From Words To Intelligence

NLU and Association Rules for Cyber Threat Analysis

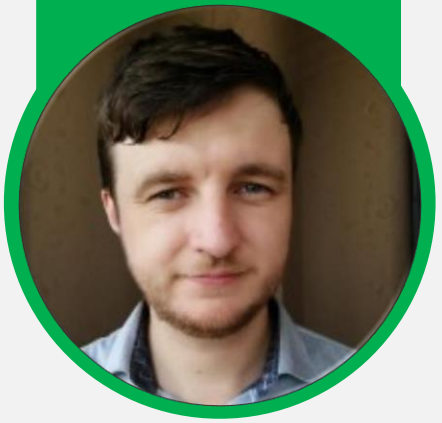
XRATOR
CONQUER YOUR RISK

XRATOR
CONQUER YOUR RISK

XRATOR

Ronan Mouchoux
Threat Intelligence Specialist

XRATOR



François Moerman
Offensive Security Specialist

XRATOR
CONQUER YOUR RISK

What is nice with plan is that …

The Threat Actor

On a mission

With objective clear, skills at hand,
I stand organized like the shifting sand.

The Target

You've things I want

I've delved deep, tracing every line,
Crafting a plan with patience and time.

The Arsenal

Ready for you

Forging weapons of digital might,
to unleash with expert insight.

"We are what we do repeatedly, so excellence is not an act but a habit."
Aristotle

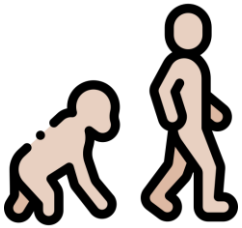
XRATOR CONQUER YOUR RISK

… it can be ~~easily~~ modified.

Obstacles

I didn't see it coming

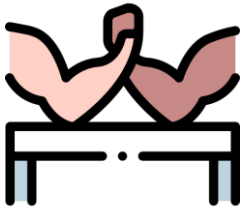
Unforeseen hurdles on my path to conquer,
but relentless I persist and grow stronger.



Adaptation

Twist this and that

Obstacles arise, go back at the source,
find compatible techniques to stay the course.



Wrestling

Constantly adapt

Adapting their ways to breach defenses,
threat actors proceed with their preferences.

"The great art is to change during the battle. Woe betide the general who comes into battle with a system."
Napoléon

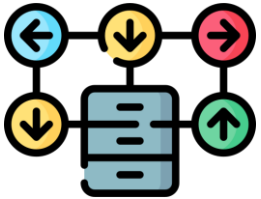
XRATOR CONQUER YOUR RISK

From Traces to Adversary's Actions

Abduction

Solving Digital Crimes

Forensic & IR must interpret digital traces with a partial view of opponent's actions.



Ambiguity

Interpretation & Vocabulary

Two people may look at the same artefact and link it to different MITRE ATT&CK® techniques.

Reproductibility

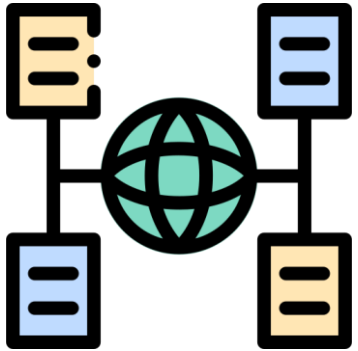
Automation & Explainability

Explainable automation enable scaling and consistence into ATT&CK mapping.

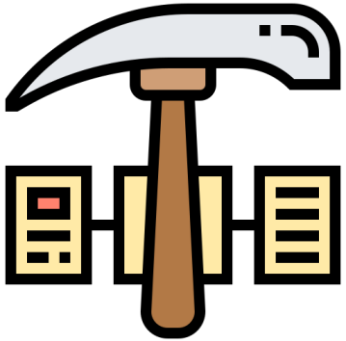
"Without adequate contextual technical details to sufficiently describe and add insight into an adversary behavior, there is little value to ATT&CK mapping." Best Practices for MITRE ATT&CK® Mapping, CISA, June 2021.

XRATOR
CONQUER YOUR RISK

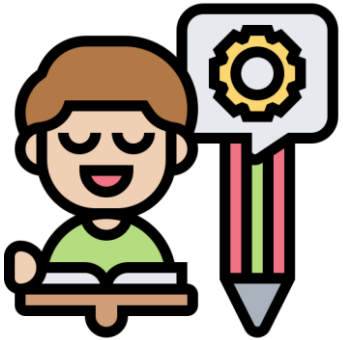
Agenda



Data Collection And Preparation



Association Rules Mining



Association Rules Qualitative Review

For more information, read the full paper :

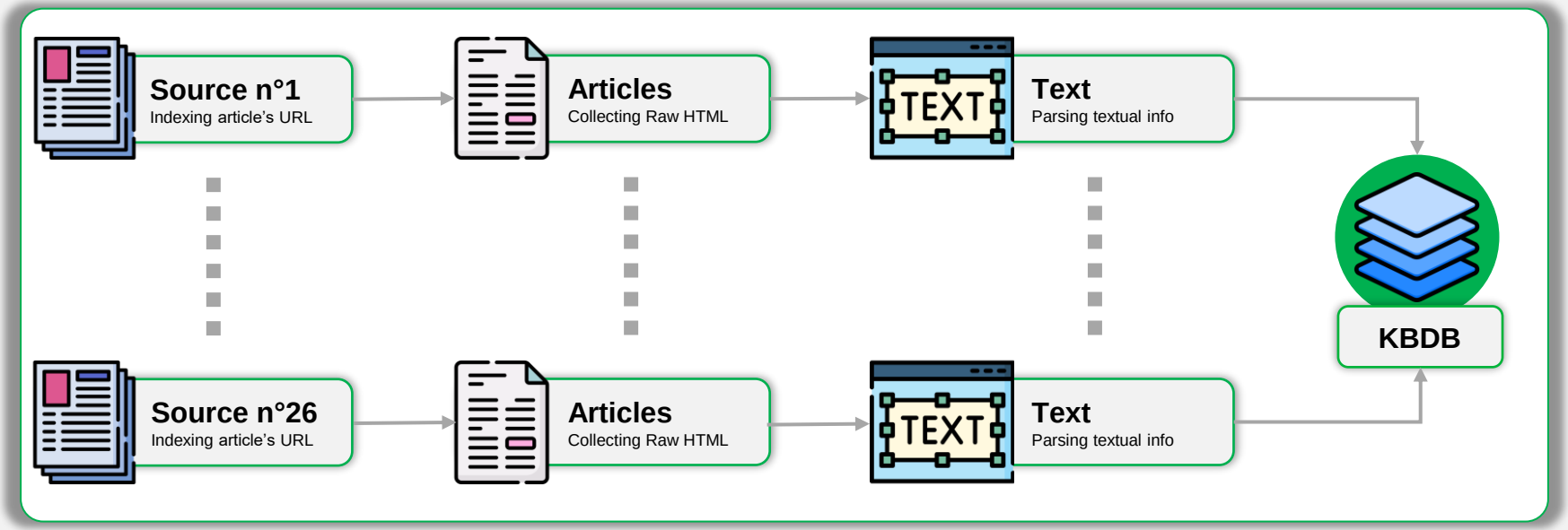
« *From Words to Intelligence: Leveraging the Cyber Operation Constraint Principle, Natural Language Understanding, and Association Rules for Cyber Threat Analysis*»

XRATOR
CONQUER YOUR RISK

Collect & Prepare

XRATOR CONQUER YOUR RISK

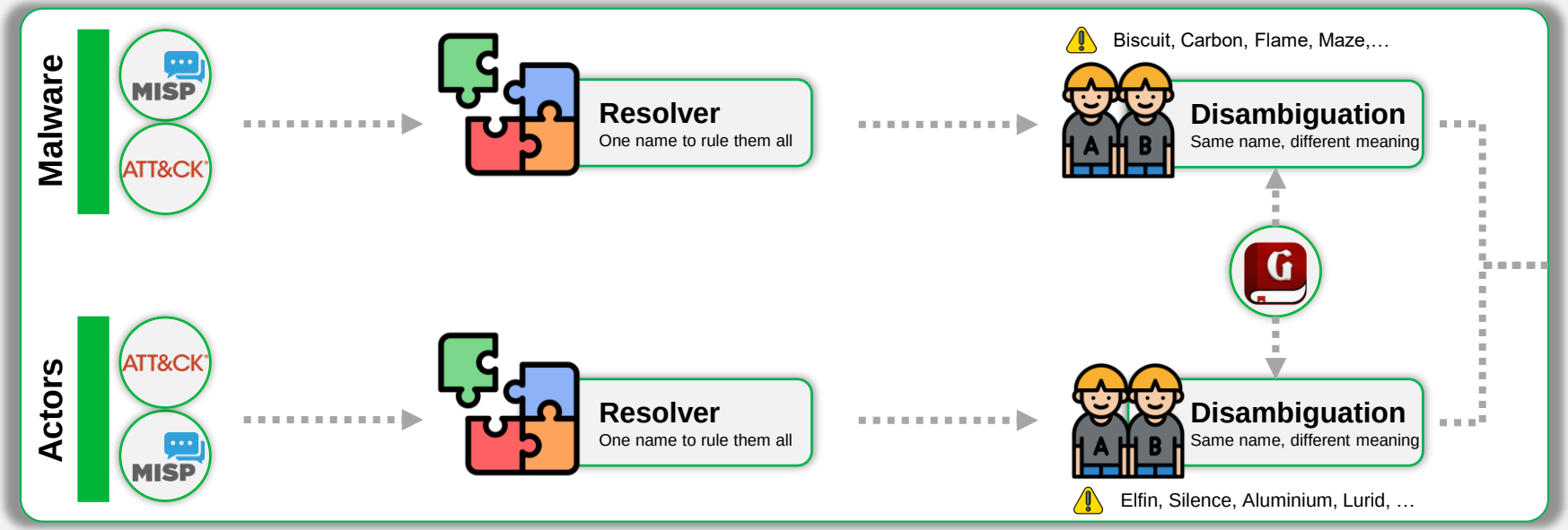
Extracting Article from HTML



The serial process allows to adapt to each source's data structure and search for specific elements in predictable places.
This is inspired by success stories of applying NLP to medical research literature.

XRATOR CONQUER YOUR RISK

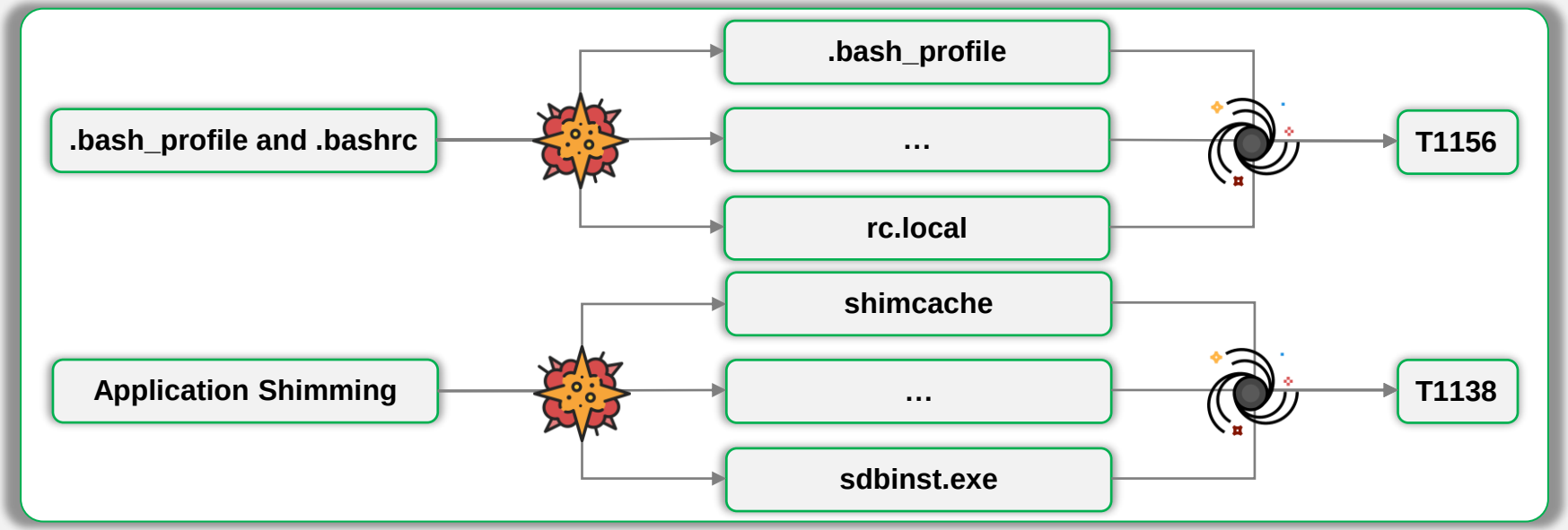
Malware & Threat Actor References



Malware and Threat Actor held several denominations. We need a key to aggerate their references across aliases and articles.
We also need to deal with denominations that collide with everyday language or between each other.

XRATOR CONQUER YOUR RISK

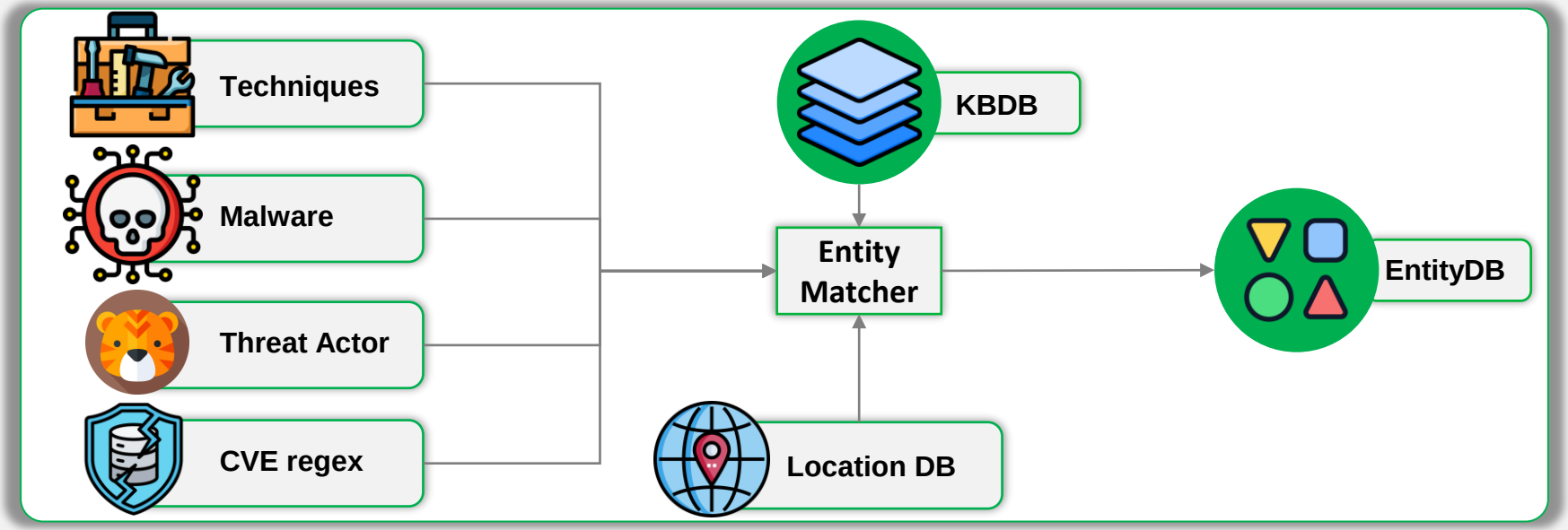
Techniques Bang & Crunch



MITRE ATT&CK (v6.3) techniques' descriptions and procedures are not enough to match or train a model.
We augment the data with synonyms.

XRATOR CONQUER YOUR RISK

Pattern Matching



Using the references' data, we match them against the textual database. Instead of machine learning, this method reduce the number of errors, enhance explainability, reproducibility and facilitate diagnosis. The results are stored in the EntityDB.

XRATOR CONQUER YOUR RISK

Pattern Matching

Article_ID	Threat Actor	Technique	Malware	…
A	/	T1, T2, T3, …	M1	…
B	TA1	T2, T5, T6, …	M2, M3	…
C	TA2	T1, T3, T7, …	/	…
D	TA1, TA3, …	/	M1, M4	…

The result is a sparse array where each article is a transaction.
We can use it as is for graph analysis or transform it into a dense array for statistical approach.

XRATOR
CONQUER YOUR RISK

Apriori Mining

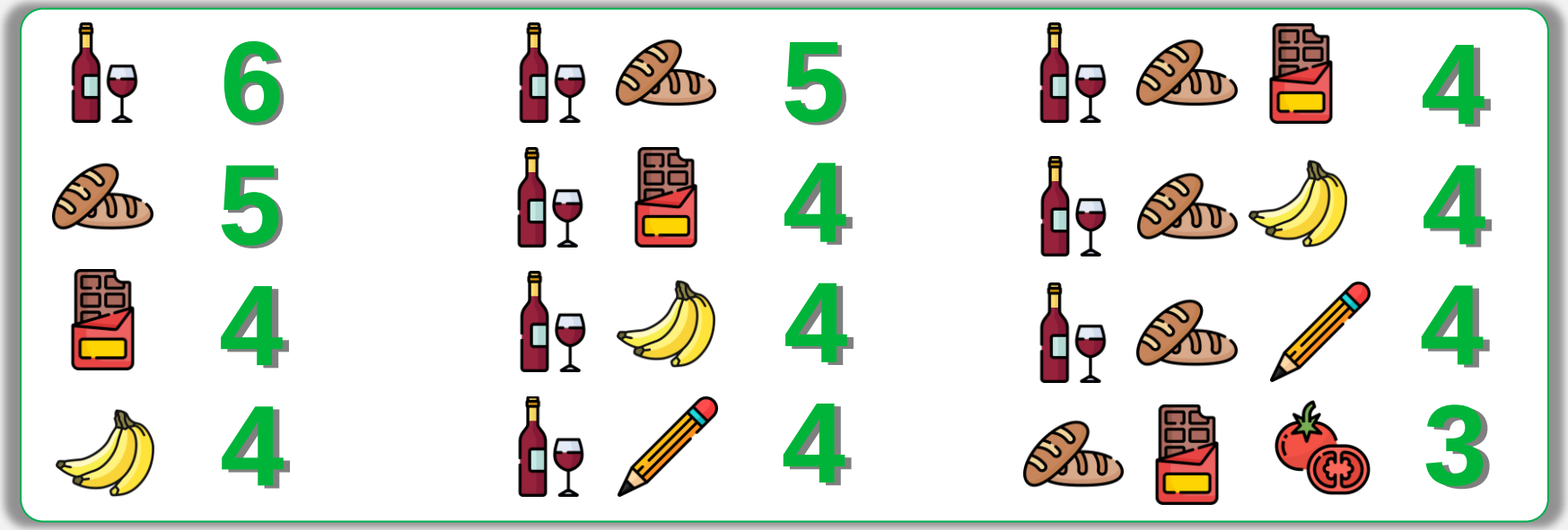
The Grocery Store Example

Transaction	Basket
A	
B	
C	
D	
E	
F	

Next time you go to the grocery store, watch for your buying habits. Search for those moments where you pick something and you say :"How, I need this too, this will be perfect with that".

XRATOR CONQUER YOUR RISK

Top Frequent Itemsets



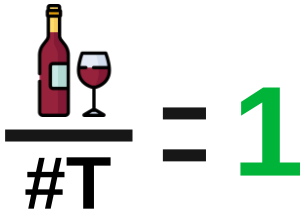
The goal is to uncover hidden patterns. The top frequencies of n-itemsets are generally obvious or just the “association by chance” of very common items. Those combinations must be examined with metrics.

XRATOR CONQUER YOUR RISK

Apriori metrics

Support

Frequency of the antecedent technique



Confidence

How likely there is the consequent technique if you have the antecedent technique

$$\frac{\text{Support}\{\text{🍷}, \text{🥖}\}}{\text{Support}\{\text{🍷}\}} = 0{,}8$$

Lift

Popularity malus

$$\frac{\text{Support}\{\text{🍷}, \text{🥖}\}}{\text{S}\{\text{🍷}\} * \text{S}\{\text{🥖}\}} = 1$$

Support {✏} = 0,67

Confidence {✏ → 📼} = 0,67

Lift {✏ → 📼} = 1,51

The goal is to uncover hidden patterns. The top frequencies of n-itemsets are generally obvious or just the “association by chance” of very common items.

XRATOR CONQUER YOUR RISK

In context

12 808
articles with one threat actor and at least two techniques.

901
Association rules with lift > 1.

1 434
Association rules spread over 73 unique threat actors.

5’51
Duration of the full process.

Winning Combo *(full dataset)*


Command Line Interface → Network Connections Discovery


Virtualization/Sandbox Evasion → Process Hollowing


Command Line Interface → Hidden Window

The test ran on a 2007-2020 dataset of 17 153 articles . We use 2-itemsets to increase interpretability.
Server: Intel W3520 - 4c/8t 2.66GHz - 32GB DDR3 ECC 1333 MHz

XRATOR
CONQUER YOUR RISK

Qualitative Review

XRATOR CONQUER YOUR RISK

APT28's top association rules

Lift	Antecedent technique	Consequent technique	Observation
5,44	Custom Cryptographic Protocol (T1024)	Logon Scripts (T1037)	Persistence with antivirus evasion (packing). Or encrypted data exfiltration
5,23	Modify Registry (T1112)	Registry Run Keys / Startup Folder (T1060)	Two-stage persistence.
4,59	Rundll32 (T1085)	Logon Scripts (T1037)	This is an implementation of the first stage of a two-stage persistence operation
4,59	Software Packing (T1045)	Windows Management Instrumentation (T1047)	Internal delivery via WMI of a packed malicious payload to avoid detection, for privileged escalation or persistence.
3,5	Process Discovery (T1057)	Peripheral Device Discovery (T1120)	Common task of a host reconnaissance and monitoring operation.

XRATOR CONQUER YOUR RISK

SANDWORM's top association rules

Lift	Antecedent technique	Consequent technique	Observation
4,59	Drive-by Compromise (T1189)	Man in the Browser (T1185)	Watering hole or communication interception plus payload injection (PRISM-like)
4,59	Remote Access Tools (T1219)	Input Capture (T1417)	Keylogger
4,59	Clipboard Data (T1115)	Input Capture (T1417)	Monitor some input (like CTRL+C) to trigger the inspection of the content of the clipboard.
3,94	External Remote Services (T1133)	Remote Services (T1021)	Using valid account on a VPN, RDP, or external accessible services.
3,94	System Information Discovery (T1426)	System Firmware (T1019)	Information gathering for persistence or privilege escalation (Rootkit-like).

XRATOR CONQUER YOUR RISK

EQUATION's top association rules

Lift	Antecedent technique	Consequent technique	Observation
3,74	**Input Capture (T1417)**	**Clipboard Data (T1115)**	Monitor some input (like CTRL+C) to trigger the inspection of the content of the clipboard.
3,59	**Remote Access Tools (T1219)**	**File and Directory Discovery (T1420)**	The attacker is using the RAT to performs its host reconnaissance.
3,19	**Credential Dumping (T1003)**	**Clipboard Data (T1115)**	Stole credential that pass through the clipboard.
3,16	**Software Packing (T1045)**	**Execution through Module Load (T1129)**	First phase of a two-stage persistence operation.
2,99	**Execution through API (T1106)**	**Execution through Module Load (T1129)**	Defense evasion to load external payload using a low-level Windows API (e.g.: "CreateProcessA").

XRATOR
CONQUER YOUR RISK

Conclusion

XRATOR CONQUER YOUR RISK

Association Rules Signatures

Threat Actor	SAND WORM	EQUATION	TURLA	LAZARUS
APT28	0.08/0.17	0.17/0.11	0.15/0.08	0.12/0.15
SAND WORM	x	0.05/0.08	0.06/0.05	0.09/0.07
EQUATION	x	x	0.05/0.08	0.07/0.06
TURLA	x	x	x	0.12/0.09

Jaccard and Sørensen-Dice similarities metrics displays very low overlaps between top threat actors association rules.

Winning Sharings *(group dataset)*

- Deobfuscate/Decode Files → Software Packing
- Obfuscated Files or Information → Deobfuscate/Decode Files
- Software Packing → Deobfuscate/Decode Files
- System Information Discovery → Process Discovery
- Obfuscated Files or Information→ Data Compressed

~75% of association rules are unique to a Threat Actor.

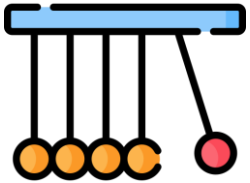
XRATOR CONQUER YOUR RISK

Lessons learned

The Threat Actor

Beyond the Kill Chain

Modus Operandi is indeed a differential factor among sophisticated actors.



Habits & Preferences

403 Forbidden

We can't tell apart which techniques is the preferred one.

Collection

Feed it with love

The best AI model can't produce any value without careful feeding.

Diachrony and synchrony linguistics analysis is the key challenge. As modern conflicts are partly conducted through zero and one, we are talking about the ability for future generation to recollect and write *The History*.

XRATOR
CONQUER YOUR RISK

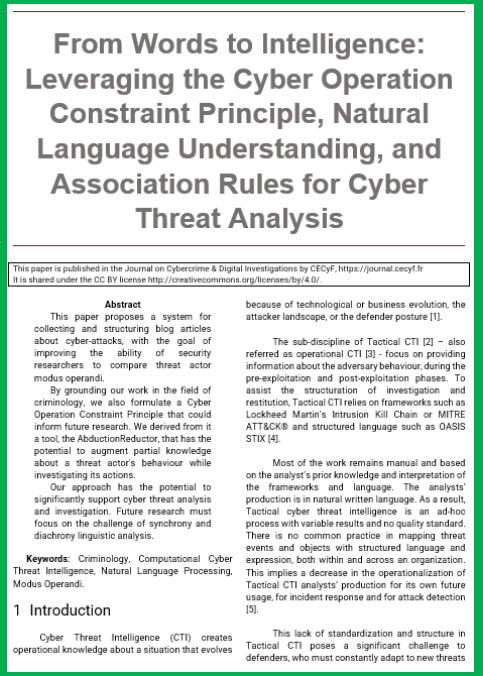
Keep hunting!

François MOERMAN
Chief Executive Officer

francois@x-rator.com

Ronan MOUCHOUX
Chief Product & Engineering Officer

ronan@x-rator.com



From Words to Intelligence: Leveraging the Cyber Operation Constraint Principle, Natural Language Understanding, and Association Rules for Cyber Threat Analysis

This paper is published in the Journal on Cybercrime & Digital Investigations by CECyF, https://journal.cecyf.fr
It is shared under the CC BY license http://creativecommons.org/licenses/by/4.0/.

Abstract

This paper proposes a system for collecting and structuring blog articles about cyber-attacks, with the goal of improving the ability of security researchers to compare threat actor modus operandi.

By grounding our work in the field of criminology, we also formulate a Cyber Operation Constraint Principle that could inform future research. We derived from it a tool, the AbductionReductor, that has the potential to augment partial knowledge about a threat actor's behaviour while investigating its actions.

Our approach has the potential to significantly support cyber threat analysis and investigation. Future research must focus on the challenge of synchrony and diachrony linguistic analysis.

Keywords: Criminology, Computational Cyber Threat Intelligence, Natural Language Processing, Modus Operandi.

1 Introduction

Cyber Threat Intelligence (CTI) creates operational knowledge about a situation that evolves because of technological or business evolution, the attacker landscape, or the defender posture [1].

The sub-discipline of Tactical CTI [2] – also referred as operational CTI [3] - focus on providing information about the adversary behaviour, during the pre-exploitation and post-exploitation phases. To assist the structuration of investigation and restitution, Tactical CTI relies on frameworks such as Lockheed Martin's Intrusion Kill Chain or MITRE ATT&CK® and structured language such as OASIS STIX [4].

Most of the work remains manual and based on the analyst's prior knowledge and interpretation of the frameworks and language. The analysts' production is in natural written language. As a result, Tactical cyber threat intelligence is an ad-hoc process with variable results and no quality standard. There is no common practice in mapping threat events and objects with structured language and expression, both within and across an organization. This implies a decrease in the operationalization of Tactical CTI analysts' production for its own future usage, for incident response and for attack detection [5].

This lack of standardization and structure in Tactical CTI poses a significant challenge to defenders, who must constantly adapt to new threats

From Words to Intelligence

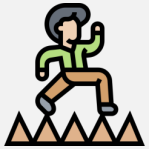
In 1915, Fosdick wants to go beyond the Bertillon system to augment crime detection. In 1996, borrowing the "Script" concept from cognitive science, Cornish argues that the knowledge about the procedural aspects and procedural requirements of crime commission has the potential to enhance situational crime prevention.

The Creative Dodger

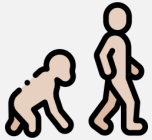
With objective clear, skills at hand,
I stand organized like the shifting sand.

I've delved deep, tracing every line,
Crafting a plan with patience and time.

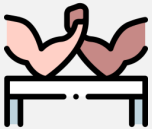
Forging weapons of digital might,
to unleash with expert insight.



Unforeseen hurdles on my path to conquer,
but relentless I persist and grow stronger.



Obstacles arise, go back at the source,
find compatible techniques to stay the course.



Adapting their ways to breach defenses,
threat actors proceed with their preferences.