HOW EXTRAHOP REVEAL(X) ENTERPRISE WORKS

BEHIND THE CURTAIN ON NETWORK DETECTION AND RESPONSE WITH EXTRAHOP REVEAL(X)

STREAM PROCESSING

Full-content analysis and decryption at the speed of the data center

The backbone of our technology is the real-time stream processor that transforms unstructured packets into structured wire data at up to 100 Gbps. Architected for parallel processing, the stream processor splits processing tasks across multiple computing cores—and will scale as more cores are added to new generations of server processors—so you get deeper insight at a fraction of the cost per Gbps of analysis compared to other real-time analytics platforms.

Reveal(x) offers both virtual and physical sensors that can be deployed at multiple points in order to provide a single, consolidated view of your environment. In highly virtualized infrastructures, Reveal(x) virtual appliances can capture inter-VM traffic, and virtual sensors are available for AWS and Azure. ExtraHop also integrates with Kubernetes and Open vSwitch, so whatever the nature of your architecture you'll be able to access full payload analysis with no added complexity.

Once the real-time stream processor receives a copy of network traffic from a tap or port mirror, here's what goes on beneath the hood:

1. Line-Rate Decryption

For encrypted traffic, the stream processor decrypts traffic at line rate, including Active Directory protocols and cipher suites that support perfect forward secrecy (PFS) such as TLS. This bulk decryption can scale to 64,000 transactions per second (TPS) using 2048-bit keys, which no other real-time analytics engine can match in a single unified offering. To learn more, read the white paper on how decryption is necessary for security visibility.

2. High-Performance TCP State Machines

Starting at the most fundamental level, the stream processor recreates the TCP state machines for every sender and receiver communicating on the network. A prerequisite for deeper application-protocol and universal payload analysis, this allows the platform to understand all TCP mechanisms and their impact. Because TCP is where the network and application meet, this approach helps you clearly identify whether problems are a network or an application issue right from the start.

3. Wire-Protocol Decoding and Full-Stream Reassembly

The real-time stream processor decodes IP-based protocols (skip to Protocols We Decode) in order to understand, define, and act on that protocol's unique application boundaries. This allows the processor to construct complete flows, sessions, and transactions for total application fluency, which in turn allows for higher-order content analysis through full-stream reassembly into wire data (derived from the wire protocol itself).

While in a perfect world this would all run pretty smoothly from start to finish, in reality traffic patterns like microbursts might result in packet loss from the tap or SPAN; in those cases the processor will automatically resynchronize and recover.

4. Full-Content Analysis

After reassembling packets into full streams, the stream processor analyzes the payload and content from layers 2-7, auto-discovering and classifying any device or client communicating on the network. The processor also continuously maps the relationships between all clients, applications, and infrastructure communicating on the network with over 4,700 metrics measured and recorded out-of-the-box.

Full-content analysis supports dozens of protocols, providing key performance indicators such as database methods used and their process time, file access by user, storage access time and errors, DNS response time and errors, web URI processing time and status codes, SSL certificates with expiration, and load-balancer and firewall latency. The stream processor also gathers sophisticated network metrics such as receive-window throttles, retransmission timeouts, and Nagle delays.

We get that not everyone is interested in knowing every detail about every layer of their environment, however, so don't worry—while the full analytics capabilities are always available to you, it's also easy to tailor your experience so you only see the precise metrics and insights you need.

5. Fully Programmable Insights

Once the stream processor has done its thing and begun supplying wire data metrics, it's time to take control of which insights you see and at what depth.

ExtraHop uses an event-driven programmable interface called Application Inspection Triggers to connect you to the stream processor and all stream transactions. Triggers allow you to programmatically extract wire data events and correlated metrics that are specific to your business, infrastructure, network, clients, and applications.

With Application Inspection Triggers, you can be as surgical or as verbose as you want and extract nearly anything from a header to the full application payload. For example, with HTTP payloads, this data can include revenue, order IDs, unique user IDs found in cookies or URIs, and even titles for web pages or error descriptions embedded by a developer in a 500 status code. And it doesn't matter what's traversing HTTP—it could be SOAP/XML, REST, JSON, AJAX, JavaScript, or HTML5.

The same principle and functionality holds true for all of our natively decoded protocols. You can also use triggers to extract, measure, and visualize data from defined fields, or to decode proprietary protocols based on TCP and UDP.

MACHINE LEARNING

Cloud-scale behavioral analytics guided on wire data metrics

Our cloud-scale machine learning service tracks detections in eight categories across your environment:

- Authentication, authorization, and access control
- Network file system
- Network infrastructure
- Database
- Email server
- Web server
- Remote access servers and methods
- Internet Communications and Telephony

Within each of these categories, our ML evaluates several protocols and hundreds of ExtraHop metrics, all with custom logic, in order to find and correlate active problems.

Architecture Overview

Unlike typical SaaS solutions, with our machine learning service only de-identified metadata is sent to the cloud. That means no payloads, filenames, strings, or other data categories that could contain sensitive data will leave your premises. ExtraHop has received SOC 2, Type 1 compliance certification for our machine learning tech, which you can learn about on our Compliance page.

ExtraHop uses the following combination of on-premises tech and cloud services to support the full ML process:

1. An on-prem device, controlled entirely by you, analyzes network traffic to extract and store 4,500+ metrics including IP addresses, URIs, database queries, CIFS filenames, VoIP phone numbers, and other potentially sensitive data; you can configure this device to collect custom metrics as you choose.
2. When the ML service is enabled, a subset of these on-prem metrics are de-identified and sent to a customer-dedicated cloud-computing instance in Amazon Web Services, which is operated by ExtraHop.

3. ExtraHop ML then builds predictive models for how we expect devices and applications to behave, and detects significant deviations from these predictions as anomalies.
4. Anomaly events are sent back to your on-premises device, although you can also opt-in to receive email alerts (which don't include any sensitive data). Once events are back inside your environment, you can re-identify and decrypt them with your private key for alerting and investigation.

Algorithms in Use

Reveal(x) includes the following machine learning algorithms to provide advanced network traffic analysis:

- Hundreds of self-adaptive unsupervised attack-detection models leveraging proprietary time series analysis and outlier detection algorithms
- Inference engine for inferring entity importance and entity network privilege level based on observed behaviors and innovative graph analytics
- Entity clustering engine for identifying behaviorally similar devices
- Peer group outlier detection engine
- Risk score estimation engine that combines domain expertise and customer base telemetry

Why the Cloud in Cloud-Scale Matters

Read this blog from an ExtraHop engineer to learn more about the unique benefits of our machine learning service as well as how Reveal(x) uses cloud-scale ML in practice.

DATA INDEXING AND STORAGE

Three complementary formats to index and store your data

ExtraHop uses three complementary formats to index and store your wire data:

1. Correlated, cross-tier metrics in the streaming datastore

Optimized for time-sequenced telemetry, the streaming datastore enables customizable dashboards that can be populated with more than 4,700 possible metrics in real time. This way you can easily see all communications across your entire environment, or narrow your focus to specific datasets.

As metrics are indexed in the datastore, newly discovered devices are automatically classified based on heuristic analysis of machine information and behavior, and ExtraHop begins building activity baselines for all systems, applications, and networks.

You can use your existing NAS infrastructure to extend the streaming datastore for long-term lookback, which is helpful for capacity planning, proving compliance efforts or

continuous improvement, and analyzing business activity over time. By default, your datastore will store fast (30-second), medium (5-minute), and slow (1-hour) metrics locally. You can, however, store 5-minute, 1-hour, and 24-hour metrics externally.

The datastore also allows you to create alerts based on current or past behaviors and events such as unusual payload size or expiring SSL certificates.

2. Transaction, message, and flow records

ExtraHop allows you to conduct multidimensional analysis of your wire data even if you don't know any query languages. Think of this like the search capabilities you'd find in a log analytics platform, except you're searching and analyzing wire data—a much richer, more consistent, and more reliable source of information than machine logs can provide.

There are two basic types of records in the ExtraHop UI: flow and transaction. Flow records show network-layer communications between two devices over an (L3) IP protocol, while L7 records show details from individual messages or transactions over any of the three types of supported L7 protocols (transactional, message-based, and session-based). ExtraHop allows you to search and filter for L7 traffic only, or to query both flow and L7 records.

You can learn how ExtraHop collects and stores built-in records, as well as more details about record types and formats, over here in our documentation.

Your transaction, message, and flow records are all stored in a resilient cluster built on scalable Elasticsearch technology so you can easily add nodes as your data grows.

3. Packets for the full payload

You can either begin with individual metrics, users, devices, or packets associated with a particular transaction, or easily drill down to that information from a high level view. ExtraHop supplies packets that offer the full payload, which you can download and analyze further as needed.

ExtraHop enables extended forensic lookback at a much more affordable price than any other real-time analytics platform, considering you can add up to 1920 TB of extended storage units per deployment with zero data tax from us.

DATA VISUALIZATION AND EXPLORATION

Intuitive querying, live activity maps, and open data stream

One of the most challenging aspects of real-time analytics at enterprise scale is, well, the scale itself. At ExtraHop, we do our best to make this easy for you as a user to parse the

immense wealth of information that is wire data and derive meaningful insights no matter which perspective you're coming from.

We start you off with a simple, intuitive user interface that includes automatically populated role-based dashboards for teams across your organization. These dashboards function on a drag-and-drop model so you can customize them further with unique widgets; if you want to create your own widget, all you have to do is select your desired data source and metrics, pick a visualization type, and save it to your dashboard of choice. You can quickly and easily export charts and background data points to PDF, Excel, or CSV.

No Scripting? No Problem.

Our visual query language gives you the power to refine or change your search queries by clicking UI elements that control everything from grouping, to filtering, to time-range selection. Whether you stick with the hundreds of built-in record attributes or branch out and define your own, this functionality means any user can quickly answer performance and security questions without needing to learn a query language.

For example, if you sort SQL messages by query string, you can identify attempted SQL injection attacks, pinpoint the malicious IP, and then instantly pivot to see all the activity of that client on the network over a given length of time. By exporting that information to Excel, CSV, or a visualization tool like Tableau or Qlik, you'll have a step-by-step map of what the attacker did so you can both respond with confidence and be better equipped to prevent future attacks.

Live Activity Maps

Along with traditional methods of data visualization like charts and graphs, ExtraHop uses live activity maps to present a dynamic and intuitive view of your environment. Instead of manually creating and updating network diagrams as your IT environment changes, you can use live activity maps to visualize protocol-based connections between devices and applications in real time.

By allowing you to filter by time interval and broaden or narrow your scope as needed, activity maps make it easy to answer multi-part questions like, "How are devices interacting within a certain tier, and how have those devices been interacting across the network in the last hour?" Anomalous behavior detections also appear on live activity maps, so you can see the context of the detection before clicking down into the transaction or even into the precise packets straight from the map.

This blog post goes into a lot more detail about the latest capabilities of live activity maps and provides some more ideas about how you might use them in your day to day.

Open Data Stream

While we supply rich query and investigation workflows within the ExtraHop interface, we also make it easy for you to integrate wire data metrics with the other data stores, querying tools, and analytics platforms in your stack. Open Data Stream allows you to

merge data from multiple sources into a single, rich set that can be queried and visualized using whatever tools your team prefers.

Visit our Technology Partners page to learn about specific integrations such as our partnerships with Phantom, Elastic, MongoDB, and many more.

PROTOCOLS WE DECODE

Over 70 enterprise protocols decoded in real time

ExtraHop decodes the following enterprise protocols with real-time fluency at the application layer. Protocol modules offer varying levels of analysis, starting with L7 classification, and Application Inspection Triggers allow you to create a custom metric.

AAA: RADIUS	ActiveMQ	HTTP-AMF
AAA: DIAMETER	AJP	DSCP
CIFS	DICOM*	iSCSI
Citrix ICA*	HL7*	MS-RPC
DHCP	LLDP	WebSocket
DNS	MSMQ	SSL
Database: DB2	Telnet	SSH
Database: Informix	SMTP	SMPP*
Database: Microsoft SQL	VoIP: SIP*	VoIP: RTP*

Database: MongoDB	VoIP: RTCP XR*	VoIP: RTCP*
Database: MySQL	Redis	POP3
Database: Oracle	NFS	Memcache
Database: Postgres	LDAP	Kerberos
Database: Sybase	VNC	Syslog
Database: Sybase IQ	ICMP	IBM MQ
FIX*	STP	SNMP
FTP	HTTP/S	RDP
SMBv3	MSRPC	WINRM

** Add-on module (not included in base license)*

Of particular interest to SecOps analysts, Reveal(x) analyzes application-layer metadata for databases, Active Directory, web, SSL, and storage systems:

Database: RDBMSs: Oracle, Microsoft SQL Server, MySQL, PostgreSQL, Informix, Sybase, and DB2. NoSQL databases: MongoDB, Memcached, Redis, Riak. Metadata extracted include transaction timing, table/user access patterns, query errors, SQL queries and responses, and system-level commands.

Identity and Access Management: Active Directory visibility, including NTLM, Kerberos, LDAP, MSRPC, WINRM, SMBv3, and DNS monitoring for privileged identities and service accounts allows you to improve detection and facilitate audits. Reveal(x) extracts metadata including user/computer account activity, invalid or expired passwords, new privileged access, privileged access errors, DNS SRV lookups, LDAP binds, plain-text HTTP authentications, unknown SPNs, and forged Kerberos ticket detection.

Web Transactions: Full HTTP payload analysis of user activity, SOAP/XML, JSON, Javascript, APIs, etc. Extracted metadata includes URI, query parameters, host headers, and user agent, among others.

Storage: Metadata extraction for all NAS and SAN transactions (iSCSI, NFS, and CIFS) enables machine learning detections based on actual file details and equips security analysts to track file access patterns and detect ransomware activity by examining file extensions and WRITE operations.

THREAT INTELLIGENCE

Real-time correlation of wire data metrics and threat intelligence information

ExtraHop Reveal(x) correlates real-time network data with threat intelligence information shared in the industry standard Structured Threat Information eXpression (STIX) file format to help you view, prioritize, and respond to indicators of compromise.

We encourage you to curate your own threat feeds but ExtraHop also maintains a list of recommended free feeds which can be added to each Reveal(x) deployment, including Anomali Limo, AlienVault OTX, and the SANS Top 100 Suspicious IP Addresses feed.

Using STIX Files in Reveal(x)

ExtraHop currently supports STIX 1.0 - 1.2. Once you upload the STIX files you care about the most to your Reveal(x) system, Reveal(x) will match the following observables (schema components specifying a suspicious object) to objects found in your wire data in real time:

- Hostnames
- IP addresses
- URIs

Reveal(x) provides visual context for specific observables by displaying indicators of time range or information source as well as indicators of compromise within the ExtraHop Web UI. Correlations also automatically contribute to the risk score calculus within Reveal(x), providing a rich launch point for triage and threat hunting. Learn more about STIX and Reveal(x) here.

The Security Dashboard

Your Security dashboard contains protocol metric data organized by region. You can create your own custom dashboard with metrics that are specifically relevant to you, or you can stick with our built-in dashboard for the following information:

- A link to the Security Overview page where you can find high-risk detections, trending metrics, and activity maps with network activity by protocol
- Recent alerts in your environment
- Suspicious hostnames, IP addresses, or URIs from uploaded threat intelligence information
- Weak SSL ciphers along with which clients and servers are participating in affected sessions
- Self-signed, wildcard, expired, and soon-to-expire SSL certificates
- Which DNS servers are most active plus the total number of reverse DNS lookup failures those servers have encountered