Solution Brief

NetSec ISVs, Enterprises, Network Security

Sangfor, Intel Accelerate Extended Detection and Response Cyber Security

Tests show that upgrading Sangfor Extended Detection and Response (XDR) analytics to 4th Gen Intel® Xeon® Scalable Processors with Hyperscan optimization can bring up to six times more performance[i] for XDR data analytics workloads

Enterprises face more and increasingly sophisticated cyberattacks as hackers adopt advanced tactics, techniques and procedures (TTPs) that evade traditional security measures. The security systems that enterprises operate on are often not fully integrated and are reactive.

What's needed is a solution that uses data from multiple security tools to improve visibility into the enterprise network.

Extended detection and response (XDR) is a cyber security threat detection and incident response platform that natively integrates data from multiple security products into a cohesive security operations system.

XDR unifies security-relevant endpoint detection with telemetry from security and business tools such as network analysis and visibility (NAV), email security, identity and access management, cloud security, security information and event management (SIEM), and more. It is an evolving security platform built on big data infrastructure to provide security teams with visibility, flexibility, scalability, and opportunities for automation.

Data analytics on big data is one of the core building blocks of XDR. For enterprise customers, it is critical that their XDR platform can process and analyze large amounts of real-time data fast and efficiently.

Sangfor, an Intel® Partner Alliance Titanium Tier member, worked with Intel to test several data analytics use cases based on the Sangfor XDR platform. The test results show that the introduction of Hyperscan with the ClickHouse database and accelerated with Intel® Advanced Vector Extensions 512 (Intel® AVX-512) can achieve up to around four times performance acceleration on 3rd Gen Intel® Xeon® Scalable Processor-based servers. Upgrading from 3rd Gen Intel Xeon Scalable Processor without Hyperscan to servers based on 4th Gen Intel® Xeon® Scalable Processor with Hyperscan can bring up to a more than six times performance boost[i].

Accelerating XDR Data Analytics using Hyperscan with ClickHouse

An XDR platform includes a database that integrates, correlates, and contextualizes data and alerts from multiple security prevention, detection and response components. XDR can be delivered on-premises or as a SaaS offering. ClickHouse is a fast open-source column-oriented database management system (DBMS) that generates analytical data reports in real-time using SQL queries. It has become a popular design choice for data analytics and search queries being used by Sangfor and other leading XDR solutions. In XDR, data analytics is a critical workload with ClickHouse consuming up to 40% of the overall computing resources in an XDR pipeline.

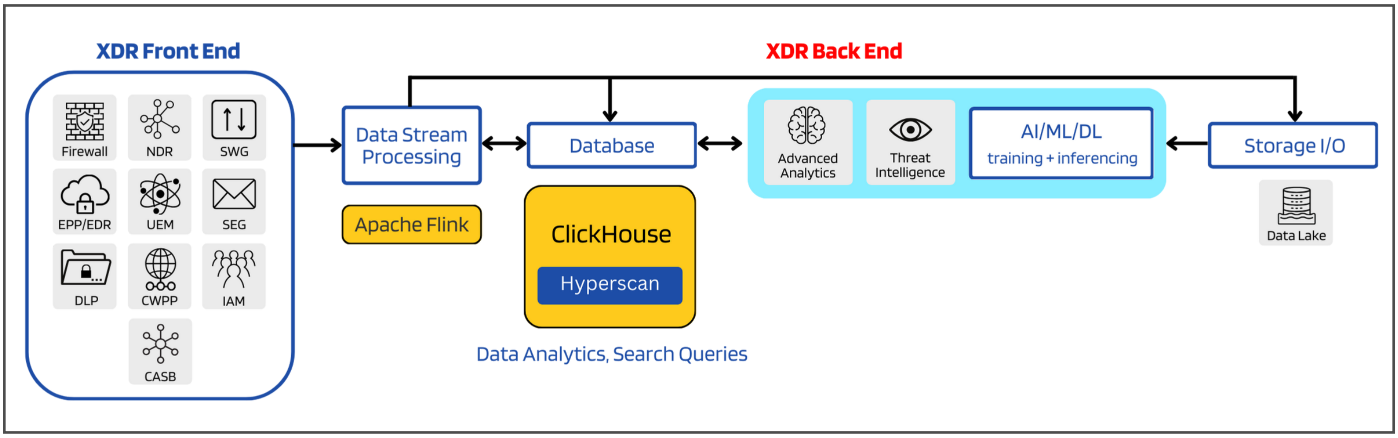

Figure 1. XDR architecture overview.

Hyperscan is an open source, high-performance multiple regex matching library developed by Intel that has been integrated with the latest version of ClickHouse. Enabling Hyperscan can speed up XDR data search queries to gain a significant performance improvement on Intel® Xeon® Scalable Processor-based platforms.

An XDR platform has a front end and a back end. The front end includes inputs from multiple security points, such as firewalls, secure web gateways (SWG), endpoint protection platforms (EPP)/endpoint detection and response (EDR), secure email gateways (SEG), and more. These are the data sources for the XDR system.

Typically, the front end will be monitored and the relevant data, such as server logs and security scanning results, will be sent to the XDR back end. The back end will ingest data from multiple front-end devices and is responsible for processing the data and performing advanced analytics to detect any anomalies occurred and further generate security responses or recommendations. The high-level overview of an XDR system is illustrated in Figure 1.

Sangfor Technologies is a leading vendor of cyber security and cloud computing solutions. The Sangfor XDR platform goes beyond traditional XDR by implementing a real integrated security solution, providing a holistic response to malware infections and APT breaches across the entire organization's network, with ease of management, operation, and maintenance.

Sangfor XDR core technical capabilities include:

- **XStream technology**: Sangfor's innovative XStream technology integrates multiple AI technologies, including an automatic ingestion engine, threat detection engine, etc. It supports 3,000+ data sources. When using XStream AI instead of traditional data ingestion solutions, data ingestion time is reduced from seven days to five minutes.
- **E+N+X deep correlation analysis technology**: Sangfor XDR uses E+N+X deep correlation technology to establish multi-source data correlation analysis, and to efficiently leverage the true value of the security data.
- **Security Orchestration, Automation, and Response SOAR**: Sangfor XDR features a built-in SOAR function that supports efficient automation of the security operations.
- **Open XDR technology**: Sangfor XDR leverages Open XDR technology to significantly improve the consistency of integrating with third-party components and native components.
- **Attack surface management (ASM) technology**: Sangfor XDR implements asset management and vulnerability management based on potential attack scenarios.

Sangfor XDR platform core engines and services:

- **Security GPT AI large language model (LLM)**: AI assistant for threat and attack detection. This large language model efficiently lowers the cost of security operations.
- **Managed detection and response (MDR)**: 24x7 year around on-line service, security monitoring and response service, based on big data infrastructure and AI algorithms.

Key Sangfor XDR capabilities:

- AI-enabled threat detection
- Significant noise reduction
- Innovative threat categorization
- Analysis and reconstruction of security events and attacks
- AI-assisted security operations
- Security GPT AI assistance tested in real-world environments and adopted by more than 100 enterprise customers
- Asset protection and risk management

While this paper focuses on using Hyperscan to accelerate XDR data analytics on Intel Xeon Scalable Processor-based platforms, these servers also have the capability to optimize the data stream processing component of the XDR pipeline. A previously published solution paper[1] has demonstrated that XDR data streaming workloads using Apache Flink software can achieve up to ~38% performance improvement on 4th Gen Intel Xeon Scalable Processor-based platform when upgrading from 3rd Gen Intel Xeon Scalable Processor-based server.[i]

Intel Xeon Scalable Processors

Intel Xeon Scalable Processors are engineered to deliver exceptional performance and efficiency for demanding workloads. These processors feature a robust architecture that includes a high core count, large cache sizes, and Intel AVX-512 advanced vector instruction set for accelerating Hyperscan. These features enable the CPUs to excel in tasks such as data analytics, artificial intelligence, and cloud computing. Additionally, Intel Xeon Scalable Processors offer built-in accelerators and optimized platforms to accelerate specific workloads, such as AI inference and training. These processors also prioritize security with hardware-based features that protect sensitive data and mitigate threats.

XDR Data Analytics using ClickHouse

In the XDR pipeline, real-time data needs to be stored in an in-memory database to perform advanced data analytics.

ClickHouse is an open-source columnar database management system (column-oriented DBMS) for online analytical processing (OLAP) that allows users to generate analytical reports using SQL queries in real-time. The typical data analytics workload can be performed with a ClickHouse client sending advanced search queries to the ClickHouse server. The communications between the ClickHouse client and the ClickHouse server use TCP locally or remotely. The general architecture of the ClickHouse search queries is illustrated in Figure 2.

Accelerating Data Analytics using Hyperscan

Hyperscan is a high-performance multiple regex matching library developed by Intel that is optimized for Intel Xeon Scalable Processors.

Hyperscan provides a flexible and easy to use library that enables the matching of large numbers of patterns simultaneously with high performance and good scalability, as well as providing unique functionality for network packets processing.

In ClickHouse, to utilize Hyperscan optimization for search queries, the underlying Hyperscan library needs to be invoked through the "**multiMatchAny()**" function API. These tests used the latest Hyperscan v5.6.1 (under the Intel's Outbound License.) Intel provides the software patch to apply Hyperscan v5.6.1 to the latest stable version of ClickHouse.

To use the Intel AVX-512 instruction set on 3rd Gen Intel Xeon Scalable Processors and newer Intel Xeon Scalable platforms, use the command below to change the gcc compile flag:

```
sed -i "s|-march=corei7|-march=icelake-server|g"
../hyperscan-cmake/CMakeLists.txt
```

Build ClickHouse by using the following command:

```
cmake -S . -B build

cmake --build build --target clickhouse
```

Examples of how to leverage Hyperscan to accelerate SQL queries are given below:

```
Query: select sum(multiMatchAny(URL,['/t[0-9]+-',
'/questions/7{9}[0-9]+'])) from datasets.hits_v1


Query: select count() from datasets.hits_v1 where
multiMatchAny(URL,['афиукд','берлик','fab','ru','www'
,'ьфьын','маиси','mam','amsy','маммси','амси','vfvc','/
t0-','/t1-','/t2-
','/questions/7777777770','faberl','febirl','фибер','ф
ибеп','фибел','фибэр','фибэп','фибэл','фибар','фибап','
фибал','/q0','/q1','/q2','/q3','/q4','/q5','/
questions/0','/questions/1','/questions/2','/
questions/3','/questions/4','/questions/5'])
```

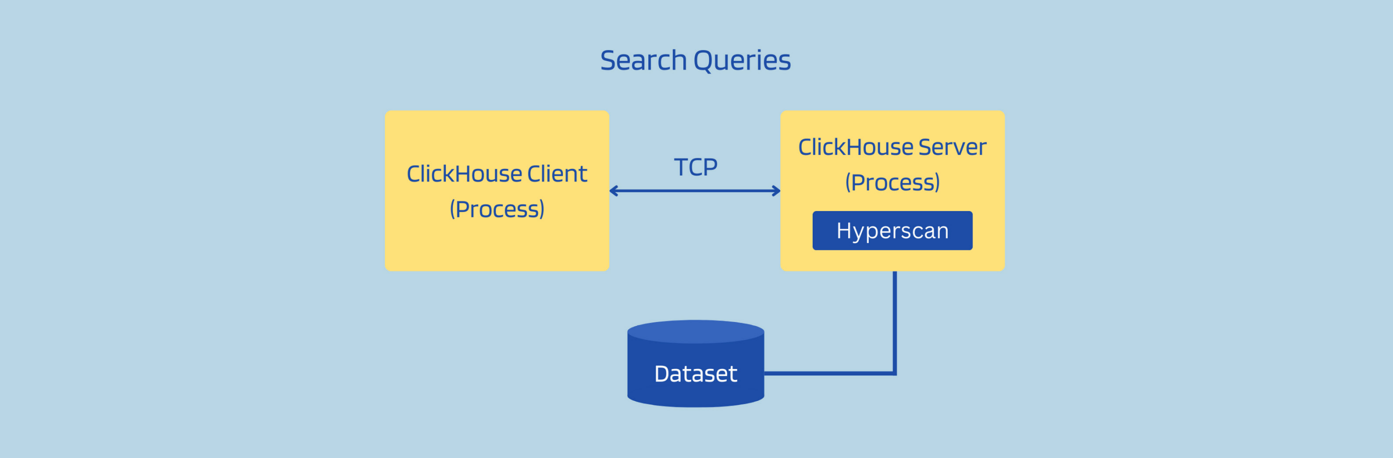

Figure 2. Data analytics using ClickHouse.

Test Case	Query Type	SQL
Q1	l1	`SELECT count() FROM table1 WHERE (url LIKE '%ghi%') OR (url LIKE '%jkl%') OR … (39 matching patterns)`
	h1	`SELECT count() FROM table1 WHERE multiMatchAny(url, [‘ghi’, ‘jkl’, ‘mno’ … ‘example5’])`

Table 1. Sample Sangfor XDR data analytics test case.

Test Set up and Performance Results

The Sangfor XDR team conducted a performance test using a customer dataset and customer-defined test cases, which were representative of a typical XDR data analytics workload. The use of real life data set makes the test results more credible.

Dataset and Test Cases

The customer dataset contains 286.6 million rows in the ClickHouse database with a total size of 85.52GB. Data analytics search queries were performed over the dataset using six customer defined test cases, a sample of which is shown in Table 1. In each test case, query type “l” represents the search query without using Hyperscan, while query type “h” represents the search query accelerated by Hyperscan.

The Hyperscan acceleration was enabled by invoking the “multiMatchAny()” function call.

Test Results

Search query latency performance evaluation results are shown in Figure 3. The baseline for the tests is a server based on the Intel® Xeon® Silver 4310 Processor that is a part of the 3rd Gen Intel Xeon Scalable Processor family. The average query latency of that server without the use of Hyperscan is about 28s. When Hyperscan is engaged on that server, latency is reduced by nearly four-times. Performance is even better using 4th Gen Intel Xeon Scalable Processor servers with Hyperscan. The company tested a server based on Intel® Xeon® Gold 5420+ Processor and saw a greater than six times latency reduction.

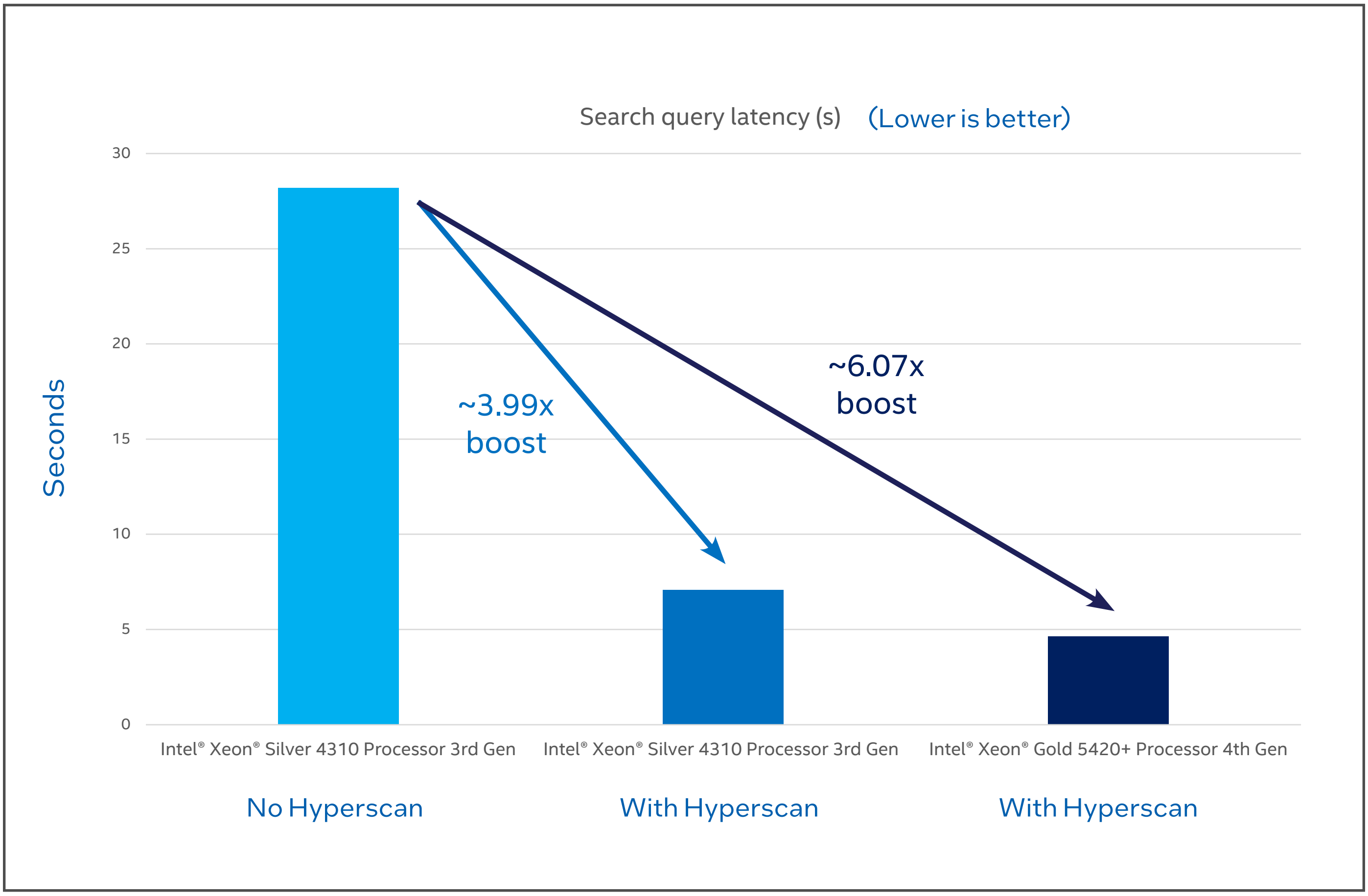

Figure 3. ClickHouse search query acceleration with Hyperscan on Intel Xeon Scalable platforms.

Conclusion

XDR is a rapidly evolving technology and key domain for network security. Data analytics is an essential pillar of an XDR workload. ClickHouse open-source software is a popular design choice for data analytics and search queries, being used by Sangfor and other leading XDR solutions. Hyperscan running on Intel Xeon Scalable Processors is a high-performance multiple regex matching library that can significantly accelerate search queries in ClickHouse.

In testing the impact of Hyperscan on various processors, these tests have shown that ClickHouse can achieve around four times performance improvement when accelerated by Hyperscan on 3rd Gen Intel Xeon Scalable Processors, and that upgrading to 4th Gen Intel Xeon Scalable Processor with Hyperscan results in a greater than six times performance boost.

Learn More

Sangfor Technologies

ClickHouse

Intel Hyperscan GitHub

Introduction to Hyperscan

Extended Detection & Response (XDR) - Accelerate XDR Data Streaming using Apache Flink* on Intel® Platforms Technology Guide

4th Gen Intel® Xeon® Scalable Processor

Intel Partner Alliance

Appendix - Terminology

Abbreviation	Description
CPU	Central Processing Unit
XDR	Extended Detection and Response
SQL	Structured Query Language
SWG	Secure Web Gateway
EPP	Endpoint Protection Platform
EDR	Endpoint Detection and Response
SEG	Secure Email Gateway
NDR	Network Detection and Response
UEM	Unified Endpoint Management
DLP	Data Loss Prevention
CWPP	Cloud Workload Protection Platform
IAM	Identity and Access Management
CASB	Cloud Access Security Broker

[i] Config1 (Intel® Xeon® Silver 4310 Processor -based SUT): 1-node, 1x Intel® Xeon® Silver 4310 Processor with 12 cores and 24 threads.; microcode is 0xd0003b9; Intel Hyperthreading - enabled; Turbo-enabled. Total DDR4 memory was 128 GB (3200 MT/s). OS: Ubuntu 22.04.6LTS; kernel is 5.4.0-176-generic; Benchmark/workload version: ClickHouse v24.4; Compiler: gcc 9.4.0 (Ubuntu 9.4.0-1ubuntu1~20.04.2); Library: Hyperscan v5.3. Test run by Sangfor on May 21, 2024.

Config2 (Intel® Xeon® Gold 5420+ Processor -based SUT): 1-node, 2x Intel® Xeon® Gold 5420+ Processor with 28 cores and 56 threads.; microcode is 0x2b0004d0; Intel Hyperthreading - enabled; Turbo-enabled. Total DDR5 memory was 256 GB (4800 MT/s). OS: Ubuntu 22.04.4 LTS; kernel is 5.4.0-176-generic; Benchmark/workload version: ClickHouse v24.4; Compiler: gcc (GCC) 12.3.0; Library: Hyperscan v5.3. Test run by Sangfor on May 21, 2024.

[ii] For more details, please refer to "Accelerate XDR Data Streaming using Apache Flink* on Intel® Platforms." https://networkbuilders.intel.com/solutionslibrary/extended-detection-response-xdr-accelerate-xdr-data-streaming-using-apache-flink-on-intel-platforms-technology-guide

Notices & Disclaimers

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

Performance varies by use, configuration and other factors. Learn more on the Performance Index site.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. No product or component can be absolutely secure.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

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

No product or component can be absolutely secure.

Your costs and results may vary.

0924/TM/DJA/PDF Please Recycle 361910-001US