LETSDEFEND.IO

SIEM FOR BEGINNERS

PREPARED FOR SOC TEAMS

LetsDefend

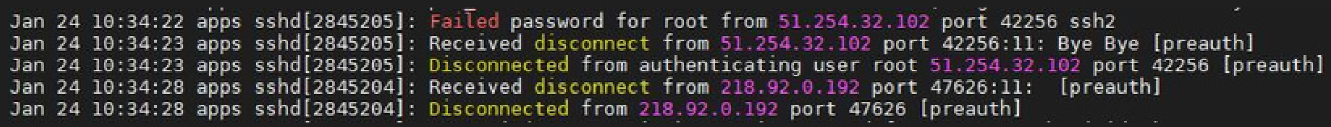
Table of Contents

LetsDefend

SIEM and Log Collection

Log Collection

It contains a basic log, time, source system and a message. For example, when we look at the content of the "/var/log/auth.log" file on an Ubuntu server, we can see the source, time and message information.



```
Jan 24 10:34:22 apps sshd[2845205]: Failed password for root from 51.254.32.102 port 42256 ssh2
Jan 24 10:34:23 apps sshd[2845205]: Received disconnect from 51.254.32.102 port 42256:11: Bye Bye [preauth]
Jan 24 10:34:23 apps sshd[2845205]: Disconnected from authenticating user root 51.254.32.102 port 42256 [preauth]
Jan 24 10:34:28 apps sshd[2845204]: Received disconnect from 218.92.0.192 port 47626:11:  [preauth]
Jan 24 10:34:28 apps sshd[2845204]: Disconnected from 218.92.0.192 port 47626 [preauth]
```

Logs are generally collected in the following 2 ways:

- Log Agents
- Agentless

Log Agents

In order to implement this method, a log agent software is required. Agents often have parsing, log rotation, buffering, log integrity, encryption, conversion features. In other words, this agent software can take action on the logs it collects before forwarding them to the target.
For example, with the agent software, we can divide a log with "username: LetsDefend; account: Administrator" into 2 parts and forward it as:

- message1 = "username: LetsDefend"

- message2 = "account: Administrator"

SIEM and Log Collection

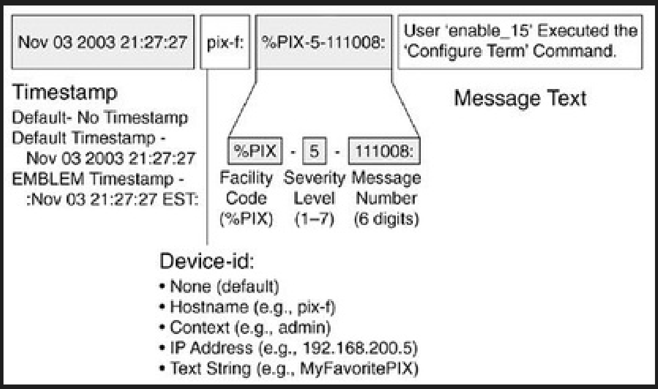
Log Agents

In order to implement this method, a log agent software is required. Agents often have parsing, log rotation, buffering, log integrity, encryption, conversion features. In other words, this agent software can take action on the logs it collects before forwarding them to the target.
For example, with the agent software, we can divide a log with "username: LetsDefend; account: Administrator" into 2 parts and forward it as:

- message1 = "username: LetsDefend"
- message2 = "account: Administrator"

Syslog

It is a very popular network protocol for log transfers. It can work with both UDP and TCP, and can optionally be encrypted with TLS. Some devices that support syslog: Switch, Router, IDS, Firewall, Linux, Mac, Windows devices can become syslog supported with additional software.
If you want to forward your log with Syslog, you will need to parsing in syslog format.
Syslog Format:
Timestamp - Source Device - Facility - Severity - Message Number - Message Text



```
Nov 03 2003 21:27:27 pix-f: %PIX-5-111008: User 'enable_15' Executed the 'Configure Term' Command.
```

Timestamp
Default- No Timestamp
Default Timestamp - Nov 03 2003 21:27:27
EMBLEM Timestamp - :Nov 03 21:27:27 EST:
%PIX - 5 - 111008:
Facility Code (%PIX)
Severity Level (1–7)
Message Number (6 digits)
Message Text
Device-id:
• None (default)
• Hostname (e.g., pix-f)
• Context (e.g., admin)
• IP Address (e.g., 192.168.200.5)
• Text String (e.g., MyFavoritePIX)

Also, the maximum packet size that can be sent with Syslog UDP is 1024 bytes. For TCP it is 4096 bytes.

SIEM and Log Collection

3. Party Agents

Most SIEM products have their own agent software. 3rd party agents have more capabilities than syslog because of the features they support. Some agents:
Splunk: universal forwarder
ArcSight: ArcSight Connectors
These agents are easy to integrate into SIEM and have parsing features.

Open Source Agents

They are generally agents that provide basic needs comfortably. However, it may not be as effective as the agent of the SIEM product itself. (Ease of installation, integration, additional features etc.)

Popular open source agents:

Beats https://www.elastic.co/beats/
NXLog https://nxlog.co/

Agentless

Agentless log sending process is sometimes preferred as there is no installation and update cost.
Usually, logs are sent by connecting to the target with SSH or WMI.
For this method, the username and password of the log server are required, therefore there is a risk of the password being stolen.
Easier to prepare and manage than the agent method. However, it has limited capabilities and credentials are wrapped in the network.

SIEM and Log Collection

Manual Collection

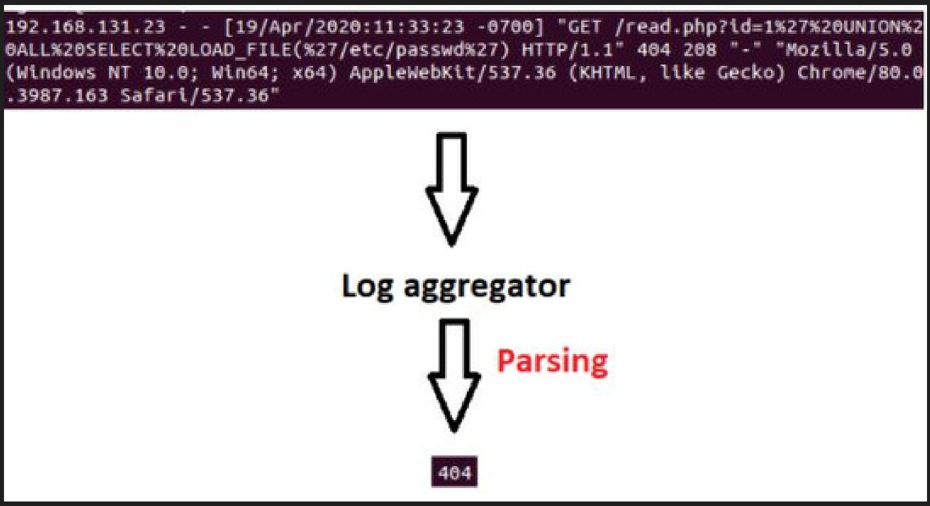
Sometimes there are logs that you cannot collect with existing agent software. For example, if you cannot read the logs of a cloud-based application with the agent, you may need to write your own script

Summary

As you can see, there are various ways to collect logs. These are agents and agentless. In cases where the agents on the market are not sufficient, you should write your own scripts.

Log Aggregation and Parsing

The first place where the generated logs are sent is the log aggregator. We can edit the logs coming here before sending them to the destination. For example, if we want to get only status codes from a web server logs, we can filter among the incoming logs and send only the desired parts to the target.



```
192.168.131.23 - - [19/Apr/2020:11:33:23 -0700] "GET /read.php?id=1%27%20UNION%2
0ALL%20SELECT%20LOAD_FILE(%27/etc/passwd%27) HTTP/1.1" 404 208 "-" "Mozilla/5.0
(Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0
.3987.163 Safari/537.36"
```

```
404
```

Aggregator EPS

What is EPS?

EPS is an event per seconds. The formula is Events/Time period of seconds.
As the EPS value increases, the aggregator and storage area that should be used also increases.

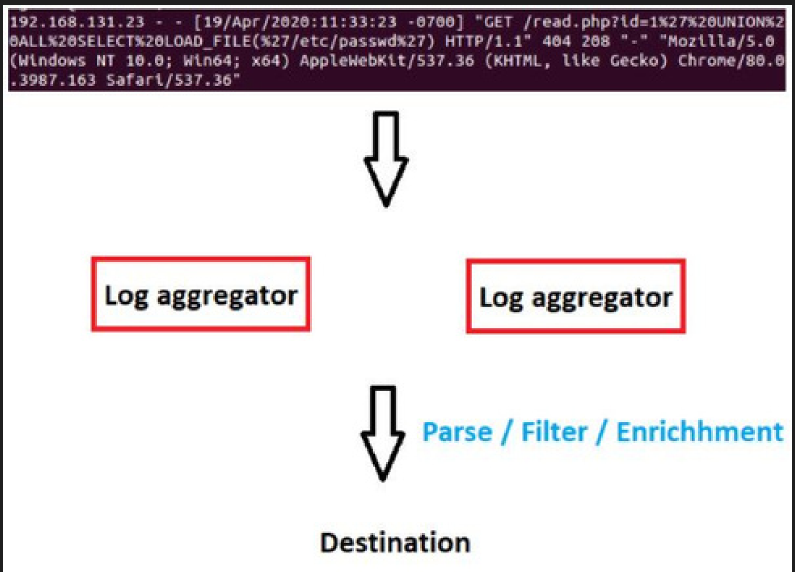
Scaling the Aggregator

More than one aggregator can be added so that the incoming logs do not load the same aggregator each time. And sequential or random selection can be provided.

Log Aggregation and Parsing

Log Aggregator Process

The log coming to the Aggregator is processed and then directed to the target. This process can be parsing, filtering, enrichment.



```
192.168.131.23 - - [19/Apr/2020:11:33:23 -0700] "GET /read.php?id=1%27%20UNION%2
0ALL%20SELECT%20LOAD_FILE(%27/etc/passwd%27) HTTP/1.1" 404 208 "-" "Mozilla/5.0
(Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0
.3987.163 Safari/537.36"
```

Log Modification

In some cases, you need to edit the incoming log. For example, while the date information of most logs you collect comes in the format dd-mm-yyyy, if it comes from a single source as mm-dd-yyyy, you would want to convert that log. Another example, you may need to convert UTC + 2 incoming time information to UTC + 1.

Log Enrichment

Enrichment can be done to increase the efficiency of the collected logs and to save time.
Example enrichments:

- Geolocation
- DNS
- Add/Remove

Log Aggregation and Parsing

Geolocation

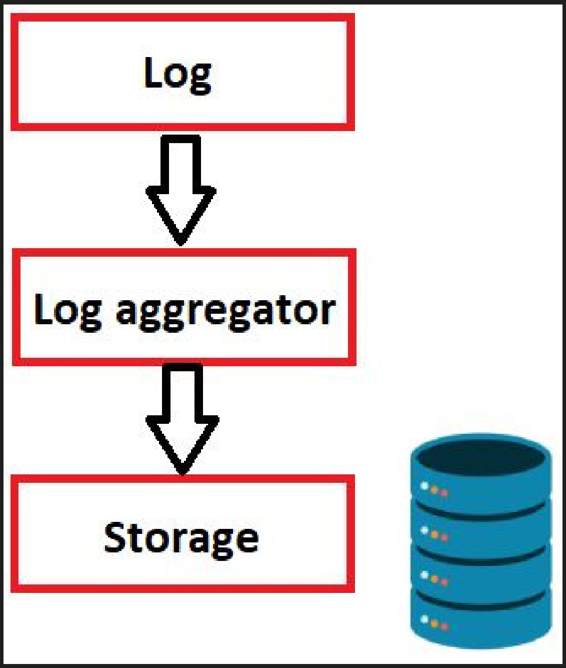
The geolocation of the specified IP address can be found and added to the log. Thus, the person viewing the log saves time. It also allows you to analyze location-based behavior.

DNS

With DNS queries, the IP address of the domain can be found or the IP address can be found by doing reverse DNS.

Log Storage

In our previous articles we talked about logs and log aggregators. The next step is to store incoming logs.



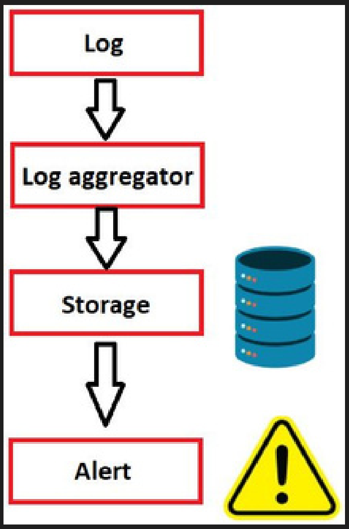
One of the common mistakes made in SIEM structures is to focus on storage size. High-sized storage is important, as well as the speed of accessing this data.

When we look at the popular storage technologies in the market (Example: mysql), we see that it is focused on adding, editing, and deleting data. But our focus is on indexing the data, we do not intend to edit the stored log later. Our purpose is to access data as quickly as possible. For this, WORM (write once read many) based technologies are more suitable to be used in SIEM.

More info about worm once read many: https://en.wikipedia.org/wiki/Write_once_read_many

Alerting

We have collected, processed and stored logs up to this point. Now, we need to detect abnormal behavior using the data we have and generate alerts.



Timely occurrence of alerts varies depending on our search speed. For a log created today, we want to create a warning immediately instead of generating a warning after 2 days. Therefore, as we mentioned in our previous article, a suitable storage environment should be created.

The alarms we will create for SIEM will usually be suspicious and need to be investigated. This means that the alarm must be optimized and not triggered in large numbers (except in exceptional cases).

Here are some ways to create an alert:

- By searching stored data
- Creating alarms while taking logs
-

Example alarms that can be created:

- New user added to global administrator
- 15 Login failed in 3 minutes with the same IP address
-

In order to create a quality alarm, you must understand the data you have. Some of the techniques for making better log searches are blacklisting, whitelisting and long tail analysis.

Alerting

Blacklist

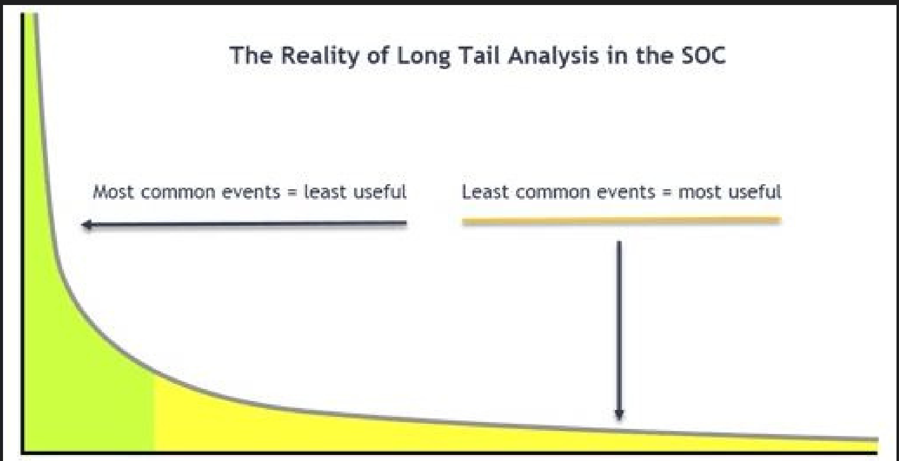
It can be used to catch undesirable situations. For example, we can collect the prohibited process names (Example: mimikatz.exe) and write them to a list. Then, if a process in this list appears in the logs, we can create an alert. Similarly, an alarm can be generated when there is a device that creates and accesses a banned IP list.
It is easy to manage and implement, but very easy to bypass. For example, if the name mimikatz2.exe is used instead of mimikatz.exe, no alarm will occur.

Whitelist

Unlike blacklist, it is used for desired situations. For example, a list of IP addresses with normal communication can be kept. If communication is made with an address other than this list, we can generate an alarm.
This method is highly effective but difficult to manage. The list needs to be constantly updated.

Long Tail Log Analysis

This method assumes that the behaviors that occur constantly are normal. In other words, if an "Event ID 4624 An account was successfully logged on" log is constantly occurring on a device, with this method we should take it as normal and approach the least occurring logs with suspicion.


The Reality of Long Tail Analysis in the SOC
Most common events = least useful
Least common events = most useful