MWS 2020
Augma 2020 Dataset

nao_sec

Yosuke Chubachi - Active Defense Institute, LTD.

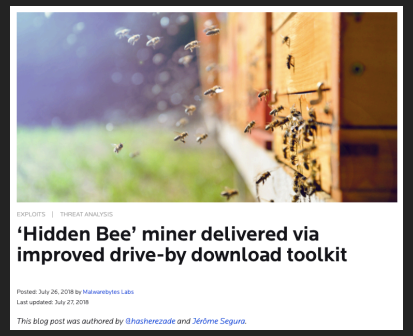
Rintaro Koike - NTT Security (Japan) KK

Shota Nakajima - Cyber Defense Institute, Inc.

Introduction

DbD Threat Landscape

2018-07
Underminer



2018-08
Fallout

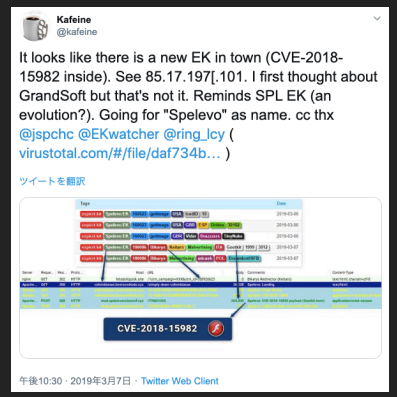
Hello "Fallout Exploit Kit"

2018-09-01

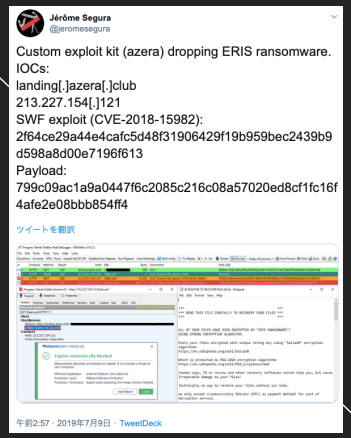
First

At the end of August 2018, we observed a new Exploit Kit. Its behavior (code generation using html) and URL pattern are similar to Nuclear Pack Exploit Kit. Therefore we named it "Fallout Exploit Kit". Fallout Exploit Kit is using CVE-2018-4878 and CVE-2018-8174. That code is distinctive and interesting.

2019-03
Spelevo



2019-07
azera



Weak Drive-by Download attack with "Radio Exploit Kit"

2019-07-15

First

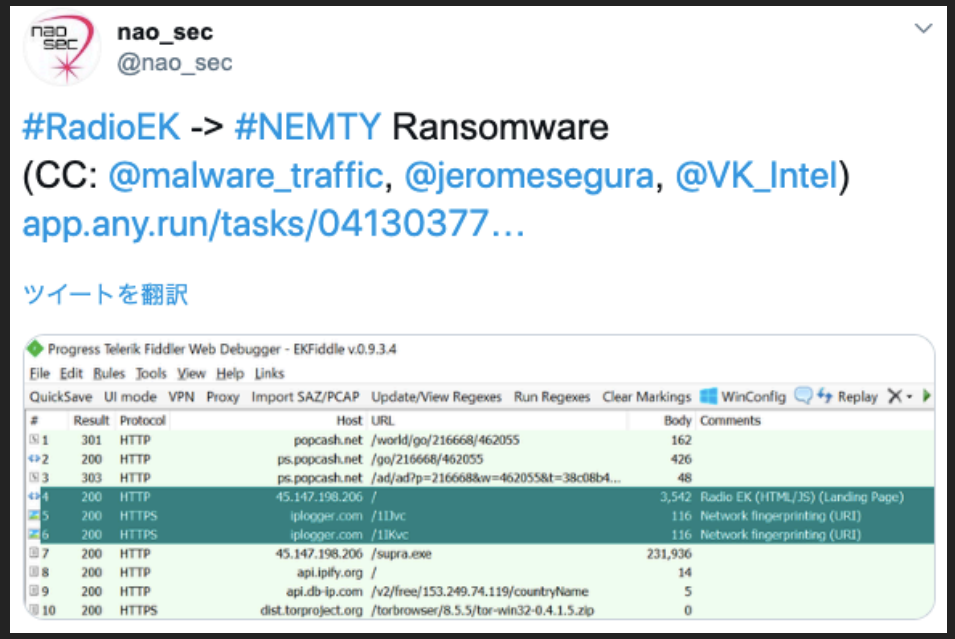
Since July 11 2019, we have observed a new Drive-by Download attack. It is redirected from the ad-network. It does not use a conventional Exploit Kit such as RIG or Fallout, but uses its own exploit kit. We call this “Radio Exploit Kit”.

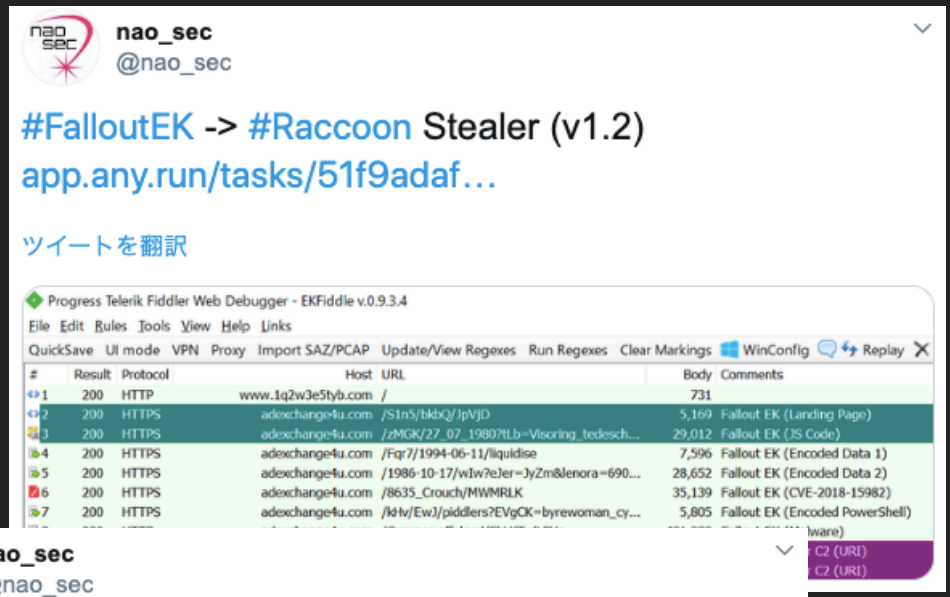
2019-07
Radio

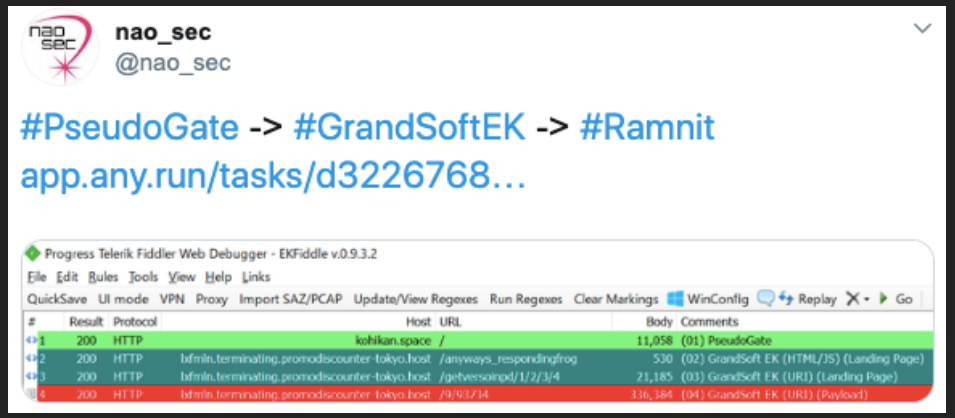
2019-08
Lord

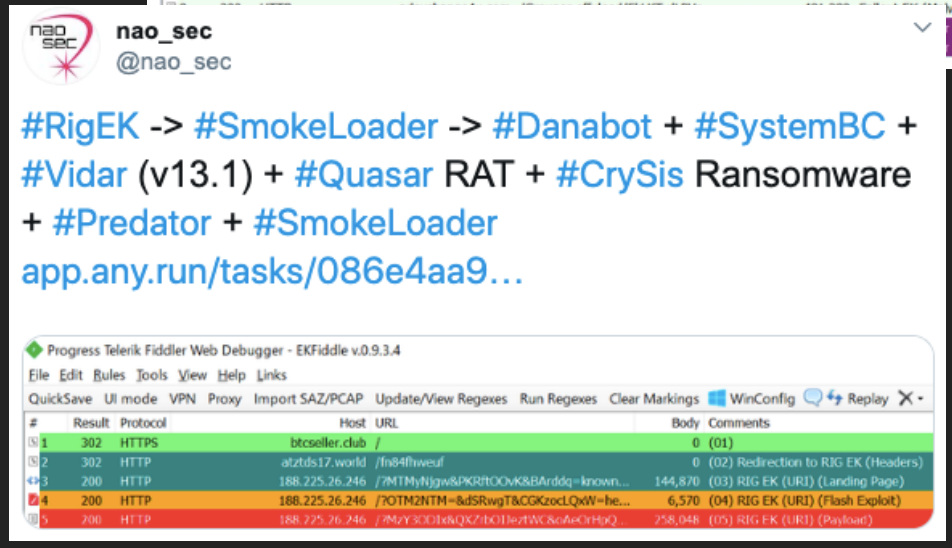
[1] https://blog.malwarebytes.com/threat-analysis/2018/07/hidden-bee-miner-delivered-via-improved-drive-by-download-toolkit/
[2] https://nao-sec.org/2018/09/hello-fallout-exploit-kit.html
[3] https://twitter.com/kafeine/status/1103649040800145409
[4] https://twitter.com/jeromesegura/status/1148289957716344832
[5] https://nao-sec.org/2019/07/weak-dbd-attack-with-radioek.html
[6] https://twitter.com/adrian__luca/status/1156934215566536705

Our Research









Motivation of Development ”Augma“

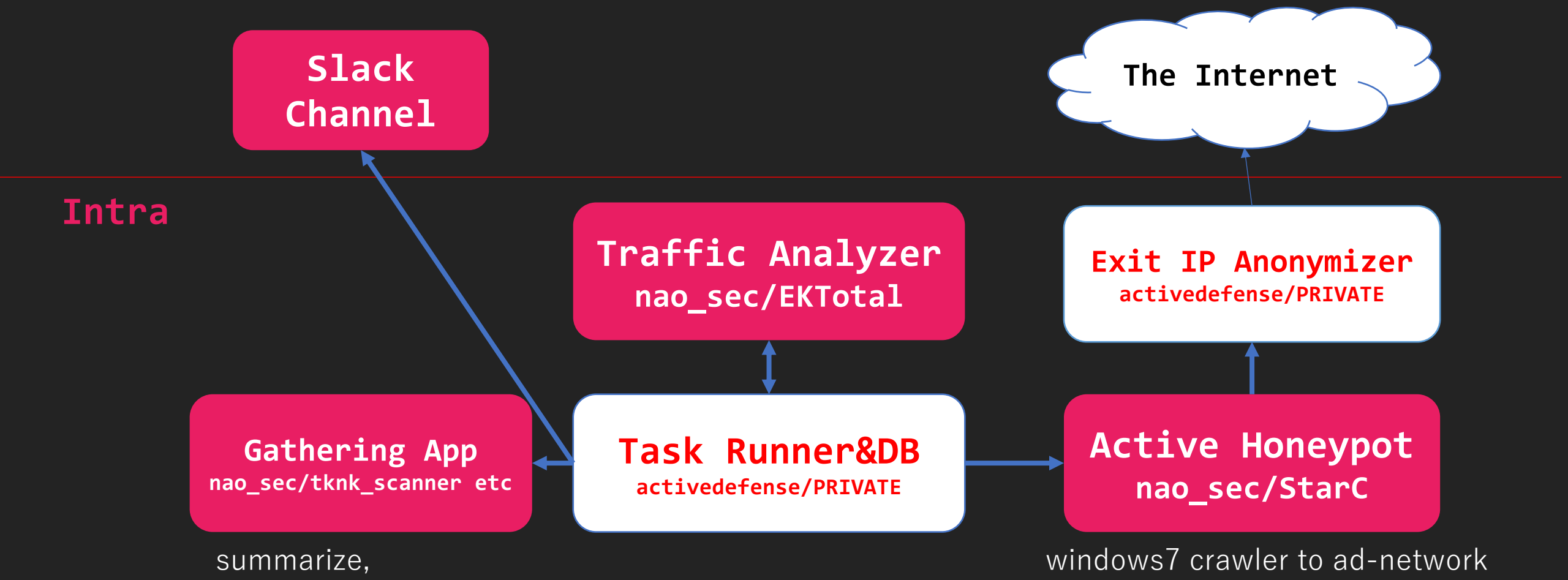
- **A drive-by threat is still “ACTIVE”**
  - Many attack campaigns and EKs have appeared
- **Manual drive-by observation is too hard**
- **We want to research the latest threat trends automatically**
  - Active Observation + Analysis + Extraction

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:
An automated active observation platform

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

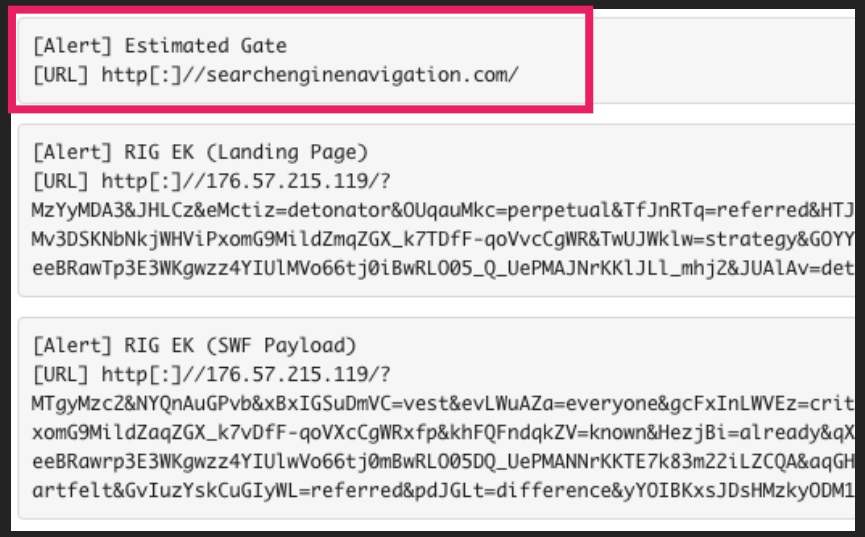
Traffic Analyzer (EKTotal)

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

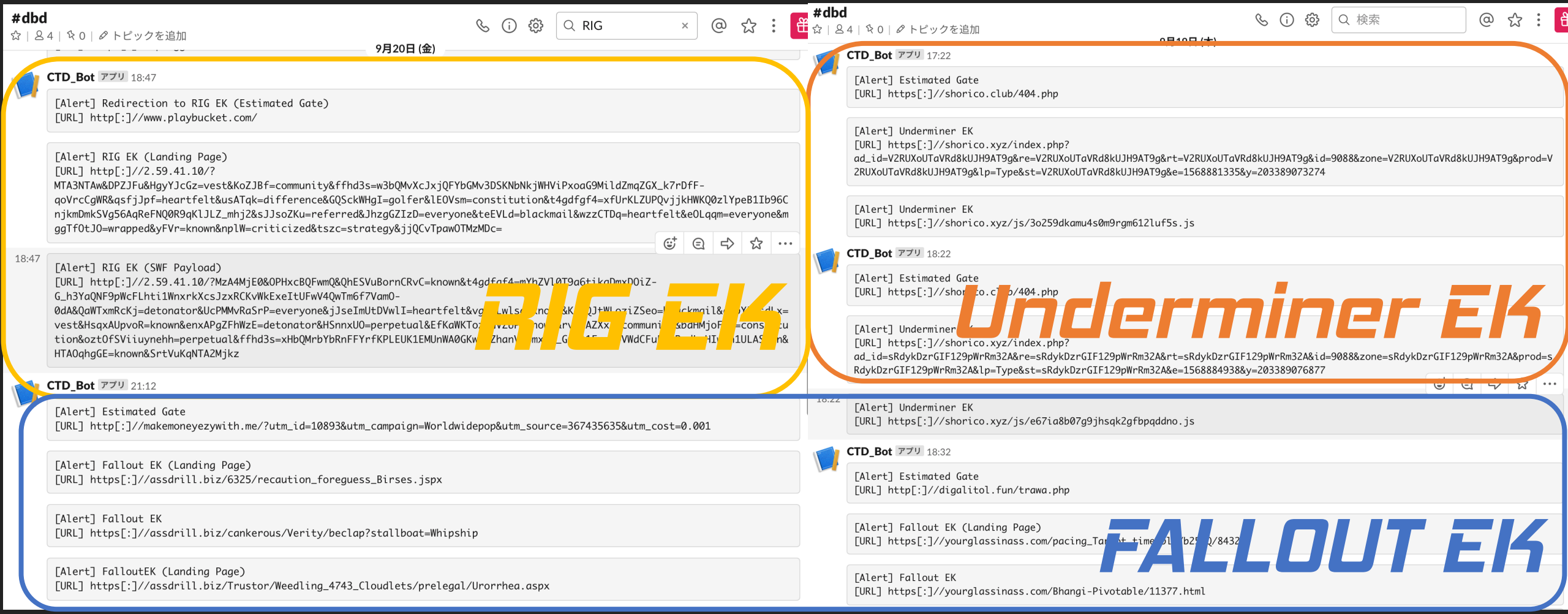
  - Input a pcap or saz, EKTotal analyze traffic data
    - Identify campaign & EK
    - Extract some information
      - Encode key
      - CVE Number
      - SWF file
      - Malware
    - Detecting with EKFiddle’s rules and Augma custom rules
      - https://github.com/malwareinfosec/EKFiddle
    -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



Finally...



Augma 2020 Dataset Overview

Augma 2020 Dataset

- **Period: 2019/05-2020/04**
- **Data:**
  - Captured malicious traffic as Fiddler saz format and pcapng format.
- **Files and Directories:**
  - augma2020/augma2020dataset.description.en.txt
  - augma2020/augma2020dataset.description.ja.txt
  - augma2020/ek/ek.metadata.tsv
  - augma2020/ek/samples/{*.pcap and * .saz}
  - augma2020/tss/tss.metadata.tsv
  - augma2020/tss/samples/{*.pcap and * .saz}
  - augma2020/unidentifiedek/samples/

Statistics

(Unit: file)

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

pcap: 10020
- ek: 7418
- tss: 2544
- unidentifiedek: 58

saz: 10008
- ek: 7419
- tss: 2544
- unidentifiedek: 45

Our Dataset Covered almost EKs in the wild

	Underminer EK	Fallout EK	GrandSoft EK	RIG EK	Spelevo EK	Capesand EK	Radio EK	Lord EK	Bottle EK	Custom EK(azera)	KaiXin	月次総計
Total	5176	583	520	443	103	59	35	19	9	4	1	6952
May-19	449	0	17	72	0	0	0	0	0	0	1	539
Jun-19	682	30	13	49	8	0	0	0	0	0	0	782
Jul-19	456	11	10	10	0	0	0	0	0	4	0	491
Aug-19	0	3	57	101	86	0	0	19	0	0	0	266
Sep-19	16	189	39	42	0	0	35	0	0	0	0	321
Oct-19	263	182	40	9	9	0	0	0	0	0	0	503
Nov-19	601	25	94	2	0	17	0	0	0	0	0	739
Dec-19	812	32	121	27	0	42	0	0	0	0	0	1034
Jan-20	1040	20	118	27	0	0	0	0	0	0	0	1205
Feb-20	255	51	11	32	0	0	0	0	5	0	0	354
Mar-20	510	39	0	72	0	0	0	0	4	0	0	625
Apr-20	92	1	0	0	0	0	0	0	0	0	0	93

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.

Any Questions?

Twitter: @nao_sec
Email: info@nao-sec.org