IDN Homographs

SSAC | ICANN63 | October 2018

Introduction

Security and Stability Advisory Committee (SSAC)

Who We Are

◉ 39 Members

◉ Appointed by the ICANN Board

What We Do

Role: Advise the ICANN community and Board on matters relating to the security and integrity of the Internet’s naming and address allocation systems.

What is Our Expertise

- Addressing and Routing
- DNS & DNSSEC
- Registry & Registrar Operations
- ISP & Network Operations
- DNS Abuse & Cybercrime
- Internationalization
- ICANN Policy and Operations

How We Advise

103 Publications since 2002

Agenda

1
Introductions

2
Internationalized Domain Names

3
Unicode and DNS Labels

4
IDN Homographs

5
Detection and Mitigation

6
Q&A

Panelists / Presenters

- Tim April
- Merike Kaeo
- Rod Rasmussen
- Suzanne Woolf

Acknowledgments

- Mike Schiffman, Farsight Security
- Sam Erb, Akamai Technologies

Internationalized Domain Names

Why IDNs?

“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

Translation → “We deal with identifiers, not words”

IDNs

- Internationalized Domain Names in Applications (IDNA2008)
- A way of representing characters other than Basic Latin in the DNS
- Internationalized Domain Names (IDNs) consist of Unicode characters

Cyrillic: правительство.рф

Korean: 스타벅스코리아.com.

:Arabic موقع.وزارة-اتصا.ت.مصر.

IDNs: Definitions

RFC6365: Terminology Used in Internationalization in the IETF

Language
A way that humans communicate

Script
A set of graphic characters used for the written form of one or more languages

Writing System
A set of rules for using one or more scripts to write a particular language

Character
The smallest unit of a writing system, the name of the encoded entity itself

Glyph
An image of a character that can be displayed

IDNs: Homoglyphs and Homographs

- **Homoglyph** One of two or more glyphs with shapes that appear identical or very similar

a ã

- **Homograph** One of two or more strings that appear identical or very similar

facebook
fãcebook

Unicode and DNS Labels

Unicode

Unicode One character set (repertoire) with as a goal to contain every written character in every language. Like other character sets, it provides a unique number for every code point, not a unique code point per character.

F	U+0046	*latin capital letter F*
A	U+0041	*latin capital letter A*
R	U+0052	*latin capital letter R*
S	U+0053	*latin capital letter S*

∞	U+221E	*infinity*
Ю	U+042E	*cyrillic capital letter yu*
П	U+041F	*cyrillic capital letter pe*
ᄫ	U+112B	*hangul choseong kapyeounpieup*

U-labels and A-labels

- The DNS can carry any value in each octet in a label
- DNS labels are interpreted as ASCII, not Unicode
- Unicode form is called the U-label
- Unicode can be encoded as Letter Digit Hash (LDH) ASCII in DNS labels
- ASCII form is called the A-label (begins with "xn--")
  - An IDN can have one, some, or all labels A-label encoded
  - There is a 1:1 mapping between A-label and U-label
  - The ASCII encoding is known as punycode

U-labels	**A-labels**
правительство.рф	xn--80aealotwbjpid2k.xn--p1ai

IDN Homographs

ASCII Look-alikes vs IDN Homographs

- ASCII Look-alike: One of two or more **ASCII** strings that appear identical or very similar
- Solutions exist for detecting some ASCII look-alikes that do not exist for IDN Homographs

acme.example

acrne.example

IDN Homographic Attacks

- Humans are really good at pattern recognition
- Many glyphs originating from the Unicode repertoire look similar or even identical to others depending on the font
- So… register an IDN that is a homograph of a well-known (usually non-internationalized) domain name
- Extort, camp, cash-park, phish, distribute malware, or do other antisocial things by using the IDN in a URL
- ???
- Profit

Examples

Real Site	Homograph	A-label
easyjet.com.	easyjeṭ.com.	xn--easyje-n17b.com.
delta.com.	deǀta.com.	xn--deta-1kb.com.
ryanair.com.	ryanaiṛ.com.	xn--ryanai-1x7b.com.
poloniex.com.	polonìex.com.	xn--polonex-3ya.com.
bittrex.com.	bītťrex.com.	xn--btrex-m3a12b.com.
linkedin.com.	lìnkedin.com.	xn--lnkedin-zya.com.

Courtesy of
Mike Schiffman,
Farsight Security

Observed via Passive DNS

ƒacebook.com.ƒacebọọk.com.
ƒacebook.tk. ƒɑcebook.com.
ḟacebook.com. fácebook.com.
fàcebook.com. fâcebook.com.
fåcebook.com. fäcebook.com.
fãcebook.com. fȧcebook.com.
fącebook.com. fācebook.com.
fạcebook.com. fḁcebook.com.
fàćebook.com. faĉebook.com.
fačebook.com. faċebook.com.
façebook.com. faƈebook.com.
facébook.com.

apple.com. appɩė.com.
âpplê.cf. ápple.com.
ăpple.com. åpple.com.
äpple.com. ąpple.com.
apṗle.com. appĺe.com.
applé.com. applè.com.
àpplè.com. applĕ.com.
ăpplĕ.com. ápplê.com.
àpplê.com. âpplê.com.
applĕ.com. applë.com.
äpplë.com. applė.com.
åpplė.com.

ñetflix.com.
ņetflix.com.
nétflix.com.
nėtflix.com.
neṭflix.com.
netflíx.com.
netflìx.com.
netflîx.com.
netflïx.com.
netflịx.com.
netflıx.com.
netƒlix.com.

ġoogle.xyz. goôgle.com.
ĝoogle.com. googĺe.com.
gọọglē.com. googlè.tk.
googlę.com. googlè.com.
googlé.com. ġooģle.com.
goơgle.com. googlè.com.
googlé.com. goơgle.com.
gooĝle.com. gooĝle.com.
gooĝle.com. gooĝle.com.
gooĝle.com. gooĝle.com.
gooĝle.com. gooĝle.com.
gooģle.com.

ɓankofamerica.com. baŋkofamerica.com.
baŋkofamerica.net. bąnkofamerica.com.
bankôfamerica.com. banköfamerica.com.
bankofamerîca.com. bänkofämericä.com.
banĸofamerica.com. banĸofamerica.net.
bɑnkofɑmericɑ.com.

wėllsfargo.com.
welłsfargo.com.
wellsfárgo.com.
wellsfårgo.com.
wellsfargó.com.
wellsfargọ.com.
wellsfɑrgo.com.

çhase.com.
chàse.com.
chäse.com.
chasé.com.
chasë.com.
chɑse.com.
chase.com.

Courtesy of
Mike Schiffman,
Farsight Security

Observed in the Wild

- 1,936 impersonation domains observed in a review of Certificate Transparency logs (2017) [1]
- Farsight January 2018 research [2]
  - Examined 125 brand names
  - In a 3 month period observed 116,113 homographs
  - Discovered 10+ live phishing sites
  - 382 impersonation domains reported from Passive DNS logs

[1]: https://github.com/CyberMonitor/defcon-25-Packet-Hacking-Village/blob/master/YOU'RE%20GOING%20TO%20CONNECT%20TO%20THE%20WRONG%20DOMAIN%20NAME%20phv2017-serb.pdf
[2]: https://www.farsightsecurity.com/2018/01/17/mschiffm-touched_by_an_idn/

Observed in the Wild (cont)

- Farsight October 2018 research [3]
  - Examined 509 brand names
  - In a 20 month period observed 11,766 unique IDN homographs
  - In same period observed 61,443 total IDNs
    - 20% in banking/finance
    - 52% in .com
    - 68% geolocate to the USA
    - 93% using IPv4

Data courtesy of Mike Schiffman, Farsight Security

What We've Seen: All The IDNs

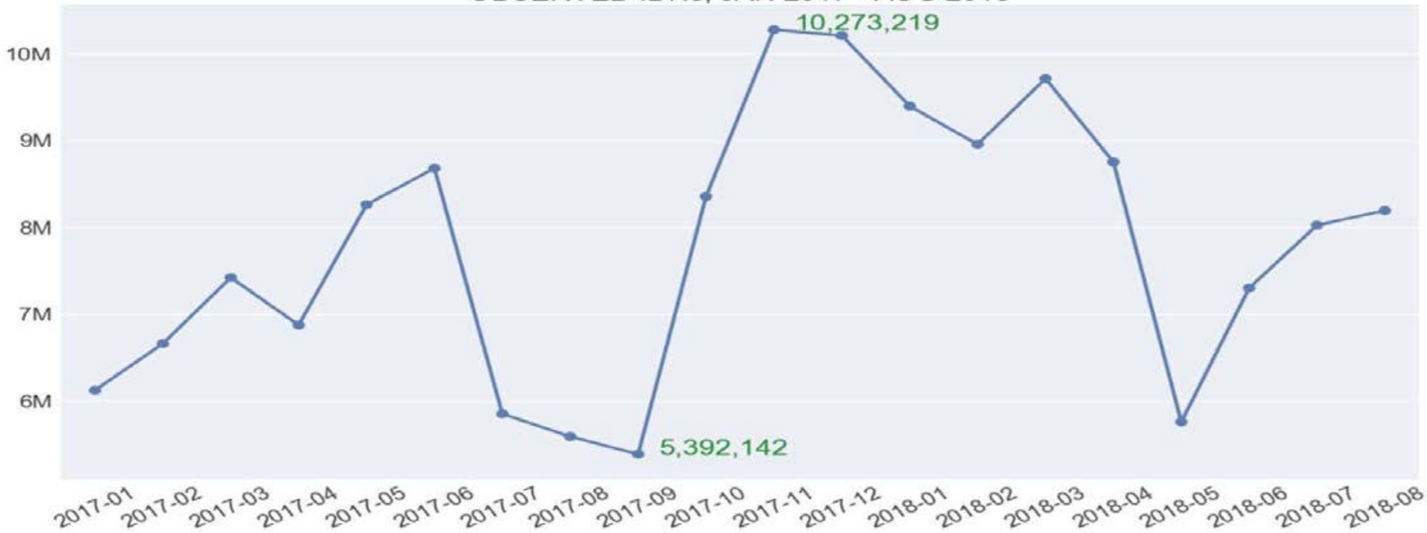
161,935,465 total IDN observations

34,460,574 total unique IDNs

Summer vacation?

Diagram courtesy of Mike Schiffman, Farsight Security

OBSERVED IDNs, JAN 2017 - AUG 2018

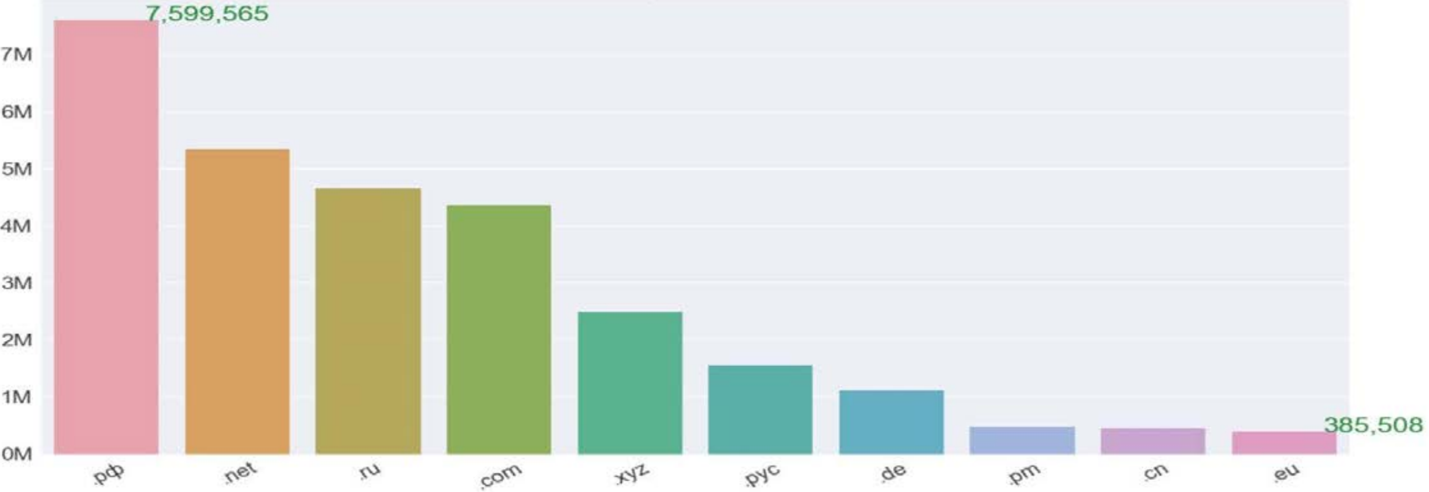


What We've Seen: Top 10 IDN TLDs

1,675 total unique TLDs

Diagram courtesy of
Mike Schiffman,
Farsight Security

TOP TEN TLDs, JAN 2017 - AUG 2018



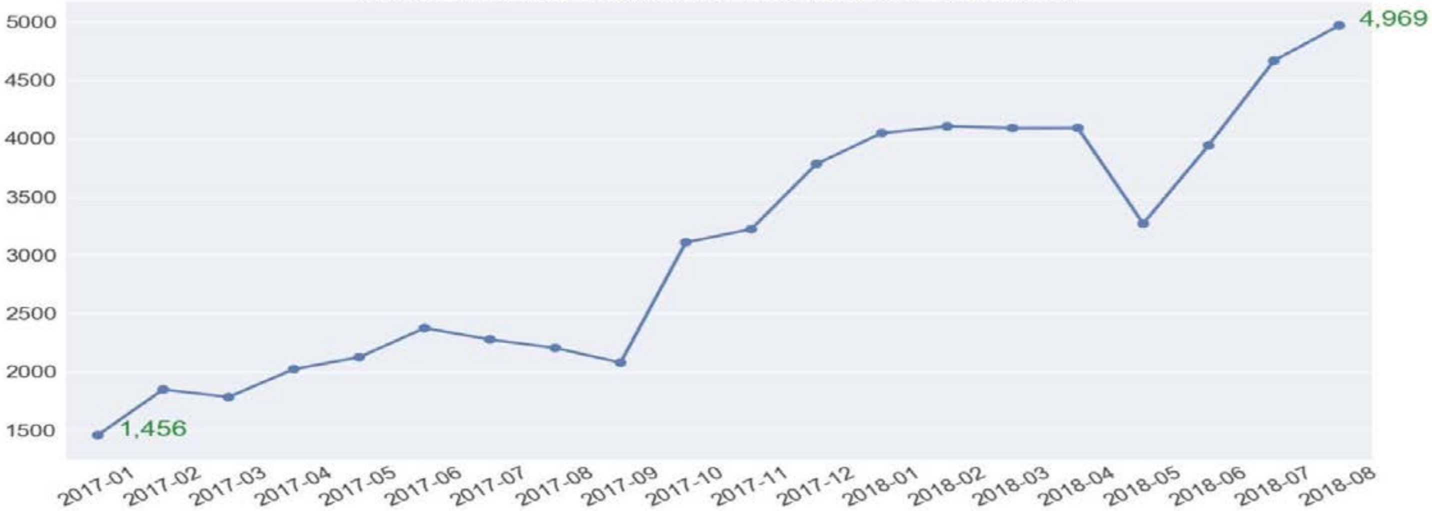
What We've Seen: IDN Homographs

61,443 total IDN homograph observations

11,766 total unique IDN homographs

Diagram courtesy of
Mike Schiffman,
Farsight Security

OBSERVED IDN HOMOGRAPHS, JAN 2017 - AUG 2018

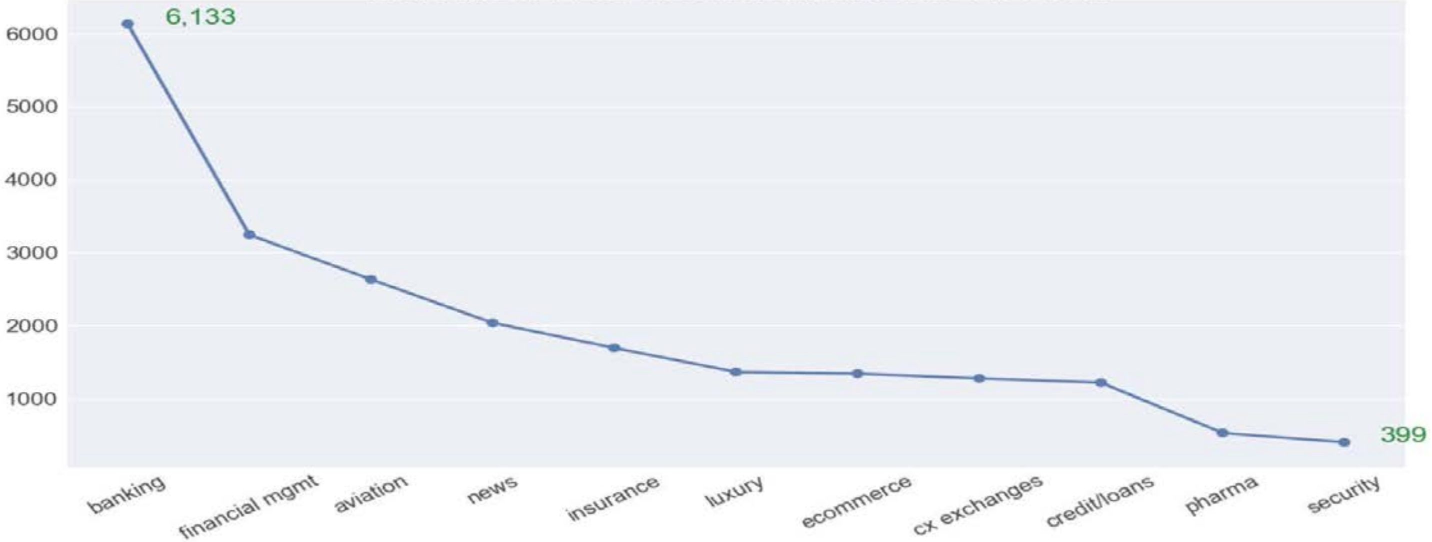


What We've Seen: IDN Homographs by Sector

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

Diagram courtesy of
Mike Schiffman,
Farsight Security

IDN HOMOGRAPHS BY SECTOR, JAN 2017 - AUG 2018



Detection and Mitigation

How to Detect Attacks

- Monitor certificate transparency logs
- Monitor DNS zone files
- Utilize passive DNS services
- Detecting IDN homographs reliably typically requires human eyes

Mitigation

- Stricter rules at registry and registrar
    - Registries and registrars implement recommendations from IDNA2008 (RFC 5890-5894, specifically RFC5894)
    - Use an inclusion based process before allowing code points
        - For example, base rules on what script a code point belongs to
    - Be extremely conservative with mixed scripts within a label, and within a domain name
    - Adapt the Label Generation Rules (LGRs)
    - Mandate homographic lookup checks
- Browsers often implement homograph preventions, but with limited success

Why is this Important?

- ICANN's mission of **Security**, Stability and Resiliency of the global unique identifiers
  - Phishing, malware, malicious email
- Affects universal acceptance
  - Failure to act may result in ad-hoc blocking or other display tricks
- Business Email Compromise (BEC) is a growing problem
  - Failure to act may result in blocking of emails that use IDNs

What Can the Community do to Help?

- Opportunity for development of tools to detect IDN Homographs
  - Visualization
  - Comparison to known homographic targets
  - Facilitate brand protection
- Awareness and outreach of the potential malicious use of IDN Homographs
  - End-user awareness
  - Implementor education
  - Service provider awareness

Relevant SSAC Publications

Relevant SSAC Publications

- https://www.icann.org/groups/ssac/documents

- SAC037: Display and usage of Internationalized Registration Data Support for characters from local languages or scripts
- SAC052: SSAC Advisory on Delegation of Single-Character IDN TLDs
- SAC084: SSAC Comments on Guidelines for the Extended Process Similarity Review Panel for the IDN ccTLD Fast Track Process
- SAC088: SSAC Response to the ccNSO evaluation of SAC084
- SAC099: SSAC Response to the ICANN Internationalized Domain Name Guidelines Working Group

Panel Discussion / Q&A

Thank you