Anomaly detection with ExtraHop Addy

Published: 2019-02-09

The ExtraHop Addy™ service is a cloud-based service that applies machine learning techniques to automatically determine what is normal versus unusual behavior in your IT environment. Unlike other machine learning solutions that rely on logs or agent data, the Addy service applies machine learning technology to your wire data without requiring you to configure anything. When the Addy service is activated, you can browse and investigate anomalies and then drill down to identify the root cause of the issue.

Overall, the Addy service offers the following types of help:

- Uncover hidden issues before they create problems for your users
- Collect high-quality, actionable data to identify root causes of anomalies
- Find unknown performance issues, security issues, or infrastructure quirks
- Gain deeper insight into your network behavior

Important: The Addy service does not analyze sensitive information and data types.

Here are important considerations about anomaly detection with the Addy service:

- You must have an Addy service license.
- You must have full system privileges, access to the Admin UI, and access through any firewalls to connect a Discover appliance to the Addy service through ExtraHop Cloud Services. For more information, see Connect to the ExtraHop Addy service.
- You must have at least four weeks of wire data metrics stored on your Discover appliance before Addy can detect anomalies.
- On a Command appliance, you can access anomalies on a connected Discover appliance if that Discover appliance is connected to the Addy service.

Navigating anomaly detection

After connecting to ExtraHop Cloud Services, the Addy service automatically begins to calculate the expected range of normal metric values from four weeks of stored Discover appliance metrics, and then detects anomalies.

To browse anomalies, log into the Web UI on the Discover or Command appliance and click **Alerts** at the top of the page. The left pane contains links to the Alert History and Anomalies pages.

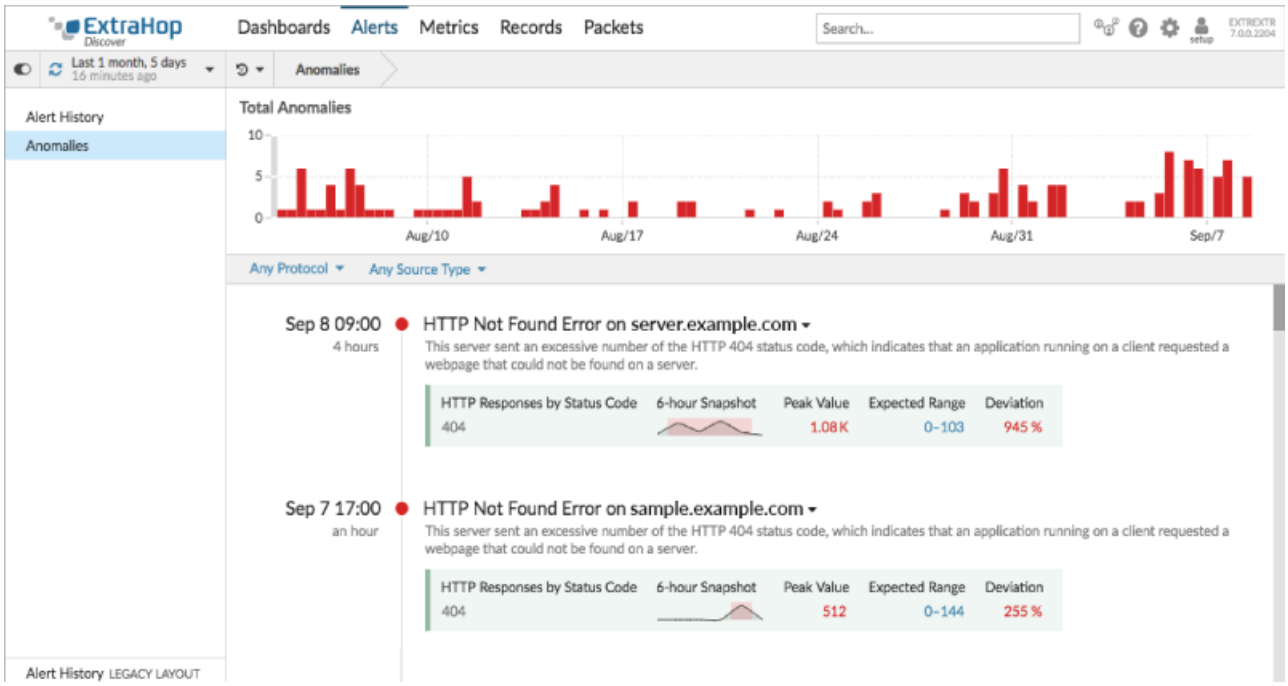
On the Alert History page, you can view the following details about anomaly alerts: name, severity, source, the most recent time.

- Alerts concepts
- Configure Addy anomaly alert settings
- Add a notification to an alert configuration to receive emails when an anomaly is generated

Note: Anomaly alerts are generated for anomalies that are detected after your alert configurations are saved.

On the Anomalies page, you can view all the anomalies that were automatically detected from your wire data by the Addy service.

The following figure shows how anomalies are displayed on the Anomalies page:



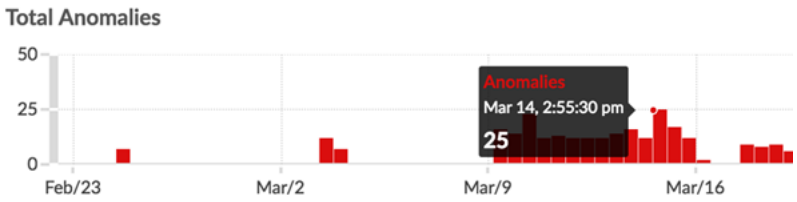
Interpret anomalies

The Anomalies page displays the total number of anomalies for the selected time interval and details about each detected anomaly.

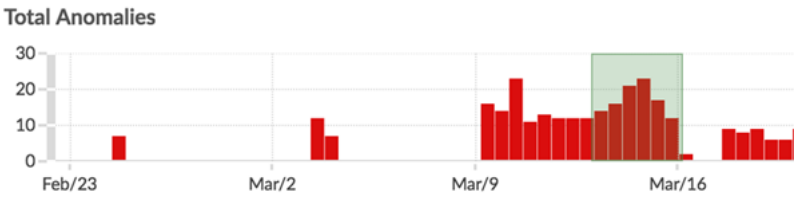
View the total number of anomalies

The Total Anomalies chart provides a summary of detected anomalies (y-axis) over time (x-axis) for the selected time interval. Each bar in the chart represents the total number of concurrent, active anomalies that were detected during a specific time period. Look for the tallest bar to determine when the most anomalies occurred in a time period.

Hover over a bar to view information, such as date, time, and the number of detected anomalies for a specific time period.



Click and drag across an area on the chart (which will become highlighted in green) to zoom in on a specific time range. The time interval in the Discover or Command appliance dynamically updates to match the new time range in the chart, and details about each anomaly that was detected in that time range are displayed below the chart.



View details for each anomaly

Each anomaly that was detected for the specific time interval appears in a list below the Total Anomalies chart. You can filter this list to find anomalies.

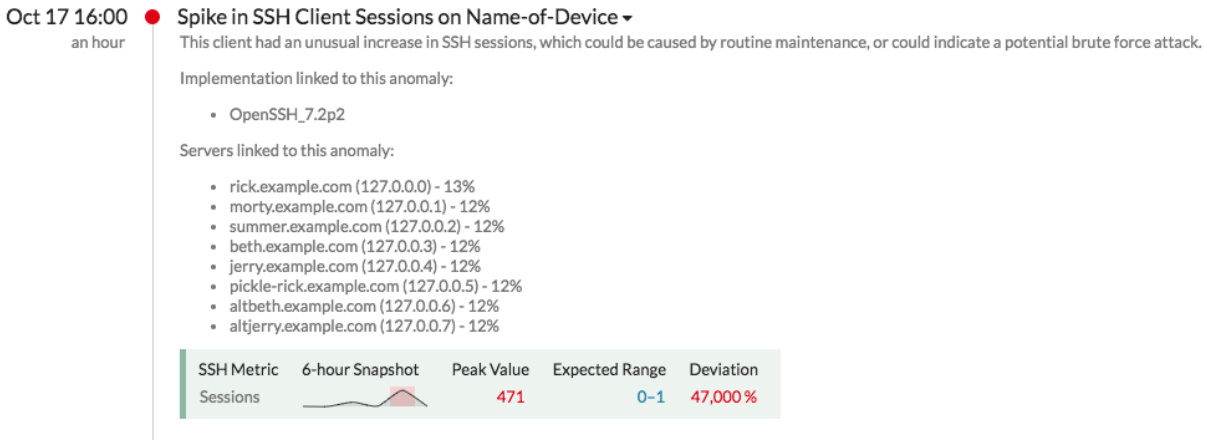
Anomaly details include the anomaly title and description, the duration of the anomaly, the anomalous metric name, a sparkline of the unusual metric activity over time, and the values associated with the anomaly.

Title

The title includes the anomalous metric and the device or application name linked to the anomaly. Click the anomaly title to navigate to the protocol page for the device or application. From the protocol page, you can investigate top-level and detail metrics. For more information, see Investigate the root cause of anomalies with the Addy service.

Description

The description provides information about what the anomaly means. For most anomalies, Addy automatically surfaces detail metrics identified with Addy's machine learning capabilities, so you can immediately begin your investigation. The following figure shows an example of this type of automated investigation. A client initiated an unusual number of SSH sessions with multiple servers. At a glance, you can learn which servers were connected to this client during the anomaly, the percentage of sessions for each server, and the name of the client implementation linked to the anomaly.



Note: Automated investigation is not available for server processing time anomalies. For these anomalies, you can investigate anomalies from protocol pages in the Discover or Command appliance.

Duration

The duration of the anomaly, listed below the date and time, indicates how long the anomalous value was detected by the Addy service. For example, the duration for the anomaly in the figure above is 2 hours.

The minimum duration of an anomaly is one hour, because the Addy service detects anomalies by analyzing metric data with 1-hour granularity. If the duration value is displayed as ONGOING, the anomalous metric is in the process of being detected.

Sparkline

Sparklines are simple line charts that show you the metric behavior that led up to the anomaly. The sparkline charts display a snapshot of metric data from the time frame around the duration of the detected anomaly (such as 6 hours), and not the overall time interval from the top of the page (such as the last 7 days).

The red area on a chart highlights the anomalous metric values, which includes the peak value, on the sparkline.

Peak Value

The maximum value from observed data that deviated from expected ranged for the duration of the anomaly.

Expected Range

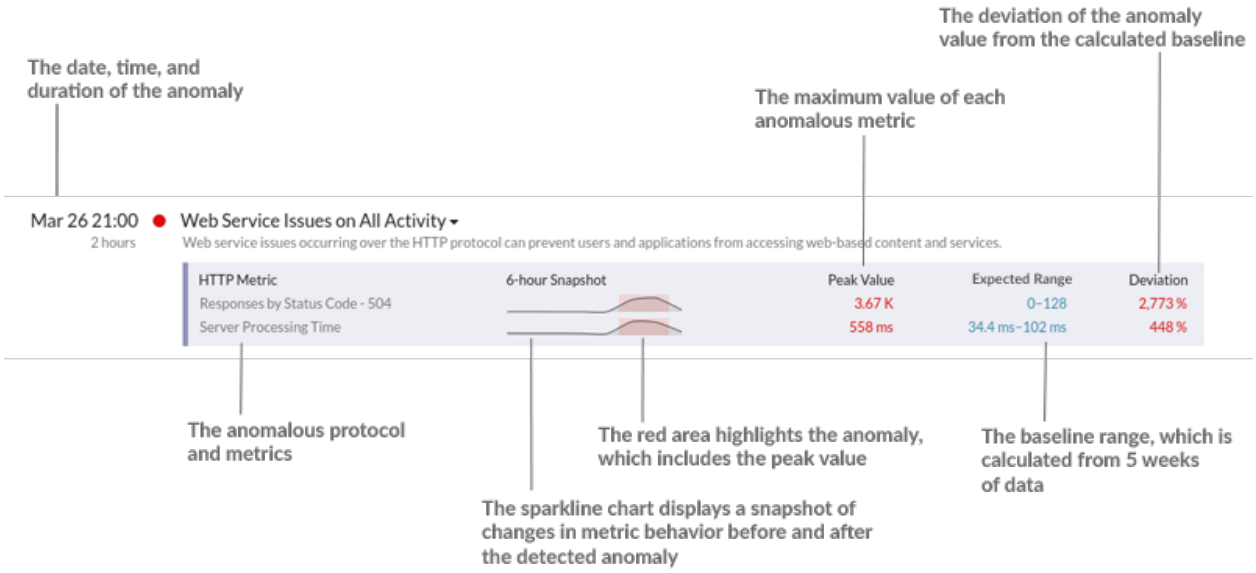
The range of values that represent a normal background level of activity, which is calculated based on 4 weeks of data. The expected range is the basis for comparison with observed values to detect changes in metric activity.

Deviation

A quantity calculated by the Addy machine learning engine to indicate the extent of change from an expected range.

More than one anomalous metric can be associated with a specific application or device. If you see concurrent anomalies, which occurred at the same time for the same device, you can investigate how the anomalous metrics contributed to an issue.

The following figure shows the type of information available for two anomalies that were detected over the HTTP protocol for a single application.



Best practices for investigating anomalies

The Addy service provides you with high-quality, actionable data about anomalies—but does not replace decision-making or expertise about your network. The following best practices explain how to determine which anomalies are worth further investigation and when to take action.

Investigate anomalies in the Discover or Command appliance

Click on an anomaly title to navigate to the device or application protocol page. This page contains the anomalous metric data observed at the time of the anomaly along with related metrics. You can then drill down on specific URIs, clients, and servers to find the source of the anomaly, and then decide how to respond.

For example, if you see an FTP server error anomaly detected for a server, you can view metrics for that server in the Discover or Command appliance, and then drill down on the anomalous error by user or client IP address to identify who is generating the error.

For more information, see Investigate the root cause of anomalies with the Addy service.

Investigate anomalies by changing the time interval

Change the time interval to view anomalies that might have occurred during a reported problem. For example, does the time frame of the anomaly coincide with a reported issue, such as slow load times or login times? You can also compare anomalies from the past month to the current date, which gives you a sense of whether the occurrence or severity of anomalies is changing over time.

For more information, see Find anomalies with the Addy service.

Investigate anomalies by protocol

Filter by protocol to quickly monitor critical protocols with a role in security, commerce, or communication processes.

For example, an FTP 530 error anomaly might indicate that someone is trying to gain unauthorized access to information on your network. Or Citrix server and client latency anomalies might indicate that clinicians cannot access patient information in a timely fashion.

Selecting different protocols can also show you how anomalies correlate to each other. An anomalous HTTP response time followed immediately by an anomalous CIFS server processing time might suggest that web servers are dependent on how quickly your file storage servers can send and receive file data.

For more information, see Find anomalies with the Addy service.

How the Addy service works

This section provides some background information on how the Addy service identifies anomalies.

Anomalies are unexpected deviations from normal patterns in device or application behavior. By detecting an anomaly as soon as it happens, you can identify and resolve a potential issue before it becomes a larger problem. You can also review historical anomaly data to investigate issues related to known security or network outage events.

In most network monitoring tools, anomalies are detected through manually-configured alerts and trend models for individual devices. However, as your network changes—because of hardware reconfigurations or the addition of applications to your network—these types of alerts and models can become quickly outdated and potentially inaccurate. The Addy service automatically delivers consistent and accurate results about anomalous metrics and protocols without requiring manual configuration for individual devices. The Addy machine learning engine analyzes the historical behavior of individual devices, and automatically adapts to each device across time when there are changes to the expected range of data in your network.

Here is how Addy anomaly detection generally works: the metrics that the Addy machine learning engine analyzes come from wire data that is collected by your Discover appliance. The Discover appliance processes this data, generates metrics, and associates the metric data with protocols, devices, and applications. The Addy service retrieves a subset of protocol metrics from the Discover appliance to analyze and report results about detected anomalies.

The algorithm that drives the machine learning engine in the Addy service calculates the expected range of normal network behavior and adapts to changing variations in protocols and metric data. Outliers, or anomalies, are then detected based on three variables:

- Observed data, collected in real-time by the Discover appliance
- Expected range data, calculated from four weeks of historical data collected by the Discover appliance
- Threshold values, which are automatically adjusted by the algorithm based on historical metric data and heuristics defined by the IT networking domain experts at ExtraHop

Note: If you need to define a specific threshold value for an anomaly, which might be associated with a service level agreement (SLA) for example, we recommend manually configuring an alert in the Discover appliance.

Essentially, an anomaly is detected when observed data deviates from the expected range of data by a significant amount. You can then view analysis results about anomalies on the Anomalies page in the Web UI of the Discover appliance. For each anomaly, the Addy service provides the measured deviation (which is the difference between the observed value and the expected range), the anomaly value, and the expected range of normal metric values at the time of the anomaly.

The Addy service also provides anomalous 50th percentile or 75th percentile values for a subset of metrics that account for server processing time.

Related topics

Check out the following resources that are designed to familiarize new users with the Addy service.

- Connect to the ExtraHop Addy service
- Find anomalies with the Addy service
- Investigate the root cause of anomalies with the Addy service