

Real-time Detection of Internationalized Brand Abuse And Other IDN-Based Misconduct

Mike Schiffman

2018-10-14

DNS-OARC 29

Amsterdam, the Netherlands

ObBio

- Mike Schiffman <mschiffm@fsi.io>
- Doing the computer security for ~25 years
- Previous: Cisco, @stake, Guardent, ISS
- Current: Engineering Team Lead for Farsight Security, Inc



“The goal of an IDN effort is not to be able to write the great Klingon (or language of one's choice) novel in DNS labels but to be able to form a usefully broad range of mnemonics in ways that are as natural as possible in a very broad range of scripts.”

-- RFC 5894



Tell Me More About How You're Into That Bitcorn

Found at <https://coinbase.com>

coinbase

Products Help Charts Sign In Sign Up

Sign in to Coinbase

Email

Password

☐ Keep me signed in on this computer SIGN IN

[Forgot password? Don't have an account? Sign Up](#)

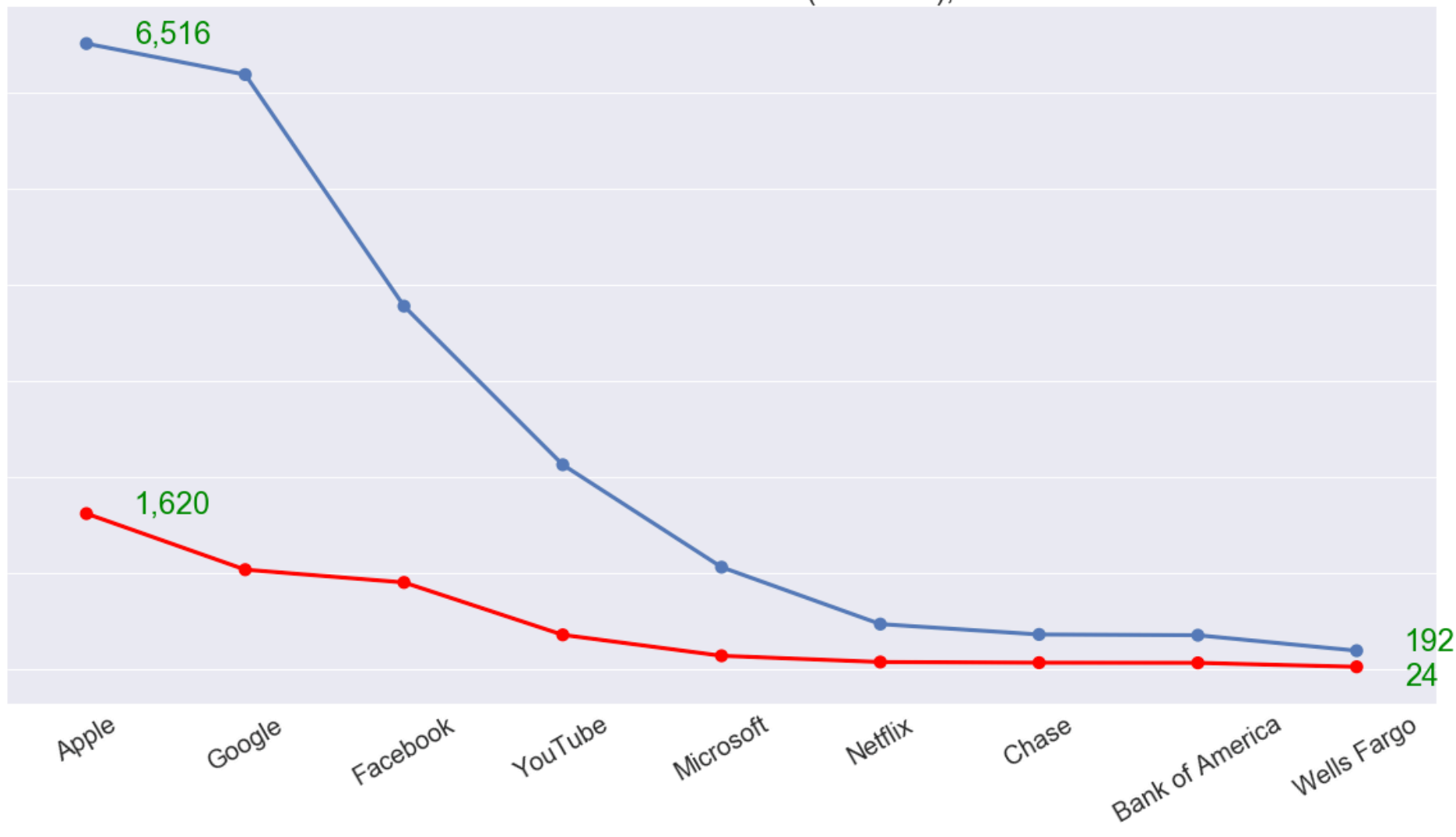
[Have an issue with 2-factor authentication?](#)

(Circa March–April 2018)

IDN Homographs: Maybe You've Heard Of These Guys?

(Snowflake count in red)

WELL-KNOWN BRAND IDN HOMOGRAPHS (TOTALS), JAN 2017 - AUG 2018



IDN Homographs: Samples From The Field

facebook.com.	apple.com.	ñetflix.com.	google.xyz.	bankofamerica.com.	wellsfargo.com.
facebøok.com.	applè.com.	netflix.com.	goôgle.com.	bankofamerica.com.	weltsfargo.com.
facebook.tk.	âplê.cf.	nétflix.com.	göogle.com.	bankofamerica.net.	wellsfárgo.com.
fäcebook.com.	ápple.com.	nètflix.com.	googlé.com.	bankofamerica.com.	wellsfårgo.com.
facebook.com.	äpple.com.	neţflix.com.	goöoglē.com.	bankôfamerica.com.	wellsfargó.com.
fácebook.com.	åpple.com.	netflíx.com.	googlè.tk.	banköfamerica.com.	wellsfargø.com.
fàcebook.com.	äpple.com.	netflix.com.	google.com.	bankofamerîca.com.	wellsfargo.com.
fâcebook.com.	apple.com.	netflîx.com.	google.com.	bänkofämericä.com.	çhase.com.
fâcebook.com.	apple.com.	netflix.com.	googlé.com.	bankofamerica.com.	chàse.com.
facebook.com.	applè.com.	netflix.com.	göoogle.com.	bankofamerica.net.	chäse.com.
fäcebook.com.	applé.com.	netflix.com.	google.com.	bankofamerica.com.	chasé.com.
facebook.com.	applè.com.	netflix.com.	googlè.com.		chasë.com.
facebook.com.	àpplè.com.		googlé.com.		chase.com.
fäcebook.com.	applë.com.		google.com.		chase.com.
facebook.com.	äpplë.com.		goôgle.com.		
facebook.com.	âpplê.com.		goôgle.com.		
fäcebook.com.	àpplê.com.		goôgle.com.		
fačebook.com.	applë.com.		goôgle.com.		
facebook.com.	applë.com.		goôgle.com.		
facebook.com.	äpplë.com.		goôgle.com.		
facebook.com.	applè.com.		goôgle.com.		
faceébook.com.	åpplè.com.		google.com.		

This font used in this presentation is Lucida Grande, a serif-free font conventionally used by many browsers, websites, and blogs (including Facebook)

Disclaimer!

All of the organizations whose homographs appear in
this slide deck were notified after discovery before any
public discussion.

Agenda

What We've Seen

What We See

How We See

What Can We All Do?

Bonus Stuff 🕒

IDN Homograph Vulnerability: Enabled By Two Things

1. Evolutionarily, humans are really good at pattern recognition

2. Many Unicode glyphs look similar or identical to others when rendered in many fonts



IDN Homograph Attacks: Touched By An IDN

- Register an IDN that is a homograph of a well-known (usually non-internationalized) site
- ...To extort, camp, cash-park, phish, distribute malware, or do other antisocial things

google.com

vs.

google.com

This “g” is Basic Latin (U+0067)

This “g” is Extended Latin (U+0261)

(The Unicode Consortium calls such code points “confusables”)

IDN Homograph Attacks: Samples From The Field

Real Site

Homograph

Punycode

easyjet.com.

easyjet.com.

xn--easyje-n17b.com.

delta.com.

delta.com.

xn--deta-1kb.com.

ryanair.com.

ryanair.com.

xn--ryanai-1x7b.com.

poloniex.com.

poloniex.com.

xn--polonex-3ya.com.

coinbase.com.

coinbase.com.

xn--coinbse-30c.com.

bittrex.com.

bittrex.com.

xn--btrex-m3a12b.com.

facebook.com.

facebооk.com.

xn--80akppap2f62a.com.

amazon.com.

amażon.com.

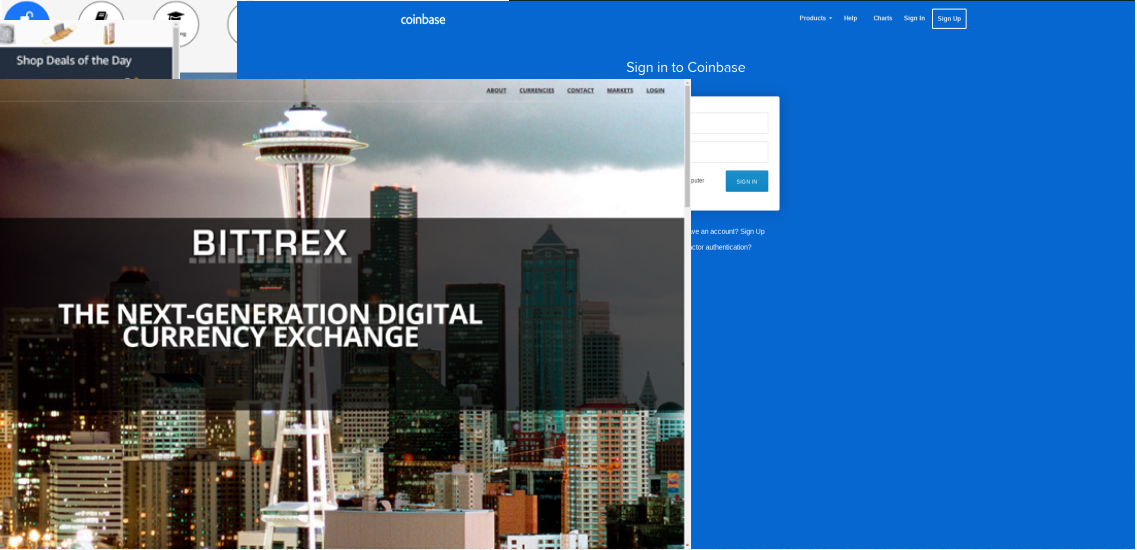
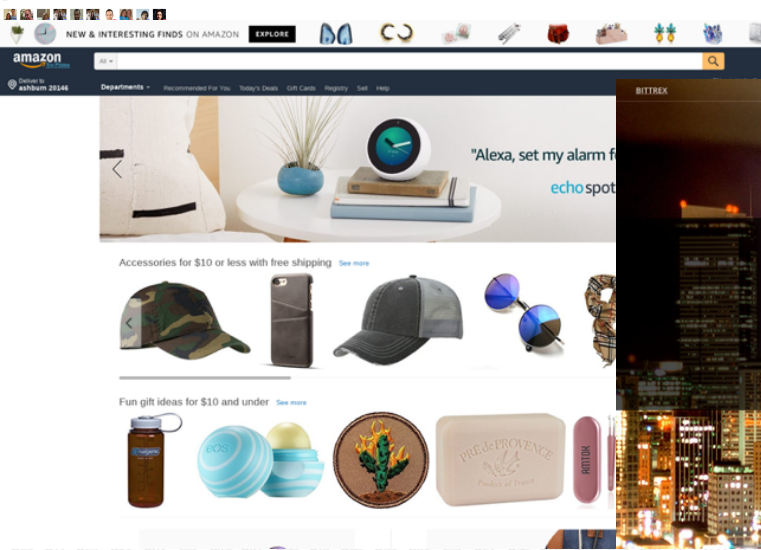
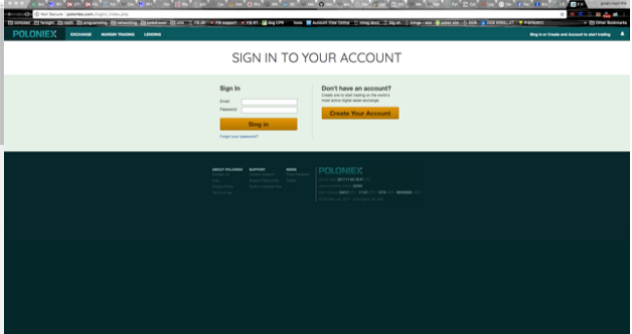
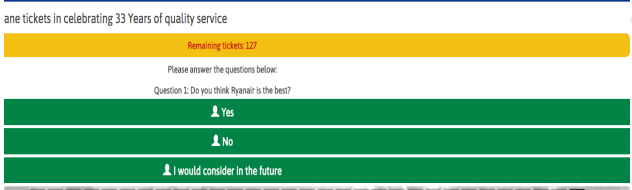
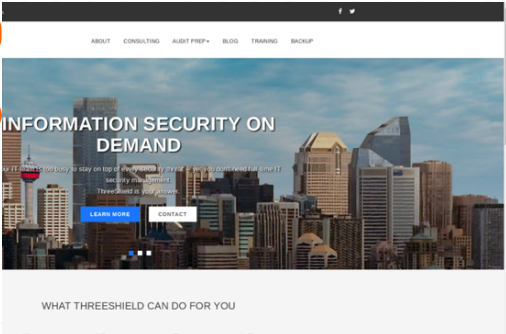
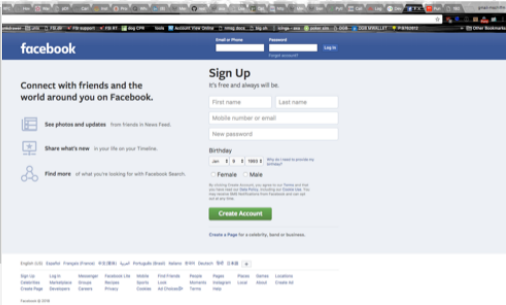
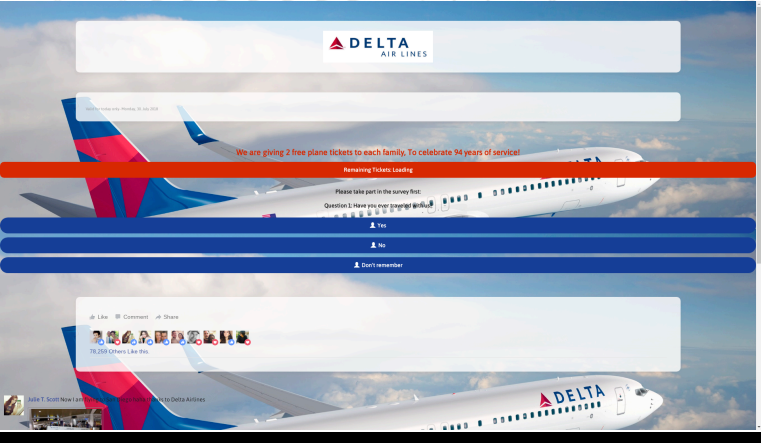
xn--amaon-7hb.com.

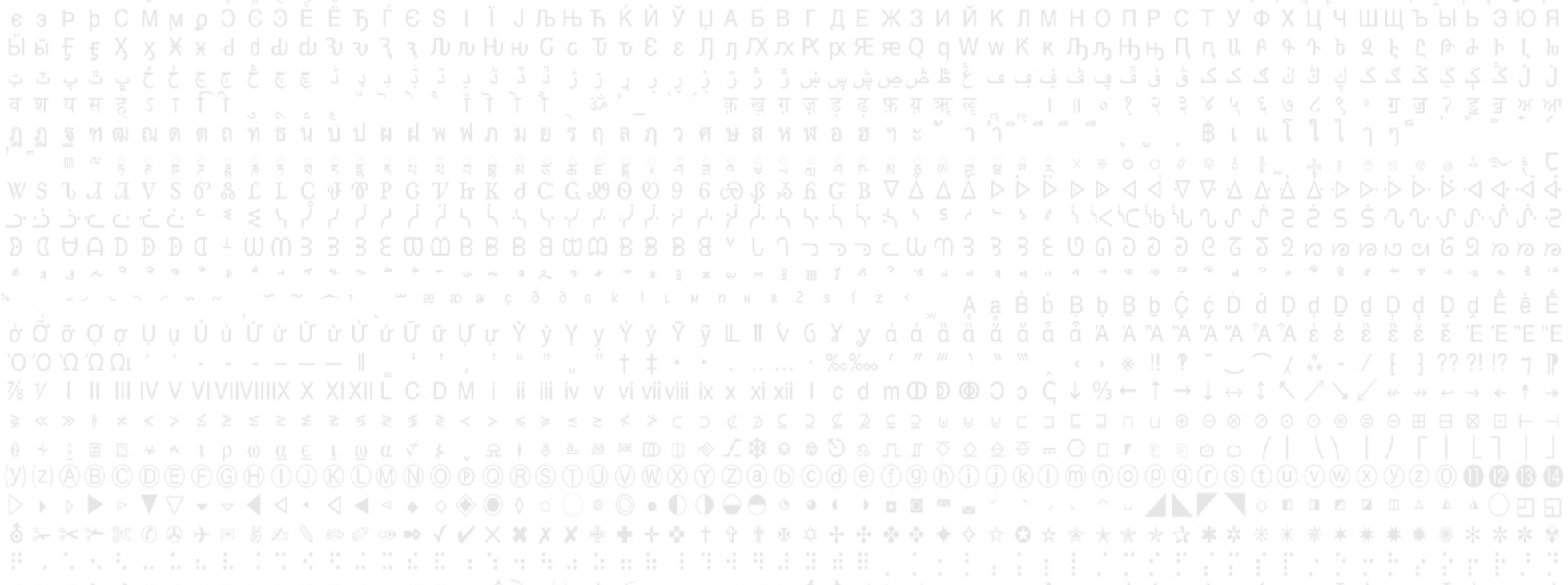
linkedin.com.

linkedin.com.

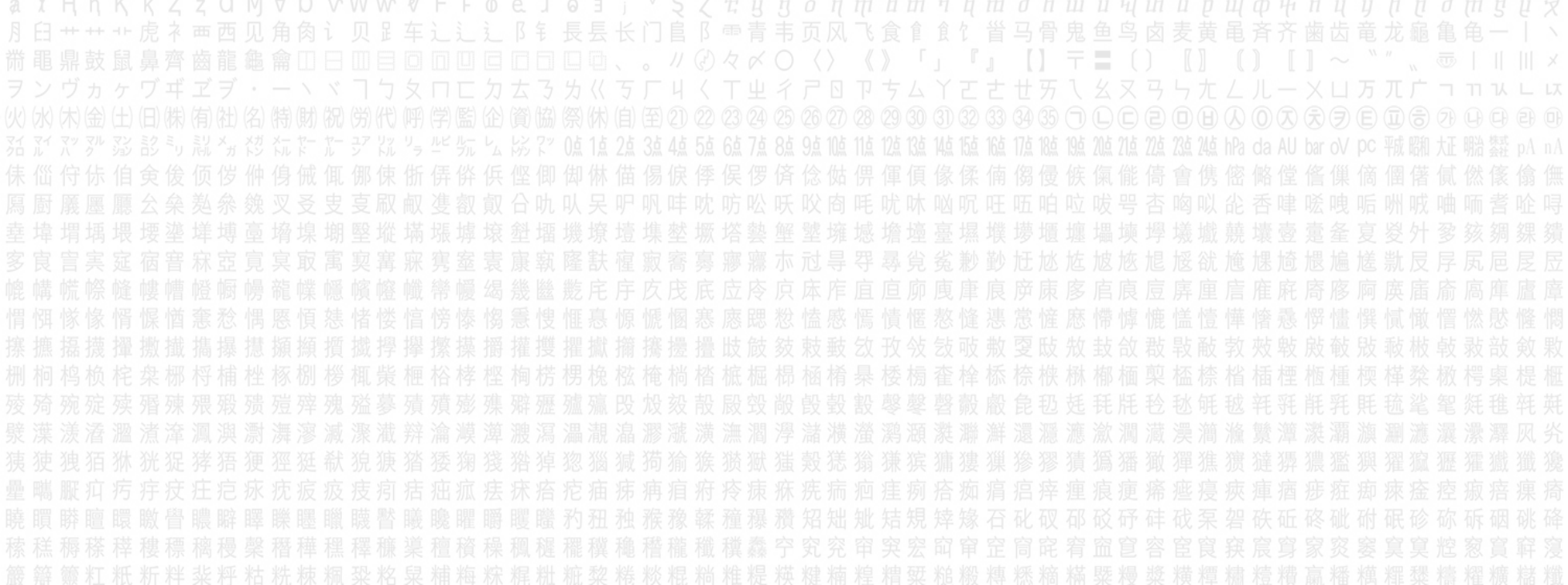
xn--lnkedin-zya.com.

IDN Homograph Attacks: Here's a Cool Collage I Made





What We've Seen / What We See



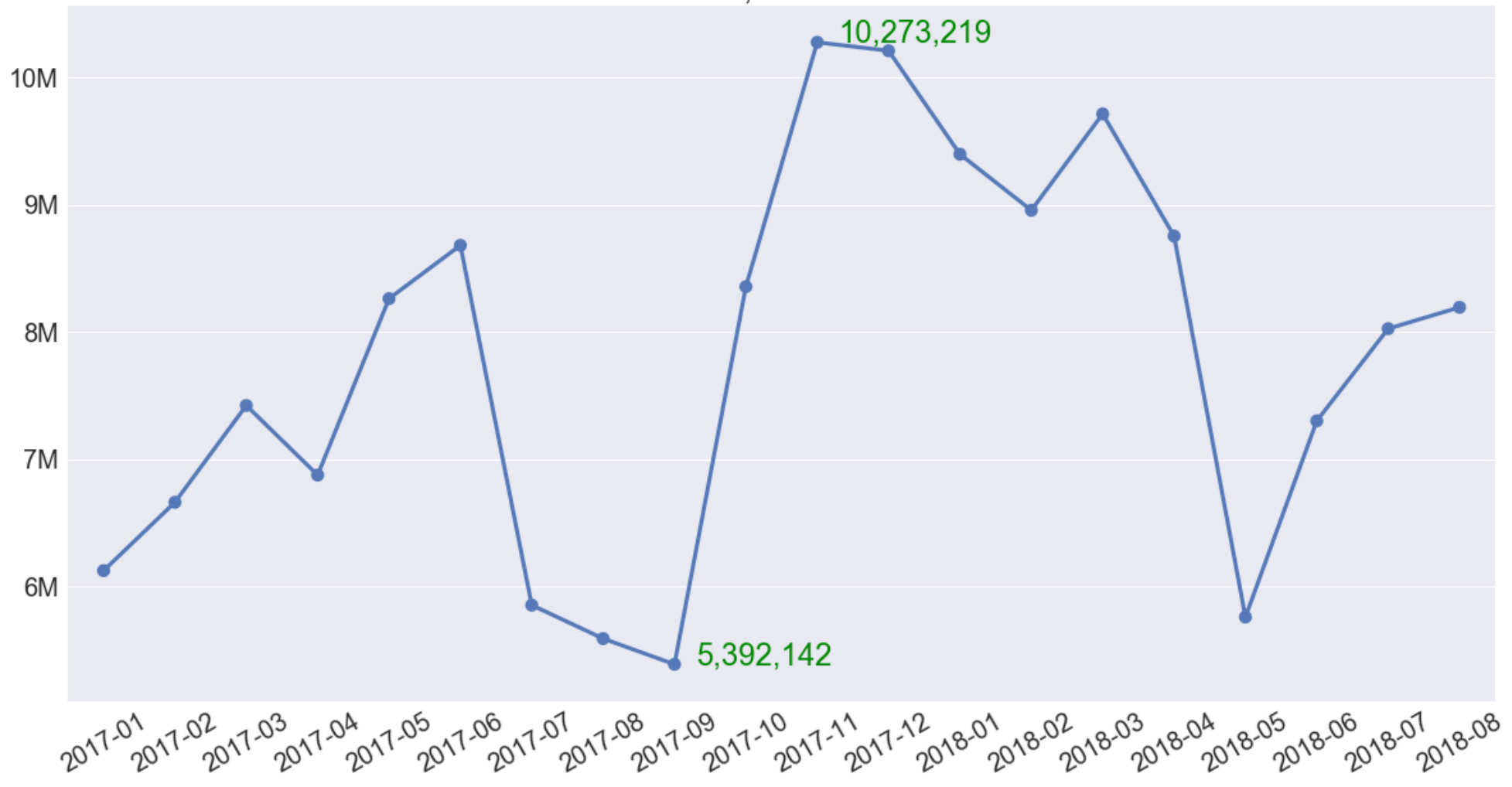
What We've Seen: All The IDNs

161,935,465 total IDN observations

34,460,574 total unique IDNs

Summer vacation?

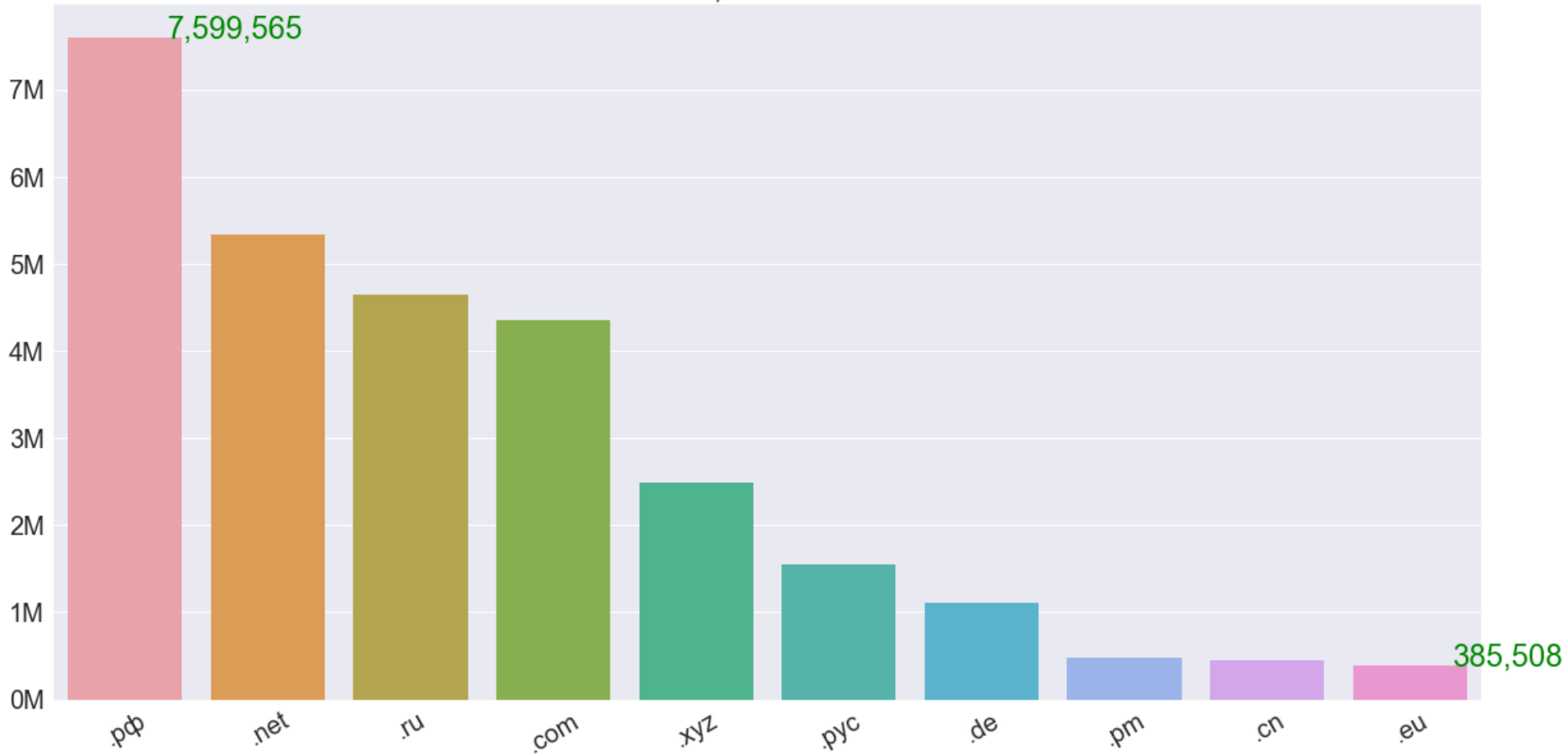
OBSERVED IDNs, JAN 2017 - AUG 2018



What We've Seen: **Top 10 IDN TLDs**

1,675 total unique TLDs

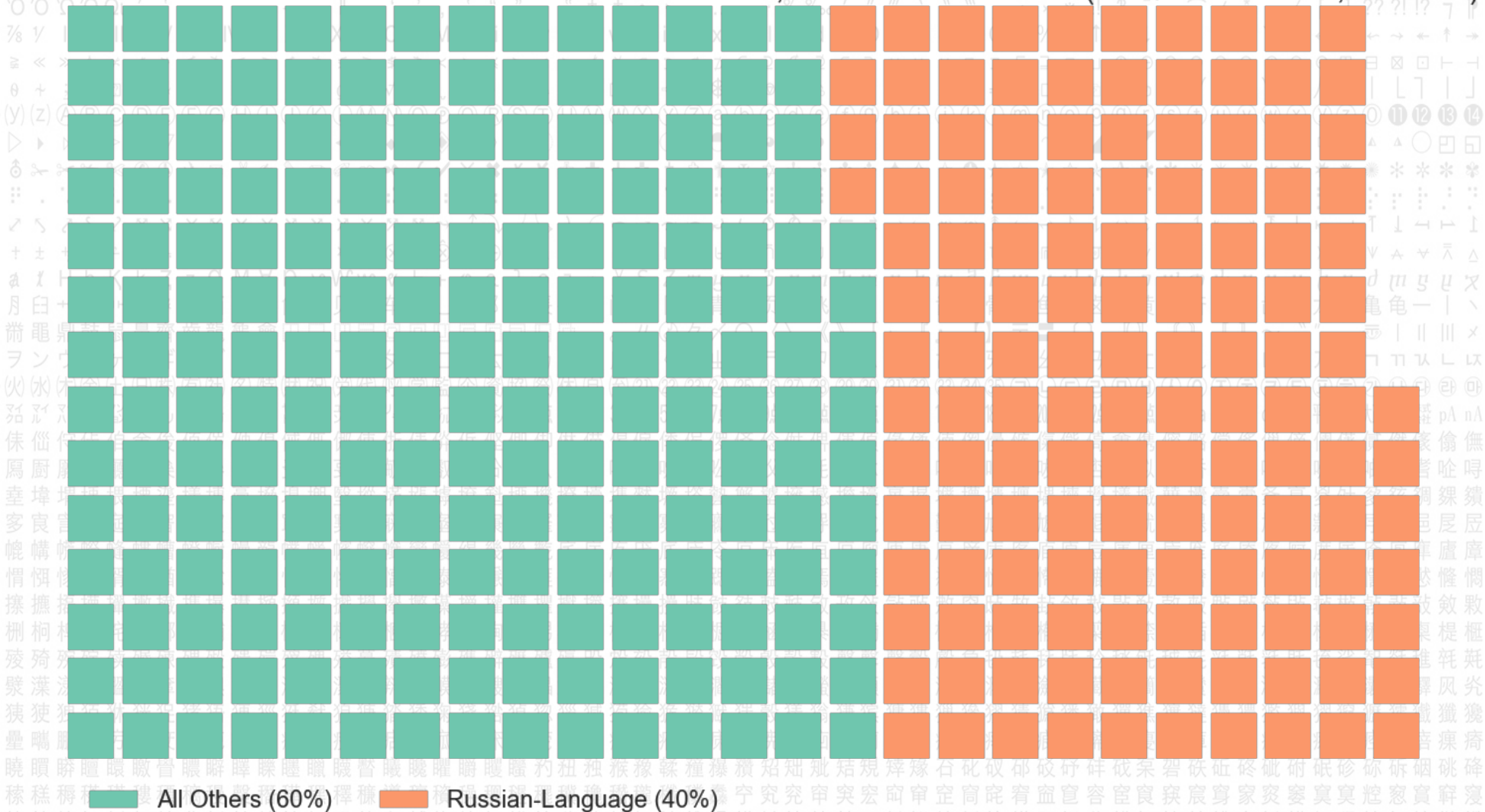
TOP TEN TLDs, JAN 2017 - AUG 2018



What We've Seen: IDN TLD, Russia vs Everyone Else

Of the 34.5M total unique IDNs, 40% belong to a Russian-language

IDN TLD DISTRIBUTION RUSSIAN-LANGUAGE V WORLD, JAN 2017 - AUG 2018 (1 SQUARE == 100,000 TLDs)



What We've Seen: What's in a Brand?

- Brands and service marks normalized as Basic Latin
- From three to 24 characters
- Collapse spaces and hyphens, remove diacritics

What We've Seen: Which Brands?

- Banks, Financial Management, Insurance, Ecommerce
- Credit & Loans, Consumer Goods, Cryptocurrencies
- Airlines, Pharma, Computer Security Companies
- etc...
- 509 total

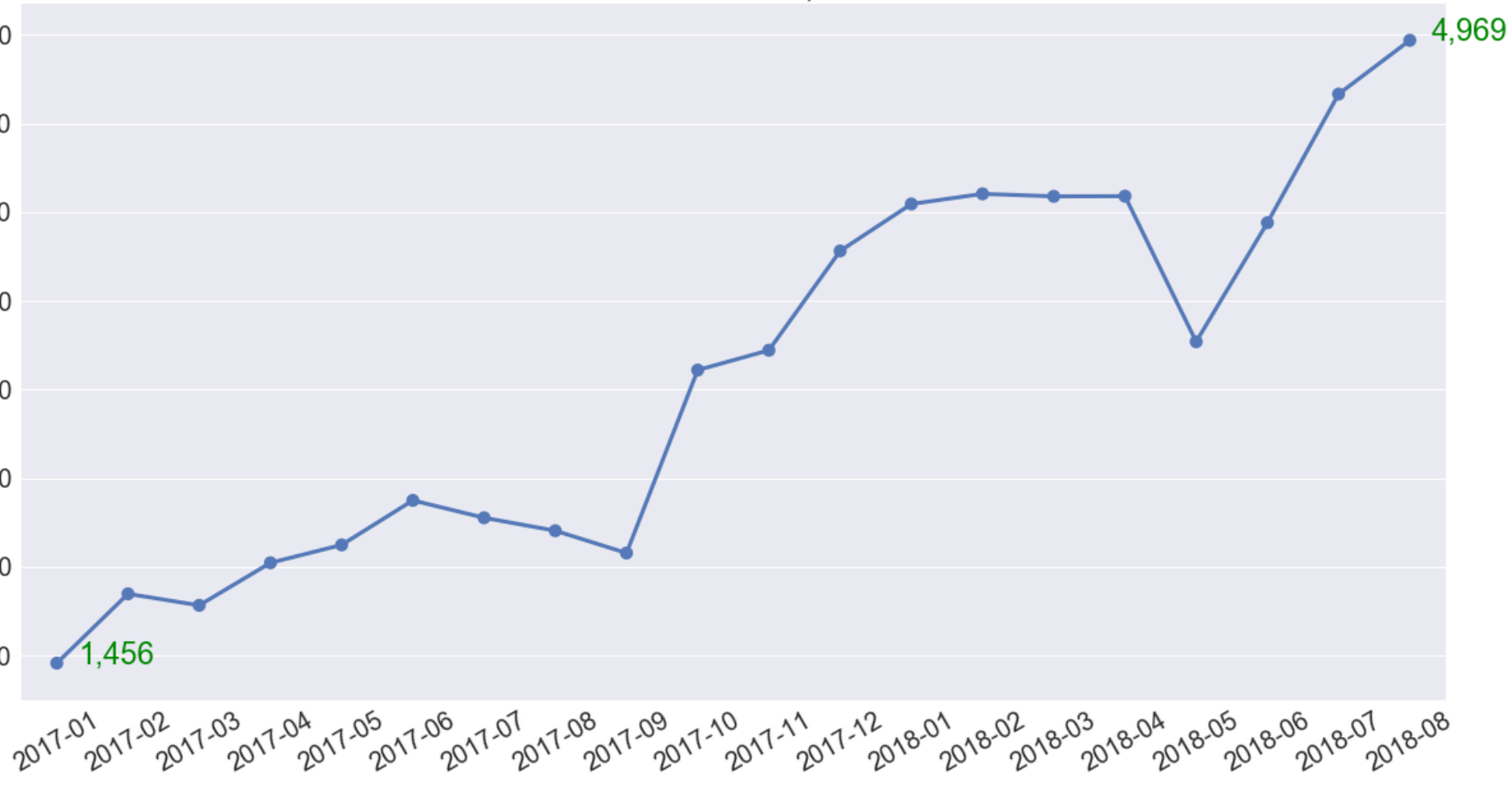
A curated and de-deuplicated blend of companies
from Alexa Top 100 and SimilarWeb.

What We've Seen: IDN Homographs

61,443 total IDN homograph observations

11,766 total unique IDN homographs

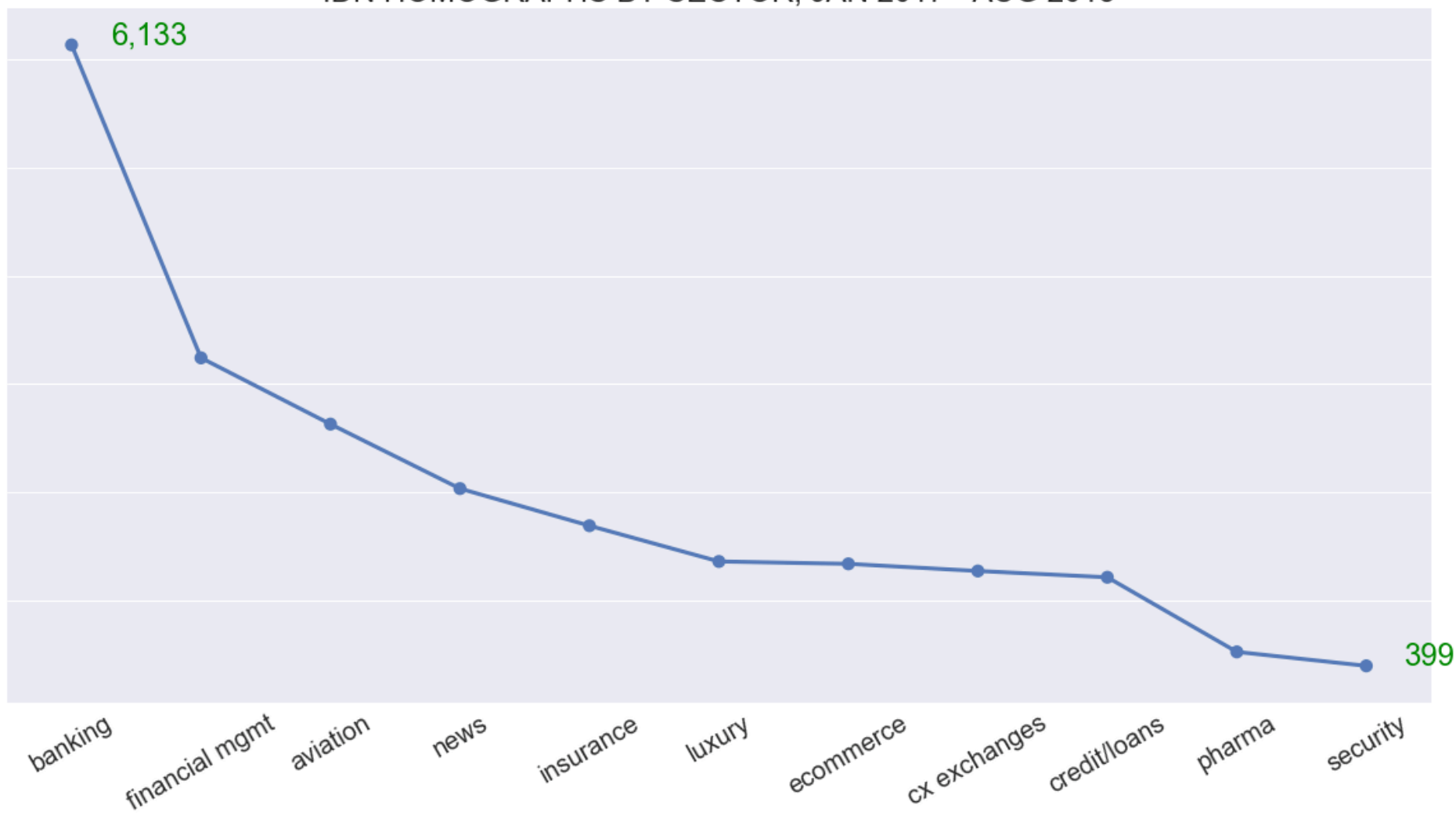
OBSERVED IDN HOMOGRAPHS, JAN 2017 - AUG 2018



What We've Seen: IDN Homographs by Sector

Of the 61,443 total IDN homograph observations,
20% are in banking/finance

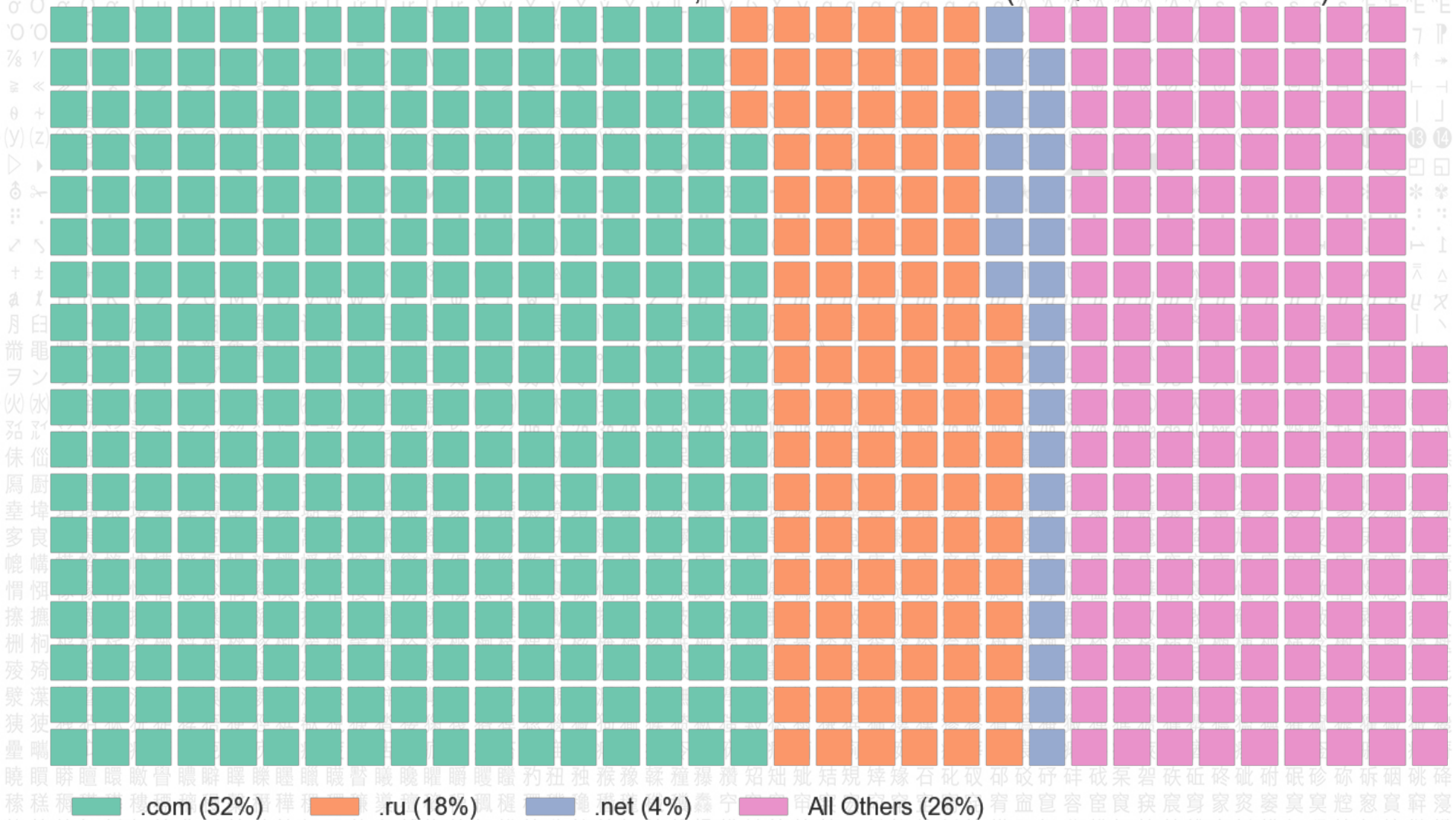
IDN HOMOGRAPHS BY SECTOR, JAN 2017 - AUG 2018



What We've Seen: IDN Homographs by TLD

52% live in .com (no surprise)

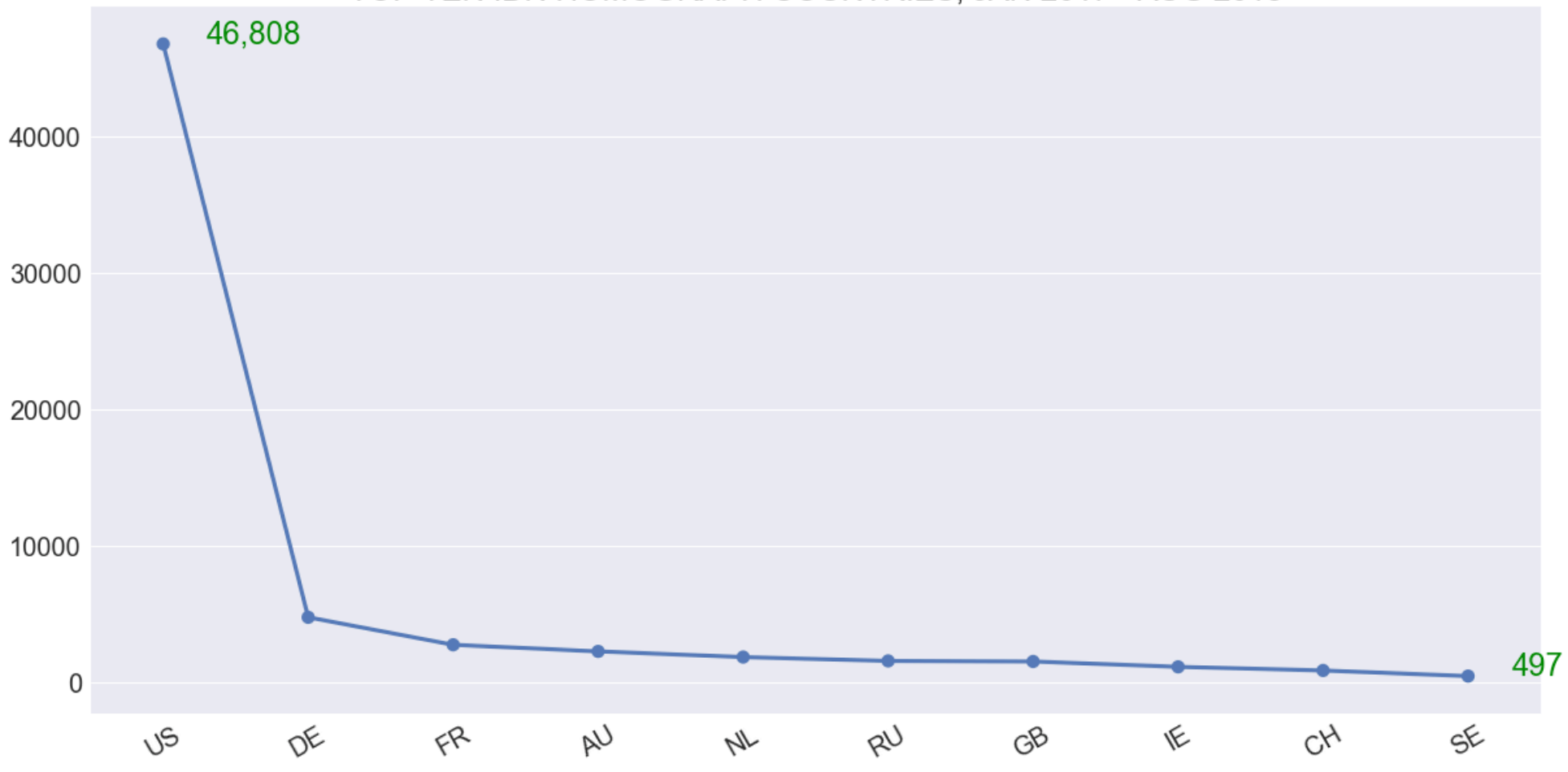
IDN HOMOGRAPH TLD DISTRIBUTION, JAN 2017 - AUG 2018 (1 SQUARE == 20 TLDs)



What We've Seen: IDN Homographs by Country

68% purport to be in the US
Cheap hosting / US-based CDNs?

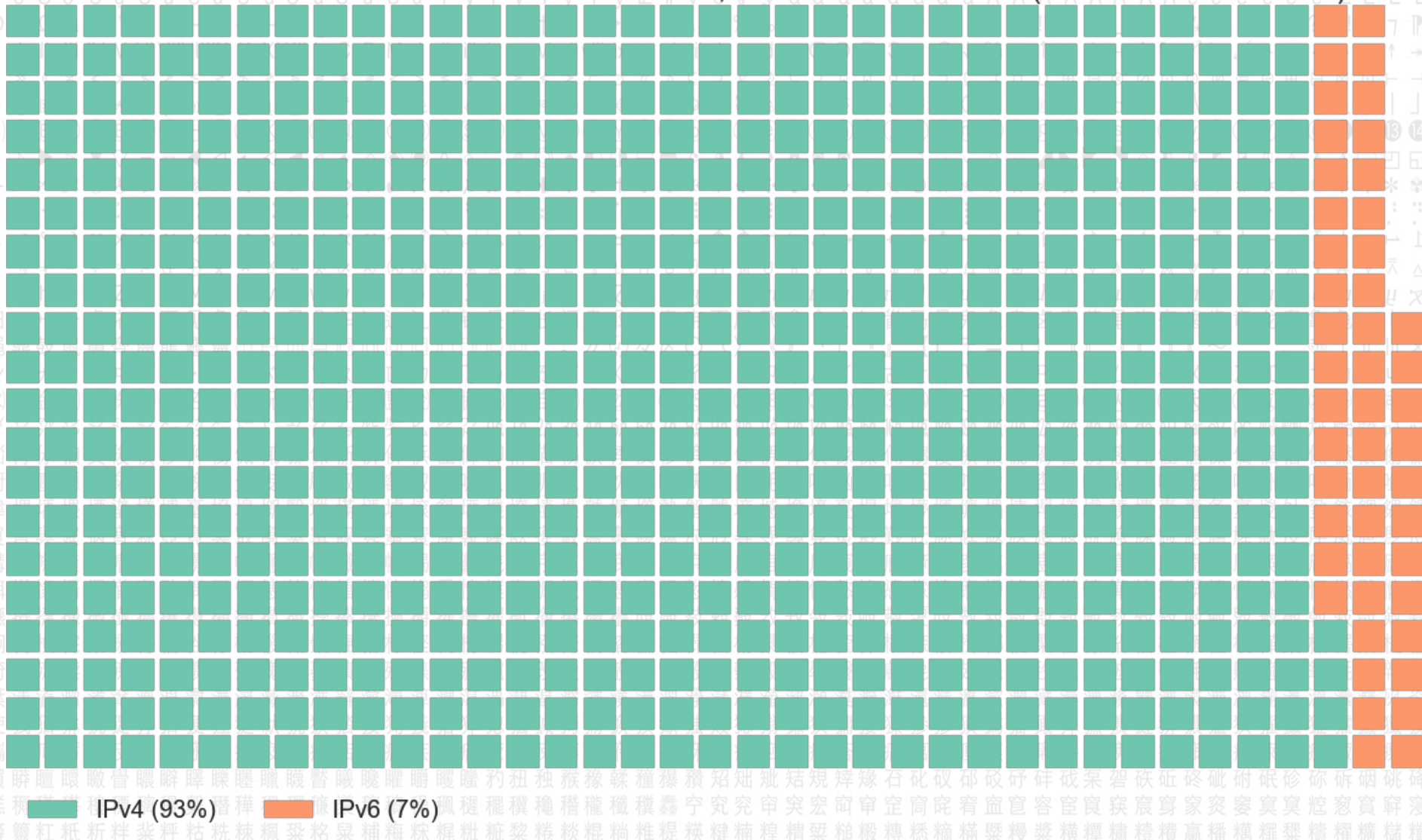
TOP TEN IDN HOMOGRAPH COUNTRIES, JAN 2017 - AUG 2018



What We've Seen: IDN Homographs by IP Version

IPv4 still reigns supreme

IDN HOMOGRAPH IPv4 vs IPv6 DISTRIBUTION, JAN 2017 - AUG 2018 (1 SQUARE == 100 IPs)



What We've Seen: A Gallery of Interesting Sites

- Back in March 2018, I had this idea...
- “Let’s treat homographs as websites and build a dashboard to display them in real-time...”

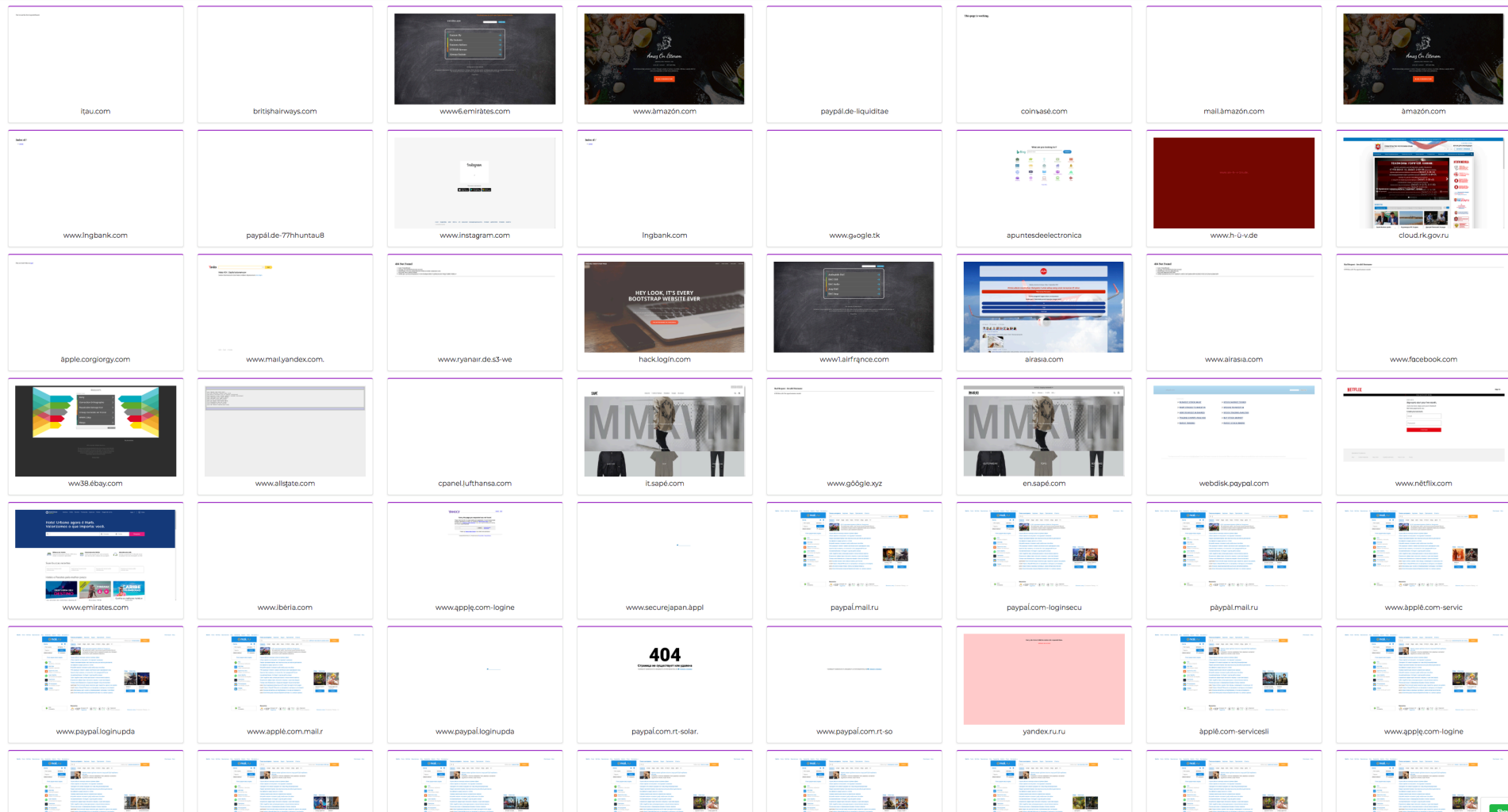


No assumption of intent on what follows!

How We See: The IDN Checker Dashboard



Farsight IDN Checker



Hey wait... One of those looks suspicious...



How We See: The IDN Checker Dashboard (zoom-in)

xn--airasa-t9a.com (airasia.com)

The screenshot displays the AirAsia website's IDN checker interface. At the top, the AirAsia logo is visible. Below it, a banner announces a promotion: "Berlaku untuk hari ini hanya - Rabu, 5. September 2018" and "AirAsia adalah memberikan tiket gratis 2 untuk setiap orang untuk merayakan 25 tahun". A progress bar indicates "Tiket tersisa: Pemuatan...". The main survey question is "Silakan mengambil bagian dalam survei pertama: Pertanyaan 1: Telah Anda pernah bepergian dengan kami?". Three response options are provided: "Ya", "Tidak", and "Tidak ingat". Below the survey, a social media feed shows a post from Julie T. Darden thanking AirAsia for two free tickets, accompanied by an image of an AirAsia ticket. The post has 78,259 likes and 28 comments.

AirAsia

Berlaku untuk hari ini hanya - Rabu, 5. September 2018

AirAsia adalah memberikan tiket gratis 2 untuk setiap orang untuk merayakan 25 tahun

Tiket tersisa: Pemuatan...

Silakan mengambil bagian dalam survei pertama:
Pertanyaan 1: Telah Anda pernah bepergian dengan kami?

Ya
Tidak
Tidak ingat

Seperti · Komentar · Berbagi

78,259 Orang lain seperti ini.

Julie T. Darden Terima kasih airasia, Untuk 2 tiket pesawat gratis

Seperti · Jawaban · 50 · 28 Menit

James Brassard Ini adalah waktu untuk pembelian, terima kasih airasia Hehe

Date Seen: 2018-09-05 05:37:44 (UTC)
Found in: white

Gallery of Interesting Sites: bittrex.com.

Toggle navigation [Bittrex](#)

- [About](#)
- [Currencies](#)
- [Contact](#)
- [Markets](#)
- [Login](#)

BITTREX

The Next-Generation Digital Currency Exchange

- 1.
- 2.
- 3.

USA Based

Proudly based in the United States, we collaborate with experts in USA financial law to ensure that we remain in compliance with the evolving legal landscape and always available to our worldwide community. Our firm dedication to being legally compliant and fully regulated means that we were one of the first companies to apply for New York's BitLicense, allowing us to proudly serve New York customers.

Security First

With over 50 years of security experience on our founder's resumes, security is our number one concern and priority. Our systems are constantly upgraded and tested to ensure that we are exceeding industry-best standards. To protect users we require two-factor authentication for all withdrawals and API usage. The entirety of Bittrex.com is protected by SSL, so you can rest easy about the safety of your funds and personal information.

Support for Algorithmic Trading

While the Bittrex.com trading interface is designed to provide an intuitive and efficient trading experience, some of our traders enjoy trading through third-party platforms or designing their own algorithmic trading bots. With our dedication to providing the best trading experience possible, we designed our APIs to allow for high-frequency trading bots, and to allow mining pools to be build upon our platform.

Unfinished phish?

Gallery of Interesting Sites: bittrex.com.



DoctoPDF Offers a Document Converter Tool and Free Web Search on your Chrome New Tab.

Free Doc to PDF Files Converter



- Convert Doc File to PDF Format in Seconds
- Quick Access to Files Converter Tool
- DoctoPDF a Simple Doc to PDF Converter Tool

Add to Chrome

By clicking the button above you agree to install DoctoPDF Chrome extension and have read and agreed to the [EULA](#) & [Privacy Policy](#)

Adware, anyone?

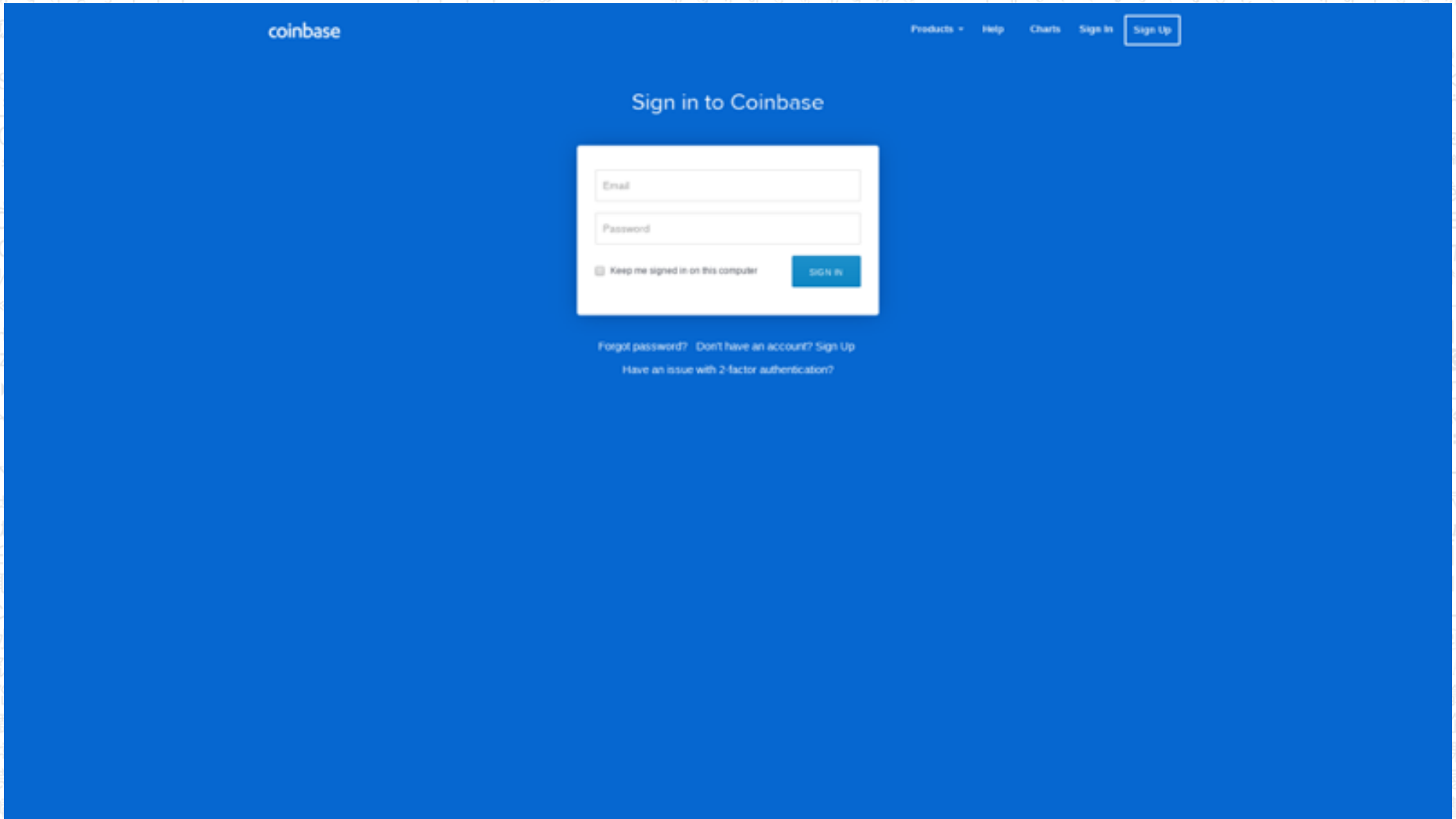
[EULA](#) [Privacy Policy](#) [How to Uninstall](#) [FAQ](#)

Gallery of Interesting Sites: bittrex.com.



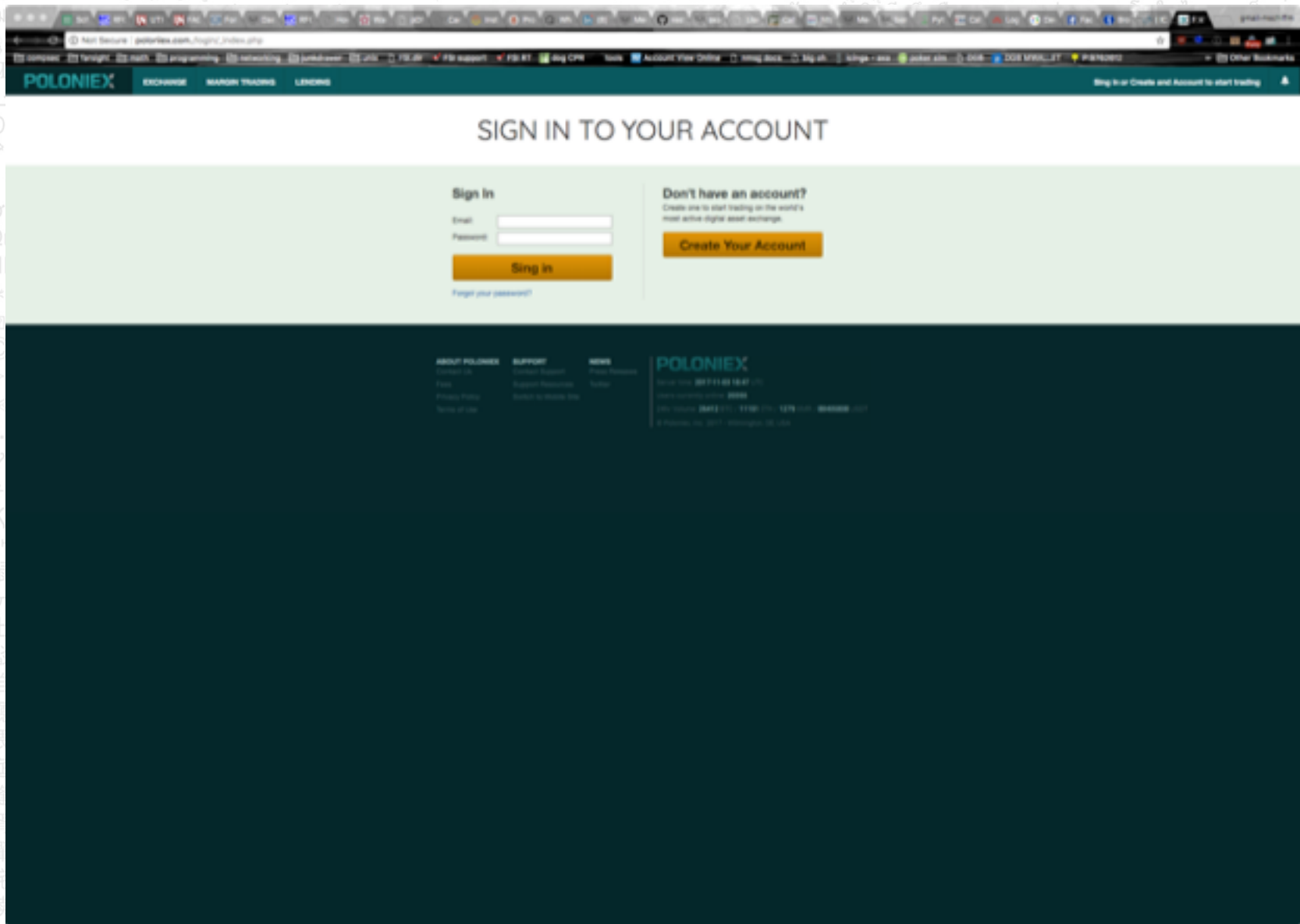
BTC PLS FTW

Gallery of Interesting Sites: **coinbase.com. coínbasé.com.**



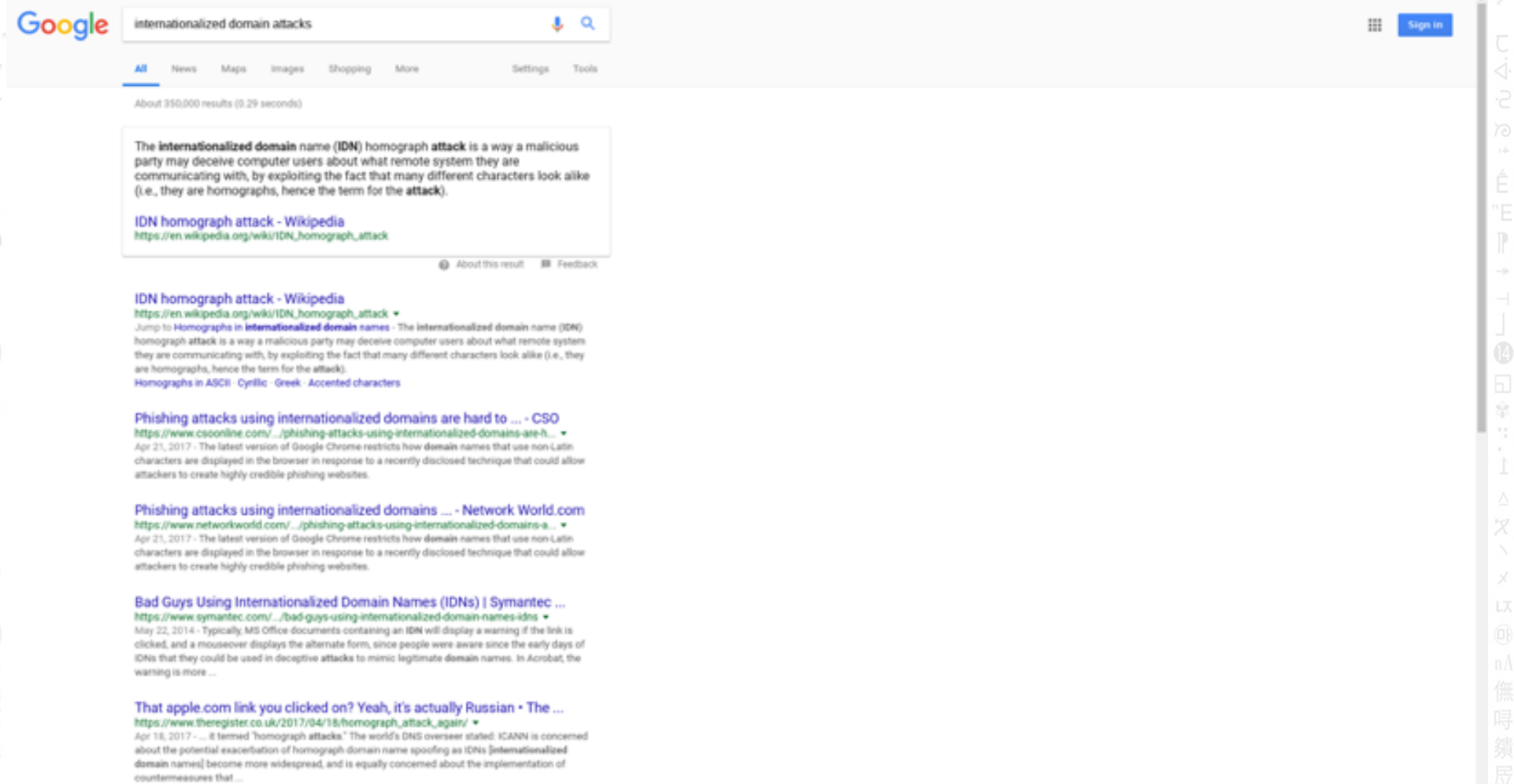
MOAR BTC PLS FTW

Gallery of Interesting Sites: poloniex.com.



WILLPHISH4ETH

Gallery of Interesting Sites: [google.com](https://www.google.com).



Google internationalized domain attacks

About 350,000 results (0.29 seconds)

The **internationalized domain name (IDN) homograph attack** is a way a malicious party may deceive computer users about what remote system they are communicating with, by exploiting the fact that many different characters look alike (i.e., they are homographs, hence the term for the attack).

[IDN homograph attack - Wikipedia](https://en.wikipedia.org/wiki/IDN_homograph_attack)
https://en.wikipedia.org/wiki/IDN_homograph_attack

About this result Feedback

IDN homograph attack - Wikipedia
https://en.wikipedia.org/wiki/IDN_homograph_attack •
Jump to **Homographs in internationalized domain names** • The internationalized domain name (IDN) homograph attack is a way a malicious party may deceive computer users about what remote system they are communicating with, by exploiting the fact that many different characters look alike (i.e., they are homographs, hence the term for the attack).
Homographs in ASCII Cyrillic Greek Accented characters

Phishing attacks using internationalized domains are hard to ... - CSO
<https://www.csoonline.com/.../phishing-attacks-using-internationalized-domains-are-h...> •
Apr 21, 2017 • The latest version of Google Chrome restricts how domain names that use non Latin characters are displayed in the browser in response to a recently disclosed technique that could allow attackers to create highly credible phishing websites.

Phishing attacks using internationalized domains ... - Network World.com
<https://www.networkworld.com/.../phishing-attacks-using-internationalized-domains-s...> •
Apr 21, 2017 • The latest version of Google Chrome restricts how domain names that use non Latin characters are displayed in the browser in response to a recently disclosed technique that could allow attackers to create highly credible phishing websites.

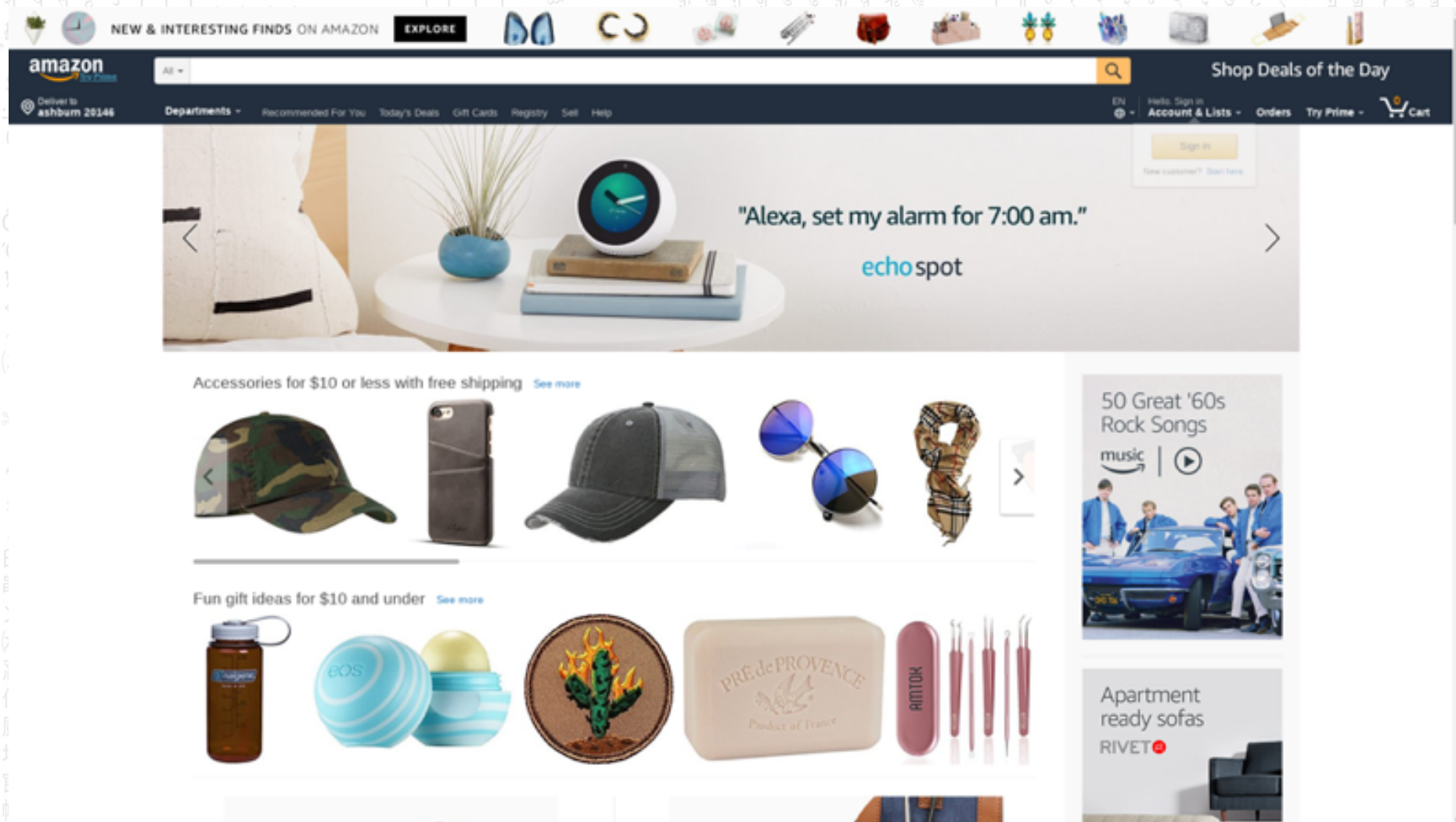
Bad Guys Using Internationalized Domain Names (IDNs) | Symantec ...
<https://www.symantec.com/.../bad-guys-using-internationalized-domain-names-idns> •
May 22, 2014 • Typically, MS Office documents containing an IDN will display a warning if the link is clicked, and a mouseover displays the alternate form, since people were aware since the early days of IDNs that they could be used in deceptive attacks to mimic legitimate domain names. In Acrobat, the warning is more ...

That apple.com link you clicked on? Yeah, it's actually Russian • The ...
https://www.theregister.co.uk/2017/04/18/homograph_attack_again/ •
Apr 18, 2017 • ... it termed "homograph attacks." The world's DNS overseer stated: ICANN is concerned about the potential exacerbation of homograph domain name spoofing as IDNs [internationalized domain names] become more widespread, and is equally concerned about the implementation of countermeasures that ...

Who's watching the watchers?



Gallery of Interesting Sites: amazon.com.



Please be entering your credit card infos

Gallery of Interesting Sites: apple.com.

Big Billion Days



Spin The Lucky Wheel!

Only this Wednesday we give our members 1 free spin for a chance to win exclusive prizes!



Only one game allowed per IP.


Like Comment Share Delete


12,068 others like this


[View more comments](#)

15 of 1,356

 **Xavier Stebin** I was lucky enough to get the fitbit surge fitness super Watch! I still can't believe this is real, hahah!
4 minutes ago Like 37

 **Shyam Mohan** I got the Lenovo Vibe K4 Note.. Thanks Thanksass.. This is the best day ever..
13 minutes ago Like 24

 **Deepu Antony** Amazing game.. Just found it. Hope I will win. EDIT: I got nothing :(
29 minutes ago Like 14

 **Tebin Joseph** Krite Friends nu onu Headset kiti, Enikkonnum kittiyilaa.. :)
35 minutes ago Like 11

 **Karan Singh** Dhanyavad, mujhe mera prize mil gaya.. :D
37 minutes ago Like 9

 **Ravi Verma** Has anyone actually won yet? I spun 2 times from both of my accounts but nothing..
41 minutes ago Like 8

 **Survival Sandhi** arrey yar mujhe kuch n ahi mila..
56 minutes ago Like 5

 **Yasmin Chaudhary** I got nothing.. why am I always so unlucky :(
1 hour ago Like 5

Spam / PUP

Gallery of Interesting Sites: [google.com](https://www.google.com).

CLICK CONFIRMATION

Your click has been flagged as suspicious - perhaps you are clicking on lots of ads today?

Please help us determine if our filters are accurate:



Adware

Gallery of Interesting Sites: **delta.com.**



Valid for today only- Monday, 30 July 2018

We are giving 2 free plane tickets to each family, To celebrate 94 years of service!

Remaining Tickets: Loading

Please take part in the survey first:

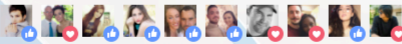
Question 1: Have you ever traveled with us?

☐ Yes

☐ No

☐ Don't remember

Like Comment Share



78,259 Others Like this.

Julie T. Scott Now I am flying to San Diego haha thanks to Delta Airlines



Spam and scam

Gallery of Interesting Sites: easyjet.com. easyjet.com.

easyJet

Friday, 3. August 2018

We are celebrating our 25th anniversary and Giving away 2 Free tickets to everyone!

Remaining Tickets: 354

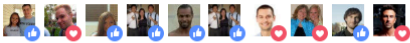
Please answer the questions below first:

Question 1: Have you ever fly with Easyjet?

Yes

No

Like Comment Share



17,259 Others Like this.



Rachel Singleton Thank you soo much Easyjet for Tickets! i got mine

Like · Reply · 50 Å· 28 mins



Julia Schröder It's time for some good fly @ Easyjet

Spam and scam

Gallery of Interesting Sites (12/12): ryanair.com.



Only today - Friday, 3. August 2018

Ryanair is rewarding everyone with 2 free plane tickets in celebrating 33 Years of quality service

Remaining tickets: 127

Please answer the questions below:

Question 1: Do you think Ryanair is the best?

☐ Yes

☐ No

☐ I would consider in the future

Spam and scam

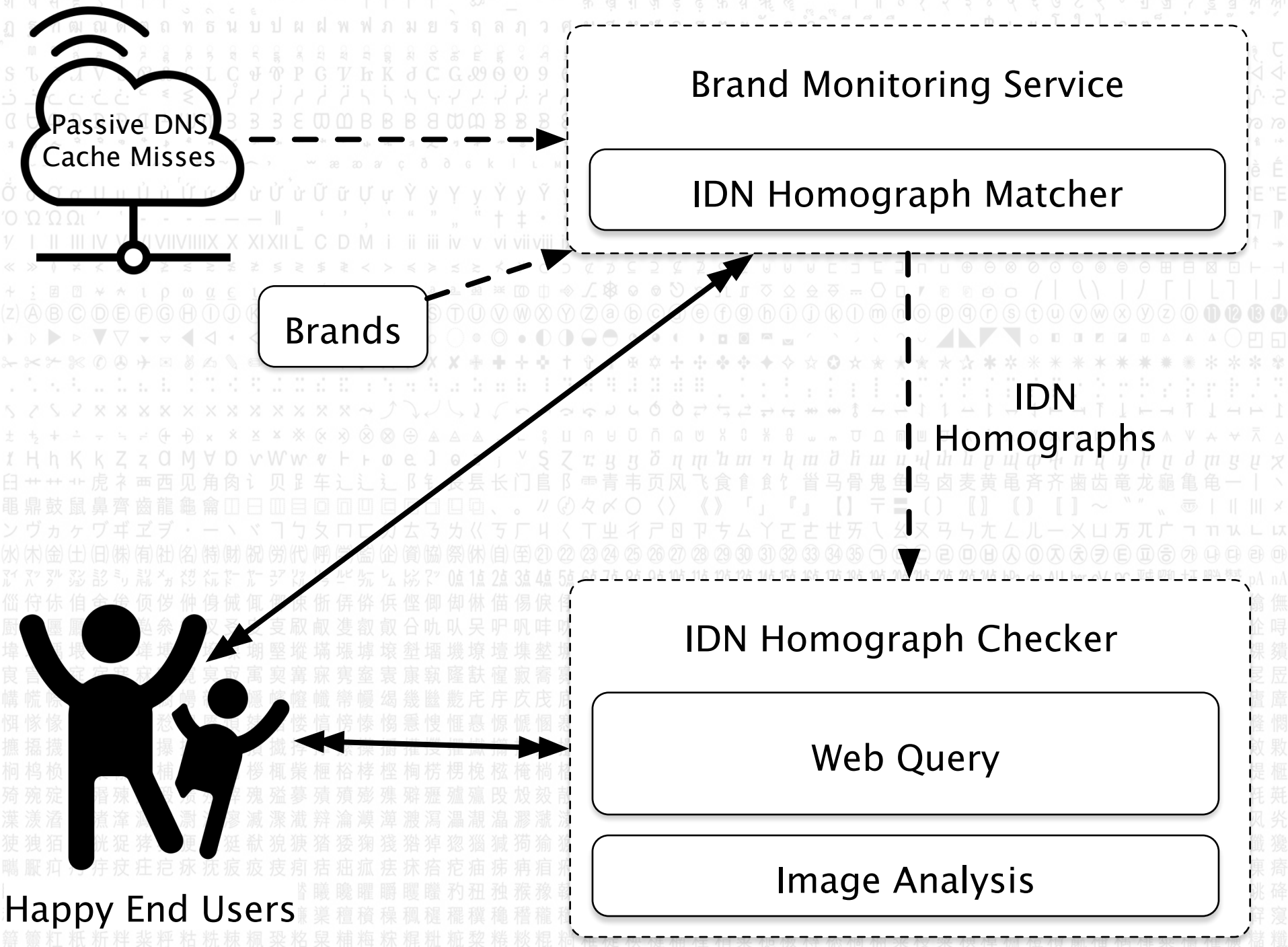
What We See: The IDN Daily News

- ~11,577,600 total IDN observations
- ~15,465 “new” IDNs
- ~3,400 homographs (from our list of 509 brands)



How We See

How We See: The IDN Homograph Pipeline



How We See: Homoglyphification

- To detect homographs we first must build a tables of homoglyphs

1. Choose font(s)

2. Generate glyphs

3. Compare glyphs

4. Prune output into final homoglyph tables

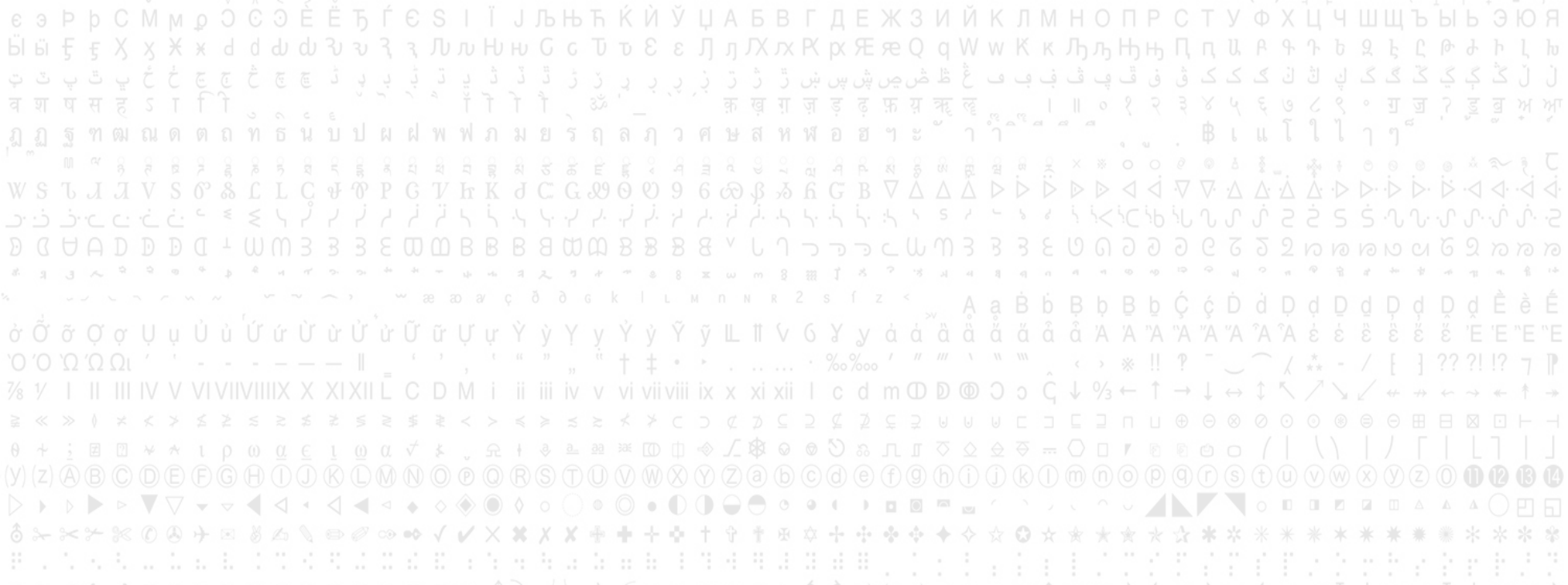
Homoglyphification: Choose Font(s)

- Basic Multilingual Plane only
- No font has full coverage of entire universe of Unicode code points
- Find something with good BMP coverage (Code2000, Unifont, Noto)
- Determine optimal glyph dimensions
- Determine desirable blocks (many can be excluded)

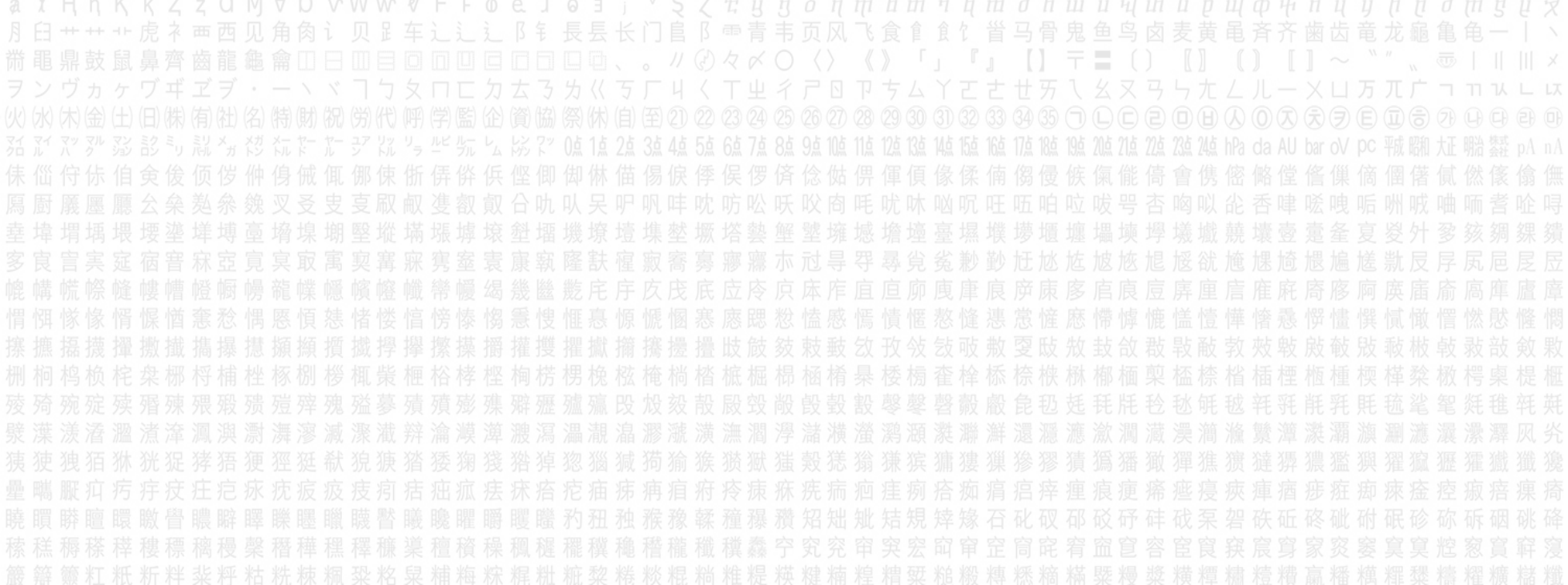
Homoglyphification: **Generate / Compare Glyphs**

- Iterate Basic Multilingual Plane blocks
- Write out glyphs
- Deduplicate (tofu)
- Compare glyphs look for homoglyphs (argh: $O(n^2)$)
- Generate htmlized view for subsequent inspection





What Can We All Do?



Mitigations: No Silver Bullet

- End-user: education and awareness
- Sysadmin: defensive registrations, defensive feeds, be aware, dnstwist, dnstwister, GFYP
- Registry/Registrar: strict(er) rules for intra-label script, commingling, mandate homograph lookup checks, “homoglyph bundling”



Bonus Stuff

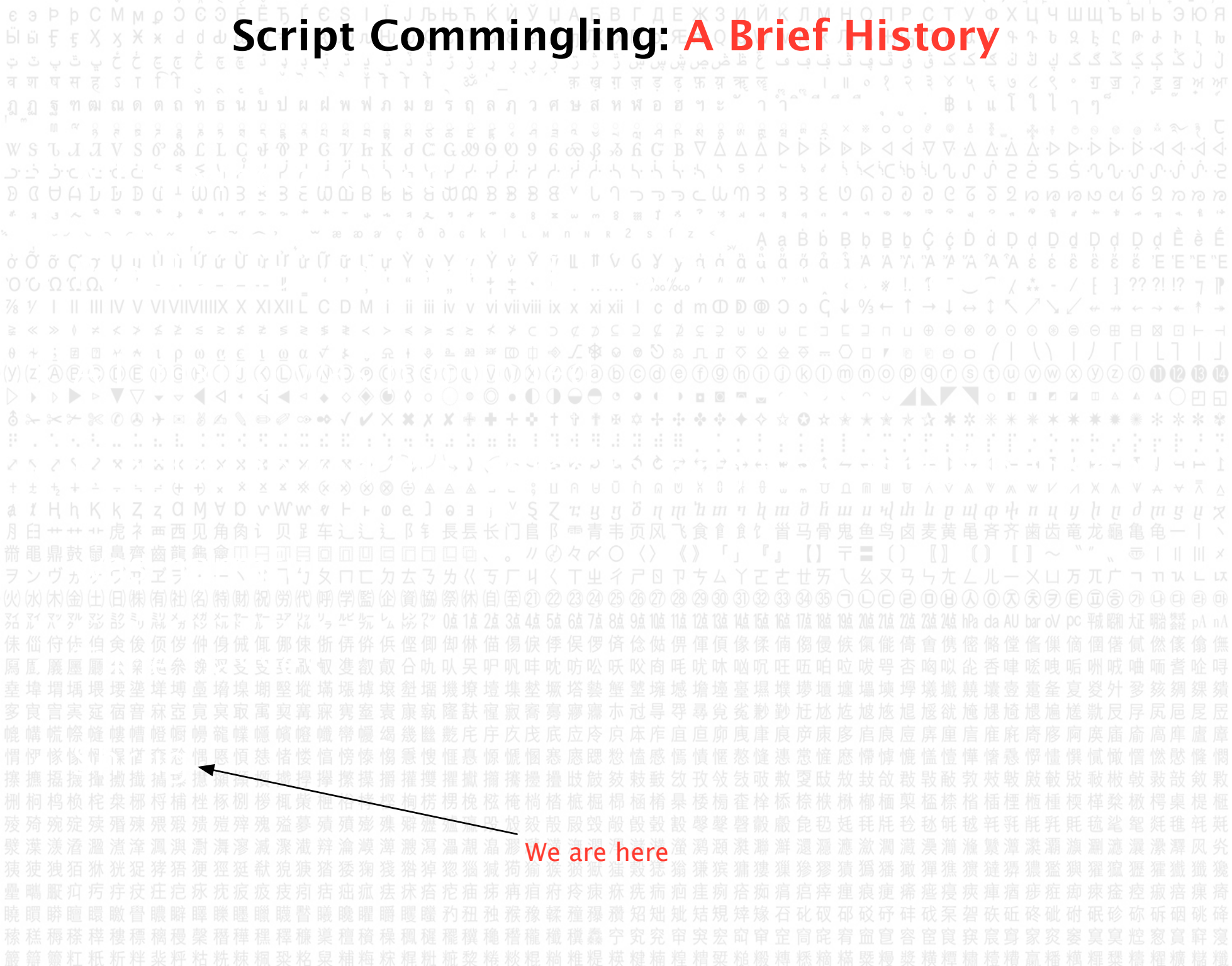
Script Commingling: It's A Problem

- The mixing of different scripts at effective second-level domain
- (Basic Latin + Cyrillic)
 - xn--pypal-4ve.com. --> pa^{U+0430}ypal.com.

U+0430

U+0061

Script Commingling: A Brief History



We are here

Script Commingling: Why Is It a Problem?

- Enables more sophisticated and difficult-to-spot homographs than any single script does (as previously shown)
- Increases the potential homographic namespace for a brand, making defensive registrations much harder

Script Commingling: Why Is It A Thing?

- “IANA Guidelines for the Implementation of Internationalized Domain Names (4.0 Final Draft)”
- Script commingling addressed in section 2.5.2 and 2.8 sub-section V
- This document coupled with several UC documents forms a foundation for detecting against-the-rules commingling
- Oh wait...

Script Commingling: It's Really Complicated

- “Exceptions to this guideline are permissible for languages with established orthographies and conventions that require the commingled use of multiple Unicode scripts”

- IDN Tables to the rescue!

- Sorta!

Script Commingling: **Our Path Forward?**

- Check TLD
- IDN Table?
- If we don't have one, fall back to the ICANN/UC rules

Errata: C Programming is Hard

- While tinkering on a PoC mixed script checker
- Found edge-case bugs in GNU libpsl, libidn, libidn2
- Libpsl: NULL pointer deref (boring)
- libidn/libidn2
 - The functions responsible for decoding Punycode into Unicode in both libidn and libidn2 can be coerced to generate invalid Unicode code point values **yet return successfully.**

Errata: Some Bugs

- The net result: all code written against libidn and libidn2 that directly calls the Punycode decoder can silently return invalid Unicode code points
- i.e.: `uint32_t all_unicode_codepoints[0x10FFFF];`
- I found this because an in-the-wild FQIDN triggered it.
- ...
- Code was patched — are you running the latest version?

Fin

Thank you very much for your time!

mschiffm@fsi.io