MWS 2021
Augma 2021 Dataset

nao_sec

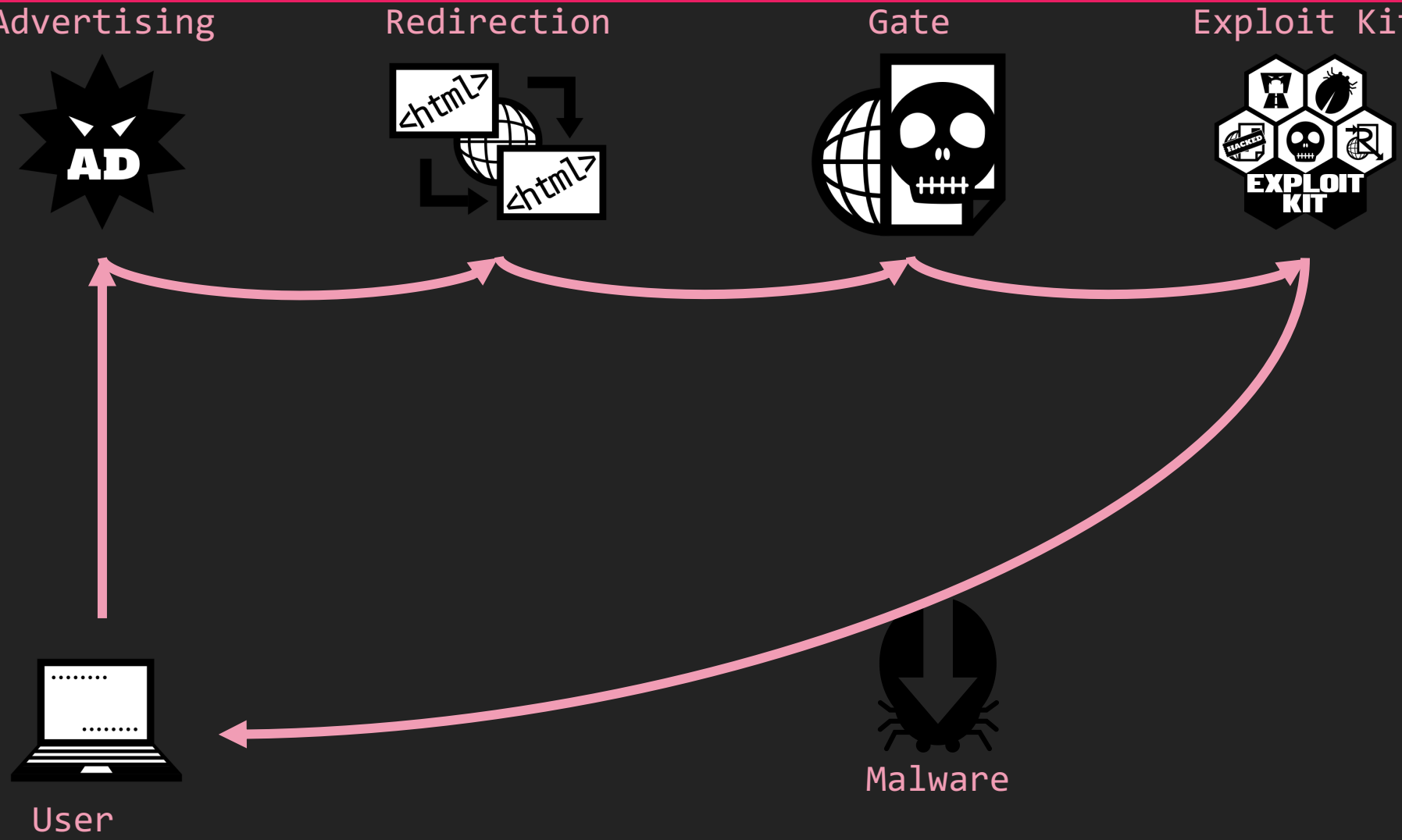
Yosuke Chubachi - Active Defense Institute, LTD.

Rintaro Koike - NTT Security (Japan) KK

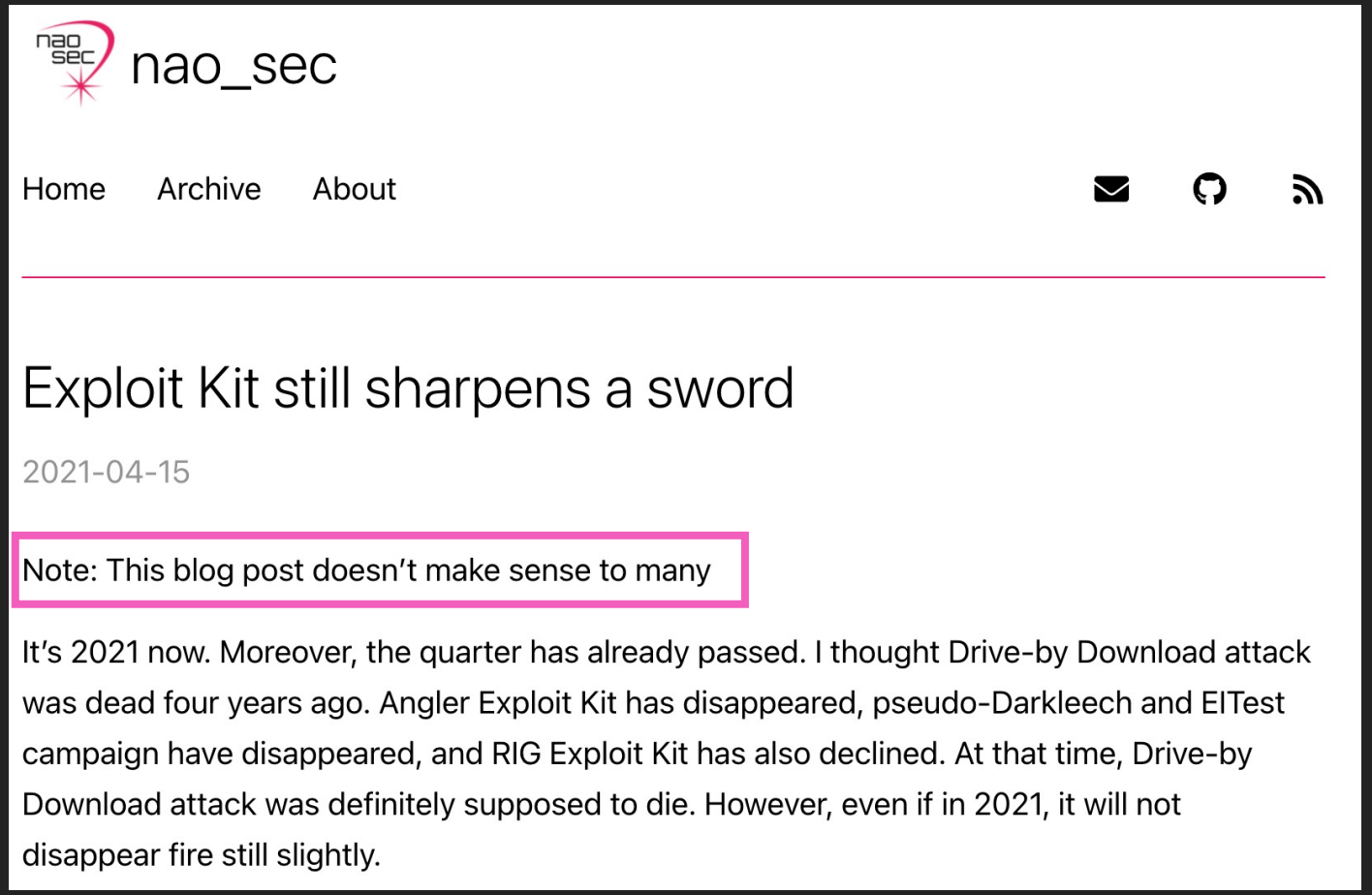
Shota Nakajima - Cyber Defense Institute, Inc.

Introduction

Malvertising – Drive by Download



Exploit Kit still sharpens a sword



nao_sec

Home Archive About

Exploit Kit still sharpens a sword

2021-04-15

Note: This blog post doesn't make sense to many

It's 2021 now. Moreover, the quarter has already passed. I thought Drive-by Download attack was dead four years ago. Angler Exploit Kit has disappeared, pseudo-Darkleech and EITest campaign have disappeared, and RIG Exploit Kit has also declined. At that time, Drive-by Download attack was definitely supposed to die. However, even if in 2021, it will not disappear fire still slightly.

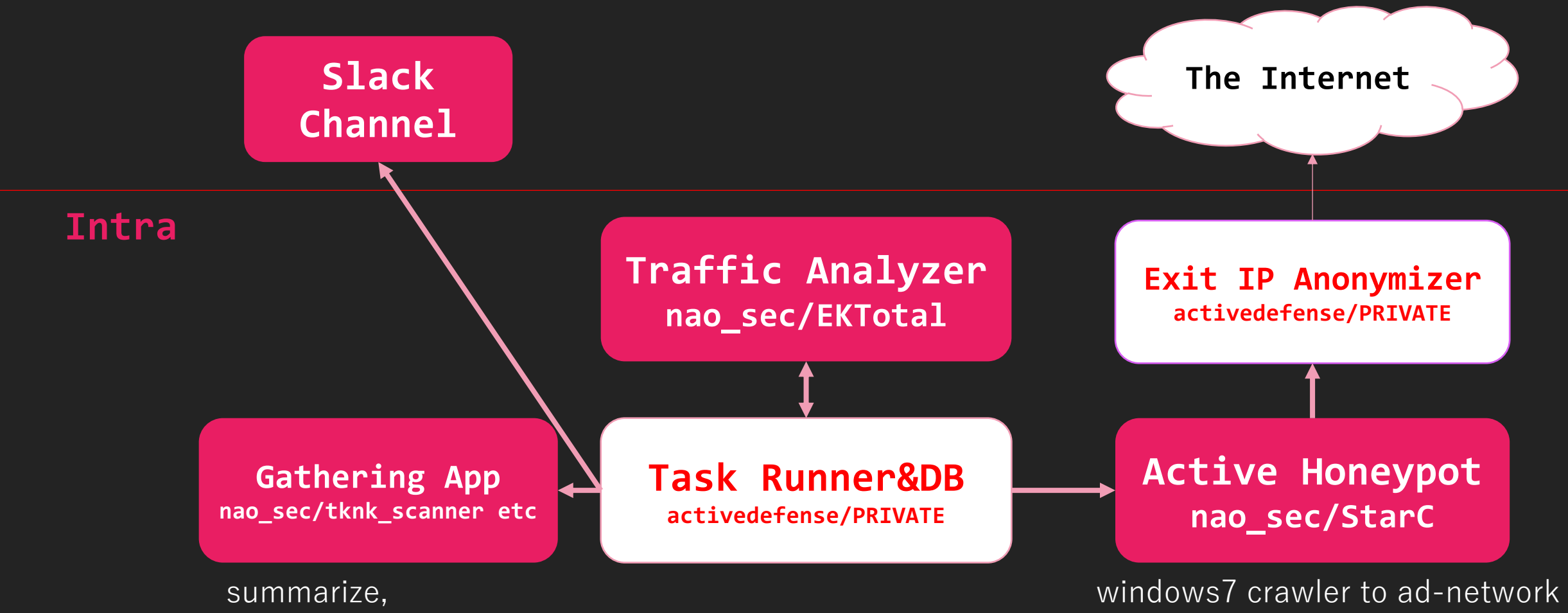
Motivation of Development ”Augma“

- **A drive-by threat is still “ACTIVE”**
  - Many attack campaigns and EKs have appeared
- **Manual drive-by observation is too hard**
- **We want to research the latest threat trends automatically**
  - Active Observation + Analysis + Extraction

Challenges of Exploit Kit Crawling

- **Anti-Cloaking**
  - EK and malware distribution infrastructure BAN specific IP address and range
  - Example, TrendMicro, Symantec and public cloud IP range is BANNED by RIG EK
- **Behaving:**
  - Evading checks of Ad-networks
- **Chasing:**
  - Crawling target selection is difficult
- **Accuracy:**
  - Need reliable detection rules

Augma Overview



Active Honeypot (StarC)

- **Simple high-interactive client honeypot**
  - https://github.com/nao-sec/starc

  - Input a URL, StarC access and collect data
    - Traffic data (pcap & saz)
    - Screenshot
    - Temp directory files

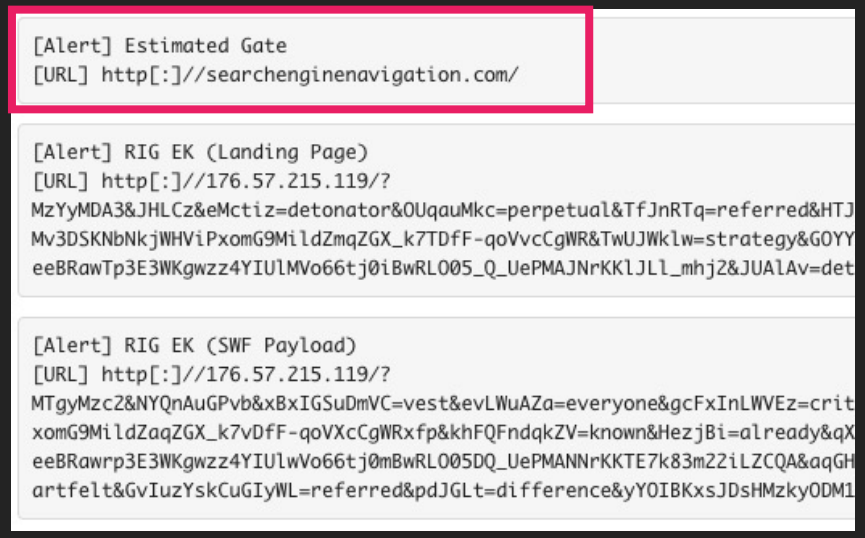
Traffic Analyzer (EKTotal)

- **Automatic DbD traffic analyzer**
  - https://github.com/nao-sec/ektotal

  - Input a pcap or saz, EKTotal analyze traffic data
    - Identify campaign & EK
    - Extract some information
      - Encode key
      - CVE Number
      - SWF file
      - Malware
    - Detecting with EKFiddle’s rules and Augma custom rules
      - https://github.com/malwareinfosec/EKFiddle
    - Lazy “Gate Estimation” added on July, 2019

Lazy “Gate Estimation”

- **Gate**
  - We call ”Gate” that campaign specific redirect server
  - EKTotal can estimate Gate
  - This function helps identify and categorize campaigns
  - Some campaigns NOT go through Gate



Augma 2021 Dataset Overview

Augma 2021 Dataset

- **Period: 2020/05-2021/04**
- **Data:**
  - Captured malicious traffic as Fiddler saz format and pcapng format.
  - ek including exploit kit traffic
  - tss including tech support scam traffic
- **Files and Directories:**
  - augma2021/augma2021dataset.description.en.txt
  - augma2021/augma2021dataset.description.ja.txt
  - augma2021/ek/ek.metadata.tsv
  - augma2021/ek/samples/{*.pcap and * .saz}
  - augma2021/tss/tss.metadata.tsv
  - augma2021/tss/samples/{*.pcap and * .saz}

Statistics – System

Period: 2020-05-01 00:05:01+00:00-2021-04-30 23:55:01+00:00

Total patrol: 95333, hit: 15342, total engage rate: 0.161

Statistics – dataset files

All threats potentially target to Japan
(Augma uses JP related IP Addresses)

(Unit: file)

pcap: 10667		
- ek:	10667	
- tss:	4682	
saz: 10008		
- ek:	10667	
- tss:	4682	

cf). Augma Dataset 2020

pcap: 10020	
- ek:	7418
- tss:	2544
saz: 10008	
- ek:	7419
- tss:	2544

Summary

date	total	Bottle EK	TSS (audio)	PurpleFox EK	RIG EK	PseudoGate (EK Redirection)	Spelevo EK	Underminer EK	Capesand EK	Fallout EK	TSS (evil cursor)	TSS (Browlock gen)	Fallout EK	CVE-2018-8174
May-20	804	0	663	0	115	0	0	0	0	0	22	2	1	1
Jun-20	1484	0	1028	0	165	101	0	89	54	40	0	7	0	0
Jul-20	1417	56	1153	18	24	76	0	30	56	4	0	0	0	0
Aug-20	1368	1234	133	0	1	0	0	0	0	0	0	0	0	0
Sep-20	2089	1725	338	20	2	0	0	0	0	0	0	0	4	0
Oct-20	4011	3599	390	0	0	0	0	0	0	22	0	0	0	0
Nov-20	2187	2098	55	34	0	0	0	0	0	0	0	0	0	0
Dec-20	136	0	95	31	1	9	0	0	0	0	0	0	0	0
Jan-21	584	0	224	360	0	0	0	0	0	0	0	0	0	0
Feb-21	483	0	312	150	0	19	2	0	0	0	0	0	0	0
Mar-21	180	0	154	0	0	0	26	0	0	0	0	0	0	0
Apr-21	282	5	96	1	0	0	147	33	0	0	0	0	0	0

Example (metadata)

TSV(Tab Separated Values):
1 line per 1 crawl

date:2019/05/01 00:10:02
saz:ek/samples/2019-05-01_00-10-02.saz
pcap:ek/samples/2019-05-01_00-10-02.pcap
name:Underminer EK
mal_url:http://27.122.57.192:9081/index.php?ad_id={snipped}
name:Underminer EK
mal_url:http://27.122.57.192:9081/js/bhfsbqqpvs9nr61tnqqou0lr8g.js

- Identifier derives from EKFiddle's rule and Augma custom rule
- Some rules define another identifier on same EK detection like GrandSoft(Checker) and GrandSoft(Landing)

Attention

- You SHOULD refer to published article[7] when you publish anything with this dataset
- You MUST pay attention to use IP, URL and payloads in this dataset. It is potentially malicious

[7] Rintaro Koike and Yosuke Chubachi, "Finding drive-by rookies using an automated active observation platform", VirusBulletin 2019, Oct. 2019.