Partnership for Cyber Security Innovation

WEBINAR

Automated Threat Actor Profiling

25 MAY 2021, 10:00 CET

HOST

REINDER WOLTHUIS

SPEAKERS

Richard Kerkdijk
TNO

Lalit Bhakuni
ABN AMRO

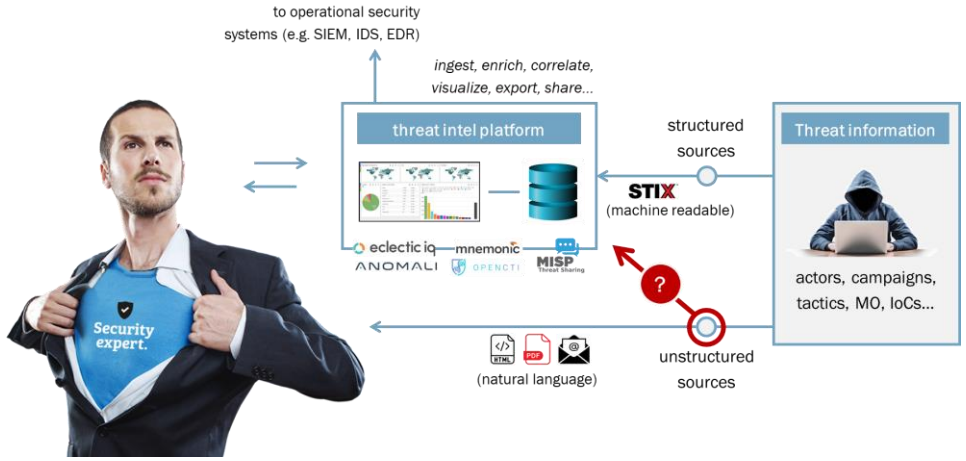
Nicole Gervasoni
TNO

Contents

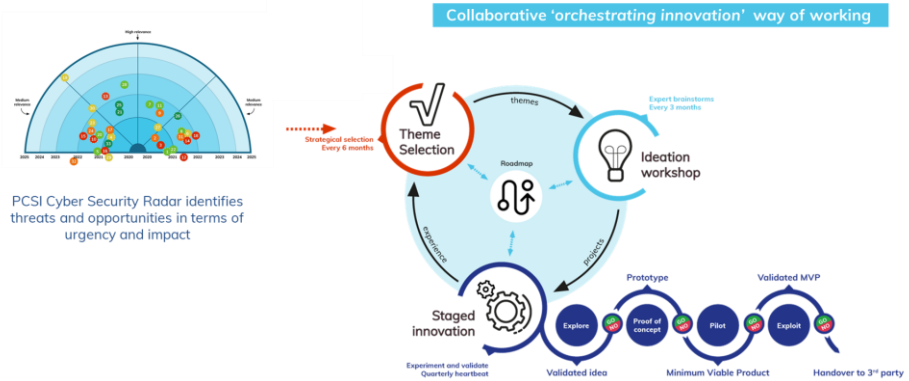
- Introduction to PCSI
- The challenge
- State of the art
- Attributes and taxonomies
- Implementation
- Demonstration
- Take aways
- Q&A

In this webinar you'll learn more about...

1. Threat Actor profiling



2. The PCSI programme



Partnership for Cyber Security Innovation

ABN·AMRO achmea ING de volksbank TNO

Practical stuff

- Attendees are muted
- If you have questions or remarks, please use the chat; the chat will be moderated
  - Questions will either be addressed directly or in the discussion section at the end of the presentation

Introduction

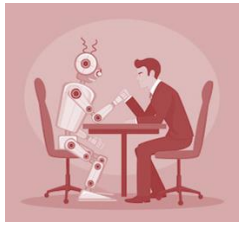
The need for cyber security innovation

> “the cybersecurity community is still far from striking the balance between defenders and attackers.”
>
> […]
>
> “the increased defence levels and expenses cannot successfully reduce levels of cyberthreat exposure.”

ENISA’s Threat Landscape Report, 2017

Increasingly complex and dynamic ICT

Increasing number and complexity of attacks



Attacks get more automated, response is human-centered

Shortage of skilled cybersecurity staff

Collaborative innovation

Isolated protection | Leaves gaps

lack of innovation | isolated efforts

Shared innovation | Creates bonds

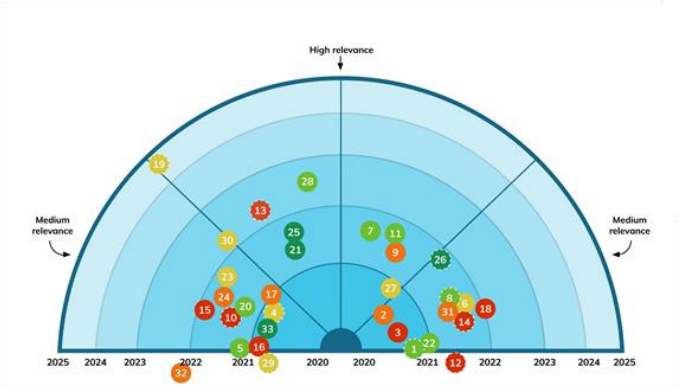
focus on innovation | joint effort

> Alone we are smart. Together we are brilliant.
>
> - Steven W. Anderson

Partnership for Cyber Security Innovation

ABN·AMRO achmea ING de volksbank TNO

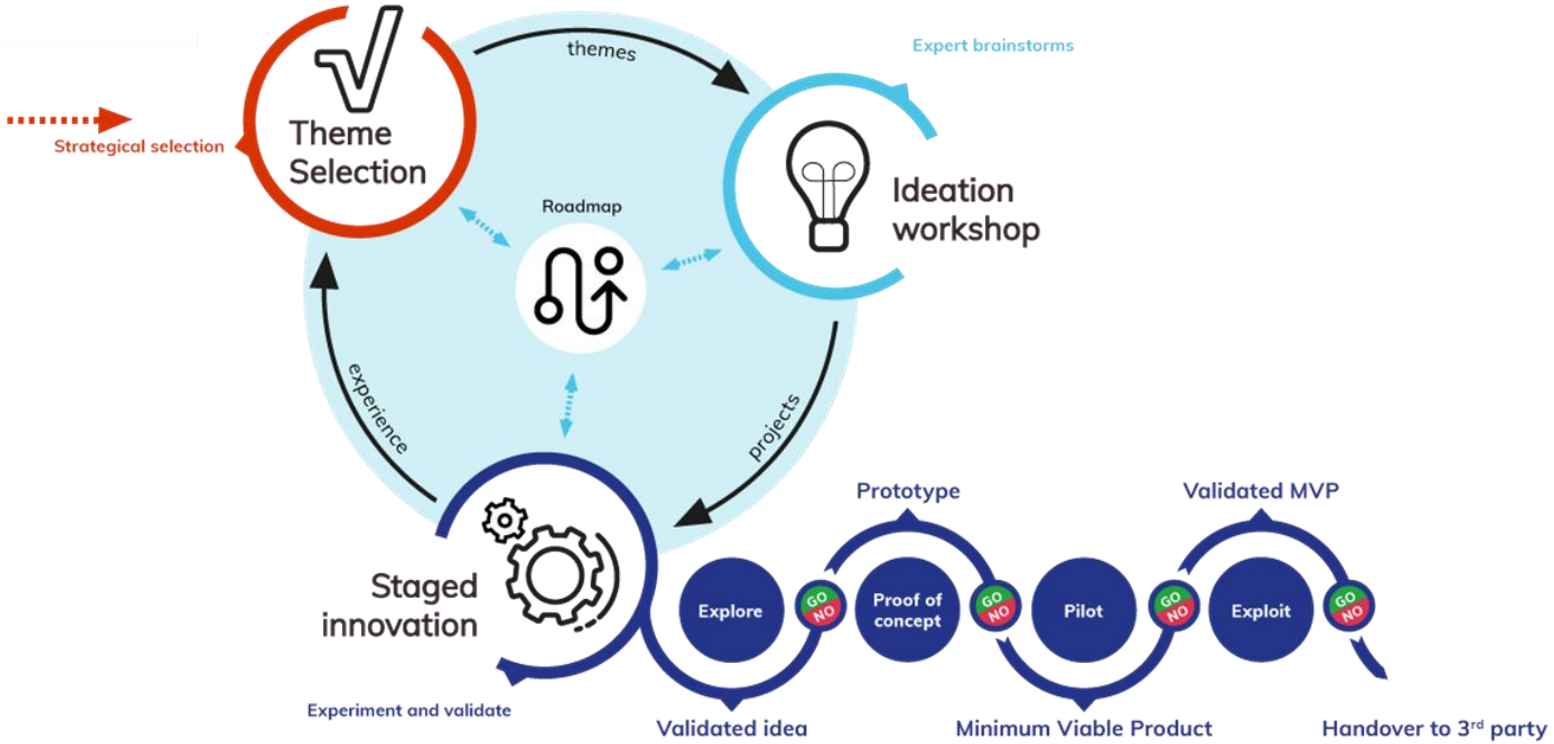
PCSI: Orchestrating innovation



PCSI Cyber Security Radar identifies threats and opportunities in terms of urgency and impact

Collaborative 'orchestrating innovation' way of working

Heartbeat of 4 months



Partnership for Cyber Security Innovation

ABN·AMRO achmea ING de volksbank TNO

Sample projects

Crystal ball DDOS detection

- Investigate whether new DDoS attacks can be predicted before they have an actual impact
- Analyse public and private data using AI techniques, to give out a warning (like a tsunami warning) before the attack actually hits

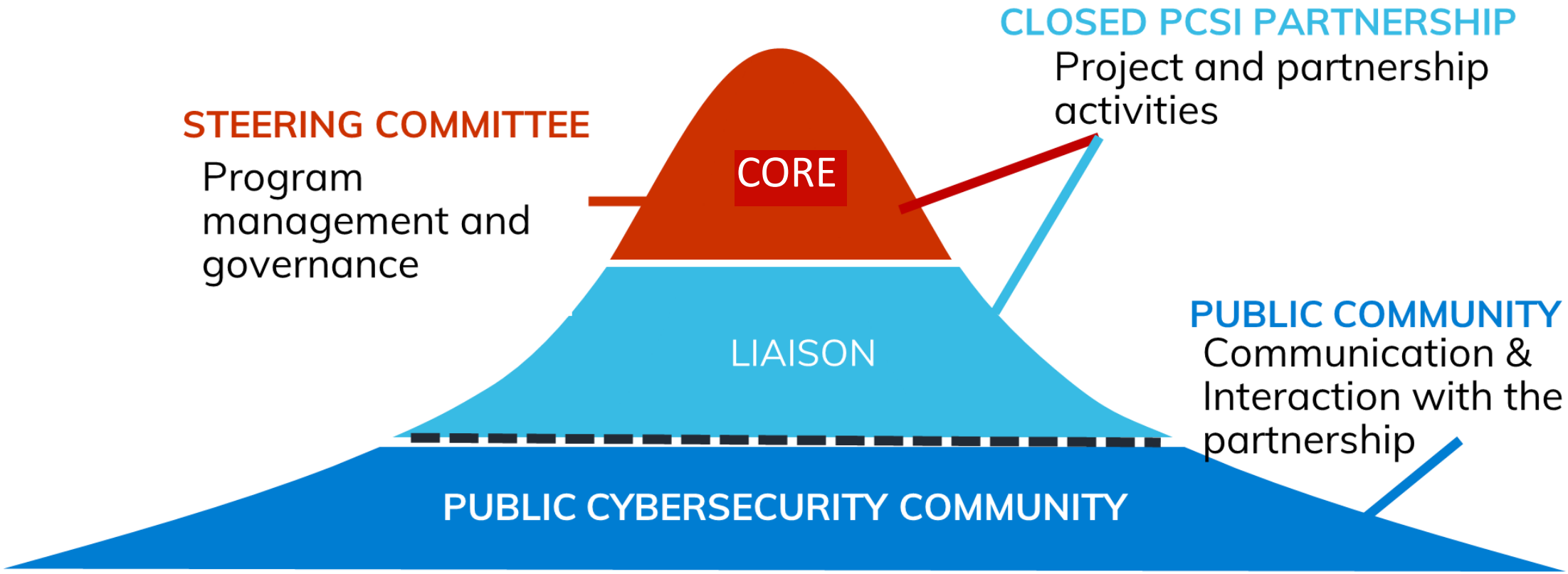
Security journey

- Investigating whether the ‘customer journey methodology’ can be used to reduce human induced security incidents
- We use data-driven AI tooling to identify internal processes, define the right steps for each role involved, and determine how to make employees security aware

Collaborative fleet

- Set up ‘honey pots’ with the specific purpose of tempting attackers to target it, so we can monitor their behavior
- We combine information from a large number of honey pots, resulting in strategic insights that would not be possible for individual organisations

PCSI Partner model



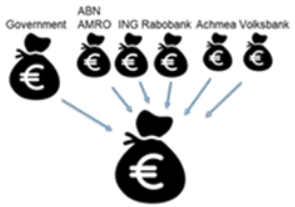
"To increase the resilience level of Dutch society and participating organisations against tomorrow's cyber threats and vulnerabilities"

(Converted into PCSI)

PCSI predecessor

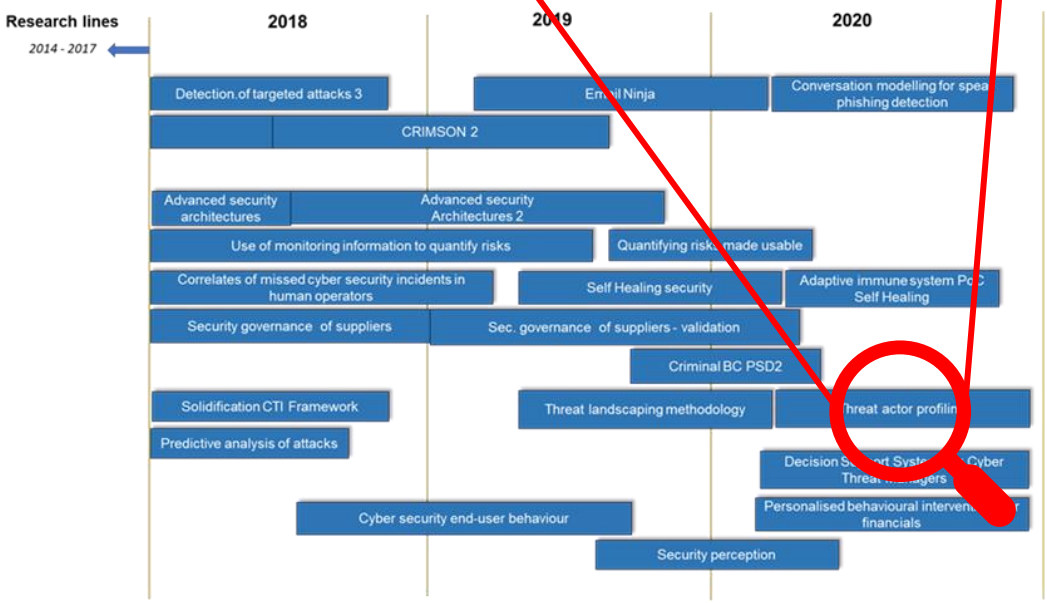
Shared working

Sharing data



Shared funding

Threat actor profiling



Shared Research Program cyber security
2014 - 2020 (2020 onwards: PCSI)
Yearly 1 million euro to innovate on cyber security

ABN·AMRO Rabobank ING achmea de volksbank TNO innovation for life

https://www.tno.nl/srpcybersecurity

Direct application of results by partners and increase resilience society by publications, presentations, spin-offs etc.

Today’s presenters

www.pcsi.nl follow us

Collecting and processing threat information

to operational security systems (e.g. SIEM, IDS, EDR)

ingest, enrich, correlate, visualize, export, share…

threat intel platform

eclectic iq · mnemonic · ANOMALI · OPENCTI · MISP Threat Sharing

structured sources

STIX (machine readable)

(not exhaustive)

Threat information

- Threat actor names, aliases and affiliations
- Threat actor objectives and known targets
- Historic and ongoing attacker campaigns
- Tools, techniques and procedures (TTP/ MO)
- Indicators of Compromise (IoCs)

?

unstructured sources

HTML · PDF · @

(natural language)

ABN·AMRO · achmea · ING · de volksbank · TNO

Challenges in processing narrative intel

Challenges

- ***Limited time to read and analyze*** - Uncountable impressive research reports/blogs by researchers in an unstructured format
- ***No industry standard for actor naming convention*** - Multiple naming conventions for same threat actor group(s)
- ***Analytical capability*** – Carving right Tactics, techniques and procedures (TTPs) is not necessarily trivial

Image Source - Forbes

Envisaged setup and key requirements

taxonomies, knowledge bases, training data

predefined (preferably acknowledged) taxonomies

source document (e.g. threat report)

heterogeneous input sources

NLP module

identify, extract & classify threat actor attributes

analyst feedback and intervention

structured threat actor attributes

MO + general characteristics

CTI platform

structure, correlate with other intel & present to threat analyst

independent of any particular TIP

STATE OF THE ART

CTR to TT

How to classify unstructured CTR data into ATT&CK Tactics and Techniques?

Cyber Threat Reports

- Unstructured text
- Several formats: pdf, html, doc

ATT&CK frameworks classification

- Tactics represent the “why” of an ATT&CK technique or sub-technique
  Tactics are treated as “tags” within ATT&CK where a technique or sub-technique is associated or tagged with one or more tactic categories.
- Several techniques represent “how” an adversary achieves a tactical objective by performing an action.
- Several sub-techniques, describing more specific means by which adversaries achieve tactical goals at a lower level than techniques.

Natural Language Processing

- Problem: It is difficult to retrieve relevant information from unstructured data.
- NLP is an AI technique that tackles this problem by combining techniques of statistics with machine learning.

State of the art solutions

TRAM - Threat Report ATT&CK® Mapping

rcATT - reports classification by adversarial tactics and techniques

Common features

- Train set from ATT&CK database
- Python libraries (scikit + nltk)
- Final human decision on the output
- Every new CTR the algorithm can be retrained

Tram

By MITRE

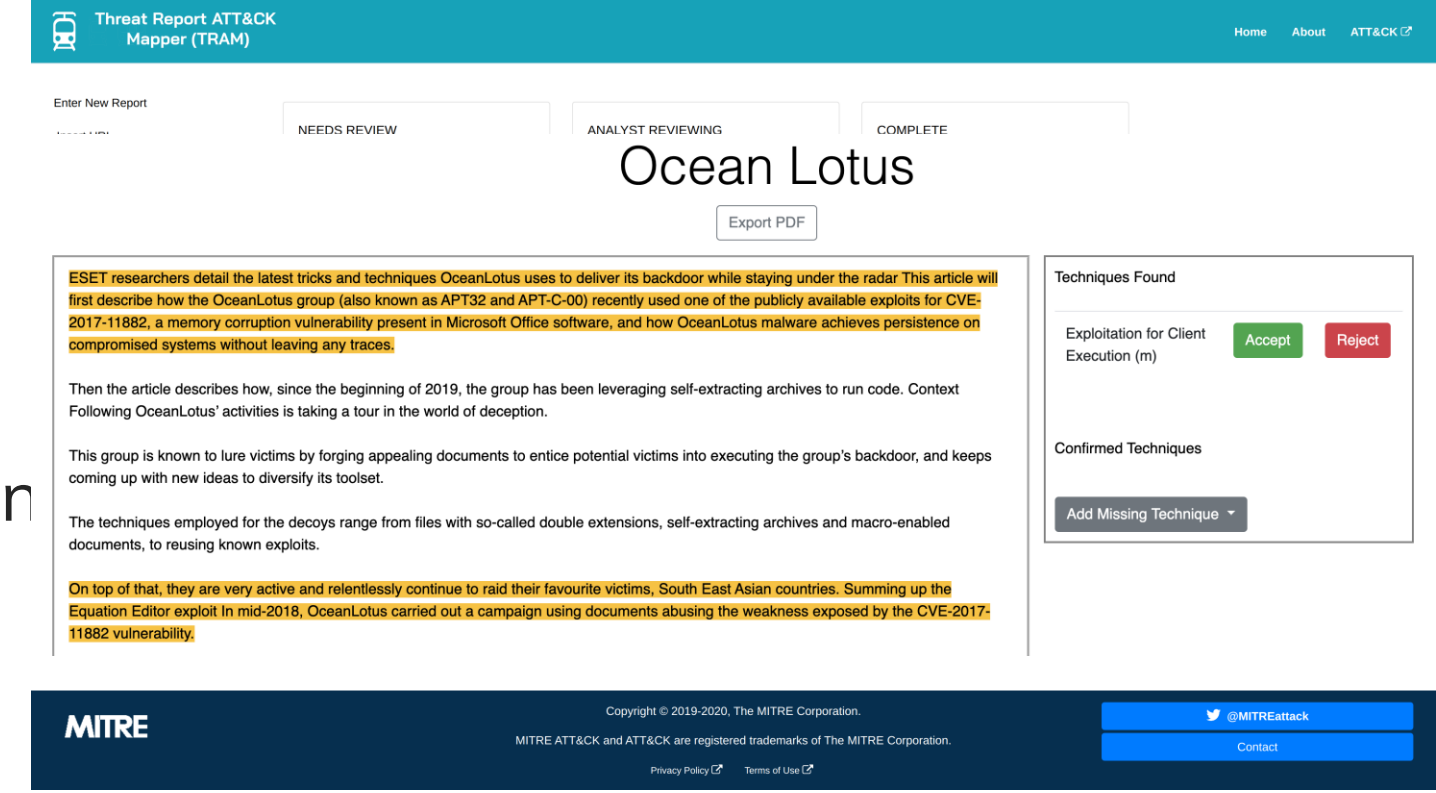
CTR webpage → pdf table (on techniques)

Approach

- Multi class logistic regression

Results

- Unknown

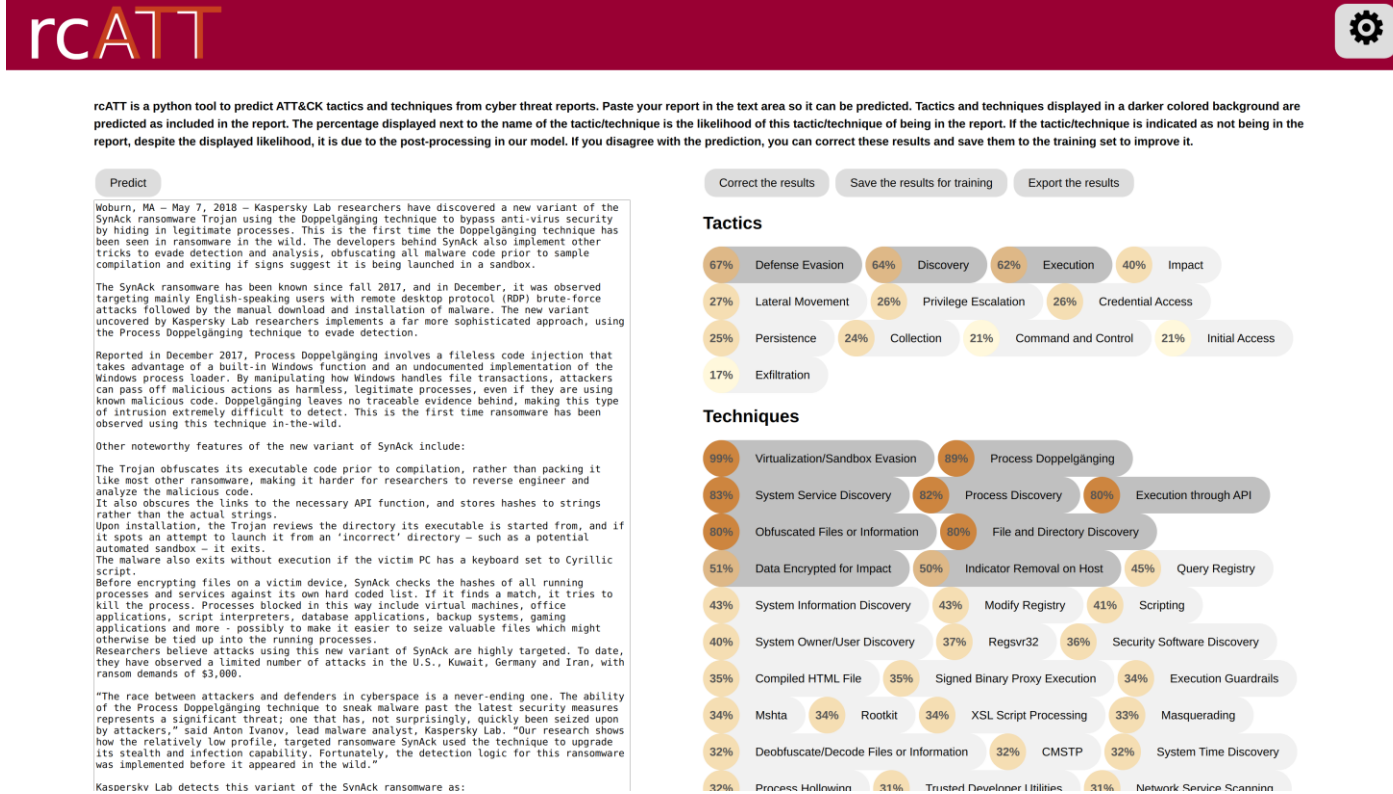


rcATT

Text reports → STIX file

Approach

- TF-IDF weighted bag-of-words + binary Linear SVC
- Post-processing techniques



Results

Possible macro-averaged $F_{0,5}$=80% on tactics prediction and $F_{0,5}$>27.5% on techniques prediction

$$F_{0.5}\left(y_t, y_p\right) = 1.25 \cdot \frac{Pr\left(y_t, y_p\right) \cdot Re\left(y_t, y_p\right)}{\left(0.25 \cdot Pr\left(y_t, y_p\right)\right) + Re\left(y_t, y_p\right)}$$

Partnership for Cyber Security Innovation

ABN·AMRO achmea ING de volksbank TNO

ATTRIBUTES AND TAXONOMIES

Attributes and Taxonomies

ATT&CK® ⇨ Tactics and techniques taxonomy

ThaiCERT Thailand Computer Emergency Response Team a member of ETDA ⇨ Actor name and motivation

Tactics and techniques

ATT&CK®

Partnership for Cyber Security Innovation

Actor name and motivation

ThaiCERT
Thailand Computer Emergency Response Team
a member of ETDA

Motivation

- Information theft and espionage
- Financial crime
- Financial gain
- Sabotage and destruction

APT group: APT 29, Cozy Bear, The Dukes

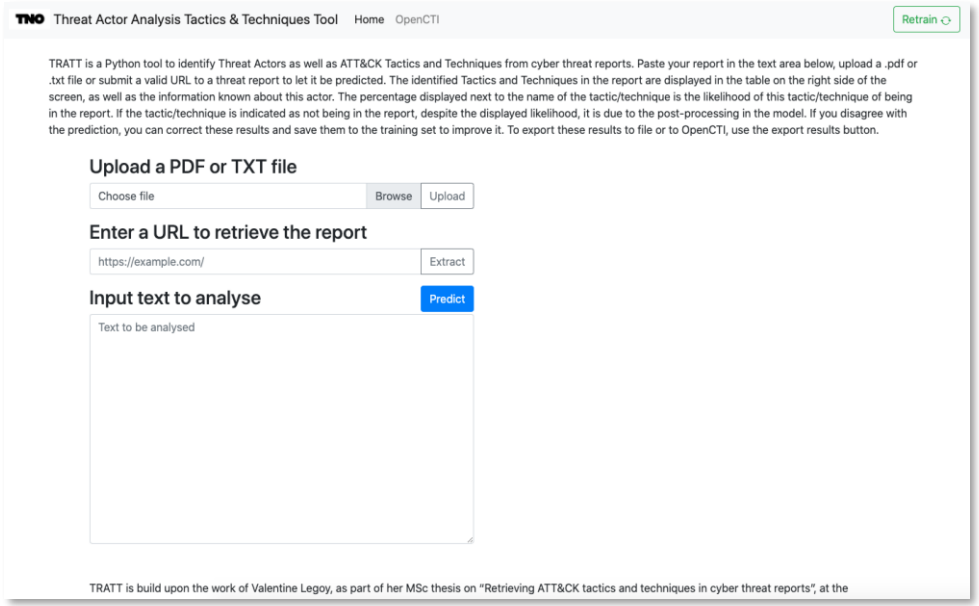
Names	APT 29 *(Mandiant)* Cozy Bear *(CrowdStrike)* The Dukes *(F-Secure)* Group 100 *(Talos)* Yttrium *(Microsoft)* Iron Hemlock *(SecureWorks)* Minidionis *(Palo Alto)* CloudLook *(Kaspersky)* ATK 7 *(Thales)* ITG11 *(IBM)* Grizzly Steppe *(US Government)* together with Sofacy, APT 28, Fancy Bear, Sednit
Country	Russia
Sponsor	State-sponsored
Motivation	Information theft and espionage
First seen	2008
Description	(F-Secure) The Dukes are a well-resourced, highly dedicated and organized cyberespionage group that we believe has been working for the Russian Federation since at least 2008 to collect intelligence in support of foreign and security policy decision-making. The Dukes primarily target Western governments and related organizations, such as government ministries and agencies, political think tanks, and governmental subcontractors. Their targets have also included the governments of members of the Commonwealth of Independent States; Asian, African, and Middle Eastern governments; organizations associated with Chechen extremism; and Russian speakers engaged in the illicit trade of controlled substances and drugs. The Dukes are known to employ a vast arsenal of malware toolsets, which we identify as MiniDuke, CosmicDuke, OnionDuke, CozyDuke, CloudDuke, SeaDuke, HammerDuke, PinchDuke, and GeminiDuke. In recent years, the Dukes have engaged in apparently biannual large-scale spear-phishing campaigns against hundreds or even thousands of recipients associated with governmental institutions and affiliated organizations

IMPLEMENTATION

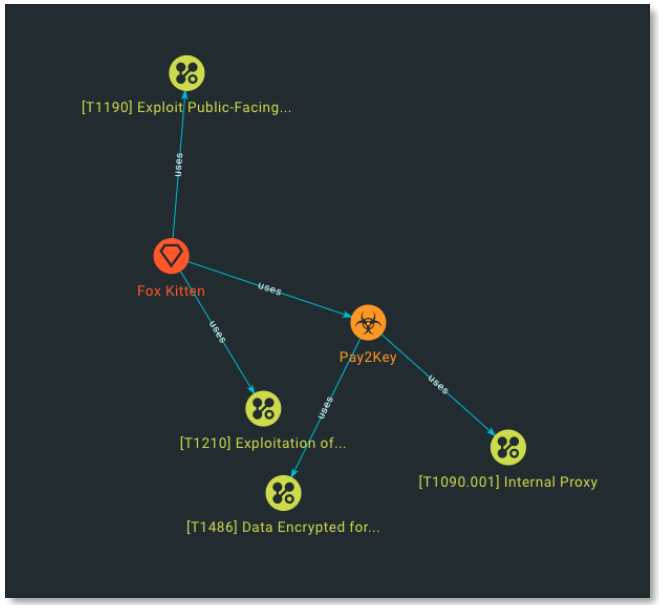
Final analyst flow

It's ATRAPP

Threat report



ATRAPP: Automated Threat Report ATT&CK maPPer



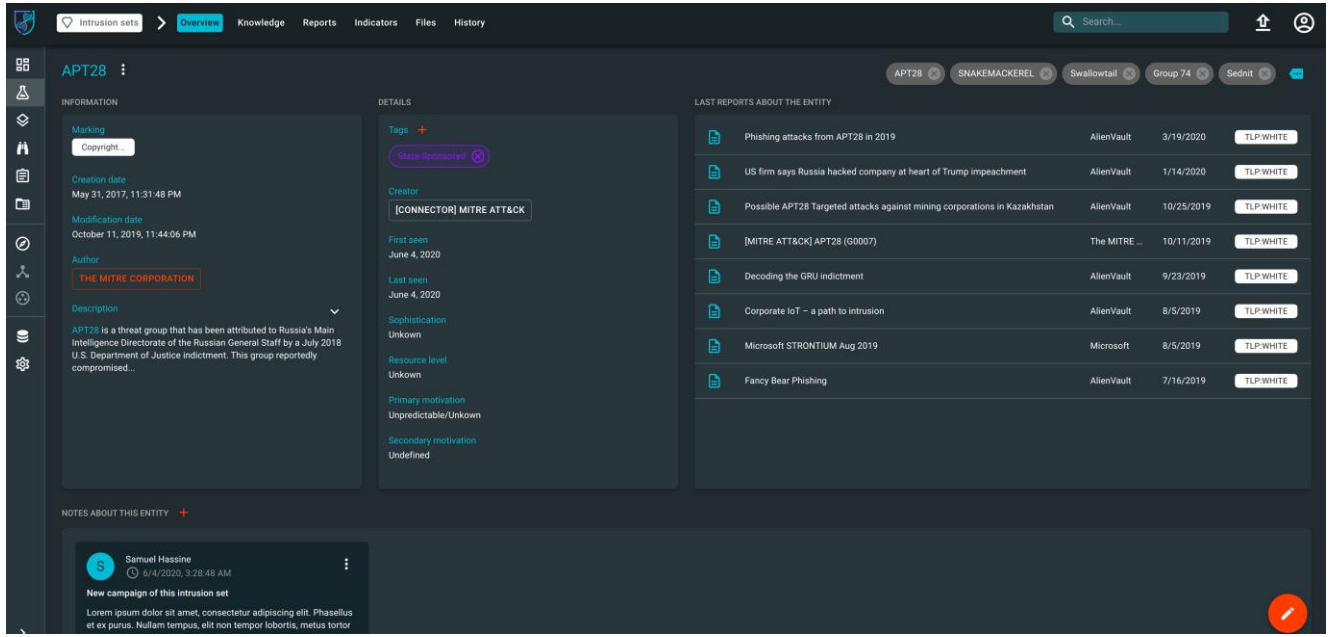
OpenCTI

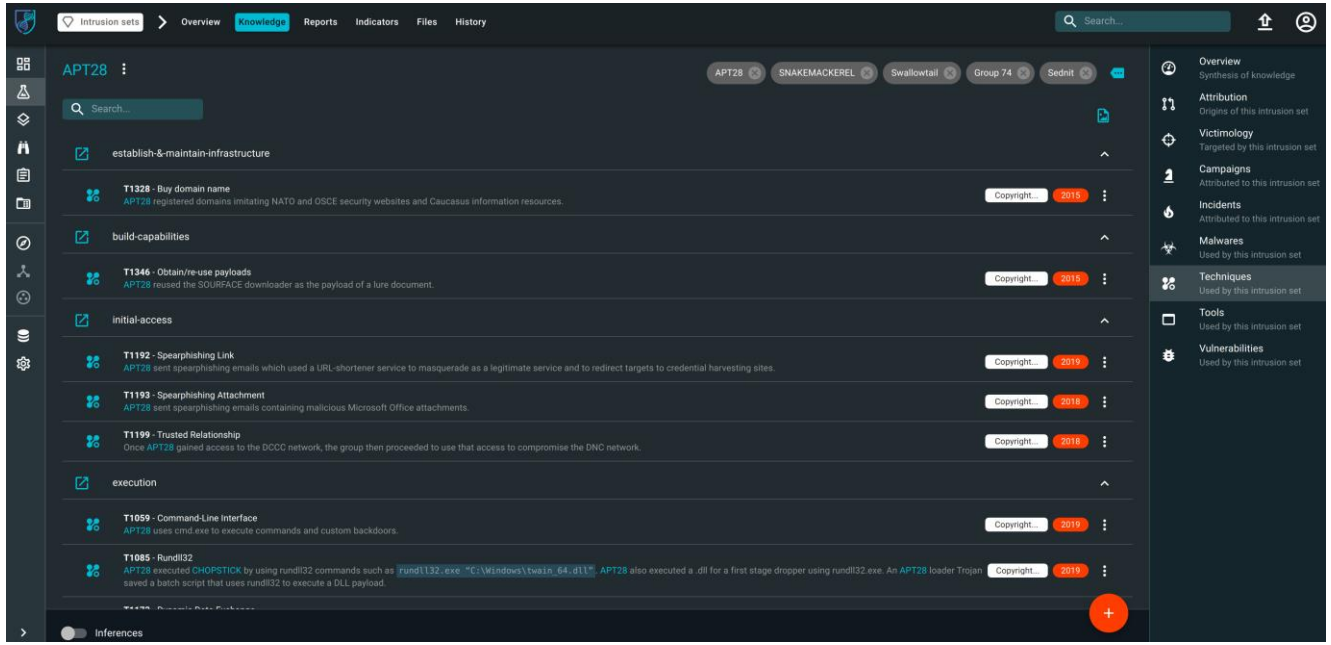
Partnership for Cyber Security Innovation

ABN·AMRO achmea ING de volksbank TNO

OpenCTI

We chose to adopt OpenCTI as our data enrichment platform due to its intuitiveness and to increasing interested that it is gaining from our financial partners.





Partnership for Cyber Security Innovation

ABN·AMRO achmea ING de volksbank TNO

Implementation

1 Feed initial training set + subtechniques and actors info

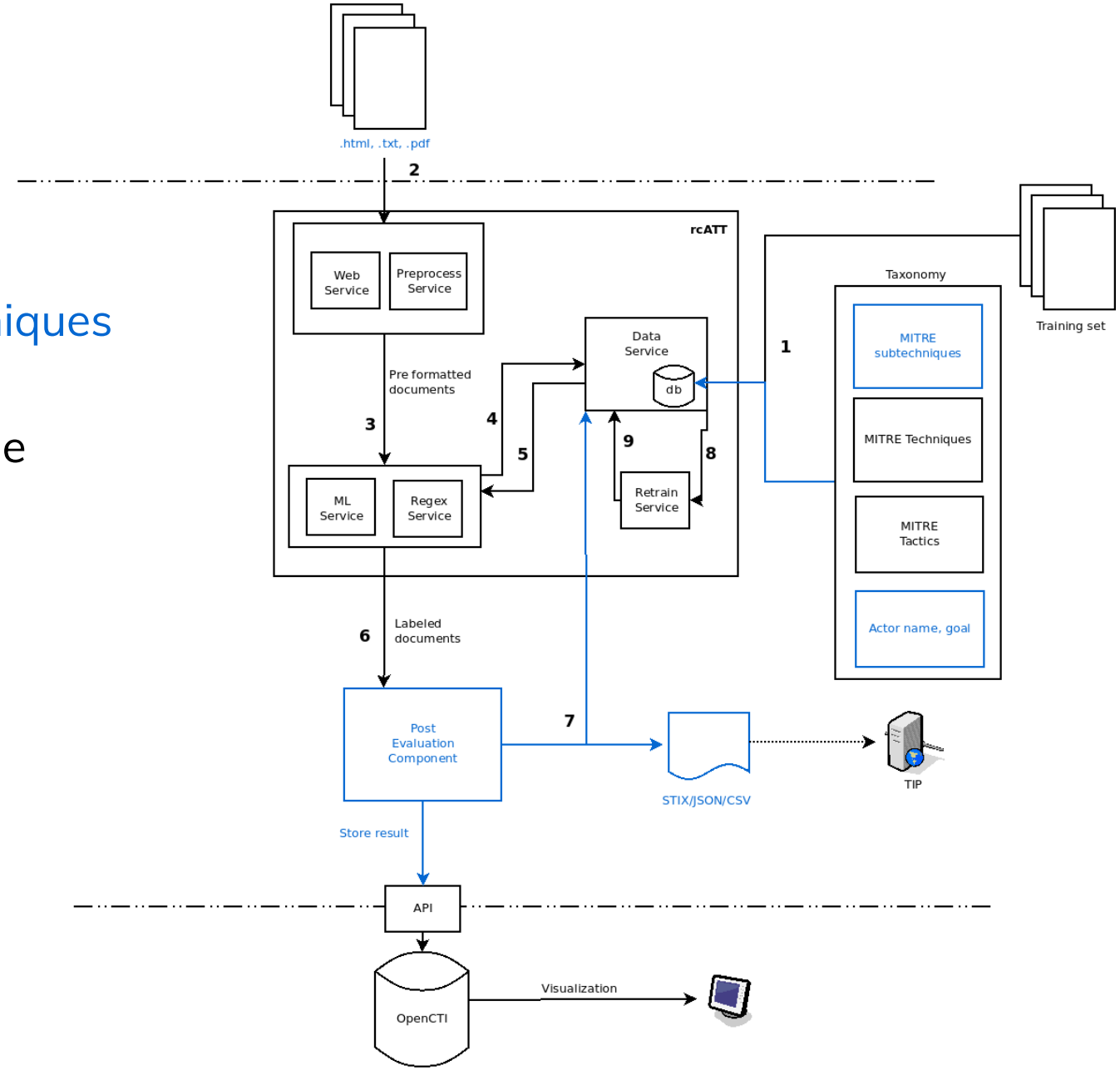
2 - 3 extract text from html, pdf, docx file

4 - 5 techniques, subtechniques, groups prediction through ML + regex algorithms

6 feed TRAM labelled output to PEC

7 store new labelled report in db

8 - 9 use new report to retrain model



Changelog

What has been done?

FUNCTIONALITIES

ATRAPP now offers a lot more functionalities to the user. We made it possible to easily interact with the application to get the most out of it and integrated it with OpenCTI.

- Dockerized application
- URL submission possible
- Change identified actor
- Easily change incorrect results
- Smooth integration with OpenCTI

TEXT PROCESSING

We examined the possibility to identify ATT&CK sub techniques and made it possible in the future. We added actor name extraction functionalities.

- Actor name, motivation and country of origin are extracted
- Submit .pdf or .txt files
- Output is now a STIX2 bundle
- Share data to OpenCTI

USER INTERFACE

A complete overhaul of the design of the application has been finished. A clean user interface with recognizable buttons and text make it easy for the user to navigate.

- Brand new UI
- Actor country flags added
- Feedback on clicks and status
- Required input fields prevent errors

DEMO

TAKE AWAYS

Some take aways

- It is feasible to automatically extract essential adversary attributes from narrative threat reports and feed these into a threat intelligence platform in a common, structured form
- The demonstrated setup seems promising as a support tool for CTI practitioners that automates much of the bulk processing and lets them focus their effort on selected reports of interest
- To realise the full potential of the envisaged support solution, its accuracy will need to be improved through more (and better) training data and processing of analyst feedback (corrections) in operational use

We intend to make the solution available as **open source** and welcome your contribution to its further development!

(webinar registrants will be informed when software is released)

Partnership for Cyber Security Innovation is a collaboration of

ING ABN·AMRO TNO

achmea de volksbank

www.pcsi.nl follow us

This project is co-funded by Holland High Tech with a PPP Grant for Research and Innovation in Top Sector HTSM

Partnership for Cyber Security Innovation is a collaboration of

ING ABN·AMRO TNO

achmea de volksbank

www.pcsi.nl follow us

This project is co-funded by Holland High Tech with a PPP Grant for Research and Innovation in Top Sector HTSM