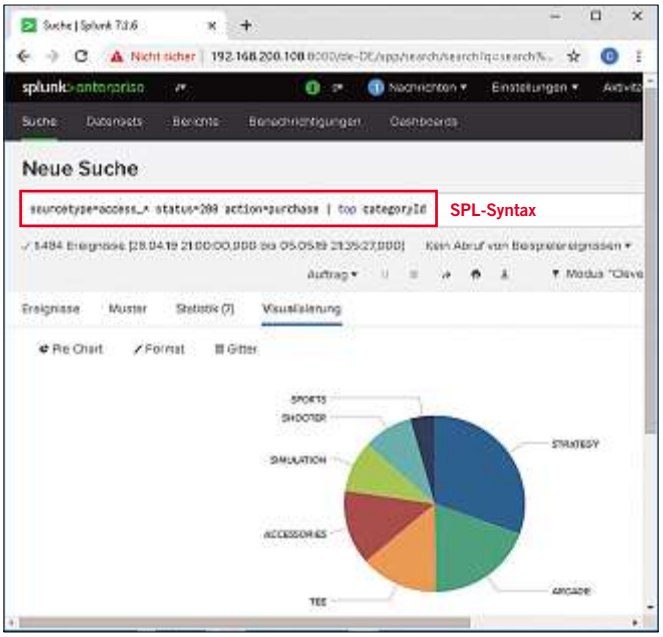
Big-Data-Analyse mit Splunk

SICHERHEIT IN IT-NETZWERKEN (TEIL 7) Am Ende des vorangegangenen Teils (»de« 15–16.2019, S. 69–70) haben wir erkannt, dass es spezieller Werkzeuge bedarf, um den riesigen Datenmengen Herr zu werden und Angriffe lokalisieren zu können. In dieser Folge stellen wir eine dafür geeignete Software detaillierter vor.



Quelle: alle Bilder C. Strobel

Bild 29: Mit Hilfe der splunk-spezifischen Such-Sprache »Search Processing Language« (SPL) lassen sich mit einfachen Befehlen Datenmuster visuell aufbereiten

Splunk Inc. ist ein US-amerikanisches Unternehmen mit Sitz in San Francisco, das seit 2012 börsennotiert ist. Mehr als 6.000 Firmen, Universitäten, behördliche Einrichtungen und Service Provider in mehr als 90 Ländern nutzen Splunk Enterprise. Es bietet verschiedene Analyse-Software für Unternehmen, die ihre Betriebs- und Geschäftsdaten auswerten möchten und generiert Umsätze im Milliarden-US-Dollar-Bereich.

Dabei hat sich das Unternehmen auf Big-Data-maschinengenerierte Daten spezialisiert, die von Webseiten, Anwendungen, Servern, Netzwerken und mobilen Endgeräten generiert werden. Eine eigene Technik wurde dafür entwickelt, um sowohl Echtzeit-Auswertungen als auch historische Maschinendaten zu überwachen, durchsuchen, analysieren und zu visualisieren.

Umgang mit der Software

Die Software ist flexibel konfigurierbar und erlaubt das Indizieren, Verarbeiten und Auswerten verschiedenartiger Datenquellen unabhängig vom Format und Speicherort wie z.B. aus herkömmlichen Datenbanken, Data Warehouse-Systemen, NoSQL-Systemen, Click-Stream-Daten, Kundentransaktionen, Netzwerkaktivitäten oder Gesprächsdatensätzen. Es durchsucht Logs, Metriken und weitere Daten von Anwendungen, Diensten oder Netzwerkgeräten. Splunk-Nutzer können so ihr Geschäfts- und Kundenverständnis vertiefen, ihren Service und ihre Betriebszeit verbessern, Kosten reduzieren und Sicherheitsrisiken minimieren.

Diese werden in Ergebnisdatenmengen in der Form zusammengestellt, dass grafische Auswertungen, Reports oder Warnungen leicht erstellt werden können. Störungen, Muster oder Auffälligkeiten sollen vom IT-Administrator leicht erkannt werden. Diagramme sollen helfen, Analyseergebnisse im Team oder vor der Geschäftsführung darzustellen. Daten verschiedener Systeme lassen sich allerdings auch zusammenfassen.

Das **Bild 29** zeigt eine grafische Auswertung in Splunk, die im Rahmen eines Splunk-Tutorials (https://www.tutorialspoint.com/splunk/index.htm) entstanden ist. Mit Hilfe der splunk-spezifischen Such-Sprache »Search Processing Language« (SPL) lassen sich mit einfachen Befehlen Datenmuster visuell aufbereiten. Die Pivot-Schnittstelle ermöglicht es Anwendern, Maschinendaten zu lesen, um umfassende Berichte zu erstellen, ohne die Suchsprache lernen zu müssen. Auch Anwender der Geschäftsebene können leicht relevante Datenerhebungen erstellen. Ereignismuster werden entlang einer Zeitachse dargestellt, um Trends, Spitzen und Abweichungen auf einen Blick festzustellen. Es gibt über 140 Befehle, die erlauben, nach Schlüsselwörtern zu suchen, beliebige Datenmengen zu filtern oder die Suche in Teilsuchen zu unterteilen. Auch die weitverbreitete »Regular-Expression-Notation« wird unterstützt. In unserem Beispiel wurde die Ergebnismenge grafisch durch ein Kreisdiagramm aufbereitet.

Die Firma Splunk möchte seine Big-Data-Werkzeuge für alle Benutzer zugänglich und nutzbar machen. So lässt sich »Splunk Enterprise« beispielsweise für die Betriebssystemplattformen Windows/ Windows Server, Linux (.deb, .rpm, .tgz) oder MacOS (.dmg, .tgz) installieren. Es gibt die Versionen »Splunk Free«, »Light«, »Enterprise« und »Cloud«. Für Einsteiger und Profis, für Anwender und Entwickler, für alleinstehende und internetbasierte IT-Umgebungen oder für dedizierte

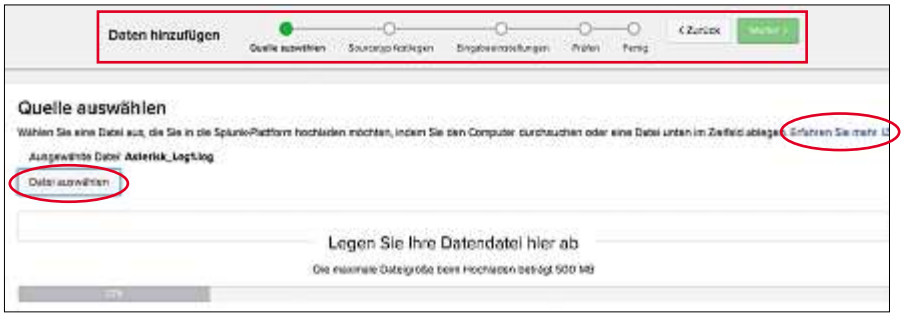


Bild 30: Schritt 1 in der Analyse mit »Splunk« – das Hinzufügen von Daten

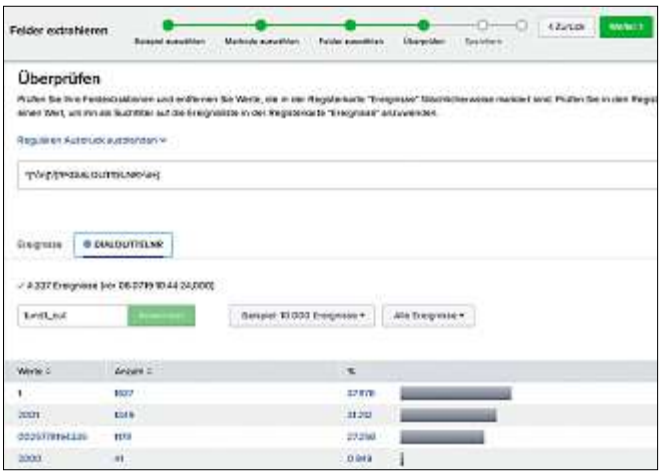


Bild 31: Eine externe Telefonnummer lässt sich mit dem Feld »DIALOUTTELNR« festlegen

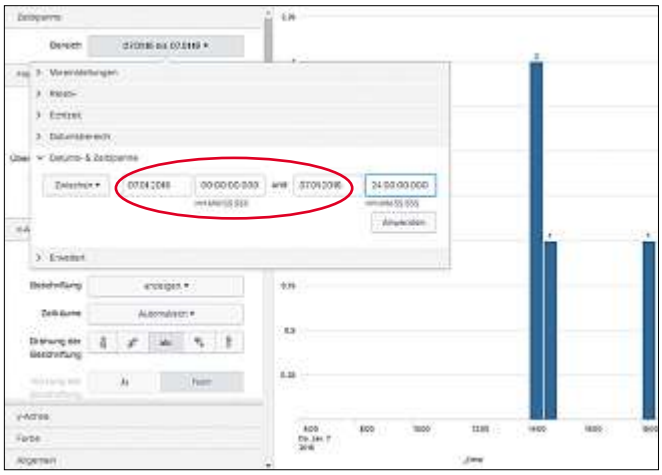


Bild 33: Die Auflösung lässt sich durch Eingabe einer Zeitspanne vergrößern – die Anrufe nach Israel werden dann als Säulen dargestellt

und cloudbasierte IT-Umgebungen werden geeignete Lösungen zur Verfügung gestellt.

Vorzüge der Software

Zusammenfassend lassen sich folgende Vorteile auflisten:

- Die Bedienoberfläche lässt sich intuitiv bedienen – Befehle lassen sich durch Maus-Klicks oder Menüs erzeugen – in den Menüs sind Direktaufrufe zum Tutorial oder Hilfeseiten integriert.
- Bei der Datensuche können vertraute bzw. verbreitete Befehle verwendet werden – so steht »*« für beliebige Zeichen, »|« für die Aneinanderreihung von Befehlen, aber auch die beliebten regulären Ausdrücke können genutzt werden.
- Suchergebnisse können in Echtzeit interaktiv ermittelt werden.
- Ereignismuster dienen zur automatischen Erkennung von Vorfällen.
- Die Personalisierung der Oberfläche ist möglich. Jeder Benutzer verfügt über seine eigene Ansicht.
- Gemäß Splunk Inc. sind Datenauswertungen für alle und alles möglich.
- Flexible und leistungsstarke Analyse von großen Datenmengen.
- Funktionen lassen sich stark anpassen und unabhängig voneinander individualisieren.

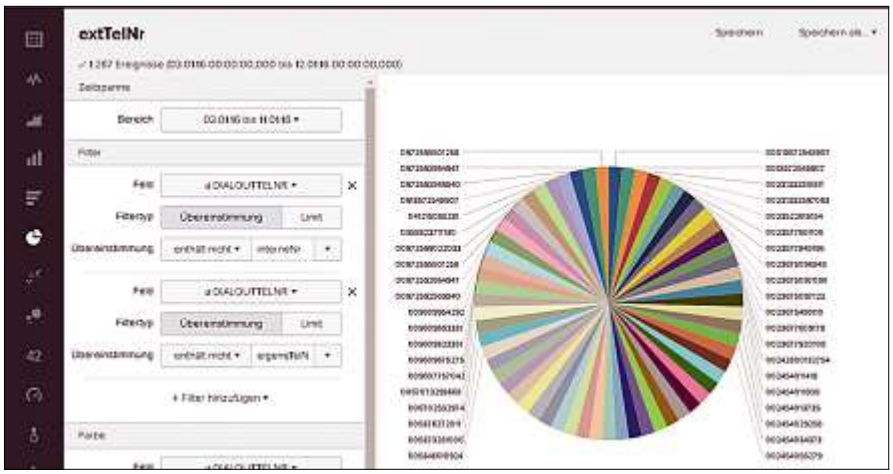


Bild 32: Grafische Aufbereitung der Ergebnisse (gewählte Nummern in das öffentliche Netz) als Kreisdiagramm

Analyse des Beispiels

Doch kommen wir zurück zum Eingangsfall. Eine Brute-Force-Angriff wurde auf einen Asterisk-SIP-Server ausgeübt und hat die Telefonkosten massiv in die Höhe getrieben. Was ist genau geschehen und in welchem Zeitraum? Hätte der Angriff nicht früher entdeckt werden können?

Schritt 1 – Daten hinzufügen

Es gibt verschiedene Wege zur Bereitstellung von Daten. Es lassen sich Dateien und Ordner überwachen, HTTP-Ereignisse sammeln, TCP-/UDP-Ports monitoren oder benutzerdefinierte Skripte ausführen. Sogenannte »forwarder« leiten verteilt Daten an Splunk weiter. Ein Assistent unterstützt Sie dabei (**Bild 30**). Der jeweilige Arbeitsschritt wird in der Übersicht dargestellt. Über den Button »Datei auswählen« öffnet sich ein Auswahlfeld, um eine Datei hochzuladen. Der Vorgang ist alt bewährt. Bei Klick auf »Erfahren Sie mehr« wird eine zusätzliche Hilfeseite online abgerufen.

Schritt 2 – Log-Daten prüfen

Über die Suchfunktion lassen sich nun alle Log-Daten übersichtlich darstellen. Die externe Telefonnummer mit der Länderkennung 00972 (Israel) wurde angerufen. In unserem Beispiel sind über 4 Millionen Einträge in der Log-Datei abgelegt. Bequem lassen Sie sich durchstöbern. Die Eingabe der Zeichenkette »@1und1_out« in das Suchfeld bringt alle betreffenden Datensätze.

Schritt 3 – Felder extrahieren

Zu den einzelnen Datensätzen lassen sich anschließend variable Bereiche festlegen. Diese Bereiche werden auch bei anderen Datensätze gesucht. Das geschieht intuitiv durch das Markieren eines Textteiles mit der Maus oder durch Festlegen von regulären Ausdrücken (**Bild 31**). Beispielsweise lässt sich die externe Telefonnummer als ein Feld mit dem Namen »DIALOUTTELNR« festlegen. In der Abbildung wird das Ergebnis angezeigt. Mithilfe von Feldern lassen sich Suchmuster erstellen. Als

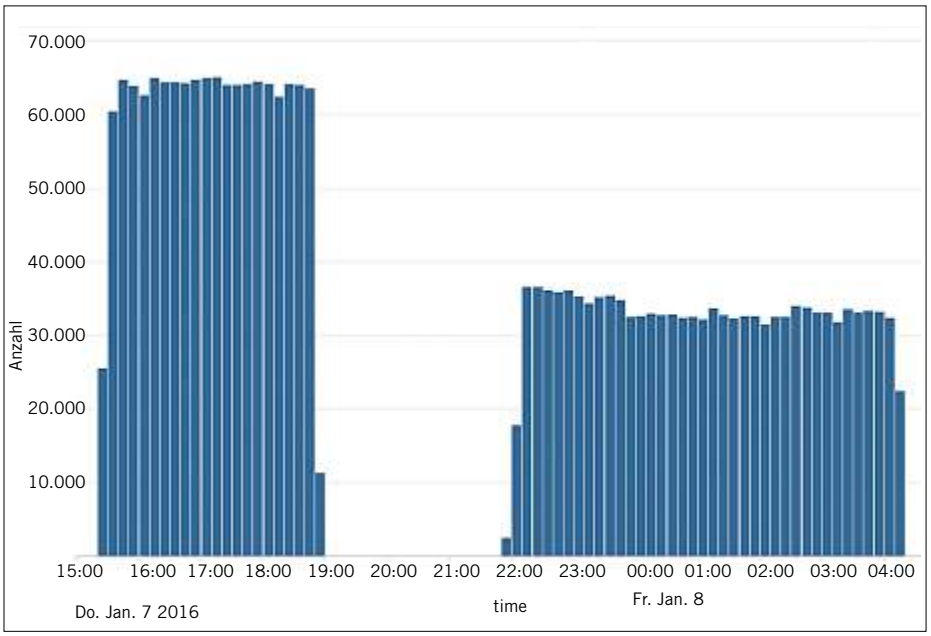


Bild 34: Binnen gut drei Stunden gab es in ca. zehnminütigen Abständen jeweils etwa 60 000 Anfragen, was der Asterisk-Server mit »wrong password« quittierte

Ergebnis erhält man alle gewählten Telefonnummern mit dem dazugehörigen Häufigkeitswert dargestellt.

Schritt 4 – Datensatz filtern

Interne und eigene Telefonnummern können nun ebenfalls aus dem Ergebnis gefiltert werden. Dies geschieht wiederum intuitiv mit einfachen Mausklicks, Menüaufrufen oder der Definition von regulären Ausdrücken. Die Datensätze mit den externen Telefonnummern bleiben übrig.

Schritt 5 – Übersicht erstellen

Splunk besitzt umfangreiche Darstellungsoptionen, ähnlich wie bei MS-Excel. Balken-, Linien-, Kreis-, Punktdiagramm u.a. sind möglich. Die Ergebnisse lassen sich speichern, aber auch in einem PDF-Format konvertieren. Als Ergebnis sind nun alle externen Nummern in einem Kreisdiagramm dargestellt (**Bild 32**), die vom Asterisk-SIP-Server aus dem lokalen Netz über den ISP-SIP-Server ins öffentliche Telefonnetz angewählt wurden.

Eine zweite Analyse soll den genauen Zeitraum der Telefonnummer »00972/ 592664947« darstellen. Dafür sind nur wenige Änderungen notwendig. Anstelle des Kreisdiagramms wird nun ein Balkendiagramm ausgewählt. Die X-Achse wird auf einen Zeitraum (**Bild 33**) eingestellt. Falls einzelne Werte nicht dargestellt werden können, zeigt Splunk automatisch die Anzahl der zutreffenden Datensätze an. Als Ergebnis geht hervor, dass am 7. Januar 2016 zwischen 14 und 15 Uhr zwei Anrufe nach 00972/592664947 (Israel) registriert wurden.

Das Filtern aller Log-Einträge auf den Terminus »wrong password« soll Aufschluss darüber geben, wie viele Brute-Force-Versuche es gegeben hat. Das **Bild 34** zeigt das Ergebnis: Das Säulendiagramm zeigt am 7. Januar die stärkste Aktivität. Zwischen 15:20 Uhr und 18:40 Uhr gibt es in 10-minutigen Zeitabständen je ca. 60 000 Anfragen. Umgerechnet sind dies ca. 100 Anfragen pro Sekunde, die etwas gemeinsam haben. Der Asterisk-Server lieferte »wrong password« bzw. »ungültiges Kennwort« an den Anfragenden zurück.

Vielleicht erinnern Sie sich noch an den Eingangsfall – ein Internet-Service-Provider kündigt Ihnen an, die nächste Monatsabrechnung mit zusätzlichen Mehrkosten zu belegen, weil von Ihrem Telefonanschluss Gespräche ins Ausland registriert wurden. Die gewonnenen Ergebnisse Ihrer Analyse helfen, dieser Situation souveräner zu begegnen.

Sie wissen nun, welche Auslandsnummern wann gewählt wurden und können den Angriff zeitlich nachbilden. Weitere Details können leicht gewonnen werden. Aufgrund der hohen Anzahl von fehlerhaften Anfragen pro Sekunde, lässt sich vermuten, dass der Angreifer eine Software bzw. Bot gewesen sein muss. Ein einzelner Mensch wäre nicht dazu imstande.

Fazit

Durch die Digitalisierung im Handwerk gehören Daten zu den wichtigen Ressourcen in einem Betrieb. Aus ihnen lassen sich bedeutende Erkenntnisse ermitteln, um neue Geschäftsmodelle zu entwickeln. Produkte lassen sich dadurch individualisieren, Strategien verbessern.

Die Herausforderung für einen Firmeninhaber ist es, Benutzungsfälle (Leitfragen) für die Big-Data-Analyse anwendungsgerecht und qualitativ festzulegen. Der Markt in diesem Umfeld wächst stetig. Vor der Einführung einer Big-Data-Technologie sollte sich ein Elektrobetrieb überlegen, welche Art und welcher Umfang notwendig sind.

Der folgende Fragenkatalog soll Ihnen helfen, Art und Umfang zu spezifizieren:

- Welche Arten von Daten liegen in dem Unternehmen vor? Sollen die Daten flexibel ausgewertet werden oder reichen Direkt-Analysen aus?
- In welchen Abteilungen bzw. Unternehmensbereichen fallen die Daten an, wer bearbeitet und wer benötigt sie?
- Stammen die Daten aus herkömmlichen Datenbanken oder sind die Quellen unstrukturiert?
- Reicht es für die Anwendung aus, dass die Daten schnell gespeichert werden und später in größeren Mengen analysiert werden oder soll die Analyse und Auswertung in Echtzeit erfolgen?
- Werden Daten aus sozialen Netzwerken für fortlaufende Analysen hinzugezogen?

Ab der kommenden Folge wollen wir Ihnen vorstellen, wie Sie die Analyse des eingehenden Netzwerkverkehrs mit Ihrer Fritz-Box durchführen können. Dabei kommen versteckte Funktionen zum Einsatz, die Sie kennen sollten.

(Fortsetzung folgt)

AUTOR

Claus Strobel
Dozent IT/ET; Schwerpunkt Netzwerktechnik,
Elektro-Technologie-Zentrum, Stuttgart