Soliton®

Soliton Dataset 2021

2021年6月2日

株式会社ソリトンシステムズ

Soliton Dataset 2021 について

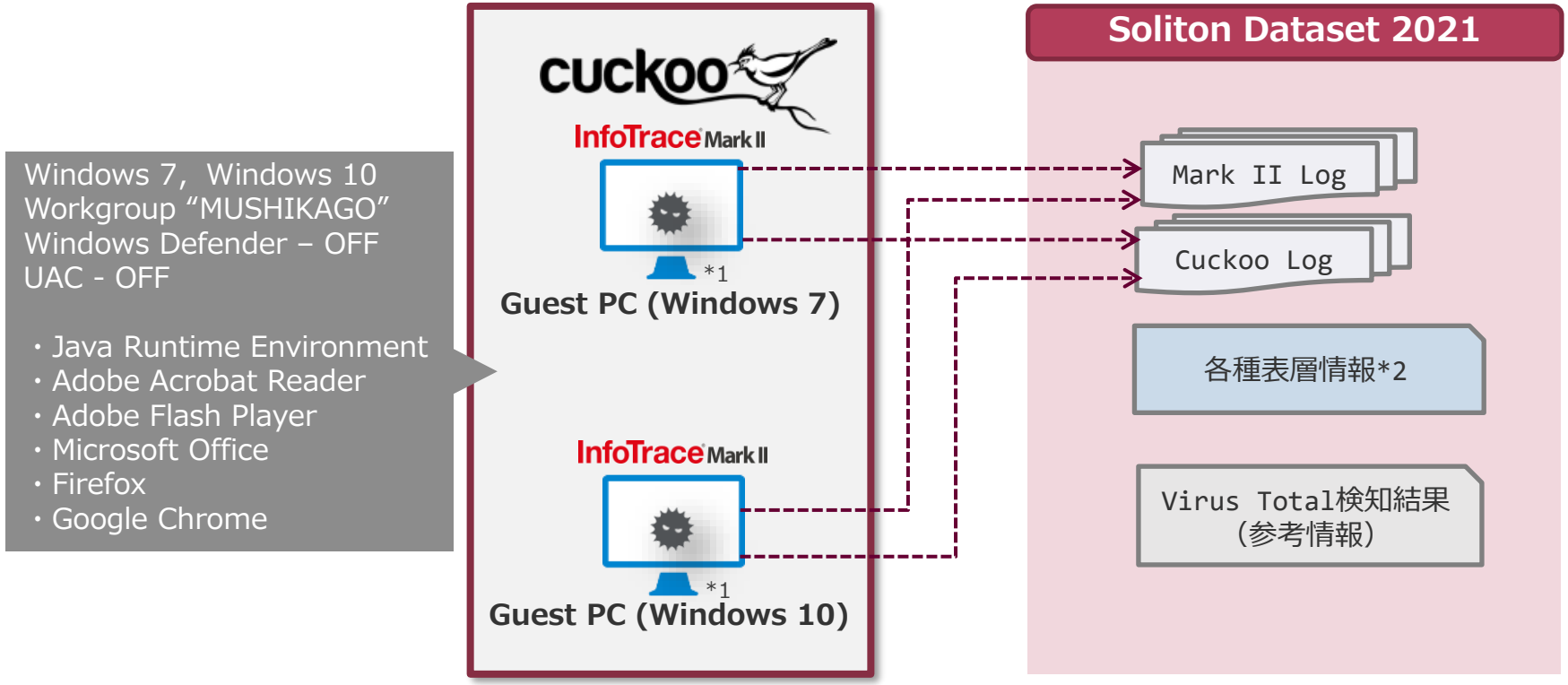
- エンタープライズ向けEDR製品であるInfoTrace Mark II（以下Mark II）は、端末における操作・挙動を記録し、サイバー攻撃や内部不正の調査を支援する製品です。
- この特性は、実際のフォレンジック現場で目にするデータに近いものとしてマルウェア対策研究に役立つと考え、マルウェアをMark II導入環境で動作させた際のログをデータセットとして提供します。

- マルウェア対策研究においては様々な観点での調査を行うため、複数種類のデータが提供されていることが望ましいと考えました。
- 動的解析システム Cuckoo Sandbox上に
Windows 7 ProベースとWindows 10 EnterpriseベースでMark IIを導入したゲスト環境を構築し、Mark IIログとCuckooログの両方を
データセットとして提供します。

検体取得方針

- 2020年4月～2021年4月に話題になったマルウェア 787検体
  - Soliton Dataset 2020では581検体
- 調査会社などから解析結果が公開されているものを中心にVirusTotalより収集
- ファイル形式にこだわらずに収集しているため、PEファイルだけではなく、スクリプトやマクロ型マルウェアなども含まれます 80検体
  - DLL型のマルウェアの実行パラメータは当該マルウェアのMark IIログをご参照ください。

マルウェア実行・ログ取得環境



*1 InfoTrace Mark IIが導入されたCuckooゲスト端末からインターネットへの通信は禁止した状態でログ取得しました。
*2 ssdeep/impfuzzy/lief/pefile/peid/TLSH/trid/stringsを含みます。詳細はSoliton Dataset 2021のREADME等をご参照ください。

提供物一覧

SolitonDataset2021
- README.txt
- ENV.txt（環境情報）
- Sysinfo/（ログ取得端末のシステム情報）
- MalwareList.csv（検体リスト）
- Data/
  - <マルウェアファミリ名>/
    - <マルウェアハッシュ値>/
      - win7_mk2.log（Windows7におけるMark IIログ）
      - win7_mk2.json（win7_mk2.logをJSONにしたログ）
      - win10_mk2.log（Windows10におけるMark IIログ）
      - win10_mk2.json（win10_mk2.logをJSONにしたログ）
      - win7_cuckoo.json（Windows7におけるCuckooログ）
      - win10_cuckoo.json（Windows10におけるCuckooログ）
      - vt_report.json（VirusTotal検知結果）
- Surface/ （各マルウェアの表層情報） ※パスワードはREADMEに記載
- Format/ （クライアントログ項目一覧）
- Tools/ （便利ツール改訂版）※2019年、2020年のものから改訂しています。

検体リスト（MalwareList.csv）

SHA256	MD5	SHA1	ファミリ	FileType	VT scan_date	positives	total	MarkII ログ	Cuckoo ロ	MarkII ログ	Cuckoo ロ	参考ページ	参考ページURL
16831855b	9a0b957f1	5122094db	APT10キャ	PE32 exec	2021/3/5 6:56	60	70	136393	836332	355426	928266	MVISION I	https://kc.mcafe
0e6f4c8cd	842134705	ed435820d	Agent Tes	PE32 exec	2020/11/23 12:29	47	71	204972	7736108	325827	955854	さまざまな	https://news.so
1e533fbcc	ac6155893	8e90a8525	Agent Tes	PE32 exec	2021/3/24 12:46	53	71	122610	8399735	340509	971070	さまざまな	https://news.so
24dd0d7b6	398024f8f6	7a23a4310	Agent Tes	PE32 exec	2020/11/24 18:30	45	70	170216	6260192	491199	969645	さまざまな	https://news.so
53997af9c1	640ca1048	58790a758	Agent Tes	PE32 exec	2019/11/29 19:01	54	69	181992	7893933	749668	947293	FFRI yarai	https://www.ffri
5ace35afb	fec8af4db	6a8b77164	Agent Tes	PE32 exec	2021/3/24 13:13	51	69	117090	1744499	407896	942515	さまざまな	https://news.so
5f1b120f7	9f02e3c48	98b45781c	Agent Tes	PE32 exec	2021/3/24 12:58	55	70	185989	7302907	384625	1002698	さまざまな	https://news.so
6e49318b8	301166ba1	7bb14616f	Agent Tes	PE32 exec	2021/3/24 13:06	57	71	102274	1724108	330583	964349	さまざまな	https://news.so
abfb0ff6f6	a83ebb631	eb7e055c3	Agent Tes	PE32 exec	2020/10/27 1:41	38	71	182369	6220327	518491	955887	さまざまな	https://news.so
af0eeca2b	c511e977e	3d77b3e09	Agent Tes	PE32 exec	2021/3/24 13:24	59	71	100257	2229472	395484	986338	さまざまな	https://news.so
b639c9915	4e507e119	2dfc88f405	Agent Tes	PE32 exec	2021/1/28 18:30	55	71	149033	4604487	479006	978225	18/01/202	https://www.tgs

- 検体ハッシュ値(SHA256，SHA-1，MD5)
- マルウェアファミリ
- ファイルタイプ
- VirusTotalスキャン日時
- 検知したアンチウイルスエンジン数(VTより)
- 使用したアンチウイルスエンジン数(VTより)
- Mark IIログファイルサイズ
- Cuckooログファイルサイズ
- 参考ページタイトル・コメント
- 参考ページURL

例）Mazeの起動（Mark IIログ）

04/01/2021 15:10:15.977 +0900 sn=384 lv=6 **evt=ps subEvt=start** os=Win
com="mws" domain="MUSHIKAGO" profile="mws"
tmid=**10dee863-ffa2-5171-7b8e-56a96e2f9f0b**
csid=S-1-5-21-1134204224-2411533656-1793949185
ip=172.24.7.101,fe80::70e6:25ab:67c9:141c mac=52:54:00:ca:68:ad usr="mws"
usrDomain="MWS" sessionID=1 psGUID={E9557369-BE3C-4956-AFA8-27C811C2692F}
psPath="C:¥Users¥mws¥AppData¥Local¥Temp¥**dee863ffa251717b8e56a96e2f9f0b41b09897d3c7cb2e8159fcb0ac0783611b**.exe" **psID=96**
parentGUID={887970AD-3804-4DBB-B7F6-E976313498E7}
parentPath="C:¥tmpd9bj1e¥bin¥inject-x86.exe" psUser="mws" psDomain="MWS"
arc=x86
sha256=**dee863ffa251717b8e56a96e2f9f0b41b09897d3c7cb2e8159fcb0ac0783611b**
sha1=31c3f7b523e1e406d330958e28882227765c3c5e
md5=c9ea6430da4e72b672ce29e56ecad603 crTime="03/08/2021 19:27:11.576"
acTime="04/01/2021 15:10:15.966" moTime="03/08/2021 19:27:11.639" size=373248
sig=None

※実行環境と実行されたマルウェアがすぐ分かるように、OSバージョンとSHA256をTMID名（端末ID）に使用しています。
また、SHA256はマルウェアのファイル名にも使用しています。

例）Mazeの起動（Cuckooログ）


```
"behavior": {
        "generic": [
            {
  “process_path“:
"C:¥¥Users¥¥mws¥¥AppData¥¥Local¥¥Temp¥¥dee863ffa251717b8e56a96e2f9f0b41b09897d3c7cb2e8159fcb0ac0783611b.exe",
          "process_name": "dee863ffa251717b8e56a96e2f9f0b41b09897d3c7cb2e8159fcb0ac0783611b.exe",
              "pid": 96,
              "summary": {
                  "file_created": [
 （省略）
},
              "first_seen": 1617289816.435221,
              "ppid": 7048
},

:
:
```

例）Mazeによるファイル暗号化（Mark IIログ）

- 04/01/2021 15:12:24.778 +0900 sn=3024 lv=7 rs=11 trs=63 rf=C16:C8:L8:R8:C9:L10:R10 **evt=file subEvt=rename** os=Win com="mws" usr="mws" usrDomain="MWS" sessionID=1 psGUID={E9557369-BE3C-4956-AFA8-27C811C2692F} psPath="C:¥Users¥mws¥AppData¥Local¥Temp¥dee863ffa251717b8e56a96e2f9f0b41b09897d3c7cb2e8159fcb0ac0783611b.exe" **path="C:¥Users¥mws¥Downloads¥upload.py"** drvType=HDD **dstPath="C:¥Users¥mws¥Downloads¥upload.py.bQfw"** dstDrv=HDD sha256=19711f879c780b8e919d071744e626459585dbaa89b304323e3507888059105a crTime="02/03/2021 17:23:39.210" acTime="04/01/2021 15:12:24.778" moTime="04/01/2021 15:12:24.778" **size=1900**

- 04/01/2021 15:12:24.778 +0900 sn=3026 lv=7 rs=11 trs=63 rf=C16:C8:L8:R8:C9:L10:R10 **evt=file subEvt=close** os=Win com="mws" usr="mws" usrDomain="MWS" sessionID=1 psGUID={E9557369-BE3C-4956-AFA8-27C811C2692F} psPath="C:¥Users¥mws¥AppData¥Local¥Temp¥dee863ffa251717b8e56a96e2f9f0b41b09897d3c7cb2e8159fcb0ac0783611b.exe" **path="C:¥Users¥mws¥Downloads¥upload.py"** drvType=HDD **read=1636 write=264 mmf=1** sha256=19711f879c780b8e919d071744e626459585dbaa89b304323e3507888059105a sTime="04/01/2021 15:12:24.778" crTime="02/03/2021 17:23:39.210" acTime="04/01/2021 15:12:24.778" moTime="04/01/2021 15:12:24.778" **size=1900** new=0

例）Mazeによるファイル暗号化（Cuckooログ）


```
{
                "category": "crypto",
                "status": 1,
                "stacktrace": [],
                "api": "CryptEncrypt",
                "return_value": 1,
                "arguments": {
                   "hash_handle": "0x00000000",
                   "buffer":
"¥u0090¥u00dd¥u00f8¥u00af¥u0014¥u008a¥u00f8i¥u00de*x¥u0019¥u00ea¥u0005¥u00141¥r8¥u00e4¥u00a7¥
u00ff]¥u00eaJ¥u00e1UB¥u009f¥u00cdj*¥u00c5¥u00c7¥u00eeC¥u00e1¥u00ac¥u00bbcn¥u0000¥u0000¥u0000¥
:
:
¥u0000¥u0000¥u0000",
                   "key_handle": "0x004d70c0",
                   "flags": 0,
                   "final": 1
                },
                "time": 1617289944.779221,
                "tid": 6824,
                "flags": {}
             },
:
:
```

Soliton Dataset 2021の利用例

動的解析の研究・学習に

- Mark IIログとCuckooログの両方を確認できます
  - Mark IIで動作の流れを把握、Cuckooで詳細把握など
- PEファイル以外の検体も含まれます（80件/787件）
- Windows7とWindows10におけるログが確認できます
- エンタープライズの実環境に近い、OS標準ソフトウェアなどの動作も含まれたログのため、実環境でマルウェア挙動を見出す研究の参考としてお使いいただけます

注意点

- Cuckoo上でMark IIを動かしているため、ログの取得時間に差があります。ランサムウェアなど、ログ量が多いと差が生じやすくなります。

ご質問・ご意見など

Slack-MWS #dataset チャネルまで
お気軽にどうぞ！