Intel® Architecture at the Edge for Greater Flexibility and Scalability

Consolidate services and packet processing workloads on a single processor architecture to improve subscriber experiences, reduce network complexity, and decrease capital and operating expenditures.

More than ever, service providers need ways to efficiently scale the network to deliver high-quality mobile and connected experiences, as well as new revenue-boosting services, while tightly managing capital and operating expenditures.

With the escalating growth in connected devices, mobile users, and video content, Internet traffic is exploding. More than ever, service providers need ways to efficiently scale the network to deliver high-quality mobile and connected experiences, as well as new revenue-boosting services, while tightly managing capital and operating expenditures (CapEx and OpEx). Fortunately, there's a flexible and scalable approach to resolving these challenges: deploying intelligent network components based on Intel® architecture at the network edge.

Adding Intelligence to the Edge

Typical network elements are based on several different processor architectures. When it’s time to add services, application-specific appliances are added, along with corresponding load balancers, switches, and backup systems. The resulting heterogeneous fixed-function infrastructure is hard to power manage, and costly to scale and enhance with new services.

An intelligent edge infrastructure based on Intel architecture (Figure 1) offers a better approach, allowing consolidation of high-value services and packet processing workloads onto a single processor architecture positioned closer to the subscriber. Bringing compute and storage capabilities to the edge allows the use of virtualization solutions that take advantage of Intel® Virtualization Technology (Intel® VT), an integrated set of hardware enhancements for virtualization. Virtualization adds astonishing flexibility to service edge solutions, enabling them to re-provision cores, blades, and systems as needed to handle changing service demands. Services can even be live migrated from node to node with minimal service interruption in response to real-time changes in demand.

With one blade capable of handling several types of workloads, Intel® processor-based servers provide a clear path to reducing CapEx costs through consolidation. They also help reduce OpEx through advanced power management features. Equally important, Intel processor-based servers at the service edge can substantially improve subscriber quality of experience. For example, delivering high-demand video content seamlessly from the edge closest to the subscribers reduces latency and jitter. What’s more, these systems provide the processing power and storage necessary to offer new managed services that create additional revenue streams.

Figure 1. Placing intelligent services at the consumer edge can improve quality of experience while reducing capital and operating expenditures.
Demonstrating the Possibilities

Intel developed an intelligent service edge proof of concept (PoC) to demonstrate how Intel® Xeon® processor-based systems positioned at the network edge can streamline the delivery of video or other priority content. The usage model for the PoC is a video-on-demand service. This type of service streams video to client devices such as smart phones, tablets, netbooks, and laptops, and is typically hosted in the core of the network. To guarantee a certain level of performance, a Quality of Service (QoS) solution assigns different priorities to different data flows. To demonstrate the on-demand scaling and power management capabilities of an Intel architecture-based solution, the Intel PoC packages a video-streaming server in a virtual machine (VM) and allows migration between the network core server and the intelligent service edge server.

Demonstration Setup

Figure 2 shows the demonstration setup.

- **Video load generator.** The rack-mount server simulates a large number of client devices by requesting thousands of video streams. The load is adjustable between low, moderate, and peak demand using a control on the demo user interface.
- **Traffic generator.** In addition to the video load generator, a traffic generator feeds the converged blade with 10 gigabits per second (Gbps) of “best-effort” traffic—data streams that compete for bandwidth with priority video streams.
- **Converged blade and service blade.** In the PoC, two servers represent the intelligent service edge. In a real implementation, these two servers would co-exist in a bladed architecture such as Advanced Telecom Computing Architecture (AdvancedTCA®). The first server, the converged blade, implements routing, deep packet inspection (DPI), and QoS. At medium demand, it also hosts the video service VM. Handling these workloads requires a high-performance multi-core processor, high-speed multi-core packet processing software, and hardware-assisted virtualization (Figure 3). The second server, the service blade, provides additional capacity for running network services in VMs if the converged blade cannot cope with demand. The service blade is powered on at peak demand and powered off when demand decreases and this capacity is no longer required—thereby reducing OpEx.

![Figure 2. The intelligent service edge combines converged and service blades.](image-url)
**Network core server.** Implemented using a rack-mount server, the network core server represents the network data center and hosts services not running at the edge of the network, such as the best-effort video workload. The network core server also hosts the priority video workload when it’s not hosted at the network edge.

**Priority video workload.** The priority video workload is a VM that is migrated across the network core server, the converged blade, and the service blade depending on demand.

**Best-effort video workload.** The best-effort video workload consists of video traffic treated with lower priority than the high-priority video. The best-effort video workload remains in the network core.

**Demo user interface client.** The demo user interface client runs the PoC management software, providing a display showing real-time traffic and workload information, and a control to adjust video workload demand. The display (see Figure 4) includes:
- Throughput of prioritized and best-effort traffic into and out of the converged blade in Gbps
- The location of the priority video workload
- Power consumption of the intelligent service edge node in watts
- Rate of DPI in Gbps

### Demonstration Scenarios

The PoC shows the intelligent service edge node’s ability to apply power management and maintain QoS for prioritized traffic while offloading the core network in response to increasing demand. In all scenarios, best-effort traffic is throttled back to avoid degrading prioritized traffic. QoS is implemented using features in 6WINDGate* multi-core packet processing software. Power consumed by the intelligent service edge node is monitored and displayed in the demo user interface.

#### Light Demand

At the start of the demo, there is light demand for the priority video service, which is in a VM running on the network core server. Video traffic is routed across the network core and through the converged blade, which acts as a router and implements QoS. The service blade is powered down to conserve power at the network edge. The DPI function runs on one core of the converged blade and uses a set of rules to scan packets for potential security threats. For the purposes of this PoC, the rule set is updated every second to show a dynamic configuration corresponding to changing network conditions and threat levels. The DPI function’s packet processing rate, displayed on the demo user interface, changes accordingly.
Moderate Demand
When the video demand control (see Figure 4) is moved to moderate load, it triggers an increase in video traffic and migration of the VM running the service from the network core server to the converged blade. The service blade remains offline. The priority video is now served from the intelligent service edge, closer to the demand, removing load and traffic from the core network server in the data center. This reduces latency and jitter.

Peak Demand
When the video demand control is moved to peak load, it prompts the service blade to power up and come online. The VM running the priority video service migrates from the converged blade to the service blade to take advantage of extra processing capacity.

Return to Moderate Demand
When demand for video decreases (video load is "dialed down"), the VM running the priority video service migrates from the service blade back to the converged blade. The service blade returns to a sleep state to conserve power.

Return to Low Demand
As demand decreases further, the VM running the video service migrates back to the data center. This further decreases power usage at the intelligent service edge node. The converged blade continues to function as a router, implementing QoS and running the DPI function.

Key Technologies

Intel® Xeon® Processor 5600 Series
Providing up to 12 high-performance cores per two-socket server, the Intel® Xeon® processor 5600 series delivers exceptional performance for packet processing and running multiple VMs simultaneously. To conserve energy, Intel Xeon processors can regulate power consumption and intelligently adjust server performance according to workload needs.

Kernel-based Virtual Machine (KVM) Virtualization
KVM is an open-source full virtualization solution for Linux*. KVM enables multiple VMs running unmodified Linux or Microsoft Windows* images. Each VM has private virtualized hardware, including a network card, disk, graphics adapter, and more.

6WINDGate* Packet Processing Software
The 6WINDGate stack from software provider 6WIND splits the networking
stack into two layers. To enable wire-speed packet-processing performance, the lower layer, typically called the fast path, processes the majority of incoming packets outside the OS environment without incurring any of the OS overheads that degrade overall performance. Only those packets that require exception processing are forwarded to the OS networking stack. Using 6WINDGate removes the complexity of integrating high-performance packet processing with the Linux environment by fully synchronizing the fast path and OS while preserving Linux APIs. 6WINDGate’s fast path modules include high-level APIs to interface with hardware features, such as queue management for QoS, to more effectively differentiate and manage best effort and priority traffic.

**Intel® Data Plane Development Kit (Intel® DPDK)**
To further increase packet processing and scaling performance, solutions like 6WINDGate use the services and optimizations of a dedicated multi-core “executive” like Intel® DPDK that enables optimized packet processing outside the OS environment. Using Intel DPDK optimizations, 6WINDGate can provide up to 10x the packet processing performance of a standard networking stack. Intel DPDK services include specific I/O drivers for network interfaces; crypto-engines and similar low-level software to distribute packets between cores; memory management; and tools to compile software in this environment.

**Sensory Networks HyperScan**
HyperScan is a pattern-matching engine that performs regular expression (regex) matching and DPI—a security measure that can consume large compute resources. While the PoC only uses a small rule set for simplicity, HyperScan scales linearly and provides exceptional performance, supporting millions of simultaneous patterns and matches concurrently when run on multi-core Intel Xeon processors.

**Results**

**Flexibility**
The PoC shows the ability of an Intel Xeon processor-based system positioned at the edge to handle all the packet processing workloads involved in providing video and other services. It also shows how the video workload, contained in its own VM, can be seamlessly transferred from the network core server to the intelligent service edge’s converged blade or, in periods of high demand, its service blade.

**Scalability**
The PoC shows that a number of different types of workloads can be handled simultaneously on an Intel architecture-based system. This makes it much easier to scale the system versus a multi-appliance system with fixed-function elements.

**Power Management**
The intelligent service edge successfully demonstrates the ability to conserve power by migrating workloads and powering down compute elements when demand drops.

**Conclusion**
Overall, the PoC demonstrates that deploying Intel processor-based systems at the service edge gives providers a cost-effective way to improve subscriber quality of experience, as well as add processing capacity for new services that could provide additional revenue streams. These advantages will continue to grow as Intel's commitment to deliver ever-increasing performance per watt produces processors with even greater capabilities in the future, helping service providers scale and manage their infrastructure to keep up with demand.

**Learn More**
See the white paper “Consolidating Communications and Networking Workloads onto One Architecture.”
### Hardware

<table>
<thead>
<tr>
<th>Component</th>
<th>Detail Configuration</th>
</tr>
</thead>
</table>
| Converged Blade, Service Blade, and Network Core Servers | • **2U rack-mount server.** Intel® Server System SR2600URBRPR with S5520UR motherboard; dual processor Intel® Xeon® processor X5690, 12 MB cache, 3.46 GHz, 6.40 GT/s, Intel® QuickPath Interface.  
• **Memory:** 12 GB (6x 2 GB DDR3-1333 ECC registered).  
• **Storage:** 3.5" 500 GB SATA hard disks.  
• **Network Interface:** Niantic 10 Gbps dual-port NIC*, Intel® EN Server Adapter X520-DA2 (SFP+ direct-attach copper cables). |
| Video Load Server | • **1U rack-mount server.** Dual processor, Intel® Xeon processor E5520.  
• **Memory:** 4 GB DDR3-1067 ECC registered.  
• **Storage:** 3.5" 500 GB SATA hard disk.  
• **Network Interface:** Niantic 10Gbps dual-port NIC*, Intel® EN Server Adapter X520-DA2 (SFP+ direct-attach copper cables). |
| Traffic Generator | • Ixia 400T* 4U traffic generator, equipped with two LSM10G1 10 Gbps load modules and IxExplorer* software. |
| Ethernet Switch | • Dell PowerConnect* 10 Gbps switch 24 ports, 1U. |
| Power Meter | • Watts Up? Pro* power meter. |

### Software

<table>
<thead>
<tr>
<th>Function</th>
<th>Product / Version</th>
</tr>
</thead>
</table>
| Operating System | • **Service blade, network core server, and video load generator:** Fedora Core 14* (kernel 2.6.35.13-92.fc14.x86_64).  
• **Converged blade:** 6WIND stack with Linux* kernel 2.6.34.  
• **Virtual machines:** CentOS 5.6*, with kernel 2.6.18-238.19.1.el5. |
| Networking Stack with QoS | • 6WINDGate 3.27.1 running on the converged blade. 6WINDGate is a commercial, high-performance, low-latency data plane solution. |
| Intel® DPDK | • The Intel® Data Plane Development Kit (Intel® DPDK) is still in the early access phase and is only available to select customers and ecosystem partners for evaluation and feedback. |
| Video Load Generator | • Built using Apache Benchmark**--a tool for benchmarking Hypertext Transfer Protocol (HTTP) servers. |
| Video Server | • Runs in a virtual machine and uses the video-on-demand and HTTP features of VLC Media Player*. |
| DPI (Deep Packet Inspection) | • Sensory Networks HyperScan version 2.1*. |
| Media Player | • VLC Media Player is an open-source, cross-platform multimedia player and framework that plays most multimedia files as well as DVD, audio CD, VCD, and various streaming protocols. |
| Video Clients | • Several laptop and tablet clients were used to display the best-effort and priority video streams. |

Click here for more information on Intel® Xeon® processor 5600 series.