Introduction to the ExtraHop system

Published: 2020-02-24

The ExtraHop system helps you monitor network activity and all of your applications. For example, you can learn how well applications are consuming network resources, how systems and devices are communicating with each other, and how to identify transactions that are flowing across the data link layer (L2) up to application layer (L7) in your network.

This guide explains how the ExtraHop system functions so that you can understand how your data is collected and analyzed. We also provide a list of learning resources and some activities to get you started.

- First, learn about our appliances and how they work together.
- Then, learn how the Discover appliance collects data from transactions observed on your wire data capture feed or from machine data through NetFlow, sFlow, IPFIX, and AppFlow traffic on remote flow networks.
- Then, learn how devices that are actively communicating on the network are discovered and classified, which provides you with over 4,000 built-in metrics for dozens of protocols.
- Finally, learn how software deduplication removes unnecessary duplicates from your ExtraHop metric data.

ExtraHop platform architecture

The ExtraHop platform comprises a suite of appliances—Discover, Explore, Trace, and Command—that are designed to passively monitor the network traffic in your environment in real time. Each appliance provides you with different types of information about your network, which you can analyze to determine where problems in your network might be developing.

Note: The combination of available appliances in Reveal(x) subscriptions vary by plan level.

ExtraHop Discover appliance

The ExtraHop Discover appliance (EDA) provides top-level and detailed metrics about transactions and traffic between devices. The Discover appliance includes tools to analyze and visualize all of your network, application, client, infrastructure, and business data.

The Discover appliance passively collects unstructured wire data—all of the transactions on your network—and transforms this data into structured wire data.

Discover appliances are provisioned with storage to support 30 days of metric lookback. Note that actual lookback varies from appliance to appliance, depending on traffic patterns, transaction rates, and the number of active protocols.

Deploy a single Discover appliance, either physical or virtual, anywhere in your network environment.

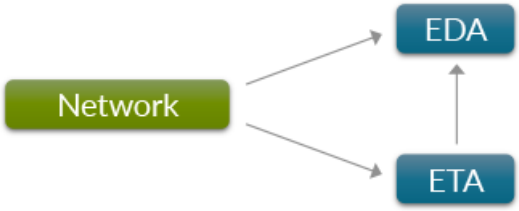
ExtraHop Explore appliance

The ExtraHop Explore appliance (EXA) integrates with the ExtraHop Discover appliance to store transaction and flow records sent from the Discover appliance. You can see, save, and search the structured flow and transaction information about events on your network with a simple, unified UI, with no modifications to your existing applications or infrastructure. Deploy a cluster of three or more Explore appliances to take advantage of data redundancy and performance improvements.

ExtraHop Trace appliance

The ExtraHop Trace appliance (ETA) continuously collects network packets and integrates with the ExtraHop Discover and Command appliances. You can quickly retrieve all packets that match a set of search criteria within a given time interval. You can then download the packet capture file for further inspection in a packet analyzer, such as Wireshark.

Deploy a Trace appliance when you need access to more than the summary data collected by the Discover appliance.



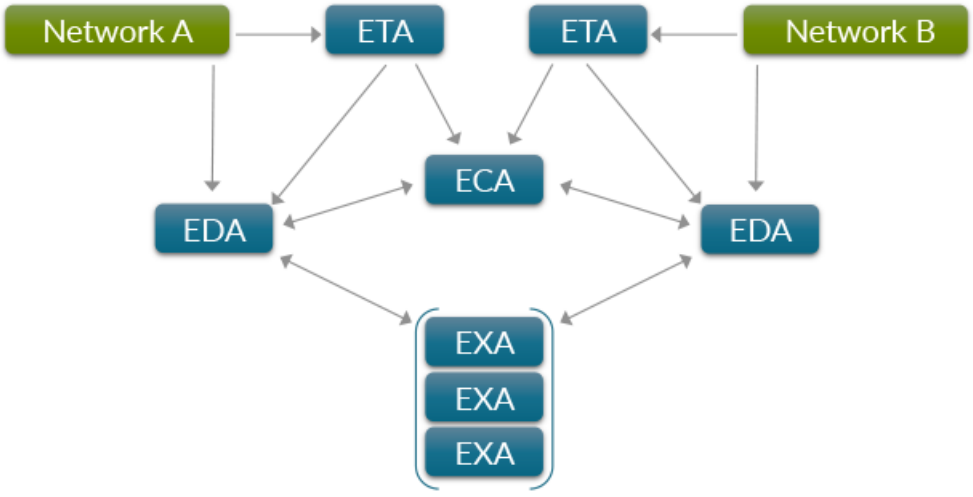
ExtraHop Command appliance

The ExtraHop Command appliance (ECA) provides centralized management and reporting across multiple ExtraHop Discover, Explore, and Trace appliances that are distributed across data centers, branch offices, and the public cloud.

You can connect an Explore appliance or cluster to multiple Discover appliances, and then query the records stored by each Discover appliance from the Command appliance.

When you add a Trace appliance, you can search, download, and analyze the collected packets to gain further insight about the information flowing across your network.

For most large ExtraHop deployments, a dedicated Command appliance is the most efficient way to manage all of your remote appliances.



Packets vs Records vs Metrics

Each appliance collects and stores a different level of information about your network interactions: the Trace appliance stores packets, the Explore appliance stores records, and the Discover appliance stores metrics.

Packets are the raw data transferred between two endpoints. Records analyze messages and transactions from raw data and then store metadata about the discrete interaction in an indexed, searchable database. Metrics aggregate observations about endpoint interactions over time.

For example, when a client sends an HTTP request to a web server, here is what each data type contains:

- The packet contains the raw data that was sent and received in the interaction.
- The related record contains the time-stamped metadata about the interaction: when the request happened, the IP address of the client and server, the requested URI, any error messages.
- The related metric (HTTP Requests) contains an aggregate of that interaction with other observed interactions during the specified time period, such as how many requests occurred, how many of those requests were successful, how many clients sent requests, and how many servers received the requests.

Both metrics and records can be customized to extract and store specific metadata with JavaScript-based triggers. While the ExtraHop system has over 4600 built-in metrics, you might want to create a custom metric that collects and aggregates 404 errors from only critical web servers. And, you might want to maximize your record storage space by only collecting transactions that occurred over a suspicious port.

Data sources in the ExtraHop system

The ExtraHop Discover appliance collects data and generates metrics from two types of data sources: wire data and machine data, such as flow data.

Wire data

Wire data is observed in real time, which provides information about what's happening on your network. With wire data, the ExtraHop system passively collects a copy of unstructured packets through a port mirror or tap and stores the data in the appliance datastore. The copied data goes through real-time stream processing, which transforms the packets into structured wire data through the following stages:

1. TCP state machines are recreated to perform full-stream reassembly.
2. Packets are constructed into flows.
3. The structured data is analyzed and processed in the following ways:
    a. Transactions are identified
    b. Devices are automatically discovered by MAC and IP address and then classified by their activity.
    c. Metrics are generated and associated with protocols and sources, and the metric data is then aggregated into metric cycles.
4. As new metrics are generated and stored, and the datastore becomes full, the oldest existing metrics are overwritten according to the first-in first-out (FIFO) principle.

Flow data

Flow data, a type of machine data, can also be collected from a network device and sent to the Discover appliance for analysis or storage. Flow data is an alternative option if wire data cannot be collected from a remote network.

Note: Reveal(x) appliances cannot be configured to collect flow data.

A flow is a set of packets that are part of a single transaction between two endpoints. Similar to how the ExtraHop system can identify flows from wire data, flows from machine data on remote networks can be

sent to a Discover appliance for analysis. Flows are identified through their unique combination of IP protocol (TCP/UDP), source and destination IP addresses, and source and destination ports.

The ExtraHop system supports the following types of flow data:

NetFlow v5

The Cisco proprietary protocol that defines a flow as a unidirectional flow of packets all sharing the following values: Ingress interface, source and destination IP address, IP protocol, source and destination ports, and the type of service. NetFlow v5 has a fixed record format with 20 fields and cannot be customized.

NetFlow v9

An adapted version of NetFlow v5 where the record format is template based. NetFlow v9 has 60+ fields in the records and can be customized. In the Discover appliance, these records are only partially parsed until the template packet is detected.

IPFIX

An open standard based on the NetFlow v9 standard. ExtraHop supports only the native format; formats where the Enterprise bit is set outside of a trigger are not supported.

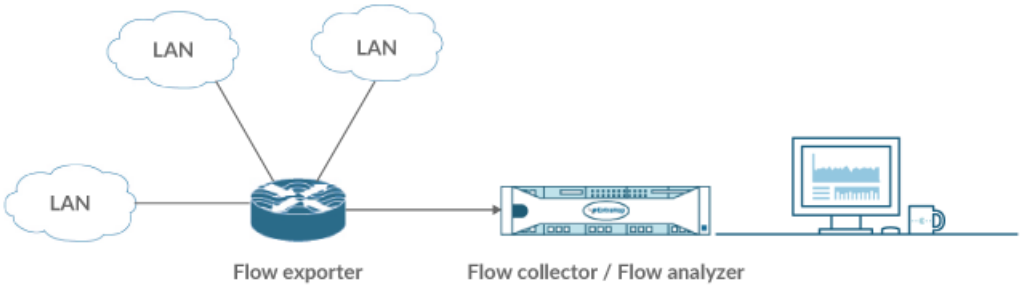
AppFlow

The Citrix implementation of IPFIX with customized extensions to include application-level information such as HTTP URLs, HTTP request methods, status codes, and so on.

sFlow

A sampling technology for monitoring traffic in data networks. sFlow samples every nth packet and sends it to the collector whereas NetFlow sends data from every flow to the collector. The primary difference between sFlow and NetFlow is that sFlow is network layer independent and can sample anything. NetFlow v5 is IP based, but v9 and IPFIX can also look at Layer 2.

The Discover appliance enables you to add any of the above flow data sources. You can then view metrics for flow networks (a network device that sends information about flows seen across the device) and their interfaces.



With the Discover appliance working as a flow collector and analyzer, you can collect the flow network traffic through the following stages:

1. Flow exporters detect and format traffic, caching information about the flow, including source and destination IP addresses, port, IP protocol, and number of bytes and packets.
2. The flow exporter sends the cached information from the flow network to the Discover appliance, which acts as a collector and analyzer for the flow data.
3. The flow network traffic is analyzed, flows are identified, and metrics are aggregated for the total number of bytes and total number of packets in each flow.

For example, when a client initiates a request to a server, the packet is sent to the router, which directs the packet to the destination server through the network topology. If that router is configured to be a flow network exporter, information about the flow is then formatted and sent to the Discover appliance for analysis.

By analyzing flows of network traffic, such as NetFlow traffic, an administrator can identify the top network flows (most bytes consumed), top network talkers (highest throughput), total number of bytes, and the total number of packets per router interface.

Metrics vs Records vs Packets

Packets are the raw data transferred between two endpoints. Records analyze messages and transactions from raw data and then store metadata about the discrete interaction in an indexed, searchable database. Metrics aggregate observations about endpoint interactions over time.

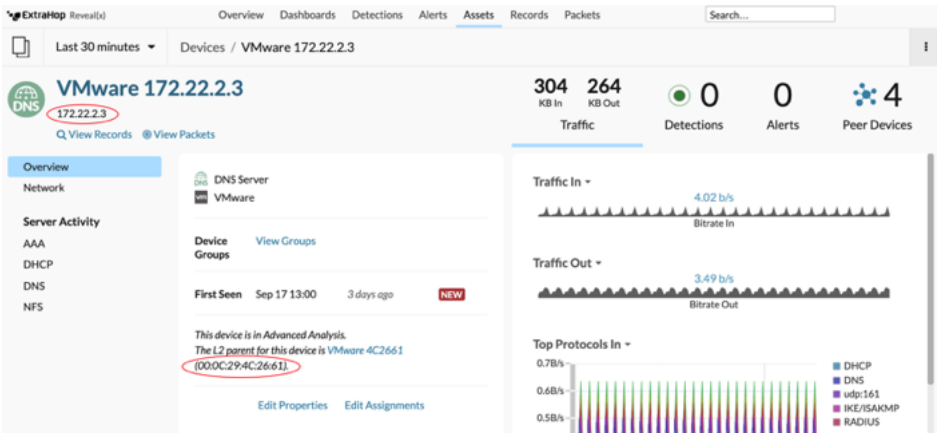
For example, a packet might contain an HTTP request from a client to a web server. The packet includes the raw data that was sent and received in the interaction. The related record contains the time-stamped metadata about the interaction: when the request happened, the IP address of the client and server, the URI, any error messages. The related metric (HTTP Requests) aggregates that interaction with other observed interactions for the specified time period, such as how many requests occurred, how many of those requests were successful, how many clients sent requests, and how many servers received the requests.

Both metrics and records can be customized to extract and store specific metadata with JavaScript-based triggers. While the ExtraHop system has over 4600 built-in metrics, you might want to create a custom metric that collects and aggregates 404 errors from only critical web servers. And, you might want to maximize your storage space by only storing URIs that resulted in 500-level server errors.

Device discovery

The ExtraHop system automatically discovers and classifies devices based on what is happening on the network. The default discovery mode is L3 device discovery, also known as Discover by IP.

First, the ExtraHop system creates an L2 device entry for every locally observed MAC address over the wire. Then, the ExtraHop system creates an L3 device entry for every locally observed IP address included in an Address Resolution Protocol (ARP) response. If the MAC address and IP address are associated with the same device, the Discover appliance links the parent L2 device and the child L3 device. The IP address and MAC address for a device are displayed in search results and on the device Overview page, as shown in the following figure.



Here are some important considerations about L3 device discovery:

- To discover L3 devices outside of your network, you can create a custom device or enable remote device discovery.
- If a router has proxy ARP enabled, the ExtraHop system creates an L3 device for each IP address that the router answers ARP requests for.
- L2 metrics that cannot be associated with a particular child L3 device (for example, L2 broadcast traffic) are associated with the parent L2 device.

- L2 devices that are not gateways or custom devices do not count towards your licensed analysis capacity. These L2 devices receive L2 Analysis and are exempt from analysis priorities and the watchlist.

After a device is discovered, the ExtraHop system begins to collect metrics for the device based on analysis priorities. You can search for L2 and L3 devices in the ExtraHop system by their IP address, MAC address, or name (either a hostname observed from DNS traffic or a custom name that you assign to the device).

Device discovery modes

The default discovery mode is L3 discovery, which is also known as Discover by IP in the ExtraHop Admin UI. When you disable Discover by IP, all locally observed IP addresses that are associated with a MAC address are aggregated into one L2 device. It is important to note that disabling Discover by IP changes the number of devices that are discovered by the ExtraHop system.

For more information, see Discover new devices by IP address in the Admin UI Guide.

Device names and roles

After a device is discovered, the ExtraHop system tracks all of the wire data traffic associated with the device. The ExtraHop system discovers device names by passively monitoring naming protocols, including DNS, DHCP, NETBIOS, and Cisco Discovery Protocol (CDP). A device can be identified by multiple names, which are all searchable. If a name is not discovered through a naming protocol, the default name is derived from device attributes (MAC address for L2 devices and the IP address for L3 devices). You can also create a custom name for a device.

Note: If a device name does not include a hostname, the ExtraHop system has not yet observed naming protocol traffic associated with that device. The ExtraHop system does not perform DNS lookups for device names.

Based on the type of traffic associated with the device, the ExtraHop system assigns a role to the device, such as a gateway, file server, database, or load balancer. Not all roles can be automatically assigned to a device, however, you can manually change a device role to any of the following roles:

Icon	Device Role
	Custom Device
	Database
	DHCP Server
	DNS Server
	Domain Controller
	File Server

Icon	Device Role
	Firewall
	Gateway
	Load Balancer
	Medical Device
	Mobile Device
	PC
	Printer
	VoIP Phone
	Vulnerability Scanner
	Web Proxy Server
	Web Server
	Other

Remote device discovery and custom devices

The ExtraHop system automatically discovers local L3 devices based on observed ARP traffic that is associated with IP addresses. By default, all IP addresses that are observed outside of locally-monitored broadcast domains are aggregated at one of the incoming routers in your network.

Note: If you have a proxy ARP configured in your network, the ExtraHop system might automatically discover remote devices. For more information, see this ExtraHop forum post.

To identify and learn about individual devices located outside of local routers, complete one of the following options:

- Add a remote IP address range in the ExtraHop Admin UI to discover L3 devices for IP addresses that are outside of the local network. An L3 device is created for each IP address that is observed within the remote IP address range.
- Create a custom device to collect metrics for a remote IP address or a range of IP addresses into one device. A single device is created for all of the IP addresses observed within the remote IP address range. For example, you can create a single device that collects metrics for several known IP addresses that belong to remote sites or cloud services.

Network locality

By default, any device with an RFC1918 IP address (included in a 10/8, 172.16/12, or 192.168/16 CIDR block) that the ExtraHop system automatically discovers is classified as an internal device. You can then monitor internal network connections to devices outside of your network with ExtraHop metrics and detections. These metrics and detections can help you determine if unauthorized devices are attempting to access your internal network. However, because some network environments include non-RFC1918 IP addresses as part of their internal network, you can change the internal or external classification for IP addresses from the Network Localities page.

For more information, see Specify the locality for IP addresses.

Software frame deduplication

The ExtraHop system removes duplicate L2 and L3 frames and packets when metrics are collected and aggregated from your network activity by default. L2 deduplication removes identical Ethernet frames (where the Ethernet header and the entire IP packet must match); L3 deduplication removes TCP or UDP packets with identical IP ID fields on the same flow (where only the IP packet must match).

The ExtraHop system checks for duplicates and removes only the immediately-previous packet both on the flow (for L3 deduplication) or globally (for L2 deduplication) if the duplicate arrives within 1 millisecond of the original packet.

By default, the same packet traversing different VLANs is removed by L3 deduplication. In addition, packets must have the same length and the same IP ID, and TCP packets also must have the same TCP checksum.

L2 duplication usually only exists if the exact same packet is seen through the data feed, which is typically related to an issue with port mirroring. L3 duplication is often the result of mirroring the same traffic across multiple interfaces of the same router, which can show up as extraneous TCP retransmissions in the ExtraHop system.

The System Health page in the ExtraHop Web UI contains charts that display L2 and L3 duplicate packets that were removed by the ExtraHop system. Deduplication works across 10Gbps ports by default and across 1Gbps ports if software RSS is enabled. L3 deduplication currently is supported only for IPv4, not IPv6.

Related topics

Check out the following guides and resources that are designed to familiarize new users with our top features.

- Learn how to monitor website performance in our dashboard walkthrough.
- Learn how to identify potential DNS server issues in our metrics walkthrough.
- Learn about wire data fundamentals (online training)
- Learn about getting started with ExtraHop (online training)
- Visit our forums to communicate with other ExtraHop users.

- Contact ExtraHop Support if you need additional help.