From evidences to actionable information

E.206

CIRCL Computer Incident Response Center Luxembourg

MISP Project
https://www.misp-project.org/

March 21, 2022

Objectives of this module

- How evidences can be useful for defense
- Why is contextualisation important
- What options do we have in MISP
- Best practises to encode and contextualise
- How can context be leveraged
- How to structure non-technical information
  - Practical case: Conti analysis

How evidences can be useful for defense

How evidences can be useful for defense

The most common recommendations to protect people and assets from cyber attacks are usually:

1. Maintaining softwares up to date
2. Staff awareness
3. Reliable Backups
4. Endpoints protection tools (IDS or SIEM)

How evidences can be useful for defense

- We can only help endpoints protection tools
- With the proper knowledge and methods, it is possible the maximize their accuracy and performance

These systems can rely on information extracted from

- Log files
- Network captures
- Disk forensic
- ...

However, from a MISP user perspective the hardest part in not to encode the raw evidences, it is to encode them so that they become **actionable**

Why is contextualisation important

Why is contextualisation important

- Allow the distinction between information of interest and raw data
- provide guidance on how to use this information can be used for for protection
- Filter out noise from information unrelated from the use-case or activity
- Enable risk assessment based on attack type, TTP and threat actor
- Allow triage in large volume of data
- Allow false-positive management

Expectations of the recipients

Most common expectations of recipients when receiving information

- Being able to **consume** the data
- Find information is **relevant** for them and their partners
- Being able to **understand** the data and its classification
- Assess the **credibility**, likelyhood and origin of the data

What do recipient hope to do with the data

Most common expectations of recipients for handling the data

- Being able to **filter** data efficiently for different use-cases
- Obtain as much **knowledge** out of the data as possible
- Know how this data was produced and where its **origin**
- Deduce why is the data **relevant** for them and how **critical** it is

Is context really that important?

- Raw data **is** useful but useless if you don't know what it is about
- That's why it should carry how and why it's relevant

```
1.2.3.9
137.221.106.104
28c643a1f69f9fca9481a4bc9f3f38f3
904afe59f6438848be96fd26fdeab01267070d25
evil.org
accounting.xlsx.exe
cat.jpg.exe
```

- In MISP, all data intrinsically have some context
  - **Type**: `ip-src` / `sha1` / `domain`
  - **Category**: `network-activity` / `payload-delivery` / `external-analysis`
  - **to_ids**: `yes` / `no`

Is context really that important?

- Sometime, more contextual information is not needed as data inherently convey its context:
  - Tor exit nodes
  - Botnet / C2 trackers
  - Ransomwares' bitcoin addresses
  - ...
- But most of the time, **context is essential**

What sort of context is pertinent

- To what kind of user this data is for
- What type of action can be performed with it
- Estimtaion on accuracy, reliability and likelyhood
- What are the impacts
- For threat actors:
  - Who is it? What tools were used?
  - What are their motivations? Who are their targets?
- How can we prevent/detect/block/remediate the attack

What options do we have in MISP

What options do we have in MISP

MISP offers mutliples means to contextualise

- Taxonomies
- Galaxies and Galaxy Clusters
- MITRE ATT&CK
- MISP Objects and relationships
- Sightings and `first_seen` / `last_seen`

Let's have an overview of each of them

Taxonomies

- Simple labels **standardised** on vocabularies
- Taxonomy tags often **self-explanatory**
  - `workflow:state="draft"`
  - doesn't need more explanation
- Triple tag system: `namespace:predicate="value"`
- Different organisational/community cultures require different nomenclatures
  - JSON libraries that can easily be defined without the involment of the MISP-project team

Tag	Events	Attributes	Tags
workflow:state="complete"	11	0	workflow:state="complete"
workflow:state="draft"	0	0	workflow:state="draft"
workflow:state="incomplete"	55	10	workflow:state="incomplete"
workflow:state="ongoing"	0	0	workflow:state="ongoing"

Galaxies and Galaxy Clusters

- `Galaxy`: Container to group galaxy clusters of the same type
- `Galaxy Cluster`: knowledge-base item with complex meta-data aimed for human consumption

- Community driven **knowledge-base libraries used as tags**
- Including descriptions, links, synonyms, meta information, etc.
- **Flexible** and **reusable**
- Works the exact same way as taxonomies but with more **meta-data**
  - `misp-galaxy:ransomware="CryptoLocker"`
  - Contains description, reference, documentation and other meta-data

Ransomware galaxy

Galaxy ID	373
Name	Ransomware
Namespace	misp
Uuid	3f44af2e-1480-4b6b-9aa8-f9bb21341078
Description	Ransomware galaxy based on...
Version	4

Value	Synonyms
.CryptoHasYou.	
777	Sevleg
7ev3n	7ev3n-HONE$T

MITRE ATT&CK and Galaxy Matrices

- MITRE ATT&CK is one of the best knowledge base of **adversary TTPs**
- **Widely used** and supported by a lot of tools
- The catalogue includes a **matrix-like** interface
- Offers clear visualisation for the kill chain

- MISP Fully support ATT&CK and embraced it's matrix structure
- Multiples matrices for other concerns are available:
  - `Badhra`: Similar to ATT&CK but for telecom operators
  - `attck4fraud`: Regrouped clusters related to fraud actions

MITRE ATT&CK and Galaxy Matrices

Pre Attack - Attack Pattern | **Enterprise Attack - Attack Pattern** | Mobile Attack - Attack Pattern

0 — 11 — Show all

Initial access	Execution	Persistence	Privilege escalation	Defense evasion	Credential access	Discovery	Lateral movement	Collection	Exfiltration	Command and control
Spearphishing Attachment	Scripting	Screensaver	File System Permissions Weakness	Process Hollowing	Securityd Memory	Password Policy Discovery	AppleScript	Data from Information Repositories	Exfiltration Over Alternative Protocol	Standard Application Layer Protocol
Spearphishing via Service	Command-Line Interface	Login Item	AppCert DLLs	Code Signing	Input Capture	System Network Configuration Discovery	Distributed Component Object Model	Data from Removable Media	Exfiltration Over Command and Control Channel	Communication Through Removable Media
Trusted Relationship	User Execution	Trap	Application Shimming	Rootkit	Bash History	Process Discovery	Pass the Hash	Man in the Browser	Data Compressed	Custom Command and Control Protocol
Replication Through Removable Media	Regsvcs/Regasm	System Firmware	Scheduled Task	NTFS File Attributes	Exploitation for Credential Access	Network Share Discovery	Exploitation of Remote Services	Data Staged	Automated Exfiltration	Multi-Stage Channels
Exploit Public-Facing Application	Trusted Developer Utilities	Registry Run Keys / Start Folder	Startup Items	Exploitation for Defense Evasion	Private Keys	Peripheral Device Discovery	Remote Desktop Protocol	Screen Capture	Scheduled Transfer	Remote Access Tools
Spearphishing Link	Windows Management Instrumentation	LC_LOAD_DYLIB Addition	New Service	Network Share Connection Removal	Brute Force	Account Discovery	Pass the Ticket	Email Collection	Data Encrypted	Uncommonly Used Port
Valid Accounts	Service Execution	LSASS Driver	Sudo Caching	Process Doppelgänging	Password Filter DLL	System Information Discovery	Windows Remote Management	Clipboard Data	Exfiltration Over Other Network Medium	Multilayer Encryption
Supply Chain Compromise	CMSTP	Rc.common	Process Injection	Disabling Security Tools	Two-Factor Authentication Interception	System Network Connections Discovery	Windows Admin Shares	Video Capture	Exfiltration Over Physical Medium	Domain Fronting
Drive-by Compromise	Control Panel Items	Authentication Package	Bypass User Account Control	Timestomp	LLMNR/NBT-NS Poisoning	Network Service Scanning	Remote Services	Audio Capture	Data Transfer Size Limits	Data Obfuscation
Hardware Additions	Dynamic Data Exchange	Component Firmware	Extra Window Memory Injection	Modify Registry	Credentials in Files	File and Directory Discovery	Taint Shared Content	Data from Network Shared Drive		Connection Proxy
	Source	Windows Management Instrumentation Event Subscription	Setuid and Setgid	Indicator Removal from Tools	Forced Authentication	Security Software Discovery	Application Deployment Software	Data from Local System		Commonly Used Port
	Space after Filename	Change Default File	Launch Daemon	Hidden Window	Keychain	System Service Discovery	Third-party Software	Automated Collection		Data Encoding

MISP Objects

Atomic attributes are great, but are lacking a way to express that some can be related to others.

MISP Objects are there to fill the gap:

- **Template system** to build complex structures composed of attributes
- Logically **group attributes** that are contextually linked between each others
  - A *file* object can contain: a size, name, content, cryptographic hashes, etc.
  - A *car* object can contain: a brand, a model, a license plate, etc.

MISP Objects

A *file* object

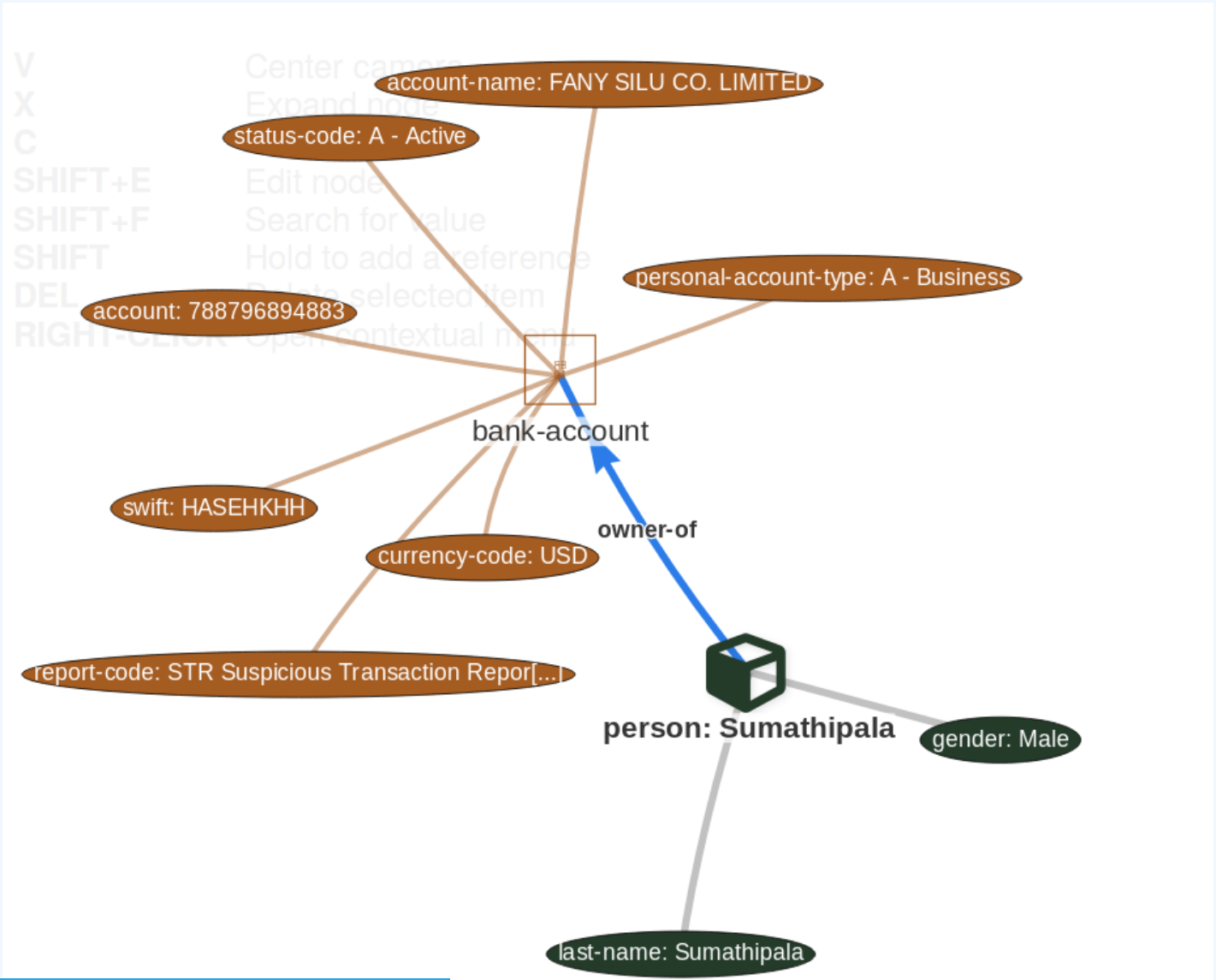
2018-03-27	**Name:** file **References:** 1			
2018-03-27	Payload delivery	**filename:** filename	putty.exe	+
2018-03-27	Other	**size-in-bytes:** size-in-bytes	774200	+
2018-03-27	Other	**entropy:** float	6.7264597226	+
2018-03-27	Payload delivery	**md5:** md5	b6c12d88eeb910784d75a5e4df954001	+
2018-03-27	Payload delivery	**sha1:** sha1	5ef9515e8fd92a254dd2dcdd9c4b50afa8007b8f	+
2018-03-27	Payload delivery	**sha256:** sha256	81de43198730467613413870 5fc1c21188ad7f27edf6b77a6551aa693194485e	+
2018-03-27	Payload delivery	**sha512:** sha512	e174ecf4fffb36d30c2cc66b37f82877d421244c924d5c9f39f2e0f37d85332b7d107d5ac5bd19cb7ffdcdbdd8b506d488faa30664ef610f62f3970c163cca76	+
2018-03-27	Payload delivery	**malware-sample:**	putty.exe	+

Relationships

- Analysts want more than a table of atrtibute, they want to see how each of them **interact** with the others
- Relationships are essentials to describe scenarios or stories with the data
- MISP allow these relationship to be built between objects

Relationships

A relationship betwen a person and its bank account



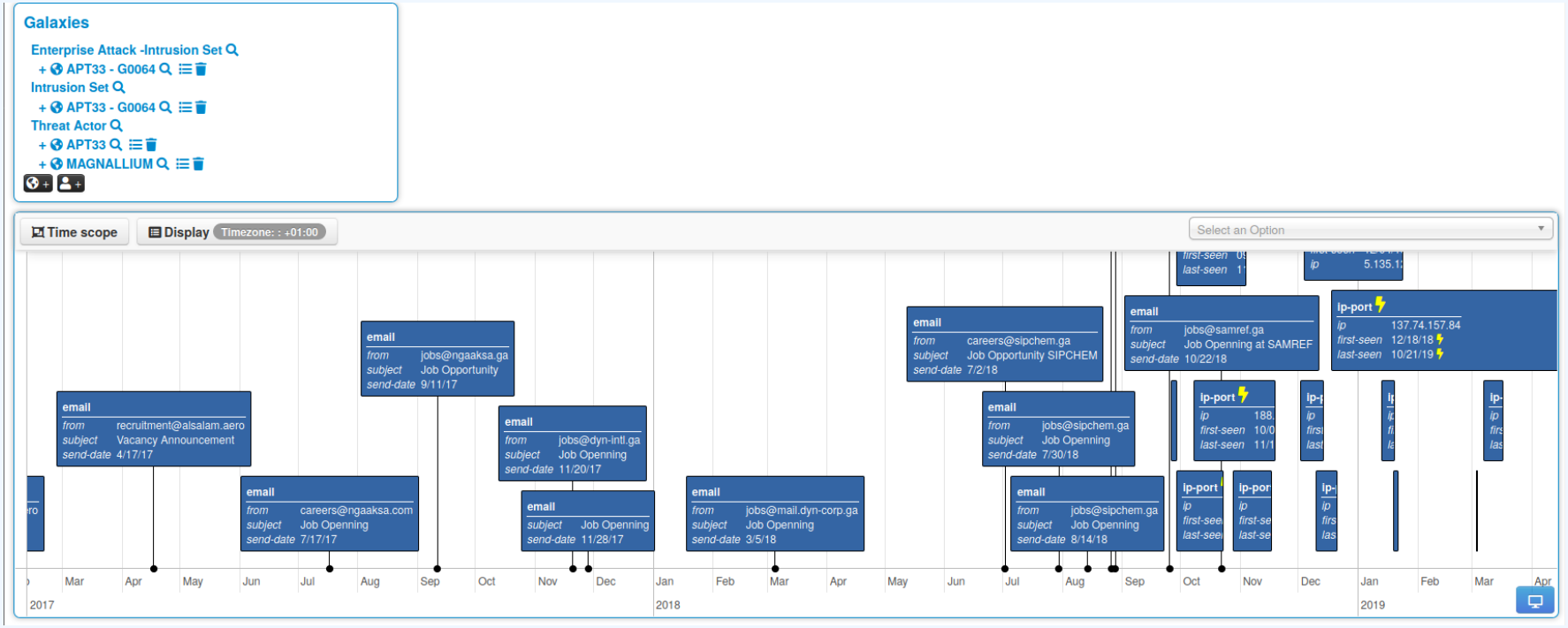
Timeliness with Sightings and `first_seen` / `last_seen`

Adding **Temporality** os a good way to avoid having the data frozen in time

- Sightings
  - Allows to signal the fact that an indicator was **sighted**
  - They can record the time and where they were the sighting was seen
  - E.g.: Sight C2 servers or phishing websites
- `first_seen` / `last_seen`
  - These two data-points allow to set when the specified item was first and last seen
  - Enables the visualisation of data timeframe with a timeline
  - *e.g: Track the duration of a campaign or duration for which something was online*

Timeliness with Sightings and first_seen / last_seen

Screenshot of the timeline widget when viewing a MISP event



Best practises to encode and contextualise

Encoding: Event

Always keep in mind that the recipient is a human:

- Include a self-explanatory title
- Make it concise
- Include a report along with the machine parsable data
  - It can either be included as an attribute or as an event-report

It will make the live of the analyst easier: That analyst might end up being you!

Encoding: Attributes and objects

Prefer the use of object rather than attributes for attributes intrinsically linked together.

Atomic data by themselve rarely exists: They are often related to something else

- Interactions between between elements are frequent
  - They can often be described by using verbs: `connects-to`, `contain-within`, ...
- A story can be inferred without the need to put it into words
  - *"`file` was attached to `email` which when extracted contained a `malware` connecting to `ip-address` which was used `C2`"*
- Properly encoding these relationships turns flat data into a **connected graph**

Contextualisation: Distributions and permissible actions

Adding context on **what** actions can be done on the data and **who** can it be shared with

- Permissible actions taxonomies:
  - *PAP*: Permissible Actions Protocol
  - *IEPF*: Information Exchange Policy (IEP) Framework
  - `pap:white` No restrictions in using this information
- Sharing level taxonomies:
  - *TLP*: Traffic Light Protocol
  - *IEPF*: Information Exchange Policy (IEP) Framework
  - `tlp:green`: Limited disclosure, restricted to the community

Contextualisation: Attributes and their context

- Each data point has a meaning and tells a part of the story
- One should try to capture the answer to these question when contextualising:
  - In what context was this IoC seen?
  - Is it related to compromision? Does it tell us anything about the adversary infrastructure?
  - Was it used to exfiltrate data? Did it acted as a C2?
  - Did it perform subsequent actions?
  - ATT&CK can procure even more knowledge

Contextualisation: Attributes and their context

However, think twice before tagging:

- If a tag applies to the whole content of the event, it should be attached on the event instead
- If the tag offers no real utility or hinder your ability to analyse the whole dataset, it should probably be ignored

Contextualisation: Origin, likelyhood and reliability

- The source of information has an impact on how people evaluates its trust
  - Data without a source / origin might be considered unreliable
  - *i.e: A research paper without citing its sources is useless*
- MISP bridges people and and communities
  - The more one is connected, the greater the quantity and diversity of data
  - Not everything you read on the internet is true!

Contextualisation: Origin, likelyhood and reliability

If you can't share the source, provide the trust in the source

- Include the reliability and the credibility of the information
  - Taxonomy: `admiralty-scale`
  - *i.e:* *`admiralty-scale:source-reliability="Usually reliable"`*
- Include the quality and likelyhood
  - Taxonomy: `estimative-language`
  - *i.e:* *`estimative-language:likelihood-probability="very likely"`*

Contextualisation: Make the attribution

- The purpose is not to blame but to identify the attacker's **intent**
- Knowing the intent greatly help to:
  - Know the objectives
  - Understand what are the targeted assets
  - Deduce the treat level
- It allows to identity behaviors
  - Might speed up the next investigation
  - Might boostrap the analysis procdess

Contextualise: Provide advices on how to protect themselves

To help recipients to better protect themselve, additional information can be provided.

- Indicate what can be done with the data
  - Use it to feed an IDS
  - Perform historical search with a SIEM to find a potential compromision
  - Inform your peers against a new type of threat
- Provide additional supporting materials
  - The original report form which the data is coming from
  - Home-brew scripts
  - Sigma rules for SIEM searches
  - Context and configurations under which the analysis was done

How can context be leveraged

Leveraging the context

Let's make use of this well-structured, context-rich data

- Incorporate all contextualisation options into API filters

```
{
    "AND": [
        "admiralty-scale:source-reliability=\"Reliable\""
    ],
    "OR": [
        "threat-actor=\"Sofacy\"",
        "sector=\"Chemical\"",
        "country=\"Luxembourg\"",
    ]
}
```

Leveraging the context

- On-demande potential false positive exclusion
- Warninglist system helps to exclude known false-positives reducing alert-fatigue

LIST OF KNOWN IPV4 PUBLIC DNS RESOLVERS

Id	89
Name	List of known IPv4 public DNS resolvers
Description	Event contains one or more public IPv4 DNS resolvers as attribute with an IDS flag set
Version	20181114
Type	string
Accepted attribute types	ip-src, ip-dst, domain|ip
Enabled	**Yes** (disable)

Values

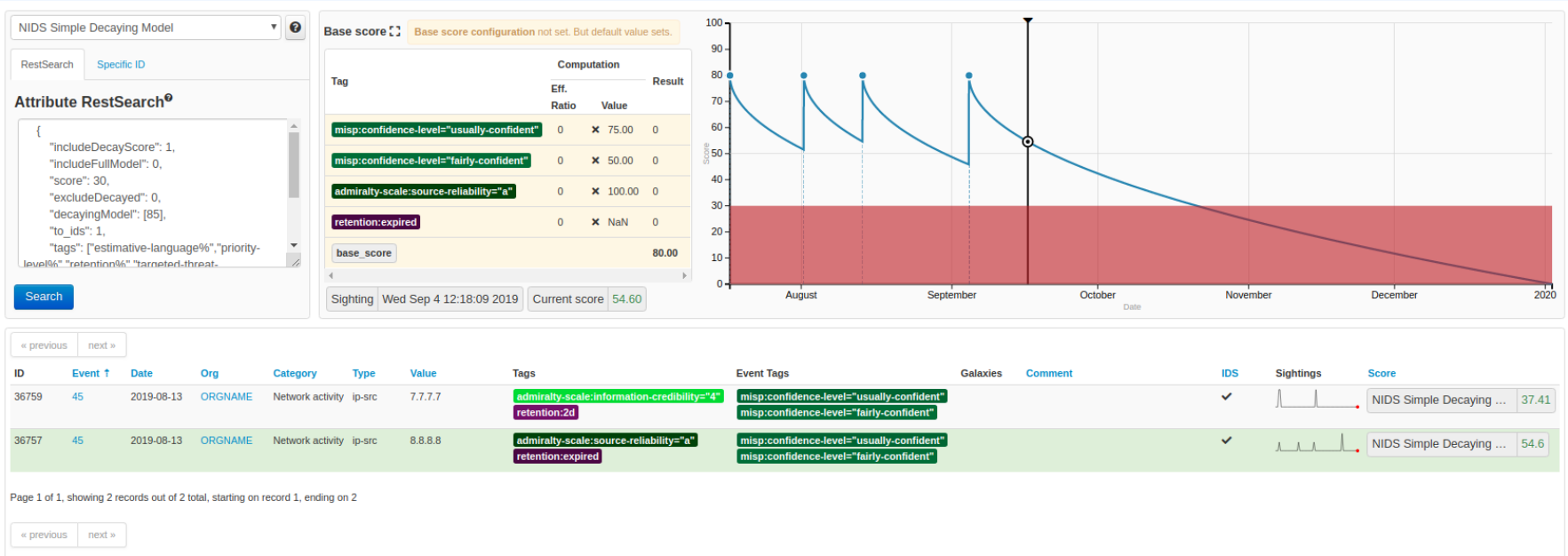
1.0.0.1
1.1.1.1
1.11.71.4

Warning: Potential false positives

List of known IPv4 public DNS resolvers
Top 1000 website from Alexa
List of known google domains

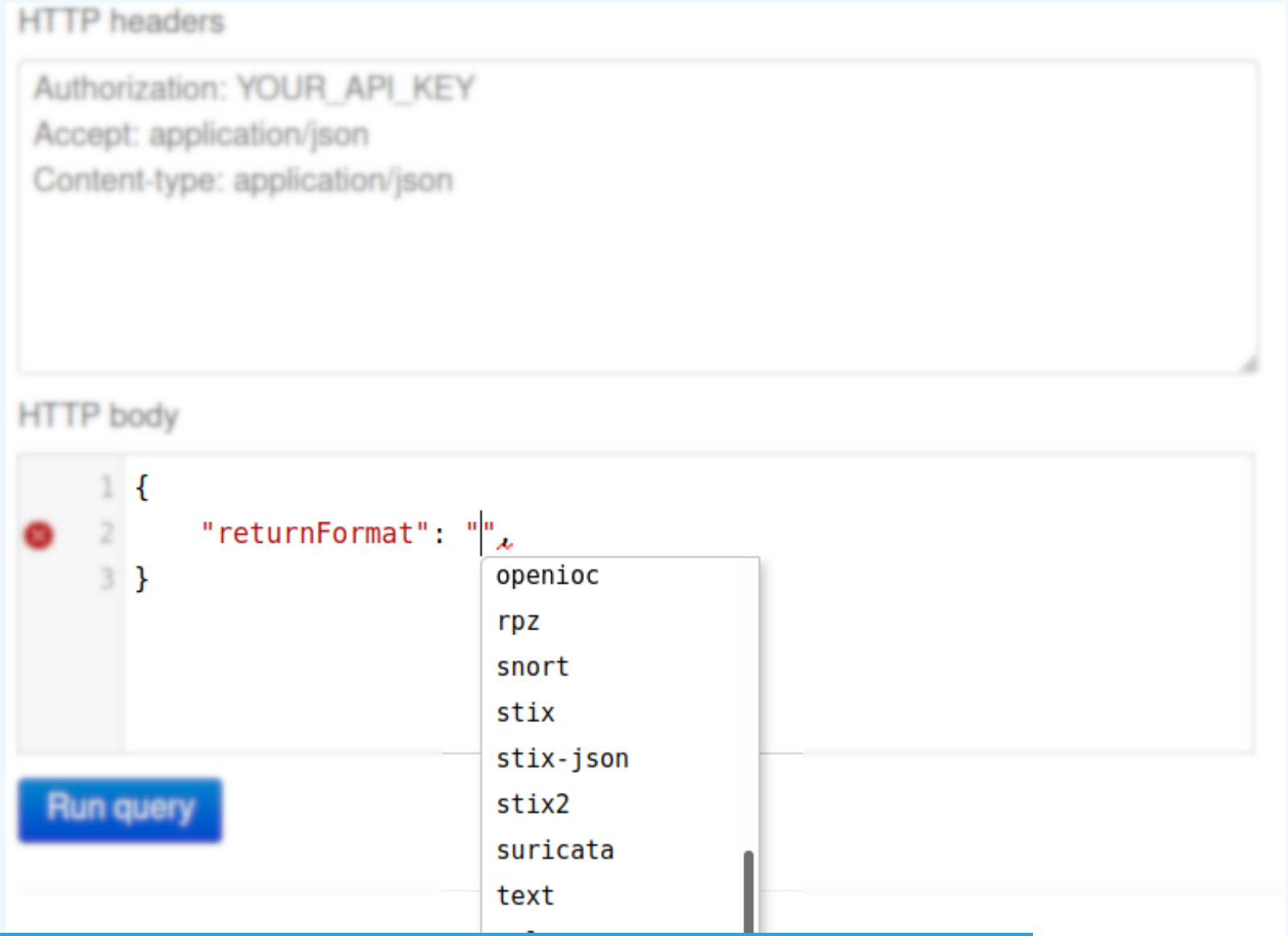
Leveraging the context

- IoC prioritization and lifecycle management
- Integrate decay models to filter out expired/unrelevant data



Leveraging the context

- Allow users to build their own export module



Enabling common user profiles to better perform their tasks

How does different user profiles benefits to most of well-structured, context-rich data

- **incident responder**: Self-explanatory data relieves pressure and reduces the change of misunderstanding it
- **SOC operator**: Reduce alert-fatigue and energy to filter unwanted data
- **ISP**: Ease the task to decide if the data is fit for blocking based on trust and context the data was seen in
- **threat analyst**: Provide insight on the modus operandi and goals of attacker
- **risk analyst**: Help highlighting potential security gaps and formulate advices on preventive actions
- **decision maker**: Guide resources allocation based on current/emerging threats for their region and sector

How to structure non-technical information

Objectives

- Identify non-technical data that can be useful for an investigation,
- Illustrate how non-technical and technical data can interact to produce meaninful insights,
- Model these interactions,
- Outline what Socio-Technical intercations are useful to share.

We live in Socio-Technical Systems

Computer and their security is linked to human activities:

- Technical traces show human activities,
- Technical traces can convey human intent,
- Human interactions can explain and give context to Technical traces,
- CyberCrime requires infrastructures and logistics that are discussed between humans,
- TTPS are discussed and exchanged,
- Human interaction can help attributing attacks to threat actors,
- Human interaction can help deciphering intent and motives, and discriminate human error from sabotage.

Plan

Use OSINT and data leaks to:

- bring context to other ransomware cases,
- better understand the gang day to day operations,
- get insights on events' timeline,
- confirm or invalidat previous hypotheses,
- select relevant information to share and produce an intelligence report.

Conti ransomware group leak analysis

Ransomware Jabber chats leak

Published on Twitter:

Contained XMPP server logs:

```
{
  "ts": "2020-09-08T00:28:49.471678",
  "from": "ceram@q3mcco35auwcstmt.onion",
  "to": "stern@q3mcco35auwcstmt.onion",
  "body": "Проинструктируйте меня. Что делать?"
}
```

Ransomware Jabber chats leak in AIL

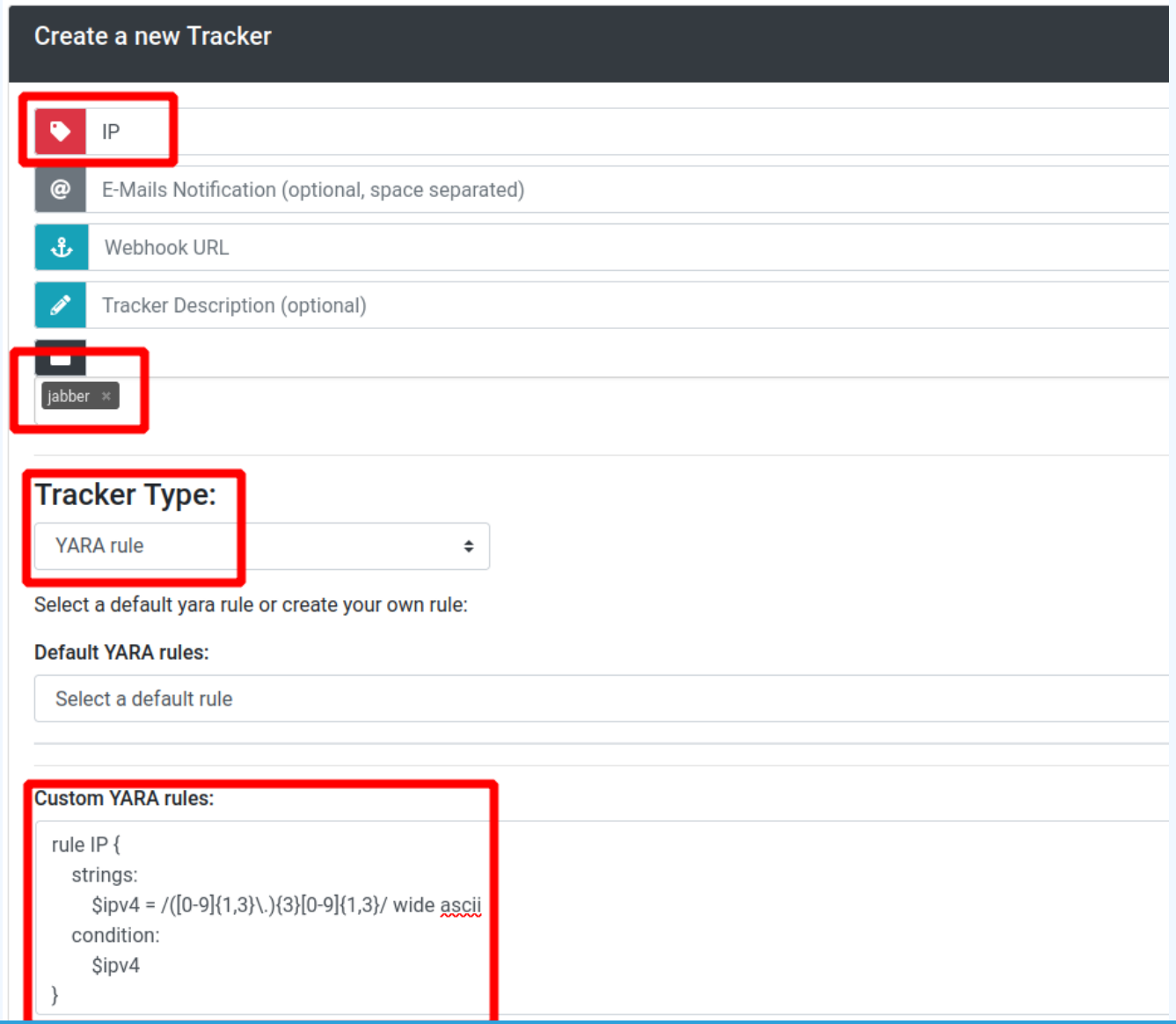
We use AIL[1] to dig into the data:

- AIL processes the data and search for relevant information
  - PGP keys,
  - Bitcoin addresses, maybe others,
  - onion hidden services,
  - IP addresses.
- Once we find relevant information we push it into MISP,
- we use MISP correlation engine to find relevant past cases.

[1] https://ail-project.org/

Ransomware Jabber chats leak in AIL

Prepare AIL to detect IPv4 addresses by creating a 'tracker':



Dig

use MISP to correlate this information from past cases,

- TODO

Dig

ease attribution of these past cases through IoC,

- TODO

add context to past cases from the chats logs in order to ease their investigation,

- TODO

Dig

support colleagues and other CSIRTs with real Intelligence.

- TODO

To sum it all up

- Given the growth and diversification and maturity of users, **contextualisation is becoming essential**
- Well-structured, context-rich data is good as it enables better **decision making**
- It will rise user capabilities and thus **improve protection**
- MISP has a format and tools designed to support contextualised data

Acknowledgment

Provide sources along with the data!

- Turning data into actionable intelligence - advanced features in MISP supporting your analysts and tools (CIRCL.lu)
  - https://www.enisa.europa.eu/events/2019-cti-eu/2019-cti-eu-bonding-eu-cyber-threat-intelligence
- Colouring Outside the Lines (Andras Iklody & Trey Darley)
  - https://www.first.org/conference/2020/recordings
- MISP Training Materials
  - https://github.com/MISP/misp-training