DREAMLAB
TECHNOLOGIES

Dreamlab Technologies

Taking a look at the *.ch zone with a DGA detector

2021

Shout out

- Jeroen van Meeuwen (Kolabnow.com)
- Sinan Sekerci (Dreamlab.net)

> whoami

- Mischa Obrecht, 34
- IT-Security since 10 years, analyst, engineer, project / security manager, tester
- Pentester and project manager for Dreamlab
- Involved in the SAPPAN-project as project coordinator for Dreamlab

To get in touch please use:

- mischa.obrecht@dreamlab.net

Agenda

[-] Introduction

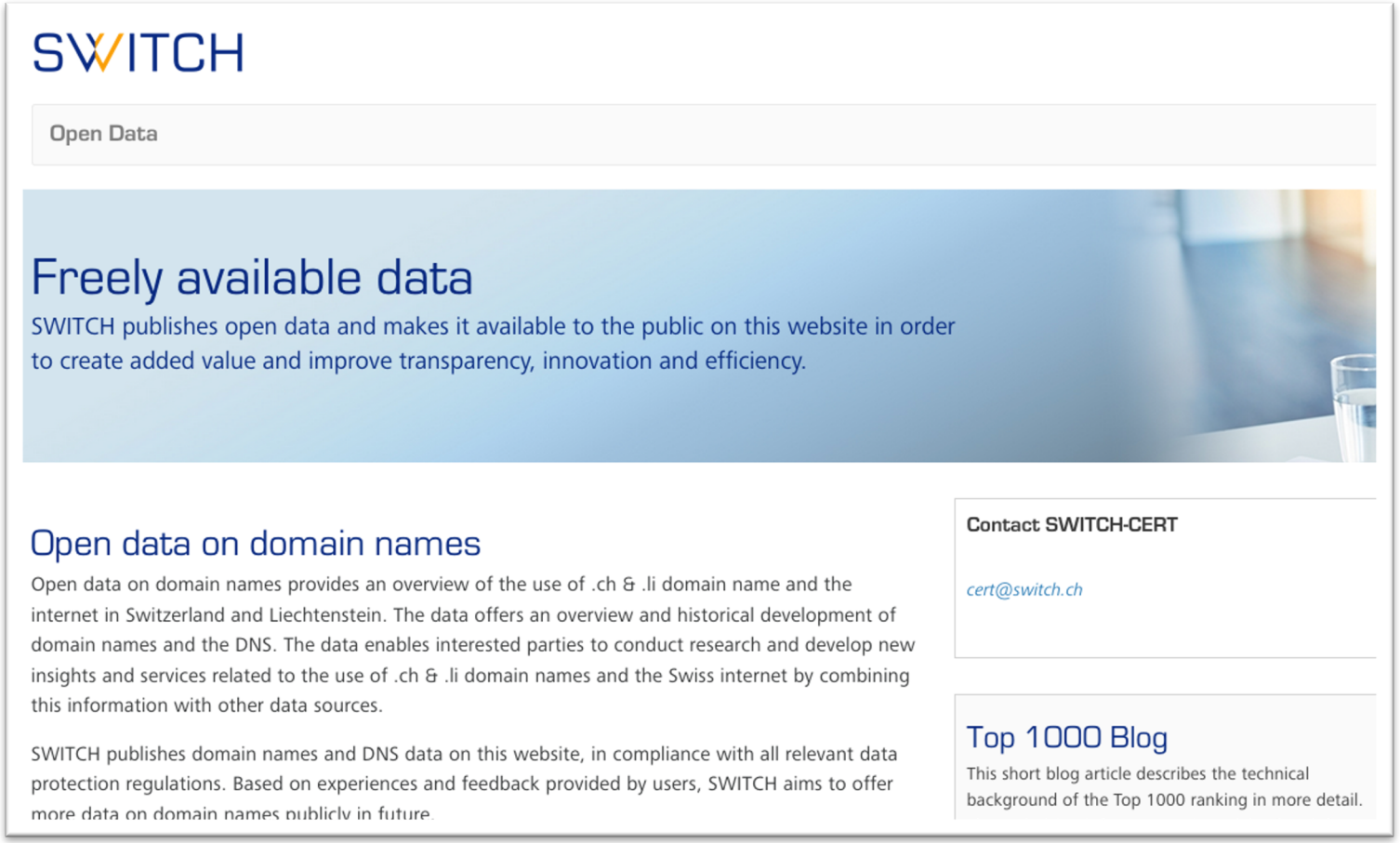
- *.ch Zonefile
- Domain Generation Algorithms
- NN DGA classifiers

[+] Methodology

[+] Results

Introduction

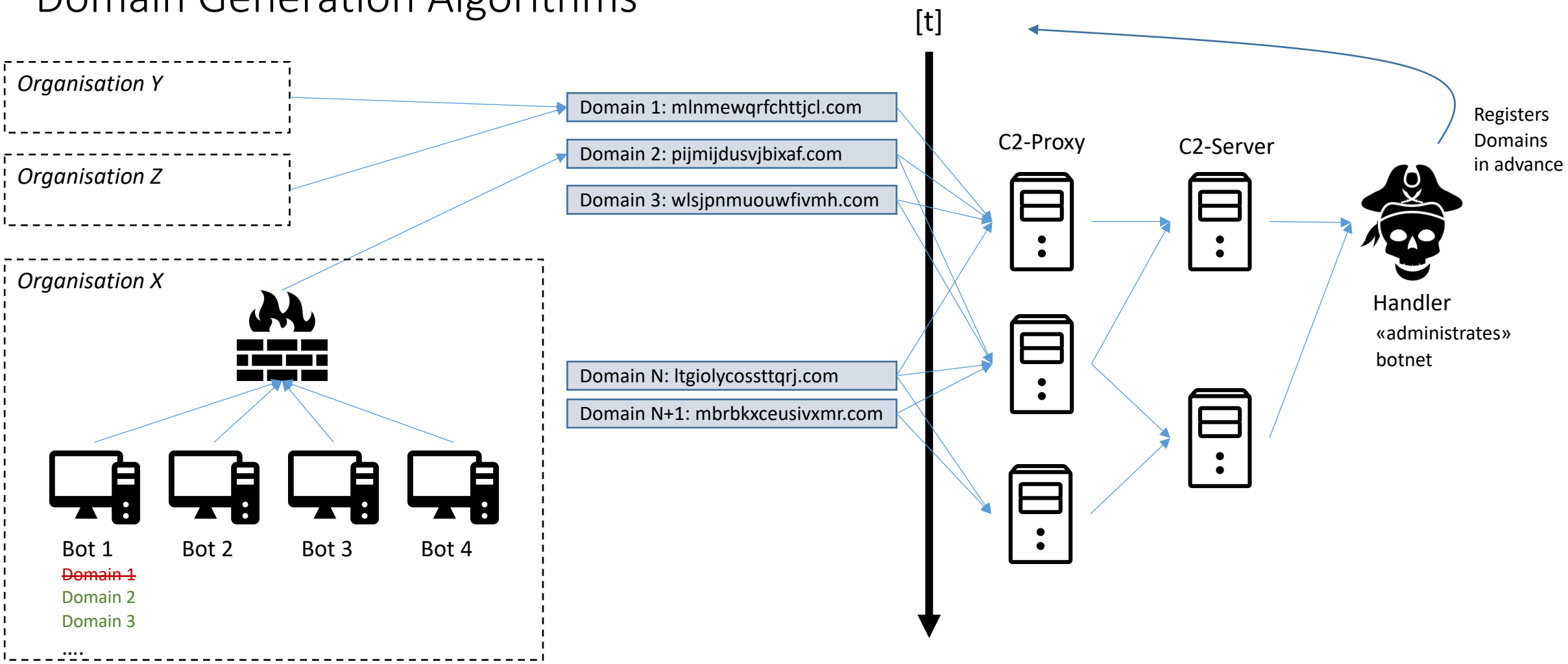
switch.ch publishes *.ch and *.li zonefiles



2260606
Entries for .ch

Introduction

Domain Generation Algorithms



➔ Bin Yu, Jie Pan, Jiaming Hu, Anderson Nascimento, and Martine De Cock. 2018.Character Level Based Detection of DGA Domain Names. InInternational JointConference on Neural Networks. IEEE, https://faculty.washington.edu/mdecock/papers/byu2018a.pdf

Introduction

(n-ary) NN based DGA detectors are investigated in context of SAPPAN-

Analyzing the Real-World Applicability of DGA Classifiers

Arthur Drichel
RWTH Aachen University
drichel@itsec.rwth-aachen.de

Ulrike Meyer
RWTH Aachen University
meyer@itsec.rwth-aachen.de

Samuel Schüppen
Siemens CERT
samuel.schueppen@siemens.com

Dominik Teubert
Siemens CERT
dominik.teubert@siemens.com

ABSTRACT

Separating benign domains from domains generated by DGAs with the help of a binary classifier is a well-studied problem for which promising performance results have been published. The corresponding multiclass task of determining the exact DGA that generated a domain enabling targeted remediation measures is less well studied. Selecting the most promising classifier for these tasks in practice raises a number of questions that have not been addressed in prior work so far. These include the questions on which traffic to train in which network and when, just as well as how to assess ro-

1 INTRODUCTION

Bots need to be able to establish a conne
and control (C2) server in order to obtain u
this end, they often rely on domain genera
which generate a large amount of pseudo
using a seed. The botnet herder knows the
predict the algorithmically generated dom
and to register a small subset of these dom
AGDs one-by-one trying to obtain a vali
server. The majority of these queries result

Table 2: Binary Classification: Mixed DGAs

Classifier	ACC	TPR	TNR	FNR	FPR
FANCI	0.99764	0.99744	0.99784	0.00256	0.00216
B-Endgame	0.99891	0.99969	0.99813	0.00031	0.00187
B-NYU	0.99907	0.99976	0.99838	0.00024	0.00162
B-ResNet	0.99916	0.99978	0.99853	0.00022	0.00147

Introduction

```
$head malicious.txt
nymaim  chvkhnfqisg.net
nymaim  fnqqhplfm.com
nymaim  mwjvbwj.info
nymaim  ridfga.com
nymaim  xzzwrbnhd.com
nymaim  znrijgzfg.com
nymaim  guqthb.com
```

Goal: Find malicious/untrustworthy domains in .ch-zone

Assumption: DGA classifiers introduced above can be used to detect malicious / Untrustworthy domains

Available data:

- Available datasets for DGA samples mostly feature *.com, *.org, *.biz, “.info, etc.
- There is no list of DGA samples for .ch-TLD, we can:
  - Create our own
  - Try to generalize (or rather specialize) from existing lists

➔What will we find applying classifiers that were trained on ”generic” data to the *.ch zone?

➔Can we find a self-contained way to specialize from existing lists and add some locality to the classifier?

Agenda

[+] Introduction

[+] Methodology

[+] Results

Methodology

Model 1 (B-NYU)

Model 2 (B-ResNet)

Training Iteration 1

- Benign Training-Set: Cisco[1] Popularity List Top 1m
- Malicious Training-Set: Netlab OpenData Project's DGA Domain List[2]

Analysis Iteration 1

- Test Dataset: *.ch-Zonefile

Training Iteration 2

- Benign Training-Set: Cisco Popularity List Top 1m + Part of *.ch Zonefile with existing MX-record
- Malicious Training-Set: Netlab OpenData Project's DGA Domain List

Analysis Iteration 2

- Test Dataset: *.ch-Zonefile
- Manual verification of results

[1]Cisco Umbrella Project - https://s3-us-west-1.amazonaws.com/umbrella-static/index.html
[2]https://data.netlab.360.com/dga

Methodology
Metrics on Datasets

Dataset	$ wc -l
Cisco Popularity List Top 1m	999’999
Netlab DGA Domain List	1’369’253
*.ch Zonefile	2’260’606
*.ch-Domains with MX-records	1’640’145

```
$ grep ".ch" cisco-benign.txt | wc –l
51567

$ grep ".ch" ch.zone.txt | wc -l
2260606
```

Agenda

[+] Introduction

[+] Methodology

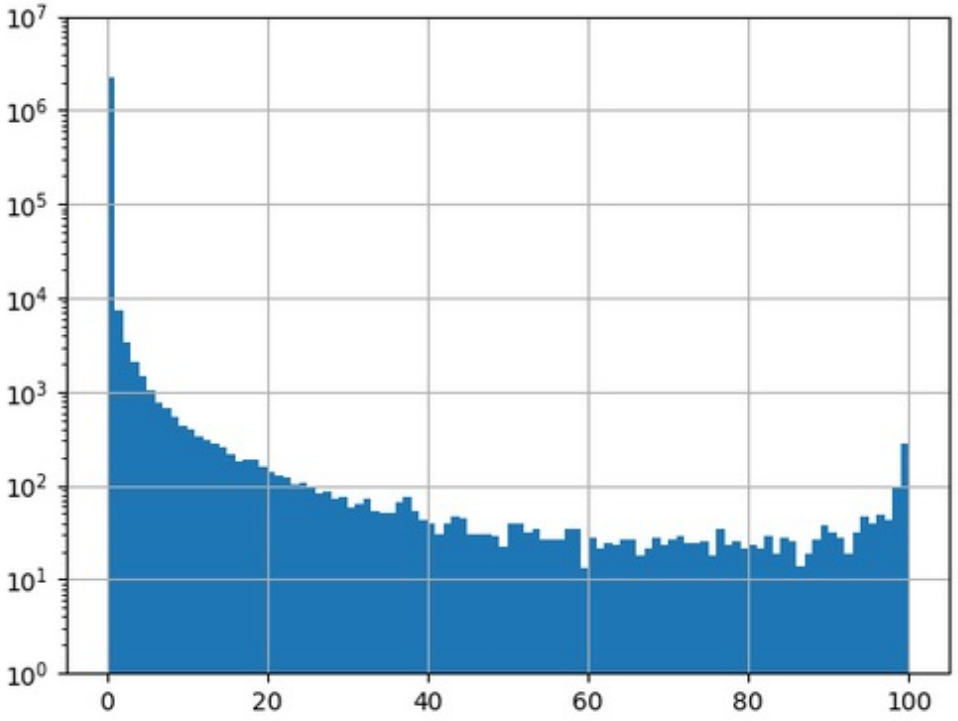
[-] Results

- Analysis Iteration 1
- Analysis Iteration 2
- Conclusion / Outlook

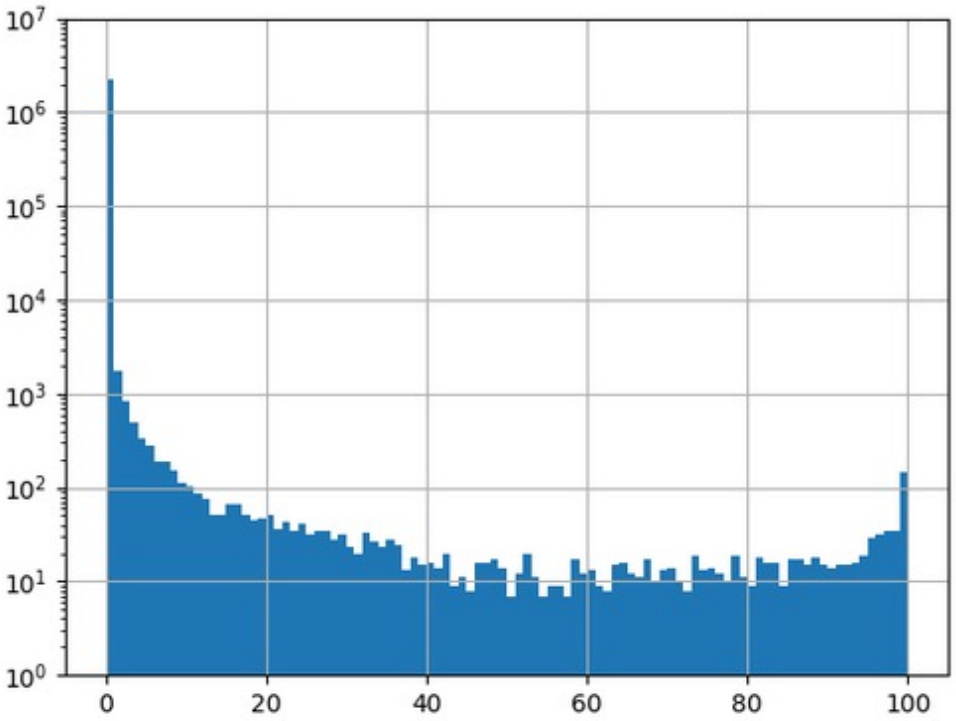
Results

Analysis Iteration 1 – Distributions of results

B-NYU Distribution:
(0, 25) 2257416
(25, 50) 1347
(50, 75) 675
(75, 100) 1024



B-ResNet Distribution:
(0, 25) 2259070
(25, 50) 524
(50, 75) 299
(75, 100) 569



Results

Analysis Iteration 1 – Obvious False Positives

B-NYU Distribution:
(0, 25) 2257416
(25, 50) 1347
(50, 75) 675
(75, 100) 1024

Has MX Record	NO MX Record
08d749013a84d810edacd03c338e2b2e.ch	0767676767.ch
19zweiundsiebzig.ch	1augustbrunch.ch
1nf02m4c0253845714.ch	2millefeuille.ch
365prosicherheitsmanufaktur.ch	3nkh5crxol.ch
4897055461203223.ch	4aumf9m40jpe.ch
50jahrefcteufen.ch	59dcb8ad64e5d3b35fb590.ch
6822beb0714d4417a5090fddaec2fa5f.ch	6qs8d6zd3d69s6.ch
aachmuehle.ch	7182818284590452353602874713526624977572 47.ch

…

B-ResNet Distribution:
(0, 25) 2259070
(25, 50) 524
(50, 75) 299
(75, 100) 569

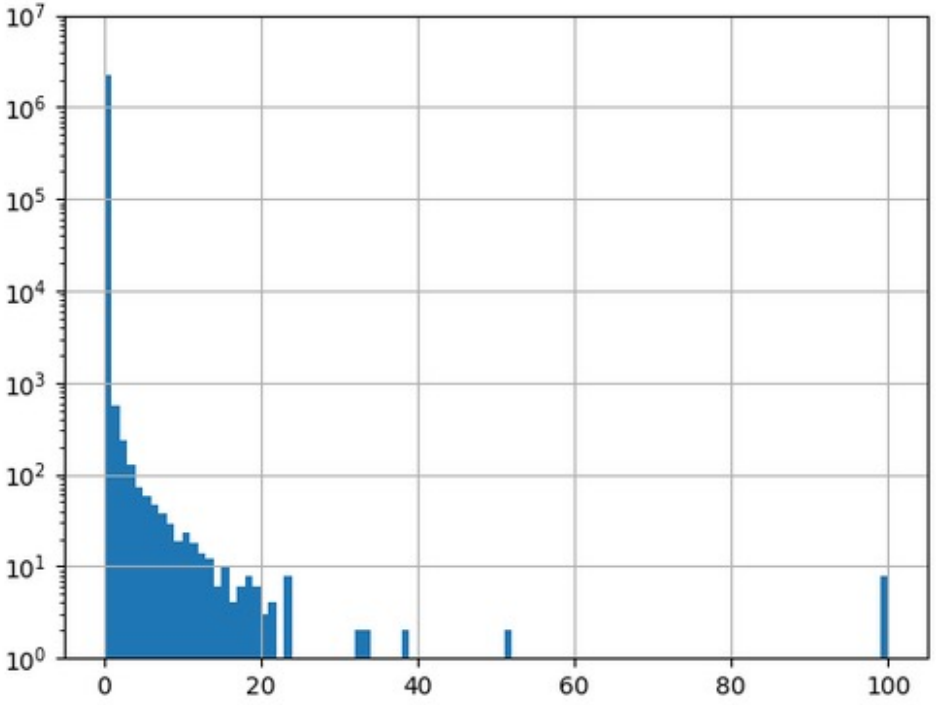
Has MX Record	NO MX Record
08d749013a84d810edacd03c338e2b2e.ch	00000000019d6689c085ae165831e934ff763ae46a2a6c
19zweiundsiebzig.ch	125jahrefeuerwehrverbandluzern.ch
1nf02m4c0253845714.ch	3epilier-assurance.ch
365prosicherheitsmanufaktur.ch	3epilier-prevoyance.ch
4897055461203223.ch	4aumf9m40jpe.ch
50jahrefcteufen.ch	6qs8d6zd3d69s6.ch
6822beb0714d4417a5090fddaec2fa5f.ch	7182818284590452353602874713526624977572 47.ch
aachmuehle.ch	7hkjuf95uenl.ch
aaeeiioouu.ch	8qswldnsrvb73xkczdyj.ch

…

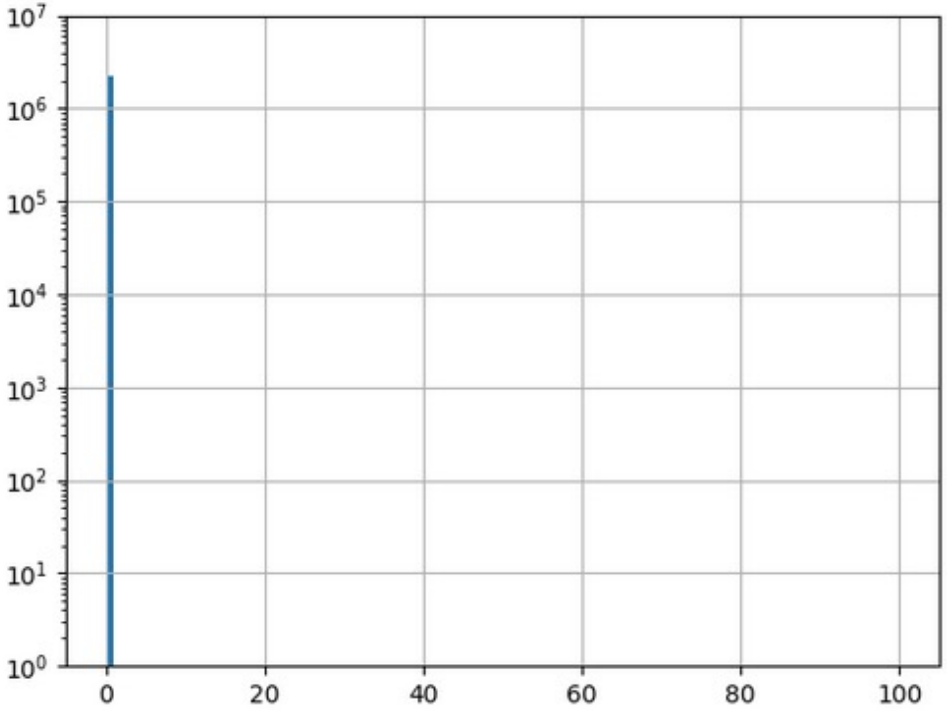
Results

Analysis Iteration 2 – Distributions of results

B-NYU Distribution:
(0, 25) 2260430
(25, 50) 13
(50, 75) 6
(75, 100) 12



B-ResNet Distribution:
(0, 25) 2260462
(25, 50) 0
(50, 75) 0
(75, 100) 0



Results

Analysis Iteration 2 – Interesting detections 1/2

Domain	Certainity	Analysis / Conclusion
abcdefghijklmnopqrstuvwxyz.ch	100%	**Likely malicious**, currently resolves to suspicious IP
adslkfalkfjlkfjdsalkfafljflsa.ch	100%	**Unclear**, no IP resolution
8qswldnsrvb73xkczdyj.ch	99.9%	**Likely malicious**, Currently resolves to suspicious IP, used to resolve to another suspicious IP involved in coronavirus-scam
rgdfgdfgdfgdf.ch	99.9%	**Likely benign**, no suscpicious observations
utitan101310bgfhnythjdukfdyjt.ch	99.8%	**Likely benign**, no suscpicious observations
sfdfgdfgdfgdfgdfg.ch	99.8%	**Unclear**, no IP resolution
n7q9ipiddq9ihtx.ch	99.1%	**Likely malicious**, currently resolves to suspicious IP
testhgfjdgdfxhgxdfhx12.ch	99.1%	**Likely malicious**, currently resolves to suspicious IP, IP has been flagged as botnet C2
oiqweurpui345345jk.ch	94.1%	**Likely benign**, no suscpicious observations
ymfvrcnwyw.ch	92.5%	**Unclear**, no IP resolution
aqddddwxszedc.ch	84.8%	**Unclear**, no IP resolution
ihjj8qltfyfe.ch	82.2%	**Likely malicious**, domain flagged by Kaspersky, currently resolves to suspicious cloudflare IP
asdfjkhdsfajdfsajhsadf.ch	77.1%	**Likely benign**, no suscpicious observations
7as6q796d6s98q6qd6sdq.ch	72.6%	**Likely malicious**, Currently resolves to suspicious IP
rggrgrgrgrgrgr.ch	66.5%	**Unclear**, no IP resolution

Results

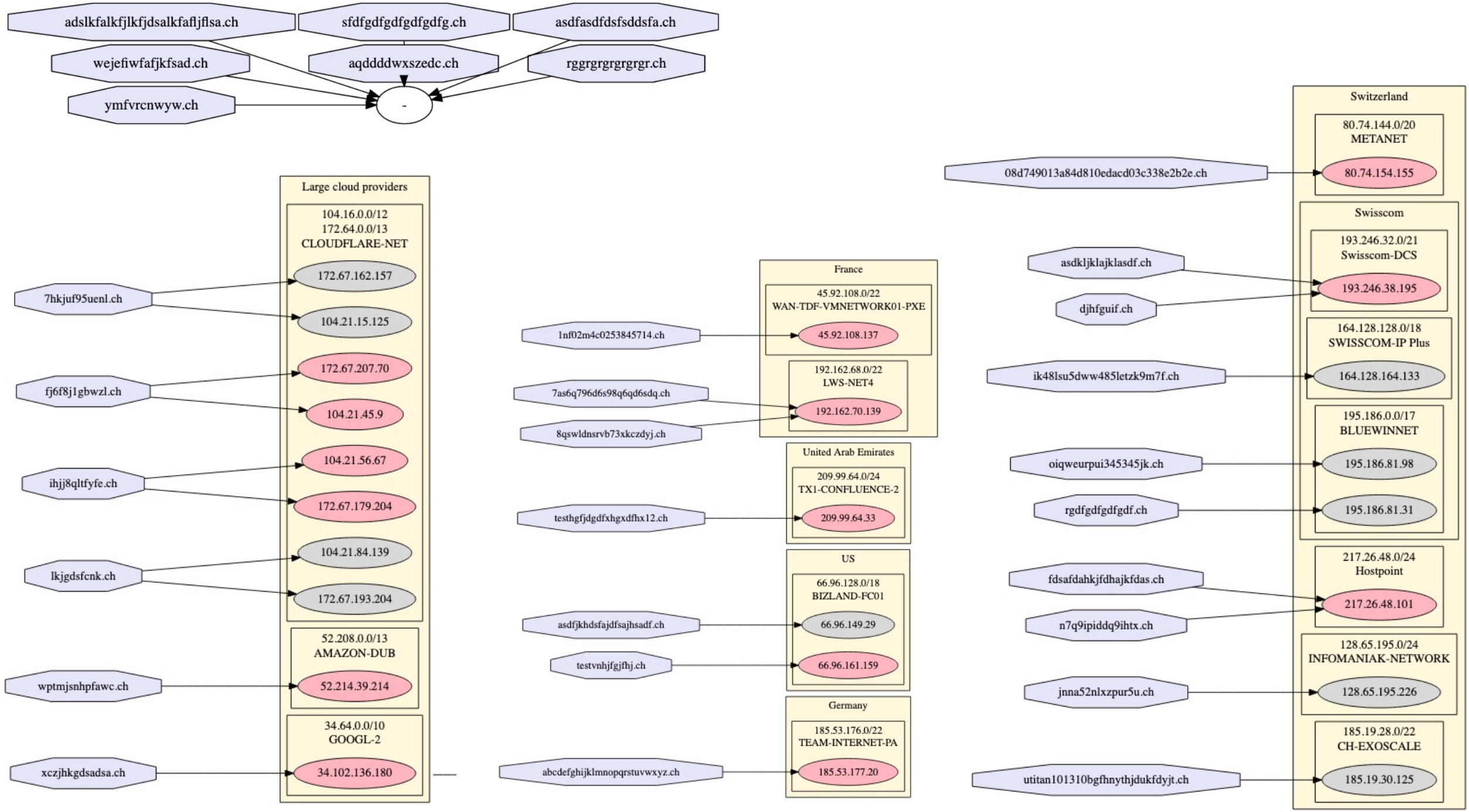
Analysis Iteration 2 – Interesting detections 2/2

Domain	Certainity	Analysis / Conclusion
fj6f8j1gbwzl.ch	54.6%	**Likely malicious**, used to resolve to suspicious cloudflare IP (with very little associated URLs)
fdsafdahkjfdhajkfdas.ch	52.2%	**Likely malicious**, Currently resolves to suspicious IP
xczjhkgdsadsa.ch	51.3%	**Likely malicious**, Currently resolves to suspicious IP
ik48lsu5dww485letzk9m7f.ch	51.1%	**Likely benign**, no suscpicious observations
asdkljklajklasdf.ch	48.3%	**Likely malicious**, Currently resolves to suspicious IP
1nf02m4c0253845714.ch	47.9%	**Likely malicious**, Currently resolves to suspicious IP
jnna52nlxzpur5u.ch	43.4%	**Likely benign**, no suscpicious observations
wptmjsnhpfawc.ch	38.9%	**Likely malicious**, Currently resolves to suspicious IP
wejefiwfafjkfsad.ch	38.8%	**Unclear**, no IP resolution
asdfasdfdsfsddsfa.ch	33.8%	**Unclear**, no IP resolution
testvnhjfgjfhj.ch	33.7%	**Likely malicious**, Currently resolves to suspicious IP
08d749013a84d810edac….e2b2e.ch	32.8%	**Likely malicious**, Currently resolves to suspicious IP
djhfguif.ch	32.6%	**Likely malicious**, Currently resolves to suspicious IP
7hkjuf95uenl.ch	30.1%	**Likely benign**, no suscpicious observations
lkjgdsfcnk.ch	28.8%	**Likely benign**, no suscpicious observations
neue-webermuehle-neuenhof.ch	25.9%	**Likely benign**, no suscpicious observations

Results

Analysis Iteration 2 – Conclusion

Certaininty of URL being malicious	Number of URLs	Verification result
75%-100%	13	5 urls are likely malicious, 4 urls do not resolve, 4 urls are likely benign, all urls do not "look" trustworthy
50%-75%	6	4 urls are likely malicious, 1 urls do not resolve, 1 urls are likely benign, all urls do not "look" trustworthy
25%-50%	12	6 urls are likely malicious, 2 urls do not resolve, 4 urls are likely benign, all but one urls do not ”look” trustworthy
0%-25%	2260430	not inspected


adslkfalkfjlkfjdsalkfafljflsa.ch
sfdfgdfgdfgdfgdfg.ch
asdfasdfdsfsddsfa.ch
wejefiwfafjkfsad.ch
aqddddwxszedc.ch
rggrgrgrgrgrgr.ch
ymfvrcnwyw.ch
-
Large cloud providers
104.16.0.0/12
172.64.0.0/13
CLOUDFLARE-NET
172.67.162.157
104.21.15.125
172.67.207.70
104.21.45.9
104.21.56.67
172.67.179.204
104.21.84.139
172.67.193.204
52.208.0.0/13
AMAZON-DUB
52.214.39.214
34.64.0.0/10
GOOGL-2
34.102.136.180
7hkjuf95uenl.ch
fj6f8j1gbwzl.ch
ihjj8qltfyfe.ch
lkjgdsfcnk.ch
wptmjsnhpfawc.ch
xczjhkgdsadsa.ch
France
45.92.108.0/22
WAN-TDF-VMNETWORK01-PXE
45.92.108.137
192.162.68.0/22
LWS-NET4
192.162.70.139
1nf02m4c0253845714.ch
7as6q796d6s98q6qd6sdq.ch
8qswldnsrvb73xkczdyj.ch
United Arab Emirates
209.99.64.0/24
TX1-CONFLUENCE-2
209.99.64.33
testhgfjdgdfxhgxdfhx12.ch
US
66.96.128.0/18
BIZLAND-FC01
66.96.149.29
66.96.161.159
asdfjkhdsfajdfsajhsadf.ch
testvnhjfgjfhj.ch
Germany
185.53.176.0/22
TEAM-INTERNET-PA
185.53.177.20
abcdefghijklmnopqrstuvwxyz.ch
Switzerland
80.74.144.0/20
METANET
80.74.154.155
08d749013a84d810edacd03c338e2b2e.ch
Swisscom
193.246.32.0/21
Swisscom-DCS
193.246.38.195
asdkljklajklasdf.ch
djhfguif.ch
164.128.128.0/18
SWISSCOM-IP Plus
164.128.164.133
ik48lsu5dww485letzk9m7f.ch
195.186.0.0/17
BLUEWINNET
195.186.81.98
195.186.81.31
oiqweurpui345345jk.ch
rgdfgdfgdfgdf.ch
217.26.48.0/24
Hostpoint
217.26.48.101
fdsafdahkjfdhajkfdas.ch
n7q9ipiddq9ihtx.ch
128.65.195.0/24
INFOMANIAK-NETWORK
128.65.195.226
jnna52nlxzpur5u.ch
185.19.28.0/22
CH-EXOSCALE
185.19.30.125
utitan101310bgfhnythjdukfdyjt.ch

Conclusion / What's next?

Conclusion:

- Applying classifiers based on generic training data to a language specific zone results in too many false positives
- Extending the training dataset with MX-domains helps to specialize a generic list of benign and malicious domains for a given country-TLD, this can be done in a self contained way
- The CNN based binary classifier works well for identifying suspiciously-looking domains, more than 50% of which are very likely malicious, thresholding remains tricky

Outlook:

- Sweden also publishes zonefile ➔ repeat analysis to confirm results
  - Initial results indicate, that method seems to work even better in the case of .se-TLD
- Further improve accuracy by using more features


Thank
You!