"High-touch packet processing is a critical piece for any telco equipment manufacturer trying to win business in today's highly competitive market. There's a lot of complexity in this software layer, and traditionally our customers did all that work themselves. Having the Intel® Data Plane Development Kit (Intel® DPDK) means customers don't have to reinvent the wheel, and they can focus on their differentiated value add," said Robert Pettigrew, Director of Marketing – Embedded Computing Business at Emerson Network Power.

pre-allocates fixed size buffers, which are stored in memory pools for fast, efficient cache-aligned memory allocation and de-allocation from NUMA-aware memory pools. Each core is provided a dedicated buffer cache to the memory pools, which is replenished as required. This provides a fast and efficient method for quick access and release of buffers without locks.

- **The Queue Manager** implements safe lockless queues instead of using spinlocks that allow different software components to process packets, while avoiding unnecessary wait times.
- **The Ring Manager** provides a lockless implementation for single or multi producer/consumer enqueue/dequeue operations, supporting bulk operations to reduce overhead for efficient passing of events, data and packet buffers.
- **Flow Classification** provides an efficient mechanism for generating a hash (based on tuple information) used to combine packets into flows, which enables faster processing and greater throughput.
- **Poll Mode Drivers** for 1 GbE and 10 GbE Ethernet controllers greatly speed up the packet pipeline by receiving and transmitting packets without the use of asynchronous, interrupt-based signaling mechanisms, which have a lot of overhead.

The Intel DPDK also includes examples to demonstrate use cases and contains many other libraries such as multi-process library, timers, logs, atomic functions, debugging and cryptography. Intel plans to continually add more functions and make enhancements to further increase performance. One noteworthy technique, called software pre-fetching, increases performance by bringing data from memory into the cache before it's needed, thereby significantly reducing memory latency. It's used in many places, including the classification library for n-tuple exact match lookups and the poll mode drivers, enabling them to speed up classification by preloading the packet header, typically the first 64 bytes, into cache.

Intel DPDK Fundamentals

- Implements a run to completion model
- Accesses all devices by polling without a scheduler
- Runs in 32-bit and 64-bit mode with/without NUMA
- Scales from Intel Atom processors to Intel Xeon processors
- Supports an unlimited number of processors and processor cores
- Optimizes packet allocation across DRAM channels
- Allocates memory from the local node where possible
- Ensures all data structures and objects are cache-aligned for greater performance

The Intel DPDK fast path application is created as a single image, which when

loaded, allocates resources, initializes setup, and assigns core affinity and the starting points of the execution code to each logical core allocated (handled by Pthread Affinity in Linux model), as depicted in Figure 6.

The Intel DPDK image can access multiple physical cores, as well as multiple logical cores, when Intel HT Technology is enabled. Each core in the image contains a dispatch loop that processes packets as they arrive. There are a number of models that may be used, based on the user's requirements. In many cases, a run-to-completion model is used; however, depending on the number of cycles to process a packet, a pipeline model may also be used by passing packets via a ring to another core for additional processing.

At least one run-to-completion core will be used for packet I/O, and there are various ways to distribute packets amongst the cores for processing ingress traffic. The simplest of these is for a core to handle as much ingress traffic as possible and load balance the remaining packets to other cores to perform the work required. Many NICs also have mechanisms to direct traffic to cores based on filters or hashes.

One such mechanism is receive side scaling (RSS), which assigns packets into different descriptor queues in order to efficiently distribute packet processing between several processor cores, thus increasing performance on a multi-processor system. RSS assigns an RSS index, resulting from the hash of the IP data, to each received packet. Packets are then routed to one of a set of Rx queues based on their RSS index; all packets from the same flow are sent to the same core per the hash. Via the pole mode drivers, the core reads packets from the NIC's queue and performs the required processing on the packets. For example, packet flows from a 10GbE interface are distributed and processed by multiple cores, each performing the same run-to-completion actions.

Besides RSS, some NICs have flow filters, which may discard packets or forward packets to specific receive queues handled by cores for processing. In the case of the Intel® 82599EB 10 Gigabit Ethernet Controller, the filters in the flow director identify flows or sets of flows and route them to specific queues, which are processed by cores. The flow director supports two types of filtering modes:

- Perfect match filters — the hardware checks for a match between the masked fields of the received packets and the programmed filters. The Intel® 82599 10 Gigabit Ethernet controller supports up to 8 K of perfect match filters.
- Signature filters — the hardware checks for a match between a hash-based signature of the masked fields of the received packet. Up to 32 K

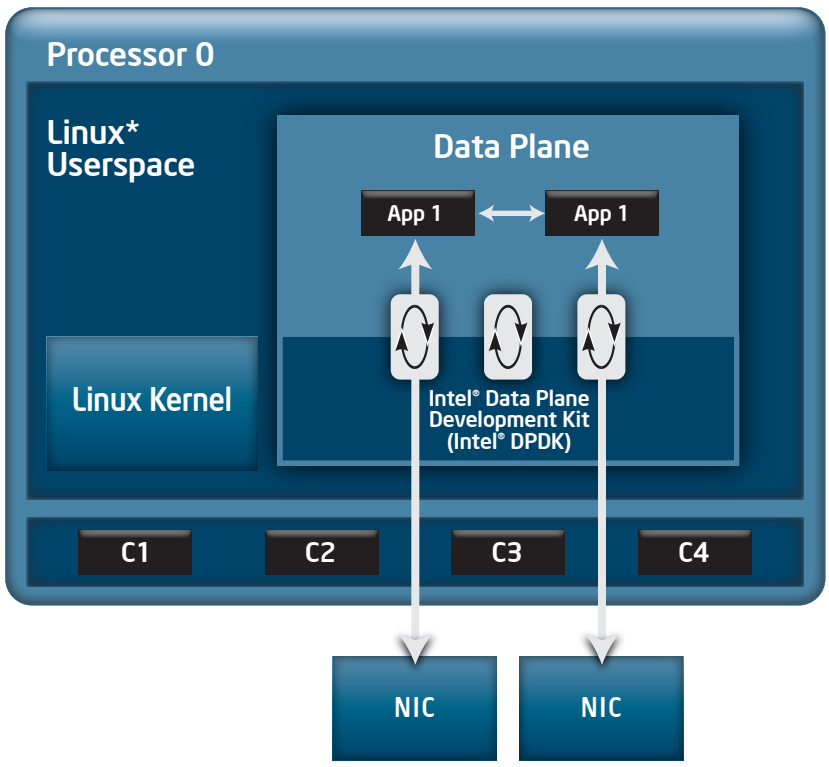

Figure 6. Linux* Model

signature filters are supported.

In a typical networking scenario, packets are received and classified, and subsequent action may follow. In the case of security applications, packet flows are recognized, actions are taken based on the flow and further processing may ensue based on the policies enforced by the application. Therefore, it is essential to process the traffic as quickly as possible to minimize the overall time needed to deliver value-added services. Visit **www.intel.com/go/DPDK** to find out more.

Migrating to a Single Architecture

Delivering the flexibility, scalability and capacity needed for next generation telecom and data networks, Intel architecture-based platforms enable a highly adaptable alternative to fixed-function hardware. They also provide greater workload consolidation that can lower design challenges around cost, software development, power consumption, time to market and team communication. Platforms will become even more powerful, following through on Moore's Law, further accommodating the workload consolidation onto one platform, which lowers architectural complexity, reduces BOM cost and speeds up product development.

Appendix: Using the L3 forwarding example in the Intel® Data Plane Development Kit (Intel® DPDK) for 64 byte packets, a server with

- dual Intel® Xeon® processor E5540 (2.53 GHz, 4 core) processed 42 Mpps.
- dual Intel® Xeon® processor E5645 (2.4 GHz, 6 core) system processed 55 Mpps.
- a single Intel® Xeon® Processor E5-2600 (2.0 GHz, 8 core) processed 80 Mpps (with Intel® Hyper-Threading Technology (Intel® HT Technology) disabled).
- dual Intel Xeon Processor E5-2600 (2.0 GHz, 8 core) processed 160 Mpps (with Intel HT Technology disabled) and 4x 10GbE dual port PCI Express* Gen. 2 NICs on each processor.

[1] Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit Intel Performance Benchmark Limitations

Copyright © 2012 Intel Corporation. All rights reserved. Intel, the Intel logo, Intel Core, Intel Atom and Xeon are trademarks of Intel Corporation in the United States and/or other countries.

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

Printed in USA MS/VC/0412 Order No. 327263-001US