Driving business value

Leveraging ingestion based security for smarter investments

RelianceCyber

THOUGHT LEADERSHIP E-BOOK SERIES

The context

I'd like to frame the discussion by outlining some fundamental principles, none of which are likely to be particularly contentious.

Firstly, any business that relies on technology is at risk of cyberattack. The higher the business' dependence on technology, the more impactful that a major cyber-attack would become. Secondly, preventative controls such as firewalls, multifactor and endpoint security products are important, but are not sufficient to either detect or deter modern threat actors. It is difficult to configure them all correctly, and prohibitive to monitor them continuously for all but enormous companies. The higher the number of controls, the more complex and difficult their configuration and monitoring becomes.

Enter Security Incident and Event Management (SIEM) platforms. SIEMs allow for centralised visibility of multiple data points such as cloud platforms, network appliances, endpoint logs, applications and much much more. Deployed, configured and monitored correctly, they provide a highly scalable single pane of glass through which to monitor any environment.   They are not, however, without their challenges, which are:

1. Integrating multiple different types of technologies with a SIEM is difficult, particularly when many businesses use niche or sometimes unique applications or platforms to perform their core business – and that these technologies are generally hugely important to the overall business goals

2. SIEMs work based on rules. Alerts are generated when conditions are met (or are absent) in logs. Creating those rules, tuning them to the business' needs and ensuring that they're constantly updated is non-trivial and requires significant expertise and a rich picture of the threat landscape

3. SIEMs alone don't assist in response. Secondary tooling or processes are still required to take an action based on an alert, requiring SIEM alerts to trigger an automated or manual response to contain or eradicate an identified threat

4. SIEMs can get expensive quickly, particularly if one takes the view of "the more data, the better", which is commonplace in the industry

The latter point is the focus of this discussion.

A move to SaaS and data ingestion charges

The continuous march to the cloud has afforded the Security Operations (SecOps) discipline some major advantages – reducing the complexity involved in hosting a SIEM and affording near infinite scalability.

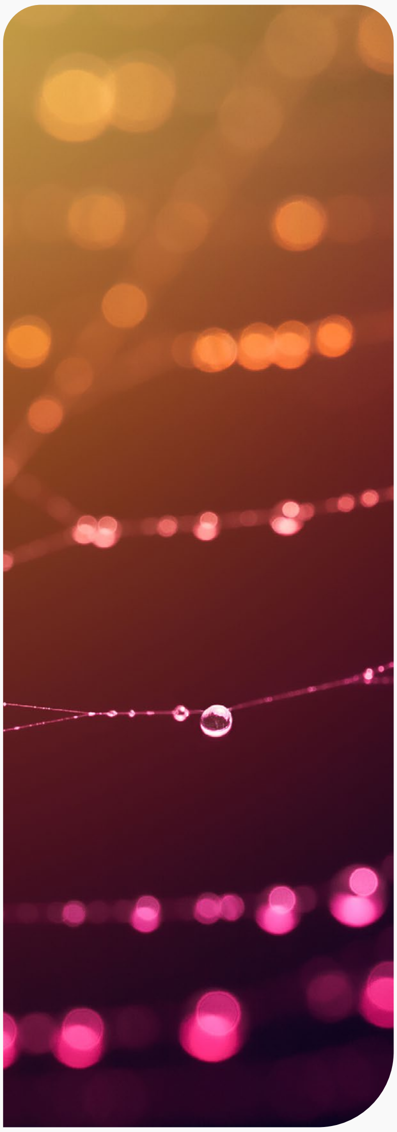
Since data and power requirements for businesses are only increasing over time, and since there is a growing reluctance to increase on-premises footprints, cloud-based SIEMs will truly be essential to maintaining a clear view of what is happening within our information environments.

From a vendor's perspective, a move to the cloud has been attractive, since they can focus on re-selling a product/platform without the need to continuously concern themselves with assisting on-premises deployments – or manually collating usage and licensing telemetry on platforms over which they have no direct visibility or control, as is the case with on-premises SIEMs. All in all, SaaS based SIEMs are generally a no brainer, and a win-win from the customer and vendor's perspective.

Nevertheless, a bill must be paid. Since vendors must now supply the network infrastructure, the compute and the data storage, they need a way of reducing their exposure to cost risk as the platforms become widely adopted. Select any top tier vendor – Microsoft Sentinel, Google Chronicle, Splunk Cloud and LogRhythm's new Axon platform to name just a few – and you'll note that all offer (and in some cases exclusively offer) an ingestion based model.

In theory, this makes sense. For the vendor, they can use data volume to quite accurately estimate the compute, network bandwidth, energy and storage required to deliver their product, increasing the price of consumption as the Gigabytes of ingestion increase – whilst generally offering reductions in price per Gigabyte as a business ascends through predefined tiers of data ingestion.

For the customer, they pay for what they need – sending the correct amount of Gigabytes of data to the SIEM will ensure that they remain within budget and that the platform will be performant. Furthermore, whilst there may be some other less obvious caps (such as number of rules that can be added, number of alerts per hour/day that can be generated, number of API calls that can be made etc), the main cost and performance consideration is firmly focussed on ingestion.



The implications for security

So what's the issue? This all sounds good, right? Well, not really – or at least, not always. The fundamental problem, which we'll unpick further below, is that ingestion-based SIEM licensing focuses the conversation on the volume rather than the value of data. The model has some significant issues, that **mostly place the risk on the customer**. These are:

1

There is a temptation to prioritise the budget over the outcome. For example, one might look at cost options and determine that it is affordable to consume 50 Gigabytes of logs per day in a given SIEM solution. This becomes a target, and often areas of the infrastructure go unmonitored until budgets can be increased – which may never happen.

2

Estimating the correct amount of Gigabytes of data that must be sent is non-trivial. There is no such thing as an accurate average amount of logs from – for example – a server or a firewall, because there is no such thing as an average server or firewall. Sure, there are handy calculators out there, but these offer an approximation at best, which are often wrong in practice given a wide variety of environmental variables.

3

The use of such calculators fails to consider any meaningful security value. For example, the fact that 50 servers may send 5 Gigabytes of logs per day conveys nothing of the importance of those servers to the business or whether or not they should be monitored. One should be asking, where is the important data stored? What things are likely to be true if our critical systems are breached? Again, the emphasis on ingestion costs forces one to think about the volume rather than the value of data.

To be clear, I do not mean to suggest that vendors have behaved in any way unethically. Their exposure to risk is simply far higher than that of individual businesses, since an unaffordable cost model which incurs huge losses might risk a discontinuation of their SaaS solution. This could lead to a return to legacy on-premises SIEMs, which is in nobody's interest. Furthermore, the vendors will generally happily accommodate an increase in ingestion if the original estimate was low.

Going the other way can sometimes be a slightly more difficult conversation.

This focus on volume over value is played out perfectly in traditional commercial processes. As an MSSP, we answer our fair share of RFPs from businesses seeking a SIEM/SOC solution. Invariably, these offer a reasonable description of the business goals, prior to providing some metrics around the environment (we have x servers, x firewalls, x cloud environments) whilst never explaining the data involved, its security value or how it is typically stored, accessed or protected. Bidders are expected to use the data provided in the RFP to estimate ingestion and to provide a cost of their solution and service, which will be directly scored against competitors, despite key pieces of the puzzle being absent. Similarly, marketplaces and commercial frameworks request per user or per GB pricing, without any discussion of security context or value.

How can this be addressed?

To succeed, it is essential to reverse the core problem that I outlined above. To switch the focus from the volume to the value of data.

One might be tempted to group data sources into 'buckets' of high, medium or low value and high/medium/low volume such as below. We see our competitors do this regularly. Once numbers of users/platforms can be assigned to each data source, one can provide optionality around what to/not to ingest based on cost/value. High value and low or medium volume data is a no brainer to ingest, whereas low value and high volume data can likely be ingested only if budgets are high.

Data source	Value	Volume
M365	High	Medium
Azure Entra	High	Medium
Endpoint Detection and Response	High	High
Firewalls	Medium	High
Core Servers	Medium	Medium
Development Servers	Low	High
Network Switches and APs	Low	High

(This is not intended to be a comprehensive list)

Whilst this is a step in the right direction, it is an overly crude simplification that is perhaps globally true on average, but unlikely to be specifically true for any given business. For some businesses for example, critical test data or intellectual property might exist within development environments, whereas for others no useful data might exist. Furthermore, there are likely to be sub-categories within data sources that are useful to unpick – for example certain firewalls might be protecting critical assets, whereas others are simply a means for office workers to access the internet, whose data is otherwise encrypted or protected. In this case we might ingest only important firewalls, accepting that any gap left from others is filled elsewhere.

There is also some nuance around specific data sources and how high the volume needs to be. Take firewalls, a generally widely-considered 'noisy' data source. It is often possible to ingest only critical Intruder Detection/Protection System or Anti-Virus telemetry, whilst disregarding the raw traffic logs which comprise the overwhelming majority of the volume. Or take servers, one can adapt a collection plan to capture only a subset of the highest value events, rather than simply ingesting everything from each server.
This should be based on the criticality of the asset and the likely threats – and regularly reviewed based on new threat intelligence as it arises.

The Reliance Cyber approach

Our preferred methodology is to undertake a formal process called Threat Modelling.

Threat Modelling is a small number of in-depth workshops, led by our consultants and Senior SOC Analysts and Engineers, which steps through a customer's environment from an attackers perspective. We prepare in advance, researching the customer's assets and profile in our Threat Intelligence platforms. The workshops are dynamic, but explore three key themes:

1. **How does the business work? In which platforms or applications does data reside, and how is it stored, protected and accessed? What systems and services are fundamental to business operations, and which perform a supporting role? How much data transits these platforms?**

   These questions allow us to understand the criticality and value of various data sources, and what applications and platforms we would need to ingest logs from to monitor to detect a threat to them. In short: This establishes what to monitor.

2. **What do our Threat Analysts assess are the key threats to the business? What are the entry points that an attacker would target and what would the attack chain look like? What conditions would be present in the event of the realisation of these threats and how would they manifest in the data sources?**

   These questions inform the SIEM rule, response playbooks, and supplemental Threat Intelligence, that must be added to the overall solution to monitor any threat.

3. **What controls are already in place to defeat and deter them? How well are they configured and who manages them?**

   This allows us to establish how well the existing controls will defeat likely attackers and any additional supplementary detection and response actions must be in place to fill any gaps.

Answering these questions allows us to produce a collection plan. We know where the essential data resides, how it is normally accessed and what a threat actor would likely need to do to violate their confidentiality, availability or integrity. We can then break data sources into three streams:

1. **Critical** - Data sources that are essential to detecting likely threats against critical assets, or which are central repositories hosting business critical data
2. **Core** - Data sources which would form part of the kill chain and/or which would trigger SIEM rules in the event of any compromise
3. **Non-Essential** - Supplemental data which might be useful post-incident, but which does not need to be ingested into the SIEM platform.

For **Critical** data sources, we would ingest as much telemetry as possible, to provide the greatest potential to detect threats against things which absolutely must be detected.

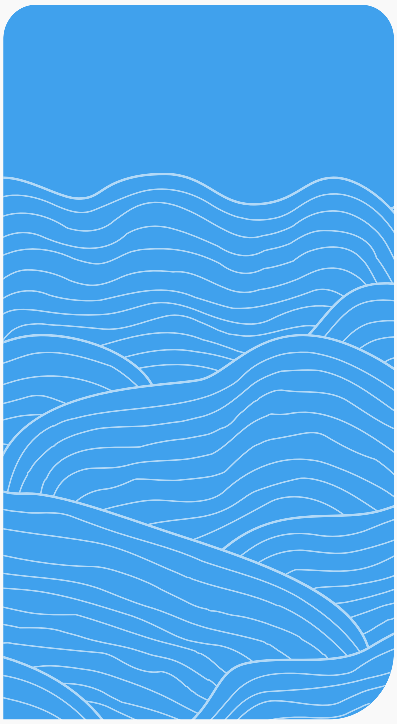
The volume of the sum of critical data would be discovered during Threat Modelling and added to cost models to inform SIEM licensing requirements. For example, for one of our retail customers, Salesforce is the primary CRM platform which hosts all customer and transactional data. We would ingest as much data as possible from this platform, and add rules from our SIEM library to provide comprehensive monitoring.

For **Core** data sources, we would intelligently select the types of data that would need to be ingested and which would be valuable in generating SIEM alerts, or providing context to SIEM alerts generated from Critical Data sources. Again, the volume of the data can be carefully calculated in advance by our experts to produce accurate SIEM licensing pricing. This might, for example, include monitoring a specific subset of events on Windows servers, or certain log types from firewalls or applications. We need not ingest everything – just the events and data of value to support the monitoring of the threats to the environment.

For **Non-Essential** log sources, we generally aim to store these in a less expensive storage mechanism than the SIEM.   These could include raw traffic logs from core switches, inter-VNET traffic in Azure or VPC flow logs in AWS. This data is particularly high volume, but generally not useful in generating SIEM alerts. Having it available in a separate container for use during investigations ensures that we can perform comprehensive analysis, but without pushing SIEM ingestion costs to an unacceptable threshold.

We repeat this process with all of our customers annually or following significant changes to their environment – ensuring that we have a constant rich picture of events and likely threats.

The overall outcome of this process is that only data of value is ingested into the SIEM, that the SOC is not drowned with unnecessary noise, and that we save precious budget to spend on higher value areas.



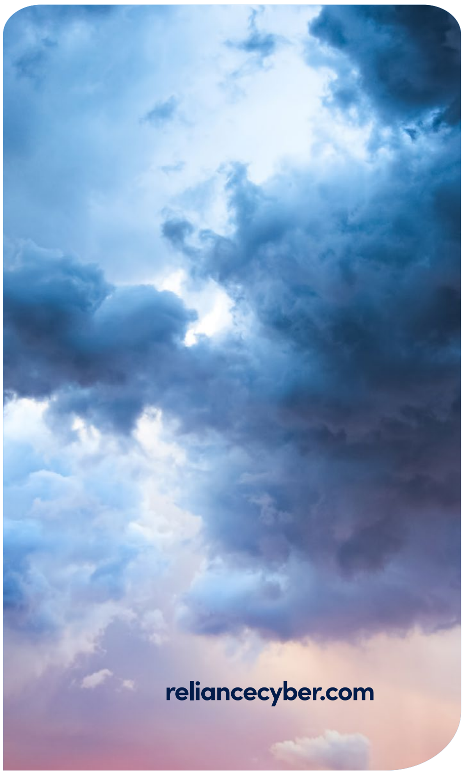
Real world impact on budgets and monitoring

Below are two cases in which Reliance Cyber have consulted on existing SIEM solutions which had been implemented by other partners.

Sector	Previous ingestion	Previous SIEM rules	Post Threat Modelling ingestion	Post Threat Modelling SIEM rules
Manufacturing	200 GB/day	25 basic rules	45 GB/day	400 foundational rules 25 custom rules
Retail	165 GB/day	40 basic rules	28 GB/day	400 foundational rules 12 custom rules

In both cases, the customer pays far less for the SIEM platform, and simultaneously extracts far higher value from the platform due to superior SIEM rules and custom use cases which are tailored to their environment via Threat Modelling.

Only data of value is ingested into the SIEM so that the SOC is not drowned with unnecessary noise and we save precious budget to spend on higher value areas.



reliancecyber.com

Impact on Mean Time to Detect and Mean Time to Respond

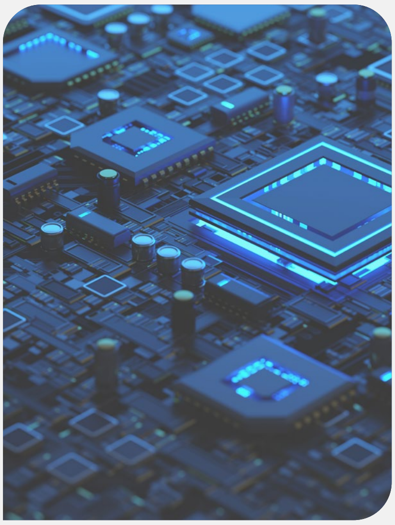
Another closely related and hugely important benefit of reducing volume and focussing on value, is improving our Mean Time to Detect (MTTD) and Mean Time to Respond (MTTR) to security incidents when they occur.

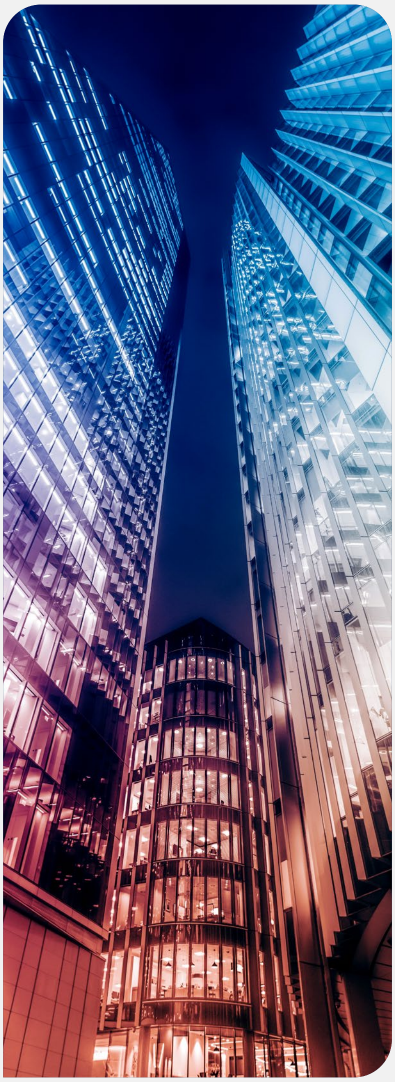
Threat Modelling allows us to understand the environment in detail and to focus our monitoring efforts on the scenarios which are likely to be present in the event of a threat being materialised. Through tailored SIEM rules and response playbooks, we can ensure that everything is in place in advance to detect and respond to these incidents when they arise. This provides a far faster MTTD and MTTR – reducing the attacker's opportunity to impact customer systems and data often by factors higher than 10. Furthermore, by reducing superfluous logs and noise via volume, our searches and hunts are more performant, our analysts are less focussed on trivial and transactional 'alerts' and we can focus our energy on creating better rules and playbooks, conducting granular threat hunts, correlating multiple high value events and producing detailed incident reports with robust recommendations and root cause analysis.

In summary

In summary, reducing the volume of logs and focussing on the value of the data can significantly enhance the security posture of an organisation by improving the detection and response capabilities, as well as reducing costs and complexity.

Customers who are dissatisfied with their existing ingestion costs and concerned about the value they are getting from their logging solution should engage with Reliance Cyber to review their current setup and identify opportunities for improvement. Reliance Cyber has extensive experience in Threat Modelling, SIEM rule development, response playbook creation, and threat hunting, and can help customers achieve a more efficient and effective security monitoring and response strategy.





Leveraging ingestion based security for smarter investments

Explore the leadership series

RelianceCyber

reliancecyber.com