FIVE SECURITY ANALYTICS PITFALLS
AND HOW TO AVOID THEM

VARONIS

CONTENTS

> “Organizations are failing at early breach detection, with fewer than 20% of breaches detected internally.[1]

Gartner

VARONIS

INTRODUCTION

Many organizations have started to consider and implement security analytics to help bolster their detective capabilities.

Those beginning their exploration of security analytics and log collection technologies often think they will be able to detect breaches by simply sending logs to a central server for analysis. It's true that logs are required, but it takes a lot more effort to make sense of them than most people realize.

Here are five pitfalls organizations encounter when they try to *proactively* mine their logs and investigate security incidents:

1. There are a lot of logs – most organizations will need to store hundreds of millions of events per day. Network devices, endpoints, security systems, applications, storage devices, proxies – they all write events prolifically, and each device type and vendor writes logs in their own way.

2. They aren't usable in their raw form. In order to use logs, one must parse them, or recognize the objects they describe. This is a user, this is a device, this is a logon event, etc. Until logs are parsed, there is no hope of relating the objects in one log to the objects in another– which is required for both forensics and proactive analytics. This is harder because logs don't come in a standard format -- since they're all different, each format must be parsed. Additionally, some logs take a few lines to describe a single "event" and some are written out of order. These logs are better off combined for both humans and analytics technologies.

3. Even after the logs are adequately parsed, they lack context. security analysts need to prioritize and investigate logs that end up as alerts. Who is the user and what do they do? Is this their machine? What office are they in? Security analysts spend a lot of time chasing this sort of information down in order to determine the nature of the security event, or if it qualifies as a security event in the first place.

4. Individual events have no context. They don't show connections to events that have happened before or events happening on different systems, and security incidents can occur over weeks, months or longer. They don't indicate the user's role, if this is their usual workstation, their normal location, if any data they access is sensitive, or if anything else is unusual about the event. Analysts must frequently review thousands of events to answer these questions and build enough context to understand and respond to a single incident.

5. The trail often goes cold once it comes to the most critical question: *Is our data safe?* This is because data access activity frequently isn't captured, stored or analyzed. For example, many organizations neither capture nor store any information on how users interact with files or emails – the subject of many data breaches.

These pitfalls help explain why raw logs yield relatively few meaningful alerts, and investigating them takes a lot of skill and time. This document describes how security analytics can overcome these pitfalls to reduce false positives, accelerate investigations, and stop more attacks more quickly.

> SIEM is not log collection, where the goal is to capture and store all logs from all devices and applications without discrimination. Yet a common mistake is to approach it this way, thinking it will be easy to make sense of all of this data once it is in the SIEM system. The predictable result is that what should be an exercise in reducing noise actually amplifies it and generates more of it. Finding a needle in the haystack does not benefit from increasing the amount of hay.[1]

Gartner

VARONIS

INTELLIGENT COLLECTION

Collecting logs doesn't seem like it should be more complicated than configuring your devices to write to a syslog server, but device logs are noisy and write many lines that describe a single "event."

The lines aren't always in the right order, and from a security perspective, they aren't all relevant. Sometimes they write to multiple log files that must be combined. To make things more complicated, devices are different, their logs are different and they change between versions. Each vendor logs different things, and they use different formats for things like user names, host names and domains.

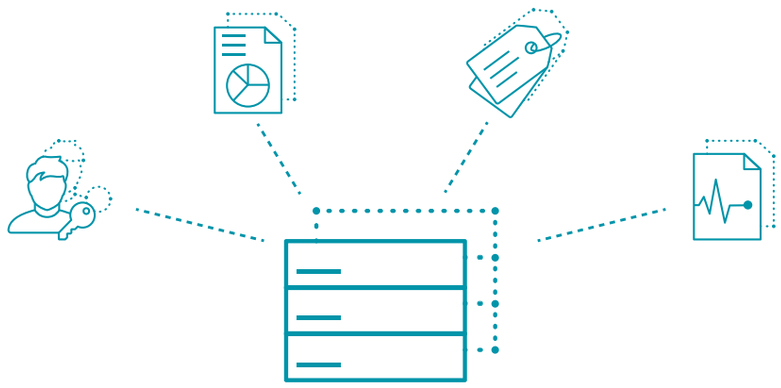
For example, the start of a remote VPN session will be spread across 10-20 individual log events, which often aren't in order as many users are interacting with the system simultaneously.

Here’s what a raw log looks like for a single VPN connection from one VPN vendor:

```
Dec  6 13:07:52 127.0.0.1  2017-12-06T11:08:09Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:09 - ive - [84.229.12
0.164] oshezaf(VaronisCertificate)[Varonis-User, PM] - Key Exchange number 1 occurred for user with NCIP 172.16.248.93

Dec  6 13:07:52 127.0.0.1  2017-12-06T11:08:09Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:09 - ive - [84.229.12
0.164] oshezaf(VaronisCertificate)[Varonis-User, PM] - VPN Tunneling: User with IP 172.16.248.93 connected with ESP tran
sport mode.

Dec  6 13:07:48 127.0.0.1  2017-12-06T11:08:05Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:05 - ive - [84.229.12
0.164] oshezaf(VaronisCertificate)[Varonis-User, PM] - Starting dsagentd session.

Dec  6 13:07:48 127.0.0.1  2017-12-06T11:08:05Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:05 - ive - [84.229.12
0.164] oshezaf(VaronisCertificate)[Varonis-User, PM] - VPN Tunneling: User with IP 172.16.248.93 connected with SSL tran
sport mode.

Dec  6 13:07:48 127.0.0.1  2017-12-06T11:08:05Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:05 - ive - [84.229.12
0.164] oshezaf(VaronisCertificate)[Varonis-User, PM] - VPN Tunneling: Session started for user with IPv4 address 172.16.
248.93, hostname OSHEZAF-LT

Dec  6 13:07:48 127.0.0.1  2017-12-06T11:08:05Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:05 - ive - [84.229.12
0.164] oshezaf(VaronisCertificate)[Varonis-User, PM] - Agent login succeeded for oshezaf/VaronisCertificate from 84.229.
120.164 with Pulse-Secure/8.3.3.1021 (Windows 10) Pulse/5.3.3.1021.

Dec  6 13:07:48 127.0.0.1  2017-12-06T11:08:05Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:05 - ive - [84.229.12
0.164] oshezaf(VaronisCertificate)[] - Host Checker realm restrictions successfully passed for oshezaf/VaronisCertificat
e , with certificate 'CN=Ofer Shezaf, OU=Herzliya, OU=IL, OU=Users, OU=Varonis, DC=varonis, DC=com'

Dec  6 13:07:48 127.0.0.1  2017-12-06T11:08:05Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:05 - ive - [84.229.12
0.164] oshezaf(VaronisCertificate)[] - Primary authentication successful for oshezaf/CertificateServer from 84.229.120.1
64

Dec  6 13:07:46 127.0.0.1  2017-12-06T11:08:03Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:03 - ive - [84.229.12
0.164] oshezaf()[] - Host Checker policy 'Varonis' passed on host '84.229.120.164' address 'a4-34-d9-ec-4d-66'  for user
 'oshezaf'.

Dec  6 13:07:46 127.0.0.1  2017-12-06T11:08:03Z  vpn-isr.varonis.com PulseSecure: 2017-12-06 11:08:03 - ive - [84.229.12
0.164] oshezaf()[] - Host Checker policy 'Domain' passed on host '84.229.120.164' address 'a4-34-d9-ec-4d-66'  for user
 'oshezaf'.
```

Though a syslog server may happily collect all of your raw logs, they'll take up a lot of disk space, take a lot of processing power to do anything with, and ultimately won't do you much good. VPN logs are sizable; DNS and Proxy logs are even more voluminous.

It's both more efficient and more advantageous to do a bit of processing, pruning and parsing upstream, especially if you're eventually sending these logs to a system that charges by the megabyte. Security Analytics solutions can prune the raw logs at the point of collection, potentially reducing the amount of data by 70-80%. By parsing and resolving some of the logs upstream (this is a user logging in, this is a host, etc.), these logs are prepared for quick central analysis. An intelligent collector that smartly parses, prunes and aggregates raw events can even perform some analytics and alerting at the point of collection. For example, an intelligent collector might generate an alert in near-real time when a particular user tries to log into a VPN (instead of waiting for the raw logs to be backhauled and analyzed by a central server).



> During research, the majority of SIEM providers told Gartner that the mass of their installed base (approximately 85%) is not using advanced threat detection or analytics features today.[2]

Gartner

VARONIS

ENRICHMENT AND ANALYTICS

Let's say you've implemented intelligent collection, pruning and parsing so your logs are in pretty good shape. At this point you have a far better solution for forensics than combing through raw logs, but you still have some work to do before you can derive meaningful, proactive insights with analytics. Effective analysis requires context about users, systems, and data.

A user may be an executive with access to sensitive data, an administrator with access to key infrastructure, or someone who recently resigned. A system may be a critical server, a workstation, or a test system. Files may contain personal information or critical IP. Some files are just pictures of cats.

Without this context, it's very hard to tell the difference between something important and something inane. A non-administrative user running an administrative tool, like a sniffer (and generating a bunch of DNS queries) is probably cause for an immediate lockdown of account and workstation; a known administrator running a sniffer (and generating a bunch of DNS queries) might merit an email or a phone call. A massive download to an unusual location really matters if the user or workstation recently accessed personal data or critical IP. Pictures of their kids, probably not so much.

Not only do the events need to be enriched with context for efficient processing, but context must be built and refined over time. Users access different data sets on different systems, from different workstations, at different hours from different places. This is where machine learning can be really effective – building and maintaining baselines of what normal behavior looks like for the interactions between all users, systems and data.

One final point about security analytics – it only works well if it has enough of the right data, or metadata, to analyze. If data is the asset you're most concerned about protecting, you'll need to understand whether anyone has actually been able to access your data. If you have critical data stored in file and email systems, file and email system activity makes up the heart of the story – without it, you can't answer the most important security question of all: "is our data safe?"

Unfortunately, raw logs for file and email activity are frequently unavailable. When they are available, they are raw and voluminous, like perimeter telemetry. If the safety of your data matters, a data-centric technology designed to provide context about data usage and data sensitivity – across as many of your core data stores as possible – will be a huge advantage.

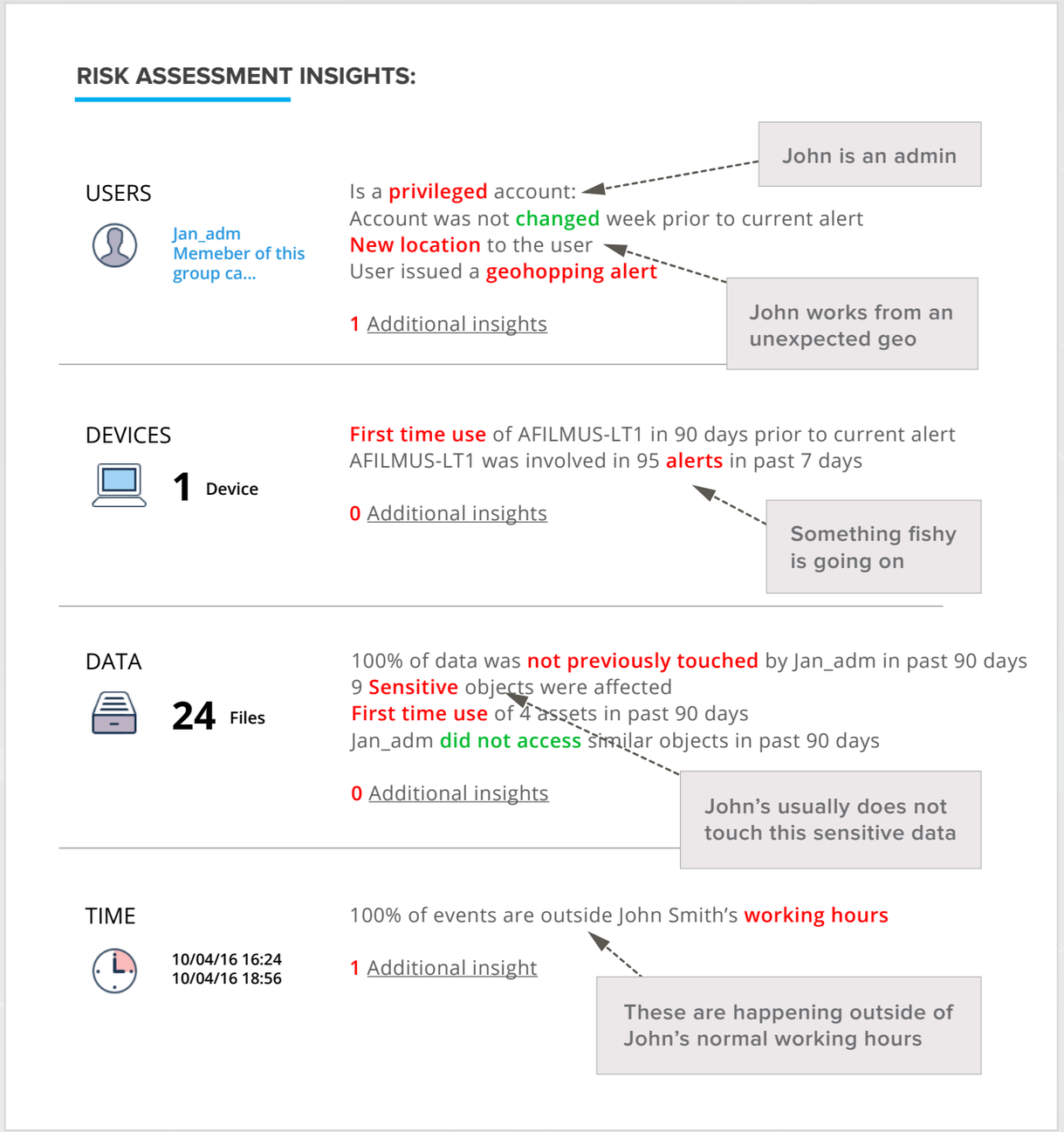
RESPONSE

Security analysts have no shortage of alerts – malware detected on a workstation, an account locked out, a successful login from the South Pole. Not only do raw, unparsed events generate more alerts, each alert takes far longer to investigate. In order to know whether or how to respond, the analysts needs to correlate the events manually – a painstaking, time-consuming process.

For example, let's say an analyst, receives an alert from a malware detection system: "malicious file detected at 10.10.150.12." The first step might be to identify the workstation, call its owner, and then see if it has been really been infected with malware. If it has, then a next step might be to query the proxy logs to determine where the malware came from, whether any connections have been made to unusual locations, and/or large uploads initiated. If they have, then worries about whether any sensitive data has been accessed start to cause dyspepsia, and the investigation continues.

Security analytics speeds this process up dramatically. Analysts get fewer alerts, they're more meaningful, and they're easier to analyze – especially if all the correlation and context is presented along with the alert.

To build an investigative timeline and determine the extent of an incident, analysts must know about the user account, the device, the data, and the time of the alert(s). Security analytics reveal whether the user is accessing the network from a normal location (for them), if the account is privileged, if sensitive data was accessed, and if the event occurred during a user's normal time window. This context helps them determine whether an alert represents a real compromise or an insignificant anomaly.



Event Time	Event Type	SAM Accou	Event Statu	Blacklisted.	Country	Connection Type	Upload Siz.	Download .	Session Du	IP Address	External IP Addr.
12/05/2017 4:15 PM	VPN login request		✔			Unkonwn					192.168.200.89
12/05/2017 4:15 PM	VPN login request		✔			Tunneling				172.16.212.150	192.168.200.89
12/05/2017 4:18 PM	VPN logout request		✔			Tunneling	265349	41638	158	172.16.212.150	192.168.200.89
12/07/2017 9:48 AM	VPN login request	dpnini	✔	–	Israel	AccessApplica...					89.139.198.93
12/07/2017 10:02 AM	VPN logout request	dpnini	✔	–	Israel	Unkonwn					89.139.198.93

If the analyst, using this context, decides to take further action, clean events will be available and connected to the incident.

CONCLUSION

Security analytics that combines intelligent collection of the right metadata, smart parsing and enrichment with machine learning reduces the overall number of alerts, and decreases the time it takes to investigate them. With fewer, more meaningful alerts, analysts have a far better chance of catching real security incidents more quickly, and seconds count when it comes to cyber security.

A user on a watch-list that uploads sensitive data to a website immediately after accessing it during non-work hours will be at the head of the investigative queue, along with the administrator that's reading the CEO's emails and marking them as unread over the VPN from somewhere warm and sunny. An account that's supposed to be running your database will light up when it suddenly starts accessing patient data, but a user updating dozens of files at the end of the month during normal workhours from their regular workstation doesn't make a sound, because that's just what they do.

Whether you're just considering a log consolidation or SIEM project, or you find yourself falling into feeling that your solution is too dumb or too slow, you might consider trying a security analytics solution. In addition to increasing your chances of catching important security events, an organization that employs good security analytics will reduce the time spent per investigation, processing overhead, disk space requirements (and associated consumption costs) and address compliance requirements more easily.

ABOUT VARONIS

Varonis is a pioneer in data security and analytics, fighting a different battle than conventional cybersecurity companies. Varonis focuses on protecting enterprise data on premises and in the cloud: sensitive files and emails; confidential customer, patient and employee data; financial records; strategic and product plans; and other intellectual property.

The Varonis Data Security Platform detects insider threats and cyberattacks by analyzing data, account activity and user behavior; prevents and limits disaster by locking down sensitive and stale data; and efficiently sustains a secure state with automation.

With a focus on data security, Varonis serves a variety of use cases including governance, compliance, classification, and threat analytics. Varonis started operations in 2005 and, as of December 31, 2017, had approximately 6,250 customers worldwide — comprised of industry leaders in many sectors including technology, consumer, retail, financial services, healthcare, manufacturing, energy, media, and education.

Live Demo

Set up Varonis in your own environment. Fast and hassle free.

info.varonis.com/demo

Data Risk Assessment

Get a snapshot of your data security, reduce your risk profile, and fix real security issues.

info.varonis.com/start

[1] Gartner, Using SIEM for Targeted Attack Detection, Oliver Rochford & Kelly M. Kavanagh, 12 March 2014
[2] Gartner, Summer of SIEM 2017 Coming..., Anton Chuvakin, 11 July 2017

ABOUT DATALERT

DatAlert automatically analyzes the information collected by our Data Security Platform to detect, notify and respond to threats to your data in near-real time. With DatAlert, get notified when something needs urgent attention – like someone accessing or encrypting a bunch of sensitive files, reading an executive's email, or making changes to group policy outside of normal change control hours.

LEARN MORE

"Varonis is a Fantastic Solution

Gartner peerinsights™

★★★★★

VARONIS

ABOUT VARONIS EDGE

Varonis Edge analyzes perimeter devices like DNS, VPN, and Web Proxy to correlate events at the perimeter to detect malware, APT intrusion, and exfiltration. DatAlert and Edge detect suspicious activity and prevent data breaches across platforms, visualize risk, and prioritize investigations.

LEARN MORE

Thousands of the world's top enterprises trust Varonis to manage and protect their data.

NASA ING Nasdaq CHAMPAGNE BOLLINGER MAISON FONDÉE EN 1829 DELL EMC

TOYOTA LUXEMBOURG INSTITUTE OF HEALTH RESEARCH DEDICATED TO LIFE L'ORÉAL

VARONIS