Beginner's Guide to **Log Correlation**

UNDERSTANDING THE MOST POWERFUL FEATURE OF SIEM

www.alienvault.com

Q: What’s in the Logs?
What’s In the Logs?!!

A: The information you need to answer
“**Who’s attacking us today?**” and
“**How did they get access to all our corporate secrets?**”

We may think of Security Controls as containing all the information we need to do security, but often they only contain the things they have detected – there is no ‘before and after the event’ context within them. This context is usually vital to separate the false positive from true detection, the actual attack from merely a misconfigured system.

Successful attacks on computer systems rarely look like real attacks except in hindsight – **if this were not the case, we could automate ALL security defenses without ever needing to employ human analysts.** Attackers will try to remove and falsify log entries to cover their tracks – having a source of log information that can be trusted is vital to any legal proceeding from computer misuse.

It’s Always in the Logs....

84% of organizations that had their security breached in 2011 had evidence of the breach in their log files.

Most system log files don’t contain entries that say, “Help! Help! I’m being attacked!” Yet when a skilled human reads those log files, they can show the sequence of events that indicates the machine being compromised – **especially when they cross-reference them against logs from other systems...**

…Except When it Isn’t In the Logs.

Logs vary greatly from system to system. With rare exceptions, there are few standards that software and systems adhere to for what is logged and to what level of detail:

- Some logs are in plain language, while others contain cryptic status codes.
- System logs don’t say “Help! Help! I’m being broken into with a compromised account!” – they say “Successful Login from Authenticated User”
- System logs by themselves are dependent on human analysis to make interpretations from.

The Blind Men and The Elephant

Log correlation is about looking at what's happening on your network through a larger lens than can be provided via any one security control or information source. Each system and control on a network has a particular type of tunnel vision – **it sees the world through a particular lens.**

- Your Intrusion Detection only understands Packets, Protocols & IP Addresses
- Your Endpoint Security sees files, usernames & hosts
- Your Service Logs show user logins, service activity & configuration changes.
- Your Asset Management system sees apps, business processes & owners

While a Network Intrusion Detection systems sees packets and streams, an Application log sees sessions, users and requests – each is logging the same activity, but from a different viewpoint. Many systems are entirely ignorant of the business processes they serve – It is possible that a web application sees "Joe from Accounting", the underlying webserver only sees "jdobson1."

None of these by themselves, can tell you what is happening to **your business** in terms of securing the continuity of your business processes… But together, they can.

Fixed Points In Time.

Since there are few standards in what information is logged,
and to what detail, most logs describe events as discrete points in time.

```
User jdobson1 disconnected
```

is encountered far more often than

```
User jdobson1 disconnected
after 3hrs 15mins from a session
originating from 192.168.1.10
```

All too often, human analysis is required to extract the things that are inferred within log data, not stated outright.

Time & Space

What may be described in a single sentence in natural language often requires many log entries across a period of time and from multiple sources. For example: "Joe Dobson logged into the time tracking system and updated five people's account information" could require a hundred log entries to demonstrate it happened.

This sequence of events however, can be described too. For example, first finding Joe's session ID in the web application, and then matching that session ID to database change logs in the following time period, correlating them together into a single section of logs.

Highly Logical, **CAPTAIN**

Log correlation is about constructing rules that look for sequences and patterns in log events that are not visible in the individual log sources.

They describe analysis patterns that would require human interpretation otherwise, tied together by Logical Operators. Example:

IF a new user
IS created on the domain
AND a new change control ticket
IS NOT created in the change control database

I’d Like a Second Opinion

No Security Control is perfect.
We talk of “False Positives” and “Tuning” of security controls,
but when something actually does happens,
there will usually be more than one record of the occurrence.

If Event B only happens if Event A occurs first, and we see
Event A followed by Event B, we know that Event A actually did occur.

“Web Proxy detected possible Malware from a site was downloaded to a host.
Antivirus on that Host reports malware was detected and removed”
– we can absolutely confirm that this site is serving malware.

Everybody is Different
Everybody is the Same.

Log correlation allows for the creation of alerts that represent what is important to your business processes and security risks. Events from different sources can be combined and compared against each other to identify patterns of behavior invisible to individual devices. They can also be matched against the information specific to your business. Correlation allows you to automate detection for the things that should not occur on your network.

Done correctly, Log Correlation is the difference between reacting to:

```
POSSIBLE-EXPLOIT: mssql improperly formed packet headers
```

or

```
User In Accounting Department seen logging into
Financial Database from a workstation in Customer
Support Department
```

The Short Version

Log correlation monitors incoming logs for logical sequences, patterns and values to identify events that are invisible to individual systems.

They can perform analysis that would otherwise be done by repetitive human analysis.

They can identify things happening that are unusual for your business processes.

By comparing events from multiple sources they can provide more context and certainty as to what is happening on your infrastructure.

They can prioritize investigation and analysis work by prefiltering log events into meaningful alerts.

But Wait, **There's More!**

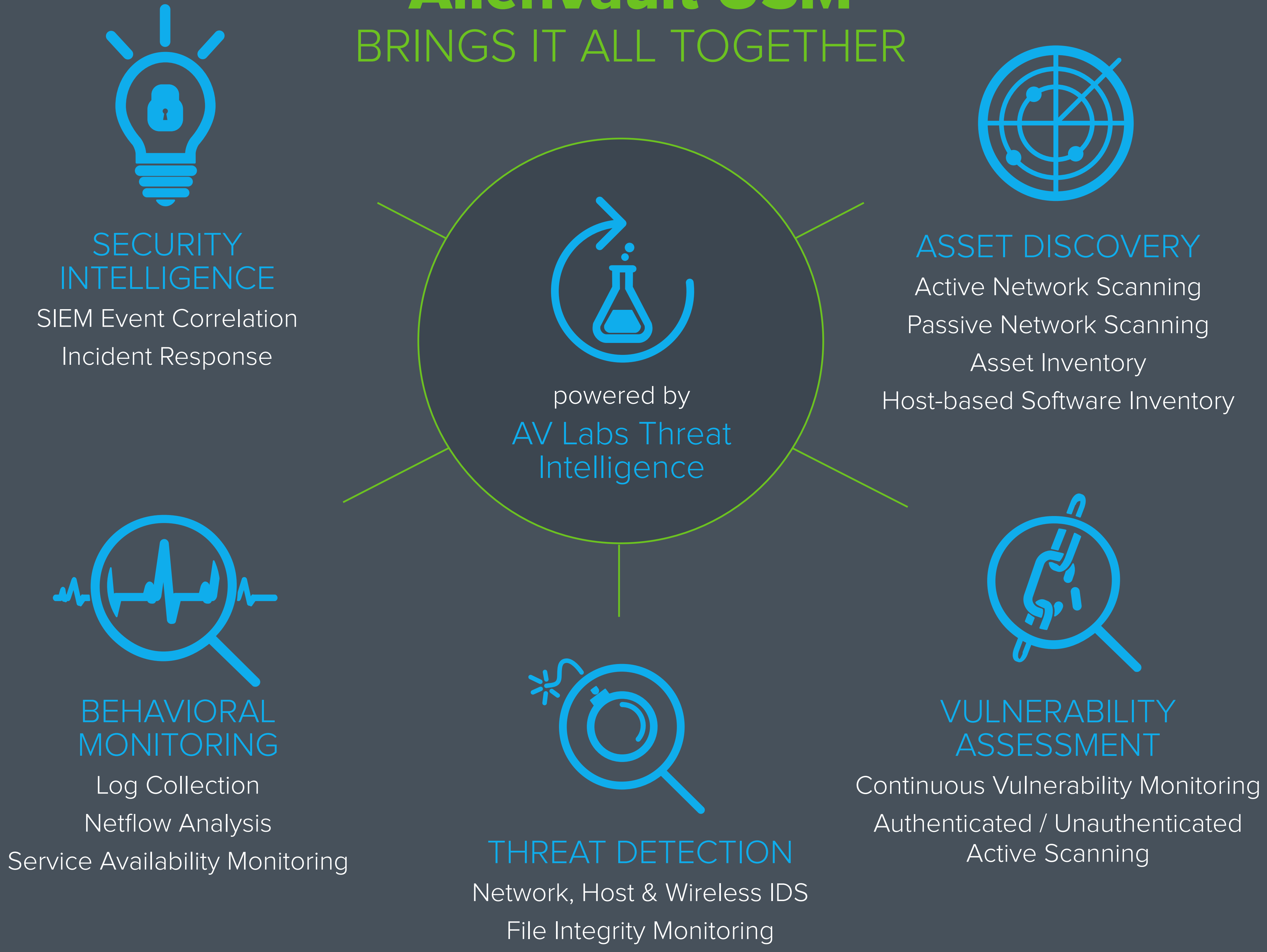
So, now you understand that log correlation can help you:

- Compare events from multiple sources to track users and processes across systems
- Track events across time periods to look for sequences of activity that should not normally occur
- Encode human knowledge about what is and is not normal for a system, or indicates a probable attack, into automatic monitoring

But, to keep up with today's threat landscape, knowing what to look for to detect threats is a constantly moving target.

AlienVault USM

BRINGS IT ALL TOGETHER



powered by

AV Labs Threat Intelligence

SECURITY INTELLIGENCE

SIEM Event Correlation

Incident Response

ASSET DISCOVERY

Active Network Scanning

Passive Network Scanning

Asset Inventory

Host-based Software Inventory

BEHAVIORAL MONITORING

Log Collection

Netflow Analysis

Service Availability Monitoring

THREAT DETECTION

Network, Host & Wireless IDS

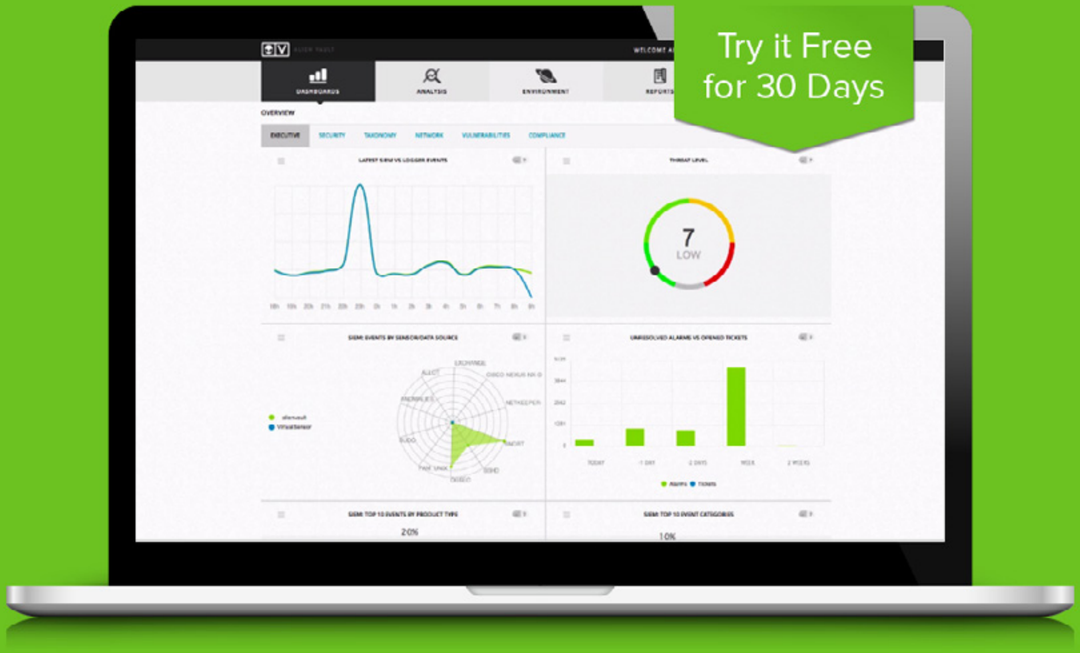
File Integrity Monitoring

VULNERABILITY ASSESSMENT

Continuous Vulnerability Monitoring

Authenticated / Unauthenticated Active Scanning

Next Steps: Play, share, enjoy!



Try it Free
for 30 Days

- Watch our 3-minute overview video
- Play in our product sandbox
- Start detecting threats today with a free 30-day trial
- Compare USM to traditional SIEM
- Join the Open Threat Exchange

www.alienvault.com