DGArchive

A deep dive into domain generating malware

Daniel Plohmann

daniel.plohmann@fkie.fraunhofer.de

2015-12-03 | Botconf, Paris

Fraunhofer
FKIE

About me

- **Daniel Plohmann**
  - PhD candidate at University of Bonn, Germany
  - Security Researcher at Fraunhofer FKIE
  - Focus: Reverse Engineering / Malware Analysis / Automation

- **Projects**
  - ENISA Botnet Study 2011 [1]
  - Analysis Tools
    - PyBox, IDAscope, DGArchive, …
  - Botnet Analysis
    - Gameover Zeus / P2P protocols [2]
    - DGA-based Malware

[1] http://www.enisa.europa.eu/act/res/botnets/botnets-measurement-detection-disinfection-and-defence
[2] http://christian-rossow.de/publications/p2pwned-ieee2013.pdf

Agenda

- Intro: Domain Generation Algorithms / DGArchive
- Comparison of DGA Features
- Registration Status of DGA Domain Space
- Case Studies

Intro

Domain Generation Algorithms

Domain Generation Algorithms

Definitions

- Concept first described ~2008: Domain Flux
- Domain Generation Algorithm (DGA)
  - An algorithm producing Command & Control rendezvous points dynamically
  - Shared secret between malware running on compromised host and botmaster
- Seeds
  - Collection of parameters influencing the output of the algorithm
- Algorithmically-Generated Domain (AGD)
  - Domains resulting from a DGA

Domain Generation Algorithms

Origin & History

- Feb 2006 Sality: dynamically generates 3rd-level domain part
- July 2007 Torpig: Report by Verisign includes DGA-like domains
- July 2007 Kraken: VirusTotal upload of binary using DDNS

- April 2008 Kraken DGA first publicly mentioned
- 3 big events in November 2008 – April 2009:
  - Szribi: Takedown, but botmaster regained control through DGA
  - Conficker: Well, you probably know about that one…
  - Torpig: DGA-based takeover

http://www.mcafee.com/threat-intelligence/malware/default.aspx?id=138354
http://fserror.com/pdf/Torpig.pdf
https://isc.sans.edu/forums/diary/Kraken+Technical+Details+UPDATED+x3/4256/
https://www.fireeye.com/blog/threat-research/2008/11/technical-details-of-srizbis-domain-generation-algorithm.html
https://seclab.cs.ucsb.edu/media/uploads/papers/torpig.pdf

Domain Generation Algorithms

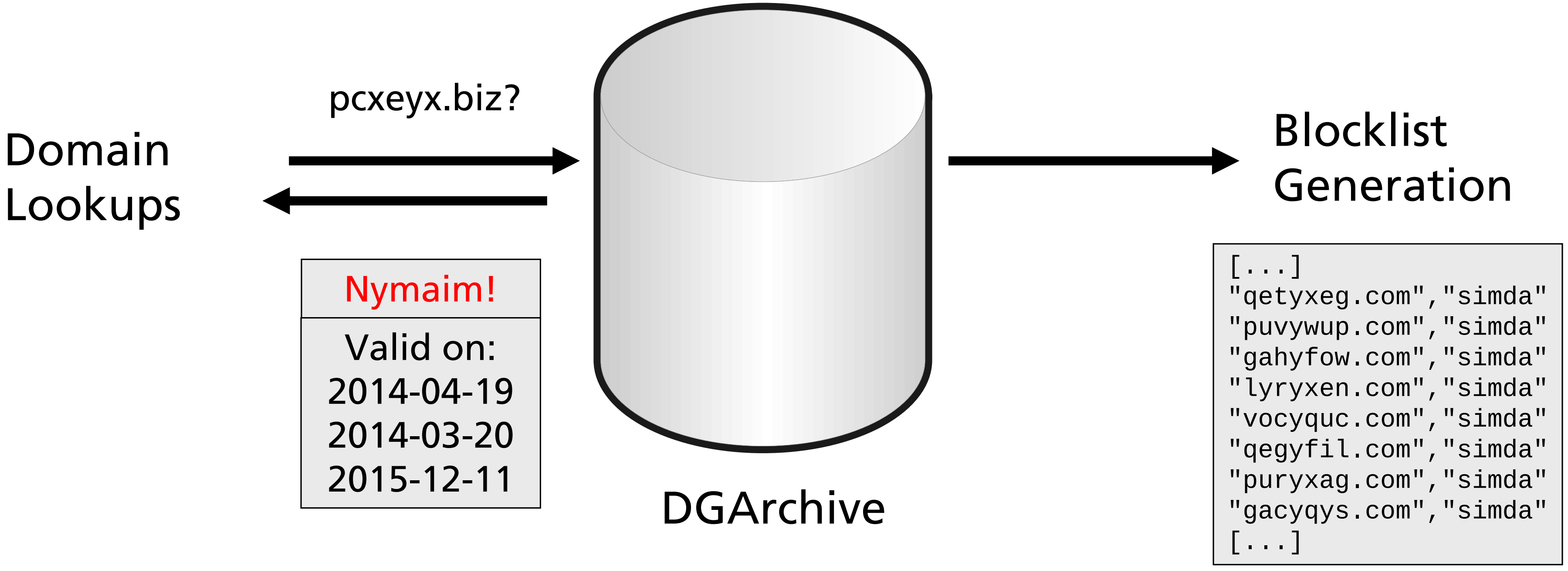
Motivation for Usage

- Aggravation of Analysis
  - No hardcoded domains / dumping -> code analysis needed
- Evasion
  - Many DGAs have short-lived domains -> avoid domain reputation
- Backup
  - Registration only when needed
- Asymmetry
  - Attacker needs one domain, defender needs to prohibit access to all
- Feasability of DGAs
  - Domains are cheap (compared to profits)

DGArchive

The idea

- **DGAs are annoying! :(**
- Idea:
  - Reverse DGA, then generate and archive all possible domains since first spotting of a malware family



Domain Lookups

pcxeyx.biz?

Nymaim!

Valid on:
2014-04-19
2014-03-20
2015-12-11

DGArchive

Blocklist Generation

```
[...]
"qetyxeg.com","simda"
"puvywup.com","simda"
"gahyfow.com","simda"
"lyryxen.com","simda"
"vocyquc.com","simda"
"qegyfil.com","simda"
"puryxag.com","simda"
"gacyqys.com","simda"
[...]
```

DGArchive

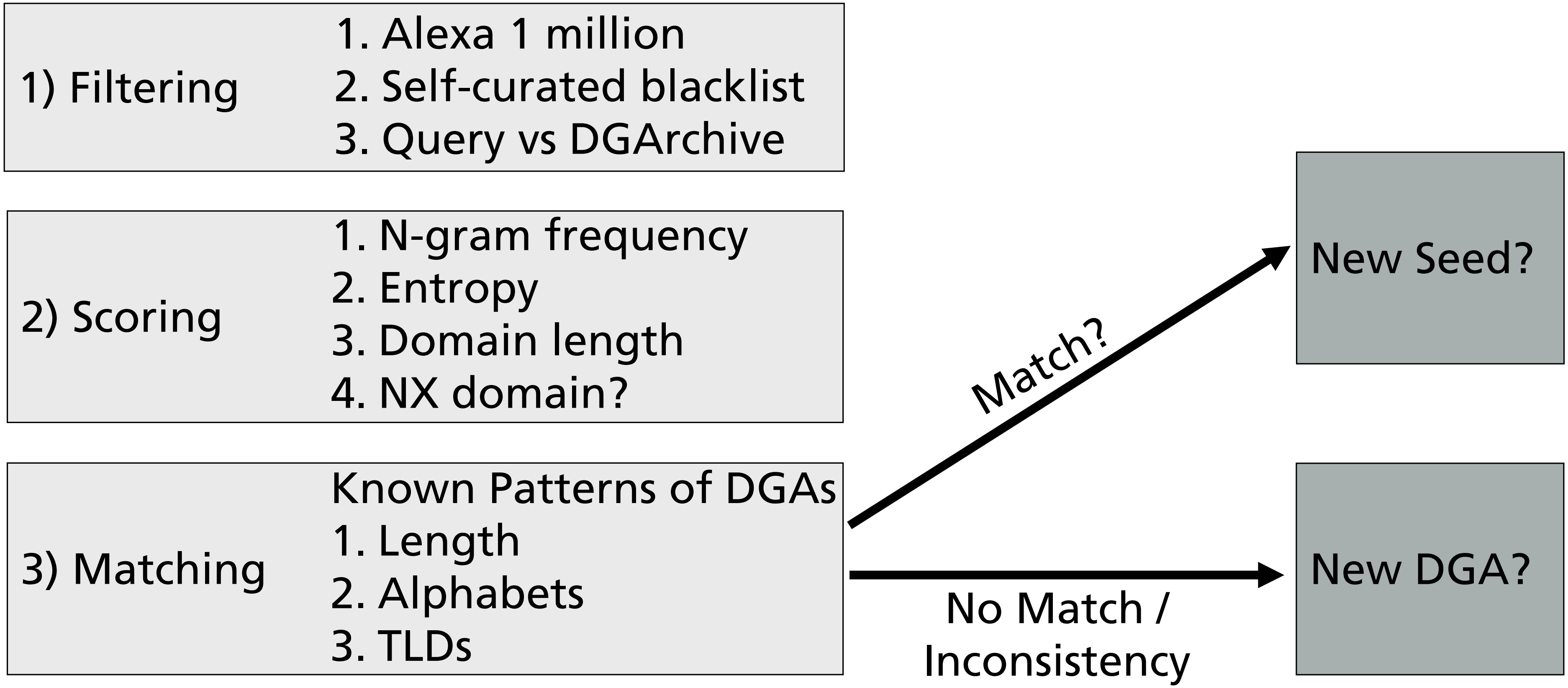
Status

- Botconf 2014: Lightning talk
  - 8 families, ~20 seeds, ~ 4 million domains
- DGArchive Today
  - 43 families/variants, ~280 seeds, 20+ million domains

Finding DGAs

Mining a Sandbox DNS feed

- Remix of academic approaches and common sense
  - Input: List of domains, queried during a sandbox run
  - DNS Feed by shadowserver ←THANK YOU!!!
  - 1,235,443 sandbox runs; 15,660,256 DNS queries (959,607 unique)



1) Filtering
1. Alexa 1 million
2. Self-curated blacklist
3. Query vs DGArchive

2) Scoring
1. N-gram frequency
2. Entropy
3. Domain length
4. NX domain?

3) Matching
Known Patterns of DGAs
1. Length
2. Alphabets
3. TLDs

Match?

New Seed?

No Match / Inconsistency

New DGA?

Parameter Extraction

Automate all the things!

- Customized sandboxing system for selected malware families
  - Processing shadowserver malware feeds (<- THANK YOU)

Part of TinyBanker DGA config in memory:

```
0000000: f9 b0 20 f3 aa 61 e8 00 00 00 00 58 2d 1b 68 40  .. ..a.....X-.h@
0000010: 00 ff 75 10 ff 75 0c ff 75 08 ff 90 33 4d 40 00  ..u..u..u...3M@.
0000020: 83 c4 0c c9 c3 90 90 90 90 90 90 90 90 90 90 90  ................
0000030: 73 70 61 69 6e 65 73 2e 70 77 00 00 00 00 00 00  spaines.pw......
0000040: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 50  ...............P
0000050: 2f 45 69 44 51 6a 4e 62 57 45 51 2f 00 00 00 00  /EiDQjNbWEQ/....
0000060: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000070: 47 50 51 61 74 5a 37 79 43 6b 4c 78 73 54 76 46  GPQatZ7yCkLxsTvF
0000080: 30 30 30 30 30 30 30 32 70 77 00 30 30 30 2c 01  00000002pw.000,.
0000090: 0a 00 e8 03 8c 00 00 00 45 ce a3 46 7d 32 b9 cc  ........E..F}2..
00000a0: 1a 55 80 de f2 8e f3 a7 e4 53 60 ca 11 6f 08 55  .U.......S`..o.U
00000b0: 14 ad 76 a6 12 67 8f 7e dd 49 fe 04 b0 b5 08 c8  ..v..g.~.I......
```

Regex for extraction of relevant fields:

```
regex_config = (
    r"\x90{4,16}"
    r"(?P<domain_name>[\S\s]{30})"
    r"(?P<unknown_word>[\S\s]{2})"
    r"(?P<uri>[\S\s]{32})"
    r"(?P<rc4_key>[a-zA-Z0-9]{16})"
    r"(?P<unknown_str>[\S\s]{8})"
    r"(?P<dga_tld>[\S\s]{6})"
    r"(?P<unknown_dword>[\S\s]{4})"
    r"(?P<num_dga_domains>[\S\s]{2})"
    r"(?P<static_config_len>[\S\s]{4})")
```

Comparison of

DGA Features

DGA Features

Intro

- Examples of DGA characteristics
  - DGA class and generation scheme (+ use of well-known algorithms)
  - Domain structure (length, alphabet) and TLDs
  - Domain validity period and domains per cycle (covered indirectly)
  - Domain randomness
  - C&C Priority
    - In short: DGA is basically always „last priority“
    - but 28 of 40 families use DGA as only C&C rendezvous method!
      (5 of them have hardcoded but basically unused domains)

DGA Features

Taxonomy and Generation Schemes

- DGA Classes (Taxonomy by Barabosch et al. [1]):

Type	Time dependent	Deterministic	Example
TID	☒	☑	Kraken, TinyBanker
TDD	☑	☑	Conficker, Gameover Zeus
TDN	☑	☒	Torpig, Bedep
TIN	☒	☒	-

- Generation Schemes

Type	Example Family	Example Domain
Arithmetic (A)	DirCrypt	vlbqryjd.com
Wordlist (W)	Matsnu	termacceptyear.com
Hashing (H)	Bamital	b83ed4877eec1997fcc39b7ae590007a.info
Permutation (P)	VolatileCedar	dotnetexplorer.info

[1] https://net.cs.uni-bonn.de/fileadmin/user_upload/wichmann/Extraction_DNGA_Malware.pdf

Bamital
Banjori
Bedep
Conficker
CoreBot
Cryptolocker
DirCrypt
Dyre
Fobber
Geodo
Gameover DGA
Gameover P2P
Gozi
Hesperbot
Kraken
Matsnu
Mewsei
Murofet 1
Murofet 2
Necurs
Nymaim
Pushdo
Pushdo TID
Pykspa 1
Pykspa 2
QakBot
Ramdo
Ramnit
Ranbyus
Redyms
Rovnix
Shifu
Simda
Suppobox
Szribi
Tempedreve
TinyBanker
Torpig
UrlZone
VolatileCedar

Names contain clickable links to references for these families.

Bamital TDD	Fobber TID	Mewsei TDD	Pykspa 2 TDD	Simda TID
Banjori TID	Geodo TDD	Murofet 1 TDD	QakBot TDD	Suppobox TDD
Bedep TDN	Gameover DGA TDD	Murofet 2 TDD	Ramdo TID	Szribi TDD
Conficker TDD	Gameover P2P TDD	Necurs TDD	Ramnit TID	Tempedreve TID
CoreBot TDD	Gozi TDD	Nymaim TDD	Ranbyus TDD	TinyBanker TID
Cryptolocker TDD	Hesperbot TID	Pushdo TDD	Redyms TID	Torpig TDD / TDN
DirCrypt TID	Kraken TID	Pushdo TID TID	Rovnix TID	UrlZone TID
Dyre TDD	Matsnu TDD	Pykspa 1 TDD	Shifu TID	VolatileCedar TID

Classes: 22 (55%) TDD, 16 (40%) TID, 2 (5%) TDN

Bamital H (MD5)	Fobber A (LCG)	Mewsei A (LCG)	Pykspa 2 A(LCG)	Simda A
Banjori A	Geodo A	Murofet 1 A (MD5)	QakBot A (Mersenne)	Suppobox W
Bedep A	Gameover DGA A (MD5)	Murofet 2 A (MD5)	Ramdo A	Szribi A
Conficker A	Gameover P2P A (MD5)	Necurs A	Ramnit A (LCG)	Tempedreve A (LCG)
CoreBot A (LCG)	Gozi W (LCG)	Nymaim A (Xorshift)	Ranbyus A	TinyBanker A
Cryptolocker A	Hesperbot A	Pushdo A (MD5)	Redyms A	Torpig A
DirCrypt A (LCG)	Kraken A	Pushdo TID A (LCG)	Rovnix A (LCG)	UrlZone A
Dyre H (SHA256)	Matsnu W	Pykspa 1 A	Shifu A (LCG)	VolatileCedar P

Classes: 34 (85%) A, 3 (7.5%) W, 2 (5%) H, 1 (2.5%) P

The Linear Congruential Generator (LCG)

- Pseudo-Random Number Generator (PRNG)
  - $X_{n+1} = (a * X_n + c) \bmod m$
  - Numerous variants of LCG with regard to parameters (a, c, m)
    - Numerical Recipes, MSVC, Park & Miller, own values, …
- Trivial example DGA: Pushdo TID

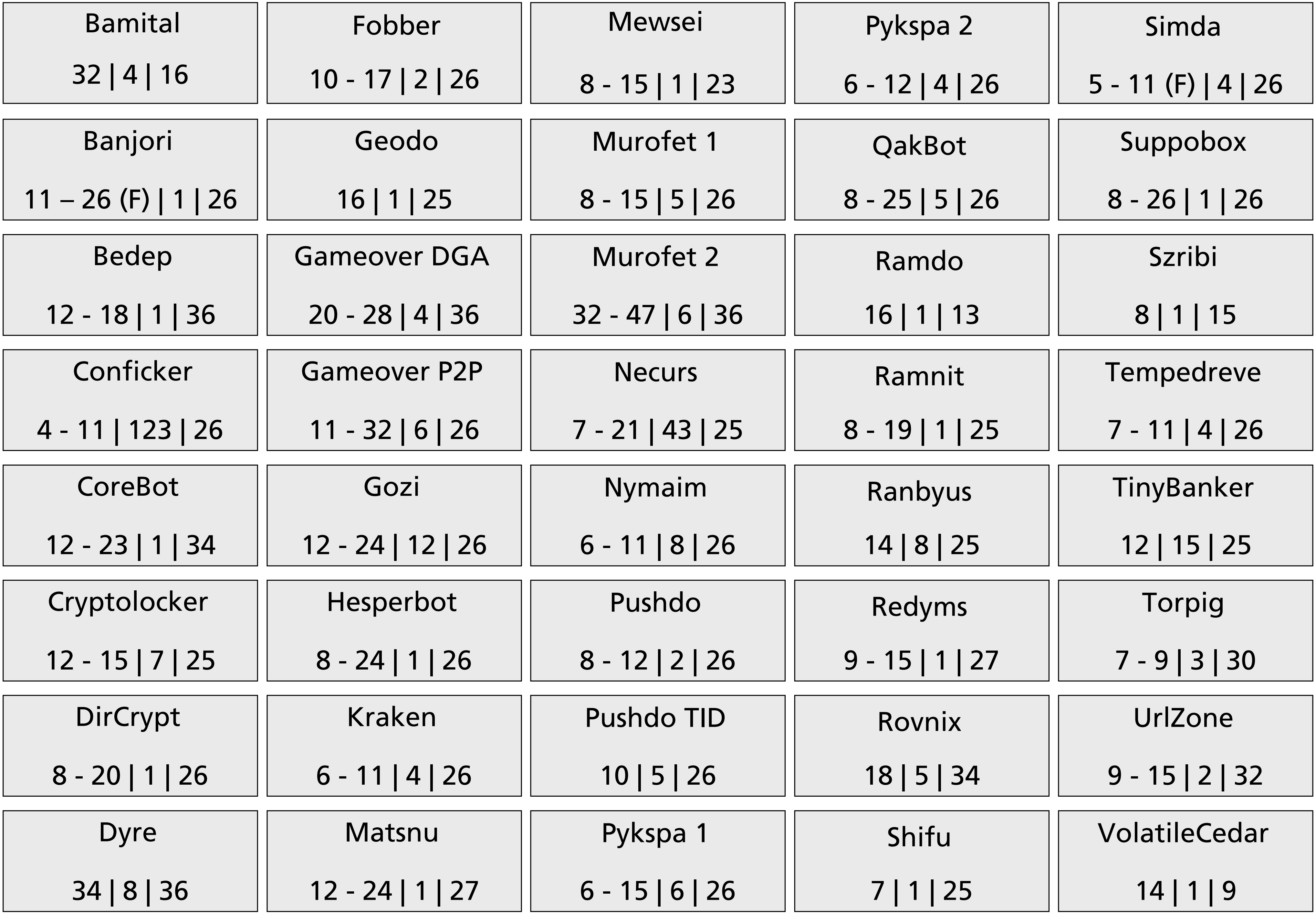
```
def generateDomain():
    domain = ""
    tlds = [".com", ".net", ".org", ".ru", ".tv"]
    for i in xrange(10):
        domain += chr(0x61 + lcg() % 26)
    domain += tlds[lcg() % 5]
    return domain

#################
"xirgbebore.tv"
"bsbuhapqbw.org"
"pgdudgjypi.ru"
```

Digression Time!

Bamital	Fobber	Mewsei	Pykspa 2	Simda
Banjori	Geodo	Murofet 1	QakBot	Suppobox
Bedep	Gameover DGA	Murofet 2	Ramdo	Szribi
Conficker	Gameover P2P	Necurs	Ramnit	Tempedreve
CoreBot	Gozi	Nymaim	Ranbyus	TinyBanker
Cryptolocker	Hesperbot	Pushdo	Redyms	Torpig
DirCrypt	Kraken	Pushdo TID	Rovnix	UrlZone
Dyre	Matsnu	Pykspa 1	Shifu	VolatileCedar

Domain Structure.



<table>
<tr><td>Bamital
32 | 4 | 16</td><td>Fobber
10 - 17 | 2 | 26</td><td>Mewsei
8 - 15 | 1 | 23</td><td>Pykspa 2
6 - 12 | 4 | 26</td><td>Simda
5 - 11 (F) | 4 | 26</td></tr>
<tr><td>Banjori
11 – 26 (F) | 1 | 26</td><td>Geodo
16 | 1 | 25</td><td>Murofet 1
8 - 15 | 5 | 26</td><td>QakBot
8 - 25 | 5 | 26</td><td>Suppobox
8 - 26 | 1 | 26</td></tr>
<tr><td>Bedep
12 - 18 | 1 | 36</td><td>Gameover DGA
20 - 28 | 4 | 36</td><td>Murofet 2
32 - 47 | 6 | 36</td><td>Ramdo
16 | 1 | 13</td><td>Szribi
8 | 1 | 15</td></tr>
<tr><td>Conficker
4 - 11 | 123 | 26</td><td>Gameover P2P
11 - 32 | 6 | 26</td><td>Necurs
7 - 21 | 43 | 25</td><td>Ramnit
8 - 19 | 1 | 25</td><td>Tempedreve
7 - 11 | 4 | 26</td></tr>
<tr><td>CoreBot
12 - 23 | 1 | 34</td><td>Gozi
12 - 24 | 12 | 26</td><td>Nymaim
6 - 11 | 8 | 26</td><td>Ranbyus
14 | 8 | 25</td><td>TinyBanker
12 | 15 | 25</td></tr>
<tr><td>Cryptolocker
12 - 15 | 7 | 25</td><td>Hesperbot
8 - 24 | 1 | 26</td><td>Pushdo
8 - 12 | 2 | 26</td><td>Redyms
9 - 15 | 1 | 27</td><td>Torpig
7 - 9 | 3 | 30</td></tr>
<tr><td>DirCrypt
8 - 20 | 1 | 26</td><td>Kraken
6 - 11 | 4 | 26</td><td>Pushdo TID
10 | 5 | 26</td><td>Rovnix
18 | 5 | 34</td><td>UrlZone
9 - 15 | 2 | 32</td></tr>
<tr><td>Dyre
34 | 8 | 36</td><td>Matsnu
12 - 24 | 1 | 27</td><td>Pykspa 1
6 - 15 | 6 | 26</td><td>Shifu
7 | 1 | 25</td><td>VolatileCedar
14 | 1 | 9</td></tr>
</table>

Scheme: Min$_{length}$ - Max$_{length}$ (Fixed per seed) | TLDs | Size of Alphabet

Bamital 32 | 4 | 16	Fobber 10 - 17 | 2 | 26	Mewsei 8 - 15 | 1 | 23	Pykspa 2 6 - 12 | 4 | 26	Simda 5 - 11 (F) | 4 | 26
Banjori 11 – 26 (F) | 1 | 26	Geodo 16 | 1 | 25	Murofet 1 8 - 15 | 5 | 26	QakBot 8 - 25 | 5 | 26	Suppobox 8 - 26 | 1 | 26
Bedep 12 - 18 | 1 | 36	Gameover DGA 20 - 28 | 4 | 36	Murofet 2 32 - 47 | 6 | 36	Ramdo 16 | 1 | 13	Szribi 8 | 1 | 15
Conficker 4 - 11 | 123 | 26	Gameover P2P 11 - 32 | 6 | 26	Necurs 7 - 21 | 43 | 25	Ramnit 8 - 19 | 1 | 25	Tempedreve 7 - 11 | 4 | 26
CoreBot 12 - 23 | 1 | 34	Gozi 12 - 24 | 12 | 26	Nymaim 6 - 11 | 8 | 26	Ranbyus 14 | 8 | 25	TinyBanker 12 | 15 | 25
Cryptolocker 12 - 15 | 7 | 25	Hesperbot 8 - 24 | 1 | 26	Pushdo 8 - 12 | 2 | 26	Redyms 9 - 15 | 1 | 27	Torpig 7 - 9 | 3 | 30
DirCrypt 8 - 20 | 1 | 26	Kraken 6 - 11 | 4 | 26	Pushdo TID 10 | 5 | 26	Rovnix 18 | 5 | 34	UrlZone 9 - 15 | 2 | 32
Dyre 34 | 8 | 36	Matsnu 12 - 24 | 1 | 27	Pykspa 1 6 - 15 | 6 | 26	Shifu 7 | 1 | 25	VolatileCedar 14 | 1 | 9

TLDs: 10+, 5+, … (per seed) – thankfully, most use only few TLDs

Bamital
32 | 4 | 16

Banjori
11 – 26 (F) | 1 | 26

Bedep
12 - 18 | 1 | 36

Conficker
4 - 11 | 123 | 26

CoreBot
12 - 23 | 1 | 34

Cryptolocker
12 - 15 | 7 | 25

DirCrypt
8 - 20 | 1 | 26

Dyre
34 | 8 | 36

Fobber
10 - 17 | 2 | 26

Geodo
16 | 1 | 25

Gameover DGA
20 - 28 | 4 | 36

Gameover P2P
11 - 32 | 6 | 26

Gozi
12 - 24 | 12 | 26

Hesperbot
8 - 24 | 1 | 26

Kraken
6 - 11 | 4 | 26

Matsnu
12 - 24 | 1 | 27

Mewsei
8 - 15 | 1 | 23

Murofet 1
8 - 15 | 5 | 26

Murofet 2
32 - 47 | 6 | 36

Necurs
7 - 21 | 43 | 25

Nymaim
6 - 11 | 8 | 26

Pushdo
8 - 12 | 2 | 26

Pushdo TID
10 | 5 | 26

Pykspa 1
6 - 15 | 6 | 26

Pykspa 2
6 - 12 | 4 | 26

QakBot
8 - 25 | 5 | 26

Ramdo
16 | 1 | 13

Ramnit
8 - 19 | 1 | 25

Ranbyus
14 | 8 | 25

Redyms
9 - 15 | 1 | 27

Rovnix
18 | 5 | 34

Shifu
7 | 1 | 25

Simda
5 - 11 (F) | 4 | 26

Suppobox
8 - 26 | 1 | 26

Szribi
8 | 1 | 15

Tempedreve
7 - 11 | 4 | 26

TinyBanker
12 | 15 | 25

Torpig
7 - 9 | 3 | 30

UrlZone
9 - 15 | 2 | 32

VolatileCedar
14 | 1 | 9

Size of Alphabet?

Not sure if intentional or bugs…

```
def generateDomain():
    domain = ""
    tlds = [".com", ".net", ".org", ".ru", ".tv"]
    for i in xrange(10):
        domain += chr(0x61 + lcg() % 26)
    domain += tlds[lcg() % 5]
    return domain
```

- PushdoTID has an alphabet size of 26!
- However, if you use modulo **25**…
  - CryptoLocker, Geodo, Necurs, Ramnit, Ranbyus, Shifu, Tinybanker
- Or do this twice (on vowels and consonants, or chars and numbers) …
  - CoreBot (34), Mewsei (23), Rovnix (34)
- Or do something even more special …
  - Ramdo (13), Szribi (15), Torpig (30), UrlZone (32)

Size of Alphabet: Some DGAs use truncated alphabets.

Bamital

Banjori

Bedep

Conficker

CoreBot

Cryptolocker

DirCrypt

Dyre

Fobber

Geodo

Gameover DGA

Gameover P2P

Gozi

Hesperbot

Kraken

Matsnu

Mewsei

Murofet 1

Murofet 2

Necurs

Nymaim

Pushdo

Pushdo TID

Pykspa 1

Pykspa 2

QakBot

Ramdo

Ramnit

Ranbyus

Redyms

Rovnix

Shifu

Simda

Suppobox

Szribi

Tempedreve

TinyBanker

Torpig

UrlZone

VolatileCedar

Domain length analysis out of scope

DGA Domain Space and

Registration Status

DGA Domain Space

as seen by DGArchive

- So we reversed many DGAs and extracted a lot of seeds…
  - How many potential DGA domains are there?
  - Are there collisions between DGAs?
  - How many of these domains are registered?

- **No ground-truth available :(**

DGA Domain Space

as seen by DGArchive

- Study conducted on data set fixed on 22nd September 2015
  - Domains generated from first spotting of family until 31.12.2015
- Eternal thanks to Michael Klatt & DomainTools!
  - Provided historic WHOIS data for all domains in DGArchive
- Evaluation of WHOIS features for majority of DGAs
  - Identified characteristics about domains
    - Sinkholes
    - Mitigations (registration turned to sinkhole at later point)
    - Pre-registrations (registration before appearence of the family)
    - Domain Parking
  - Random fact: 25 / 40 DGAs surfaced 2013 and later!

Bamital
Banjori
Bedep
Conficker
CoreBot
Cryptolocker
DirCrypt
Dyre
Fobber
Geodo
Gameover DGA
Gameover P2P
Gozi
Hesperbot
Kraken
Matsnu
Mewsei
Murofet 1
Murofet 2
Necurs
Nymaim
Pushdo
Pushdo TID
Pykspa 1
Pykspa 2
QakBot
Ramdo
Ramnit
Ranbyus
Redyms
Rovnix
Shifu
Simda
Suppobox
Szribi
Tempedreve
TinyBanker
Torpig
UrlZone
VolatileCedar

How many domains are generated by these DGAs?

Bamital 197,000 | 1	Fobber 2,000 | 2	Mewsei 1,984 | 1	Pykspa 2 775,342 | 2	Simda 11,528 | 12
Banjori 421,390 | 30	Geodo 90,232 | 2	Murofet 1 4,063,680 | 2	QakBot 385,000 | 1	Suppobox 98,304 | 3
Bedep 3,806 | 4	Gameover DGA 6,182,000 | 2	Murofet 2 262,000 | 1	Ramdo 3000 | 3	Szribi 2,949 | 1
Conficker 125,118,625 | 3	Gameover P2P 262,000 | 1	Necurs 3,551,232 | 6	Ramnit 18,000 | 18	Tempedreve 204 | 1
CoreBot 18,160 | 2	Gozi 16,963 | 9	Nymaim 65,040 | 3	Ranbyus 64,400 | 7	TinyBanker 81,930 | 90
Cryptolocker 1,108,000 | 1	Hesperbot 178 | 3	Pushdo 124,021 | 4	Redyms 34 | 1	Torpig 17,610 | 2
DirCrypt 420 | 14	Kraken 300 | 1	Pushdo TID 6,000 | 1	Rovnix 10,000 | 1	UrlZone 10,009 | 6
Dyre 592,000 | 1	Matsnu 3,346 | 2	Pykspa 1 22,764 | 1	Shifu 1,554 | 2	VolatileCedar 170 | 1

Sum of unique domains: 143,584,257 or 18,465,647 without Conficker

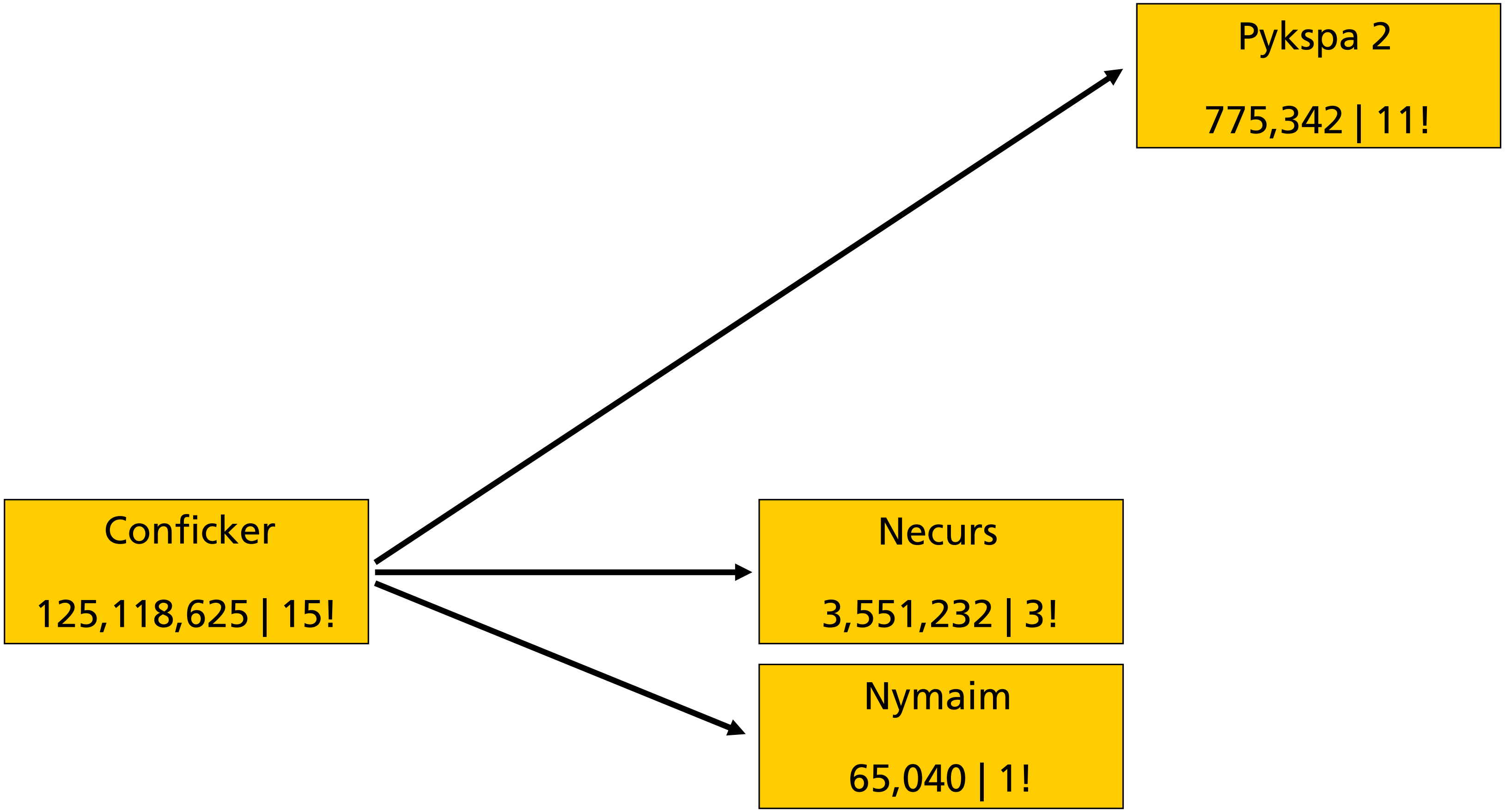
DGA Domain Space

Domain Collisions

- DGA domains may collide
  - Within a DGA
  - With other DGAs

Bamital 197,000	Fobber 2,000	Mewsei 1,984	Pykspa 2 775,342	Simda 11,528
Banjori 421,390	Geodo 90,232	Murofet 1 4,063,680	QakBot 385,000	Suppobox 98,304
Bedep 3,806	Gameover DGA 6,182,000	Murofet 2 262,000	Ramdo 3000	Szribi 2,949
Conficker 125,118,625	Gameover P2P 262,000	Necurs 3,551,232	Ramnit 18,000	Tempedreve 204
CoreBot 18,160	Gozi 16,963	Nymaim 65,040	Ranbyus 64,400	TinyBanker 81,930
Cryptolocker 1,108,000	Hesperbot 178	Pushdo 124,021	Redyms 34	Torpig 17,610
DirCrypt 420	Kraken 300	Pushdo TID 6,000	Rovnix 10,000	UrlZone 10,009
Dyre 592,000	Matsnu 3,346	Pykspa 1 22,764	Shifu 1,554	VolatileCedar 170

How many domain collisions between Conficker and the other DGAs?



Conficker

125,118,625 | 15!

Pykspa 2

775,342 | 11!

Necurs

3,551,232 | 3!

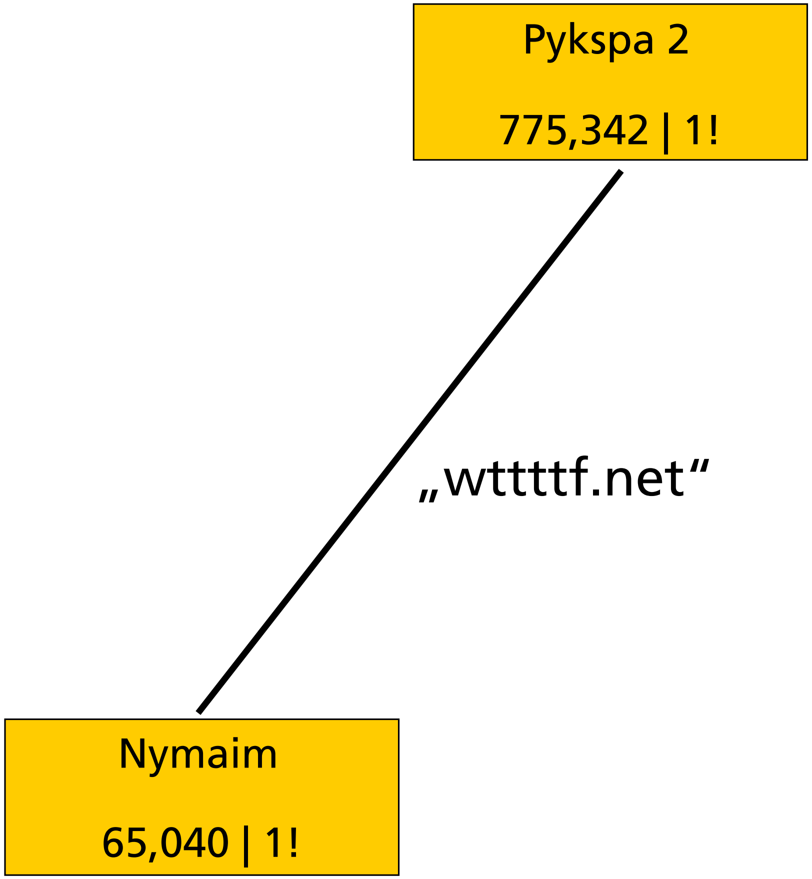
Nymaim

65,040 | 1!

Not that many.

Bamital 197,000	Fobber 2,000	Mewsei 1,984	Pykspa 2 775,342	Simda 11,528
Banjori 421,390	Geodo 90,232	Murofet 1 4,063,680	QakBot 385,000	Suppobox 98,304
Bedep 3,806	Gameover DGA 6,182,000	Murofet 2 262,000	Ramdo 3000	Szribi 2,949
Conficker 125,118,625	Gameover P2P 262,000	Necurs 3,551,232	Ramnit 18,000	Tempedreve 204
CoreBot 18,160	Gozi 16,963	Nymaim 65,040	Ranbyus 64,400	TinyBanker 81,930
Cryptolocker 1,108,000	Hesperbot 178	Pushdo 124,021	Redyms 34	Torpig 17,610
DirCrypt 420	Kraken 300	Pushdo TID 6,000	Rovnix 10,000	UrlZone 10,009
Dyre 592,000	Matsnu 3,346	Pykspa 1 22,764	Shifu 1,554	VolatileCedar 170

And now without Conficker?



Pykspa 2

775,342 | 1!

„wttttf.net“

Nymaim

65,040 | 1!

One single collision

Conclusions:

- So if there are so few collisions between DGAs…
  - using pre-calculated AGDs to identify malware is pretty accurate!
  - Actually for both family and campaign!

Inter-DGA domain collisions: Basically non-existent!

DGA Domain Space

Registrations and Domain Collisions

- DGA domains may collide
  - With already registered (benign) domains

- So first: How many domains are registered?

Bamital	197,000
Banjori	421,390
Bedep	3,806
Conficker	125,118,625
CoreBot	18,160
Cryptolocker	1,108,000
DirCrypt	420
Dyre	592,000
Fobber	2,000
Geodo	90,232
Gameover DGA	6,182,000
Gameover P2P	262,000
Gozi	16,963
Hesperbot	178
Kraken	300
Matsnu	3,346
Mewsei	1,984
Murofet 1	4,063,680
Murofet 2	262,000
Necurs	3,551,232
Nymaim	65,040
Pushdo	124,021
Pushdo TID	6,000
Pykspa 1	22,764
Pykspa 2	775,342
QakBot	385,000
Ramdo	3000
Ramnit	18,000
Ranbyus	64,400
Redyms	34
Rovnix	10,000
Shifu	1,554
Simda	11,528
Suppobox	98,304
Szribi	2,949
Tempedreve	204
TinyBanker	81,930
Torpig	17,610
UrlZone	10,009
VolatileCedar	170

Registrations?

Bamital 8,340 (4.22%) 197,000	Fobber 13 (0.65%) 2,000	Mewsei DDNS 1,984	Pykspa 2 1,927 (0.25%) 775,342	Simda 379 (3.29%) 11,528
Banjori 683 (0.16%) 421,390	Geodo 107 (0.12%) 90,232	Murofet 1 3,172 (0.08%) 4,063,680	QakBot 1,088 (0.28%) 385,000	Suppobox 11,338 (11.53%) 98,304
Bedep 654 (17.18%) 3,806	Gameover DGA 1,081 (0.02%) 6,182,000	Murofet 2 559 (0.21%) 262,000	Ramdo 47 (1.57%) 3000	Szribi 54 (1.83%) 2,949
Conficker - 125,118,625	Gameover P2P 74,755 (28.53%) 262,000	Necurs 295 (0.01%) 3,551,232	Ramnit 939 (5.22%) 18,000	Tempedreve 20 (9.80%) 204
CoreBot DDNS 18,160	Gozi 305 (1.80%) 16,963	Nymaim 656 (1.01%) 65,040	Ranbyus 98 (0.15%) 64,400	TinyBanker 1,733 (2.12%) 81,930
Cryptolocker 3,820 (0.34%) 1,108,000	Hesperbot 15 (8.43%) 178	Pushdo 453 (0.37%) 124,021	Redyms 11 (32.35%) 34	Torpig 139 (0.79%) 17,610
DirCrypt 86 (20.48%) 420	Kraken DDNS 300	Pushdo TID 245 (4.08%) 6,000	Rovnix 1 (0.01%) 10,000	UrlZone 127 (1.27%) 10,009
Dyre 850 (0.14%) 592,000	Matsnu 610 (18.23%) 3,346	Pykspa 1 455 (2.00%) 22,764	Shifu 11 (0.71%) 1,554	VolatileCedar 13 (7.65%) 170

Registrations: 115,079 (0.62%) of 18,465,427 unique domains we had data for.

Bamital
7,891 / 8,340
197,000

Gameover P2P
72,713 / 74,755
262,000

Cryptolocker
2,899 / 3,820
1,108,000

Takedowns actually account for 72,56% of all considered DGA registrations.

DGA Domain Space

Domain Collisions

- DGA domains may collide
  - With already registered (benign) domains

Bamital 0 / 8,340 197,000	Fobber 0 / 13 2,000	Mewsei	Pykspa 2 757 / 1,927 (39.28%)	Simda 66 / 379 (17.41%) 11,528
Banjori 0 / 683 421,390	Geodo 0 / 107 90,232	Murofet 1 0 / 3,172 4,063,680	QakBot 0 / 1,088 385,000	Suppobox 8.434 / 11,338 (74.39%)
Bedep 0 / 654 3,806	Gameover DGA 0 / 1,081 6,182,000	Murofet 2 0 / 559 262,000	Ramdo 0 / 47 3000	Szribi 0 / 54 2,949
Conficker	Gameover P2P 0 / 74,755 262,000	Necurs 10 / 295 (3.34%) 3,551,232	Ramnit 0 / 939 18,000	Tempedreve 0 / 20 204
CoreBot	Gozi 48 / 305 (15.74%) 16,963	Nymaim 70 / 656 (10.67%) 65,040	Ranbyus 0 / 98 64,400	TinyBanker 0 / 1,733 81,930
Cryptolocker 0 / 3,820 1,108,000	Hesperbot 0 / 15 178	Pushdo 3 / 453 (0.66%) 124,021	Redyms 0 / 11 34	Torpig 2 / 139 (1.44%) 17,610
DirCrypt 0 / 86 420	Kraken	Pushdo TID 0 / 245 6,000	Rovnix 0 / 1 10,000	UrlZone 0 / 127 10,009
Dyre 0 / 850 592,000	Matsnu 244 / 610 (40.00%) 3,346	Pykspa 1 12 / 455 (2.64%) 22,764	Shifu 0 / 11 1,554	VolatileCedar 0 / 13 170

Pre-Registrations: Wordlist-DGAs and short domains cause the most collisions.

Conclusions:

- Breakdown of Pre-Registrations (9,646)
  - Wordlist-DGAs: 8,726 (90.46%)
  - Remainder: 920
    - „Short“ domains (length 5-6): 856 (93.04%)
    - Remainder: 64
      - Accidentally „real“ words: „veterans.kz“
      - Pronounceable „words“: „kankanana.com“
      - „kandilmed.com“
- Basically no collisions with non-Wordlist DGAs or „long“ domains
  - Using DGArchive for blocking -> very low FP rate for blocking!

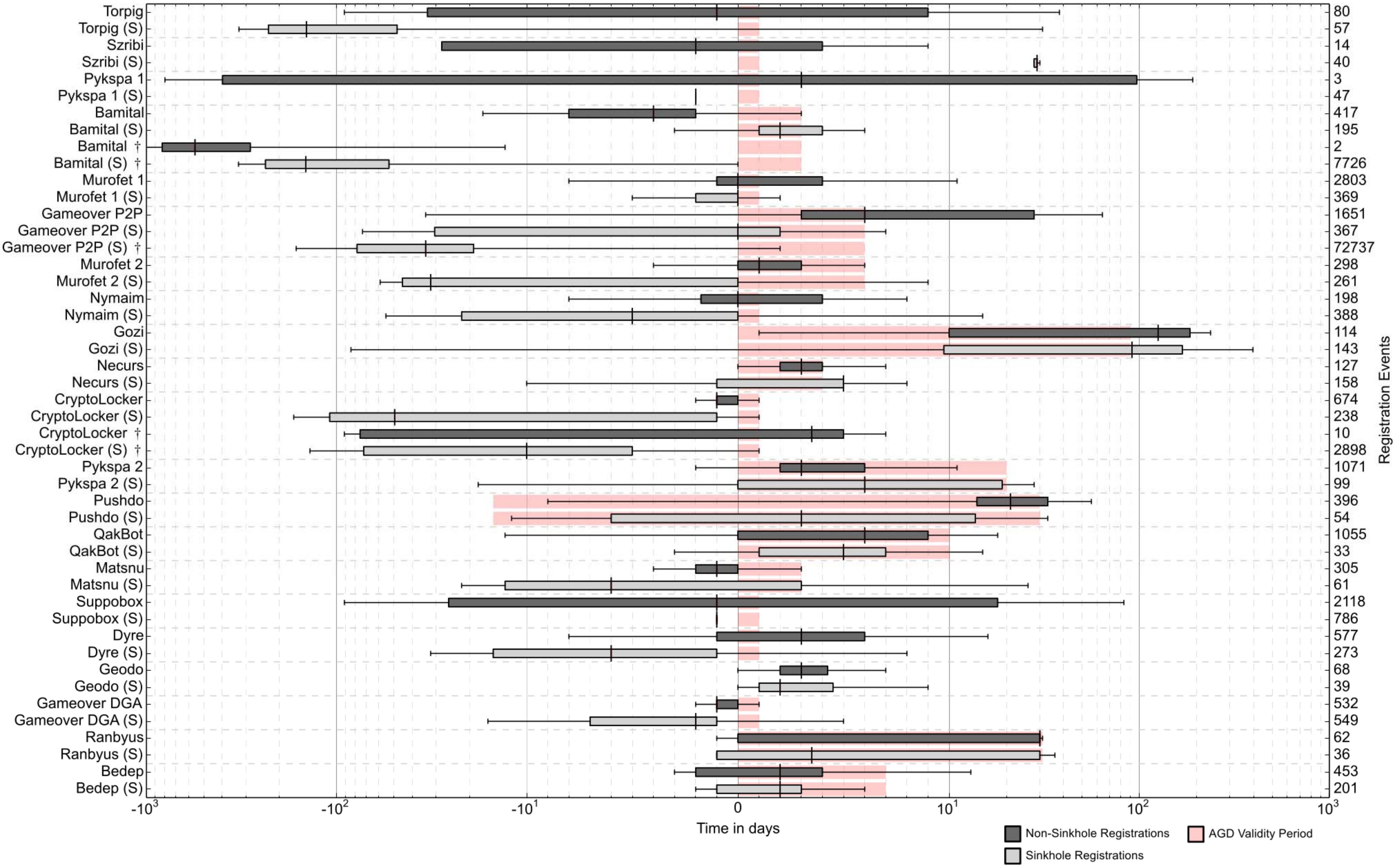
592,000	3,346	22,764	1,554	170

Pre-Registrations: With some exclusions basically non-existent.

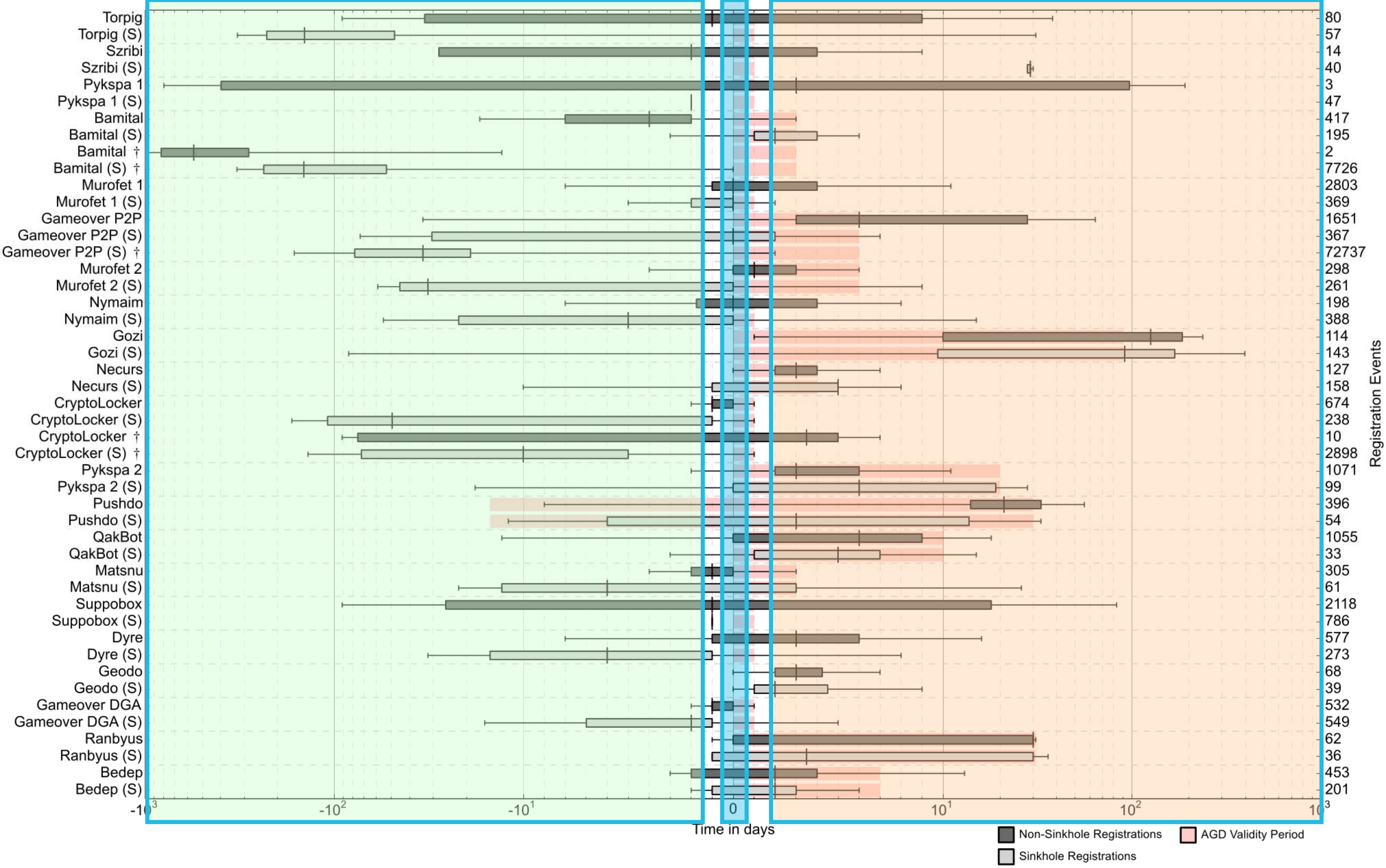
DGA Domain Space

Registration Timings

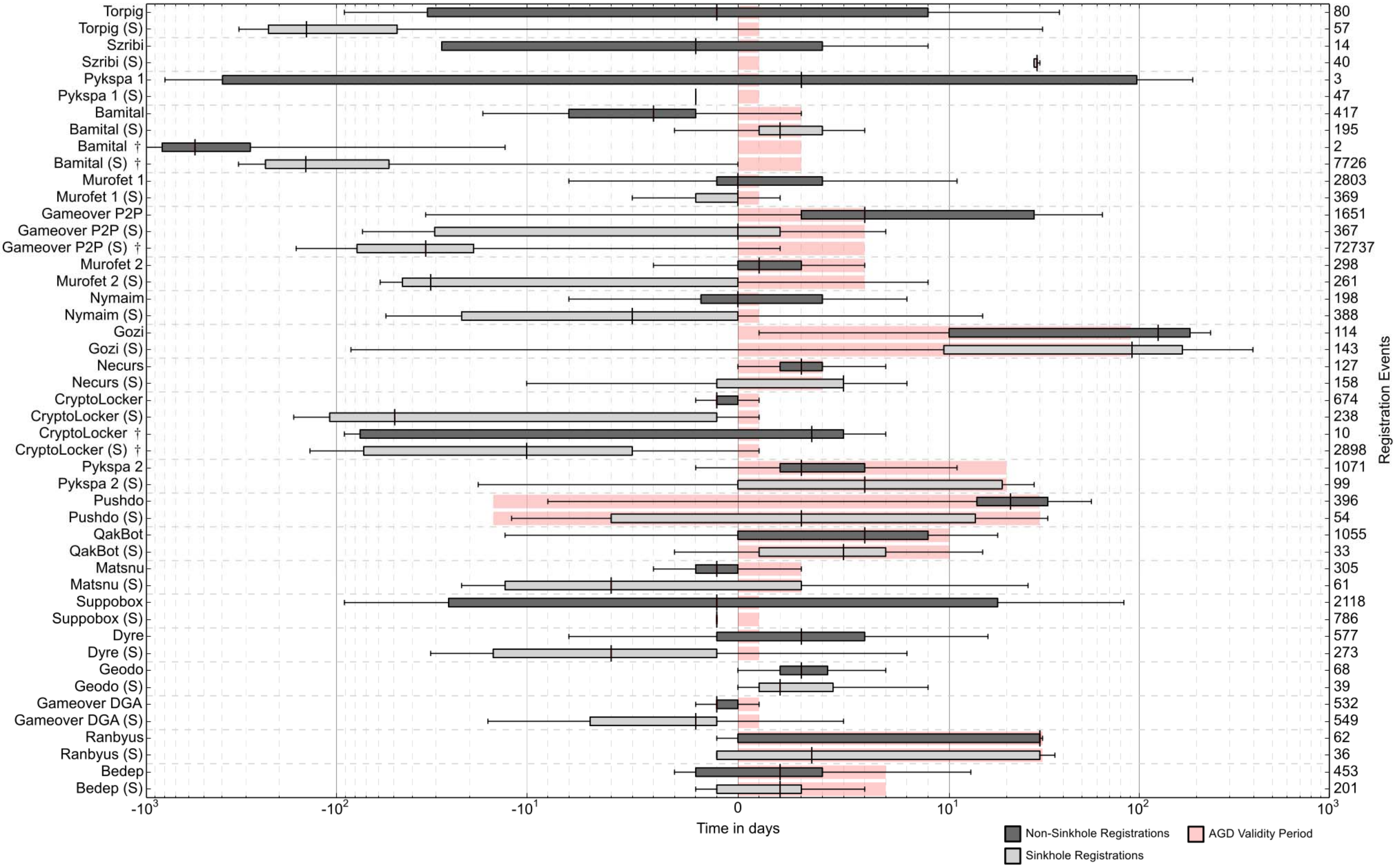
- Consider time-dependent DGAs
  - Sets of domains have a window of validity!
- What is „registration lookahead“?
  - Relative „offset“ between start of validity and registration time
- In the following: Evaluation of registration lookaheads
  - For sinkholes
  - For „non-sinkholes“



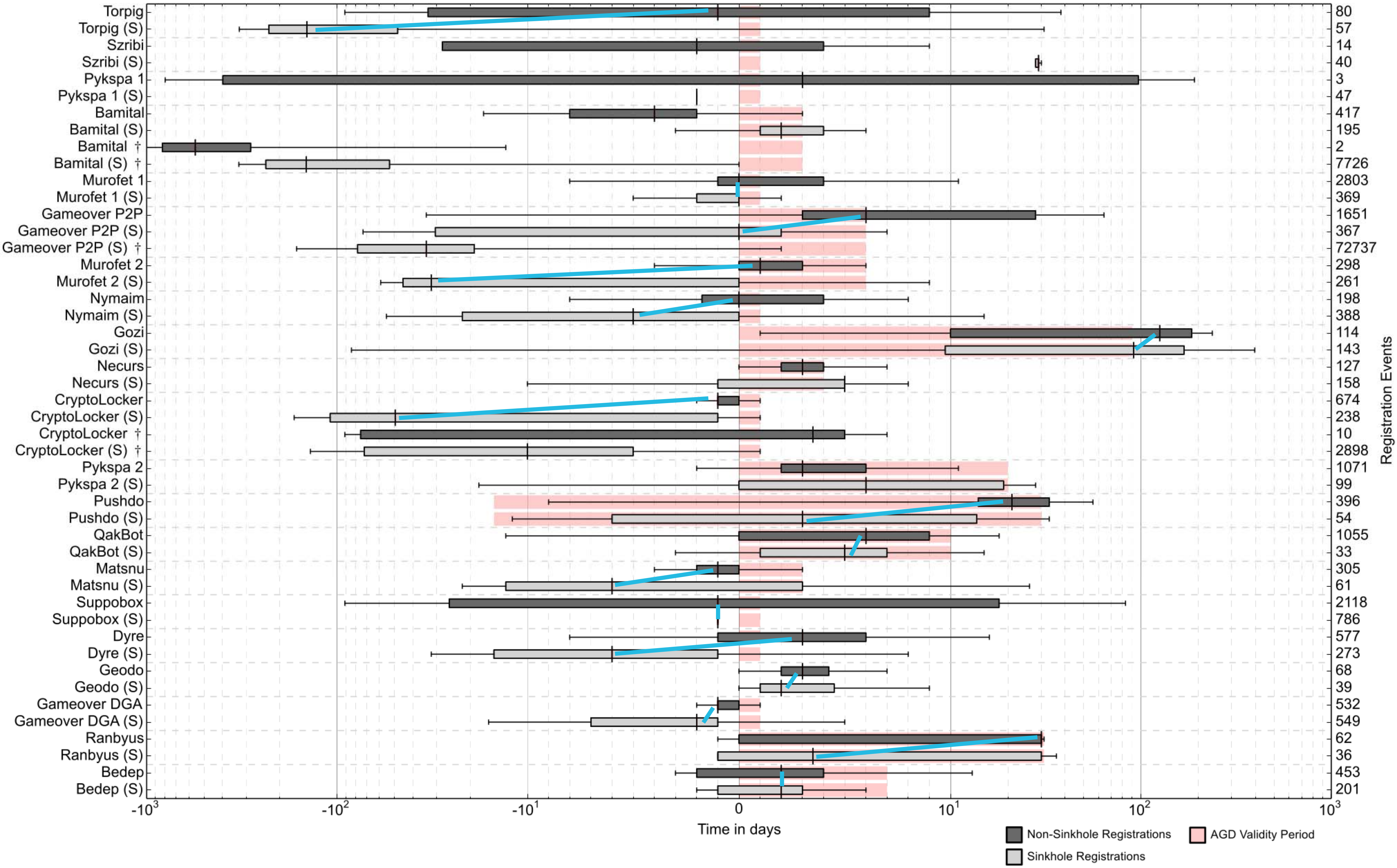
Overview of registration lookaheads



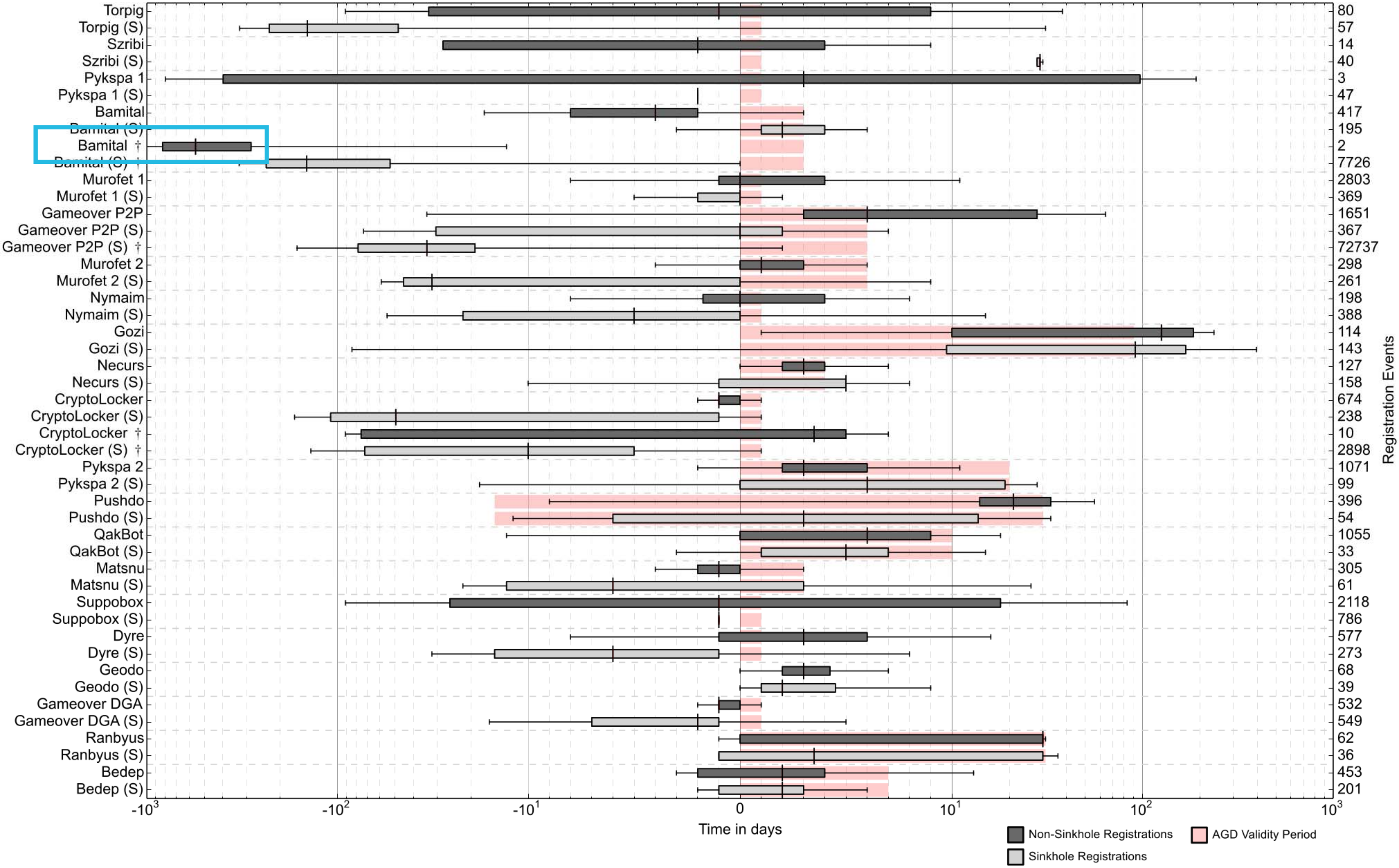
Domains registered BEFORE, ON, AFTER the first day they became valid in the DGA



However, red bars show how long domains REMAIN valid



Observation: Sinkholes are often registered earlier than „non“-sinkholes

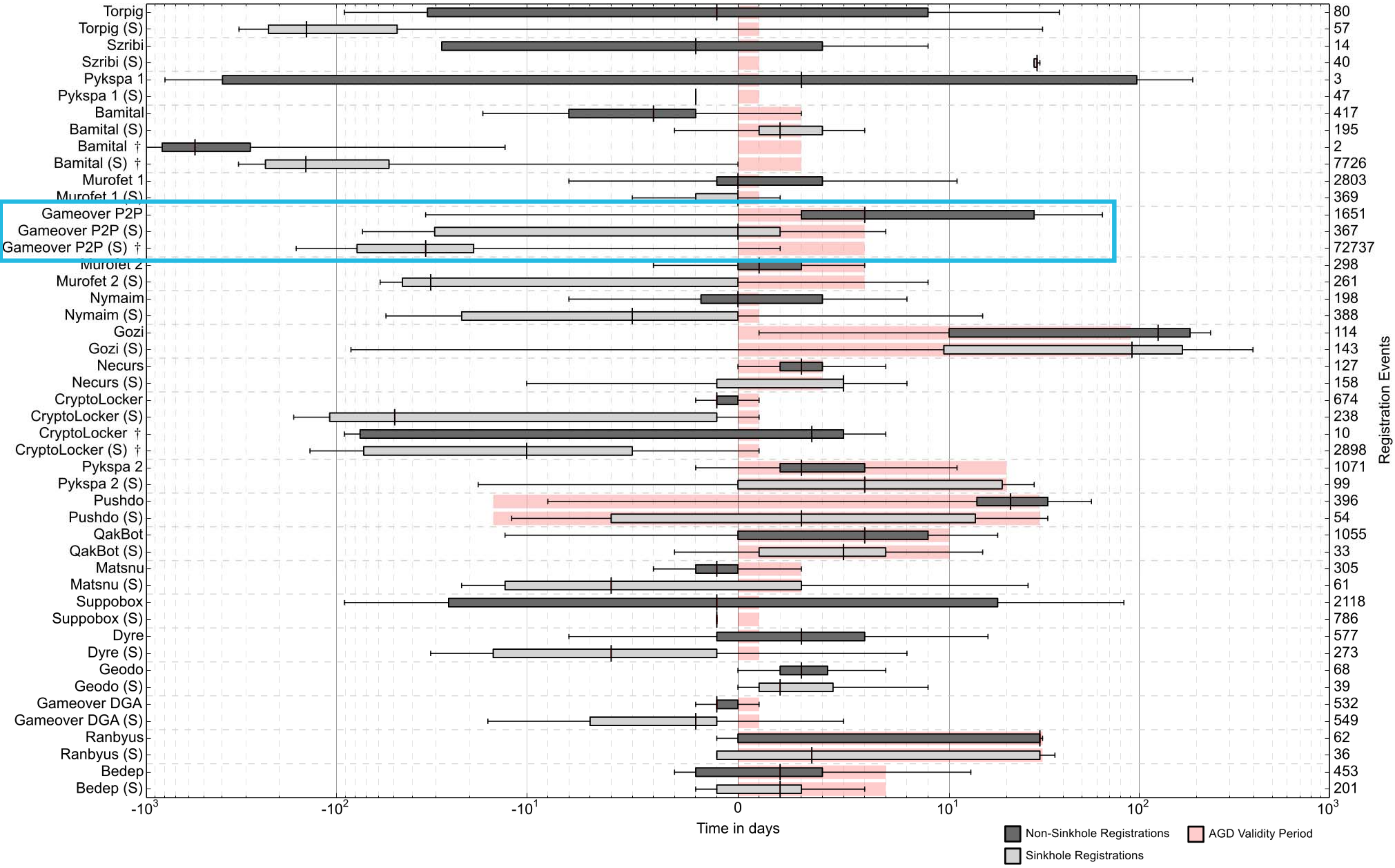


Observation: Some domains are registered far into the future

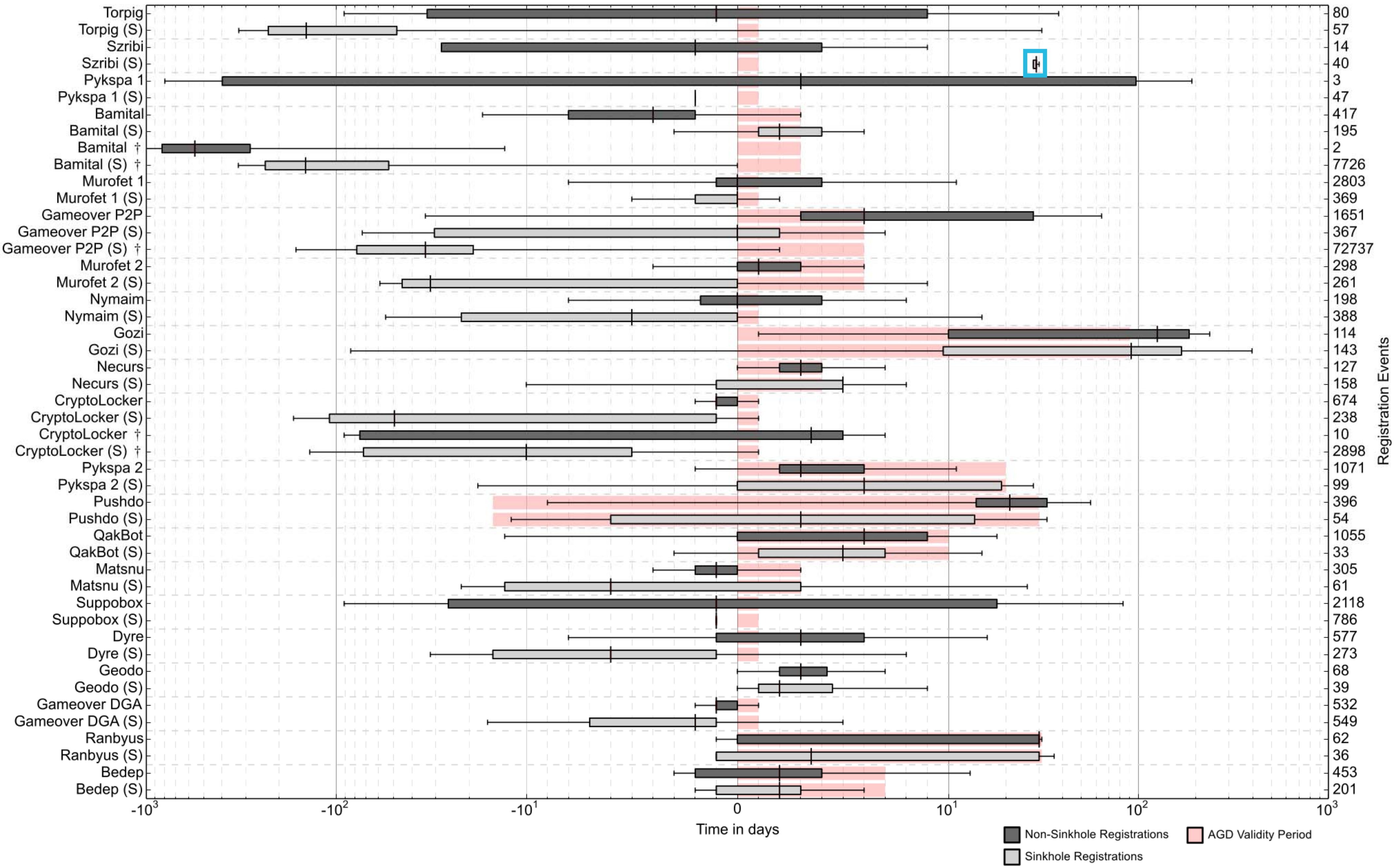
Bamital: Same day registrations for domains 1, 2, 3 years in advance

Nymaim: Same day registrations for domains 1, 2 years in advance

Murofet: Same day registrations for domains 1, 2, 3, 4 years in advance



Observation: Gameover P2P was killed completely! Yay! :)



Observation: Reversing Fail by some sinkholer?

32 domains registered exactly one month after validity?

Wrapping up

Conclusion

Conclusion

Wrapping it up

- DGArchive
  - Looking for more users / contributors!
  - Request free access: daniel.plohmann@fkie.fraunhofer.de
  - Required: basic proof of identity (e.g. no freemailer) or vetting
- Future plans
  - Document everything in detail (paper in preparation)
  - Heuristical Domain Classifier
  - More automation
  - DGA Hunting Collaboration / Community?

DGArchive
Thanks for your contributions

- Johannes Bader, Michael Klatt
- Chris Baker, John Bambenek, Thomas Barabosch, Adam Brunner, Steffen Enders, Christopher Kannen, Peter Kleissner, Felix Leder, Thorsten Jenke, Jason Jones, Alexandr Matrosov, Sandor Nemes, Isaac Palmer, Dennis Schwarz, Brett Stone-Gross, Tillmann Werner, Zhang Zaifang
- Anubisnetworks, Checkpoint, DomainTools, GovCERT.ch, Quarantainenet.nl, Shadowserver, SWITCH.ch, Symantec