Funny Accents: Exploring Genuine Interest in Internationalized Domain Names

Victor Le Pochat, Tom Van Goethem, Wouter Joosen

PAM 2019, 29 March 2019

KU LEUVEN DistriNet

What do these brands have in common?

What do these brands have in common?

Internationalized Domain Names (IDNs) allow Unicode characters in domain names

User agent		DNS
google.com		google.com
köln.de	Punycode	xn--kln-sna.de
яндекс.рф		xn--d1acpjx3f.xn--p1ai

Strategic Objective on Unique Identifier System: **3. Evolve the unique identifier systems to continue to serve the needs of the global Internet user base.**

Strategic Goals

3.1 Encourage readiness for Universal Acceptance, IDN implementation, and IPv6 by increasing awareness to enable more end users to use the Internet.

3.2 Improve understanding of and responsiveness to new technologies by greater engagement with industry, academia, standards development organizations, and other relevant parties.

3.3 Continue to deliver and enhance the IANA functions with operational excellence.

3.4 Plan a properly funded, managed, and risk-evaluated new round of gTLDs.

26 comments received:

- 2 support / 13 edits / 11 concerns & other suggestions

Topics where we would welcome added community input:

- Evolution of new identifier systems to ensure universal resolution
- Support the adoption of IDNs
- Worldwide deployment of IPv6

IDNs can be abused due to visual similarity

www.google.com	≠	www.goog**l**e.com
www.google.com	≠	www.go**ọ**gle.com
www.google.com	≠	www.go**o**gle.com
www.nestle.com	?	www.nestl**é**.com

[Hol06, Liu18]

Brands may want to use IDNs with *genuine interest* ...

- **corresponds** to brand
- **easier** to read and understand

... but malicious actors might want to do so too

- **corresponds** to brand
- **easier** to read and understand

- more **difficult to distinguish** legitimate site from **phishing**
- **abuse typed** domain with accents

Generating candidate domains

Ownership, use and abuse

User agent behavior

Generating candidate domains

Ownership, use and abuse

User agent behavior

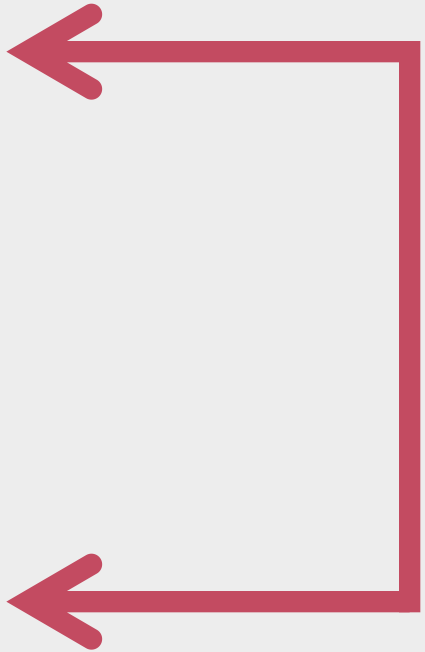
nestle.com	*Original domain*
Home | Nestlé Global	*Root page title*
home nestlé global	*Convert to lowercase, remove punctuation*
home nestle global	*Remove accents*
(köln → koeln)	*(Apply substitutions)*

nestle.com

Home | Nestlé Global

home nestlé global

home **nestle** global



nestle.com

Home | Nestlé Global

home **nestlé** global

home **nestle** global

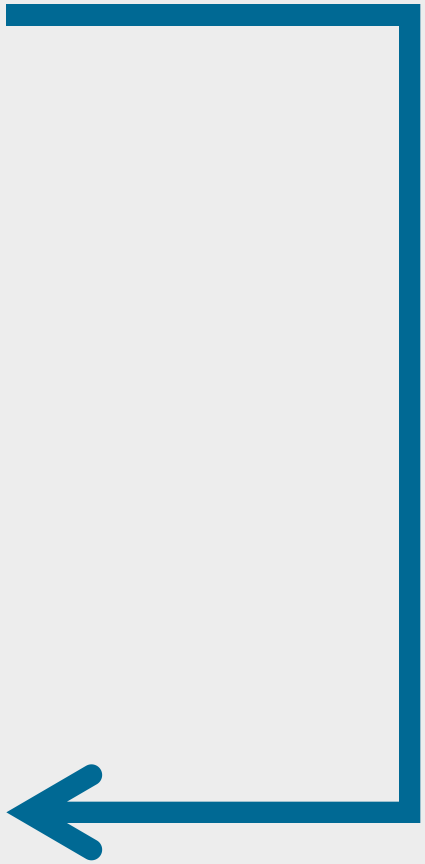
nestle.com

Home | Nestlé Global

home **nestlé** global

home **nestle** global

nestlé.com



Generating candidate domains

Ownership, use and abuse

User agent behavior

Have these IDNs already been registered?

Tranco
1 000 000
↓

15 276
candidates

12 087 (79.1%)
unregistered

1 363 (11.3%)
non-compliant with TLD policy

4 116 (34.1%)
unavailable/ additional restrictions

6 608 (54.7%)
readily available

3 189 (20.9%)
registered

[LeP19]

Who owns the registered IDNs?

59.1%

(likely) same

34.6%

different

How are the registered IDNs being used?

41.6%
same content

23.5%
parked/for sale

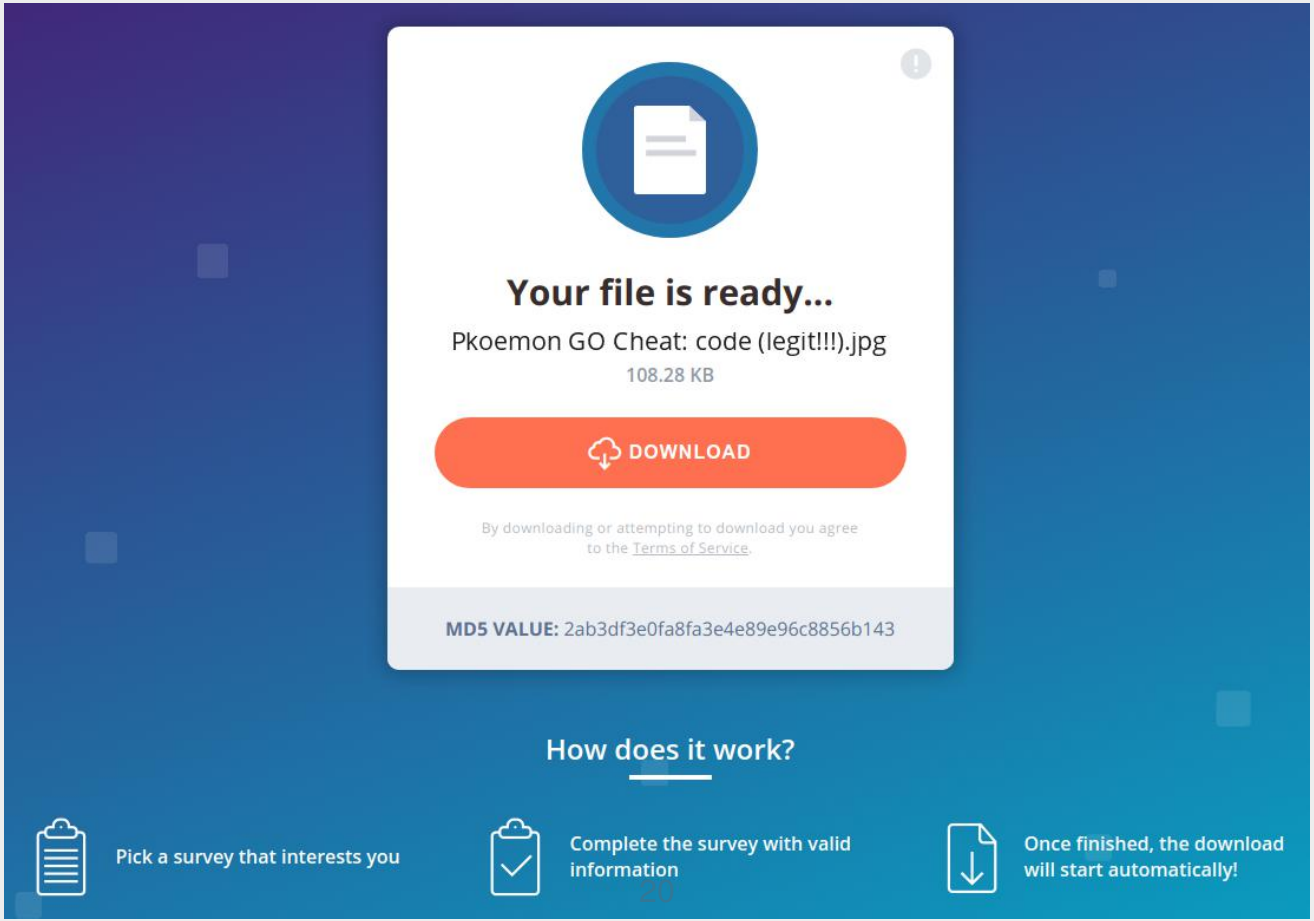
26.8%
'forgotten'

Are the registered IDNs being abused?

- No **known** malicious activity (blacklists)
  - Phishing domains can **evade blacklisting**
  - Parked domains **only sometimes redirect** to malicious content
- Some **questionable behavior**

[Vis15, Tia18]

pokémongo.com



Generating candidate domains

Ownership, use and abuse

User agent behavior

Browsers display IDNs differently (even on popularity)

Unicode	Unicode unless popular	Punycode
 	 	  
pokémon.com		xn--pokmon-dva.com

Email clients: similar inconsistencies, even within vendors

IDNA standard revision introduced “deviations”

straße.de

IDNA2003 strasse.de

```
A    89.31.143.1
```

≠

IDNA2008 xn--strae-oqa.de

```
A  81.169.145.78
```

IDNA standard revision introduced “deviations”

straße.de

IDNA2003 strasse.de

A 89.31.143.1

≠

IDNA2008 xn--strae-oqa.de

A 81.169.145.78

iOS Mail before 12.1.1 was vulnerable to phishing

Awesome Email Client

From: victor@straße.de

Subject: Test of IDN support by Victor

Hello
This is a test for IDN support by email

```
From: <victor@xn--strae-oqa.de>
Date: Tue, 2 Oct 2018 14:22:27 +0200
Subject: Test of IDN support by Victor
```

[CVE-2018-4429]

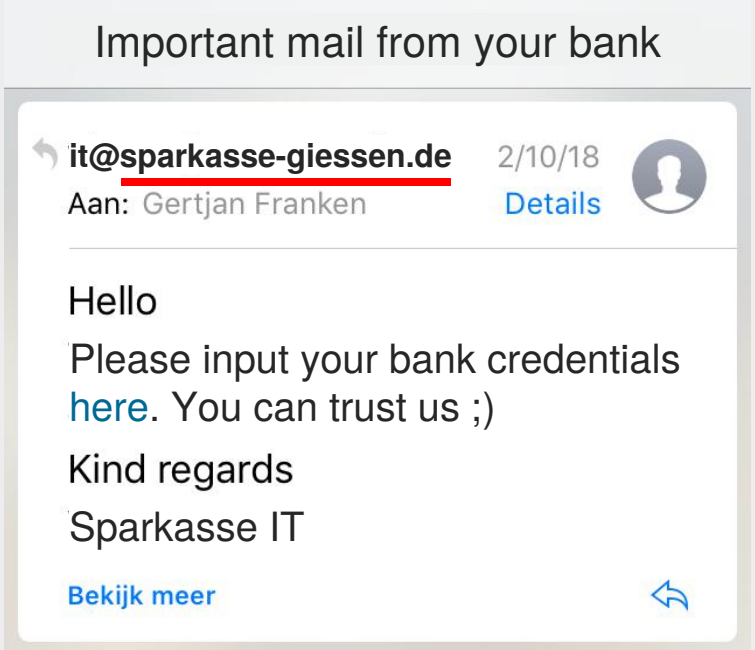
iOS Mail before 12.1.1 was vulnerable to phishing

Awesome Email Client

From: it@sparkasse-gießen.de

Subject: Important mail from your bank

Hello
Please input your bank credentials here.




```
From: <it@xn--sparkasse-gieen-2ib.de>
Date: Tue, 2 Oct 2018 14:22:27 +0200
Subject: Important mail from your bank
```

Shortcomings of key actors limit IDN uptake

- Registries: guidelines to **prohibit** or **limit** registrations of IDNs
  but **not widely** implemented

- Brand owners: some **own** their 'genuine interest' IDNs
  but they sometimes **'forget'** them
  and many also leave them to **squatters**

- User agents: primary point of **interaction** with IDNs for **users**
  but **inconsistent** support

Datasets: https://osf.io/s96dg/

Thank you!

Victor.LePochat@cs.kuleuven.be

References

1. [Hol06] Holgers, T., Watson, D.E., Gribble, S.D.: Cutting through the confusion: a measurement study of homograph attacks. In: USENIX Annual Technical Conference, pp. 261–266. USENIX Association (2006)
2. [Liu18] Liu, B., et al.: A reexamination of internationalized domain names: the good, the bad and the ugly. In: 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, pp. 654–665 (2018). https://doi.org/10.1109/DSN.2018.00072
3. [LeP19] Le Pochat, V., Van Goethem, T., Tajalizadehkhoob, S., Korczyński, M., Joosen, W.: Tranco: a research-oriented top sites ranking hardened against manipulation. In: 26th Annual Network and Distributed System Security Symposium, February 2019. https://doi.org/10.14722/ndss.2019.23386
4. [Vis15] Vissers, T., Joosen, W., Nikiforakis, N.: Parking sensors: analyzing and detecting parked domains. In: 22nd Annual Network and Distributed System Security Symposium. Internet Society (2015)
5. [Tia18] Tian, K., Jan, S.T.K., Hu, H., Yao, D., Wang, G.: Needle in a haystack: tracking down elite phishing domains in the wild. In: Internet Measurement Conference, pp. 429–442. ACM (2018). https://doi.org/10.1145/3278532.3278569