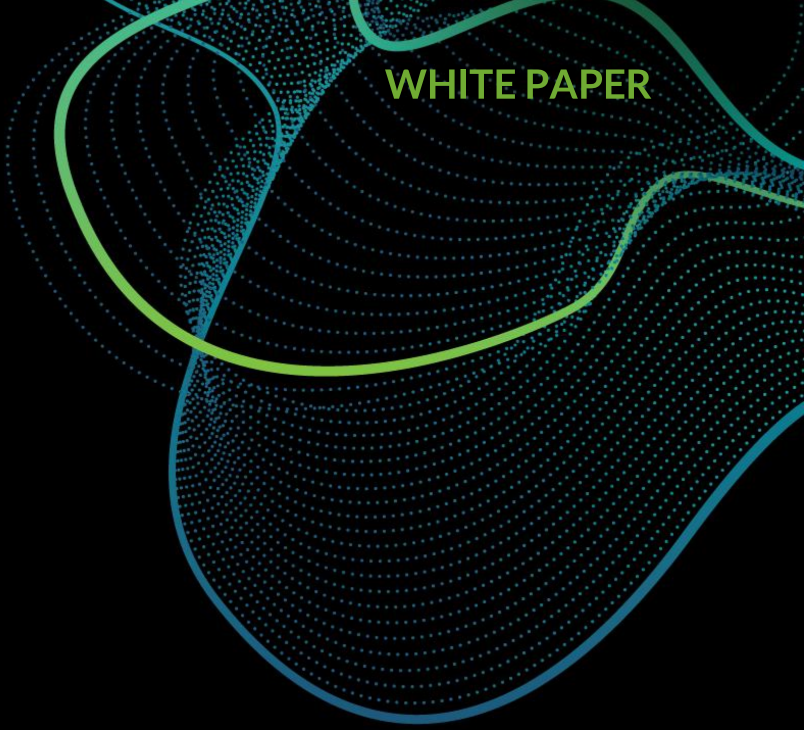
WHITE PAPER



Reveal(x)

Technical Architecture

Under the Hood

ExtraHop Reveal(x) takes advantage of proven streaming analytics technology and advanced machine learning to provide unprecedented visibility, immediate insights, and definitive answers. This white paper details the technical underpinnings of the Reveal(x) architecture and explains how these design concepts enable radically more efficient Security Operations.

TABLE OF CONTENTS

Overview

Modern Security Operations teams have access to vast amounts of data, but this has not translated into greater effectiveness. The goal of ExtraHop Reveal(x) is to not only eliminate blind spots with unprecedented visibility, but to also cut through the noise of alerts with high-fidelity behavioral analytics. In addition, Reveal(x) dramatically reduces the time required to take action, from days to seconds, through automated investigations. To achieve these goals, Reveal(x) takes advantage of new machine learning and network traffic analytics technology.

At its core, Reveal(x) is powered by an open, programmable, and extensible real-time streaming analytics platform and cloud-based behavioral analytics. This powerful combination of stream processing and machine learning technology allows organizations to mine insights about their security and operations from all the data in motion within the IT environment. By focusing advanced behavioral analytics on critical assets throughout the security operations workstream, the Reveal(x) platform replaces, automates, or minimizes much of the effort and delay of network threat detection, investigation, and response.

The ExtraHop streaming analytics engine transforms unstructured packets into structured wire data at line rate, up to 100Gbps, meaning that it can keep up with modern enterprise networks and datacenter speeds. This data, classified and indexed, is analyzed along many dimensions, yielding high-fidelity insights about threat activities and unusual, potentially malicious behavior. Reveal(x) then enriches the original data and insights with asset intelligence, threat intelligence, and risk context.

As they investigate detections highlighted by Reveal(x) or their own hunches, analysts navigate rapidly through investigation to root cause using optimized workflows and on-demand forensic data. Partnerships and an open data platform support integrated interactions to place ExtraHop Reveal(x) within a productive workstream of detection, containment, response, and forensics.

The design of Reveal(x) provides a highly scalable architecture, improves productivity, and reduces operational expense. For instance, the product relies on an “analysis first” approach that applies deep analytics to all network traffic without having to store packets “just in case.” By using this metadata for analysis, we can dramatically increase the amount of practical lookback without a heavy storage penalty. Extensive automation also decreases errors and latency, assisting in the move to real-time operations.

This paper provides technical details, architectural concepts, and some key ideas to help you understand Reveal(x) technology.

Unique Network Traffic Analytics

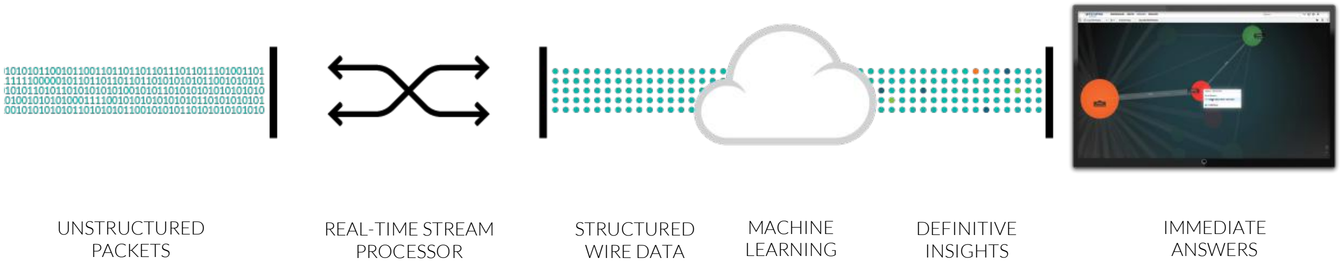
ExtraHop Reveal(x) network security analytics observes north-south (external-internal) and east-west (internal only) network traffic to provide visibility, detect suspicious behavior, and improve investigations. Several parts of our approach are unique and patented, contributing to processing speed and capacity that network traffic analytics (NTA) models have not approached:

- **Relevant Insights:** Reveal(x) captures detailed data directly from your enterprise about the systems, services, and users currently active in your environment. By concentrating analytics and resources on the assets critical to your business, Reveal(x) helps you focus on the threats and issues most relevant to your organization's risk, infrastructure, and posture.
- **Timeliness:** Reveal(x) provides real-time visibility and analysis of traffic in flight, taken directly from the network itself. Network traffic is analyzed in memory through stream processing, not written to disk first. This "analysis first" model permits immediate detection and response, instead of requiring retroactive data correlation and post hoc queries.
- **Accuracy:** Reveal(x) employs agentless asset discovery and classification with accurate feature extraction from live traffic through fluency in 50+ L2-L7 data center protocols. The consistent structure of these protocols allows feature extraction to be consistent across all assets that utilize them, so the analysis and conclusions are maximum fidelity.
- **Reliability:** With a passive, out-of-band deployment, Reveal(x) provides objective observation of network transactions results—empirical and immutable evidence of asset behavior. Since no *a priori* assumptions are made about what characteristics an asset should have or the relationships between assets, the results of this analysis are inherently complete.

Extracting Wire Data at Scale

Reveal(x) performs passive analysis, avoiding any disruption of network traffic or the burden of agents. Upon receiving a copy of network traffic from a tap or port mirror, the stream processor performs line-rate decryption (optional), multi-protocol decoding, and full-stream reassembly to reconstruct every conversation taking place in the data center and extract relevant metadata—also known as *wire data*—at sustained rates of 100 Gbps. This processing engine is the key enabler for decryption at scale, feature extraction, imposing structure on highly unstructured packet data, and committing the data to the data store. As Reveal(x) works, it streams structured transaction records into the index for free text search, and also captures the packets in parallel to all of this.

The ExtraHop Reveal(x) platform architecture is optimized for parallel processing. That means that the real-time stream processor efficiently splits the task of processing the streams across multiple computing cores, and it will scale as cores are added to new generations of server processors. The result: customers get deeper and more meaningful insight at a fraction of the cost per Gbps of analysis compared to other real-time analytics platforms.



What Gartner Says About Wire Data

At the core of the product is a real-time stream processor that recreates TCP state machines for every client and server and then reassembles the complete sessions, flows, and transactions to extract L2–L7 metrics. The resulting information is a crucial data source that Gartner Research calls wire data.

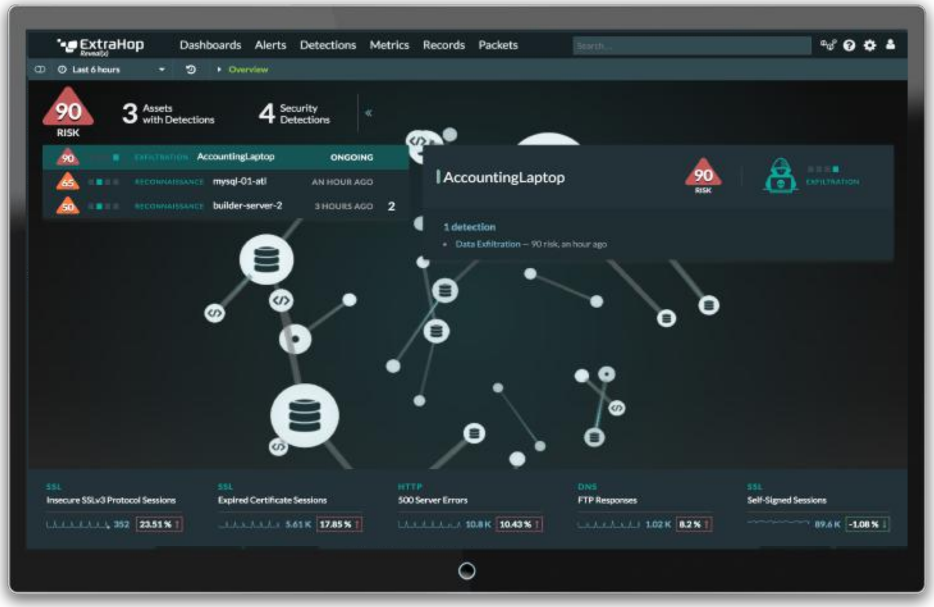
"While log data will certainly have a role in future monitoring and analytics, it is Wire data—radically rethought and used in new ways—that will prove to be the most critical source of data for availability and performance management over the next five years."

Use Data- and Analytics-Centric Processes with a Focus on Wire Data to Future-Proof Availability and Performance Management, Will Cappelli | Vivek Bhalla, Gartner, March 2016

Machine Learning for Behavioral Analysis

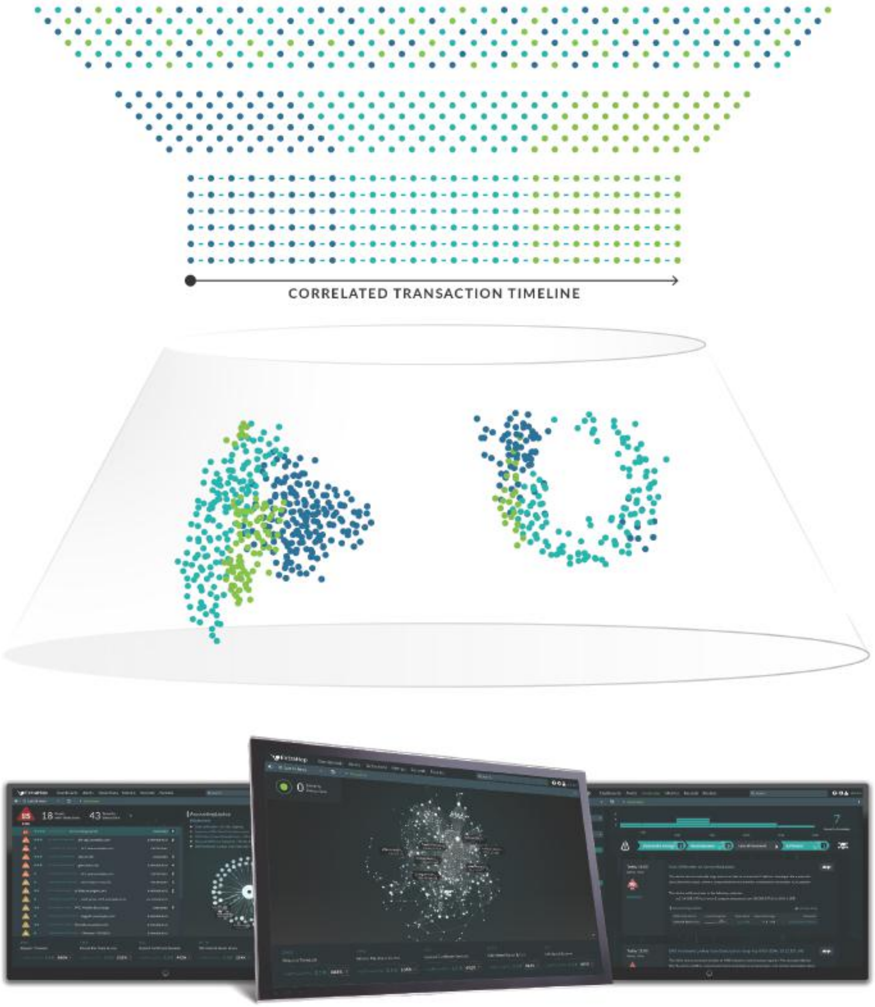
ExtraHop Reveal(x) goes beyond rule-based detection and statistical modeling to perform behavioral analysis using machine learning. ExtraHop's advanced machine learning is a highly secure and scalable technology that uses strong detectors based on dimensionality reduction and outlier detection to identify new and suspect behavior in real time.

Essentially, these techniques help us build more effective machine learning models. Scientists look at the features (also called dimensions) that are available in the data and assess the fit and impact of each feature on the model's results. The features with the greatest impact are used in the model so analytics processes can be most efficient.



Machine learning enables Reveal(x) to surface relevant behavioral anomalies that affect your most critical assets.

The rich metadata we extract from the network provides more than 4,600 features to train our machine learning models. Since 2014, our data scientists and cybersecurity experts have been refining the systems for the highest precision. In addition, users can provide feedback on detections to continually improve accuracy. The wire data produced by the the Reveal(x) platform provides excellent structure for advanced analysis.



DATA

Decryption
50+ Protocols
L2-L7 Classification
Reconstructed
Sessions
4600+ Features

MATH

Feature Engineering
Dimensionality
Reduction
Outlier Detection

RESULTS

Unprecedented
Visibility
Definitive Insight
Immediate Answers

ExtraHop Reveal(x) appliances send a relevant and pseudonymized subset wire data to the cloud for detection of behavioral anomalies. This combination of network appliances with cloud-based machine learning is a deliberate design choice, and overcomes the problem of "data gravity" by extracting the machine learning features near where the data is generated (on the network) and then sending that lightweight metadata to where scalable computing resources exist (in the cloud). Other cloud-based security analytics products require organizations to send raw data up to the cloud for anomaly detection, which limits scale. In contrast, the Reveal(x) architecture is highly scalable and can apply a suite of machine learning algorithms to more than a petabyte of data each day, given that a single appliance can analyze a sustained 100 Gbps of traffic.

Each customer has their own dedicated cloud instance, so no data is commingled. In addition, data is pseudonomized on-premises before sent to the cloud and then re-identified once the anomalies are sent back to the appliance to protect sensitive information and customer privacy, a critical step in meeting global data privacy regulations like GDPR.

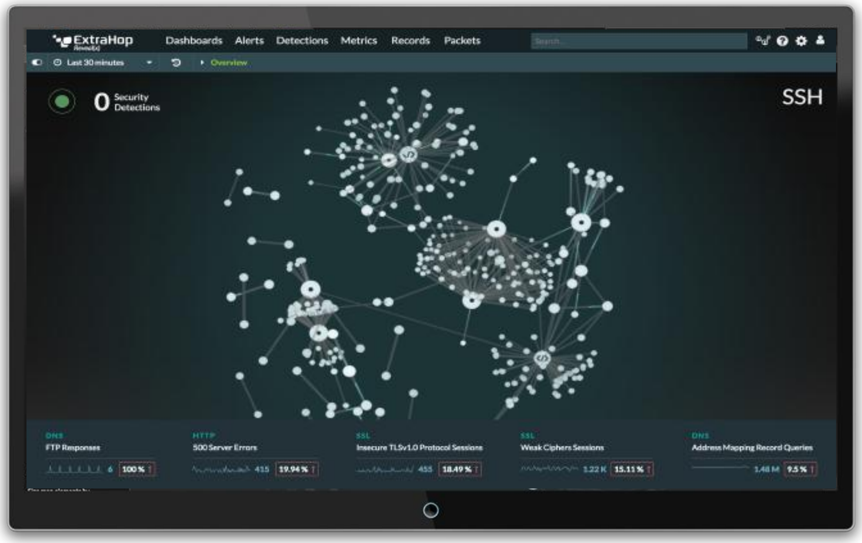
Dynamic Discovery, Classification, and Mapping

As metrics are indexed, Reveal(x) classifies newly discovered devices based on heuristic analysis of machine information and behavior. For example, if a machine responds to database requests, then it is classified as a database server. This automated model continuously validates the completeness of your understanding of the environment, using real-time traffic analysis to provide objective and consistent reporting of asset behavior, regardless of the asset's ability to self-report data.

Keeping an up-to-date inventory of assets can be difficult using traditional methods. Reveal(x) can stream data about newly discovered devices to CMDBs and other asset inventory systems to keep them up to date, and enable SecOps teams to identify unauthorized activity, such as "Shadow IT" SaaS applications, rogue DNS servers, cryptocurrency mining, and botnet operations.

The platform automatically builds activity baselines for all systems, applications, and networks. The platform also continuously maps the relationships between all clients, applications, and infrastructure communicating on the network. A live activity map visualizes this information and lets you drill down to devices, individual transactions, and even the exact packets in just a few clicks.

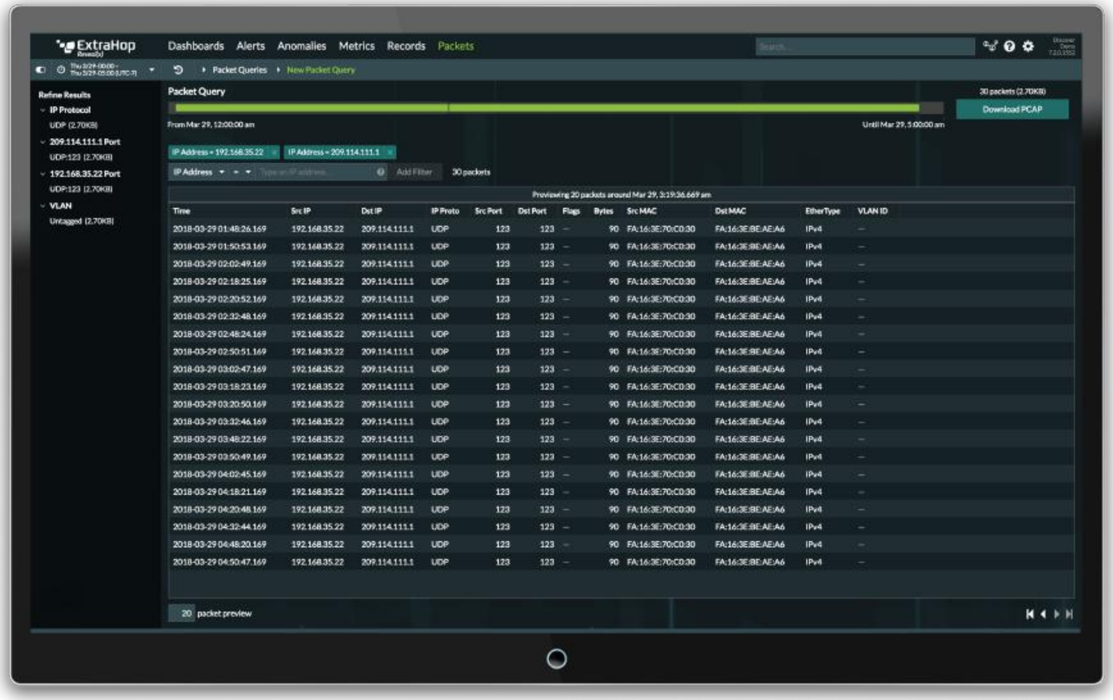
You can create alerts based on behaviors and events that are indexed and stored, either now or in the past. These can be based on behaviors like anomalous network activity, error messages, unusual payload size, or expiring SSL certificates.



With live activity maps in Reveal(x), you can easily compare time periods to discover which device communications are new.

Enabling Natural and Rapid Workflows

Traditionally, security teams have relied on packet capture technologies to extract insight from the network. In the standard approach, packets are recorded directly to storage and then analyzed later using a packet analysis tool such as Wireshark. Since only experienced security analysts and network engineers are trained to interpret this data, packet analysis has become a bottleneck for efficient operations. However, packets are still regarded as the most granular and empirical evidence that something happened.



Packet Query

30 packets (2.70KB)

From Mar 29, 12:00:00 am

Until Mar 29, 5:00:00 am

Refine Results
- IP Protocol
  - UDP (2.70KB)
- 209.114.111.1 Port
  - UDP:123 (2.70KB)
- 192.168.35.22 Port
  - UDP:123 (2.70KB)
- VLAN
  - Untagged (2.70KB)

IP Address = 192.168.35.22 × IP Address = 209.114.111.1 ×

IP Address = 30 packets

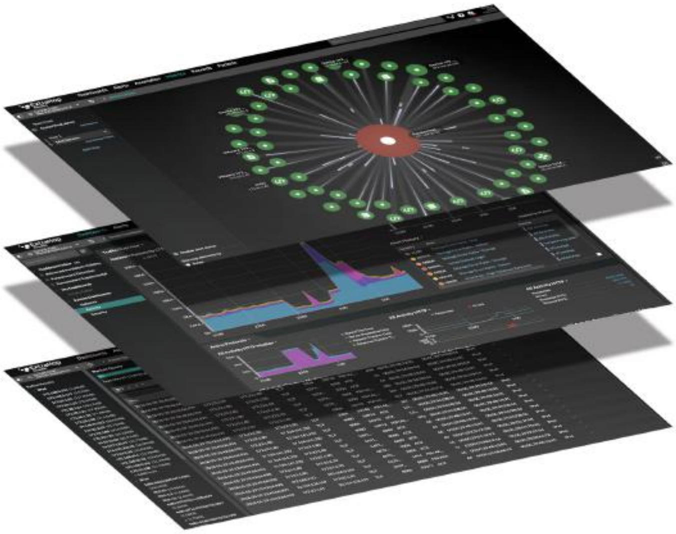
Previewing 20 packets around Mar 29, 3:19:36.669 am

Time	Src IP	Dst IP	IP Proto	Src Port	Dst Port	Flags	Bytes	Src MAC	Dst MAC	EtherType	VLAN ID
2018-03-29 01:48:26.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 01:50:53.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 02:02:49.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 02:18:25.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 02:20:52.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 02:32:48.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 02:48:24.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 02:50:51.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 03:02:47.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 03:18:23.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 03:20:50.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 03:32:46.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 03:48:22.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 03:50:49.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 04:02:45.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 04:18:21.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 04:20:48.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 04:32:44.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 04:48:20.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—
2018-03-29 04:50:47.169	192.168.35.22	209.114.111.1	UDP	123	123	—	90	FA:16:3E:70:C0:30	FA:16:3E:BE:AE:A6	IPv4	—

20 packet preview

In the screen shown above, Reveal(x) displays the 30 packets that comprise a specific transaction. Normally, isolating these packets would involve time-consuming and skilled investigation.

ExtraHop turns the packet-first approach on its head. Instead of analyzing only a small segment of network traffic after it has been recorded to disk, the ExtraHop platform analyzes all network traffic as it passes over the network and then records the key insights—port scans, remote login failures, database methods used, files transferred, and much more—in a format that everyone can understand. This analysis of data in motion as opposed to data at rest is crucial, because it enables much more natural and rapid workflows for analysts. Within five clicks, users can drill down from high-level dashboards to see transaction records and then download the packets that comprise those transactions.



Deep Dive: Mining Network Data for Security

Reveal(x) reconstructs and indexes all network transactions in real time to make them immediately searchable and useful for analytics. The system carries the burden for data collection, parsing, normalization, correlation, and retrieval efforts for every device communicating on the network, effectively automating many stumbling blocks to effective SOC operations.

A Reveal(x) deployment will include a range of physical, virtual, or cloud-based appliances connected to a cloud-based machine learning system. Each appliance performs a different function to index and store your wire data in three complementary formats:

- **Correlated, cross-tier metadata is produced by the ExtraHop Discover appliance (EDA)**, featuring a streaming datastore that is optimized for time-sequenced telemetry. More than 4,600 metadata metrics are used to perform statistical analysis and also provide features for cloud-based machine learning models. In addition to these out-of-the-box metrics, you can define custom metrics gathered by the parsing engine. Metrics are indexed and made available for analysis in real time, providing analysts with immediate visibility into all communications across the entire environment through dashboards and alerts. The appliance is configured for a minimum of 30 days storage for metadata lookback, and external network storage can be used for additional lookback.

- **Transaction, message, and flow records are created in the ExtraHop Explore appliance (EXA)**. Built on scalable Elasticsearch technology, the Explore appliance indexes the metadata in a noSQL datastore that allows you to use free text search as well as a visual query language to conduct multidimensional analysis of your wire data. This record search and query ability is similar to what is provided by log indexing and analysis tools, except applied to transactions observed on the network instead of logs recorded by systems.

- **Forensic evidence in the form of packets are captured by the ExtraHop Trace appliance (ETA)**. Our purpose-built design performs continuous packet capture, correlating wire data metrics with the underlying packets in real time. When you want the digital forensic evidence for root cause analysis or to fulfill legal chain of custody requirements, you can click through to select just those packets. You can also compose a new packet query, filtering down to just the kilobytes of packet capture you care about.

Solutions for Every SOC

Reveal(x) includes the infrastructure required for specific use cases that align with SOC maturity, sized and licensed using a subscription model based on critical assets being analyzed:

- **Standard:** Discovery, monitoring, and incident response for modest security programs

- **Premium:** Adds decryption and integration support for SOCs with SIEMs and other formalized tools

- **Ultra:** Adds packet capture for forensic and threat hunter teams

The collection system can be placed anywhere in your environment: at the core switch, in the DMZ, in the cloud, or at branch offices. Larger deployments can take advantage of clustering, redundancy, centralized management, and other enterprise-class features. [Learn more about deployment and detection details] For larger distributed ExtraHop deployments, an optional dedicated Command appliance is the most efficient way to manage all of your remote appliances.

Discovery, Standard and Advanced Analysis

Reveal(x) discovers and classifies every device observed on your network, providing non-stop accuracy and visibility by automatically identifying new devices as they start to use the network. The base Discovery mode analysis is applied to observed devices and captures records, packets, and information about protocol activity.

In addition, two classes of analysis can be performed based on the critical asset value of the devices using your network. The difference between the two additional analytic levels is the detail in the data analyzed, either L2–L7 or L2–L3.

- **Advanced Analysis** calculates and presents charts with L2–L7 protocol metrics, records, packets, activity maps, and information about protocol activity. Advanced Analysis is commonly applied to databases, sensitive laptops, DNS servers, Active Directory servers, and other devices that contain or have access to high-value data.
- **Standard Analysis** produces charts with L2–L3 metrics, records, packets, activity maps, and information about protocol activity. Standard Analysis is helpful for devices that you want to map communications for, such as employee workstations.
- **Discovery Mode** collects records, packets, and information about protocol activity for all devices observed on the network. Discovery Mode will allow you to see all devices on the network and drill into the packet detail of their communications.

Your Reveal(x) solution will be sized based on the total Advanced Analysis capacity. You can programmatically target specific device groups or activity groups for Standard Analysis as needed, based on their importance to your network. For example, you can rank groups in an ordered list to let Reveal(x) know which devices are the most important to you. To maintain the accuracy of the analytics, you can define a watchlist and the system will automatically and dynamically prioritize newly discovered individual assets that are classified as that asset type for Advanced Analysis.

Asset values vary by business, risk of attack, and utility in attack techniques. Most companies will have a range of device groups and activity groups designated for advanced analysis.

Up-level Your SOC with ExtraHop Reveal(x)

From real-time processing and full-stream reassembly of wire data to optimized investigation workflows and automated response, Reveal(x) offers potent technology for demanding enterprise security operations. Our design reflects scale, manageability, efficiency, and flexibility to help you modernize and enhance your SOC at your pace. See for yourself: Extensive demos, walkthroughs, and documentation can help you dig into the details. Find it all at extrahop.com/revealx.

ABOUT EXTRAHOP
ExtraHop is the first place IT turns for insights that transform and secure the digital enterprise. By applying real-time analytics and machine learning to all digital interactions on the network, ExtraHop delivers definitive insights and instant answers that help IT improve security, performance, and the digital experience.

ExtraHop Networks, Inc.
520 Pike Street, Suite 1600
Seattle, WA 98101 USA

http://www.extrahop.com/
info@extrahop.com
T 877-333-9872
F 206-274-6393