Threat Detection Bad Trips: Log Everything!

Alright, this is a public service for SIEM engineers. How many times you've heard that answer when someone asks 'What should be logged?' and how hard is it today to predict the outcome of over-logging?

Disclaimer: all opinions here are applicable to enterprise, corporate environments only. That's where most of my experience comes from. It doesn't apply to SMBs or military environments, for instance.

# Know your recipes

When you go shopping for ingredients to make food, you typically have an idea of what you want to prepare because you already know your recipes.

You may be able to slightly change your mind in case you find a promotion or a super fresh ingredient you were not counting on initially. That’s fine.

However, if you are cooking for a large group of people, it is likely that you plan ahead by determining which dishes you will prepare and what ingredients you will need to fulfill the requirements.

You guessed it correctly! How can I define what needs to be logged if I don’t know the detections/indicators, reports and other use cases for those logs?

Rotten potatoes or unattended, untouched logs?

# Where has it started?

There are two common answers in consulting engagements you should be very familiar with, even if you are just starting a career:

* It depends.
* Define (or give me) the requirements.

Regardless of the case, you know you are in for some research and much work before figuring out what needs to be done.

Needless to say, in case someone hasn't decided to follow the rabbit hole, that's likely when Log Everything pops up.

Vendors love that take, BTW!

Whether you realized logging everything is not a good idea because the system was overloaded or because most logs doesn't provide any value for threat detection, one thing is for sure: you are paying the bill.

Some SIEM vendors still charge per data ingestion, which means it's not something they will easily oppose. You need more inputs.

There are many talented sales teams who know how to balance that (value focus), but in the end of the day, they all need to meet their target$.

So, what should be the answer to ‘What should be logged?’ That’s the wrong question, you should reply!

# The Consequences

Just in case you haven't committed that mistake (yet), I'm going to list below some of the issues one might encounter by going the 'Log Everything' way:

## Data Acquisition costs

On-boarding logs is sometimes a long and costly process. And it’s not simple tied to storage and licensing costs. Consider the amount of time the teams need to chase data owners and fine-tune parsers/extractors.

## Performance Degradation

On agents and servers. Depending on how scalable or elastic is your setup is, adding more data has a direct impact on query performance. That might lead to other issues such as low user/consumer satisfaction and false-negatives from skipped jobs, timeouts, etc.

## Bad Expectations

If you log everything, you should be able to deliver every single detection use case, correct? Of course, not. It takes more than data to deliver a solid use case. Also, bothering stakeholders (server admins) asking for more and more data without demonstrating any value is not a good strategy in the long term.

In the end, here's the most important message: logging everything will lead to more time spent on data readiness, which slows down time-to-value!

# The Compliance/Accountability/Forensics infamous use case

If the only use case applicable is around post-mortem, then consider storing the data somewhere else cheaper, but not in an Analytics Platform. That's a MUST in case you have Data Pipeline Management in place which makes it easier to route and later decide where logs should be forwarded to.

# Getting data is hard, making sense of it, is harder

Many organizations are still not able to plan and deliver detection use cases, so a logical first step is to start collecting data. However, that gives a false sense of progress with the consequences exposed earlier here.

Be sure to understand what drives a good detection use case, what problem needs to be solved and where that fits in your Threat Detection program. After that, log data requirements become clearer and more engaging.

That's all for this short one! Feedback always welcome!