



DIPL.-ING. ANDREAS  
MERTZ, CISSP

## SIEM & SEARCH – Intelligente Verarbeitung von Security Events

Schwierigkeitsgrad:



Angesichts des erdrückenden Logaufkommens sind Security-Analysten heute kaum mehr in der Lage, mit manuellen Methoden oder eigenentwickelten Skripten Logdaten und Sicherheitsmeldungen zeitnah aufzubereiten und auszuwerten. Folglich finden, wenn überhaupt, nur Stichproben statt.

### Logfluten und Ressourcenmangel

Angesichts des erdrückenden Logaufkommens sind Security-Analysten heute kaum mehr in der Lage, mit manuellen Methoden oder eigenentwickelten Skripten Logdaten und Sicherheitsmeldungen zeitnah aufzubereiten und auszuwerten. Folglich finden, wenn überhaupt, nur Stichproben statt.

Ohne eine zeitnahe Analyse von Logevents, ohne zentralisierte Archivierung, ohne Alerting und Trendanalysen, fehlt der gesamtheitliche Überblick über die aktuelle Sicherheitslage im Unternehmen.

Die Security-Abteilung kann nur noch reaktiv handeln und ihre Maßnahmen greifen zu spät. Es kommt noch schlimmer: Weil Policy- und Compliance-Richtlinien nicht oder nur ungenügend erfüllt werden, läuft das Unternehmen Gefahr, das nächste Audit nicht zu bestehen.

Was fehlt, sind geeignete Werkzeuge, um mit Datenmengen von mehreren Terabyte effizient zu interagieren. Was fehlt, ist der integrative Ansatz, alle Logdaten an einem zentralisierten Punkt zu erfassen, zu aggregieren, zu korrelieren, zu visualisieren, zu alerten und zu archivieren.



Abbildung 1. Terabyte-Handling 2009

### Status Quo im SIEM-Dschungel

SIEM-Lösungen (Security Information & Event Management) können heute Analyse und Korrelation von Sicherheitsereignissen in Echtzeit (also unmittelbar zum Zeitpunkt des Auftretens) automatisiert durchführen. Durch Erfassung, Normalisierung, Aggregation und Korrelation der Logevents unterschiedlicher Systeme verschiedener Hersteller (Cross-Device & Cross-Vendor Data) können aus zehntausenden von Events diejenigen identifiziert werden, die eine tatsächliche Bedrohung kritischer Anwendungen und Daten darstellen. Das schafft Transparenz und spart Aufwand.

Leider funktioniert dieser Ansatz nicht immer. Klassische SIEM-Lösungen arbeiten Datenbank- / Schema-orientiert und erfordern deshalb die Normalisierung von Logdaten. Wenn Unternehmen jedoch in hohem Umfang eigenentwickelte Applikationen einsetzen, gibt es in der Regel dafür keine vorgefertigten Parser bzw. Konnektoren. An dieser Stelle prallen zwei Philosophien aufeinander: Index-orientierte Suche versus Datenbank-orientierte Korrelation. Für Novizen dieser Thematik ein schwer zu durchdringender Dschungel.

### Strukturierung

Die Lösungsansätze der Hersteller im Bereich Log-Analyse können derzeit in zwei wesentliche Gruppen unterteilt werden – Search und SIEM.

#### Lager 1: IT-Search

Der bestimmende Aspekt von IT-Search-Lösungen besteht in der zentralen Erfassung, Indexierung und Speicherung von Logs sowie der Bereitstellung von Such- und Analysemöglichkeiten. In der

technischen Umsetzung verzichten diese Log-Management-Lösungen folglich auf aufwendiges Parsen und Normalisieren von Logs und setzen stattdessen auf eine teilweise oder vollständige Blind-Indexierung des Logtextes (Universal Indexing).

Einige dieser Lösungen sind in der Lage, wiederkehrende Terme und Muster zu lernen und beherrschen dabei eine Volltextindizierung in Echtzeit (z.B. ArcSight Logger, LogLogic, Splunk, etc.).

IT-Search-Lösungen sind auf hohe Durchsätze getrimmt und speichern Logdaten mit Kompressionsraten von bis zu Faktor 10:1 in sogenannte indexed Flatfiles.

Sie verwenden keine Datenbanken zum Ablegen der Informationen und arbeiten somit schemalos. Stattdessen werden die Logdaten i.d.R. in einem offenen Format (gzip) gespeichert und signiert, um Manipulationssicherheit zu gewährleisten.

Sie beherrschen keine komplexen Korrelationen in Echtzeit. Die Möglichkeiten der Loganalyse orientieren sich an einer forensischen Analyse erfasster Daten.

**Lager 2: SIEM**

SIEM-Lösungen kommen ursprünglich aus dem Security Incident Management und wurden für SOCs entwickelt. Die Verarbeitung von Security Events von Firewalls, Proxies, IDS/IPS, etc. beherrschen sie heute perfekt und warten mit intelligenter Ereigniskorrelation, Workflow-Unterstützung, Collaboration-Tools, Live Channels und

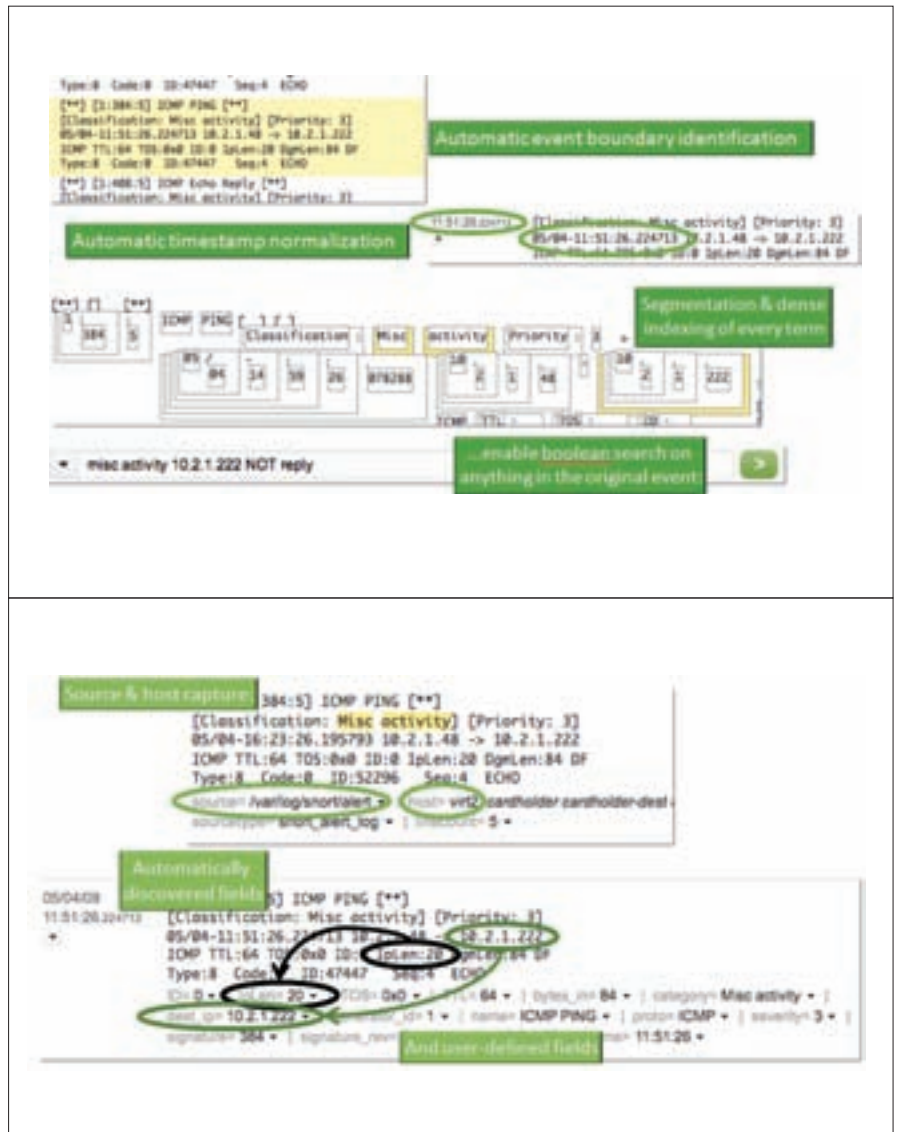


Abbildung 2. Universal Indexing am Beispiel splunk

WERBUNG

ITFS24 GmbH gehört zu den bundesweit zugelassenen Bildungsträgern, die entsprechend der Anerkennungs- und Zulassungsverordnung Weiterbildung (AZWW) zur Annahme von Bildungsgutscheinen berechtigt sind. Aktuell haben Arbeitsuchende sowie Festangestellte die Möglichkeit, auf diesem Weg die komplette Übernahme der Lehrgangsgebühr bei der Agentur für Arbeit zu beantragen.



**Wir bieten Kurse sowohl für Einsteiger als auch für Profis an. Vollzeit, Halbzeit sowie Online Weiterbildungen gehören zu unseren Angeboten.**

Entscheiden Sie sich für die Qualifizierung durch ITFS24 GmbH und profitieren Sie von den vielen Vorteilen u.a.:

- Flexibilität durch fortlaufenden Einstieg
- **100% Kostenübernahme durch die Agentur für Arbeit**
- Weltweit anerkannte Abschlüsse (CCNA, LPI, MCP, MCSA, MCSD, MCSE, MCITP,...)
- Lernerfolgkontrolle durch qualifizierten Dozenten

Sie erreichen uns telefonisch für eine kostenlose Beratung unter 06131-6696122 oder schreiben Sie uns unter [info@itfs24.com](mailto:info@itfs24.com)

# FÜR EINSTEIGER

Ticketing-Funktionen auf SIEM-Lösungen sind wesentlich schneller geworden und können heute Analyse und Korrelation von Sicherheitsereignissen in Echtzeit (also unmittelbar zum Zeitpunkt des Auftretens) mit 6.000 EPS / pro System und mehr automatisiert durchführen.

Nahezu alle führenden Hersteller von SIEM-Lösungen (z.B. ArcSight, RSA Envision, Trigeo, Q1Labs) arbeiten datenbankorientiert und benötigen eine möglichst breite Produktunterstützung durch vorhandene Konnektoren, meist in Form RegEx-basierter Parser.

Nur wenige SIEM-Lösungen bieten eine umfangreiche Normalisierung und Kategorisierung. Erfolgt dies nicht, sinkt der Informationsgehalt und damit die Aussagekraft gegenüber dem Raw Event beträchtlich. Kaum eine Lösung bietet ein wirklich leistungsfähiges SDK, mit dessen Hilfe man eigenständig Parser bzw. Konnektoren entwickeln kann um z.B. eigene Applikationen einzubinden.

Beide Ansätze verfolgen unterschiedliche Zielsetzungen und weisen jeweils spezifische Vor- und Nachteile auf. Durch Mergers & Acquisitions versuchen Unternehmen beider Lager seit ca. 12 - 18 Monaten Kompetenzen in beiden Bereichen aufzubauen und Gesamtlösungen zu entwickeln. Stellvertretend seien hier ArcSight (ESM und Logger) und LogLogic (Exaprotect) erwähnt.

## Bewertung und Einsatz

Bestimmendes Kriterium für die grundlegende Systemauswahl ist vor allem der Einsatzzweck.

Wenn es darum geht, Events von vor allem weitverbreiteten Quellen mit einem hohen Grad an Automatisierung durch komplexe Korrelationen zu verarbeiten, sind klassische, Datenbank-orientierte SIEM-Lösungen die erste Wahl. Müssen die Events einer hohen Anzahl von „legacy Applikationen“ mit großer Vielfalt der Log-Schemata, Formate und Inhaltsoptionen analysiert werden, so kann der Aufwand für die Entwicklung von Parsen schnell eine Dimension erreichen, die das Konzept grundsätzlich in Frage stellt. Darin liegt einer der wesentlichen Vorteile von Loganalyse-Systemen begründet, welche auf Universal Indexing setzen: Es sind keine

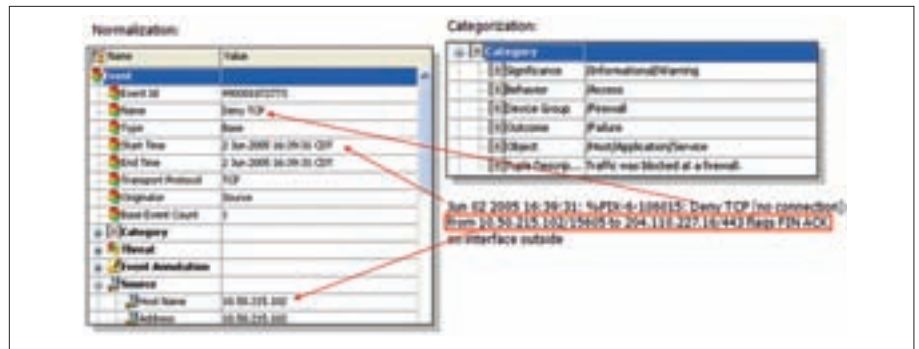


Abbildung 3. Normalisierung und Kategorisierung im SIEM-System

umfangreichen Vorarbeiten nötig, um Logs analysieren zu können.

Universal Indexing indiziert alle Arten von Logdaten, jeden Term, jedes Ereignis von allen Quellen in Echtzeit, ohne dabei Datenbanken zu verwenden, teure Konnektoren zu benötigen oder benutzerdefinierte Parser für proprietäre Anwendungen vorauszusetzen. Um es auf den Punkt zu bringen: „Universal Indexing frisst einfach alles.“ Den Analysten ermöglicht es - in Analogie zu Google - nach beliebigen Termen, Wörtern oder Wortgruppen zu suchen.

„Universal Indexing frisst einfach alles.“  
A. Mertz, IT-CUBE SYSTEMS

Die Geschwindigkeit der Suchabfragen hängt allgemein von der Anzahl der gleichzeitigen Searches, von deren Komplexität, vom zu durchsuchenden Logdatenvolumen und dessen Verteilung auf verschiedene Systeme sowie von der Hardware-Performance ab. Bei fast allen Universal-Indexing-Lösungen kann man mehr oder weniger

intelligent mit den Suchergebnissen interagieren. Die Technologieführerschaft nimmt in dieser Disziplin derzeit zweifelsohne Splunk ein. Mit einem Ajax-basiertem Web2.0-GUI sind Drill-Downs, Zoomfunktionen auf der Zeitachse, Trendanalyse und das Erkennen von Spikes oder Anomalien vorbildlich umgesetzt. Statistiken, Grafiken und andere praktische Werkzeuge, ermöglichen es in kürzester Zeit die **Nadel im Heuhaufen** zu finden.

Der Vorteil von IT-Search ist allerdings auch sein größter Nachteil: Der Analyst muss wissen, wonach er sucht. Das System korreliert nicht automatisch.

Dazu ein Beispiel: Ein Administrator legt einen neuen User in einer MS Windows Domäne an und weist diesem umfassende Zugriffs- und Administrationsrechte zu. Anschließend nutzt er diesen Account, um über einen Zeitraum von 5 Tagen vertrauliche Daten in kleinen Mengen zu kopieren. Danach löscht er den Account. Zwar stehen in den Security Event Logs Einträge über das Anlegen und Löschen dieses Accounts,

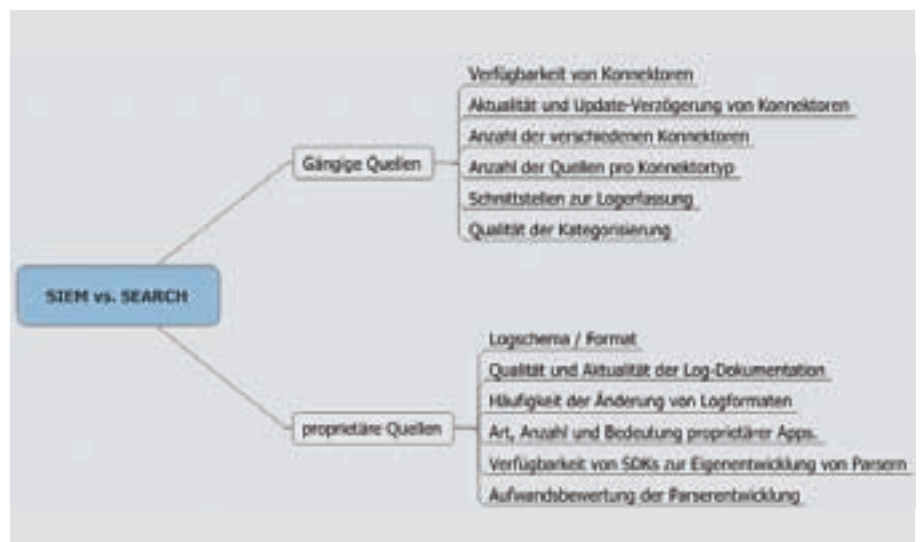


Abbildung 4. Quellenbezogene Bewertungskriterien SIEM vs. IT-SEARCH



# FÜR EINSTEIGER

10	The solution can add additional timestamps.
11	The solution persists the original unaltered log in a raw format.
12	The solution allows searches based on regular expressions.
13	The solution supports correlated searches based on boolean, nested, quoted string and wildcard searches.
14	The solution supports combined time and term searches.
15	The solution supports fields providing the common metadata to perform structured search, alerting, reporting and analysis.
16	The solution allows to rename fields or groups of events to be referred by name.
17	Specification of new fields does not cause a re-indexing of the data.
18	The solution supports scheduling and alerting of searches.
19	The solution provides altering via eMail, RSS, SNMP.
20	The solution allows to summarize search results to reports with interactive charts, graphs and tables.
21	The solution is able to store log data in a compressed way requiring 30 .. 50% of the original data size to retain data and index sets.
22	The solution supports tiered storage of log data and on demand restoring of archived data.
23	The solution supports an online retention period of 2 days for the total log volume (xxx GB / day).
24	The solution supports archiving of xx days for the total log volume (xx GB / day).
25	The solutions architecture supports distributed deployment.
26	The solutions architecture supports multi-tier deployment.
27	The solution supports the deployment of multiple index servers to be combined as a single logical data store for to distributed search.
28	The solution provides a central management for distributed deployments.
29	The solution can be deployed in high availability mode or as redundant cluster.
30	The solutions architecture supports search across multiple installations.
31	The solution supports routing of data based on characteristics and content.
32	The following data inputs are supported:
33	Proprietary (e.g. OPSEC LEA)
34	Syslog
35	SNMP
36	FTP, SCP, rsync, ftp, sftp
37	Capturing output from scripts or commands such iostat, ps, top, etc.
38	ODBC, JDBC
39	Log4J Appenders
40	Windows API and WMI
41	The solution provides an AJAX based intuitive browser-based GUI with no need for additional plugins.
42	The solution allows to create custom dashboards.
43	The solution provides encryption and authentication of communication channels and GUI / CLI access.
44	The solution provides a flexible, role-based system to build roles to be mapped policies for different classes of users which access log data or administrate the system.
45	The solution provides SSO interworking with LDAP, Active Directory and e-Directory.
46	The solution provides audit capabilities in a way that any administrative and user activities can be logged and audited figuring out who's accessing what data and when.
47	The solution provides data tampering measures in a way that single events are being signed and streams of events are being block signed. It must be in accordance to Federal Information Processing Standards (FIPS) and Common Access Card (CAC), for criminal investigation and prosecution.
48	The solution is hardened self-contained software server or hardware appliance.