

# Efficient data sharing and correlation

Pavel Kácha <ph@cesnet.cz>  
Václav Bartoš <bartos@cesnet.cz>

46<sup>th</sup> TF-CSIRT Meeting  
25<sup>th</sup> September 2015, Tallin

# Problem...

- Administrators often run their own IDS`, security probes, central syslog
  - For network health, finding compromised machines, botnet activity, malware, antispam
  - They pick just the data, important for *them*
  - ... *and throw away the rest*

# Problem...

- Administrators often run their own IDS`, security probes, central syslog
  - For network health, finding compromised machines, botnet activity, malware, antispam
  - They pick just the data, important for *them*
  - ... *and throw away the rest*

## Let's share!

# Problem...

- Administrators often run their own IDS`, security probes, central syslog
  - For network health, finding compromised machines, botnet activity, malware, antispam
  - They pick just the data, important for *them*
  - ... and throw away the rest

## Let's share!

- *Ahem... Excuse me... nice, but...  
Format? Content? Protocol? Classifications? Policies?*

# Warden<sup>3</sup>

- *Client/server* (glorified queue – transport, not storage)
- *Community*
  - Reciprocity – all your data is available to the whole Warden community...
  - ... and all the community data is available to you

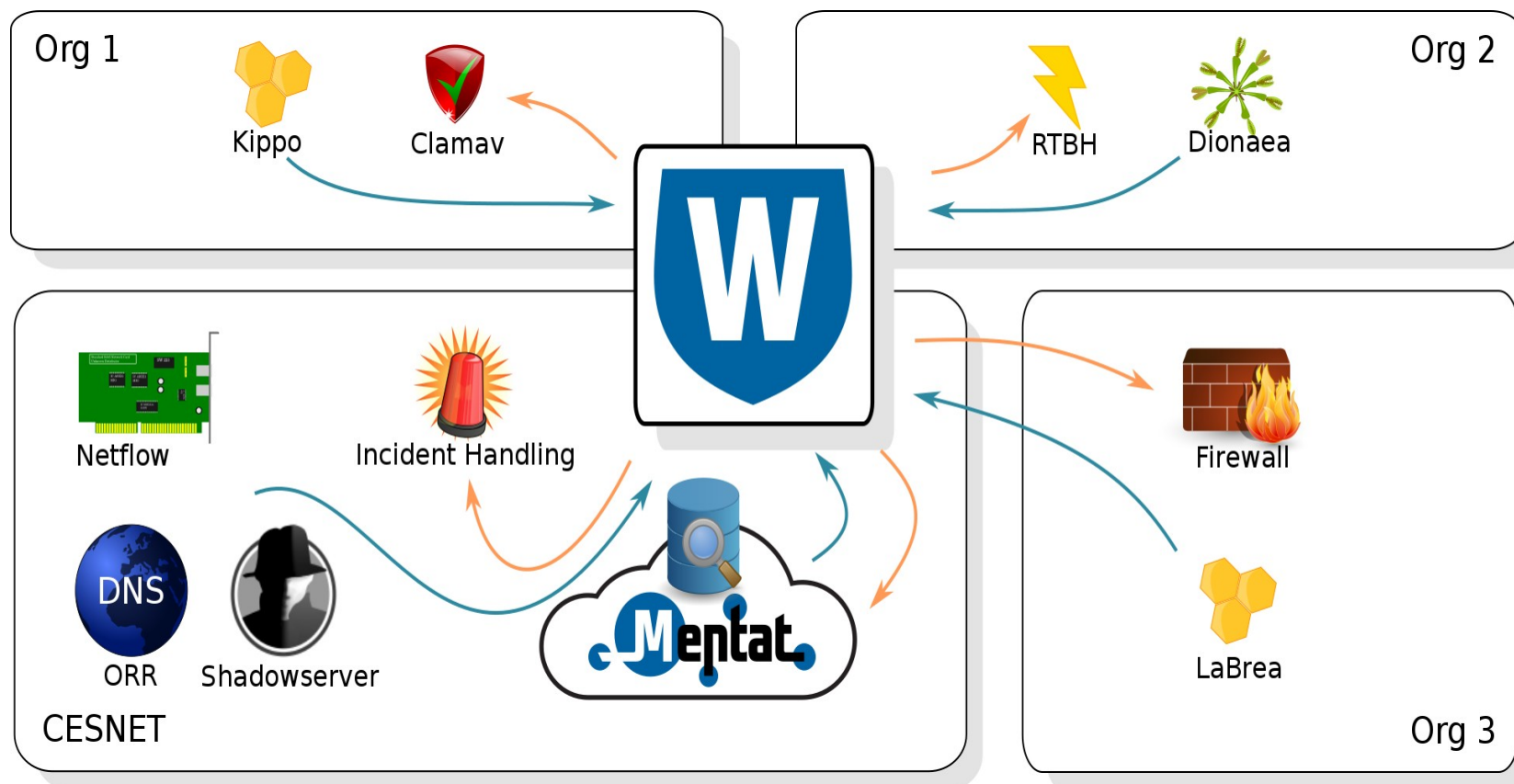
- *Bulk operations*

- *Filtering*

- *Security*

X509  
encryption  
“sanity” checks  
peer review

- *Open/Libre*



# IDEA - Intrusion Detection Extensible Alert

## Botnet C&C

```
{
  "Format": "IDEA0",
  "ID": "cca3325c-a989-4f8c-998f-5b0e971f6ef0",
  "DetectTime": "2014-03-05T15:52:22Z",
  "Category": ["Intrusion.Botnet"],
  "Description": "Botnet Command and Control",
  "Source": [
    {
      "Type": ["Botnet", "CC"],
      "IP4": ["93.184.216.119"],
      "Proto": ["tcp", "ircu"],
      "Port": [6667]
    }
  ]
}
```

## Honeypot

```
{
  "Format": "IDEA0",
  "ID": "2E4A3926-B1B9-41E3-89AE-B6B474EBOA54",
  "DetectTime": "2014-03-22T10:12:31Z",
  "Category": ["Recon.Scanning"],
  "ConnCount": 633,
  "Description": "EPMAPPER exploitation attempt",
  "Ref": ["cve:CVE-2003-0605"],
  "Source": [
    {
      "IP4": ["93.184.216.119"],
      "Proto": ["tcp", "epmap"],
      "Port": [24508]
    }
  ],
  "Target": [
    {
      "Port": [135]
    }
  ]
}
```

- JSON (NoSQL friendly), but mostly flat and typed structure (SQL friendly)
- Extensibility (*producers can use their own keys and tags*)
- Marking of anonymised, imprecise, forged data
- Able to distinct events of which we are the primary source, 3<sup>rd</sup> party events, correlated events
- Taxonomies (*mkll categories, tag based Source/Target/Detector description*)

# Warden - HTTP API

- Server – Python, WSGI (Apache), MySQL
- Protocol – HTTP + JSON

```
$ curl 'https://warden.example.com/getEvents?count=1&id=12'
```

```
{"lastid": 13,  
  "events": [  
    {"Format": "IDEA0",  
      "ID": "48fb18c4-435d-4cd8-ad8a-fb4c2998f3d0",  
      "Category": ["Test"],  
      "DetectTime": "2014-10-19T15:22:20.409128Z"}]}
```

```
$ curl --request POST --data '{#$$%^' 'https://.../sendEvents'
```

```
{"error": 400,  
  "method": "getEvents",  
  "message": "Deserialization error, cause was ValueError: Expecting  
property name: line 1 column 1 (char 1)",  
  "detail": {  
    "args": "{#$$%^"  
  }  
}
```

# Warden - Python API

```
wclient = Client(
    **read_cfg("warden_client.cfg"))

# -- or --

wclient = Client(
    url = 'https://.../warden3',
    keyfile = 'etc/key.pem',
    certfile = 'etc/cert.pem',
    cafile = 'etc/tcs-ca-bundle.pem',
    timeout = 10,
    errlog = {"level": "debug"},
    filelog = {"level": "debug"},
    idstore = "MyClient.id",
    name = "cz.cesnet.honeypot.kippo"
)
```

```
# receiving
ret = wclient.getEvents(count=10)
for e in ret:
    print e
if isinstance(ret, Error):
    print("Error: %s" % ret)

# sending
event = {
    "Format": "IDEA0",
    "ID": str(uuid4()),
    "DetectTime": isostamp(datetime),
    "Category": ["Test"]
}
ret = wclient.sendEvents([event])
if not ret:
    print("Error: %s" % ret)
```



## Search alerts

**Q Alert database search**

**Source:** 127.0.0.1 **Target:** 127.0.0.1 **AND OR**

**From:** 2012-12-12 12:12:12 **To:** 2012-12-12 12:12:12

**Detector:** --- Unspecified ---  
cesnet.au1/LaBrea  
cesnet.au1/SSERV  
cesnet.au1/X4  
cesnet.au1/USERS

**Category:** Anomaly.Traffic  
Attempt.Exploit  
Attempt.Login  
Availability.DDoS  
Availability.DoS

**Search** **Go Advance**

If you use certain queries often, you might consider saving them:

--- Personal query --- Unique name for the query **Save**

Displaying items 1 to 30 (30 items) | Page 1

**Next**

#	Detected	Source	Target	Categorization	
1	2015-09-22 13:18:05	-- undisclosed --	211.240.36.71	Availability.DoS	
2	2015-09-22 13:13:23	-- undisclosed --	193.87.171.19	Availability.DoS	

## Report M20150922M-F4amT

Unprotected access: <https://mentat-hub.cesnet.cz/mentat/unauth/report/32WvuPYwWxpyaxZAhxMo>

Severity	Abuse	Created
medium	abuse@vstecb.cz	2015-09-22 08:06:37

## Report timing

Time period	2015-09-22 06:00:00 - 2015-09-22 08:00:00 (2h)
Delay	6m 37s
Report sent	2015-09-22 08:06:37   Report mailed to abuse contact 'abuse@vstecb.cz'

## Report magnitude

Event count	400 (400 entered filtering, 0 blocked)
IP count	1 unique IP address
Diversisty	1 analyzer, 2 categories

## Report message

Vážení kolegové,

detekční systémy CESNETu zaznamenaly následující problém(y) související s Vašim rozsahem IP adres nebo Vaší doménou (uvedené časy jsou lokální):

[1] Systémy na následujících IP adresách jsou infikovány známým malware, součástí botnetu (Botnet Drone):

\* Analyzer: X4  
\* Popis: Botnet Drone  
\* Kategorie: Intrusion.Botnet/Malware

-----  
IP | Čas | # událostí  
-----  
195.113.220.250 | 2015-09-21 15:44:53 - 2015-09-22 07:14:44 | 400  
-----

# Lesson learned

## **Manpower...** lack of it

- Not only ours, but in the involved networks
- Provide potential recipients with simpler means
  - Overviews, dashboards
  - Old school mail reporting

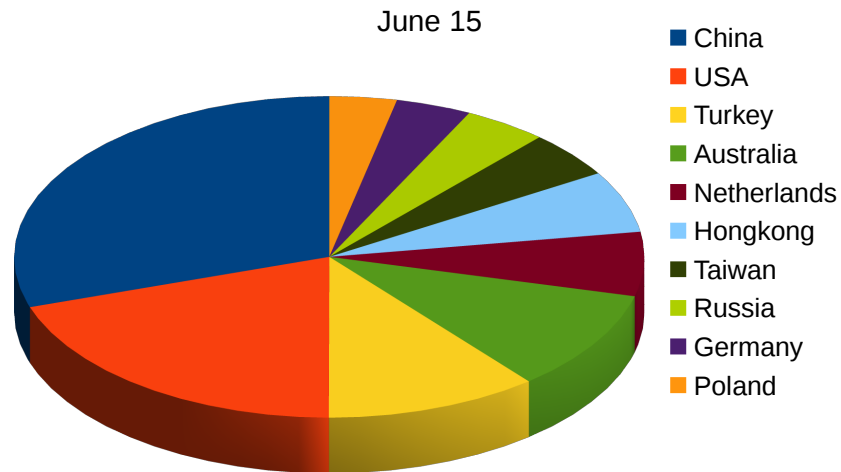
(But make sure to tune it really well, as we did in Mentat)
- Find ways to simplify spreading the reach of *your* detectors into constituency networks
  - Virtualization, docker
  - Tunneling
  - Lightweight probes

# Also...

## Next?

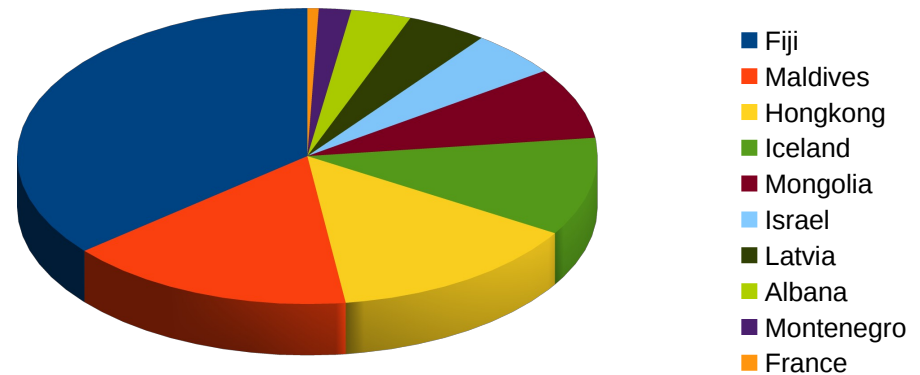
- Connectors
  - Kippo
  - Dionaea
  - BitTorrent snooper
  - RT submitter
- Warden 2 → Warden 3
- Warden-Filer – file based platform agnostic API
- General log file adaptor (fail2ban connector)
- Blacklist generators
- Visualisation generators
- Analysis, correlation
- Processes – to make data available for 3rd parties

## Incident TOP10 share by country



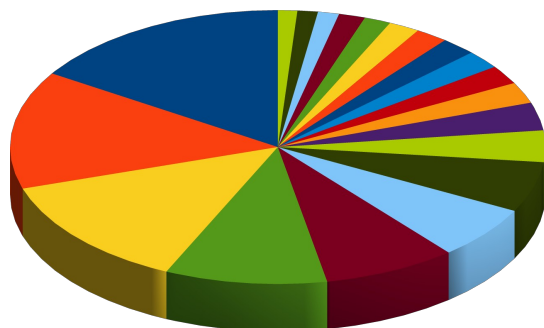
## Incident TOP10 share

according to number of incidents  
per one IP in the country  
June 2015



## TOP 20 incident share by AS

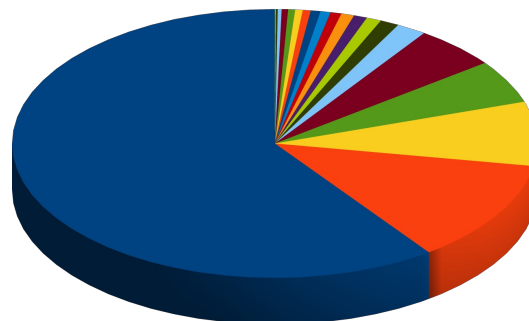
June 2015



- Chinanet CN
- Turk Telekomunikasyon Anonim Sirketi TR
- SoftLayer Technologies Inc. AU
- CNCGROUP China169 Backbone CN
- CHINANET jiangsu province backbone CN
- Ecatel LTD NL
- Data Communication Business Group TW
- CariNet, Inc. US
- SoftLayer Technologies Inc. HK
- Hurricane Electric, Inc. US
- HOT NET LIMITED HK
- PlusServer AG DE
- University of Michigan US
- Jazz Telecom S.A. ES
- Biznes-Host.pl sp. z o.o. PL
- MCI Communications Services, Inc. d/b/a Verizon Business US
- 013 NetVision Ltd. IL
- Contabo GmbH DE
- CNCGROUP IP network China169 Beijing Province Network CN
- Abovenet Communications, Inc US

## TOP 20 incident share by AS

according to number of incidents  
per one IP from AS  
June 2015



- HOT NET LIMITED
- Przedsiębiorstwo Usług Specjalistycznych ELAN mgr inż.
- Nikultsev Aleksandr Nikolaevich
- Ecatel LTD
- DELORIAN Internet Services Artur Grabowski
- Nagravision SA
- DataClub S.A.
- PE Voronov Evgen Sergiyovich
- Livenet Sp, z o.o.
- WEDOS Internet, a.s.
- Storm Systems LLC
- MediaServicePlus Ltd.
- Black Fox Limited
- CariNet, Inc.
- Iradeum Trading Ltd.
- DataWagon LLC
- DDNET SOLUTIONS SRL
- HOSTKEY B.V.
- Hosting Solution Ltd.

# Alert data analysis

- Dataset: **all alerts** reported to Warden **in one month** (June 2015)
  - 29 mil. alerts from 14 reporters

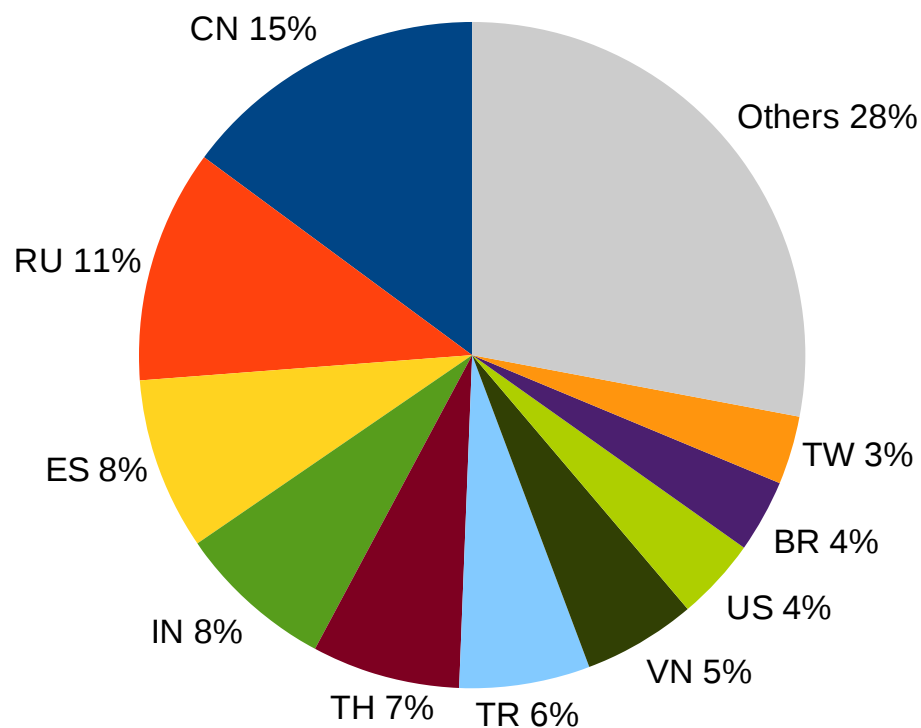
Category	# alerts	# uinq. src. addrs.
Scanning	27,295,094	814,333
Login attempt	1,718,566	4,125
Copyright	176,790	732
Botnet drone	9,981	61 (drone) / 32 (CC)
(D)DoS	7,100	134
Spam	6,943	3,593
<i>Others</i>	<i>7,147</i>	--

- Analysis focused on **numbers of source addresses**, not alerts
- Only **scan** and **login** data used

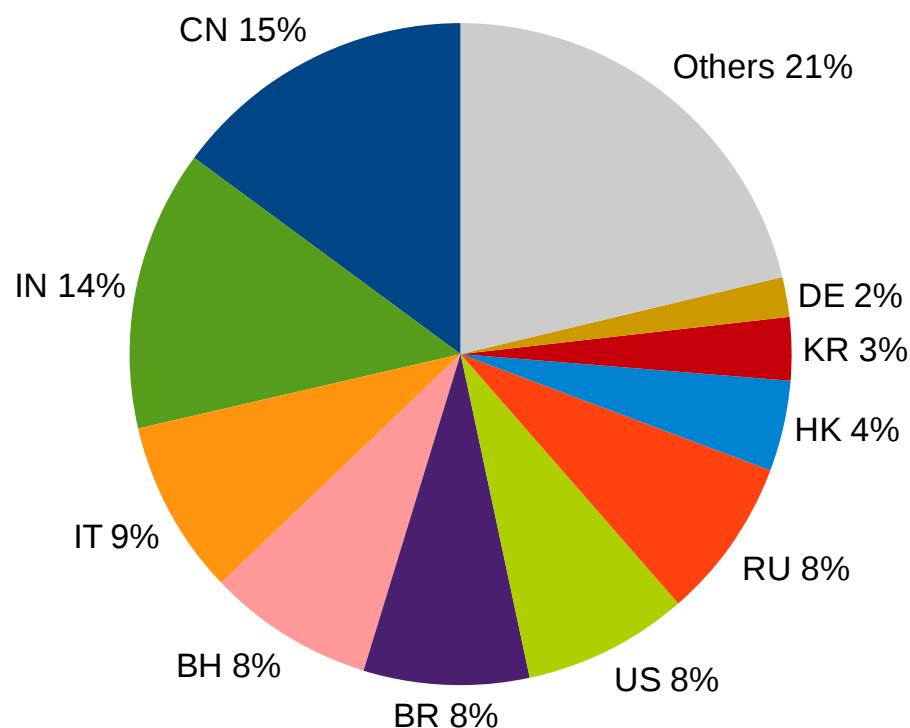
# Bad countries

- Number of source addresses by country

Top 10 countries of origin (scan)



Top 10 countries of origin (login)



# Bad (and good) countries

- **Relate** attacks to distribution of **normal traffic**
  - By number of unique addresses observed in a **sample of NetFlow data**
- **Relative badness**

$$\text{Badness}(C) = \frac{\text{portion of attacking addresses from country } C}{\text{portion of all communicating addresses from country } C}$$

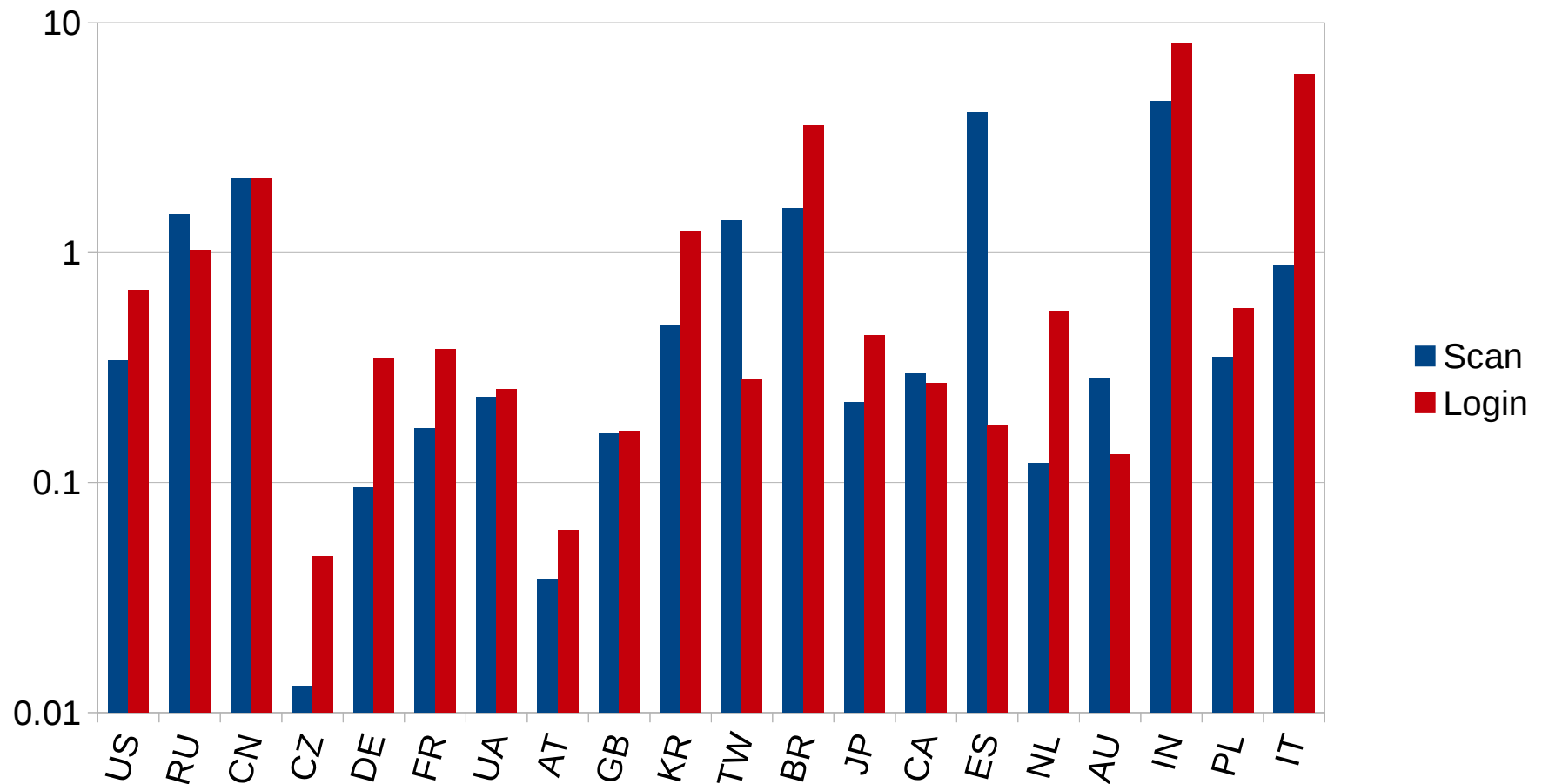
<1 good                  ≈1 average                  >1 bad

- Example:
  - From **China** comes **7.0%** of all observed addresses, but **14.8%** of scanners.  
 $\text{Badness}(\text{China}) = 14.8 / 7.0 = \mathbf{2.1}$  „more scanners than average“
  - From **USA** comes **11.7%** of all observed addresses, but only **4.0%** of scanners.  
 $\text{Badness}(\text{USA}) = 4.0 / 11.7 = \mathbf{0.34}$  „less scanners than average“



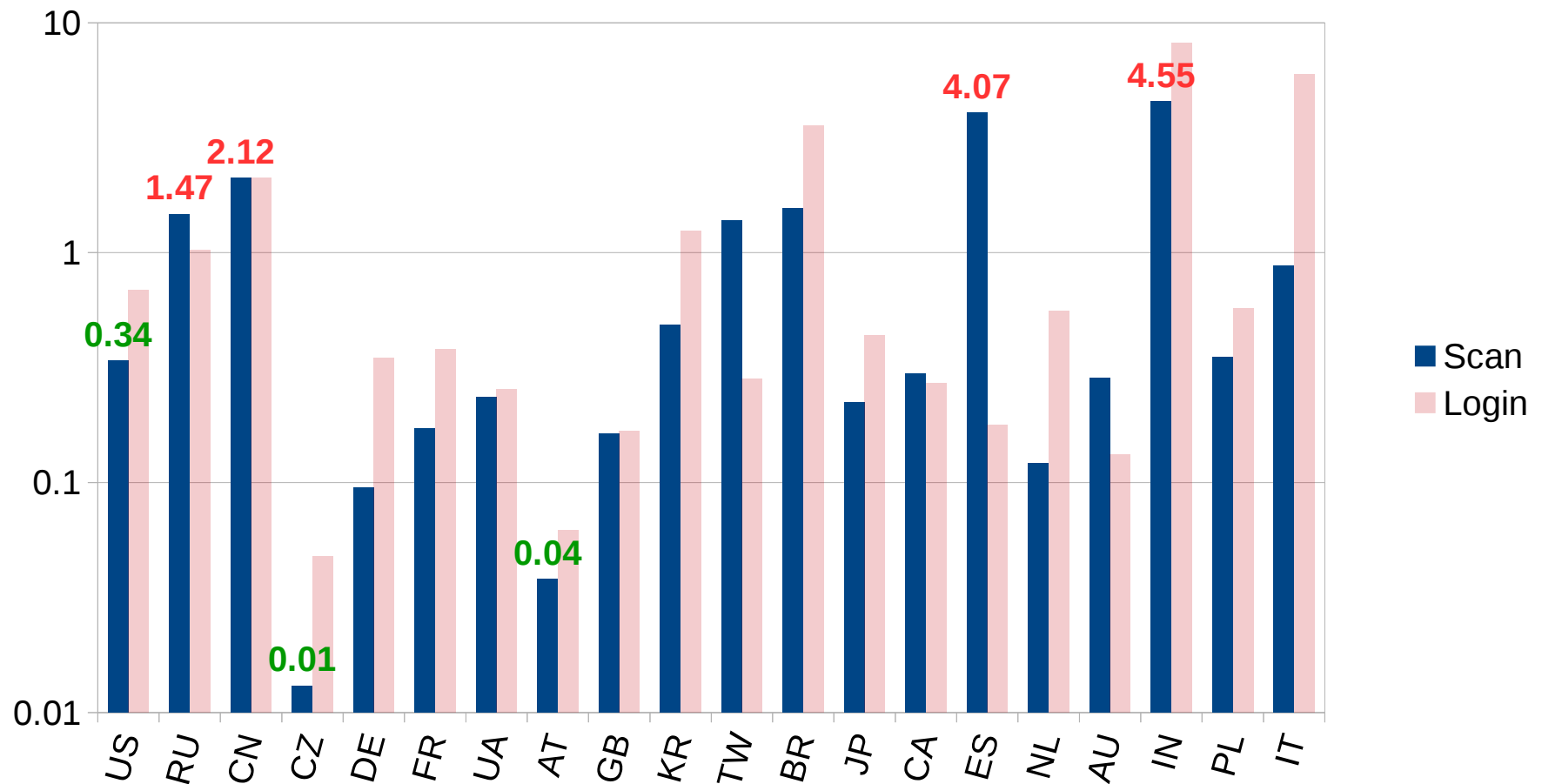
# Bad (and good) countries

- Relative badness of top 20 countries (by number of addresses)



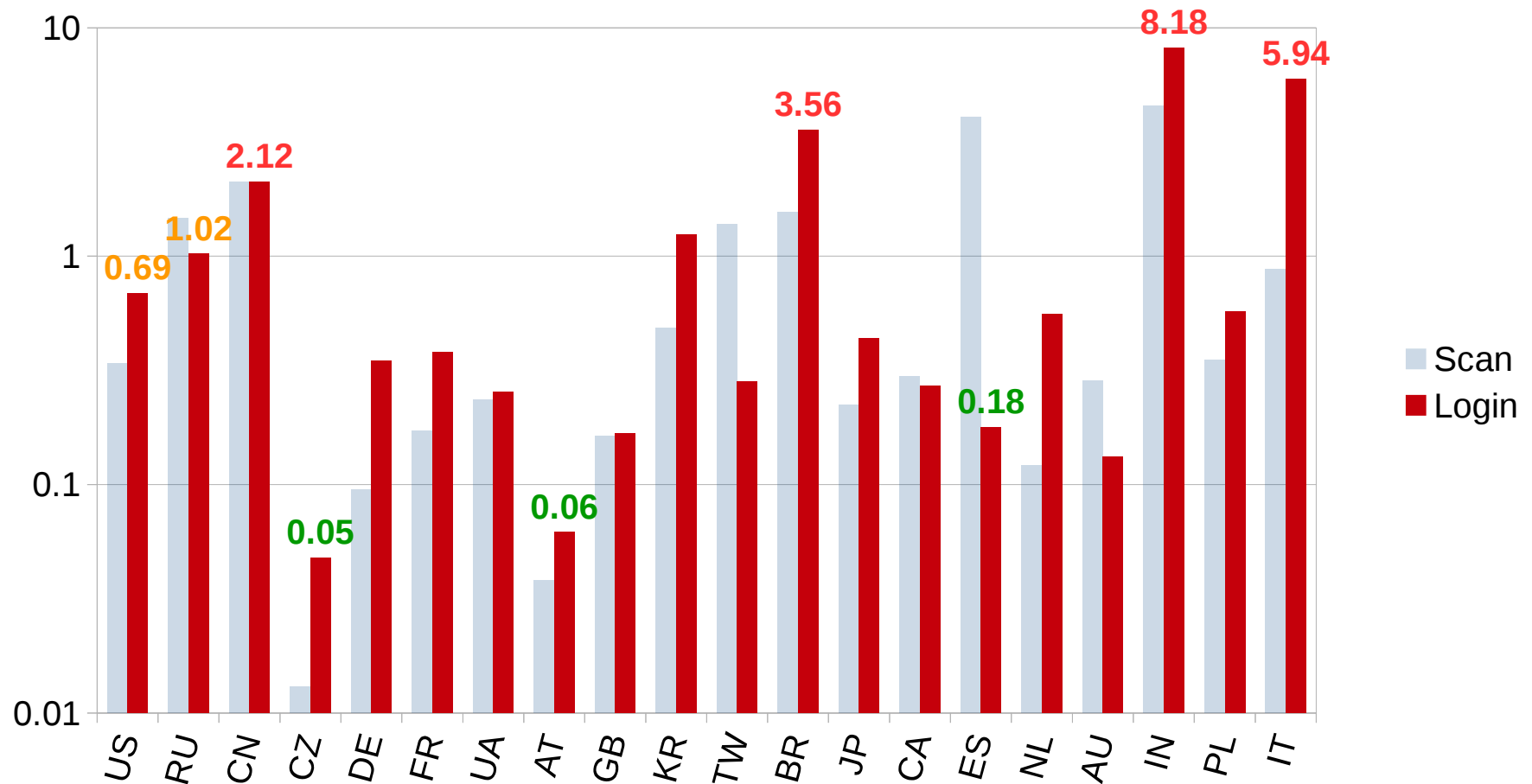
# Bad (and good) countries

- Relative badness of top 20 countries (by number of addresses)



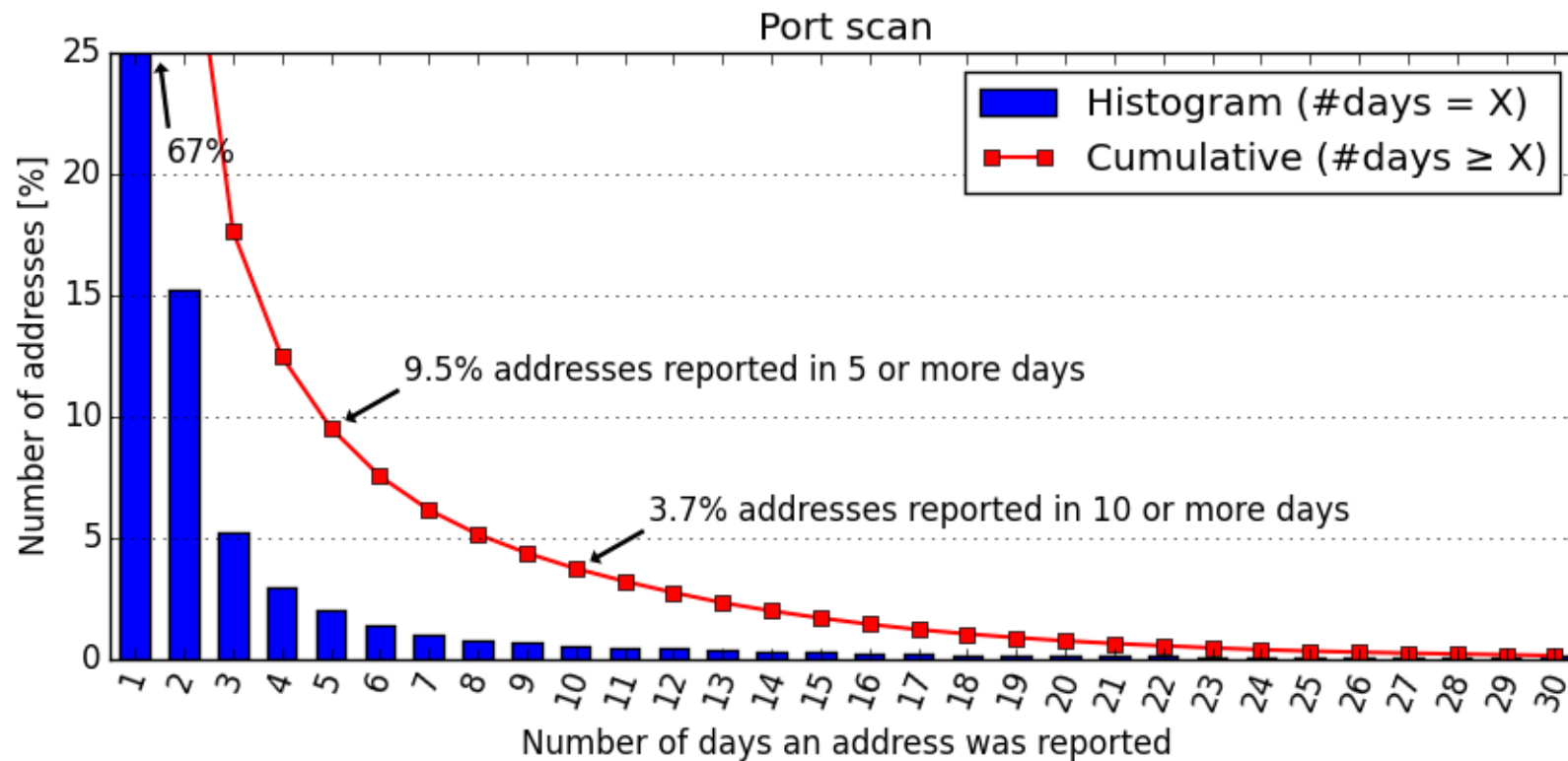
# Bad (and good) countries

- Relative badness of top 20 countries (by number of addresses)



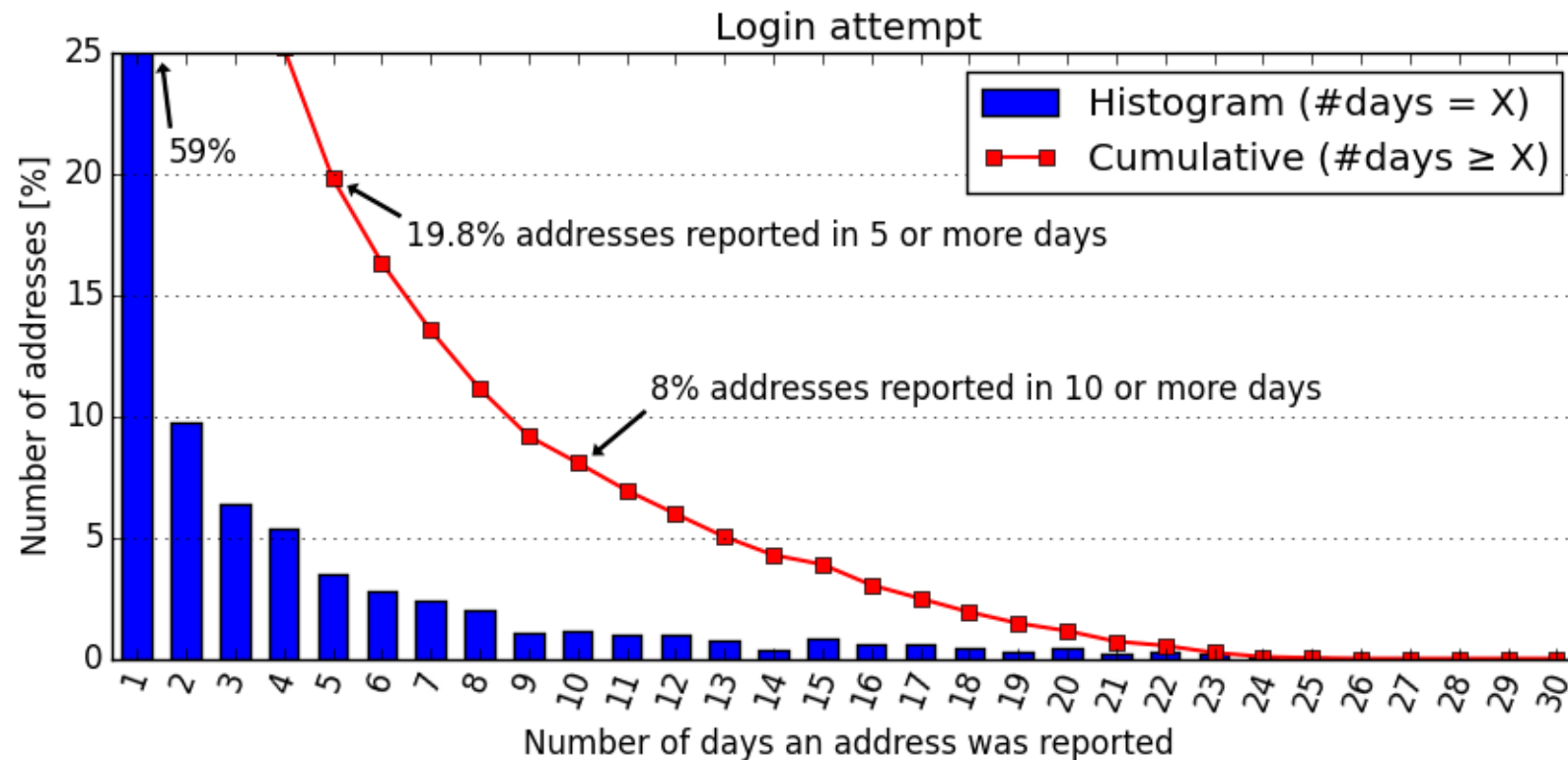
# Time correlations

- In how many days in a month an address was reported as scanning?
  - 67% reported only once
  - 9.5% reported 5 or more times - notorious attackers
    - Responsible for 62% of all alerts!



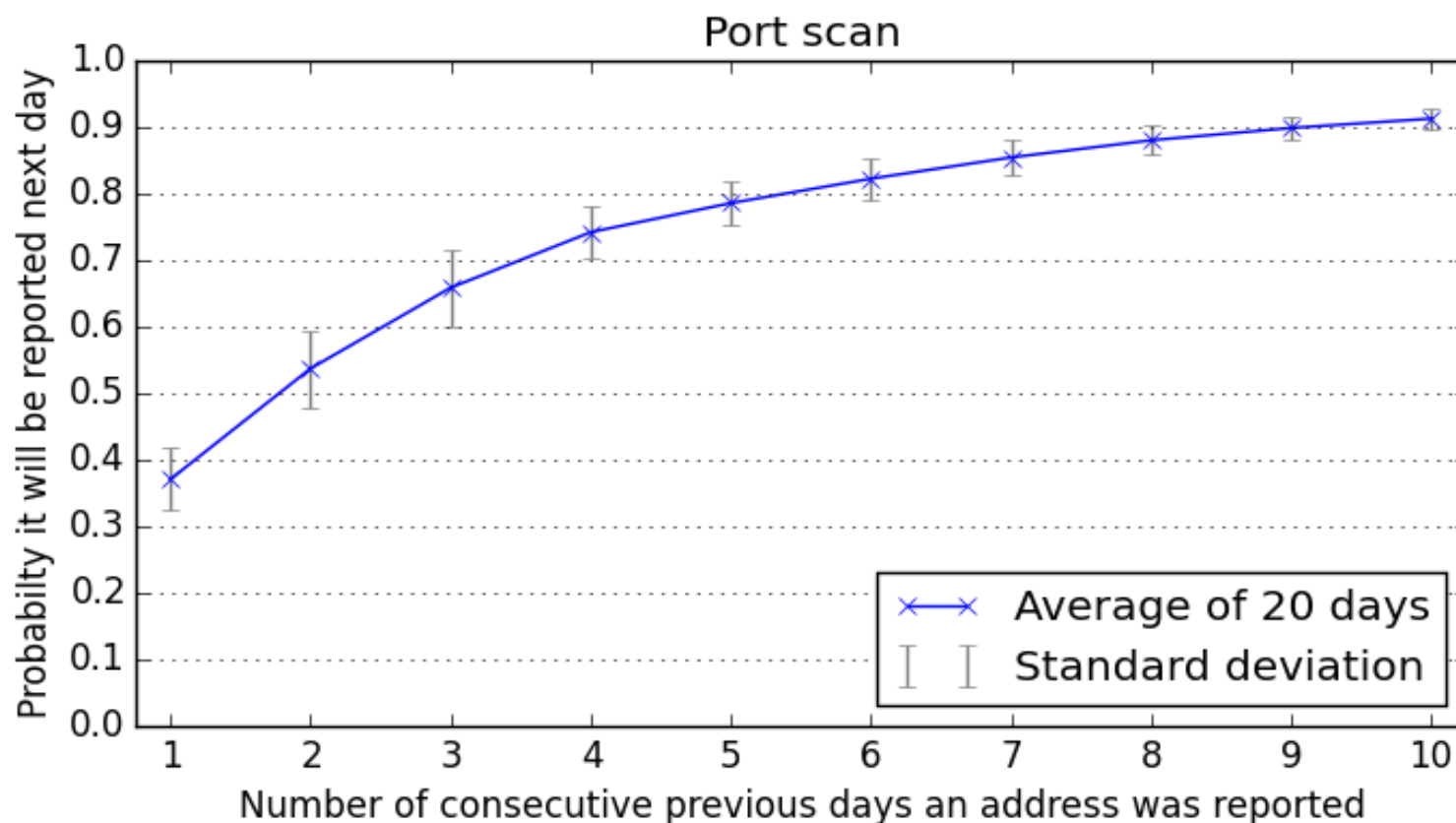
# Time correlations

- The same for **login attempt** data ...
  - 59% reported only once
  - 20% reported 5 or more times (resp. for 22% of alerts)
  - No address reported more than 24 times



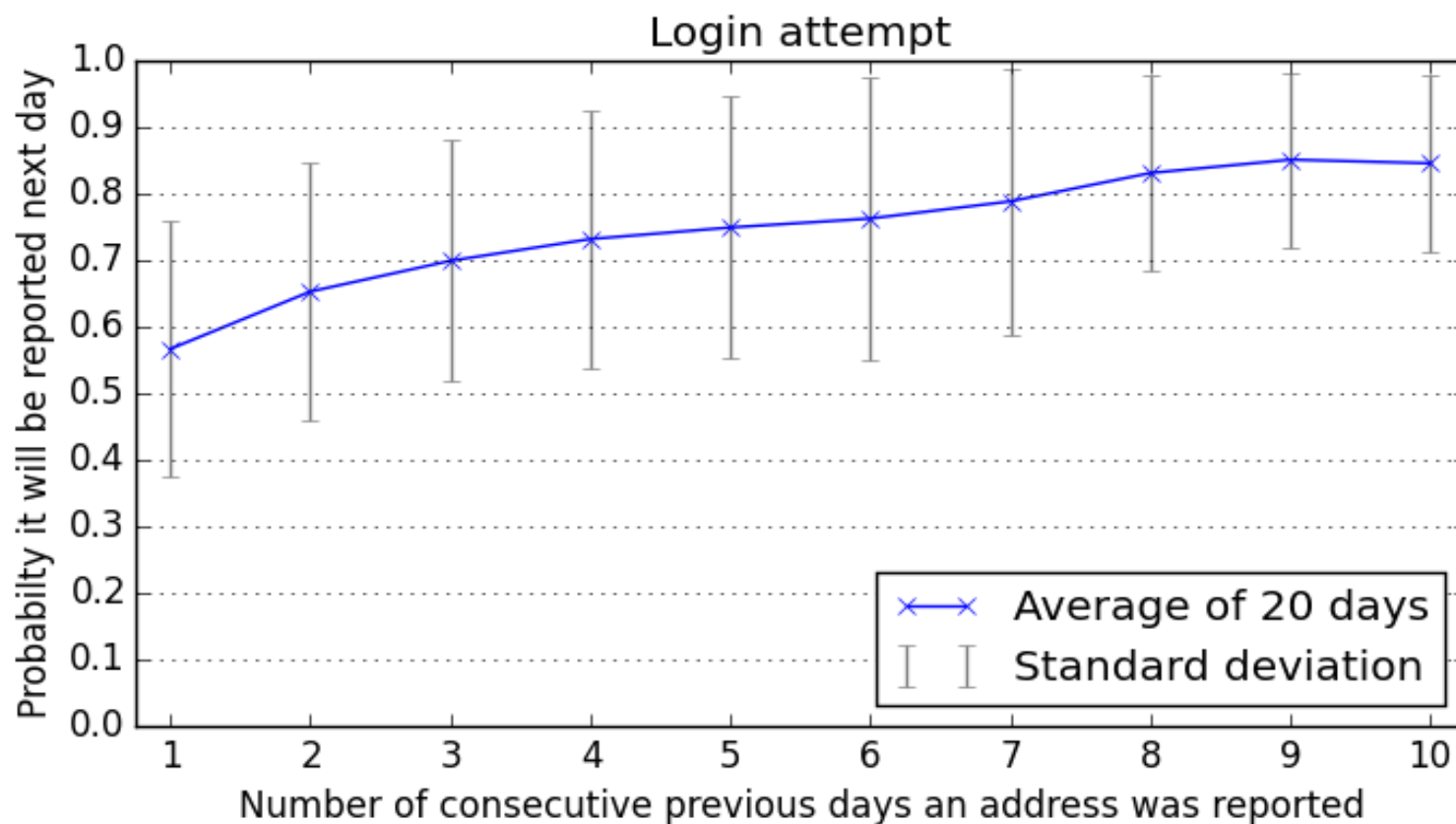
# Predictability

- If an address attacked previously, will it attack again?
  - $\text{Prob}(\text{reported at day } D \mid \text{reported at days } [D-n, \dots, D-1])$



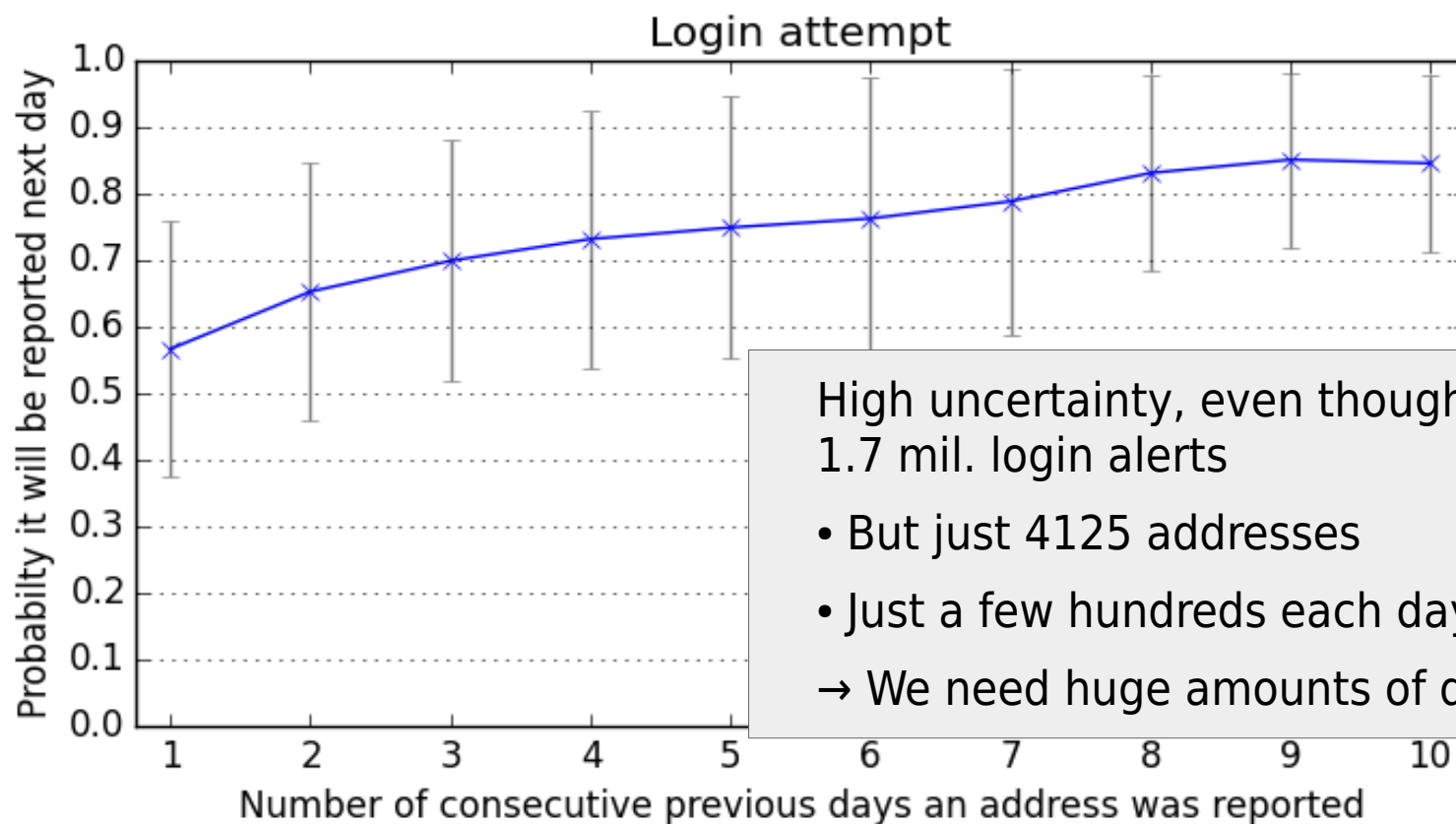
# Predictability

- If an address attacked previously, will it attack again?
  - $\text{Prob}(\text{reported at day } D \mid \text{reported at days } [D-n, \dots, D-1])$



# Predictability

- If an address attacked previously, will it attack again?
  - $\text{Prob}(\text{reported at day } D \mid \text{reported at days } [D-n, \dots, D-1])$



High uncertainty, even though we have 1.7 mil. login alerts

- But just 4125 addresses
- Just a few hundreds each day
- We need huge amounts of data



# Analysis summary

- There are many other ways the data can be analyzed
- Analyzed data comes from CESNET network only
  - Just a tiny part of the Internet

To provide more precise and more useful statistics

**we need more data ...**

# Analysis summary

- There are many other ways the data can be analyzed
- Analyzed data comes from CESNET network only
  - Just a tiny part of the Internet

To provide more precise and more useful statistics

**we need more data ...**

**... from you**

# Reputation Shield

- Reputation Shield (project within GÉANT GN4)
  - **Alert sharing** among multiple NRENs (based on Warden)
  - Centralised **analysis** of alert data
    - Aggregation, correlation, enhancement with information from external sources
    - Reputation modelling
      - List of bad actors + everything we know about them
- We are **looking for participants**
  - More data from more networks → better results
  - If you have any data to share, please contact me
- More information: <http://repsh.cesnet.cz>

# Thank you for your attention.

- Pavel Kácha: [ph@cesnet.cz](mailto:ph@cesnet.cz)
- Warden: <http://warden.cesnet.cz/>
- IDEA: <https://csirt.cesnet.cz/IDEA>
  
- Václav Bartoš: [bartos@cesnet.cz](mailto:bartos@cesnet.cz)
- Reputation Shield: <http://repsh.cesnet.cz>
  
- CESNET-CERTS: <https://csirt.cesnet.cz/>