NEXCOM

Committed to Customer Success

Solution Brief

NSA 5200 Provides Scalable AI Security for Threat Detection and Encrypted Networks

The Trend

The growing reliance on the Internet for communication, business operations, and critical infrastructure has increased the need for enhanced security and privacy in network communications. Companies are protecting their sensitive data and operations through encryption, access control, and threat detection across various cybersecurity protocols, including SASE, ZTNA, VPN, DNS over HTTPS (DoH), DNS over TLS (DoT), and QUIC. The integration of AI into these protocols enables real-time threat intelligence, automated anomaly detection, and adaptive security measures.

One of the key trends is AI-driven threat detection in encrypted traffic, where machine learning models (such as TensorFlow, PyTorch, Scikit-learn and more) analyze behavioral patterns rather than relying on traditional deep packet inspection. This approach helps identify threats hidden within encrypted channels without compromising privacy. Additionally, AI is improving risk-based authentication and access control, dynamically adjusting security policies based on user behavior, device health, and other factors.

Another significant trend is the use of AI for network performance optimization, particularly in latency-sensitive environments. AI-powered predictive analytics help dynamically route traffic, mitigate congestion, and optimize bandwidth usage in secure network protocols. Furthermore, adversarial AI threats are emerging, where attackers use AI to break through security measures, requiring constant adaptation of AI-driven cybersecurity defenses.

The Challenge

The adoption of AI-driven security solutions across DoH, SASE, ZTNA, VPNs, and other encrypted protocols is accelerating in enterprise security, cloud infrastructure, and Zero Trust architectures. However, the performance and efficiency of AI applications depend heavily on the underlying hardware and computational resources required for processing large volumes of encrypted traffic in real-time.

AI models, particularly deep learning demand substantial processing power, memory, and storage, making large-scale deployment difficult for resource-constrained organizations. Additionally, balancing low-latency performance with advanced AI-driven threat detection is complex, as real-time encryption analysis can introduce network slowdowns that impact user experience.

Another key challenge is the accuracy and resilience of AI-based threat detection. While AI helps identify anomalous traffic patterns without decryption, it remains vulnerable to false positives that could block legitimate connections, disrupting operations. Meanwhile, attackers are leveraging adversarial AI techniques to evade detection by getting around malicious activity within encrypted traffic.

To overcome these challenges, cybersecurity teams must adopt adaptive AI models, invest in hardware acceleration (e.g., TPUs, GPUs) to ensure effective and scalable security solutions.

NEXCOM Solution

NEXCOM NSA 5200 is a 1U rackmount

Scalable 1U Cybersecurity Rackmount with 14th Gen Intel® Core™ Processor

cybersecurity appliance, powered by the latest 14th Gen Intel® Core™ processors, the NSA 5200 provides high-performance capabilities for real-time attack detection and machine learning. NSA 5200 is specifically designed to handle the demanding workloads of modern security protocols without compromising network speed or operational efficiency.

The 14th Gen Intel® Core™ processor delivers powerful general-purpose computing while integrating Intel Deep Learning Boost (DL Boost) and Advanced Vector Extensions (AVX) for optimized AI performance. These enhancements enable efficient inference for AI models, making them well-suited for both edge and cloud environments.

Besides embedded Intel technologies, NSA 5200 AI performance can be further enhanced with integrated graphics (iGPU) Intel® UHD Graphics 770 that can be enabled through drivers, while the addition of an Intel® Arc™ A370M discrete GPU (dGPU) via first or second LAN module slot with PCIe 5.0 signal provides substantial additional GPU processing power, essential for accelerating heavy AI workloads, deep packet inspection, and machine learning inference. This flexibility in GPU choice ensures that the NSA 5200 can increase its capability to meet specific security needs and adapt to evolving AI models.

Equipped with four DDR5 ECC/non-ECC UDIMM slots, supporting up to 128GB of memory, NSA 5200 ensures that even the most data-intensive AI models, such as deep learning networks used for encrypted traffic analysis, can run efficiently, maintaining high accuracy in detecting potential threats without overwhelming system resources.

NSA 5200 features TPM 2.0 (Trusted Platform Module) enhancing security to enable hardware-based encryption, essential for safeguarding sensitive data during AI-driven operations.

The system offers four external extension slots that adapt a variety of NEXCOM LAN modules with up to 100GbE per slot, boosting Ethernet throughput, as well as additional expansion cards, including storage, and wireless adaptors; and mentioned above AI accelerator card.

The combination of AI acceleration, hardware-based security features, and network flexibility allows organizations to deploy AI-driven security solutions while overcoming the performance bottlenecks. This makes the NSA 5200 an ideal choice for IT and OT infrastructure managers who need a powerful, adaptable solution to protect against emerging threats in modern network environments.

TABLE I
NSA 5200 TEST CONFIGURATIONS

Item	Configuration 1	Configuration 2	Configuration 3
CPU	Intel® Core™ i9-14900 with Turbo Boost ON		
dGPU	x	x	Intel® Arc™A370M, installed in PCIe Gen5 slot
iGPU	x	Intel® UHD Graphics 770	x
Memory	2 x DDR5 SDRAM DIMM 4800 MHz, 32GB		
Storage	1 x 256GB		
Network	Intel® Ethernet Controller I210-AT Gigabit Network Connection		
OS	Ubuntu 22.04.3 LTS (6.8.0-51-generic)		

NSA 5200 AI Capability: Three Setups Tested

Test Configuration and Topology

For a full overview of NSA 5200 real-world AI performance, it has been evaluated across three configurations described in Table I. The comparison aims to determine the optimal balance of performance, efficiency, and scalability for AI-enhanced cybersecurity applications running on the NSA 5200.

The tests focus on DNS over HTTPS (DoH) with deep learning (DL) malicious URL detection as an example use case. The core of DoH application test was URLNet TADK anomaly detection. This workload involves analyzing encrypted DNS queries to detect potentially harmful domains, a compute-intensive task that benefits from AI acceleration. Performance judgment is based on queries per second (QPS) to measures throughput efficiency.

The deployment of the DoH testing architecture with client and server (DUT1/2/3) nodes is shown in Figure 1. In a same scenario DUT uses different hardware components (CPU, iGPU and dGPU) to adopt AI computation to see the performance.

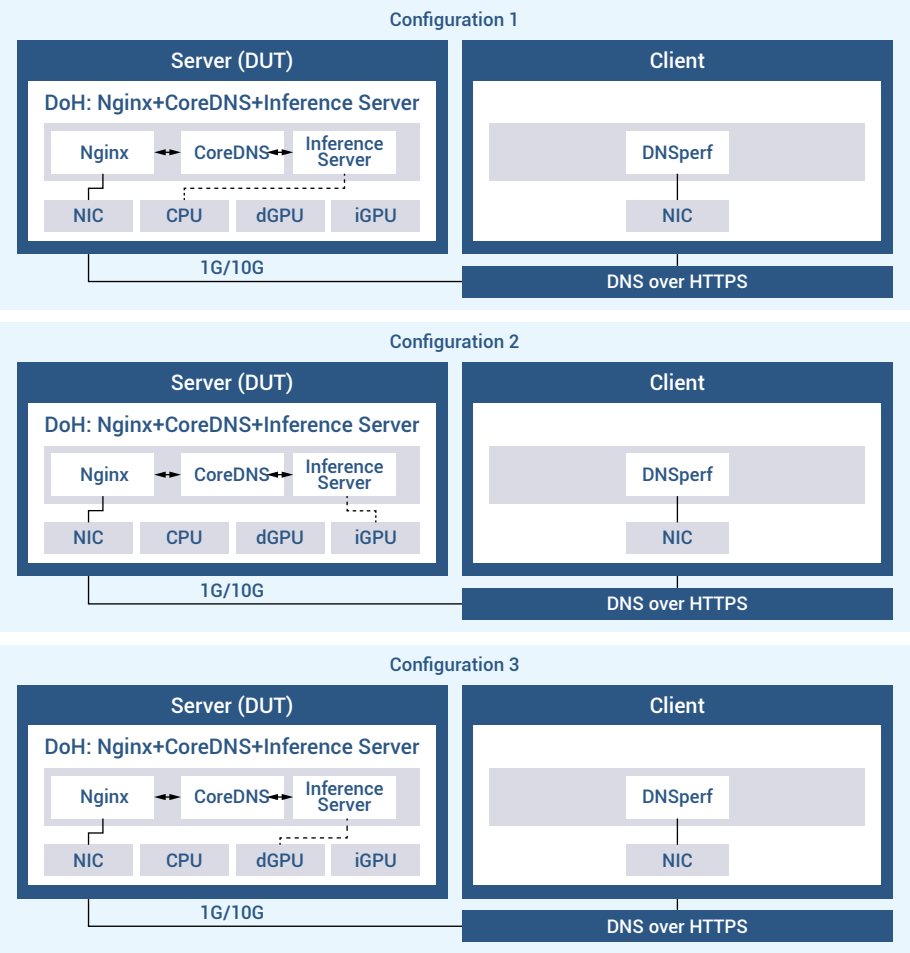

Figure 1. DoH with DL Malicious URL Detection Test Topology.

Real-World AI Performance Testing on NSA 5200

Test Results

The performance evaluation of NEXCOM NSA 5200 across different hardware configurations highlights hardware acceleration and computational efficiency. Additionally comparing different AI models across three hardware configurations is essential to understanding the optimal balance between performance, efficiency, and resource utilization.

For a fair comparison four AI models were chosen: Tensorflow Frozen, Tensorflow Frozen INC, TensorFlow Saved, OpenVINO. Each model leverages hardware differently - TensorFlow Frozen is optimized for general-purpose AI workloads and benefits from high-performance CPU cores. TensorFlow Frozen INC introduces INT8 quantization, improving inference speed and efficiency while reducing computational load. TensorFlow Saved retains flexibility for retraining and fine-tuning, making it suitable for adaptive security models. OpenVINO is specifically optimized for Intel's AI acceleration capabilities maximizing performance on Intel GPUs and enhancing real-time security applications. Detailed test results are shown in Table II.

TABLE II
TEST RESULTS. DOH DL MALICIOUS URL DETECTION WITH URLNET-TADK-AD INFERENCE MODEL

Item	TF/saved, QPS	TF/frozen, QPS	TF/frozen-inc-int8, QPS	OV/OV, QPS
Configuration 1 Intel® Core™ i9-14900	1456.57	1460.11	1215.37	801.41
Configuration 2 Intel® Core™ i9-14900 with UHD Graphics 770	4703.62	4712.85	4151.61	4796.03
Configuration 3 Intel® Core™ i9-14900 with Intel® Arc™ A370M	5019.67	5046.46	4923.49	5151.63

Across all four models, the CPU + dGPU configuration delivered the highest performance, proving the benefits of dedicated AI acceleration. This configuration leverages Intel Arc A370M's high parallel processing power, optimized matrix operations, and high memory bandwidth resulted in substantial gains for high-throughput, real-time Internet traffic analysis and AI-driven cybersecurity applications, such as next gen firewall (NGFW) or intrusion prevention services (IPS).

The CPU + iGPU configuration achieved 7–16% lower performance compared to the dGPU configuration, highlighting the difference in computational efficiency. While the integrated GPU leverages Intel UHD Graphics 770 AI acceleration, the added performance of the discrete GPU provides a measurable advantage over integrated GPU, come readily as a much affordable solution without the need to migrate to more expensive hardware platform to run Edge AI applications.

Despite delivering lower performance compared to GPU-accelerated configurations, the CPU-only configuration remains a viable solution for AI-powered workloads like visual data & image processing. With Intel's high-performance hybrid architecture, AI-optimized instructions, and Deep Learning Boost (DL Boost), the Intel i9-14900 processor can efficiently handle inference tasks, making it a strong option for organizations prioritizing lightweight AI deployments, offering a scalable path for future upgrades.

NSA 5200 Scales for Diverse AI Security Workloads

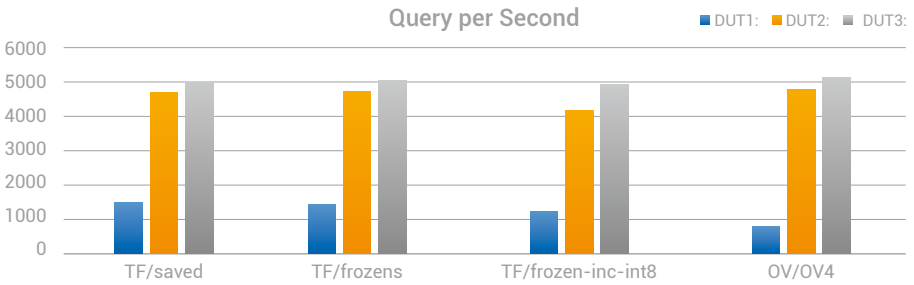

Figure 2. QPS Comparison Results between Different Configurations and Different AI models.

Test results show that each NSA 5200 configuration is optimized for different types of workloads, ensuring the highest QPS rates under different deployment conditions.

By choosing the appropriate configuration, organizations can optimize AI-driven cybersecurity workloads, balancing performance, power efficiency, and real-time inference capabilities.

TABLE III
RECOMMENDED SECURITY WORKLOADS BASED ON NSA 5200 CONFIGURATION

NSA 5200 Configuration	Best for	Key Benefits	Recommended Workloads
Intel® Core™ i9-14900	General-purpose security processing	High single-thread and multi-thread performance	Network firewall, DPI, SIEM, rule-based security detection
Intel® Core™ i9-14900 with UHD Graphics 770 dGPU	Optimized AI inference with OpenVINO	Low power, low latency, efficient for AI-driven security	AI-assisted IDS/IPS, behavior-based malware detection, SSL/TLS inspection
Intel® Core™ i9-14900 with Intel® Arc™ A370M iGPU	High-speed deep learning inference & real-time AI cybersecurity	High throughput, parallel processing, real-time detection	AI-powered DNS security, network anomaly detection, NGFW, IPS, large-scale threat intelligence

Conclusion

The adoption of AI-driven cybersecurity solutions is rapidly transforming network security. As cyber threats become more sophisticated, organizations must balance privacy, performance, and security while handling high volumes of encrypted traffic. However, AI-driven security workloads require optimizing resource utilization to ensure low latency, high throughput, and cost-effective security enforcement without compromising network performance.

NEXCOM NSA 5200 addresses these challenges by providing a scalable, AI-optimized security platform. The flexibility of CPU-driven inference, combined with GPU acceleration, enables high-performance threat detection while maintaining efficient resource allocation. The tests prove that NSA 5200 ensures fast response times, high query-per-second (QPS) rates, and enhanced network security capabilities, effectively safeguarding encrypted communications in Zero Trust, SASE, edge computing and cloud-driven environments.

Committed to Customer Success

NEXCOM, founded in 1992 and headquartered in Taiwan, stands as a distinguished global leader in edge computing and industrial IoT solutions. Demonstrating an unwavering commitment to excellence, NEXCOM provides integrated services encompassing SD-Edge Computing (software-defined edge computing) and cutting-edge MOM (manufacturing operations management) platforms. Its comprehensive solutions include network and communication, mobile computing, video surveillance, smart city and retail, digital healthcare, AIoT services, OT cybersecurity, industrial IoT and industrial robots—all developed based on open standards. As a trailblazer in the industry, NEXCOM continues to set the standard for innovation and reliability, meeting the diverse needs of its global clientele with precision and sophistication.

www.nexcom.com

NEXCOM is a Titanium member of the Intel® Partner Alliance, as a top tier of the Alliance. Intel and more than 500 global IoT partners of the Intel® Partner Alliance provide scalable, interoperable Intel® -based technologies and solutions that accelerate deployment of intelligent devices and end-to-end analytics. Close collaboration with Intel and each other enables Alliance members to innovate with the latest technologies, helping developers deliver firstin-market solutions.

Learn more at: https://www.intel.com/content/www/us/en/partner-alliance/overview.html

Intel and Atom are registered trademarks of Intel Corporation in the United States and other countries.