EXTRACTOR: Extracting Attack Behavior Graphs from Threat Reports

Kiavash Satvat

Rigel Gjomemo

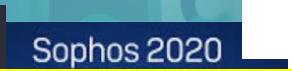
V.N. Venkatakrishnan

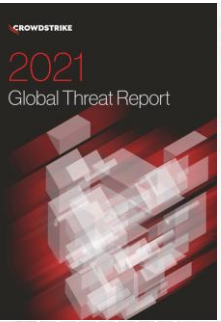
ksatva2, rgjome1, venkat@uic.edu

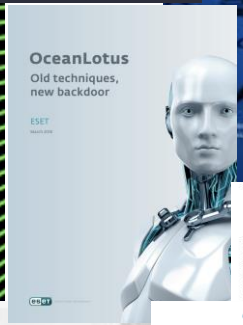
University of Illinois at Chicago

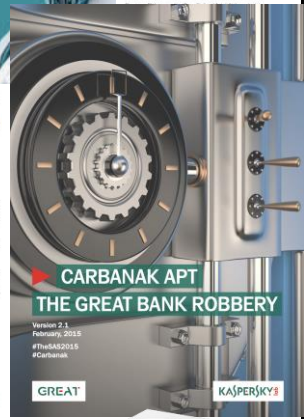
Cyber Threat Intelligence Reports

A significant amount of knowledge is available in Cyber Threat Intelligence (CTI) reports.









Human-intensive Effort for Creation of Detection Rules

CTI Reports

Intrusion Detection Systems

Hillstone NETWORKS
TREND MICRO
TippingPoint
Suricata
Open Source IDS / IPS / NSM engine
zeek
McAfee
SECURITY ONION
FIREEYE
OSSEC
VECTRA

Wouldn't it be great to automate and scale?

What is being done automatically?

- Search for fragmented Indicators of Compromise (IOC)
  - Hash values, file/process names, IP addresses, domain names

Limitation:

- Updated or re-purposed attacks and malware polymorphism
- Use of legitimate-looking names (like svchost in Windows)
- Easy for the attacker to mutate and evade detection systems!

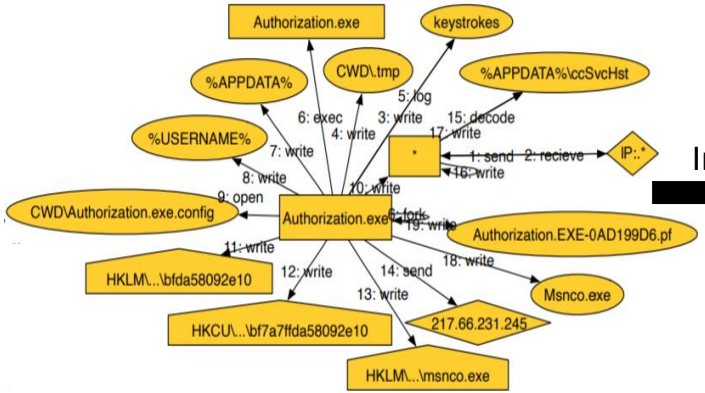
What if we can learn more than just isolated IOCs?
Something which is harder to evade!...

Problem Statement

- Extract **actionable** provenance graphs of attack **behavior** from natural language CTI reports
  - **Actionable**: Provenance graphs can be directly used to perform threat hunting
  - **Behavior**: connected events and entities, not single IOCs

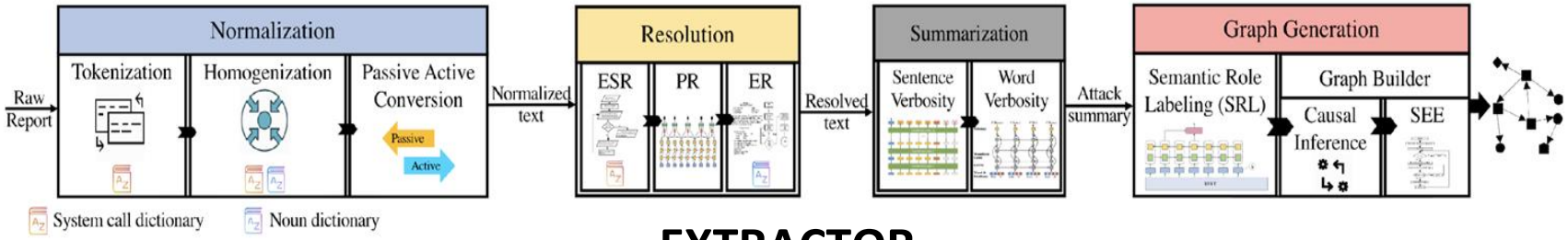
The malware connects to the Command & Control (CnC) server.
The "Authorization.exe" malware has keylogger functionality.
It stores the logged keystrokes in the following file: [CWD]\.tmp
When the "Authorization.exe" malware is executed it :
Creates a copy of itself in the following locations: %APPDATA% %USERNAME%
Tries to open the following file: [CWD]\Authorization.exe.config
Entrenches in the system for persistence in the following registry locations:
HKCU\...\bf7a7ffda58092e10 HKLM\...\bfda58092e10
Beacons to the following C2 node IP:.* over TCP port 1177:"217.66.231.245'
Makes the following modification to the registry to bypass the Windows Firewall:
HKLM\...\msnco.exe
The downloaded file is decoded, written to disk as %APPDATA%\...\ccSvcHst
The following files created when the Authorization.exe malware executed: msnco.exe
authorization.EXE-0AD199D6.pf
Msnco.exe and Authorization.EXE-0AD199D6.pf are created by Authorization.exe.

Extractor



Input

Challenges and Approach



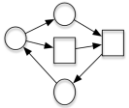
EXTRACTOR

- CTI Language Complexities
  - Domain specific vocabulary
  - Ellipsis subjects and objects
  - Pronoun
  - Passive vs active
- Verbosity
  - Inter vs Intra verbosity
- Relationships Extraction (Subject, Verb, Object)
  - Causality and flow of attack

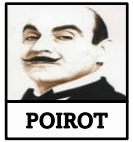
Evaluation

1.

6 public CTI reports



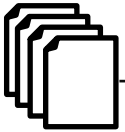
Precision	Recall	F1-Score
0.89	0.94	0.92



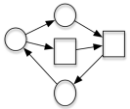
Threat Hunting Tool

Detection	False Positive
✓	0

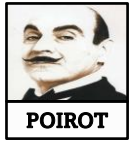
2.



8 DARPA Transparent Computing Engagement 2



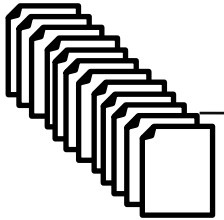
Precision	Recall	F1-Score
0.96	0.93	0.94



Threat Hunting Tool

Detection	False Positive
✓	0

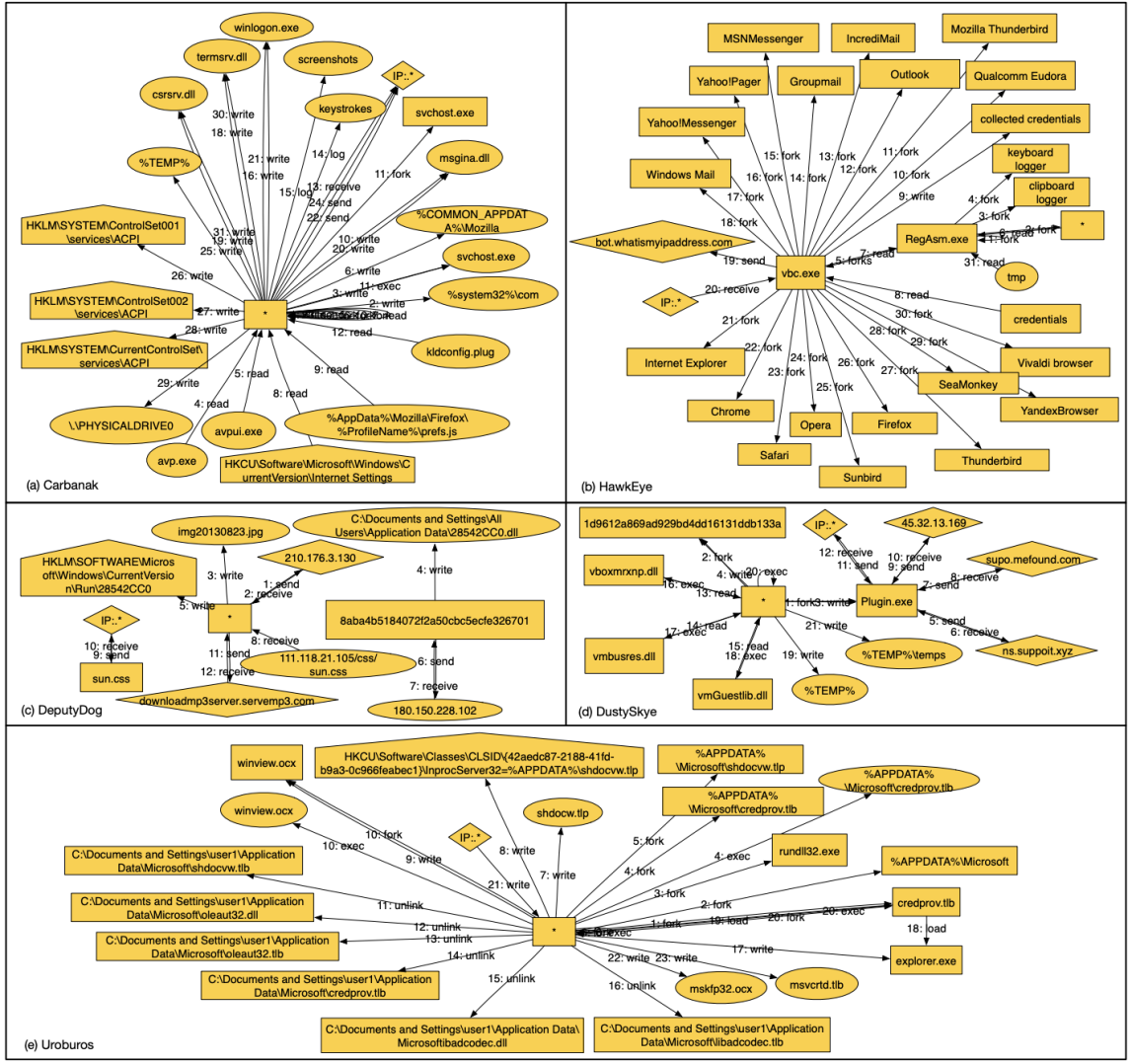
3.



Large-Scale Evaluation

Technical description VS solution section

	# reports	Similarity Score
Microsoft	4020	91%
TrendMicro	11600	85%



Sample Attack Behavior Graphs from Public CTI Reports

Source code:

https://github.com/ksatvat/EXTRACTOR

Questions?