Preserving Confidentiality When Hunting With Friends

PhD. Paolo Di Prodi (Priam Cyber AI ltd, United Kingdom)
Gabriel Bassett (Liberty Mutual, United States)
Hugo Ideler (Roseman Labs, Netherlands)
Mini agenda today

- A real example: the Trigona Campaign
- Security Operations and Data
- Incident Response frameworks and standards for sharing
- Example of cooperative intra company reporting
- Sharing more with PET frameworks
- The 3 main approaches
- Why MPC+DP are the winners?
- Real deployment in the Netherlands
- Conclusion
- Q&A
Who is Paolo aka "The DOC"

- PhD in multi agent ML
- Founder of Priam AI in UK
- Senior Data Scientist for Fortinet
- Data Scientist for Microsoft
- Contributes to several open source initiatives such as STIX 2.1 and EPSS
Who is Gabriel Bassett

- Director of Cyber Risk Advisory Services, Liberty Mutual
- Founder, Information Security Analytics LLC
- Former Lead Data Scientist, Verizon DBIR
- BoD & Game Architect, CTF Factory, INC
- Director, BSides Las Vegas Ground Truth Track
Who is Hugo Ideler

- Head of Engineering at **Roseman Labs**, a start-up specializing in Multi-Party Computation
- Lead Engineer in NCSC-NL's SecureNed Platform
- Former Senior Manager at Deloitte’s Incident Response practice
- 10 years of experience in DFIR
The Trigona campaign

CVE-2021-40539
Published: 09/07/2021
CVE Base Score: 9.8 CRITICAL

Not a Zero Day!

Our ground truth: PAN UNIT 42 and ARETE report. Campaigns: Dec 2022, Jan 2023 and Feb 2023
Total Victims: 15

The malicious operator would take its time to explore: average 4 months…

Trigona discovered in October 2022
The Trigona campaign: detections & mitigations?

IPS VirtualPatch

Threat actor gains access to the targeted environment by leveraging CVE-2021-40539 vulnerability (possible entry vector).

AppLocker ZeroTrust

ScreenConnect connection installation

EDR detection

PSEXESVC Execution

Installation of LogMein and DWAgent

Powershell Execution using Cobalt Strike Beacon to Connect with C2 Server

AV/EPP

Game over! Extorsion campaign.

Windows PS Logging & Defender

Malicious Powershell execution

Multiple malicious Powershell scripts

Anydesk Execution

FileZilla Execution (Possible Data Exfiltration)

1.exe.exe Ransomware Payload Execution

PS Policies EDR: bypass PS

SDWan Segmentation

DLP

For eye candy, add your favourite ATT&CK TTP ….
Classical Sharing Scenarios

- Push/Pull Hub/Spoke
- Mostly Push/Unidirectional

SOC 1  SOC 2  SOC 3  SOC 4

STIX 2.1  Markings TLPs

MISP/OpenCTI

MSP 1  MSP 2

SOC 1  SOC 2  SOC 3  SOC 4

VERIS Dataset  Anonymized

VERIS schema

Pub/Sub
Ground Truth and Simulation

- A Stix 2.1 package with ...
- A pool of 10 companies: 4 impacted

<table>
<thead>
<tr>
<th>Entity</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report</td>
<td>2</td>
</tr>
<tr>
<td>Intrusion Set</td>
<td>1</td>
</tr>
<tr>
<td>Attack Pattern</td>
<td>32</td>
</tr>
<tr>
<td>Campaign</td>
<td>1</td>
</tr>
<tr>
<td>Identity</td>
<td>1</td>
</tr>
<tr>
<td>Indicator</td>
<td>45</td>
</tr>
<tr>
<td>Relationships</td>
<td>99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Company Identity</th>
<th>Bundle Size</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company A</td>
<td>49</td>
<td>Got the attack vector</td>
</tr>
<tr>
<td>Company B</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Company C</td>
<td>47</td>
<td>Got the attack vector</td>
</tr>
<tr>
<td>Company D</td>
<td>47</td>
<td></td>
</tr>
</tbody>
</table>

Indicators: 30
Patterns: 7
Company 1: investigation point

One of the companies finds a suspicious behaviour from one of their security products….

Fast query: how common is this technique given the context?

Query MITRE ATT&CK Sightings and ATT&CK Top Techniques?

UEBA: anomaly?
EDR: lateral movement?
Company 2: contextual info

The activity originates from an active directory host with a windows server from their main headquarters.

More context during the investigation…
Full knowledge graph: simplified

Public facing unpatched CVE
The companies will only be able to share part of the information based on TLP levels.
How do we make red more transparent?

Assuming you have a perfect incident sharing platform with real time sharing and querying, standardized & extended formats like STIX 2.1, VERIS, ATTACK FLOW, ATT&CK, CACAO, OpenC2…. Tools like OpenCTI, MISP for exchanging.

How can I build this shared graph rapidly without worrying?
Privacy-enhancing Technologies (PETs) for cyber sharing

Most traditional techniques offer weak mathematical guarantees of privacy.

We need something more powerful and with stronger mathematical guarantees, known as Privacy-enhancing Technologies (PETs).
## Privacy-enhancing Techniques (PETs)

<table>
<thead>
<tr>
<th><strong>Fully-homomorphic encryption (FHE)</strong></th>
<th><strong>Multi-party Computation (MPC)</strong></th>
<th><strong>Differential Privacy (DP)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• High computation cost</td>
<td>• Low computational cost</td>
<td>• Very fast to compute</td>
</tr>
<tr>
<td>• Low communication cost</td>
<td>• High communication cost</td>
<td>• Support most queries</td>
</tr>
</tbody>
</table>

- High computation cost
- Low communication cost
- Low computational cost
- High communication cost
- Very fast to compute
- Support most queries
Differential Privacy Example

Challenge
You want to create a survey for your team to measure how many bagels they eat every day.

Some people in your team are afraid to participate because they are on a "diet" and they don't want to risk to be identified if future information is released.
Differential privacy randomization

- Each participant spins a dial and adds noise to their true answer.

<table>
<thead>
<tr>
<th>ID</th>
<th>True Answer</th>
<th>Randomized</th>
<th>Coin Toss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paolo</td>
<td>2</td>
<td>2</td>
<td>Head</td>
</tr>
<tr>
<td>Gabe</td>
<td>3</td>
<td>4</td>
<td>Tail</td>
</tr>
<tr>
<td>Hugo</td>
<td>5</td>
<td>3</td>
<td>Tail</td>
</tr>
</tbody>
</table>

- Your HR team then starts to query the database for bagel consumption.

<table>
<thead>
<tr>
<th>HR</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jon</td>
<td>120</td>
</tr>
<tr>
<td>Tim</td>
<td>90</td>
</tr>
<tr>
<td>Ryu</td>
<td>150</td>
</tr>
</tbody>
</table>

Total: 100

QUERY BUDGET TO BE AGREED

Over many queries the average response will be 100
Multi-party computation (MPC) example

Gabe  6  😋

Paolo  10 😋

**Challenge**
Gabe and Paolo each have a number of Montreal bagels.

They want to know how many bagels they have together, without revealing their own stacks.

How can they do this?
Multi-party computation (MPC) example

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>( N_0 )</th>
<th>( N_1 )</th>
<th>( N_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gabe</td>
<td>6</td>
<td>3</td>
<td>+ 1</td>
<td>+ 2</td>
</tr>
<tr>
<td>Paolo</td>
<td>10</td>
<td>2</td>
<td>+ 5</td>
<td>+ 3</td>
</tr>
</tbody>
</table>

Gabe and Paolo each split their stacks and give their bagels to three helpers (MPC nodes).

Note: Simplified example; in reality, the numbers should be randomly selected from a large finite field.
Multi-party computation (MPC) example

Each helper adds his bagels together. Finally, the helpers add their numbers together.

Gabe: $\begin{array}{c} \text{6} \\ \text{+} \end{array} \begin{array}{c} \text{6} \\ \text{=} \end{array} \begin{array}{c} \text{3} \\ \text{+} \end{array} \begin{array}{c} \text{1} \\ \text{+} \end{array} \begin{array}{c} \text{2} \\ \text{=} \end{array}

Paolo: $\begin{array}{c} \text{10} \\ \text{+} \end{array} \begin{array}{c} \text{10} \\ \text{=} \end{array} \begin{array}{c} \text{2} \\ \text{+} \end{array} \begin{array}{c} \text{5} \\ \text{+} \end{array} \begin{array}{c} \text{3} \\ \text{=} \end{array}

Answer: $\begin{array}{c} \text{16} \\ \text{+} \end{array} \begin{array}{c} \text{16} \\ \text{=} \end{array} \begin{array}{c} \text{5} \\ \text{+} \end{array} \begin{array}{c} \text{6} \\ \text{+} \end{array} \begin{array}{c} \text{5} \\ \text{=} \end{array}$

None of the helpers learn anything about the original amount of bagels from either Gabe or Paolo.
Cooperative threat hunting: traditional

Classical setup, some sort of federated database: a mix local and central databases.

1. Local Intel
2. Public OSINT
3. Query interface

Party A

MISP

Analysis Query 1
Analysis Query 2
Analysis Query 3
Examples with Trigona campaign

<table>
<thead>
<tr>
<th>Union</th>
<th>Join</th>
</tr>
</thead>
<tbody>
<tr>
<td>• All malware hashes</td>
<td>• All malware hashes</td>
</tr>
<tr>
<td>• Include compiled Delphi</td>
<td>• All exfiltration URL, IP, Domain</td>
</tr>
<tr>
<td>• Include command-line flags</td>
<td>• All tools used on Window</td>
</tr>
<tr>
<td>• Include ransomware TTP</td>
<td>• Count vulnerabilities involved</td>
</tr>
<tr>
<td>• Count incidents in the last month</td>
<td>• List vulnerabilities</td>
</tr>
<tr>
<td>• Count total companies</td>
<td>• List OS versions affected</td>
</tr>
<tr>
<td>• Count total records/users</td>
<td></td>
</tr>
<tr>
<td>• Total payments demand</td>
<td></td>
</tr>
</tbody>
</table>
Example queries

SELECT count(name) FROM identity

Result: 10

SELECT name, roles FROM identity
WHERE identity.roles CONTAINS 'SOC'

Response: 1, your company Contoso inc

SELECT count(identity.id) AS affected
FROM indicator AS I
JOIN ON report AS r ON i.id IN r.object_ref
JOIN ON identity AS c ON c.id IN r.object_ref
WHERE (i.name LIKE 'trigona' OR i.description LIKE 'trigona')

Response 4 out of 10
Cooperative threat-hunting: MPC

Initial cooperation between three security operation teams...

...easy to expand with additional participants

1. Input module 2. Privacy engine 3. Analyse interface

Party A
Party B
Party C

Analysis Query 1
Analysis Query 2
Analysis Query 3, etc.

Party D
Party E
Stack components

IDE
• JupyterLab
• PyCharm

Visualization
• Streamlit
• Dash

Data Operations
• Pandas
• Spark
• Others

Application
• Flask/Django
• Vue.JS
• Others

Schema/Protocols
• VERIS
• STIX2.1
• ATT&CK
• CACAO

MPC Engine
• MPyC
• Others

DP Engine
• OpenMined
• DiffPrivLib
• Others

Graph DB

STIX ORM
CACAO
Kestrel
CVE

STIX 2.1
ATT&CK
ATTCK FLOW

A lot of moving parts to orchestrate and maintain, plus performance optimizations required.
An example of a growing network: SecureNed
Anonymous collection of sensitive cyber threat intelligence
An example of a growing network: SecureNed
Anonymous collection of sensitive cyber threat intelligence

1. Inputs via surveys

"We don't want details to be traceable to our organizations"

2. Answers are encrypted at the source

"We want to gather details about cyber incidents to identify patterns and coordinate a rapid response"

3. Results are shared in non-traceable form

"We don't want details to be traceable to our organizations"
Conclusions

Strong data model

Secure computation is a reality

Share sensitive data securely
Join our community

Reach out to us
paolo@priam.ai
gabriel.bassett@libertymutual.com
hugo.ideler@rosemanlabs.com