Data Obfuscation and Field Protection in Splunk

Angelo Brancato | Security Specialist
Dirk Nitschke | Senior Sales Engineer
28 September 2017 | Washington, DC
Agenda

Protect Your Machine Data

- The Drivers
- The Solution
- The Demo
Who we are

Dirk Nitschke
Senior Sales Engineer

Angelo Brancato
Security Specialist
Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.
Why?
Minimize Risk, Enable Business
Confidential data in events can not be shared / collected
- Cross border organization
- SIEM as a Service
- Cloud

Regulation (like GDPR, PCI, HIPAA) have strict imposts on protecting personal data
How?

Transport Presentation & Data Layer Protection

On an Event Field Level
Transport, Data, and Presentation Layer Protection

Option 1
Event Data
Presentation Layer Protection

Option 2
Event Data
Data Layer Protection (In-Motion & At-Rest)
Option 1: Presentation Layer Protection

Data as is

ccn= 5105-1051-0510-5100

Anonymization (e.g. SHA256 hash)

ccn=fb415937c6f3065774810b300720cb3f2e82340a09b42b074fd13a09bc341fd939029012a

Pseudonymization (e.g. AES256 encryption)

ccn=U2FsdGVkX1+pn/g/S3aXZKlq+dMegBKii0P4H6Ge86ZjUPeYjlvAYEBfnL3XM6lyz

Format Preserving Pseudonymization (e.g. Format-Preserving-Encryption / Tokenization)

ccn=5105-0864-7332-5372

> Result Masking
Option 2: Data Layer Protection

Data Layer Protection (In-Motion & At-Rest)

```
ccn=5105-1051-0510-5100
```

- Anonymization (e.g. SHA256 hash)
  - ccn=fb415937c6f3065774810b300720cb3f2e82340a09b42b074fd13a09bc341fd939029012a

- Pseudonymization (e.g. AES256 encryption)
  - ccn=U2FsdGVkX1+pn/g/S3aXZKlq+dMegBKIC0P4H6Ge86ZjUPSeyjvAYEBfnL3XM6iyz

- Format Preserving Pseudonymization (e.g. Format-Preserving-Encryption / Tokenization)
  - ccn=5105-0864-7332-5372
Option 2: Data Layer Protection

**Forwarder**

**Indexer**

**Search Head**

**Splunk User**

Optional: + Original Event

“Always-On” Transport Layer Protection (TLS)

Data Layer Protection (In-Motion & At-Rest)

Data as is

- Anonymization (e.g. SHA256 hash)
- Pseudonymization (e.g. AES256 encryption)
- Format Preserving Pseudonymization (e.g. Format-Preserving-Encryption / Tokenization)

<table>
<thead>
<tr>
<th>ccn</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ccn=5105-1051-0510-5100</td>
<td>fb415937c6f3065774810b300720cb3f2e82340a09b42b074fd13a09bc341fd939029012a</td>
</tr>
<tr>
<td>ccn=5105-0864-7332-5372</td>
<td>U2FsdGVkX1+pn/g/S3aXZIKlq+dMegBKi0P4H6Ge86ZjUPeYjlvAYEBfnL3XM6tyz</td>
</tr>
</tbody>
</table>

Modular / Batch Processing

- cn=5105-1051-0510-5100
- cn=5105-0864-7332-5372
Option 2: Data Layer Protection

Data as is

Anonymization (e.g. SHA256 hash)

Pseudonymization (e.g. AES256 encryption)

Format Preserving Pseudonymization (e.g. Format-Preserving-Encryption / Tokenization)

ccn=fb415937c6f3065774810b300720cb3f2e82340a09b42b074fd13a09bc341fd939029012a

ccn=U2FsdGVkX1+pn/g/S3aXZKIlq+dMegBKil0P4H6Ge86ZjUPeYJlvAYEBfnL3XM6tyz

ccn=5105-0864-7332-5372

ccn=5105-1051-0510-5100

Splunk User

Search Head

Indexer

Forwarder

EP (External Processor)

External Processor

"Always-On" Transport Layer Protection (TLS)

“On-site | cloud”

Optional: + Original Event

Private Cloud

Public Cloud

On-Premises

Containers

Web Clickstreams

RFID

Smartphones and Devices

Servers

Messaging

GPS

Location

Packaged Applications

Custom Applications

Online Services

Databases

Call Detail Records

Energy Meters

Firewall

Intrusion Prevention

Software and Services
What is Format Preserving Encryption?

An Analogy...
Various Options to Protect Event Fields
Protect Field Values
Protect data at earliest stage in the process

Data source owner is responsible

Application support

May need a means to decode data again
Input Layer – Modular / Batch Processing

- Pre-process data
- Create your own data input capabilities using a modular input
  - Very flexible
  - Requires scripting, programming
- Forward raw data plus meta data to external processing engine
- Output of external processor is an input for Splunk again
- Very flexible but also complex
Built-In at Indexing Time – Regex Replace

- Typing pipeline / regex replacement processor
- Uses SEDCMD or TRANSFORMS to modify data at indexing time

- Easy to implement
- Limited flexibility, mainly anonymization
Scheduled search selects and transforms the data
- SPL and custom search commands if needed

Send modified events to a different index
- Think about collect, cefout, or a custom search command

Modified events are delayed because of scheduling

Preservation of event metadata
Hide data at presentation layer

- `<your_search> | eval user=sha256(user)` or your own custom search command

Optionally: user lockdown

- Pre-defined app with dashboard access only
- No search app, no raw search, no raw event drill down

May be good enough
## Overview – All Options

No scientific research

<table>
<thead>
<tr>
<th>Option</th>
<th>Layer</th>
<th>Latency</th>
<th>Security</th>
<th>Complexity</th>
<th>Usability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Data Source</td>
<td>Low</td>
<td>Very High</td>
<td>Medium (decoding)</td>
<td>Medium (decoding)</td>
</tr>
<tr>
<td>Modular / Batch</td>
<td>Data</td>
<td>Medium (API calls)</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>External Processor</td>
<td>Data</td>
<td>Medium (API calls)</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Regex Replace</td>
<td>Data</td>
<td>Medium</td>
<td>High (anonymization)</td>
<td>Low</td>
<td>Low (anonymization)</td>
</tr>
<tr>
<td>Scheduled Search</td>
<td>Data</td>
<td>Very High (schedule)</td>
<td>Low</td>
<td>Medium</td>
<td>Medium (due to latency)</td>
</tr>
<tr>
<td>Result Masking</td>
<td>Presentation</td>
<td>Medium</td>
<td>Low (need-to-know)</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>
1. Many possible ways – each has pros and cons
2. Qualify – what are the requirements and boundary conditions
3. Data obfuscation requires a proper concept and careful planning
4. Choose and mix – and keep it simple
How?

Demo

Just a Demo – Your Mileage Will Vary
Demo Setup

bar.log ➔ Pipeline ➔ “secret” ➔ “DPO” role ➔ “Default” role

firstname, lastname, email, cc, src_ip ➔ Scheduled Search, Custom Search Command
Demo Setup

foo.log → Pipeline → “secret” → “DPO” role

firstname, lastname, email, cc, src_ip → Clone sourcetype, transforms, output as raw → “unclassified” → “Default” role

Modify Events, send to HEC → SEDCMD
### Original Events

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-09-21 14:32:05.049960</td>
<td>component=calceng&amp;vendor_action=starting&amp;action=success</td>
</tr>
<tr>
<td>2017-09-21 14:32:03.082647</td>
<td>component=calceng&amp;vendor_action=starting&amp;action=success</td>
</tr>
<tr>
<td>2017-09-21 14:32:03.014554</td>
<td>component=calceng&amp;vendor_action=restoring&amp;action=success</td>
</tr>
<tr>
<td>2017-09-21 14:32:02.014399</td>
<td>firstname=Gerald&amp;lastname=Knight&amp;Email=<a href="mailto:knights54@jalbum.net">knights54@jalbum.net</a>&amp;cc=353255 4004575080&amp;src_ip=194.8.168.11&amp;vendor_action=purchase&amp;action=success</td>
</tr>
<tr>
<td>2017-09-21 14:31:59.881900</td>
<td>firstname=Carol&amp;lastname=Robinson&amp;Email=<a href="mailto:crobinson@bibliogateway.com">crobinson@bibliogateway.com</a>&amp;src_ip=220.216.132.30&amp;action=failure</td>
</tr>
<tr>
<td>2017-09-21 14:31:59.261594</td>
<td>firstname=Ananda&amp;lastname=Wright&amp;src_ip=95.87.185.135&amp;vendor_action=logon&amp;action=success</td>
</tr>
<tr>
<td>2017-09-21 14:31:58.610016</td>
<td>firstname=Gary&amp;lastname=Arndl&amp;src_ip=162.59.117.207&amp;vendor_action=login&amp;action=failure</td>
</tr>
<tr>
<td>2017-09-21 14:31:57.184515</td>
<td>firstname=Albert&amp;lastname=Mason&amp;Email=<a href="mailto:mason@fema.gov">mason@fema.gov</a>&amp;cc=3058019519 79478&amp;src_ip=41.51.107.49&amp;vendor_action=purchase&amp;action=success</td>
</tr>
</tbody>
</table>
Processed Events
Masked critical data
Workflow Action

Decode field values from the UI
Decoding Dashboard
Custom Search Command

Encoding and decoding
# Decryption Command Usage

Track usage of critical commands

## Decryption Command Usage

<table>
<thead>
<tr>
<th>_time</th>
<th>sid</th>
<th>username</th>
<th>command</th>
<th>comment</th>
<th>fieldnames</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-09-21 14:34:06.131</td>
<td>15059997245.34</td>
<td>dpo</td>
<td>aesencrypt.py</td>
<td>Bulk decoding</td>
<td>[&quot;firstname&quot;, &quot;lastname&quot;, &quot;email&quot;, &quot;uc&quot;]</td>
</tr>
<tr>
<td>2017-09-21 14:33:30.889</td>
<td>15059997219.34</td>
<td>dpo</td>
<td>aesencrypt.py</td>
<td>Bulk decoding</td>
<td>[&quot;firstname&quot;, &quot;lastname&quot;, &quot;email&quot;, &quot;uc&quot;]</td>
</tr>
<tr>
<td>2017-09-21 14:32:48.781</td>
<td>dpo, dpo, U0hvY29uZjwMTc_...</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Workflow Action: email</td>
<td>[&quot;email&quot;]</td>
</tr>
<tr>
<td>2017-09-21 14:31:33.002</td>
<td>dpo, dpo, U0hvY29uZjwMTc_...</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Workflow Action: email</td>
<td>[&quot;email&quot;]</td>
</tr>
<tr>
<td>2017-09-21 14:24:32.861</td>
<td>dpo, dpo, U0hvY29uZjwMTc_...</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Workflow Action: email</td>
<td>[&quot;email&quot;]</td>
</tr>
<tr>
<td>2017-09-21 13:17:39.117</td>
<td>dpo, dpo, U0hvY29uZjwMTc_...</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Workflow Action: email</td>
<td>[&quot;email&quot;]</td>
</tr>
<tr>
<td>2017-09-21 13:11:59.111</td>
<td>dpo, dpo, U0hvY29uZjwMTc_...</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Workflow Action: email</td>
<td>[&quot;email&quot;]</td>
</tr>
<tr>
<td>2017-09-21 13:11:43.895</td>
<td>dpo, dpo, U0hvY29uZjwMTc_...</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Workflow Action: email</td>
<td>[&quot;email&quot;]</td>
</tr>
<tr>
<td>2017-09-21 13:10:39.261</td>
<td>dpo, dpo, U0hvY29uZjwMTc_...</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Workflow Action: email</td>
<td>[&quot;email&quot;]</td>
</tr>
<tr>
<td>2017-09-21 13:09:44.578</td>
<td>15059992184.45</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Bulk decoding</td>
<td>[&quot;email&quot;]</td>
</tr>
<tr>
<td>2017-09-21 13:09:00.904</td>
<td>dpo, dpo, U0hvY29uZjwMTc_...</td>
<td>dpo</td>
<td>fpedecrypt.py</td>
<td>Field Decoding Dashboard</td>
<td>[&quot;email&quot;]</td>
</tr>
</tbody>
</table>
Common Dashboard

Privileged view
Common Dashboard

Normal view
Enterprise Security
Normal view
Enterprise Security

Normal view
Enterprise Security
Normal view
## Enterprise Security

### Normal view

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
<th>Action</th>
<th>Correlation Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Destination</td>
<td>SNXV:818</td>
<td></td>
<td>Endpoint - High Or Critical Priority Host With Malware - Rule</td>
</tr>
<tr>
<td>Destination Business Unit</td>
<td>emea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination Category</td>
<td>pci</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination City</td>
<td>Havant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination Country</td>
<td>UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination Latitude</td>
<td>50.8436</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination Longitude</td>
<td>-0.98451</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination NT Hostname</td>
<td>SNXV:818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination PCI Domain</td>
<td>wireless</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination Requires Antivirus</td>
<td>true (requires_av)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination Should Time Synchronize</td>
<td>true (should_time)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destination Should Update</td>
<td>true (should_update)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signature</td>
<td>EICAR-AV-Test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Additional Fields

- A high or critical priority host (SNJV:818) was detected with malware.
- Currently not investigated.

### Related Investigations:
- Ping SNXV:818
- Session Center
- Stream Capture
- Traffic Search (as destination)
- Traffic Search (as source)
- Final Event:
  - Update Search
  - Vulnerability Search
  - Web Search (as destination)
  - Web Search (as source)

### Adaptive Responses:
- Notable: saved 2017-09-22T07:30:26+0000 admin ✓ success
- Risk Analysis: saved 2017-09-22T07:30:26+0000 admin ✓ success
Enterprise Security
Normal view
Thank You

Don't forget to rate this session in the .conf2017 mobile app