Splunk® User Behavior Analytics Get Data In to Splunk User Behavior Analytics 5.0.0

Understand data flow in Splunk UBA

Generated: 1/08/2020 12:48 am
Understand data flow in Splunk UBA

Get data in to Splunk UBA from Splunk Enterprise. Splunk UBA uses this data along with automated machine learning algorithms to generate anomalies and threats about the users, accounts, devices, and applications in your environment. Some common use cases include detecting compromised accounts and devices, detecting data exfiltration, detecting insider abuse, including privileged accounts abuse, and providing context and information for investigations.

This image shows how Splunk UBA uses data from Splunk Enterprise to generate anomalies and threats. Splunk UBA can send these anomalies and threats back to the Splunk platform for further analysis using integrated products such as Splunk Enterprise Security. Each element in the image is described in the table immediately following the image.

### Step Process Description

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1    | Raw or Parsed Data from Splunk Enterprise | SPLUNK UBA can ingest parsed, CIM-compliant data or raw events from Splunk Enterprise. **• See How data gets into Splunk UBA for more information about how data gets ingested into Splunk UBA from Splunk Enterprise.**  
  **• See Add data from the Splunk platform to Splunk UBA for instructions.** |
<p>| 2    | Event Parsing and Identity Resolution | SPLUNK UBA analyzes the data from Splunk Enterprise and maps the fields in the data to Splunk UBA-specific fields. Properly mapped data fields enables SPLUNK UBA to perform real-time identity resolution to accurately define users and devices. |</p>
<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Views</td>
<td>Splunk UBA provides a level of abstraction so that you only see important, relevant data. For example, firewall logs can come in a variety of formats depending on the source vendor. Splunk UBA abstracts the variety in the data and shows you only the data relevant to firewall-related activity, such as IP, port, and information about the firewall action.</td>
</tr>
<tr>
<td>4</td>
<td>Anomaly Models</td>
<td>You can also configure generic data sources to bring raw events into Splunk UBA. These raw events are stored in data cubes and are processed for anomalies even though they do not belong to a specific view. Anomaly models analyze the data in Splunk UBA and create anomalies based on a variety of factors.</td>
</tr>
</tbody>
</table>

- Streaming models analyze ingested data in real time and determine the impact of those events over a short time window, such as the past hour. Based on this analysis, streaming models can produce a multitude of items in Splunk UBA, such as anomalies, indicators of compromise (IoCs), or analytics data.
- Batch models and anomaly rules analyze ingested data over a larger time window, such as the last 24 hours, typically running overnight due to the need to process large amounts of data. All threat models in Splunk UBA run as batch models, taking into account the aggregation of data in Splunk UBA including the data cataloged by the streaming models. There are two types of batch models in Splunk UBA:
  - Rare event models generate anomalies by detecting unusual, rare, or first time activity.
  - Time series models generate anomalies by tracking specific activities over a period of time.

You can create new batch anomaly models or clone existing models using the custom use case framework.
Anomalies are grouped into various categories such as Exfiltration, Infection, or Expansion. These categories typically correspond to stages of the kill chain and make it possible for your threat logic to place anomalies into the correct sections of the chain. Anomaly models also add metadata to anomalies. These metadata are specific to individual anomaly types and are used to filter the anomalies during investigation or anomaly hunting.

### Step 5: Anomalies

Anomalies are notable findings in the data, such as deviations from typical behavior or the detection of interesting patterns, like beaconing. A Splunk UBA operator can view anomalies and take further action as needed. Anomalies vary in scope and complexity, ranging from simple highlights of a useful alarm generated by an external product, such as a security endpoint solution or a firewall, to stealthy data exfiltration attempts requiring advanced statistical and machine learning models to detect. Anomalies are generated by the streaming models, batch models, and anomaly rules.

- You can use anomaly action rules in Splunk UBA to manage existing anomalies in the system as part of the anomaly creation flow. For example, you can delete or restore anomalies, modify the score, or add anomalies to a watchlist. See Take action on anomalies with anomaly action rules.
- You can also customize anomaly scoring rules to provide a level of control and consistency across specific anomaly types. See Customize anomaly scoring rules.

Anomalies have both types and categories. Types are specific descriptive names of anomalies, while categories are generic descriptions for anomalies. Multiple anomaly types can share a category, and one anomaly type can have multiple categories. Create an anomaly action rule or a custom threat with either anomaly types or anomaly
categories, depending on how specific you want the rule or threat to be.

Threat models in Splunk UBA build dynamic threats based on the data and anomalies in the system.

You can also create custom threat rules to identify verifiable threats in your network, like specific activities that you want to monitor for policy compliance. Create, edit, enable, and manage the custom threats that are most useful for your organization. You can create custom threats that apply to users, devices, or sessions. Several custom threats are included with Splunk UBA.

- Threat models take into account the aggregation of data in Splunk UBA, including the data cataloged by the streaming models, to generate threats. All threat models in Splunk UBA run as batch models.
- Threat rules generate threats by looking for specific anomaly patterns within a specific window of time. A threat is generated each time the anomaly pattern is found. Each rule runs on a pre-defined schedule, depending on the nature of the rule.

A threat is a collection of one or more anomalies that form a clearly defined security use case, such as Data Exfiltration. Threats are often correlated with indicators of compromise (IoC) and other supporting evidence to provide a detailed description of a series of events.

Threats can be computed in the following ways:

- Kill-chain threats examine all anomalies for a specific user or device for patterns that align with the kill-chain stages. Example kill-chain threats are Data Exfiltration by Suspicious User or Device or Data Exfiltration by Compromised Account.
- Graph-based threats are computed based on groups of similar anomalies rather than anomalies grouped by user or device. Example graph-based
**Threats**

- Data-driven threats collect data about anomalies and users or devices to determine the likelihood of a threat. Insider threats are computed by the data-driven computation process. Example data-driven threats are Lateral Movement and Data Exfiltration.
- Rule-based threats are raised from custom threat rules when a specific set of conditions are met. Example rule-based threats are Brute Force and Data Exfiltration after Data Staging.

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Threats are Public-facing Website Attack or Fraudulent Website Activity.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data-driven threats collect data about anomalies and users or devices to determine the likelihood of a threat. Insider threats are computed by the data-driven computation process. Example data-driven threats are Lateral Movement and Data Exfiltration.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rule-based threats are raised from custom threat rules when a specific set of conditions are met. Example rule-based threats are Brute Force and Data Exfiltration after Data Staging.</td>
</tr>
</tbody>
</table>