

Two Million IDNs

Tech Day | ICANN 64 | Kobe

John Levine | Standcore LLC | john.levine@standcore.com

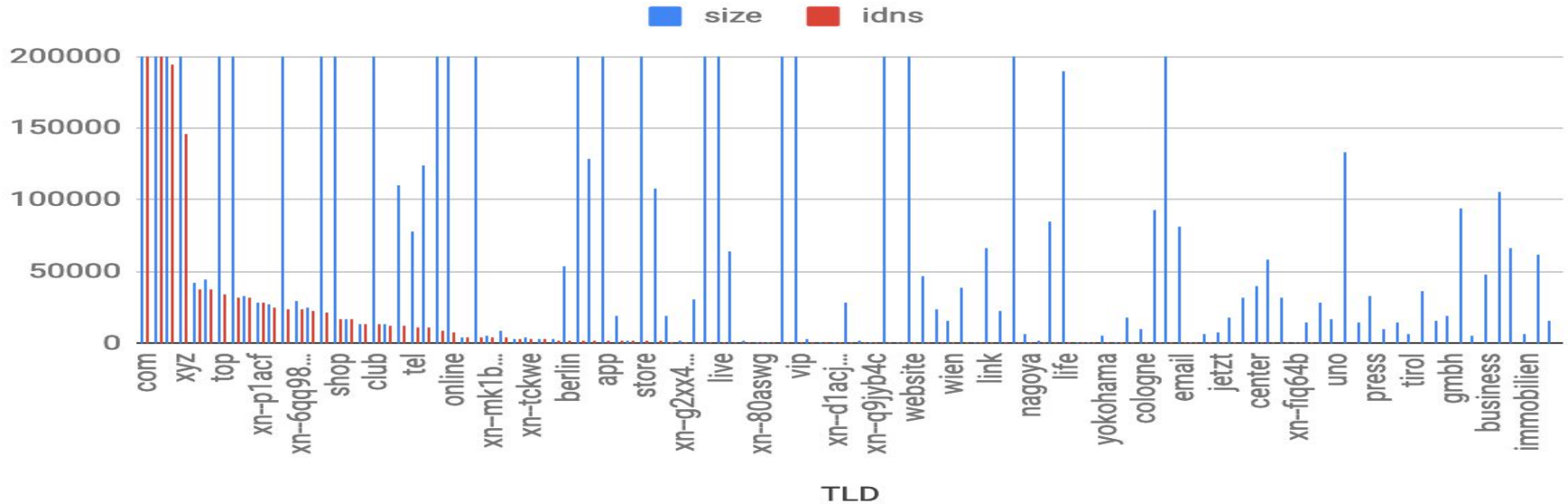


The contracted zones

- Number of zones: 1232
- Number of names: 193M
- Number of IDNs: 2M
 - Or about 1% of names

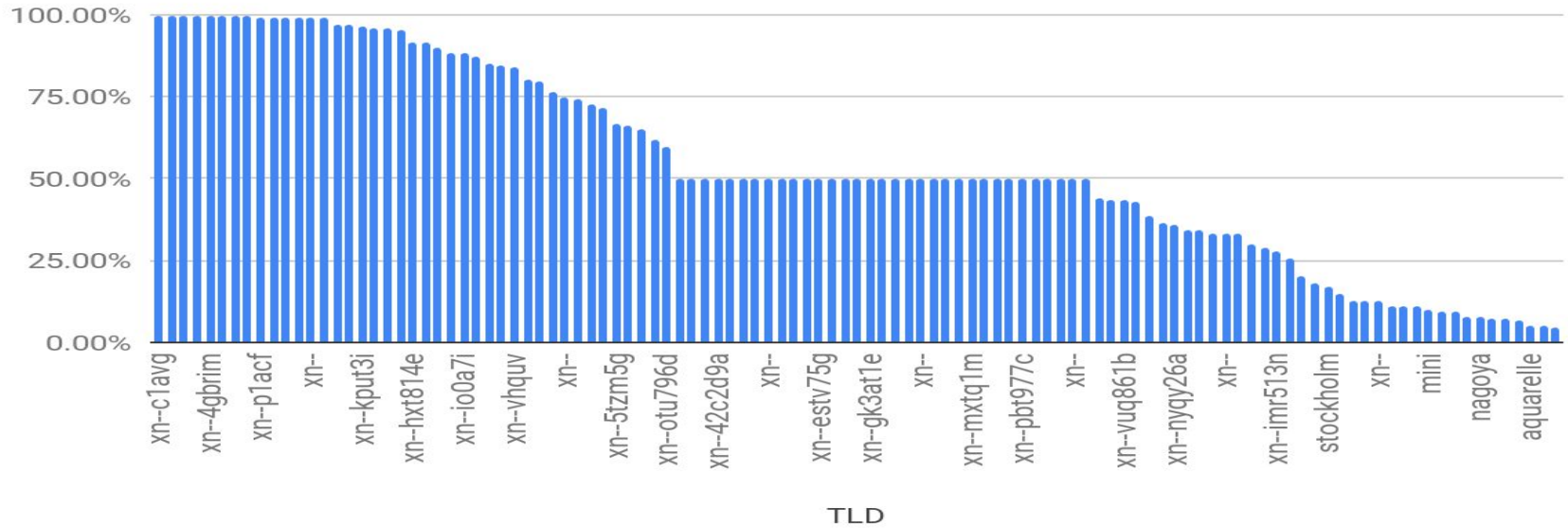
The contracted zones

size and idns



The contracted zones

IDN percent



What did I do?

- Get all the zone files, via FTP or CZDS
- Run them through a script
 - Takes about 90 minutes
- Put statistics in a database
- Find out when interesting names were registered
 - Not redacted but still pretty screwed up

What did I check?

- Is the name valid under IDNA 2003?
- Is the name valid under IDNA 2008?
- Is the name valid under the TLD's label generation rules?

Label Generation Rules: the Theory

- TLD says what languages it accepts in contracts
- TLD sends language rules to IANA in RFC3743 format
 - See <https://www.iana.org/domains/idn-tables>
- Registered names follow rules for a language/script

Label Generation Rules: the Practice

- Some TLDs register languages not in the contract
 - Notably Chinese
- Table files rarely follow 3743 syntax
 - Files are text but some files in HTML just because
- Some TLDs haven't sent in files

Dealing with LGR files

- Ad-hoc parser
 - Handles all the files, I think
- Turn each file into a set of character codes
- Make merged set of all languages in a TLD
- Doesn't capture context dependent rules

Checking the IDNs

- Check if valid under IDNA2003
- Check if valid under IDNA2008
- See if characters in name are in merged set for TLD
- If so, see if in merged set for a language in the TLD
 - Not quite right, doesn't do context rules
 - If TLD changed the rules, doesn't check date

What did I find?

- Most names are OK:
 - Valid under both IDNA2003 and IDNA2008
 - Matches some language rules
- Thousands of names are not
 - 509 names invalid IDN2003
 - 4845 names invalid IDN2008

Invalid IDN2003

- 509 names invalid IDN2003
- All appear to be new valid 2008
 - Many German with ß
 - Arabic with digits

Invalid IDN2008

- 4845 names invalid IDN2008
- Most are length errors
- About 1000 other errors
 - Pre-2008 legacy junk names
 - New gTLD naughtiness

Ancient junk

- Pre-2008 legacy junk names
- Funky punctuation
 - §sex.com, €-bank.com, 1000°.com, ...
- All appear to be parked or for sale
- Some look evil
 - xn--google-36d.com google.com

Severe sloppiness

- Names in non-contracted languages so no LGR
 - 600 Chinese name in .CLUB
- Not following IDNA rules
 - xn--taylorswift-nu5j.tokyo taylor·swift.tokyo

Silly games

```
$ whois xn--gx5aaaaaaaa...aaa.top
```

```
The queried object does not exist:
```

```
Name is reserved by Registry
```

```
Primary IDN: xn--gx5aaaaaaaa...aaa.top
```

```
$ host xn--gx5aaaaaaaa...aaa.top
```

```
xn--gx5aaaaaaaa...aaa.top has address 23.234.27.100
```

Summary

- Most IDNs are OK
- Long tail of old junk
- Some new TLDs need compliance help