Further Aspects of Passive DNS

Datamining, visualization and alternative implementations

Sebastien Tricaud (PicViz),
Alexandre Dulaunoy (CIRCL.lu),
L. Aaron Kaplan (CERT.at),
David Durvaux (CERT.be),
John Kristoff (Team Cymru)

June 19, 2012

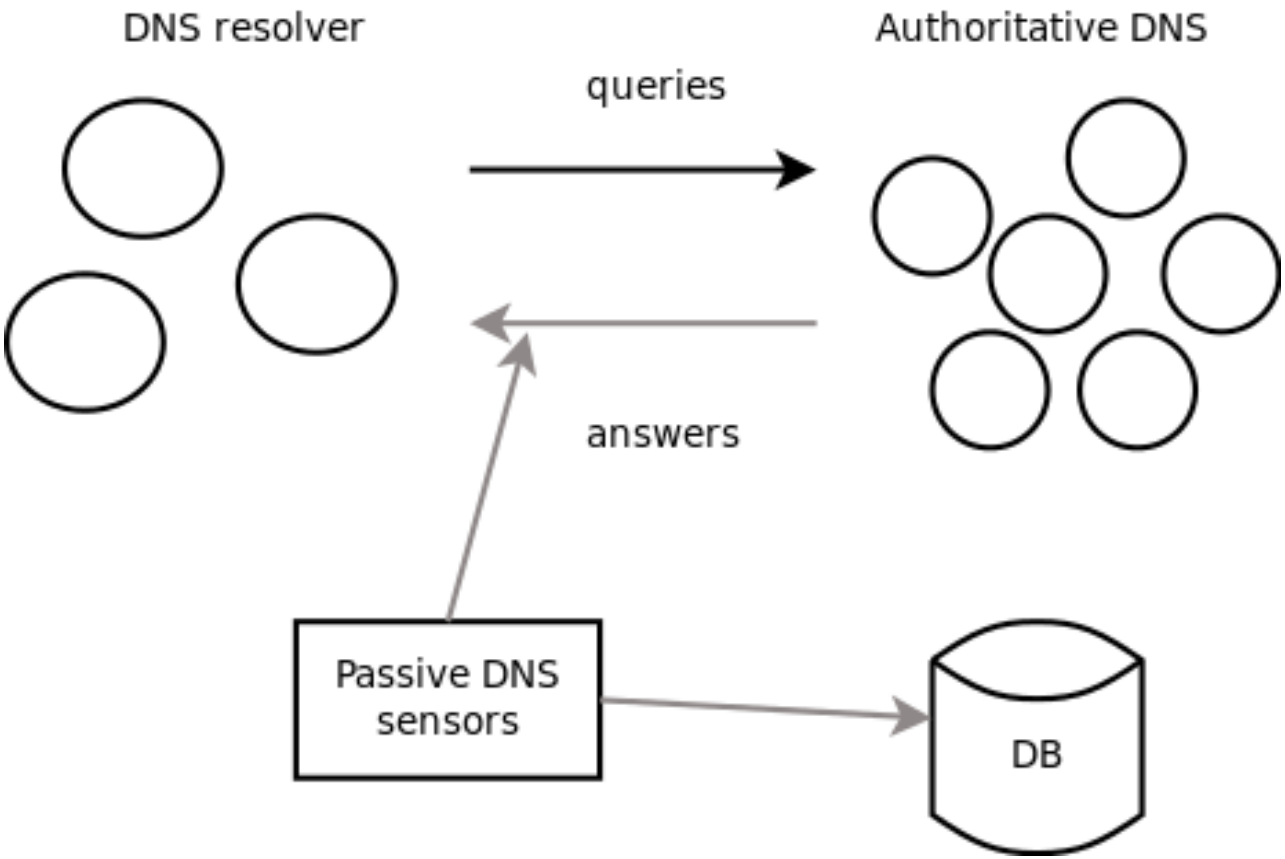
Disclaimer

- Passive DNS is a technique to collect only valid answers from authoritative or caching nameservers
- By its design, privacy is preserved (e.g. no source IP addresses from resolvers are captured[1])
- DNS data collected is only publicly known DNS data
- The research is done in the sole purpose to detect malicious IP/domains or content to better protect users
- Passive DNS implementations are subject to local rules

[1] Except if an application abuses DNS answers to track back their users.

"Passive DNS is to DNS ops as NetFlow is to net ops."
John Kristoff

Passive DNS - how it works



What's the purpose? Some examples...

Detection of shared compromised web hosting - the enisa.eu case

- Regularly malicious links are posted on compromised systems
- What are the other services or domains hosted on the same A/AAAA record?
- What happens to "infected" redirect (because the web hosting server is infected)?
- How Passive DNS can help?

EG (Egypt being offline)

- Discover non resolvable domains using nameserver in Egypt
- Interesting discovery *randomstring*.medicpills.ru ($\rightarrow$ less spam?)
- BIT.LY case is similar (when Libya was offline)
- Passive DNS helps to find interdependecies among services

Malware infection

- History of a domain name in conjunction with Netflow records
- Find shorted lived domain names
- Get back the A/AAAA records
- and find infected PCs in your Netflow.
- Quick win!

Passive DNS implementations

- BFK (F. Weimer[a]) passive dns
- **CIRCL pdns-toolkit**[b]
- **CERT.at passive dns**[c]
- CERT.ee passive dns

[a] Presented at FIRST 2005
[b] `github.com/adulau/pdns-toolkit/`
[c] access upon request

- CERT.lv passive dns
- ISC DNSDB[a]
- The University of Auckland DNS History Database Project (DHDB)
- Team Cymru passive dns

[a] `https://dnsdb.isc.org/`

Passive DNS design comparison - an ecosystem

	CIRCL pdns-toolkit	**CERT.at passive dns**
datastore	Redis	PostgreSQL + memcached
storage	memory	hybrid
exhaustive	-	+
space efficient	++	+
input	pcap, dnscap output	nmsg
open source	yes	ask

Some statistics from the CERT.at Passive DNS...

Storing Passive DNS - CIRCL.lu perspective

- Implementing the storage of a Passive DNS can be challenging
- Starting from standard RDBMS and then moved to a key-value store
- We learned to hate[2] hard disk drive and to love random access memory
- Loving memory is great especially when it's now cheap and addressable in 64bits

[2] exception → only used for data store snapshot

Passive DNS + Ranked domains - Where visualization can help

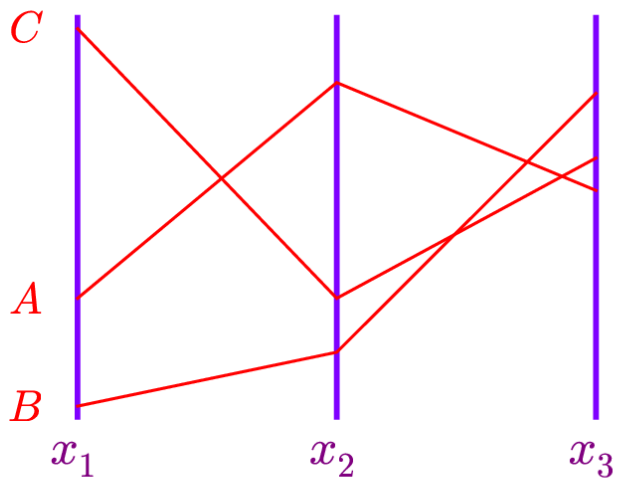
- Now, we have 50 millions lines of ranked hostname...

```
...
www.stopacta.info.
www.vista-care.com.
breadworld.com.
o-o.resolver.A.B.C.D.5xevqnwsds5zdq34.metricz.\
l.google.com.
www.thechinagarden.com.
smtp10.dti.ne.jp.
...
```

Why visualization?

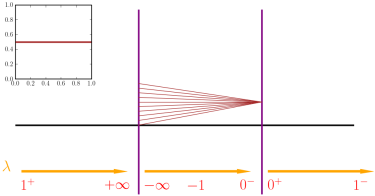
- Understand big data
- Find stuff we cannot guess

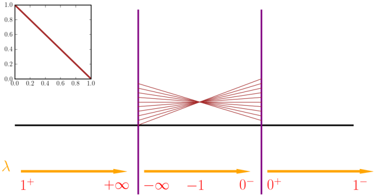
Choosing Parallel Coordinates

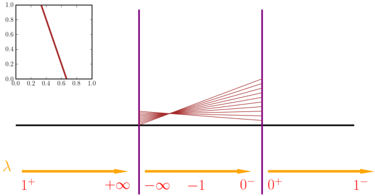


- Display as much dimensions wanted (yes, **as many**)
- Display as much data wanted (I mean it!)

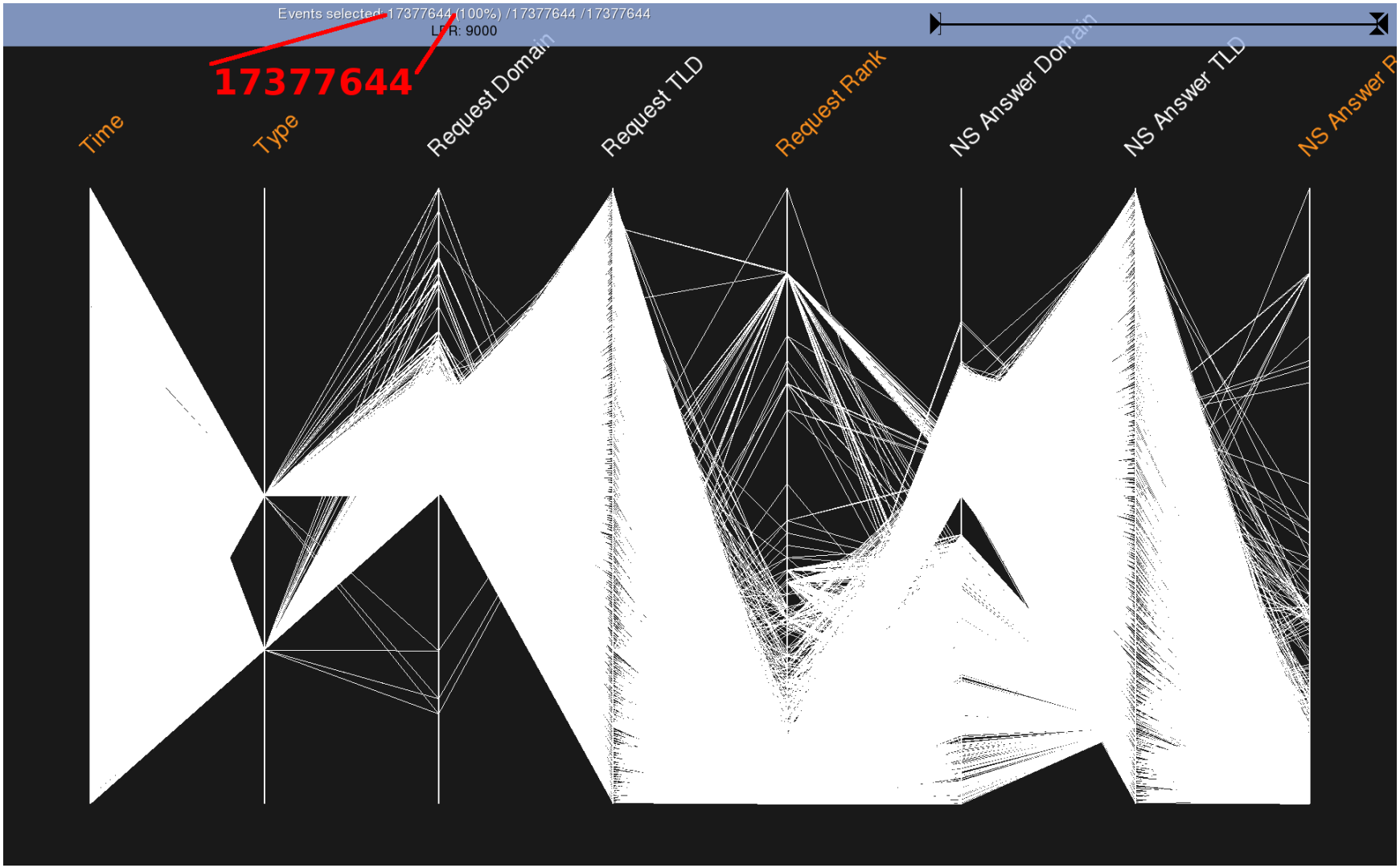
Interesting patterns



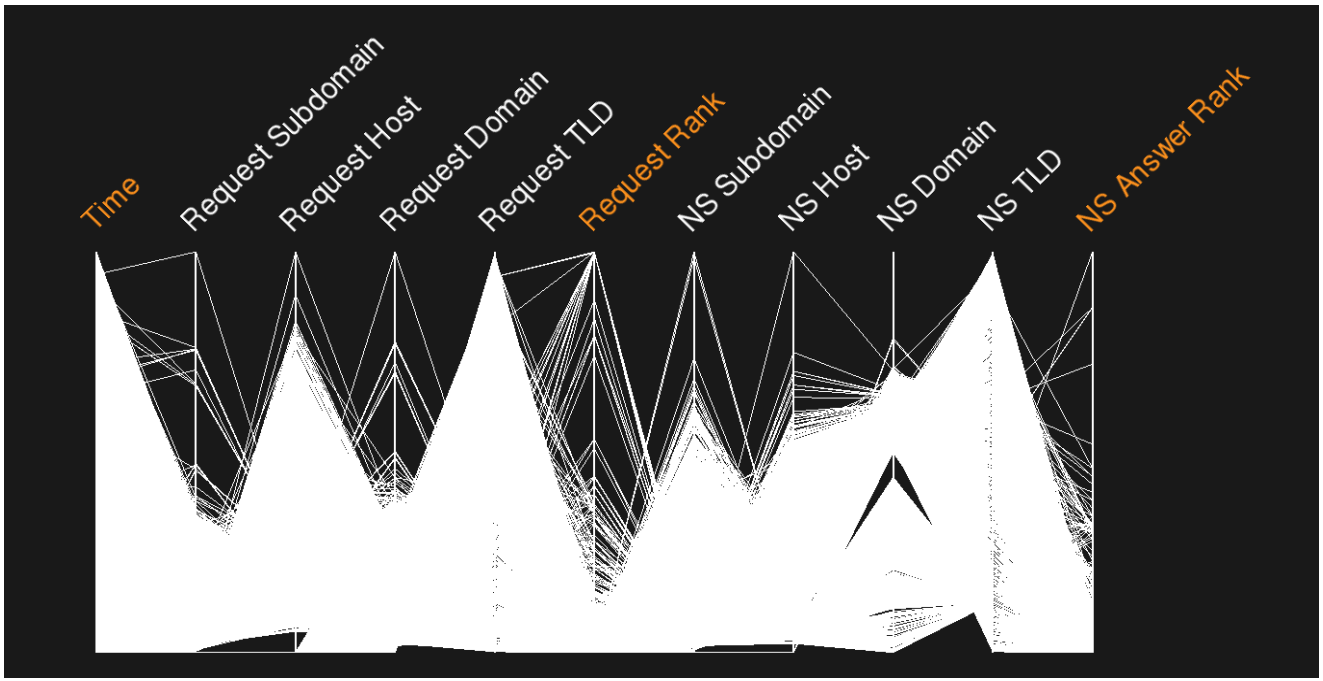




Picvizing a CIRCL passive DNS dataset



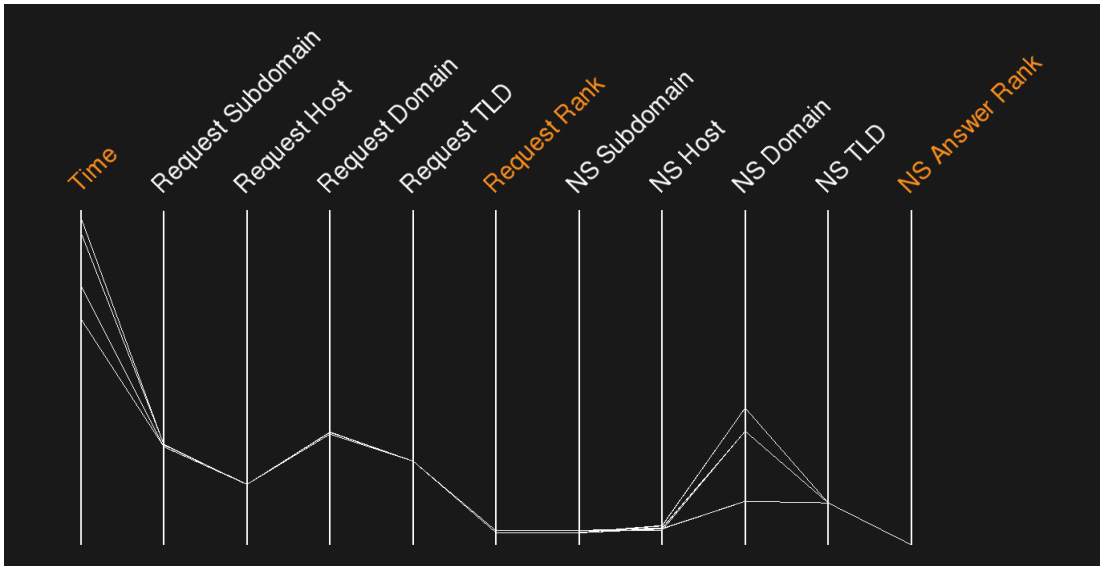
Picviz with subdomains split



Reward: highest is youtube

Subdomain entropy

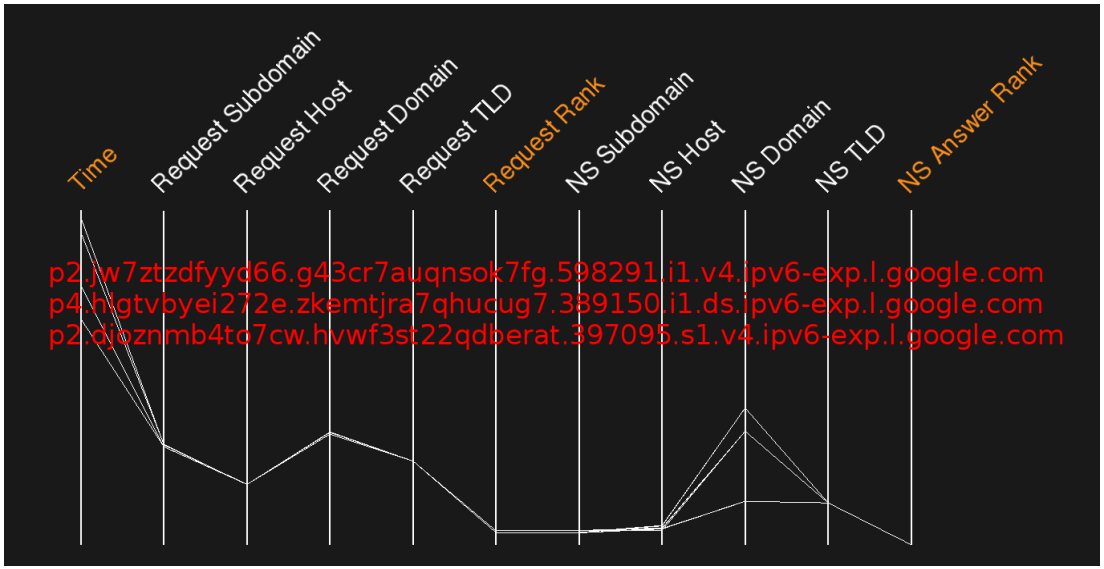
Only one sub-domain has an entropy[3] >4.8



[3] Shannon entropy

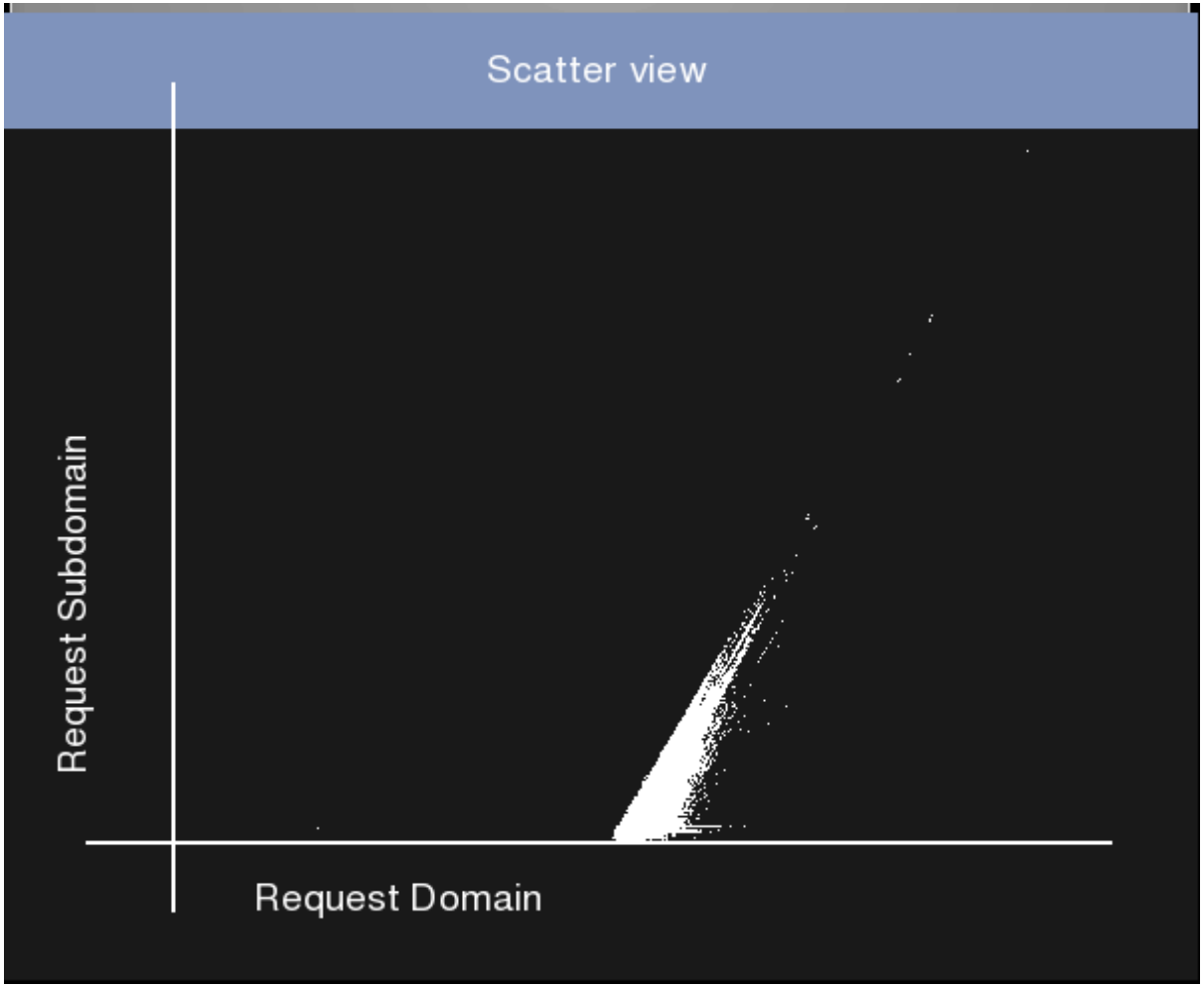
Subdomain entropy

Only one sub-domain has an entropy[4] >4.8

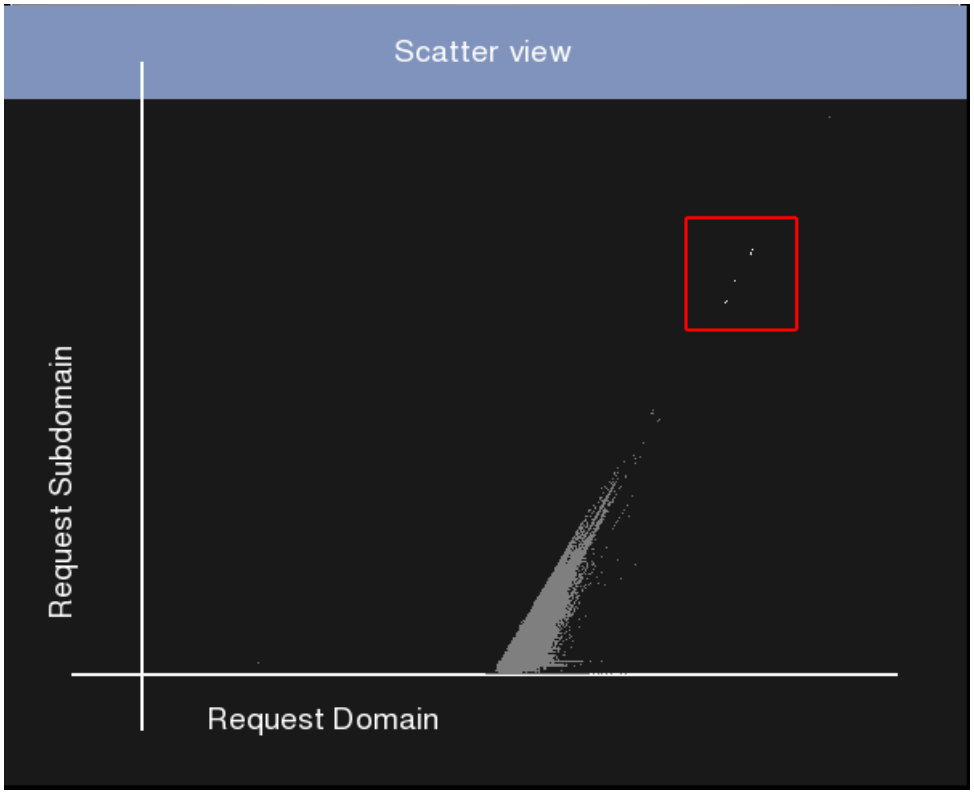


[4] Shannon entropy

Scatter plot - finding outliers



Scatter plot - finding outliers - covert channel?

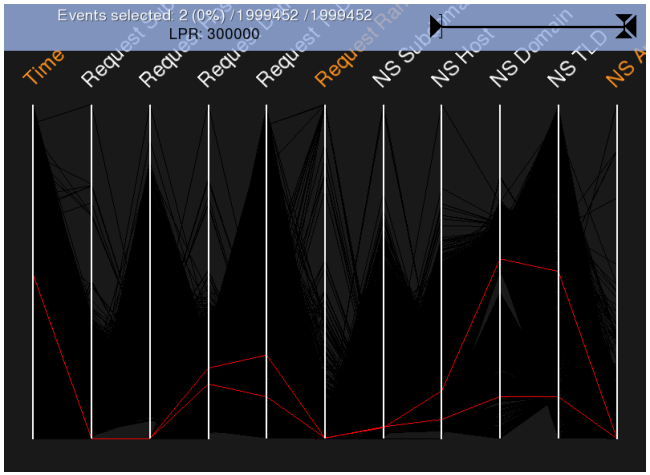


030066363663643937306531[..].36393764313333653763.lbl8.mailshell.net
t10000.u1318235395163.s203679668[..]-1329.zv6lit-null.zrdtd-1311.zr6td-null.results.potaroo.net
030030643038316639653 86[..].643065613438373465 33.lbl8.mailshell.net

Searching for Zeus

Using the broad Polish CERT regex

`[a-z0-9]{32,48}\.(ru|com|biz|info|org|net)`



- We get some cool domains:
  - cg79wo20kl92doowfn01oqpo9mdieowv5tyj.com
  - eef795a4eddaf1e7bd79212acc9dde16.net
- but more important we got a visualization profile to find outliers not matching the regexp

Conclusion

- Passive DNS is an infinite source of security data mining
- A team of passive DNS is at your services, contact us!
- (adequate) Visualization is an appropriate way to discover unknown malicious or suspicious services
- This finally helps CSIRTs to act earlier on the incidents
- Common output format for different implementations (work in progress)

Q&A

- alexandre.dulaunoy@circl.lu
- sebastien@honeynet.org
- kaplan@cert.at
- david.durvaux@cert.be
- jtk@cymru.com