ICANN61 – Tech Day
IDN Abuse

Merike Kaeo (presenting)

Research by:
Mike Schiffman, Stephen Watt

FARSIGHT SECURITY

Motivation

- Lots of Data To Play With
- Shed Light on Domain Abuse via IDN Homographs
  - IDNs allow forgeries to be nearly undetectable by either human eyes or human judgment
  - Is it well understood by the wider public?
- How Bad Is The Problem
  - Registering Internet DNS names for the purpose of misleading consumers is not news
  - Wanted to determine prevalence and reach of issue

Terminology

FARSIGHT SECURITY

Terms to know when dealing with IDNs

- Code point: A numerical value representing a Unicode character i.e.: `U+03B1`
- Plane: A contiguous set of code points (17 in total; plane 0, *The Basic Multilingual Plane* is the most important)
- Block: Logical subdivision of a plane; “Basic Latin” (ASCII `0x–0x7f`), or CJK Unified Ideographs
- UTF-8: Common scheme for variable length encoding of Unicode code points into sequences of 1 – 4 bytes (`U+0000—U+10FFFF`); is backwards compatible with ASCII
- SSIM: Structured Similarity Index; a fractional value representing the similarity between two images that can range from 0.0 (least similar) to 1.0 (identical)
- Homoglyph: One of two or more characters with shapes that appear identical or very similar (O ”oh” and 0 “zero”)
- Homograph: Same as above, but entire words are considered

Unicode

Universal Encoding

- Unicode is a universal standard for encoding language glyphs
- It provides a unique number for every character (this is a code point)
- Latest version contains 136,755 characters covering 139 modern and historic scripts

Example Unicode characters

F:	U+0046	I:	U+0049	✪:	U+272A
A:	U+0041	G:	U+0047	∰:	U+2230
R:	U+0052	H:	U+0048	ॐ:	U+0950
S:	U+0053	T:	U+0054	♥:	U+2665

Punycode

A lossless method for down sampling Unicode into ASCII

- 'Taking data that requires larger encoding space and fitting it into a smaller presentation format (“puny”)
- Punycode is an encoding to convert Unicode characters into ASCII
- Technically, into a subset of ASCII known as LDH (letters, digits, hyphens)

Example Unicode --> Punycode

αβγδεζηθικλμνξοπρστυφχψω --> xn--mxacdefghijklmnopqr0btuvwxy

IDNs represent Unicode labels and may appear as such to the end user, but over the wire they are sent encoded using Punycode

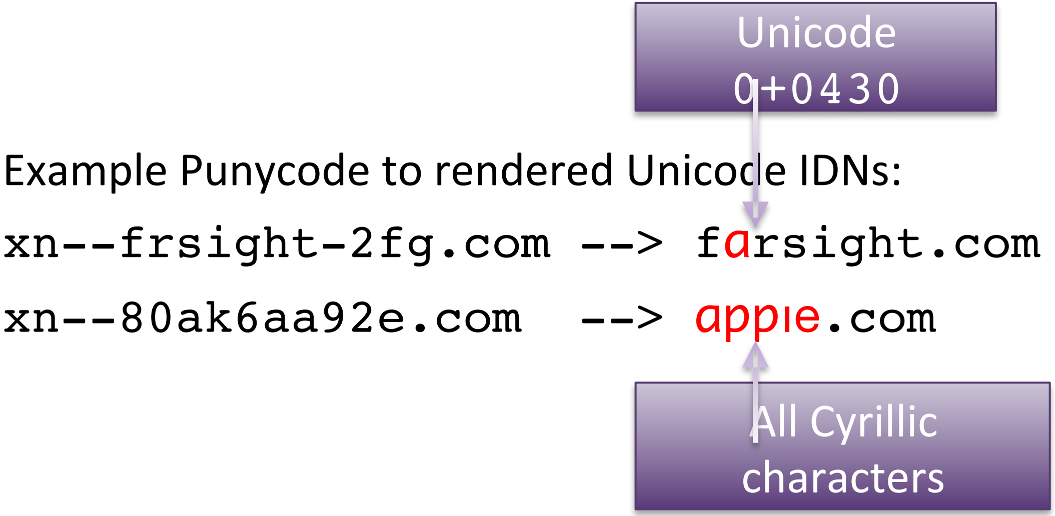
IDN Homographs

- Different letters or characters might look alike
  - Uppercase “I” and lowercase “l”
  - Letter “O” and number “0”
- Characters from different alphabets or scripts may appear indistinguishable form one another to the human eye
  - Individually they are known as *homoglyphs*
  - In the context of the words that contain them they constitute *homographs*

IDN Homograph Attacks

And this is why we can’t have nice things

- Bad actors figured out they can register IDNs and target sites using homoglyphs (or sometimes homographs)



Unicode 0+0430

Example Punycode to rendered Unicode IDNs:

```
xn--frsight-2fg.com --> fаrsight.com
xn--80ak6aa92e.com  --> аррӏе.com
```

All Cyrillic characters

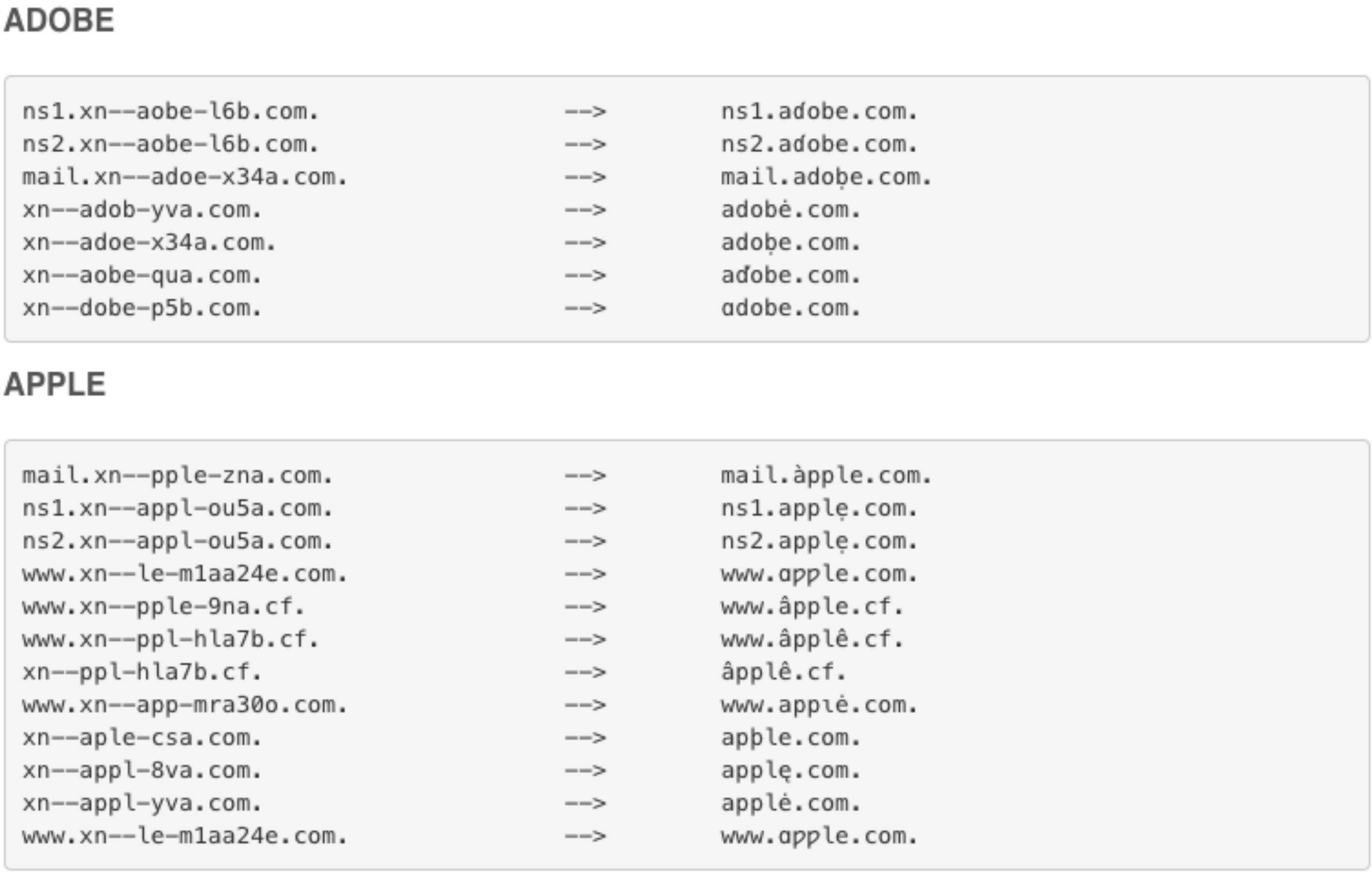
Research Done

- Examined 125 top brand domain names
  - Large content providers, social networking companies, financial websites, luxury brands, cryptocurrency exchanges, etc.
- Monitoring IDN homographs in real-time
- From 3 month observation period observed 116,113 homographs
  - 2017-10-17 23:41 UTC to 2018-01-10 19:00 UTC

Disturbing Findings

- Indepth details:
  - https://www.farsightsecurity.com/2018/01/17/mschiffm-touched_by_an_idn/
- The large number of homographs seems disturbing and may need further investigations
  - No assumption made of intent against domains or domain owners
- However, did find some live phishing sites
  - Companies were contacted to alert them of suspected phishing sites
  - Demonstrates that threat of IDN homograph impersonation is both real and actively being exploited

Suspicious IDNs



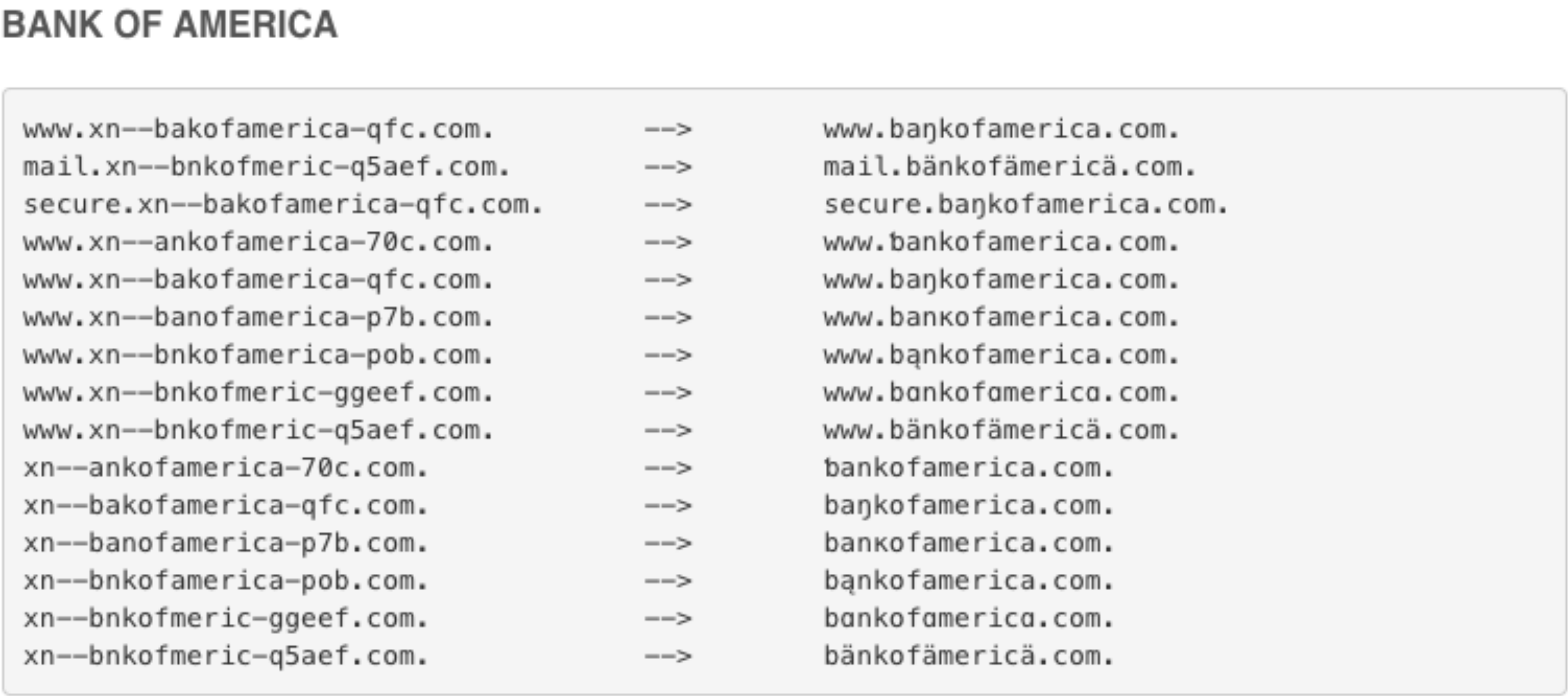
ADOBE

```
ns1.xn--aobe-l6b.com.          -->        ns1.aɗobe.com.
ns2.xn--aobe-l6b.com.          -->        ns2.aɗobe.com.
mail.xn--adoe-x34a.com.        -->        mail.adoḅe.com.
xn--adob-yva.com.              -->        adobė.com.
xn--adoe-x34a.com.             -->        adoḅe.com.
xn--aobe-qua.com.              -->        aďobe.com.
xn--dobe-p5b.com.              -->        ɑdobe.com.
```

APPLE

```
mail.xn--pple-zna.com.         -->        mail.àpple.com.
ns1.xn--appl-ou5a.com.         -->        ns1.applẹ.com.
ns2.xn--appl-ou5a.com.         -->        ns2.applẹ.com.
www.xn--le-m1aa24e.com.        -->        www.ɑρρle.com.
www.xn--pple-9na.cf.           -->        www.âpple.cf.
www.xn--ppl-hla7b.cf.          -->        www.âpplê.cf.
xn--ppl-hla7b.cf.              -->        âpplê.cf.
www.xn--app-mra30o.com.        -->        www.appɩė.com.
xn--aple-csa.com.              -->        apþle.com.
xn--appl-8va.com.              -->        applę.com.
xn--appl-yva.com.              -->        applė.com.
www.xn--le-m1aa24e.com.        -->        www.ɑρρle.com.
```

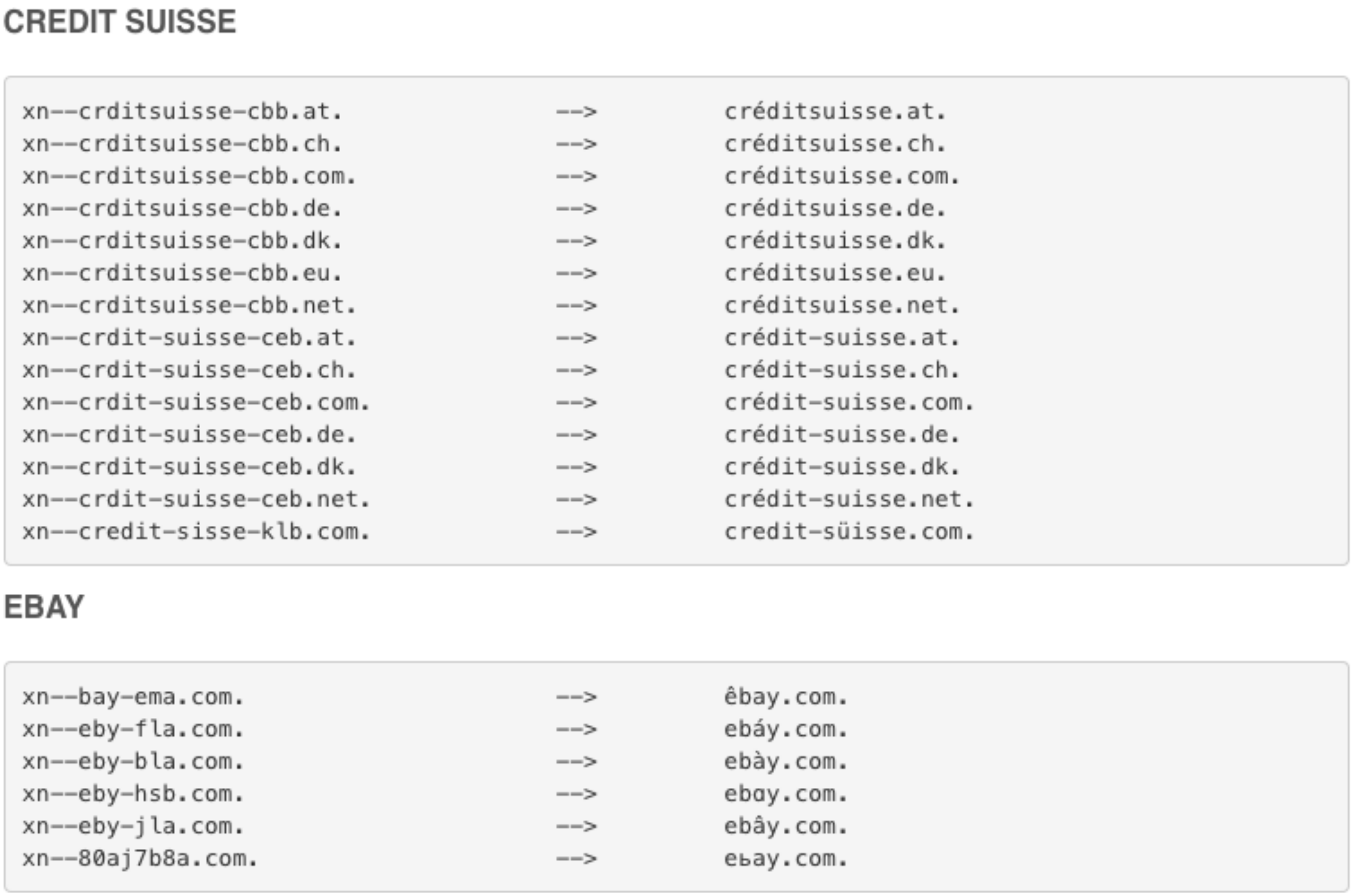
Suspicious IDNs



BANK OF AMERICA

```
www.xn--bakofamerica-qfc.com.         -->        www.baŋkofamerica.com.
mail.xn--bnkofmeric-q5aef.com.        -->        mail.bänkofämericä.com.
secure.xn--bakofamerica-qfc.com.      -->        secure.baŋkofamerica.com.
www.xn--ankofamerica-70c.com.         -->        www.ƅankofamerica.com.
www.xn--bakofamerica-qfc.com.         -->        www.baŋkofamerica.com.
www.xn--banofamerica-p7b.com.         -->        www.banĸofamerica.com.
www.xn--bnkofamerica-pob.com.         -->        www.bąnkofamerica.com.
www.xn--bnkofmeric-ggeef.com.         -->        www.bɑnkofɑmericɑ.com.
www.xn--bnkofmeric-q5aef.com.         -->        www.bänkofämericä.com.
xn--ankofamerica-70c.com.             -->        ƅankofamerica.com.
xn--bakofamerica-qfc.com.             -->        baŋkofamerica.com.
xn--banofamerica-p7b.com.             -->        banĸofamerica.com.
xn--bnkofamerica-pob.com.             -->        bąnkofamerica.com.
xn--bnkofmeric-ggeef.com.             -->        bɑnkofɑmericɑ.com.
xn--bnkofmeric-q5aef.com.             -->        bänkofämericä.com.
```

Suspicious IDNs



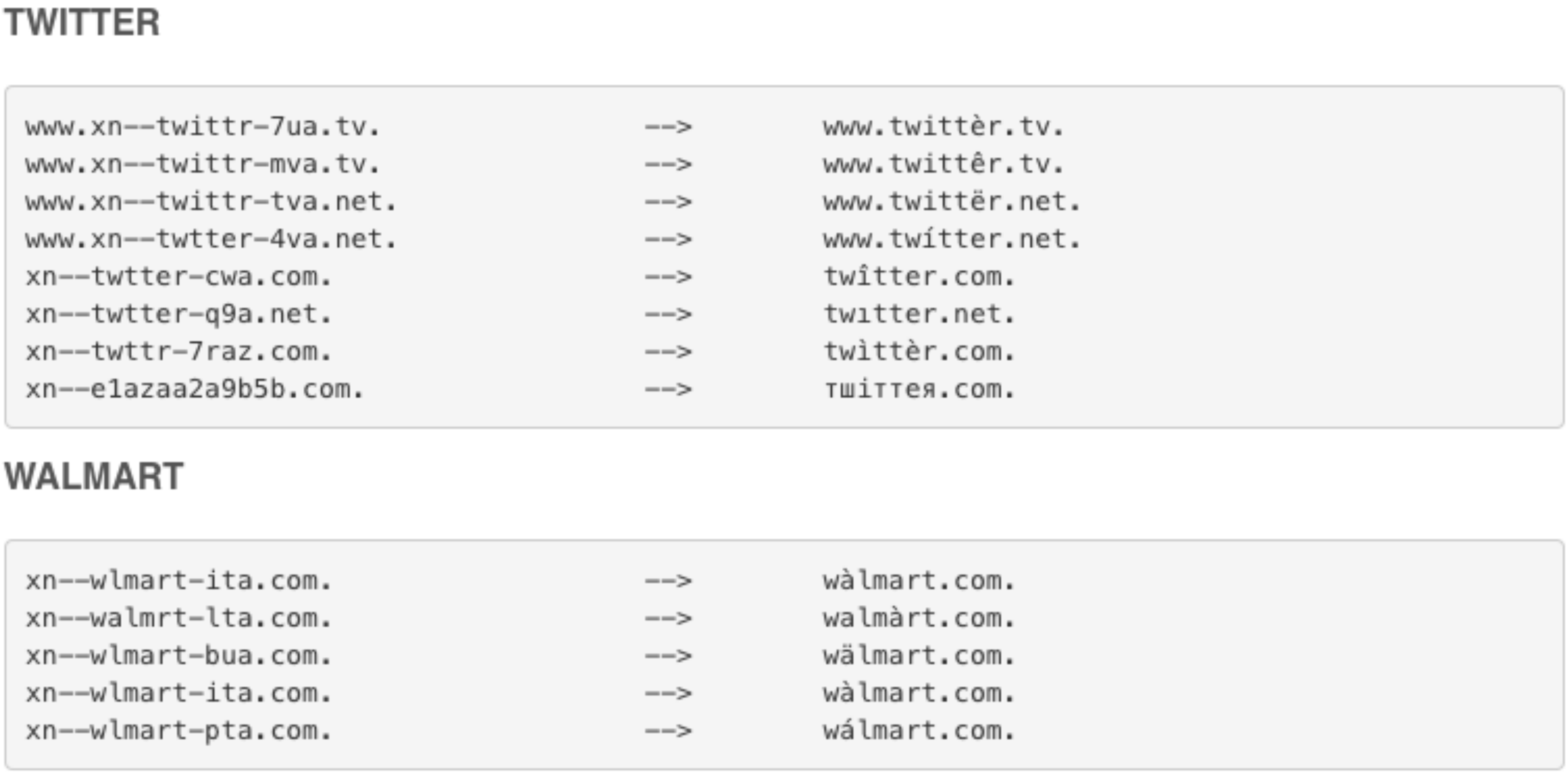
CREDIT SUISSE

```
xn--crditsuisse-cbb.at.        -->        créditsuisse.at.
xn--crditsuisse-cbb.ch.        -->        créditsuisse.ch.
xn--crditsuisse-cbb.com.       -->        créditsuisse.com.
xn--crditsuisse-cbb.de.        -->        créditsuisse.de.
xn--crditsuisse-cbb.dk.        -->        créditsuisse.dk.
xn--crditsuisse-cbb.eu.        -->        créditsuisse.eu.
xn--crditsuisse-cbb.net.       -->        créditsuisse.net.
xn--crdit-suisse-ceb.at.       -->        crédit-suisse.at.
xn--crdit-suisse-ceb.ch.       -->        crédit-suisse.ch.
xn--crdit-suisse-ceb.com.      -->        crédit-suisse.com.
xn--crdit-suisse-ceb.de.       -->        crédit-suisse.de.
xn--crdit-suisse-ceb.dk.       -->        crédit-suisse.dk.
xn--crdit-suisse-ceb.net.      -->        crédit-suisse.net.
xn--credit-sisse-klb.com.      -->        credit-süisse.com.
```

EBAY

```
xn--bay-ema.com.               -->        êbay.com.
xn--eby-fla.com.               -->        ebáy.com.
xn--eby-bla.com.               -->        ebày.com.
xn--eby-hsb.com.               -->        ebɑy.com.
xn--eby-jla.com.               -->        ebây.com.
xn--80aj7b8a.com.              -->        eьay.com.
```

Suspicious IDNs



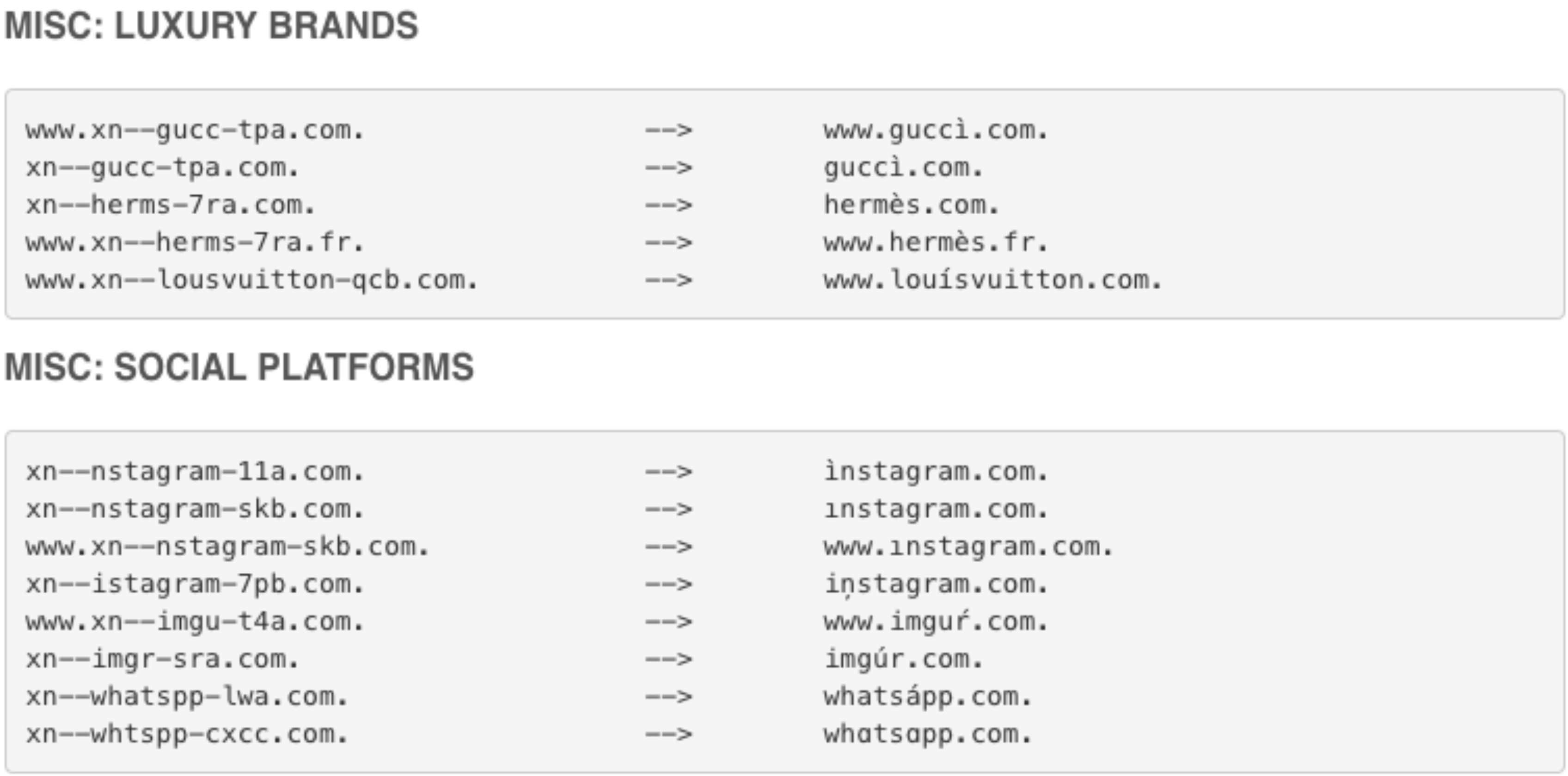
TWITTER

```
www.xn--twittr-7ua.tv.          -->       www.twittèr.tv.
www.xn--twittr-mva.tv.          -->       www.twittêr.tv.
www.xn--twittr-tva.net.         -->       www.twittër.net.
www.xn--twtter-4va.net.         -->       www.twítter.net.
xn--twtter-cwa.com.             -->       twîtter.com.
xn--twtter-q9a.net.             -->       twıtter.net.
xn--twttr-7raz.com.             -->       twìttèr.com.
xn--e1azaa2a9b5b.com.           -->       тшіттея.com.
```

WALMART

```
xn--wlmart-ita.com.             -->       wàlmart.com.
xn--walmrt-lta.com.             -->       walmàrt.com.
xn--wlmart-bua.com.             -->       wälmart.com.
xn--wlmart-ita.com.             -->       wàlmart.com.
xn--wlmart-pta.com.             -->       wálmart.com.
```

Suspicious IDNs



MISC: LUXURY BRANDS

```
www.xn--gucc-tpa.com.                -->        www.guccì.com.
xn--gucc-tpa.com.                    -->        guccì.com.
xn--herms-7ra.com.                   -->        hermès.com.
www.xn--herms-7ra.fr.                -->        www.hermès.fr.
www.xn--lousvuitton-qcb.com.         -->        www.louísvuitton.com.
```

MISC: SOCIAL PLATFORMS

```
xn--nstagram-11a.com.                -->        ìnstagram.com.
xn--nstagram-skb.com.                -->        ınstagram.com.
www.xn--nstagram-skb.com.            -->        www.ınstagram.com.
xn--istagram-7pb.com.                -->        iņstagram.com.
www.xn--imgu-t4a.com.                -->        www.imguŕ.com.
xn--imgr-sra.com.                    -->        imgúr.com.
xn--whatspp-lwa.com.                 -->        whatsápp.com.
xn--whtspp-cxcc.com.                 -->        whɑtsɑpp.com.
```

General Observations

- While IDN related abuse domains are a fraction of the overall abuse domains, they do exist
- Publicity surrounding this kind of abuse is growing which will motivate potentially more abuse
- What is role of IETF (who decides what characters can be used in an IDN) vs role of ICANN (who decides policy) ?
- Would certain policy enforcements mitigate most of the potentially harmful IDN related abuse domains ?

QUESTIONS ?