# Optimized Data Plane Processing Solutions using the Intel® DPDK

Executive Summary

End-user usage patterns and expectations are creating challenges for both service providers and network equipment suppliers with respect to network capacity, system-level cost-performance, the delivery of advanced network services and time-to-revenue. These problems can be effectively addressed only at the platform level, as opposed to the component level.

This white paper describes how Intel and 6WIND* are working together to address these key telecom platform challenges through a unique combination of the latest processors and software. This close technical collaboration has delivered a highly-optimized platform with advanced features and capabilities for next-generation networking equipment.

The integrated platform, described in this paper, enables service providers to reduce their CAPEX and OPEX, while delivering the advanced services that are critical to subscriber acquisition and retention. Moreover, network equipment providers can take advantage of the platform in order to accelerate their time-to-market for LTE solutions that provide industry-leading cost-performance.
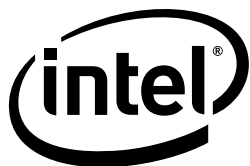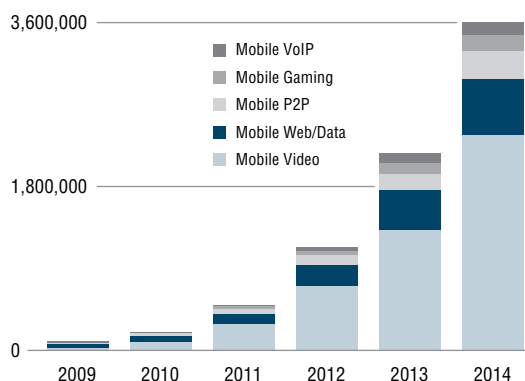
www.intel.com/go/commsinfrastructure   www.6wind.com

Packet Processing Challenges in Mobile and Fixed Networks

As part of the specification process for the LTE telecom standard, the 3GPP organization selected IP-based protocols as the standard architecture for LTE networks. This has presented telecom equipment manufacturers (TEMs) with significant opportunities to take advantage of networking elements already proven in enterprise and cloud infrastructure. Hardware platforms based on industry-standard processors and running commercially-available software can deliver the performance required for LTE networking subsystems at far lower cost and with much quicker development cycles than platforms using proprietary hardware such as ASICs, FPGAs or complex network processors. Unlike in the case of earlier circuit-switched 2G, 2.5G and 3G networks, equipment for LTE networks can benefit from the same hardware and software innovations found in enterprise and cloud systems. In fact, many "converged" networking solutions are applicable to both mobile and fixed networks.

TB per Month — **108% CAGR 2009-2014**

Legend: Mobile VoIP, Mobile Gaming, Mobile P2P, Mobile Web/Data, Mobile Video

2014 values: 4%, 5%, 8%, 17%, 66%

Source: Cisco* VNI Mobile, 2010

As designers work on next-generation equipment for converged networks, they face significant performance challenges due to end-users' usage patterns and expectations.

Dominated by mobile video and cloud-based services, Internet traffic is growing exponentially, challenging service providers to deploy ever-increasing network capacity to keep up with user demand. Average revenue per user (ARPU) is essentially flat; thus, in order to maintain an acceptable return on investment (ROI), service providers must constantly achieve cost-performance improvements as they introduce new equipment.

In parallel with the upsurge in raw traffic, subscribers increasingly expect advanced services on mobile platforms. These advanced services, such as location-based services and real-time translation, require more sophisticated processing within the network, which adds to the complexity and cost of core network equipment. On the other hand, these advanced services are crucial to the service providers' ability to monetize their networks as they shift from unlimited data plans to tiered pricing and migrate to policy-driven content distribution.

In summary, both mobile and fixed service providers operate under challenging cost constraints:

- CAPEX - low product cost is essential to support worldwide deployments of 4G networks.
- OPEX - electrical power needed to run both data center equipment and cooling is a major contributor to the overall total cost of ownership (TCO).

From the perspective of a network equipment supplier, high-performance packet processing is critical to addressing these CAPEX and OPEX challenges. In converged, all-IP networks, the performance of the packet processing subsystem determines both the overall system-level throughput (network bandwidth) and the latency, which is critical for applications as diverse as mobile gaming and financial transactions. At the same time, the advanced network services that are so important for network monetization rely on accelerated packet processing to reach the required performance levels for functions such as quality of service (QoS), deep packet inspection (DPI), video acceleration and policy enforcement.

Unified System Architecture

Until recently, processor-based network equipment was typically designed around a heterogeneous system concept, in which the networking control plane ran on one processor architecture (e.g., Intel® architecture processors) while the data plane executed on a different architecture, such as a multi-core MIPS platform, with specialized network acceleration features. In order to benefit from the level of performance attainable with such a system, developers would accept the additional complexity resulting from heterogeneous software architecture as well as the schedule penalties associated with the integration and maintenance of two different code bases.

Clearly, this is not an ideal solution, and a unified system architecture, in which the control plane and data plane run on the same processor architecture while achieving the necessary cost-performance targets, is preferable for several reasons: software development, debug and integration is simplified; processor resource utilization is improved because the control plane and data plane can be distributed among cores with greater flexibility; development schedule risk is minimized; and software maintenance is much easier with a common code base and a single development environment.

As explained in the next sections, recent generations of Intel® Xeon® processors provide the networking performance necessary to enable their use as a unified platform for converged networking equipment, especially when combined with high-performance packet processing software such as 6WIND*'s 6WINDGate* software and the Intel® Data Plane Development Kit (Intel® DPDK).
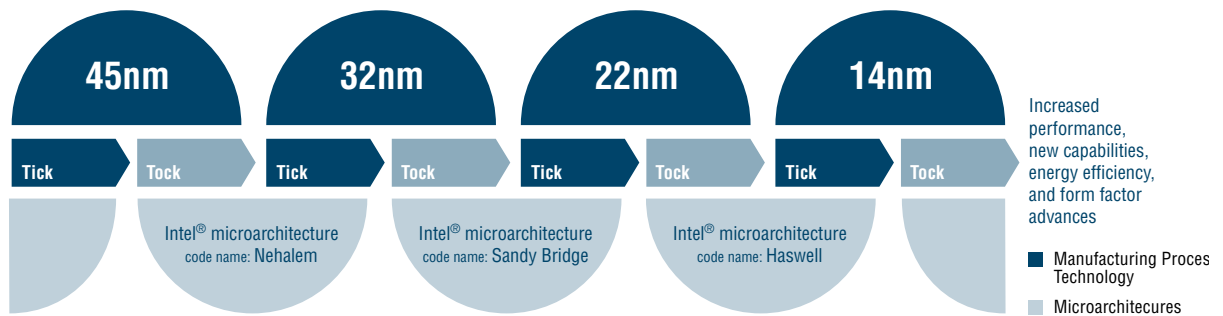
The Tick-Tock Mobel Through the Years

Figure 1. Illustration of Intel's "Tick-Tock" Model
[Reference: http://www.intel.com/content/www/us/en/silicon-innovations/intel-tick-tock-model-general.html]

Pervasive Virtualization

As service providers look to achieve further improvements in CAPEX and OPEX, they are starting to widely use virtualization techniques in their networking equipment. Virtualization allows traditional, fixed-function physical network appliances to be replaced by virtual appliances (VAs), which are pure software running on standard server blades. This enables processing resources to be allocated dynamically in line with specific traffic patterns, whether to manage the allocation of processor cores for control plane versus data plane functions, or to reconfigure the specific functions being performed by the VA (for example, a firewall function versus WAN optimization), or both.

Virtualization has long been used to manage and optimize the allocation of application workloads within data centers. Its adoption within the networking subsystems of mobile and fixed networks enables service providers to optimize both CAPEX and OPEX, while guaranteeing full network scalability in anticipation of the next "killer app" whose resource requirements are unknown today.

Intel® Architecture for Data Plane Processing

Packet processing performance has improved significantly on Intel® processor-based platforms due to the combination of software advances and Intel® microarchitecture enhancements. From the software perspective, the availability of high-performance packet processing software, such as 6WIND's 6WINDGate software, is a game changer. It is optimized for maximum throughput, integrating the Intel DPDK software libraries designed to dramatically reduce packet processing latency. With respect to microarchitecture, Intel refreshes its processors on a regular "Tick-Tock" cadence, where each new generation provides significant improvements—all while enabling developers to run their existing application software and tool suites.

At a constant beat-rate, Intel's "Tick-Tock" model predictably provides increased performance and energy efficiency, and new capabilities with each processor generation. Figure 1 illustrates the Tick-Tock model beginning with the Intel microarchitecture codenamed Nehalem on a 45nm manufacturing process, up through the Intel microarchitecture codenamed Haswell on a 14nm process. Each "tick" of the Tick-Tock model represents improvements in the manufacturing process, leading to greater transistor density for the previous "tock" microarchitecture. Alternating "tock" cycles use the previous "tick" cycle's manufacturing process technologies to introduce new capabilities in processor microarchitecture, such as hardware-assisted video transcoding and encryption/decryption. With both tick and tock cycles, Intel seeks to continually improve energy efficiency and performance.

TCO concerns are leading customers to seek a single architecture design from top-to-bottom. Intel addresses TCO and time-to-market (TTM) issues with a single architecture capable of running multiple workloads. Platforms based on Intel architecture processors and commercially available high-performance packet processing software provide the capability to consolidate application processing, control processing and data plane processing workloads onto a single platform.

This strategy is also allied to Intel's predictable Tick-Tock model and the Intel DPDK, an optimized data plane software solution developed to help unleash the full potential of an Intel architecture platform. The Intel DPDK is a set of optimized software libraries and drivers (Figure 2) that enable high-performance data plane packet processing on Intel architecture processors.
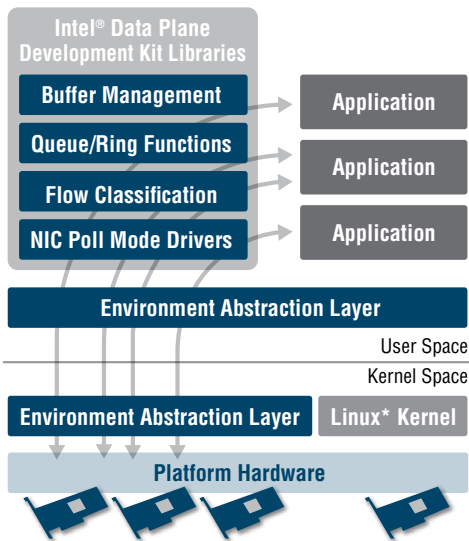
Figure 2. The Intel® Data Plane Development Kit (Intel® DPDK)

The Intel DPDK consists of the following components:

- **Memory/Buffer Manager** allocates NUMA-aware pools of objects in memory; the pools are created in huge page memory space and use a ring to store free objects. The Intel DPDK pre-allocates fixed-size buffers, which are stored in the memory pools. The manager also provides an alignment helper to ensure that objects are distributed evenly across all DRAM channels, thus balancing memory bandwidth utilization across the channels. Moreover, the manager significantly reduces the amount of time the operating system must spend allocating and de-allocating buffers.

- **Queue Manager** implements safe lockless queues, instead of using spinlocks, that allow different software components to process packets, while avoiding unnecessary wait times.

- **Flow Classification** is an efficient mechanism for generating a hash (based on tuple information) used to quickly combine packets into flows, which enables faster processing and greater throughput. The mechanism incorporates Intel® Streaming SIMD Extensions (Intel® SSE) to increase the parallelism of the packet processing.

- **Poll Mode Drivers** greatly speed up the packet pipeline for 1 GbE and 10 GbE Ethernet controllers by receiving and transmitting packets without the use of asynchronous, interrupt-based signaling mechanisms, which have a lot of overhead.

- **Environment Abstraction Layer (EAL)** provides an abstraction to platform-specific initialization code, which eases application porting effort. The 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, including the run-time libraries for launching and managing Intel DPDK software threads.

Extending Intel® Data Plane Development Kit Features Via 6WIND Enhancements

6WIND has developed a number of value-added enhancements to the Intel DPDK library that provide increased system functionality and performance compared to the baseline software. These value-added enhancements (Figure 3) include:

- High-performance software crypto support, implemented via the Intel® Advanced Encryption Standard New Instructions (Intel® AES-NI)[1] in the Intel® Xeon® processor E5600 series and E5-2600 series.

- Device monitoring and statistics functions, such as Linux* Ethtool MTU support, full RX/TX queue statistics and CRC error statistics, which enable improved system-level profiling, analysis and debug.
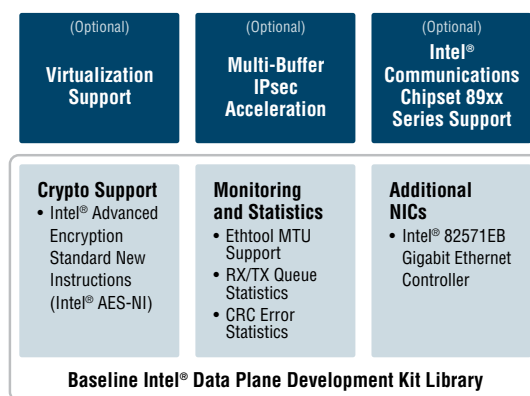

Figure 3. Value-added Enhancements

- Support for additional Network Interface Cards (NICs), such as the Intel® 82571EB Gigabit Ethernet Controller, beyond those supported in the baseline Intel DPDK library.

6WIND also provides a range of optional add-on extensions to the Intel DPDK designed to improve the cost/performance of both physical and virtual networking appliances while enabling the use of the Intel DPDK in software-defined networks. These optional add-ons include:

- IPsec acceleration, achieved through integration of the Intel® Multi-Buffer Crypto for IPSec library;

- Crypto acceleration via support of an external accelerator, the Intel® Communications Chipset 89xx series (codenamed "Cave Creek"), which is part of Intel's next-generation communications platform, codenamed "Crystal Forest".

- Virualization-related enhancements (Figure 4) that maximize system performance by removing key I/O and communication bottlenecks include:

1. **I/O Virtualization (IOv),** an industry-standard approach for increasing the performance of virtual network appliances by bypassing the virtual switch within the hypervisor, thus removing the I/O performance constraints imposed by the virtual switch.
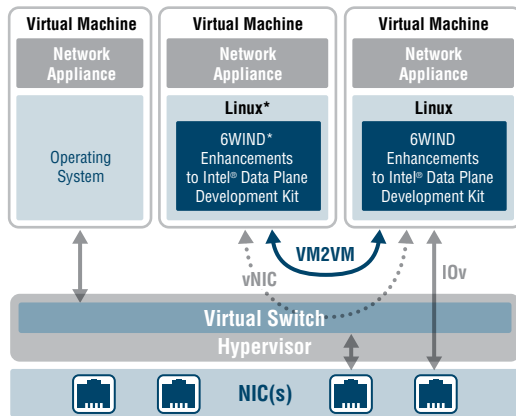
Figure 4. Virtualization Enhancements

2. A **virtual NIC (vNIC)** driver that leverages communication between virtual machines (VMs) via the virtual switch, enabling the efficient development and provisioning of systems with multiple VMs and significant East-West network traffic.

3. For systems that require the ultimate level of performance for East-West traffic between VMs, a **VM-to-VM (VM2VM)** driver enables direct VM-to-VM communication, bypassing the virtual switch while remaining fully compatible with industry-standard hypervisors.

These Intel DPDK enhancements and optional add-ons are maintained by 6WIND as private branch, regularly synchronized with Intel's on-going releases of the baseline library. They are delivered to customers either as a stand-alone library or, for applications that also require high-performance packet processing software, are integrated within the 6WINDGate software solution as described in the next section.

Maximizing Packet Processing Capabilities Via 6WINDGate* Integration

Intel Xeon processors provide an ideal platform for implementing the high-performance packet processing required for converged networking equipment. However, for developers of such equipment, it is critical to implement a software solution that best utilizes this platform in order to deliver the maximum possible networking performance at the system level.

A standard networking stack uses services provided by the operating system (OS) and is subject to significant overheads associated with functions such as preemptions, threads, timers and locking. These processing overheads are imposed on each packet passing through the system, resulting in a major performance penalty for overall throughput. Furthermore, although some symmetric multiprocessing (SMP) improvements can be made to an OS stack to support multicore architectures, performance fails to scale linearly over multiple cores for complex packet processing such as required by converged networking

equipment. For example, a processor with eight cores may not process packets significantly faster than one with two cores. Generally, a standard OS stack does a poor job of exploiting the potential packet processing performance of a multicore processor.

The 6WINDGate packet processing software solves this problem through a fast path-based architecture, while incorporating a comprehensive set of high-performance networking protocols fully optimized for Intel Xeon processor-based platforms.

With reference to Figure 5, 6WINDGate includes the following features:

1. Optimized Architecture for Packet Processing

The vast majority of packets are processed in a fast path environment, executing outside the OS kernel in a Linux userspace environment. By avoiding typical OS overheads associated with functions, such as preemptions, threads, timers and locking, this architecture maximizes data plane processing performance. Only those rare packets that require complex processing are forwarded to the OS networking stack, which performs the necessary management, signaling and control functions. Most of the processor cores can be dedicated to running the fast path in order to maximize the overall throughput of the system, while at minimum, only a single core is required to run the OS, the OS networking stack and the application's control plane. The fast path cores run in a Linux userspace environment, so the system can be reconfigured dynamically as traffic patterns change in order to optimally allocate CPU resources to the control plane and the fast path.
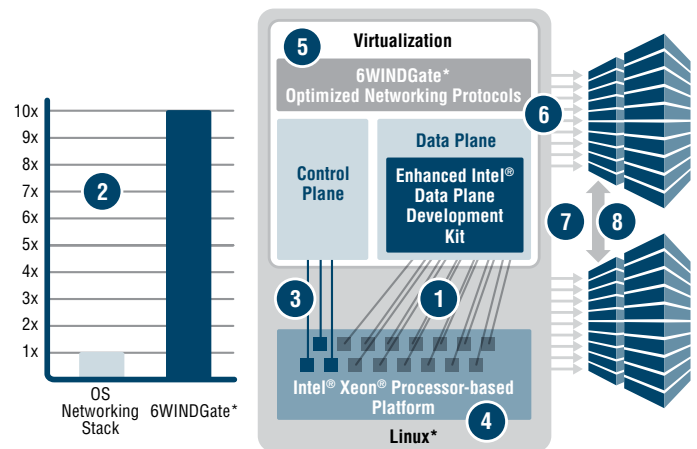

Figure 5. Features of an Intel® Xeon® Processor-based Platform Running 6WINDGate* Packet Processing Software

Splitting the networking stack in this way has no impact on the functionality of application software, which interfaces to the same OS networking stack as previously. Existing applications do not need to be rewritten or recertified, but they run significantly faster because the underlying packet processing is accelerated through the fast path environment.

2. Maximum Networking Performance

In a typical converged networking solution, such as a security gateway, 6WINDGate will deliver seven to ten times the performance of a system based on a standard OS networking stack, thanks to the fast path implementation and advanced, architecture-specific optimizations. This is described in more detail in the following.

3. Architecture Optimizations

6WINDGate has been optimized to benefit from the unique architectural features of Intel Xeon processor-based platforms, while fully exploiting the Intel DPDK data plane library, as described earlier. The enhanced Intel DPDK library is delivered pre-integrated within 6WINDGate, so there is no need for licensees themselves to perform this integration.

6WINDGate provides full support for heterogeneous systems in which the control plane runs on a different processor architecture from the data plane (for example, the control plane on an Intel Xeon processor and the data plane on a processor based on a different instruction set architecture (ISA)).
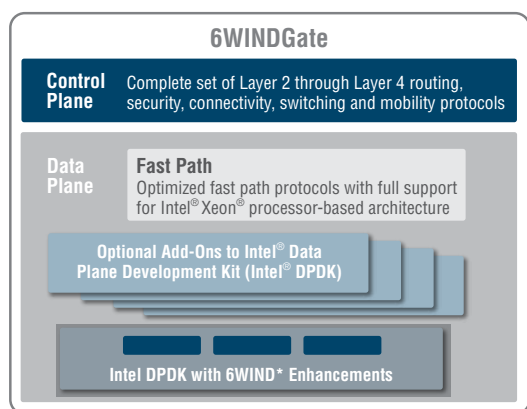

Figure 6. 6WINDGate* Pre-Integrates the Intel® Data Plane Development Kit

4. Full Linux* Compatibility

6WINDGate is fully compatible with Linux networking APIs, so standard Linux application software can be deployed on systems running 6WINDGate with no need for changes. In addition, 6WINDGate supports Linux distributions from commercial suppliers, 6WIND's processor partners and kernel.org.

5. Optimized for Virtualization

6WINDGate supports open-source and proprietary hypervisors, using the advanced techniques described previously to maximize I/O bandwidth in a virtualized environment.

6. Comprehensive Networking Protocol Suite

The 6WINDGate packet processing software comprises a comprehensive set of networking protocols optimized for converged networking equipment, such as crypto acceleration, firewall, GRE, GTP, IPv4, IPv6, IPsec, Layer 2 switch, MPLS, NAT, QoS, SCTP, SSL termination, TCP termination, UDP, virtual routing, VLAN and others.

7. Scalability

For high-end networking equipment, 6WINDGate is fully scalable across processors, blades and racks. Designers can configure 6WINDGate to run on as many cores as necessary in order to achieve the required level of system performance, without any performance penalty for the distributed configuration.

8. Carrier Grade Reliability

For systems where five-nines or zero-downtime reliability is required, 6WINDGate provides full support for high availability (HA) frameworks in order to ensure reliability through industry-standard failover modes. 6WINDGate supports active-standby control plane redundancy via HA daemons that are responsible for synchronizing control plane states.

By delivering 6WINDGate as an integrated product that includes the enhanced version of the Intel DPDK library, 6WIND minimizes the development time and schedule risk for high-end networking equipment since there is no need for licensees to perform any of the integration themselves. In fact, the presence of Intel DPDK is completely transparent to users of the integrated solution because in this configuration, the Intel DPDK exists as an internal component of 6WINDGate, and it does not need to be directly accessed by application-level software.

Use Case Example: LTE Security Gateway

While service providers worldwide are migrating their networks to the LTE standard, many are also implementing wireless offload schemes to route traffic over the Internet in order to provide increased coverage for subscribers while reducing the overall load on the wireless network.
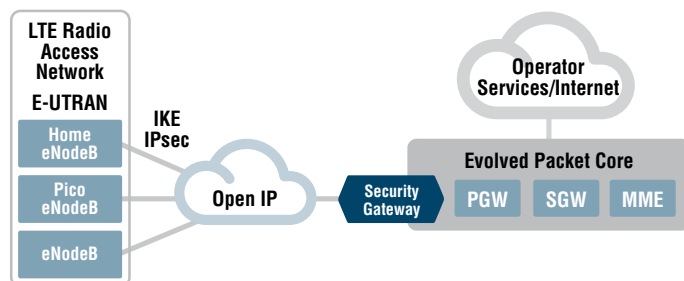

Figure 7. Wireless Offload Example

Both of these approaches leave wireless traffic more vulnerable to security threats and breaches than previous wireless network architectures. Historically, dedicated leased lines (T1/E1/J1) were used for the backhaul connections between cell sites and the core network. Conversely, LTE uses all-IP connections running over commercial broadband links. These connections are inherently more vulnerable to attacks by hackers and traffic spoofing. The flat LTE topology increases the overall security risk by providing a direct connection from cell sites to the core network.

In the wireless offload scenario (Figure 7), user traffic from WiFi access points, picocells and femtocells is connected to the core network via public Internet connections. This potentially exposes the core network to a wide variety of Internet-based attacks, so security gateways with firewalls should be deployed to provide a critical layer of protection.

LTE networking equipment must be designed to resist these complex security challenges while providing the high data rates necessary to support the growth in mobile traffic bandwidth. Given the explosion in mobile Internet traffic, the introduction of complex, advanced services and the CAPEX pressures faced by service providers, OEMs are under pressure to increase the functionality of their gateways while at the same time achieving significant improvements in cost-performance and minimizing the time-to-market for each new product generation.

In terms of the system software, the networking protocols available within **6WINDGate Mobile Edition** include a complete, integrated solution for an LTE security gateway. As shown in Figure 8, 6WINDGate integrates routing, security and mobility features that include:

- Control plane protocols: BGP, OSPF, RIP (all protocols are virtual routing-aware), DHCP, EAP, IKEv1/v2-VR, MOBIKE, Mobile IP and Radius.

- Fast path protocols: Firewall, IPv4, IPsec, QoS, Mobile IP, statistics and virtual routing (all features are IPv6-ready).

- High availability (HA) support for carrier grade security systems with five-nines or zero-downtime reliability requirements, maintaining IPsec tables in both active and standby control planes, which avoids the need to reestablish IKE control plane sessions following subsystem failures.

The high-capacity 6WINDGate IKE solution maximizes the number of access points supported by the gateway. It also includes the capability to manage an IKE instance per a virtual router.

By integrating a complete set of protocols for the control plane and the fast path, 6WINDGate provides developers with a single-vendor solution for a high-performance security gateway based on Intel Xeon processors. By removing the need for developers to integrate networking software components from multiple suppliers, 6WINDGate has been proven to accelerate their time-to-market by up to twelve months.

The following analysis of the system-level performance that is achievable with this solution used a reference platform comprising a platform based on **dual Intel Xeon processor E5-2600 series (codenamed "Crown Pass") running at 2.7 GHz**.

On this platform, the 6WINDGate solution delivers approximately **14 Mpps per core** of IP forwarding performance, while the architecture of the software ensures the performance scales linearly according to the number of cores configured to run the fast path (subject, of course, to any finite throughput limits imposed by hardware constraints).
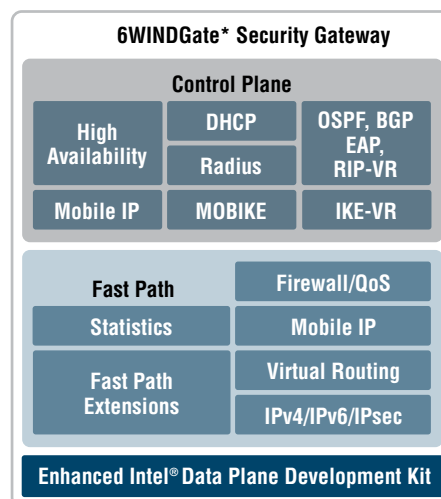

Figure 8. Security Gateway Example

6WINDGate can also process:

- **5 Gbps of encrypted IPsec traffic per core** (512 byte packets) using the Intel AES-NI instruction set and the 6WINDGate IPsec fast path module.

- **80,000 IPsec tunnels or 80,000 IKE sessions** with an establishment rate of 160 to 1,000 tunnels/sessions per second depending on the authentication methods. The number of IPSec tunnels or IKE sessions can be increased if hardware accelerators are used for IKEv2 authentication phases.

These results can be combined to estimate the performance of an IPsec security gateway using a Linux userspace implementation that dynamically allocates cores either to the control plane or the fast path.

Such a system benefits significantly from dynamic allocation of CPU resources. While IPSec tunnels or IKE sessions are initially being established, the overall traffic is low and CPU resources are allocated to accelerate the IKE negotiation process, which is a control plane function. However, once the sessions are established, the load on the control plane is diminished, and CPU resources are reallocated to maximize the fast path performance required for processing the encrypted traffic.

Using this flexible architecture, this gateway can achieve the following performance:

- **80,000 IPsec tunnels or 80,000 IKE sessions** with an establishment rate of 160 to 1,000 tunnels/sessions per second depending on the authentication methods.

- **Fast path performance of 70 Gbps** (512 byte packets). This assumes that 14 cores are dedicated to the fast path and the remaining two cores are used for the control plane.

Assuming a single LTE subscriber uses an average bandwidth of 1.2 Mbps, a board based on the reference platform can cipher the traffic

Ten data plane blades run the 6WINDGate* data plane (with integrated Intel® Data Plane Development Kit software) to maximize IPsec processing performance

Two redundant control plane blades ensure zero system downtime via 6WINDGate's High Availability features
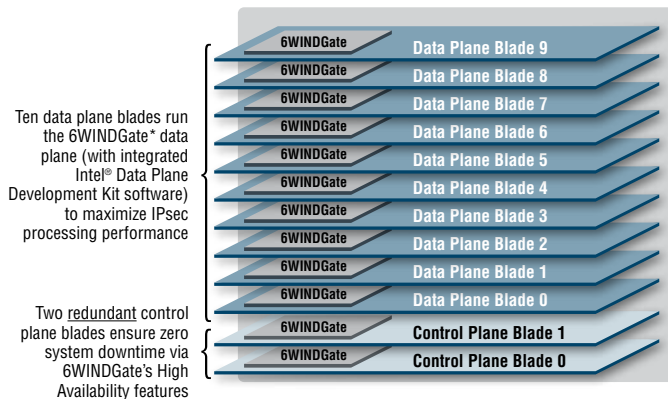
Figure 9. Partitioning the Blades of a Security Gateway

for more than 58,000 users. It can also manage a very large number of eNodeBs and Home eNodeBs (HeNBs) as only a few IPSec tunnels or IKE sessions have to be established in each eNodeB or HeNB.

Considering a chassis populated with 12 blades, each based on the reference platform described previously, where ten blades are dedicated to the fast path and the remaining two blades run a high availability control plane, the following performance can be achieved:

- **80,000 IPsec tunnels or 80,000 IKE sessions with a fully redundant architecture.**

- **Fast path performance of 700 Gbps.**

This configuration can cipher the traffic from more than **580,000 LTE users** connected through several tens of thousands of eNodeBs or HeNBs.
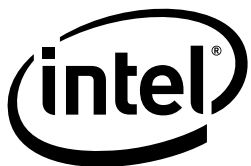
Summary of Benefits for Intel® Data Plane Development Kit -plus-6WINDGate* in High-Performance Networking

The 6WINDGate packet processing software comprises a comprehensive set of networking protocols optimized for converged networking equipment. The extensive protocol support in 6WINDGate's Mobile Edition, combined with the performance benefits and maximized resource utilization from dynamic allocation of CPU resources, provide a complete, integrated solution for 4G networks.

For high-end networking equipment, 6WINDGate is fully scalable across processors, blades and racks. Designers can configure 6WINDGate to run on as many cores as necessary in order to achieve the required level of system performance, without any performance penalty for the distributed configuration.

Intel's "Tick-Tock" model provides a constant beat-rate of improvements to its core microarchitecture and industry-leading process technologies. This, along with continued software development and enabling, ensures performance scales with the Intel architecture roadmap.

A unified system architecture, in which the control plane and data plane run on the same processor architecture while achieving the necessary cost-performance targets, is preferable to a hybrid approach for several reasons. It greatly simplifies software development as well as debug and integration. Processor resource utilization is improved because the control plane and data plane can be distributed among cores with greater flexibility. A common code base and single development environment offer minimized development schedule risk and make software maintenance much easier. These benefits can be achieved with Intel's next-generation communications platform, codenamed "Crystal Forest", which provides the networking performance necessary to serve as a unified platform for converged networking equipment, especially when combined with high-performance packet processing software such as 6WIND's 6WINDGate software and the Intel DPDK. Intel architecture platforms, combined with commercially available high-performance packet processing software, provide the capability to consolidate application processing, control processing and data plane processing workloads onto a single platform.

www.intel.com/go/commsinfrastructure

www.6wind.com