What SIEM Data Sources Should You Integrate into Your Platform?

Posted on May 20, 2021

In a perfect world, there would not be any need to mull over what data sources to integrate into an organization's security information and event management (SIEM) solution. All kinds of data that can be used and abused by threat actors should be added. After all, attacks can hide behind seemingly innocuous logs.

But in reality, each data source comes at an additional cost since most SIEM solution providers typically charge per gigabyte. Thus, organizations have to strike a balance between budget constraints and security. However, one should not necessarily suffer for the sake of the other. But that requires careful strategizing in terms of what data sources to integrate into SIEM solutions.

This post takes a deep dive into SIEM data sources to help organizations understand the following:

- What SIEM data sources are
- Factors to consider when choosing SIEM data sources to feed to solutions
- Potential data sources to integrate into SIEM solutions

What Are SIEM Data Sources?

SIEM data sources are data feeds or databases fed to a SIEM product, which processes and analyzes them. The primary goal of SIEM products is to detect possible cyberthreats so organizations can investigate and mitigate them promptly. SIEM solutions cannot do this alone—they need to ingest data that the systems process and analyze.

Hence, SIEM products are often considered only as effective as the data sources they ingest, making the selection crucial.

SIEM Data Sources to Feed to Solutions

SIEM data sources can be split into two broad categories—data logs and contextual information. We explain what these are in more detail with examples below.

Data Logs

Logs from your network infrastructure are among the necessary data sources that your SIEM platform must ingest. These include logs from switches, routers, databases, applications, domain controllers, Domain Name System (DNS) servers, and wireless access points.

For instance, detecting successive failed login attempts to your router could hint at a brute force attack or someone manually trying to guess the login credentials. The read volume logged in a database could also be considered an indicator of compromise (IoC). An increase in database read volume could mean that someone is trying to extract a massive amount of data from the database.

Aside from network device logs, the logs from an organization’s security systems can also be valuable SIEM data sources. These include:

- Firewall logs
- Web filtering logs
- Endpoint security solution logs
- Honeypot data

Contextual Information

Data logs are necessary for SIEM products to run analyses. However, the volume of data logs that these systems process often result in too many alerts, which are difficult for security teams to handle.

And this is not too far-fetched, as researchers have discovered that as many as 56% of large organizations receive at least 1,000 security alerts each day. And most of them find this volume disturbing for several reasons, including:

- Minor problems may drown critical alerts.

- They could be wasting time investigating false positives.
- Security teams are simply overwhelmed by the number of alerts.

One of the ways to reduce the burden could be to provide context to data logs so that SIEM solutions can prioritize alerts according to their urgency. Therefore, it may not be enough to feed your SIEM plain data logs. Data sources that help contextualize information could add further value.

So, what data sources can add contextual information? Here are some examples.

Threat Intelligence

Cybersecurity intelligence sources provide information about known threats, including IoCs, artifacts, and attacker tools, tactics, and procedures (TTPs).

Instead of raw data, the information that threat intelligence supplies is already processed. Alone, it can already help security teams gather information about ongoing threats. That said, integrating threat intelligence sources into SIEM products can also amplify capabilities and give a sense of priority.

For example, data logs containing multiple failed login attempts from the IP address 192[.]81[.]208[.]169 could be flagged by any SIEM system. But what if the login was successful? Your SIEM product, in its bare form, may not find anything suspicious. After all, the user was able to provide the correct login credentials.

But when threat intelligence is integrated into your SIEM solution, it would detect that the IP address is related to Hafnium, a new cyberattack group exploiting zero-day vulnerabilities in Microsoft Exchange Servers. With this information, your SIEM product can tag the alert as critical, bringing it in front of all other alerts.

Domain Ownership and Contact Information Data Sources

Say that network logs tell you that a user contacted the domain mezzo[.]cc on a certain date and time. That may not raise an alert, especially since the domain is not tagged malicious.

Integrating current and historical WHOIS data sources into your SIEM solution can give you more context by specifying who the user contacted. The historical WHOIS records of mezzo[.]com, for instance, connects it to one of the people involved in spreading GozNym malware. According to the Federal Bureau of Investigations (FBI), the cyberattack group used the malware to steal online banking credentials and gain unauthorized access to victims' bank accounts.

At the very least, SIEM products with domain ownership data sources can alert security teams when domains owned by known threat actors appear in data logs. Aside from ownership details, WHOIS data also contains registration dates and contact information, which can add more context to logged domain names.

IP Geolocation Data Sources

Logs and events can further be enriched by including IP geolocation data. The source or destination IP addresses can be traced to the cities, states, and countries where the events came from. For organizations with diverse locations, IP geolocation data can help security teams pick events that originate from outside their service areas.

Furthermore, IP geolocation enables security teams to visualize a geographic map of events and detect patterns.

Factors to Consider When Selecting SIEM Data Sources

With the vast array of data sources to choose from, here is the underlying question: what data sources should be integrated into a SIEM solution? To answer this question, you first need to identify the "what" of your SIEM product. What do you want to monitor? And, what is your SIEM's primary use case?

For example, you may primarily want to track outgoing network connections to prevent data exfiltration. So, you would need network, firewall, web filtering, web proxy, and DNS query logs to serve as SIEM data sources.

On the other hand, if you also want to detect threats as they approach your network, threat intelligence sources could be logical additions to your set of data sources.

After identifying how the SIEM product is going to be used, other considerations also play a role in selecting data sources. Among them are:

- **Ease of integration:** Integrating the data source into your SIEM solution should be uncomplicated. For one, you have to make sure that the data source comes in a compatible format.

- **Ability to be parsed:** Some data sources may include records that you don't need for a specific SIEM use case. The vendor should take out the necessary data and use a format that your SIEM solution can consume.

- **Data quality:** To reiterate, your SIEM product's effectiveness depends on the data sources fed to it. Aside from choosing the right type of data source, it's also essential for the source to be complete and up-to-date.

How WhoisXML API Can Help

WhoisXML API provides IP, WHOIS, and DNS data sources in formats that are compatible with most SIEM products. IBM QRadar SIEM is a platform that benefits from WhoisXML API's complete and well-parsed WHOIS data source.

Events logged by QRadar are enriched with complete WHOIS information, including domain ownership details, availability status, and registry dates.

WhoisXML API also provides IP data sources that can add netblock ownership and geographical context to events with daily and weekly updates respectively.

Aside from current WHOIS and IP data sources that add context to an event, WhoisXML API also provides data sources that can further enhance the threat detection capability of SIEM solutions.

Historical WHOIS data sources, for instance, enable SIEM products to look through the past ownership details of a domain name identified in an event. Domains previously owned by known threat actors may warrant alerts.

The same is true for DNS Database Download. Integrating DNS data points into SIEM systems can alert security administrators of an event's association with malicious domains, IP addresses, and other DNS records.

Typosquatting Data Feed and Disposable Email Domains Data Feed can also be used for comparing event logs. Are there incidents that involve domains found in either feed? If so, the event can be given a higher priority, as the use of typosquatting and disposable email domains are indicative of phishing or spam campaigns.

Conclusion

SIEM data sources are as important as the solution you choose. They can refer to data logs, such as network, firewall, and device logs. But for SIEM solutions to help prioritize alerts, including data sources that provide contextual information would add more value. Among such data sources are threat, WHOIS, DNS, and IP intelligence.

The primary driver in selecting what data source to feed your SIEM solution is ultimately its specific use case. Is it meant to detect threat indicators and artifacts? Do you want your SIEM system to prevent data exfiltration?

Once you identify the “what” of your SIEM platform, you can choose data sources using a number of criteria, such as ease of integration, data quality, and a vendor’s parsing ability. WhoisXML API provides well-parsed and easily integrable data feeds and APIs that can:

- Provide context to your SIEM solution’s data logs and events
- Enhance the capabilities of your SIEM platform, especially when its primary use case is threat detection