Normalization and Processing

Normalization merges events containing different data into a reduced format which contains common event attributes. Most logs capture the same basic information – time, network address, operation performed, etc. Categorization involves adding meaning to events – identifying log data related to system events, authentication, local/remote operations, etc.

Log Enrichment

Log enrichment involves adding important information that can make the overall data more beneficial for security analysts when investigating alerts or unusual activity. One example could be logs that contain public IP addresses, but not their geographical location. Performing a simple lookup to see what geographical range the IP belongs to, this can now immediately provide analysts with the country this IP is based in, which can aid investigations and help to build metrics.

Log Indexing

SIEMs can hold an absolute ton of data, and to search through all of that, especially when looking over a long period of time such as a couple of weeks, this can be extremely slow. By indexing attributes that are shared by a large amount of logs it can make searching for specific attributes across large data faster compared to having to scan every single piece of data in the SIEM storage to get the answers you need.

Log Storage

For large organizations, the amount of storage needed to support a SIEM can be a large effort on the part of infrastructure and security teams. While alternatives to on-premises servers exist, such as Amazon Web Services S3 buckets or Hadoop, it is important for teams to consider all of their options, weighing in factors such as cost, ease-of-use, and scalability.

Normalization

Different software, hardware, and devices produce their own format of logs, as there is no universal format (we wish!). This makes it difficult when sending tons of different logs to a SIEM, because the platform needs to be able to understand and sort all of the attributes to allow for analysts to perform searches through all of the data stored within the SIEM.

SIEM log normalization is the process of changing log formats into a format that is as similar as possible across all devices and log sources, giving the SIEM a break and allowing for more consistent searching and information breakdown. Obviously logs from a Windows endpoint won't look the same as logs from a Linux system, but if we can match up things the best we can, the SIEM can handle the rest.

A great example by Travis Marlette in the Splunk Best Practice book is the following: "In our example, let's say source_ip represents both the src_ip field from Cisco network devices, and the source_address from Juniper network devices." This means we can now search in our SIEM for source_ip value, and it'll check both Cisco and Juniper logs!

○ Using RDP and SSH

○ How to Start Your Exam