White Paper

Industry

Secure Services Application

Efficient Encryption to Optimize HTTPS Termination on Centralized and Edge Footprints with F5® and Intel® QuickAssist Technology (QAT)

There is a shift on the Internet today moving towards secure communication. With 85%-95%[1] of HTTPS internet traffic encrypted by browsers, the routine processing of SSL traffic, both simple SSL Offload and SSL Orchestration (decryption/inspection/re-encryption process), continuously drains valuable power and compute resources every month. To combat this situation, Intel and F5 initiated a proof of concept to test COTS hardware with Intel's QAT and F5's BIG-IP® VE for virtualized applications which delivered up to 86%[2] TCO savings with an improved performance similar to more costly dedicated hardware options.

This whitepaper will demonstrate how deploying this alternative configuration is particularly useful for private data centers seeking to maximize efficiency and reduce monthly power costs.

Contents

Authors

Felipe Pastor
Technical Specialist

Ai Bee Lim
Solution Architect

Tim Miskell
Solution Architect

Jesse Driskill
Solution Engineer III

Matt Burns
Solution Engineer III

John Bertoni
Major Account Manager

John Touart
Major Account Manager

Chris Meredith
Sr. Solution Engineer

Market Trends Impact on Security

The rapid volumetric explosion of online workloads to meet years of pandemic needs (higher bandwidth networks to support the exponential increase in the number of internet connected devices across business, home and personal use) are stressing common Central Processing Units (CPUs) in edge networking devices. The unprecedented demands on wire-speed data encryption and decryption, low latency and automation are not scaling proportionately, resulting in redundant processes which choke operational and cost efficiency. The impact is especially critical when power and CPU utilization can be a large portion of Total Cost of Operations (TCO), and continued hardware supply chain issues may be delaying upgrades, forcing software to work harder. Additionally, the increasing number of outside attacks on web applications only reinforces the need to retain a robust efficient security posture to protect applications and content, including for new HTTP/2 protocols and SSL/TLS encryption at Layer 7.

1. HTTPS encryption on the web https://transparencyreport.google.com/https/overview?hl=en
2. See below for workload details and configurations. Results may vary

F5 BIG-IP Virtual Edition (VE) at the Edge

Bolstered by decades of expertise in traffic management, F5 is a leader in multi-cloud application services, secure network and infrastructure supporting functions for enterprises and Communications Service Providers (CSPs). F5 application services deliver resilience so that applications are always optimized, secure and scale any environment and on any device. Designed for similar performance as their respected BIG-IP appliances, F5 BIG-IP Virtual Edition (VE) software products are optimized for virtualized environments and to run on Commercial off-the-shelf (COTS) servers powered by Intel architecture CPUs at an overall lower cost.

F5 solutions address multiple enterprise and service provider use cases for network and application environments in both core and edge/data centers including:

- Web Application Firewall (WAF)
- Protocol fluency and tunneling capabilities such as TCP, HTTP/2, Diameter, and IPsec
- Identity and access management
- IoT security
- Intelligent traffic management, load balancing, and DNS services
- SGi-LAN/N6 consolidation (policy enforcement, firewall, CGNAT, DDoS)

To meet high traffic demands, improve speed and responsiveness, edge expansion is an ideal greenfield use case to virtualize the network and build data center efficiencies delivering greater geographic coverage and better customer experience. However, many edge servers are optimized first for cost, and use lower performance CPUs which can reduce the amount of computing power available for encryption, impacting the overall security posture. Additionally, when CPU cores must be dedicated to crypto workloads, fewer compute cycles are available for other critical functions such as networking and application workloads.

The performance and CPU utilization of application workloads can be significantly improved by using dedicated hardware accelerators, such as Intel QAT, to compress and offload cryptographic workloads from the CPU to accelerate processing. This in turn enables increased encryption workloads thus improving virtual network function (VNF) performance, which subsequently directly positively impacts transaction volume.

To demonstrate the increased performance of encryption acceleration, F5 tested the performance of its BIG-IP VE software application delivery controller (ADC) virtual network functions (VNFs) on an Intel® architecture-based edge server utilizing Intel QAT. Network administrators may not be aware that QAT accelerating technology is ALREADY built into most Intel-powered COTS edge hardware, such as X86 servers.

The majority of encryption in today's traffic is SSL/TLS, so improving efficiency here will have a very significant impact on capacity. Further, security-conscious organizations want to avoid clear-text traffic whenever possible, which leads to workloads being decrypted for processing, then re-encrypted before sending the traffic to the next service. These decryption/re-encryption patterns not only double the number of cryptographic operations when compared to SSL Offload, they can also become quite a bit more complicated as the traffic frequently must be decrypted and re-encrypted at multiple points. Orchestrating all that decryption and re-encryption, while ensuring that security policies are adhered to is a significant challenge. SSL Orchestration (SSLO) is the process of decrypting, inspecting, and re-encrypting incoming traffic before sending it along for further processing. And providing security devices with visibility into SSL/TLS encrypted traffic is a very common use-case. SSLO supports policy-based management and steering of traffic flows to existing security devices and centralizes the SSL decrypt/encrypt function by delivering the latest SSL encryption technologies across the entire security infrastructure. Because of this SSL service chaining, this process has a higher compute requirement than basic SSL Offload—and these types of services were traditionally performed on dedicated hardware platforms.

F5 BIG-IP VE Supports both SSL Offload and SSL Orchestration Use Cases with QAT delivering a 10x improvement:

Offload: Traffic is decrypted and forwarded without additional consideration;

Orchestration: Traffic is decrypted, then sent to additional service-chain elements for inspection, then re-encrypted (potentially many times) before being forwarded to the final destination.

An Encryption Workload Perfect for QAT

Figure 1. Virtualization and SSL Orchestration showing each step of web service delivery before activating QAT

Intel Architecture Adapts for Continual Efficiency

As the leader in defining processor architecture for decades, Intel has unveiled new CPU architecture generation over generation continually adapting and improving compute performance for today and tomorrow's new challenging workloads. Intel offers many processor products to address different requirements from customer premise equipment to data center servers.

For example, Intel's 2nd Generation Intel® Xeon® Scalable Processors offer stable platform operations, agility, a smaller footprint, and more energy efficiency. Some highlighted features:

2nd Generation Intel® Xeon® Scalable Processor

2nd Generation Intel® Xeon® Scalable Processor provides architecture improvement to support higher compute requirement for various use-cases.

Intel® QuickAssist Technology (Intel® QAT)

Intel® QAT is developed to accelerate cryptographic functions include cipher, hash and authentication operations, public key functions, compression and decompression functions.

Intel® Ethernet Network Adapter XXV710-DA2

Flexible and Scalable 10/25GbE Network Adapter with Hardware Optimizations and Intelligent Offloads for Cloud and Network Virtualization Deployment.

The benefits of deploying a virtualized BIG-IP with Intel® QuickAssist Technology include:

1. **Portability:** BIG-IP can be deployed, migrated, and scaled up/down dynamically as a VNF within the datacenter.
2. **Efficiency:** Asymmetric/symmetric crypto operations can be offloaded from the CPU, (e.g., Intel Advanced Encryption Standard-New Instructions [AES-NI] set, to utilize Intel® QAT either on the Platform Controller Hub or Add-In Card).

As a result, Infrastructure-as-a-Service (IaaS) providers no longer need to purchase fixed function hardware appliances, but instead deploy a variety of workloads on general purpose servers while maintaining a high degree of performance.

F5 Lab Test Setup and Configuration

The test server, also known as the Device Under Test (DUT), is equipped with 2nd Generation Intel® Xeon® Scalable Processors along with Intel® Ethernet 700 Series Controllers and built-in Intel® QuickAssist Technology accelerator adapters. For the purpose of this test, the F5 BIG-IP VE Local Traffic Manager (LTM) Virtualized Network Function (VNF) is:

- Deployed on a Network Function Virtualization (NFV) infrastructure
- Leverages Transport Layer Security (TLS)
- Uses Public Key Exchange (PKE) for authentication

The hypothesis for this test is that by adding the incremental processing power of Intel's QAT, the SSL processing in the F5 Virtual Edition software rivals F5 industry standard, mid-range hardware platforms to solve the industry higher demand on SSL offload requirements. This results in more efficiency and improving the TCO.

The F5 Local Traffic Manager VNF was selected because it provides a benchmark to demonstrate improved throughput capabilities of the DUT in terms of HTTPS Transactions Per Second (TPS) with Intel® QAT compared to TPS capacity when using the system CPU alone. HTTPS Transactions Per Second (TPS) is the key metric because it is the simplest indicator of SSL capacity and usually of greater importance than SSL throughput.

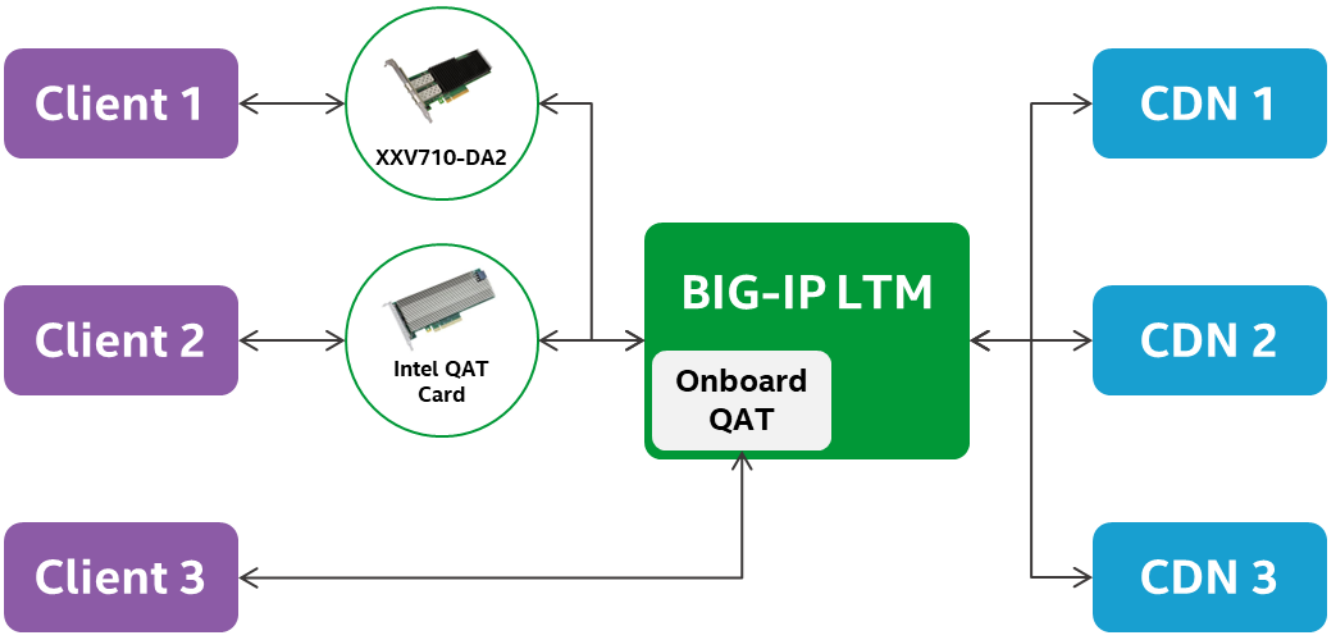

Figure 2. BIG-IP VE: Local Traffic Manager adding optional Intel SmartNICs for offload when Intel QAT is NOT built into the core processor.

Our objective is to improve SSL transaction volume without deploying additional servers. In this use case the virtualized Local Traffic Manager acts as an intermediary load balancer and performs TLS offload for a set of clients requesting HTTP web content from a set of CDN servers in a typical configuration. The LTM VNF is configured with Intel® XXV710-DA2 NICs for network connectivity along with an Intel® QAT card to add incremental hardware offload of PKE operations. To measure the TLS performance benchmark, we will demonstrate the cryptographic processing capabilities of the underlying platform for the LTM application.

Specifically:

- The client server negotiates a cipher suite
- The client performs a PKE operation to verify the authenticity of the LTM proxy
- Once authenticated the client and server will communicate over a secure channel using symmetric encryption

The test configuration was to compare QAT-accelerated performance to CPU-only (native) performance on the same server. HTTPS requests were to be sent to the BIG-IP VE LTM VNF at the highest rate it would support with a given number of CPUs, to find peak native performance. Intel QAT accelerations were then enabled, and the volume of traffic was increased to find the new peak threshold.

Pre-test settings for BIG-IP VE LTM with TLS offload

Component	DUT
DUT NICs	1x XXV710-DA2
DUT NIC VFs	2x 25 Gbps VFs (2x VFs/VM, 1x VF/PF)
DUT vCPUs	2x 6230N 20C40T (12C/12T per BIG-IP VM)
DUT Idle Cores/Threads	28C/68T
Ciphers	ECDHE-RSA-AES-128-GCM-SHA256
DUT QAT	2x AICs Rev 04 (12x VFs, 4x VFs/PF)
Traffic Generator	Oslo 21.06 (25 Gbps TX/Socket, 50 Gbps TX/Server)
Traffic Profile	HTTPS GET Requests (IPv4+TCP)
DDR Memory	12 Slots/32GB/2933 MT/s
Platform	HPE ProLiant DL380 Gen10
Host OS	Red Hat Enterprise Linux release 8.2 (Ootpa)
Host Kernel	4.18.0-193.el8.x86_64
gcc	gcc version 8.4.1 20200928 (Red Hat 8.4.1-1)
glibc	glibc-2.28-151.el8.x86_64
Host i40e driver/FW	2.8.20-k/7.00
BIG-IP/QAT Driver	15.1.2/1.7.l.4.12.0-00011
NGINX	1.18.0
BIOS	HPE U30 Revision 2.13
Microcode	0x5003102
Test By	F5 Networks
Test Date	10/05/2021

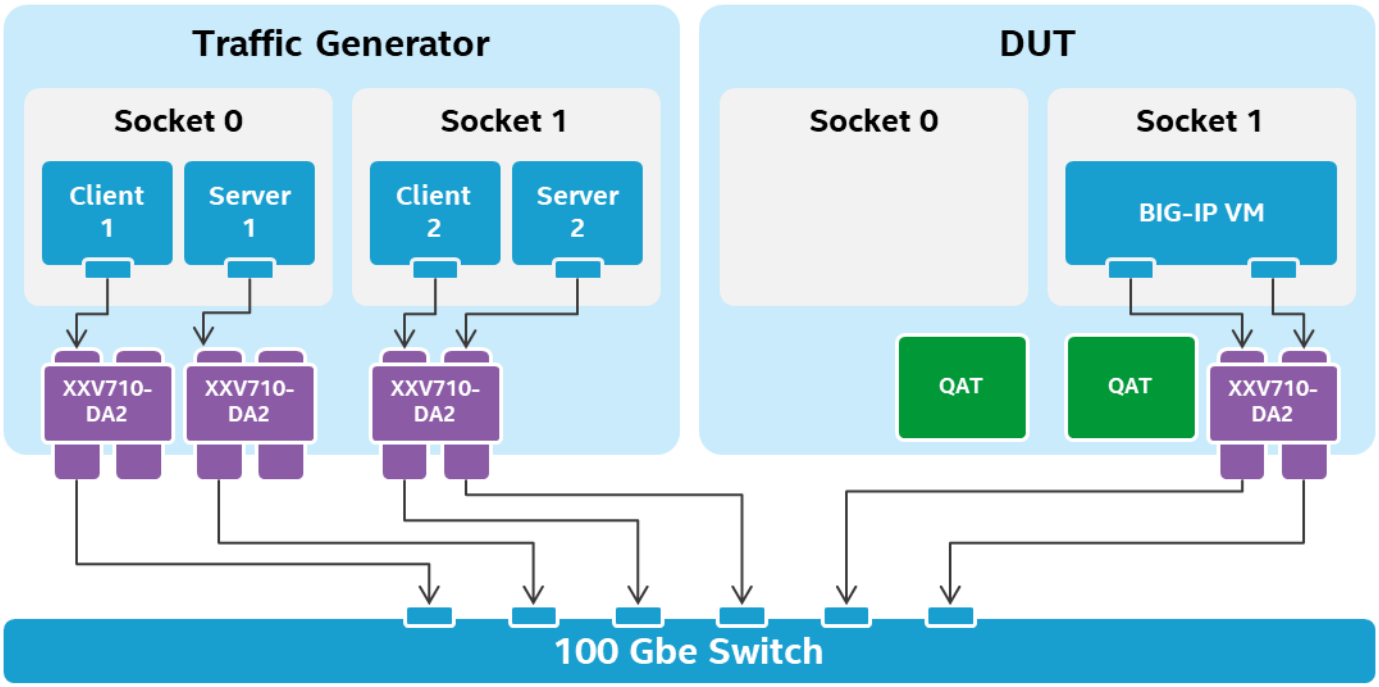

Figure 3. Test Configuration with software traffic generator designed to load the BIG-IP VE LTM with HTTPS requests

Test Results

Results show that the BIG-IP with Intel QAT can significantly handle up to 9.6x more[3] HTTPS requests and better bulk throughput/transactions per second when compared with a system that utilized the system CPU alone for cryptographic operations.

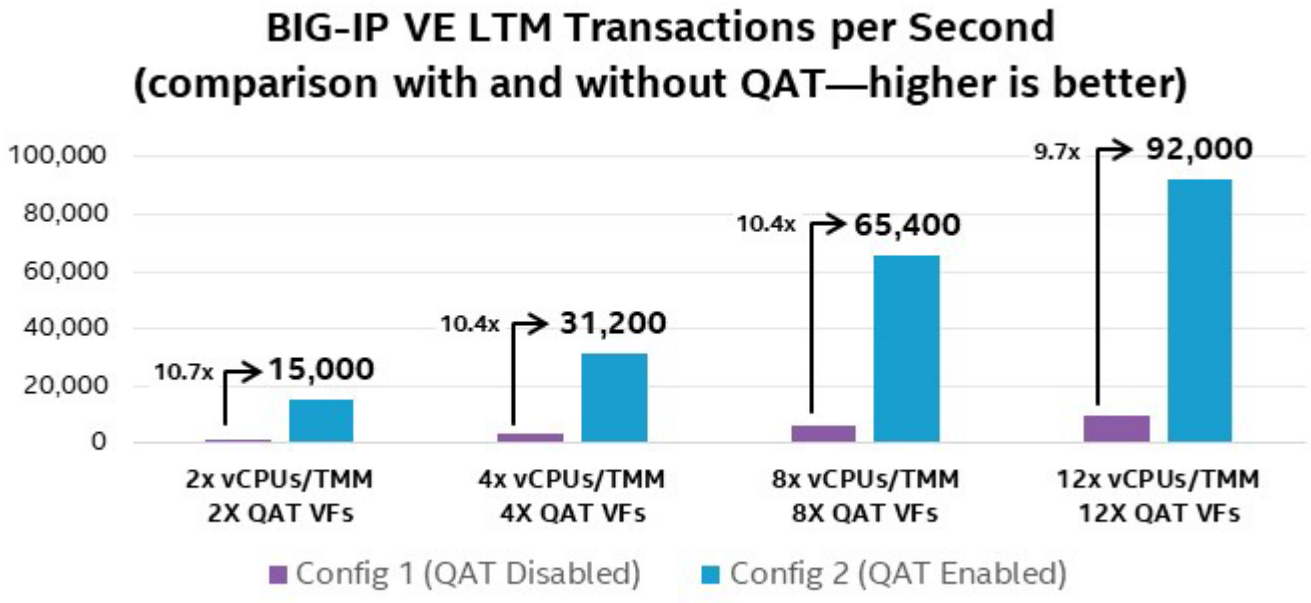

Figure 4. Transactions per Second show nearly 10x increased efficiency vs. software only[3]

Total Cost Analysis for a Data Center (using on-prem metrics)

Telecommunication Service Providers typically deploy the most efficient data center infrastructure—both physical and virtual—to only meet their current viable product (MVP) needs. There are resulting incidental inefficiencies. For

3. See below for workload details and configurations. Results may vary
Performance varies by use, configuration and other factors. Learn more on the Performance Index site.

example, some individual servers may rarely get used or remain idle, waiting on peak demand. All while consuming power and passively adding to TCO.

This section demonstrates how our test scenario models the workload of a standard 42U rack using 8 Xeon 6320N servers to document shifting impacts on aggregated traffic dependent on:

- The number of servers
- CPU characteristics (number of cores, TDP)
- The number of QAT units activated, correlated to server quantity

This paper is working under the hypothesis that telcos will enforce a maximum CPU utilization at deployment time. See how cores sitting idle can instead deliver accelerated performance improvement of nearly 10x vs Software with the same control plane and same tooling by activating QAT.

By Rack View

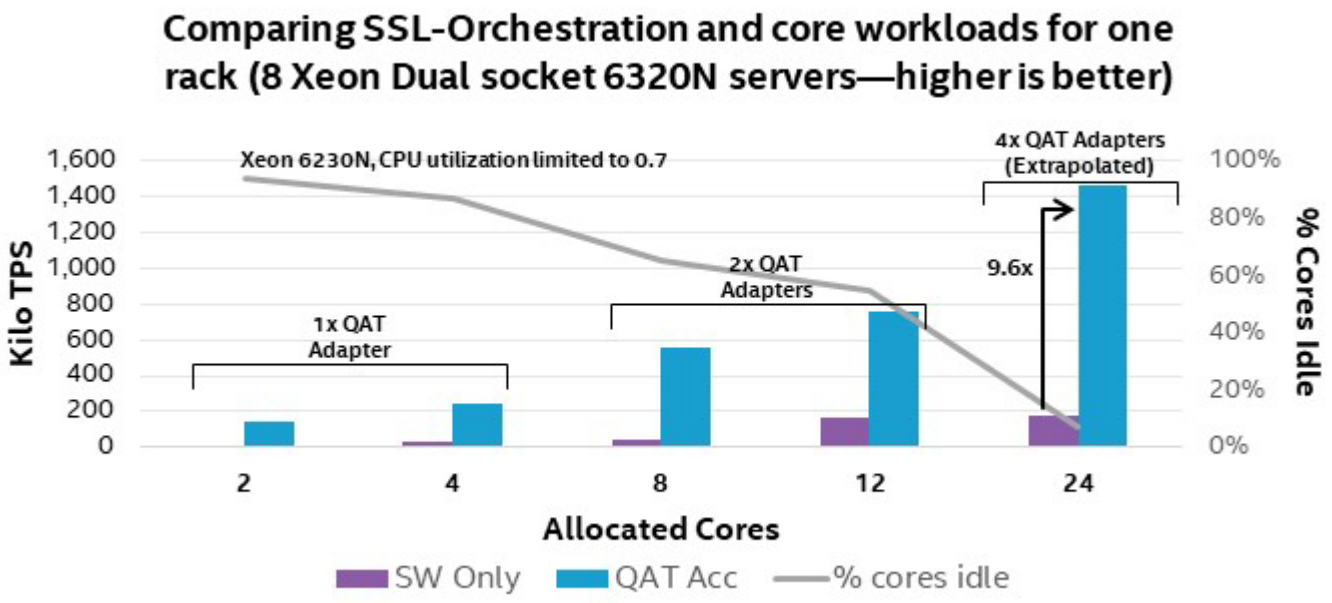

Figure 5. SSL-Orchestration Throughput vs SW + QAT design combinations in range of cores, Rack View[3]

Graph above illustrates the value of a cloud native SSL-O workload for various Key Performance Indicators:

- Choosing the flexibility F5 deployment for BIG-IP VE software facilitates balancing between SSL-O and free cores to run service delivery workloads
- Multiple design points shown from 2 to 24 cores, Software and QAT accelerated solutions are possible with the same control plane and same tooling
- Easy scaling
- By increasing number of cores dedicated to SSL-O per server = Software only configuration
- By increasing number of optional QAT cards dedicated to SSL-O per server = QAT acceleration configuration
- Increasing number of servers

By Edge Server View

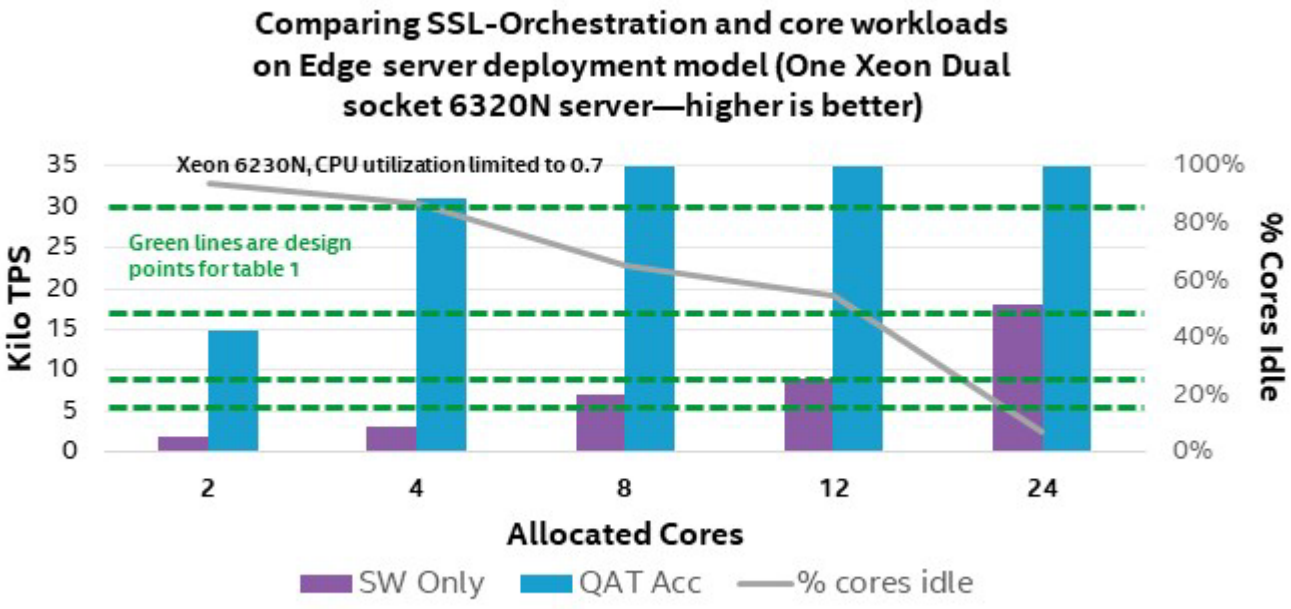

Figure 6. Mapping SLA to SSL-O KPI, Service Edge view[3]

For edge solutions that require up to a maximum 30K TPS, by embracing QAT, this test data shows that CPU idle cores stay above 85% (as noted above in Figure 5), making room for service delivery workloads on the same compute node.

When trying to serve that many transactions on a software only configuration, the CPU would be consumed just on the SSL-O workload.

The following table provides several design points and the number of cores required to implement them:

SSL TPS SLA	SW only - # cores allocated	SW only - % idle cores	QAT Acc - # cores allocated	QAT Acc - % idle cores
5K	8	69%	2	92%
9K	12	54%	2	92%
18K	26	8%	4	92%
30K	NA	NA	4	85%

Table 1 to map design points to percentage of idle cores

In the context of service providers delivering services, income per month will be compared to operation costs per month. We focused on the KPI “Cost per Kilo TPS per month”, as means to map monthly service delivery cost and measured that Intel QAT brings an average 85% (noted in Table 1) cost reduction vs an equivalent SW only solution across all configurations modeled, as shown in Figure 6.

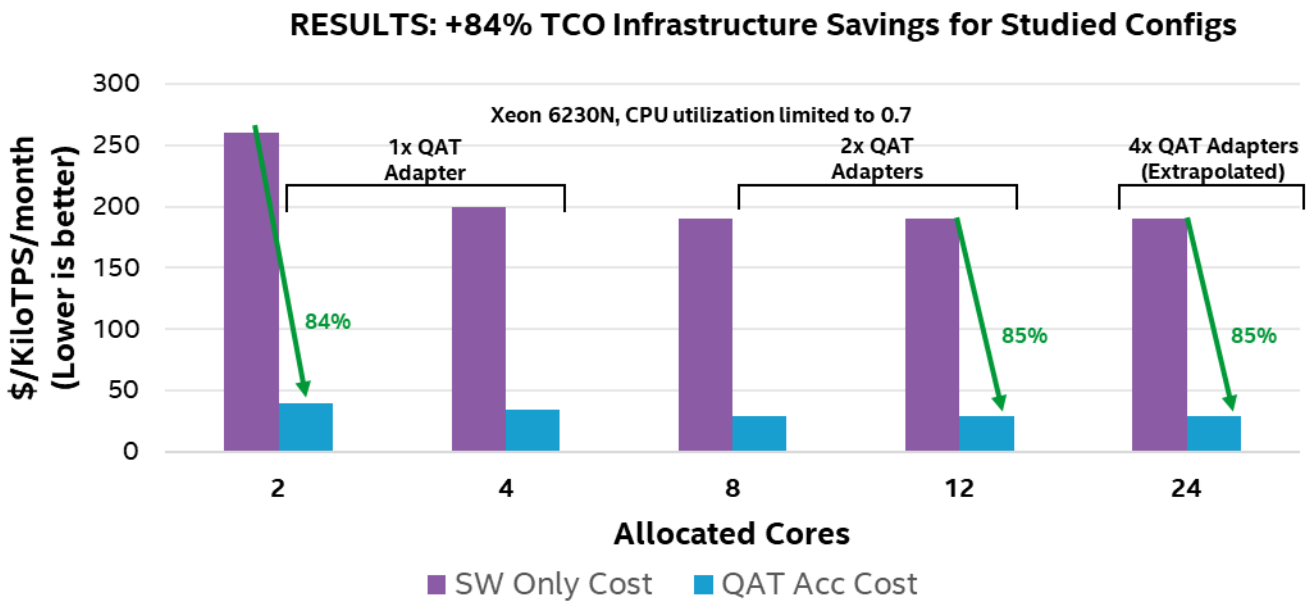

Figure 7. SSL-O service delivery cost, SW only vs. Intel QAT Accelerated[3]

QAT frees up cores to run service delivery

Before the enablement of Intel QAT, F5 software leveraged CPU-based solutions for encryption such as Intel® Advanced Encryption Standard New Instructions (Intel® AES-NI). These software-based encryption solutions deliver very high performance when taking advantage of Intel AES-NI but still consume a percentage of CPU capacity depending on the amount of data and how many requests are directed at a particular application or service. Scaling performance of software-based encryption requires distributing the load across many servers or deploying servers with higher performance CPUs or with multiple CPUs. Utilizing encryption hardware, like Intel QAT, provides increased performance and frees up CPU processor cycles for other compute needs.

The most applicable F5 BIG-IP VE use cases for Intel QAT are related to SSL Offload, with or without full SSL Orchestration. Scenarios such as DNS over HTTPS (DoH) represent simple SSL Offload and will benefit as much as complex HTTPS services, including vCDN.

Additional benefits of deploying a BIG-IP with Intel® Quick Assist Technology include:

1. BIG-IP is running within a Virtual Machine and is portable, therefore can be more easily deployed, migrated, and scaled up and down dynamically as a VNF within a datacenter.
2. The asymmetric and symmetric crypto operations can be offloaded from running on the CPU, for example with the AES-NI instruction set, to Intel® QAT either to the Platform Controller Hub or to an Add-In Card. As a result, infrastructure as a service providers will no longer need to be locked into purchasing fixed function hardware appliances, rather they can consume generic servers and run a variety of workloads on the platform all while maintaining a high degree of performance. This crypto offload enables deeper and broader use of the common server platform for application services, web servers, data analytics, and so much more.

Conclusion

The use of F5 BIG-IP VE and Intel QAT together enables users to noticeably reduce operating cost by improving efficiency for core and edge investments to simplify operations for their Web Application Firewall, SSL Offload and SSL Orchestration use cases while keeping cores free for production applications. By deploying one of these tested configurations, operators can consistently reduce monthly operational costs up to 84% (see Figure 6).

As noted above, the 9.6x SSL transaction rate using QAT equals that achieved with dedicated hardware solutions. Coupled with automated discovery and provisioning, a greatly simplified technology environment is achievable in many locations – edge, network edge or core datacenter.

In the end, this combination helps users achieve more with the same and get the maximum from their investments. Interactions with the security processes are made more efficient by offloading or orchestrating SSL transactions, which speeds the overall user experience.

Notices & Disclaimers

Performance varies by use, configuration and other factors. Learn more on the Performance Index site. Performance results are based on testing shown in configurations and may not reflect all publicly available updates. No product or component can be absolutely secure.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

TCO Cost Assumptions

- TCO calculations by Intel as of Q3'2021
- Server unit cost calculated as average across the industry: server cost: $11,289 and QAT AIC unit card cost: $748
- 8 servers in a 10KW rack, with the following server wall power requirement: server 0.6KW and per QAT AIC unit card power consumption 0.023KW
- Average cost/KWh used: $0.12
- Assumed PUE factor of 1.5

Assumed annual rack cost per RU of $75.76