OPPO Accelerates Cloud Security Gateway HTTP3-QUIC with Intel® QAT

"OPPO has been committed to delivering the ultimate experience to users through continuous innovation. In the era of 5G and AI, OPPO has partnered with Intel and utilized Intel® QAT to build an NGINX-based security gateway that supports HTTP3-QUIC, significantly lowering network latency and improving connection success rate, delivering a smoother experience for OPPO device and service users. OPPO continues to innovate in the fields of network transmission, smart routing, and protocol optimization, to improve the user experience. "

Hang Fang

OPPO Head of Security and Privacy

Overview

The proliferation of smart mobile devices and internet services has resulted in rapid increase in internet traffic, putting increasing pressure on internet systems. Meanwhile, users are also becoming more demanding about the latency and stability of internet applications, requiring internet service providers to achieve low-latency, high-throughput, and quality-of-service (QoS) aware networks. It is becoming increasingly difficult for the traditional HTTP protocol to meet the needs of internet applications. This has resulted in the rapid development of HTTP/3 operations based on Quick UDP Internet Connections (QUIC).

Compared to traditional TCP, QUIC involves many encryption and decryption operations, placing higher demands on performance, affecting the load-bearing ability of internet operations, as well as infrastructure return over investment (ROI). To accelerate HTTP3-QUIC, OPPO's Security and Privacy team has worked with Intel to achieve software offloading of encryption and decryption computing at the hardware level through Intel® QuickAssist Technology (Intel® QAT). This technology delivers up to 3 times the HTTP3-QUIC concurrent short connections compared to CPU-based software acceleration[1], while improving cost-effectiveness.

Background: Rapidly growing QUIC operations bring challenges to encryption-decryption performance

The current TCP/HTTP-based internet technology is facing numerous challenges from the rapid growth of mobile internet, such as low-latency support for interactive communications, security and privacy of user data, and development and deployment of new transport mechanisms. QUIC was born out of such needs. QUIC is a new transport protocol that combines TCP's congestion control and loss recovery characteristics, with more powerful signaling capabilities. In addition, QUIC also reduces latency by reducing the RTT (Round-Trip Time) of connected devices. QUIC integrates the key agreement feature of TLS 1.3, requiring that all connections be encrypted. Mandatory encryption not only helps protect user data, but it also helps prevent middlebox tampering of packet data[2]. HTTP/3 was designed to fully leverage the advantages of the QUIC protocol.

[1] Data from tests performed by OPPO in June 2022. Test configuration: Dual Intel® Xeon® Gold 6330 processor,512 GB total memory (16 x 32G DDR4), 480 GB SATA SSD, 6.4 TB Intel® SSD DC P4610,25G network card, CentOS Linux release 7.6, Linux 3.10.01160.31.1.el7.x86_64 x86_64, QAT1.7.L.4.14.0-00031, GCC 7.5.0, QAT_Engine v0.6.6 + BSSL support. Intel does not control or audit 3rd-party data. Consult other sources to confirm the validity of any data cited.

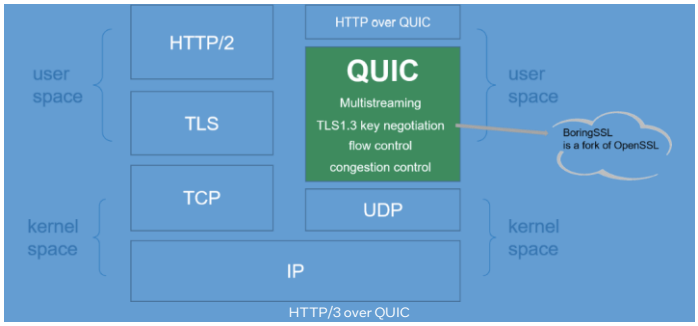

Figure 1. Next-generation transport layer protocol QUIC

The numerous improvements that the QUIC protocol provides for network transport provides significant advantages in terms of latency, reliability, and user space congestion. That is why more and more web applications are migrating from TLS to QUIC. QUIC-based HTTP/3 usage continues to grow. W3Techs data indicates that 24.9% of websites are already using HTTP/3[3] as of May, 2022.

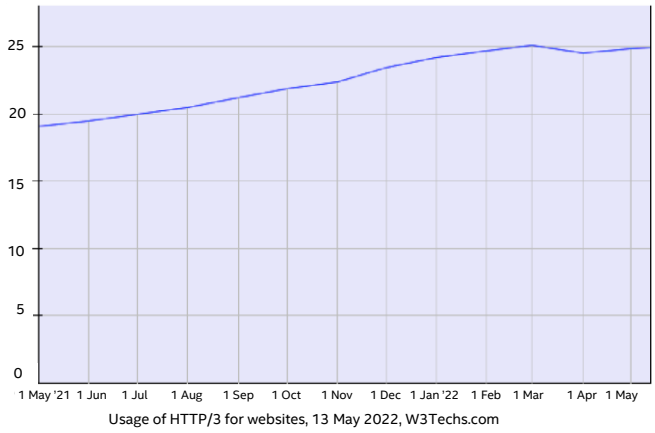

Figure 2. Percentage of websites using HTTP/3

Users that want to migrate their web applications to QUIC find that while QUIC brings numerous advantages, it also presents challenges in terms of portability, integrability, compatibility, and performance. Like TLS, the key agreement of QUIC uses a large amount of CPU resources while a connection is being established. The server and client generate a session key when establishing a connection, but the handshake is typically encrypted using asymmetric keys, requiring more computing power.

Because of the CPU resources that data encryption and decryption takes up, Internet Service Providers (ISP) typically have to increase their server resource investment in order to maintain web application QoS, which generates considerable cost pressure. In addition to software optimization and 3rd Generation Intel® Xeon® Scalable processors, existing servers can also optimize QUIC encryption and decryption by adding a dedicated Intel QAT accelerator, offloading encryption, and decryption to the Intel QAT accelerator for processing.

Solution: OPPO accelerates HTTP3-QUIC with Intel® QAT

To fully leverage the operational advantages of QUIC technology, OPPO's Security and Privacy team built a custom security gateway based on NGINX on the front-end of the unified access layer, adding configuration management and WAF capabilities while optimizing TLS offloading, as well as supporting the QUIC protocol. OPPO implemented Intel QAT to accelerate the encryption and decryption operations after the introduction of QUIC.

As shown in Figure 3, OPPO's unified access layer includes layer-4 network load balancing and layer-7 application load balancing (security gateway). Layer-4 network load balancing performs scheduling of TCP or UDP data packets. Layer-7 application load balancing is the final gateway for accessing backend services, receiving client requests (as the server-side), and ensuring network security. In traditional HTTP1/2 processing, TLS handshake offloading is achieved using OpenSSL, and while establishing a connection TLS handshake utilizes RSA asymmetric encryption.

The traditional solution uses synchronous operations, congesting an API call until the request is complete. When a parallel processing entity forms part of the execution flow, the processor may be left idle at times, wasting resources. By using a single accelerator under this API, the application can execute busy cycles while waiting for a response from the accelerator or use an execution model like pthreads to perform context switching, allowing other useful tasks to be completed while waiting. Both solutions take up additional resources.

To overcome this challenge, Intel introduced asynchronous HTTPS connection asynchronous offloading in NGINX, significantly improving parallel connection processing capability. OPPO's async-mode-NGINX-based security gateway offloads RSA asymmetric encryption to the Intel QAT accelerator, improving overall server performance.

---

2 [1] Cui Y, Li T, Liu C, et al. Innovating Transport with QUIC: Design Approaches and Research Challenges[J]. IEEE Internet Computing, 2017, 21(2):72-76. 3 Data from: https://w3techs.com/technologies/details/ce-http3

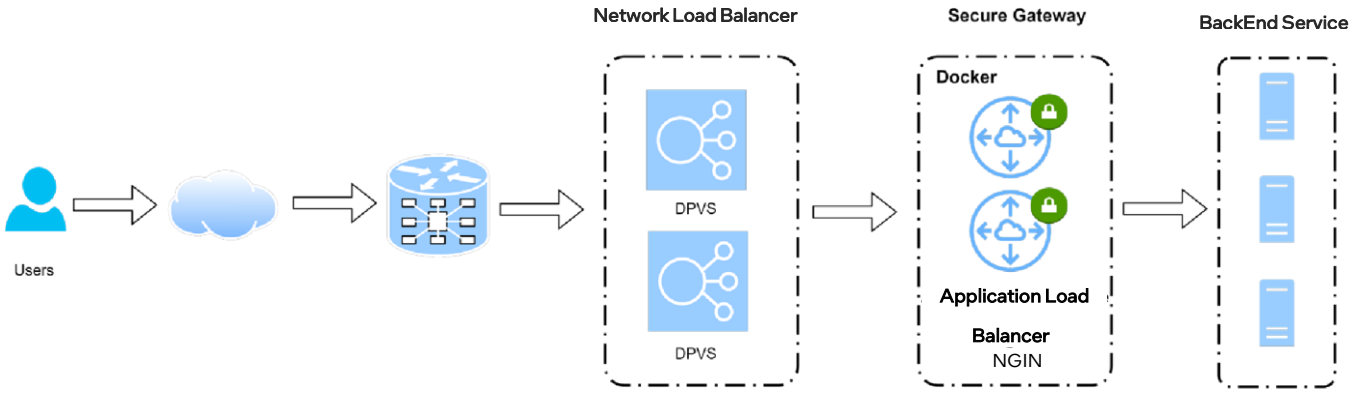

Figure 3. OPPO unified access layer architecture

Intel® QAT hardware acceleration technology is targeted at network security and data storage needs. Intel QAT focuses on data security and compression acceleration, helping to increase the performance of applications and platforms. In terms of network security applications, Intel QAT offers support for multiple forms of symmetric data encryption (e.g., AES), asymmetric public key encryption (e.g., RSA, elliptic curve) and data integrity (SHA1/2/3), accelerating operations such as data encryption/decryption and digital signatures.

Intel QAT accelerators are optimized for NGINX, allowing it to call the accelerator asynchronously. NGINX is a high-performance HTTP and reverse proxy web server that also provides IMAP/POP3/SMTP services. Asynchronous mode allows NGINX to process in parallel and reduce wait time, achieving the necessary performance while using less resources, and improving application response.

Intel QAT also supports powerful compression acceleration, delivering Intel QAT-accelerated synchronous compression API. With support for stateless parallel compression/decompression, pipeline processing based on Intel QAT asynchronous API, thread safety compression, and zero-copy mode, it can integrate multiple small data compression/decompression requests into a single Intel QAT hardware request, reducing CPU utilization and increasing throughput.

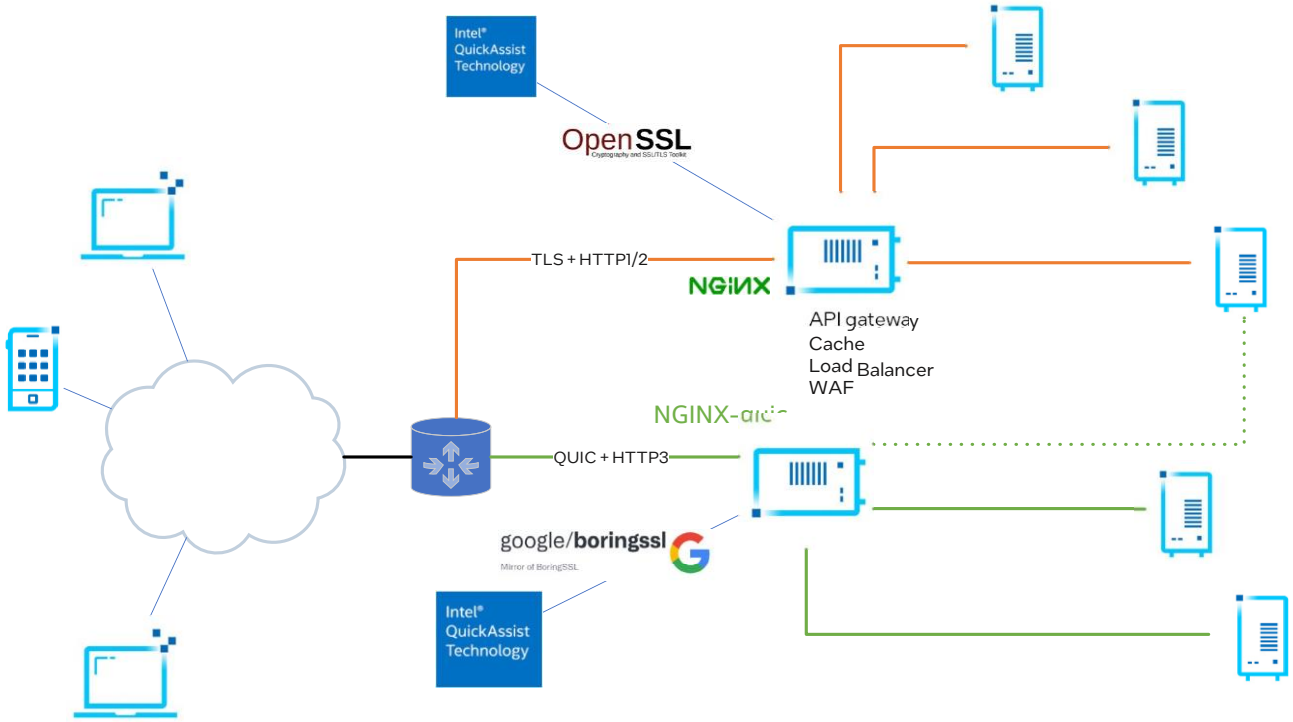

Figure 4. Intel® QAT accelerates encryption/decryption workloads under different protocols

OPPO's security gateway is customized based on NGINX, inheriting NGINX's asynchronous characteristics. The crypto engine of the mainstream QUIC protocol was developed based on BoringSSL, however. When the security gateway supports QUIC, the crypto library switches to BoringSSL, and traditional OpenSSL-based Intel QAT acceleration no longer applies.

In response, Intel proposed a new solution that includes Intel QAT acceleration adaptation for the crypto library, QUIC protocol stack, and asynchronization of the Intel QAT engine. These optimizations allow OPPO to achieve TLS 1.2/ TLS 1.3/ QUIC acceleration in parallel in a single web server, achieving support for multiple SSL libraries in a unified Intel QAT engine library, including OpenSSL and BoringSSL, and meeting the acceleration needs for OPPO's security gateway.

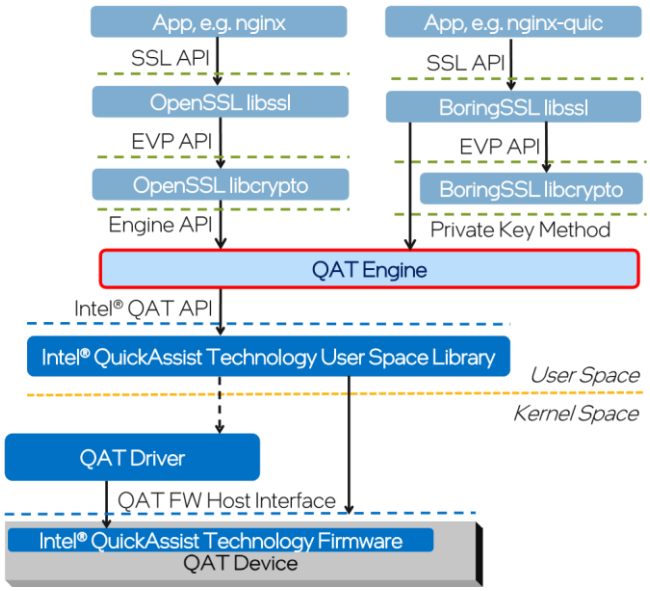

Figure 5. OPPO security gateway accelerated with Intel® QAT

Validation: 3-4x performance increase

OPPO performed tests to validate the QUIC performance increase from Intel QAT, with data shown in Figures 6 and 7.

Tests indicate that Intel QAT demonstrates significant performance advantages over software acceleration on 3rd Generation Intel® Xeon® Scalable processors, achieving 4x TLS performance and 3x QUIC performance[6].

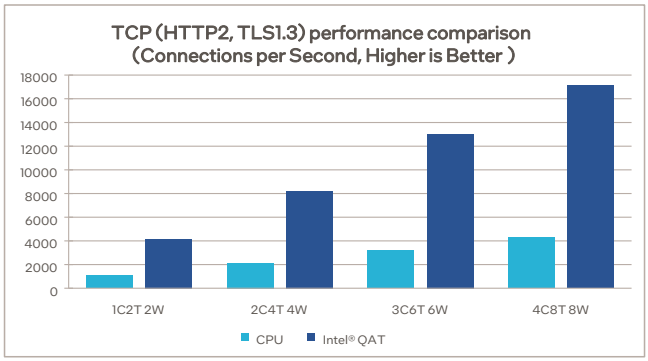

Figure 6. TCP (HTTP2, TLS1.3) performance comparison on CPU and Intel® QAT[4]

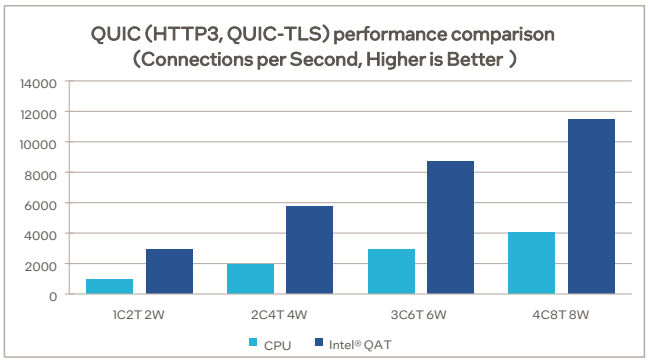

Figure 7. QUIC (HTTP3, QUIC-TLS) performance comparison on CPU and Intel® QAT[6]

Thanks to the increase in single server node performance brought on by Intel® QAT, OPPO can meet more encryption/decryption needs of its web operations without server cluster expansion or disruptive update of the existing system environment, while keeping datacenter infrastructure TCO in check.

This solution also has long-term evolvability, supporting cryptoNI technology across the entire Intel QAT accelerator and CPU lineup, meeting the optimization requirements of future TLS 1.2/ TLS 1.3/ QUIC versions on different hardware platforms. Intel QAT also helps further improve performance and affordability in other OPPO encryption and decryption environments.

4,5,6 Data from tests performed by OPPO in June 2022. Test configuration: Dual Intel® Xeon® Gold 6330 processor, 512 GB total memory (16 x 32G DDR4), 480 GB SATA SSD, 6.4 TB Intel® SSD DC P4610, 25G network card, CentOS Linux release 7.6, Linux 3.10.0-1160.31.1.el7.x86_64 x86_64, QAT1.7.L.4.14.0-00031, GCC 7.5.0, QAT_Engine v0.6.6 + BSSL support. Intel does not control or audit 3rd-party data. Consult other sources to confirm the validity of any data cited.

Prospect: Working together to accelerate development of the HTTP3-QUIC ecosystem

More and more users will be migrating their web applications to HTTP3-QUIC in the foreseeable future, further promoting the prosperity of the HTTP3-QUIC ecosystem. To help more users achieve this transformation, Intel is working to further optimize the application of Intel QAT in HTTP3-QUIC and cooperating with partners such as OPPO to optimize relevant technology stacks.

As a leader in going beyond boundaries, OPPO is working to promote the use of Intel QAT in more scenarios, accelerating data processing efficiency, increasing hardware resource utilization rate, delivering a better internet application experience for users, and building seamless user-centric ubiquitous services.

About OPPO

OPPO is a global leader in smart device manufacturing and innovation. Behind the mission of elevating life through technological mastery, we strive to be a sustainable company that contributes to a better world. Since its inception in 2004, OPPO has expanded its operations to more than 40 countries around the world, boasting more than 300 million total users.

About Intel

Intel (Nasdaq: INTC) is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com.