IDN Visual Security Deep Thinking

xisigr

Feb,2019

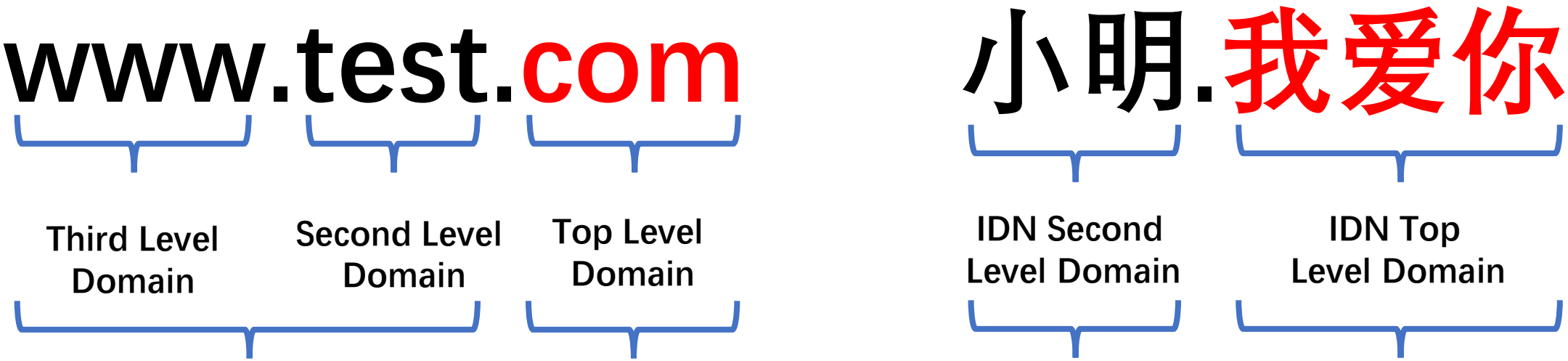
About me

- Security researcher of Tencent's Xuanwu Lab
  - https://xlab.tencent.com
- Author of Web Front-End Hacker's Handbook
  - https://www.web2hack.org/
- xisigr@Xeye team
- twitter.com/xisigr ,weibo.com/xisigr

Internationalized Domain Names (IDNs)

- IDNs
  - Long time ago, domains could only consist of the Latin letters A to Z, digits, and a few other characters from the US-ASCII coded character set.
  - IDNs allow characters from the Universal Character Set Unicode since 2003.
  - In 2018, Unicode 11.0 contains a repertoire of 137374 characters covering 146 modern and historic scripts.
  - In these more than 10 years, IDNs visual confusion security issues has never been interrupted.
  - What do IDNs Mean to you?

ASCII Domain names vs. Internationalized Domain Names (IDNs)



ASCII
Letters [a-z]
Digits [0-9]
Hyphen [-]
Label length = 63

ASCII
Letters [a-z]
Label length = 63

Unicode
Valid Unicode-Label: IDNA2008
Valid ASCII-Label: Punycode, "xn--"

Unicode
Valid Uicode-Label
Valid ASCII-label

Homoglyph Attack blowout

- 0 and O; 1, l and I

qwerty QWERTY I2345
uiopas UIOPAS 6789-
dfghjk DFGHJK "/@£_
lzxcvb LZXCVB &'()+
nm NM ¾¼⅔½
½%=$!
;:..

Please read description.

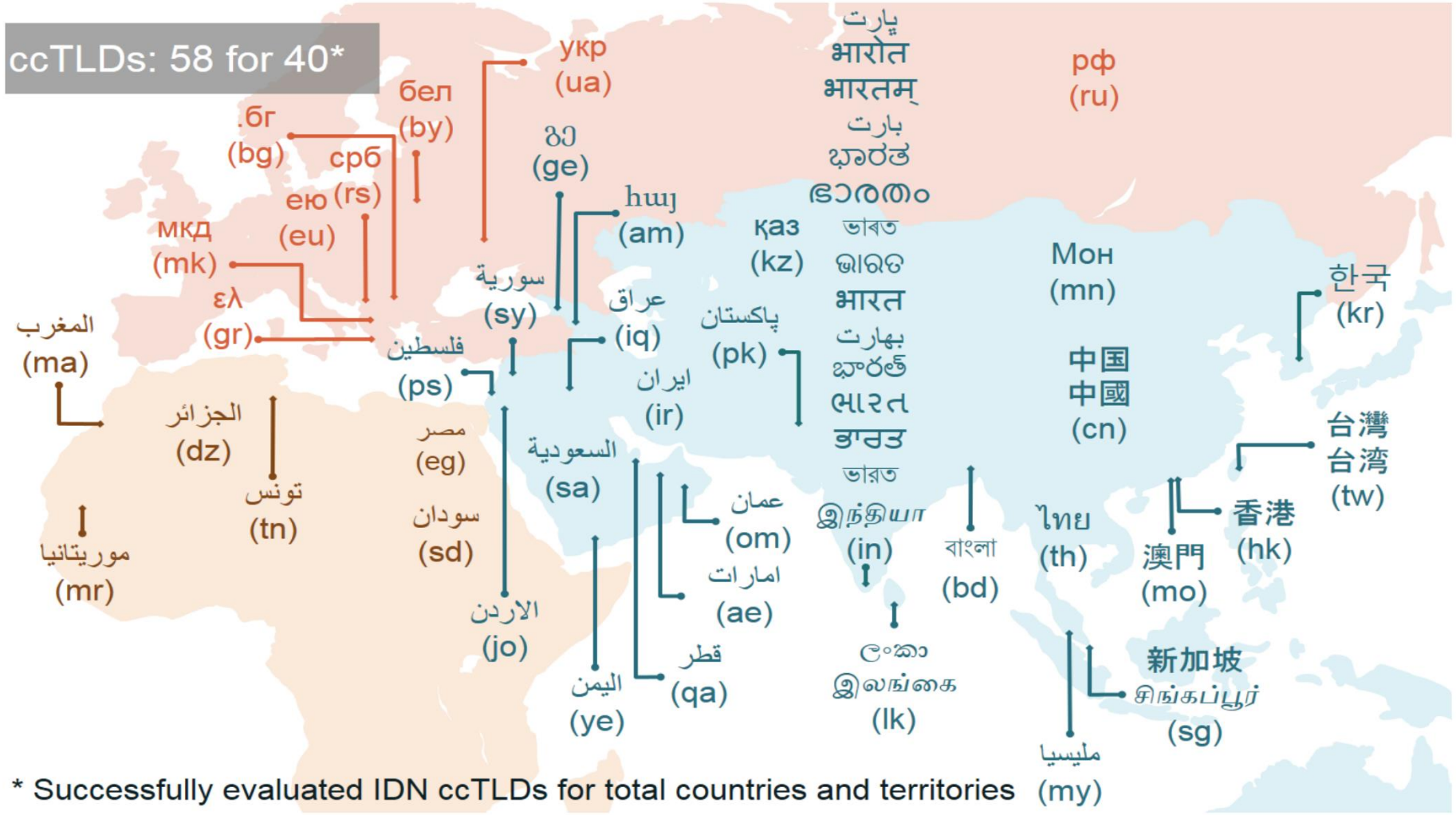
https://en.wikipedia.org/wiki/Homoglyph

Thinking

IDN TLD

- Until late 2009, TLDs were restricted to only the Latin letters a to z . After 2009, IDN TLDs were introduced in other scripts.
  - 58 IDN ccTLDs evaluated representing 40 countries/territories
  - 56 IDN ccTLDs delegated representing 38 countries/territories
  - Requests cover 33 languages in 19 scripts
- gTLD
  - com，org，net，edu，gov，mil……
  - شبكة(网络)，онлайн(在线)，グーグル(谷歌)，游戏……(IDN gTLD)
- ccTLD
  - cn，jp，nz，hr，be，cc……
  - موريتانيا(毛里塔尼亚)，新加坡，한국(韩国)，سودان(苏丹)……(IDN ccTLD)

IDN Country Code Top-Level Domains

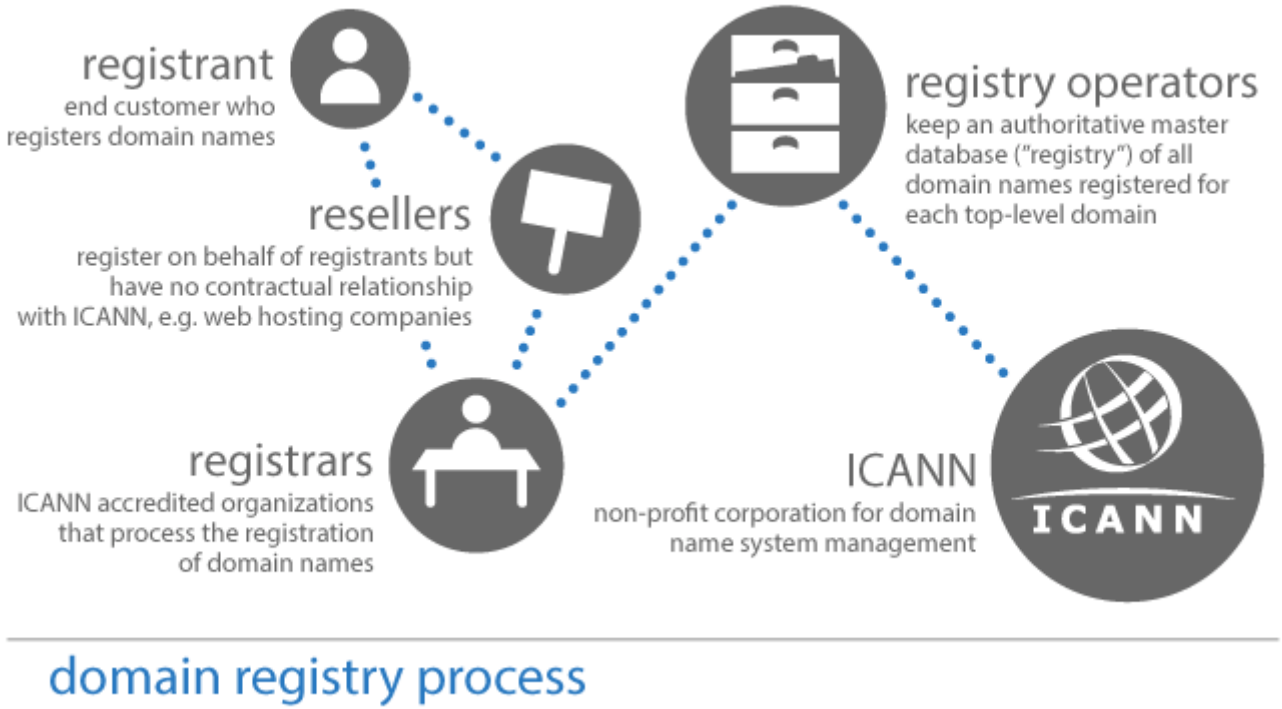


https://community.icann.org/download/attachments/84220048/IDN_update_ais2018.pdf

Think about it

- IDN TLD has been around for ten years (2009-2019). With the emergence and increasing of IDN TLD, What changes to IDN visual security?
- Whether Emoji is allowed in IDN or in IDN TLD?

Domain Name Registration Process



https://whois.icann.org/en/domain-name-registration-process

Think about it

- In the whole domain name registration process, which links are relatively weak will help us to hunt IDN spoof.

IDNs Registration Rules

- gTLD
  - .COM .NET
    - https://www.verisign.com/en_US/channel-resources/domain-registry-products/idn/idn-policy/registration-rules/index.xhtml
  - ……
    - ……
- ccTLD
  - السعودية.(Saudi Arabia.)
    - http://www.nic.sa/en/view/writing_arabic_idn_guideline
  - Association's Hebrew IDN rules
    - https://www.isoc.org.il/files/docs/ISOC-IL_Registration_Rules_v1.5_ENGLISH_-_26.6.2016.pdf
  - ……
    - ……

.COM IDNs Registration Rules

- 1. IETF Standards
  - Compliance with all of the RFC that comprise the IDNA2008 standard.
- 2. Restrictions on Specific Languages
  - All IDN registrations require a 3 letter Language Tag.
- 3. Restrictions On Commingling Of Scripts
  - All code points within an IDN must come from the same Unicode script
  - ……
- 4. ICANN’ s Restricted Unicode Points
  - ……
- 5. Special Characters
  - ……

Unicode Scripts And Associated Code Points

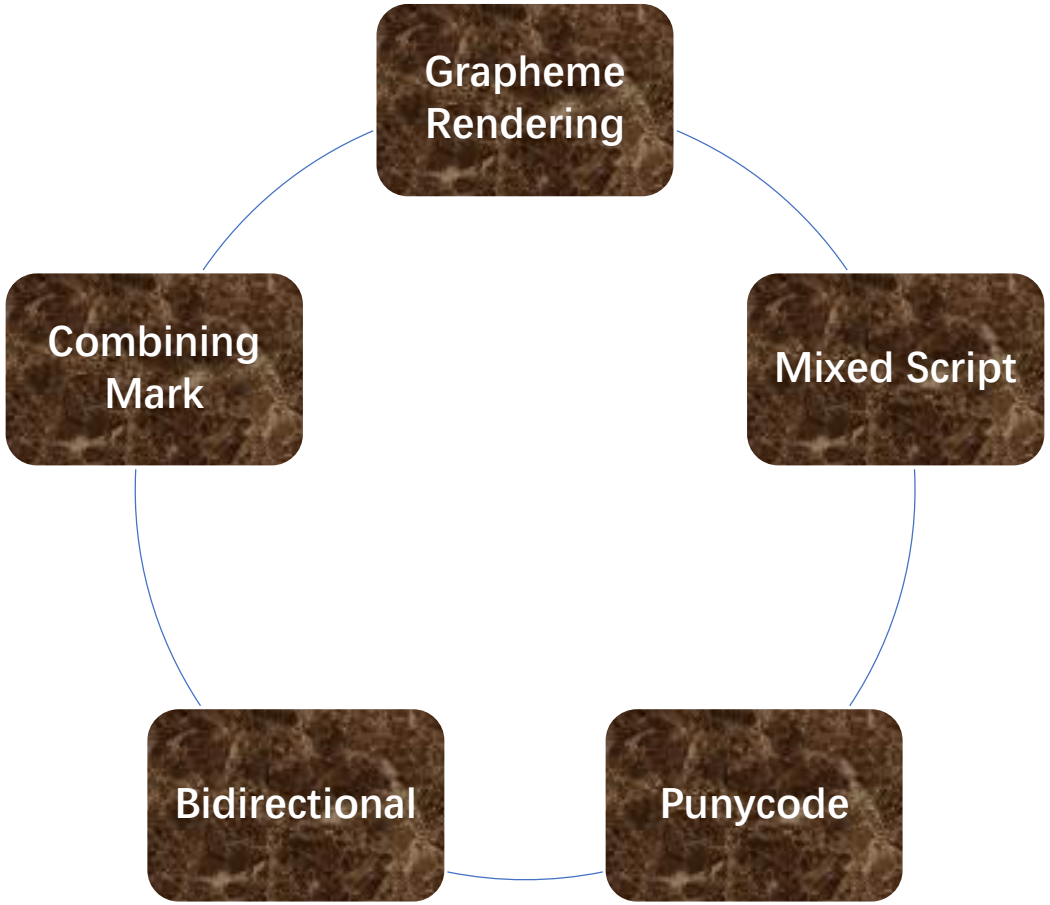
Arabic
Armenian
Avestan
Balinese
Bamum
Batak
Bengali
Bopomofo
Brahmi
Buginese
Buhid
Canadian Aboriginal
Carian
Cham
Cherokee
Coptic
Cuneiform
Cyrillic
Devanagari
Egyptian Hierogyphs
Ethiopic
Georgian
Glagolitic
Greek
Gujarati
Gurmukhi
Han
Hangul
Hanunoo
Hebrew
Hiragana
Imperial Aramaic
Inscriptional Pahlavi
Inscriptional Parthian
Javanese
Kaithi
Kannada
Katakana
Kayah Li
Kharoshthi
Khmer
Lao
Latin
Lepcha
Limbu
Lisu
Lycian
Lydian
Malayalam
Mandaic
Meetei Mayek
Mongolian
Myanmar
New Tai Lue
Nko
Ogham
Ol Chiki
Old Persian
Old South Arabian
Old Turkic
Oriya
Phags Pa
Phoenician
Rejang
Runic
Samaritan
Saurashtra
Sinhala
Sundanese
Syloti Nagri
Syriac
Tagalog
Tagbanwa
Tai Le
Tai Tham
Tai Viet
Tamil
Telugu
Thaana
Thai
Tibetan
Tifinagh
Vai
Yi

Think about it

- How many IDN rules are there in the world ?
- How do we find IDN Spoof under these rules?

Continue Thinking

Unicode Visual Security Issues



Grapheme Rendering

- Inadequate Rendering
  - An additional problem arises when a font or rendering engine has inadequate support for characters or sequences of characters.

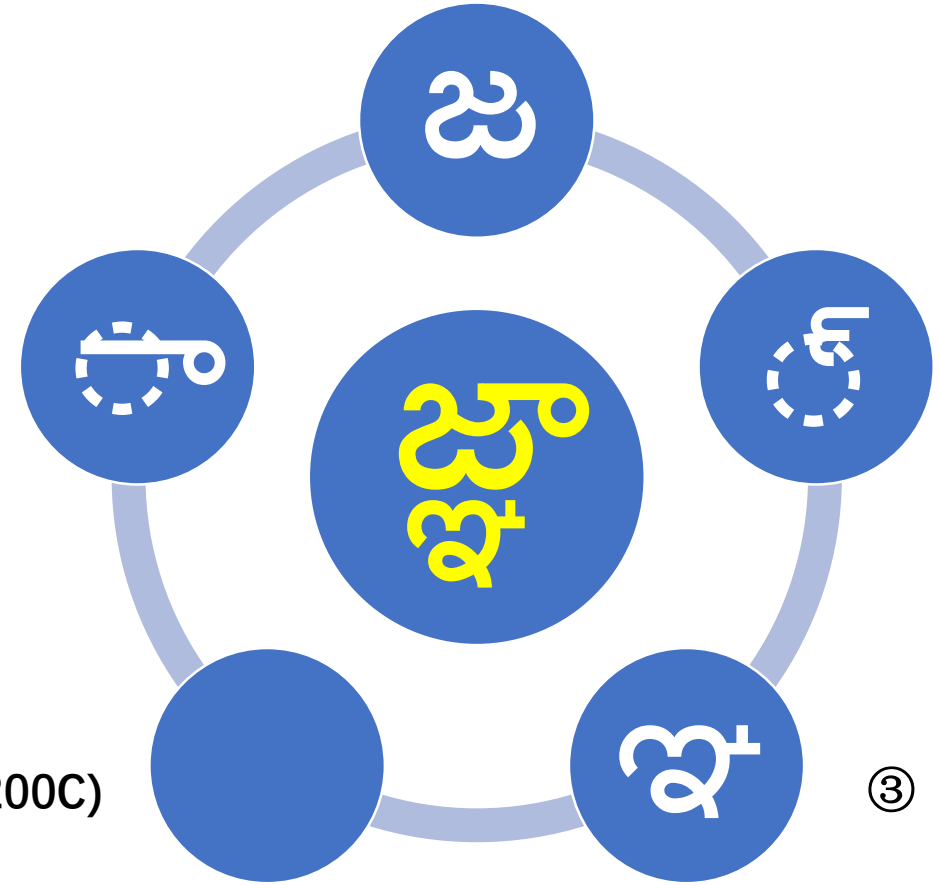
- Quick Example

String	UTF-16	Punycode
e̩l.com	0065 006C 0323 002E 0063 006F 006D	xn--e-zom.com
ẹl.com	0065 0323 006C 002E 0063 006F 006D	xn--l-ewm.com
ẹl.com	1EB9 006C 002E 0063 006F 006D	xn--l-ewm.com

http://www.unicode.org/reports/tr36/#Inadequate_Rendering_Support

CVE-2018-4124

① telugu sign virama (U+0C1C)



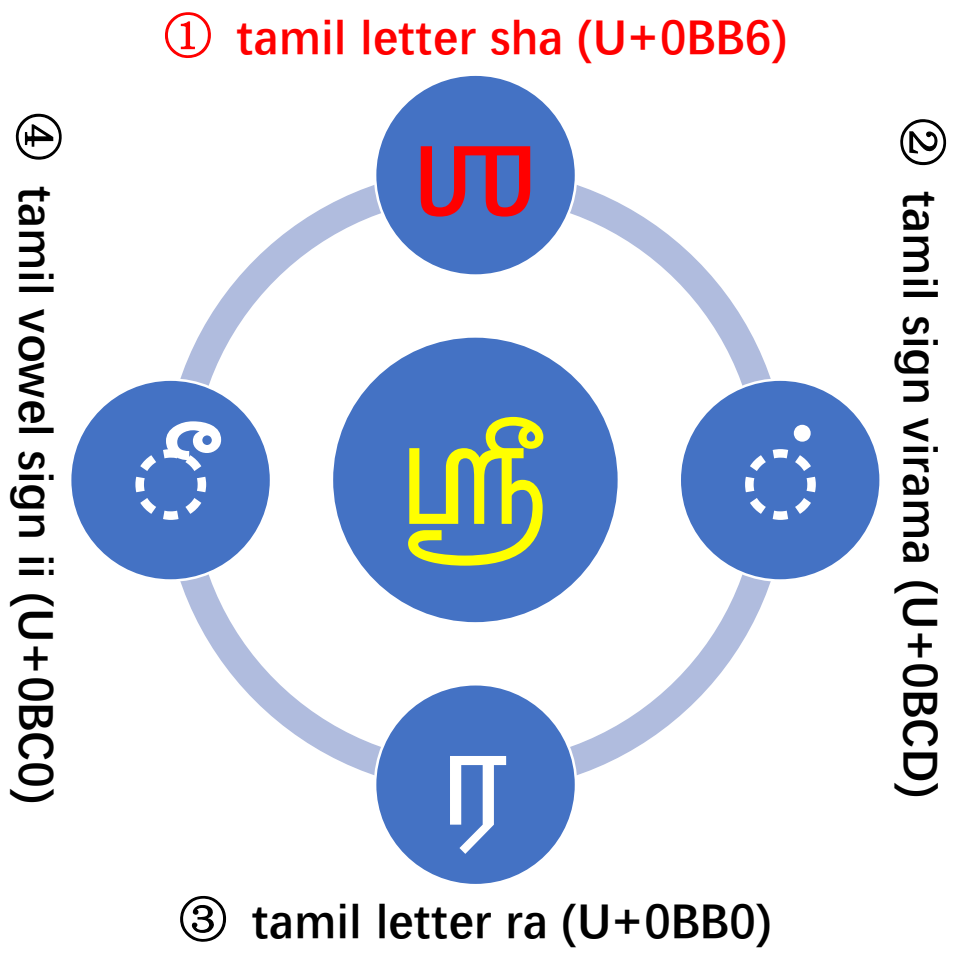
⑤ telugu vowel sign aa (U+0C3E)

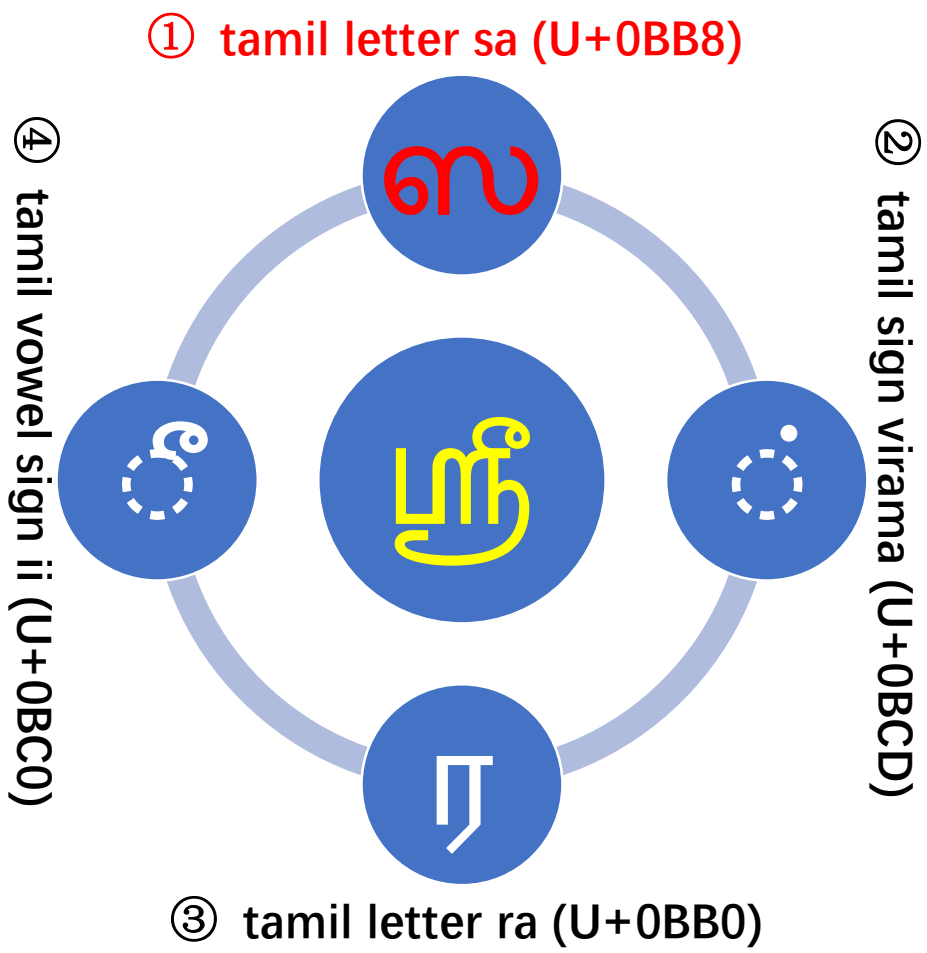
② telugu sign virama (U+0C4D)

④ zero width non-joiner (U+200C)

③ telugu letter nya (U+0C1E)

Glyphs in Complex Scripts





Mixed-Script

- Mixed-Script Spoofing
  - The characters in some scripts, though different in meaning, are usually identical or nearly identical in appearance. However, the existence of visually confusable characters across scripts offers numerous opportunities for spoofing.
- Quick Example

String	UTF-16	Punycode
top.com	0074 **03BF** 0070 002E 0063 006F 006D	xn--tp-jbc.com
top.com	0074 **006F** 0070 002E 0063 006F 006D	top.com

http://www.unicode.org/reports/tr36/#Mixed_Script_Spoofing

Punycode

- PunyCode Spoofing
  - Punycode is a special encoding used to convert Unicode characters to ASCII, which is used to encode internationalized domain names (IDN). The Punycode transformation is relatively dense. That means that it is fairly likely that arbitrary words after the "xn--" will result in valid labels.
- Quick Example
  - URL: http://藰藹護蕸.com
  - PunyCode URL: http://xn--google.com

http://www.unicode.org/reports/tr36/#TablePunycodeSpoofing

Bidirectional Text

- Bidirectional Text Spoofing
  - When characters are mixed with left-to-right text(LTR) and right-to-left text(RTL), Unicode Bidirectional Algorithm will use a precise set of rules to determine the final visual rendering. However, presented with arbitrary sequences of text, this may lead to text sequences which may be impossible to read intelligibly, or which may be visually confusable.
- Quick Example
  - Access to http://127.0.0.1/%D8%A7/example.org
  - In address bar, maybe visual rendering http://example.org/ /127.0.0.1

http://www.unicode.org/reports/tr36/#Bidirectional_Text_Spoofing

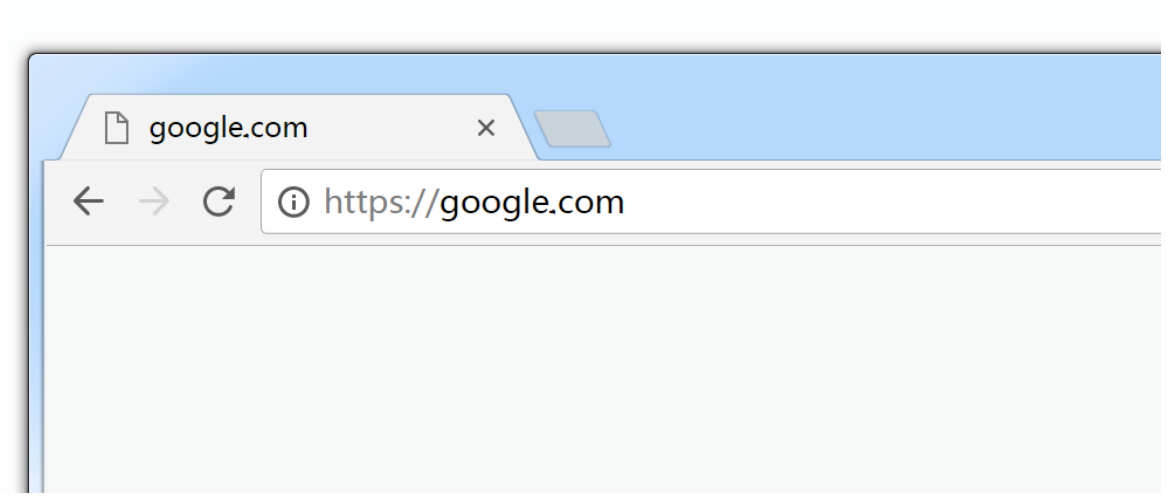
LTR vs RTL

- URL-LTR
  - Subdomain: hi
  - Domain: google
  - TLD: com
  - Path: search
- http://hi.google.com/search

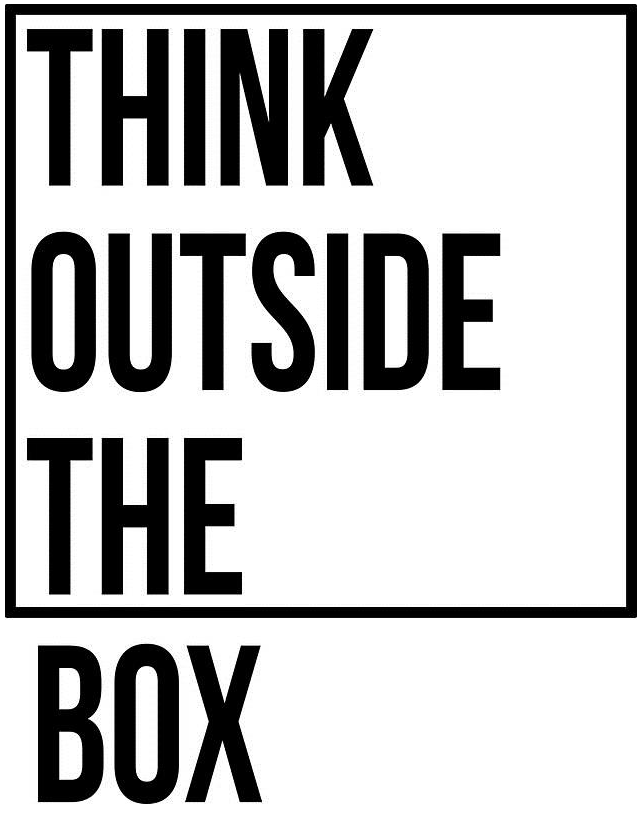
- URL-RTL(hebrew)
  - Subdomain: ף
  - Domain: ר
  - TLD: ל
  - Path: מ
- 渲染：http://מ/ל.ר.ף

Combining Mark

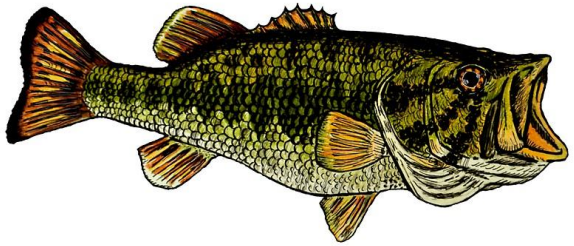
- Combing mark spoofing
  - Combining mark are characters that are intended to modify other characters. Combining character sequence maybe become visually indistinguishable with other characters .
- Quick Example
  - google◌.com
  - ◌ = U+18A9



http://www.unicode.org/reports/tr36/#TableCombiningMarkOrderSpoofing


THINK
OUTSIDE
THE
BOX

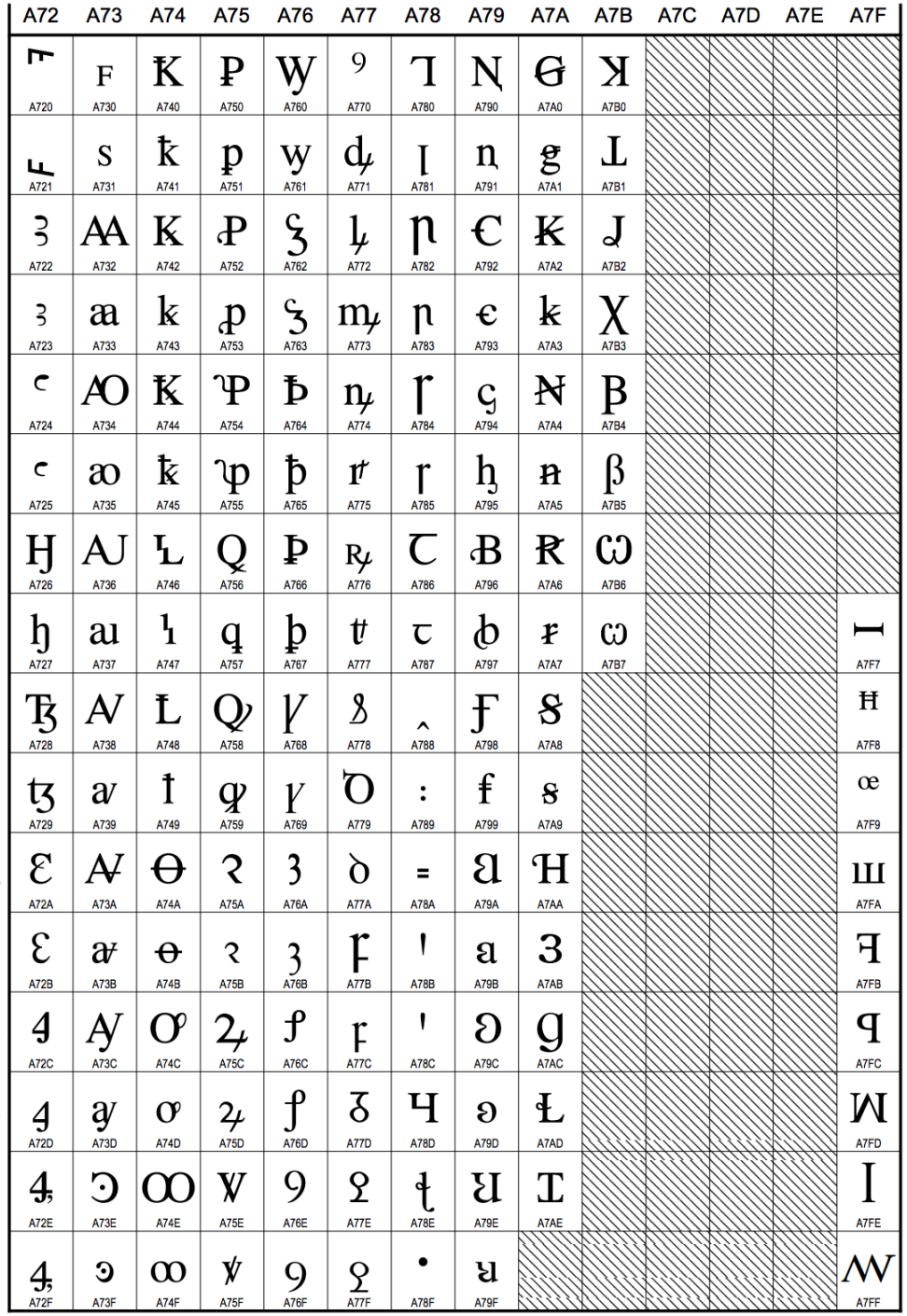
Inadequate Grapheme Rendering



Safari Address Bar Spoof Using Latin-d
CVE-2018-4277

https://xlab.tencent.com/en/2018/11/13/cve-2018-4277

Latin Extended-D



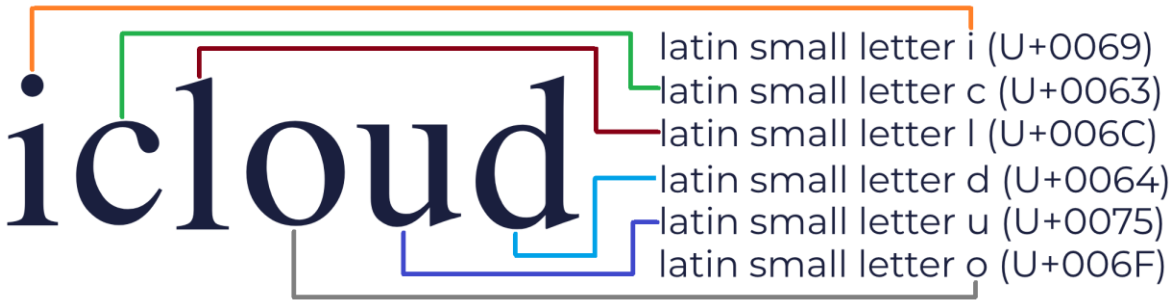
ꝸ

latin small letter dum (U+A771)

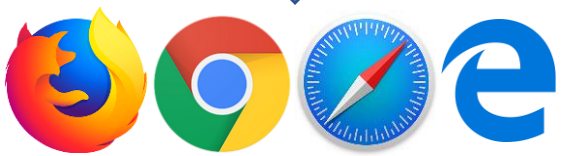
ⅆ

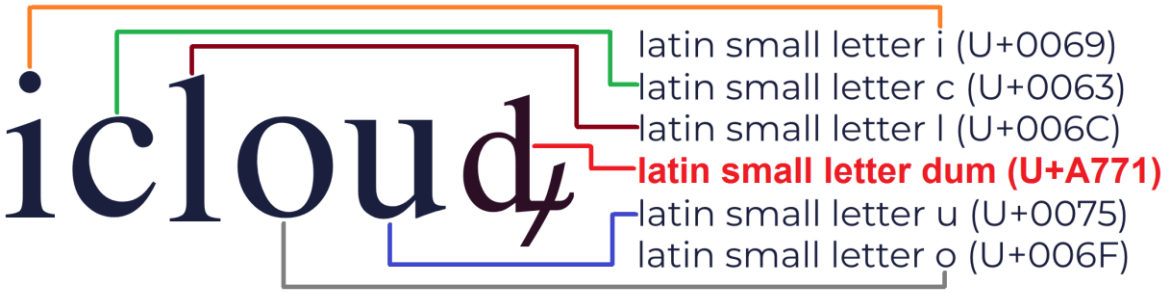
In Safari address bar

Latin: icloud.com VS Latin Extended-D: icloud.com



icloud.com

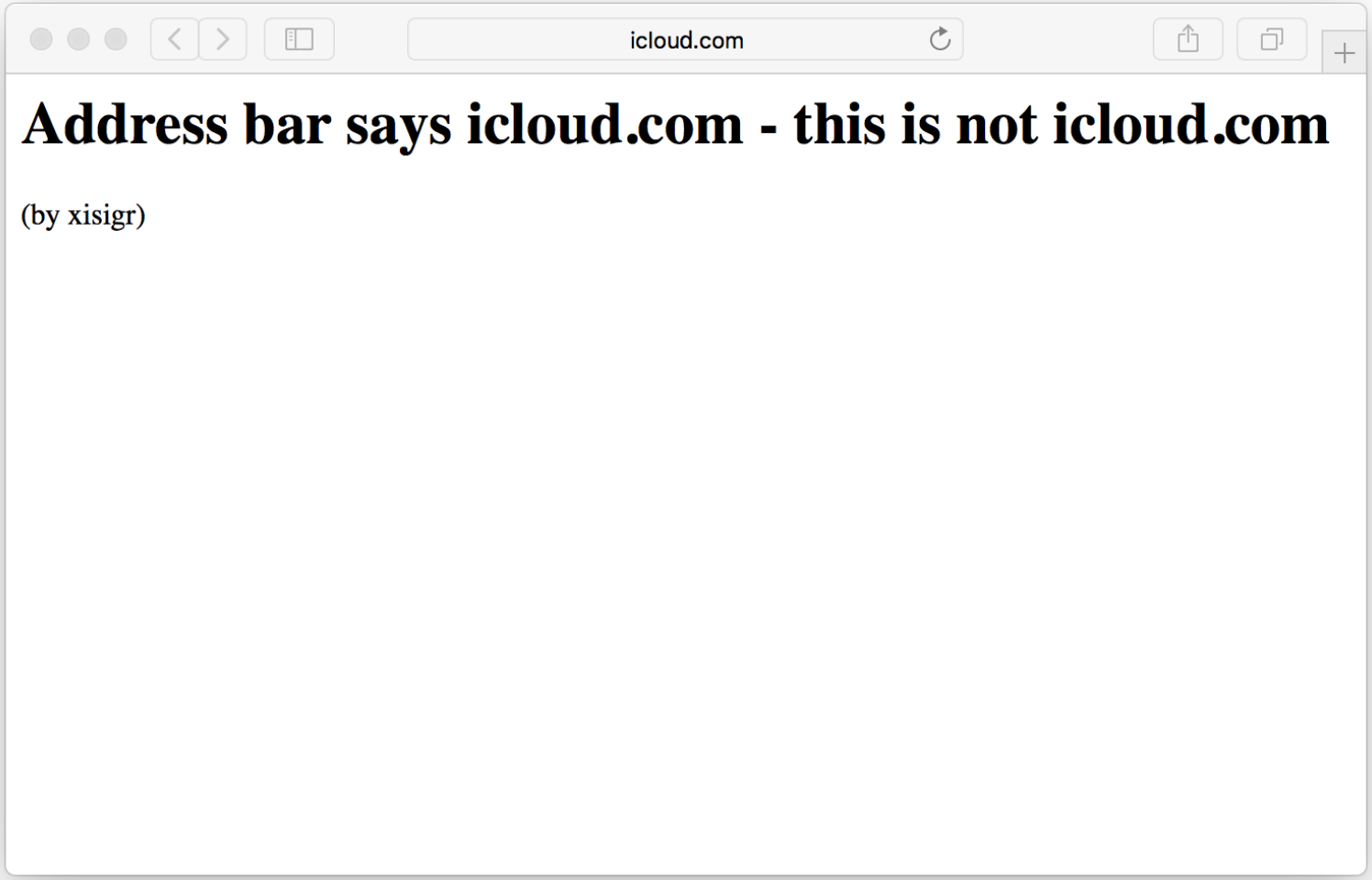




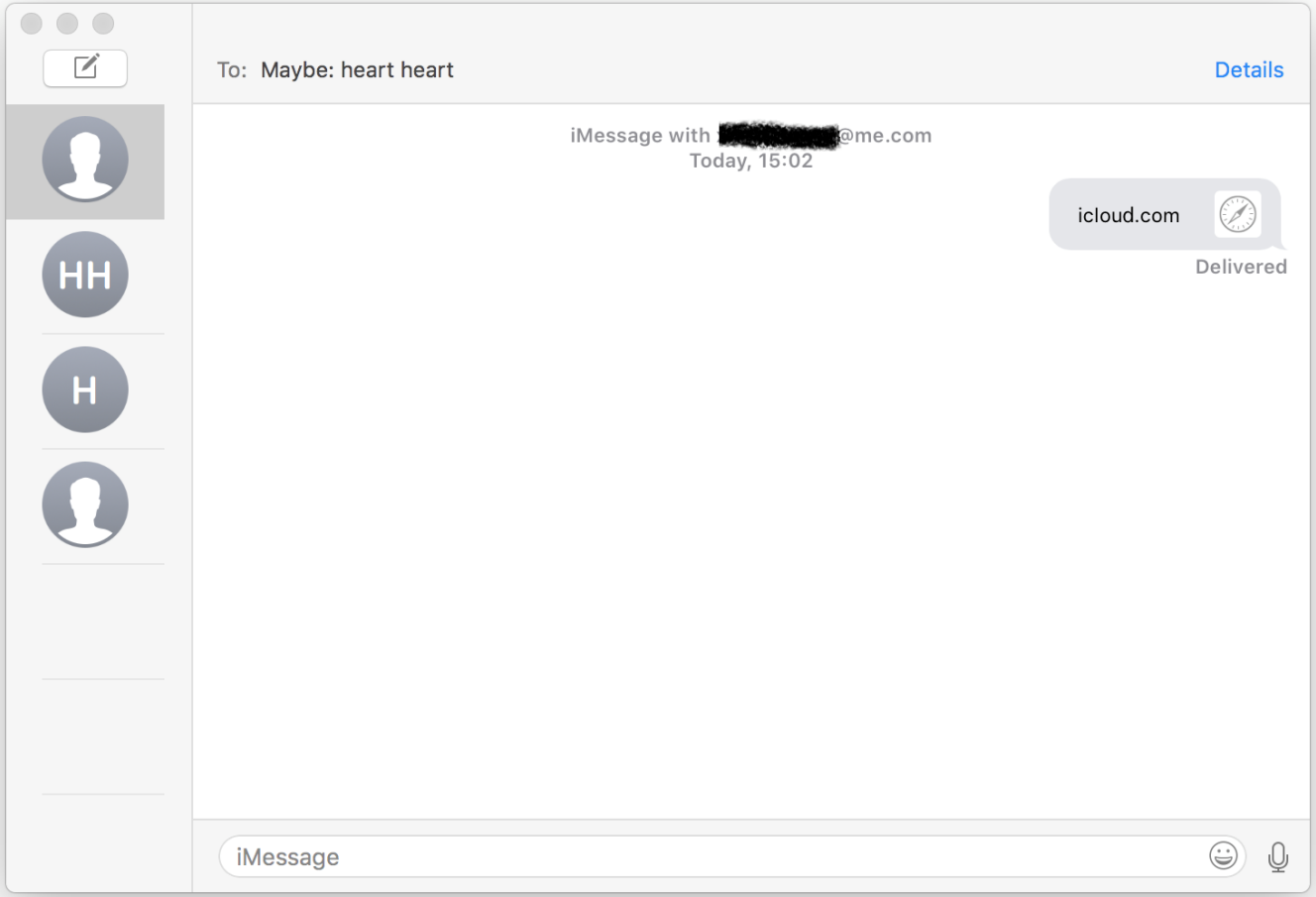
xn--iclou-rl3s.com

icloud.com

xn--iclou-rl3s.com



xn--iclou-rl3s.com



Domain include d can be spoof

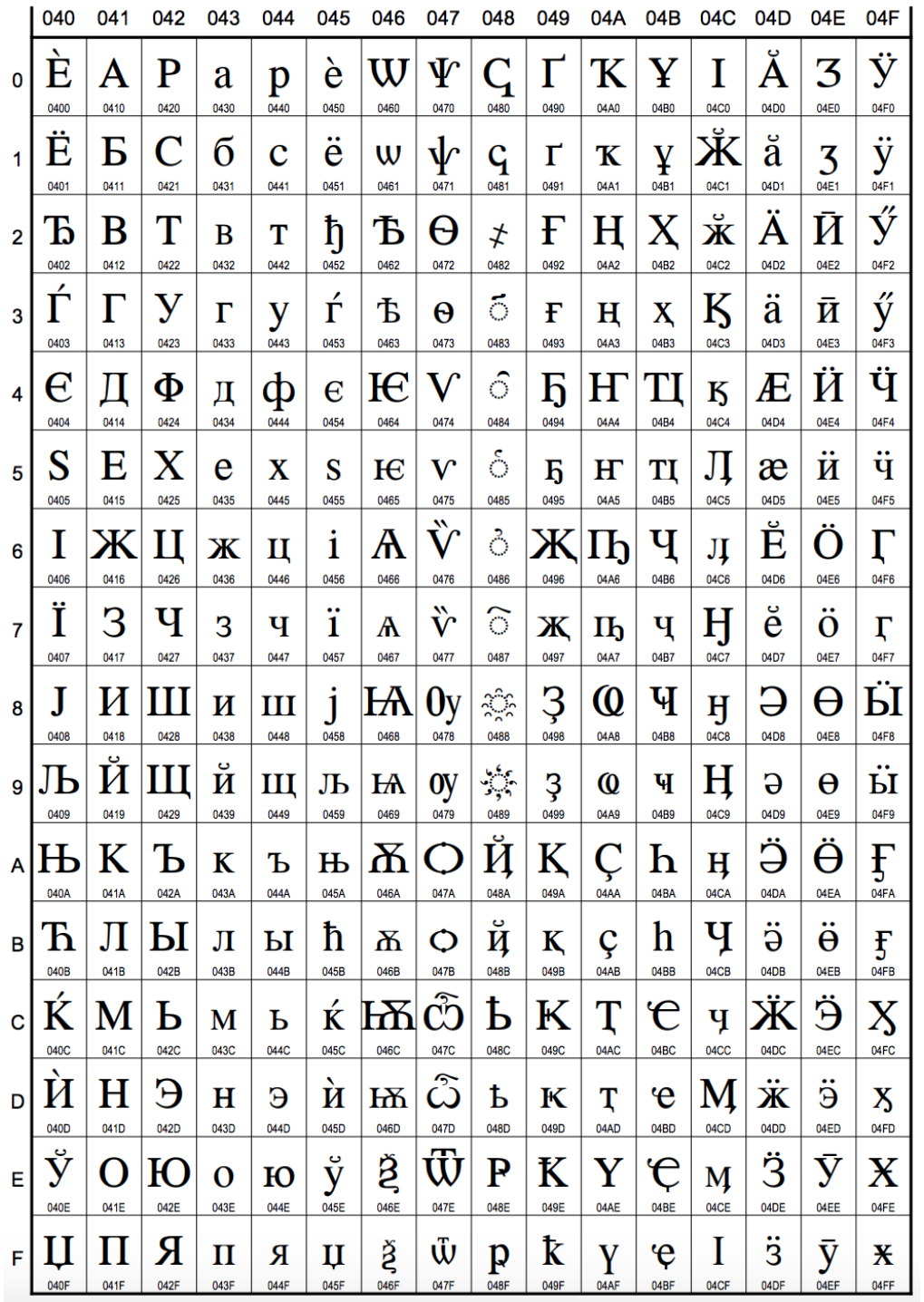
- Top 10k , domain include d > 25%
  - linkedin.com
  - baidu.com
  - jd.com
  - adobe.com
  - wordpress.com
  - dropbox.com
  - godaddy.com
  - reddit.com
  - ……

Whole-Script

Address Bar Spoof using Cyrillic
CVE-2017-5060

https://www.xudongz.com/blog/2017/idn-phishing

Cyrillic



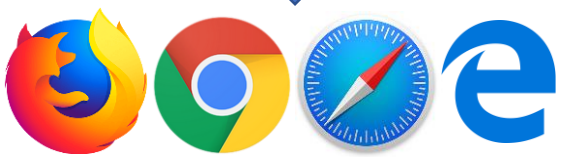
a cyrillic small letter a (U+0430)

p cyrillic small letter er (U+0440)

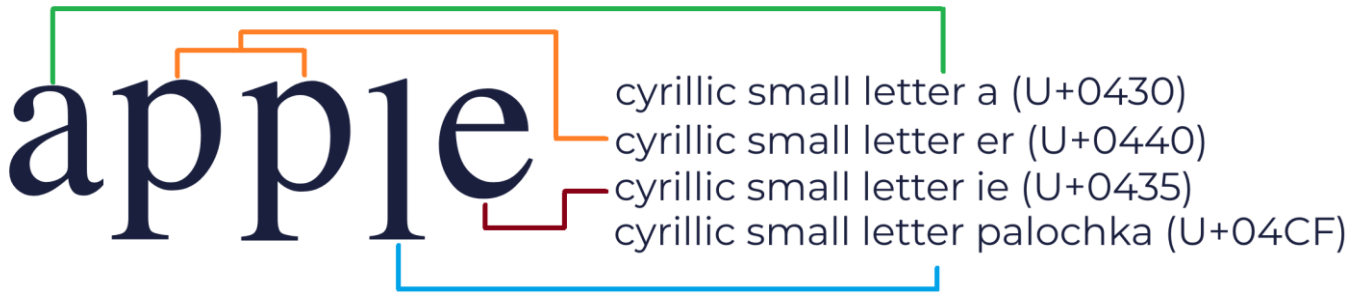
l cyrillic small letter palochka (U+04CF)

e cyrillic small letter ie (U+0435)

Latin: apple.com VS Cyrillic : apple.com



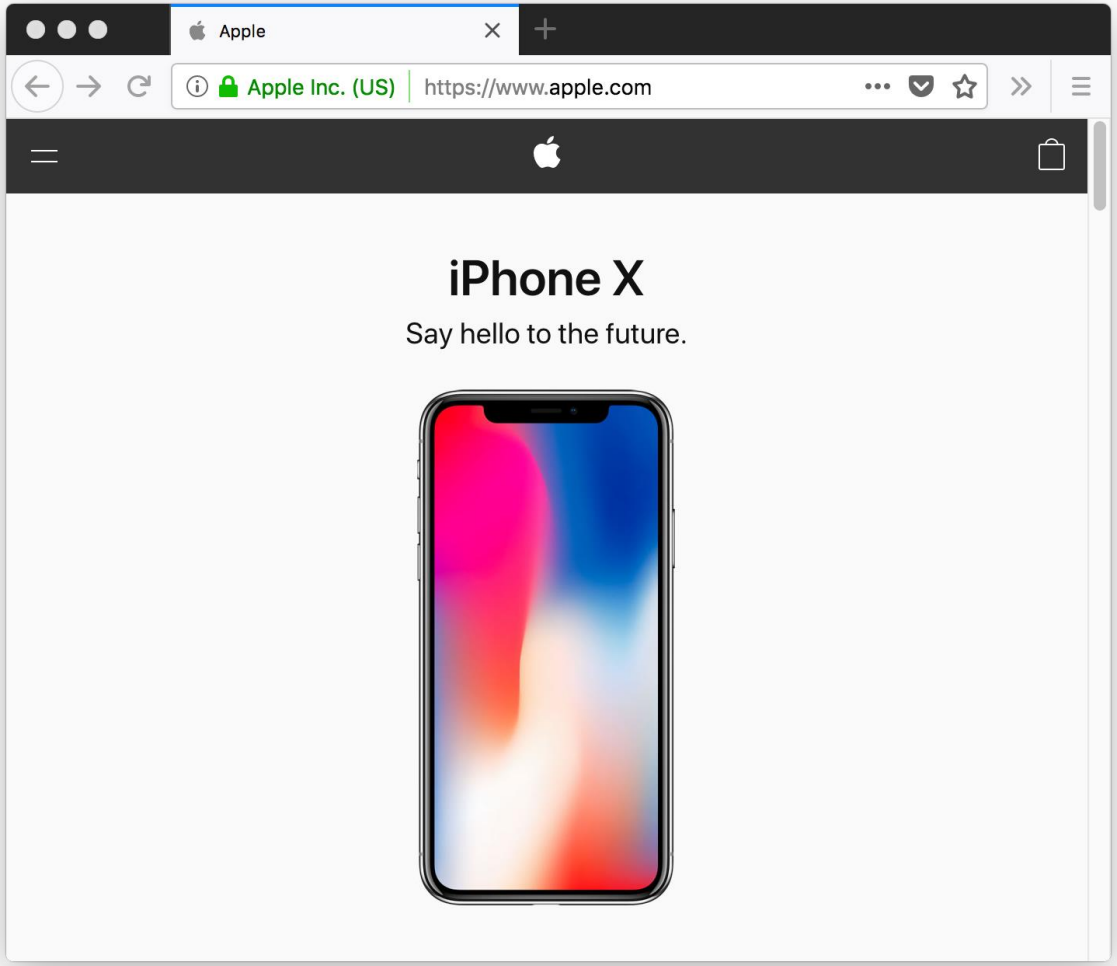
apple.com

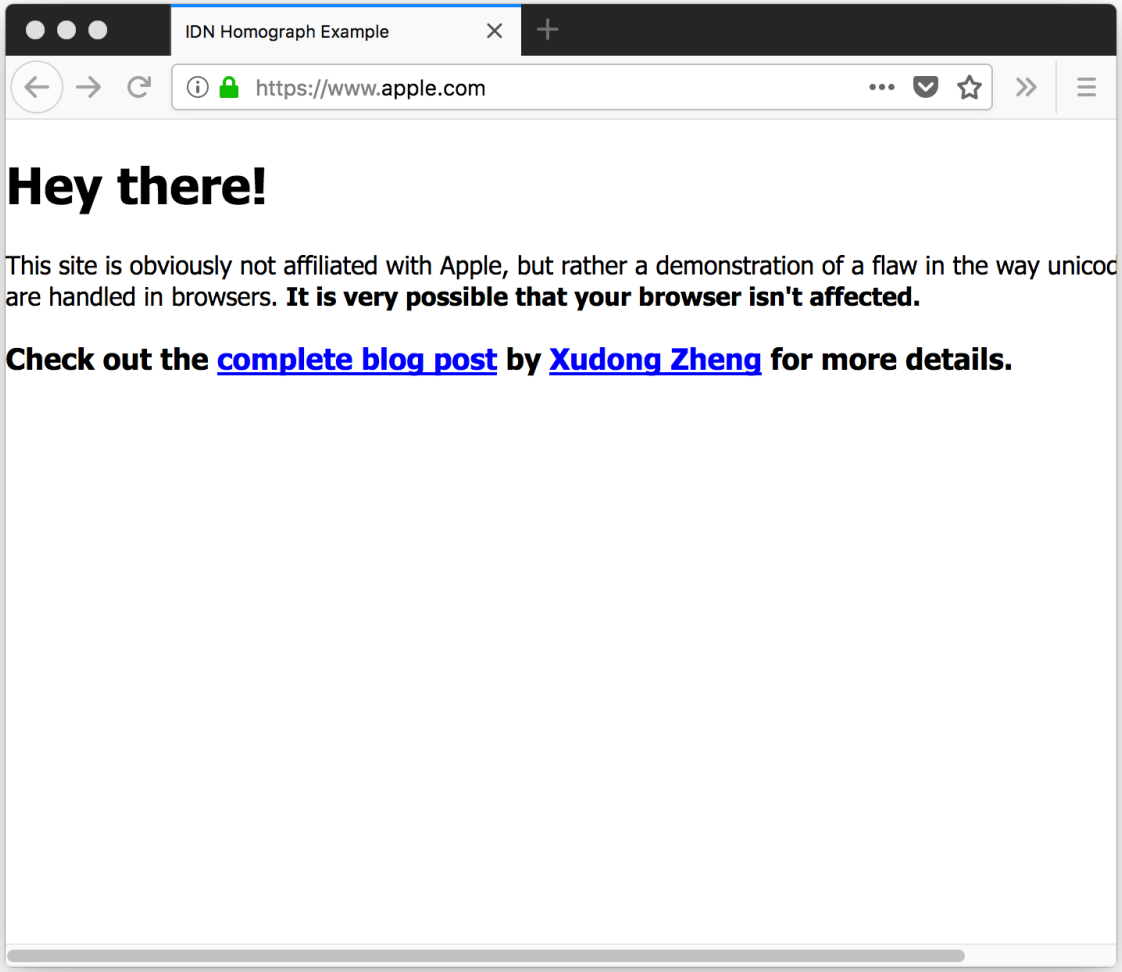


xn--80ak6aa92e.com

apple.com

Latin: apple.com VS Cyrillic : apple.com



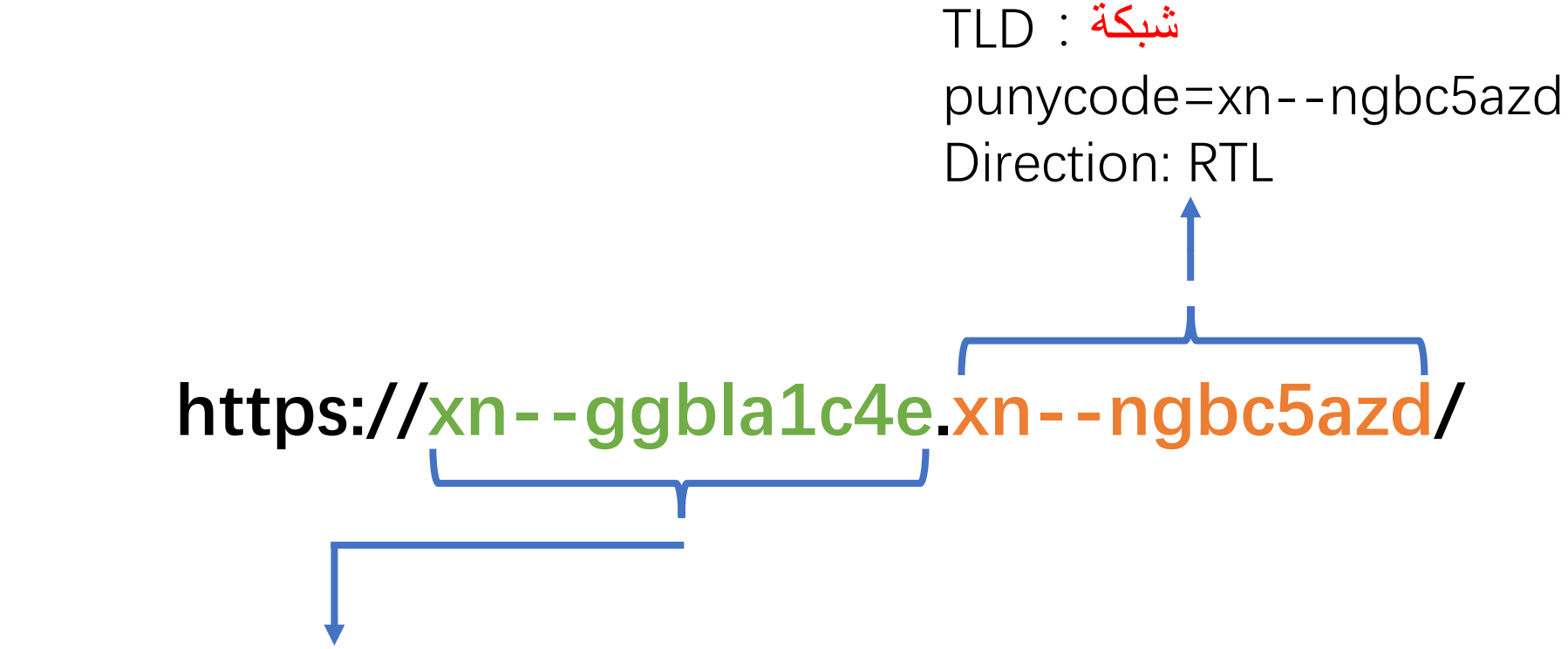


Bidirectional Text

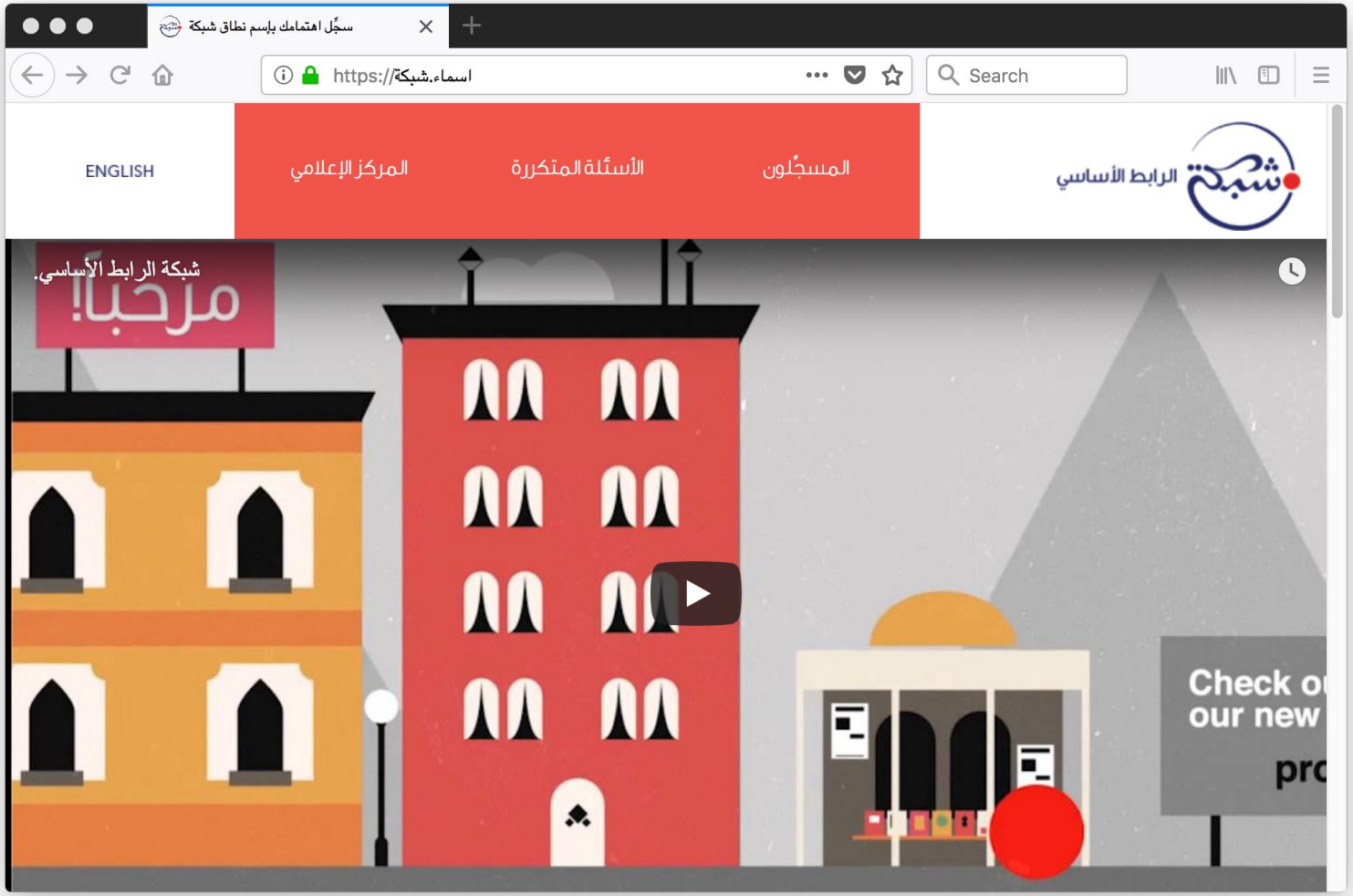
Firefox URL spoof using RTL
CVE-2018-5117

https://www.mozilla.org/en-US/security/advisories/mfsa2018-02/#CVE-2018-5117

CVE-2018-5117



https://xn--ggbla1c4e.xn--ngbc5azd



CVE-2018-5117

TLD：شبكة
punycode=xn--ngbc5azd
Direction: RTL

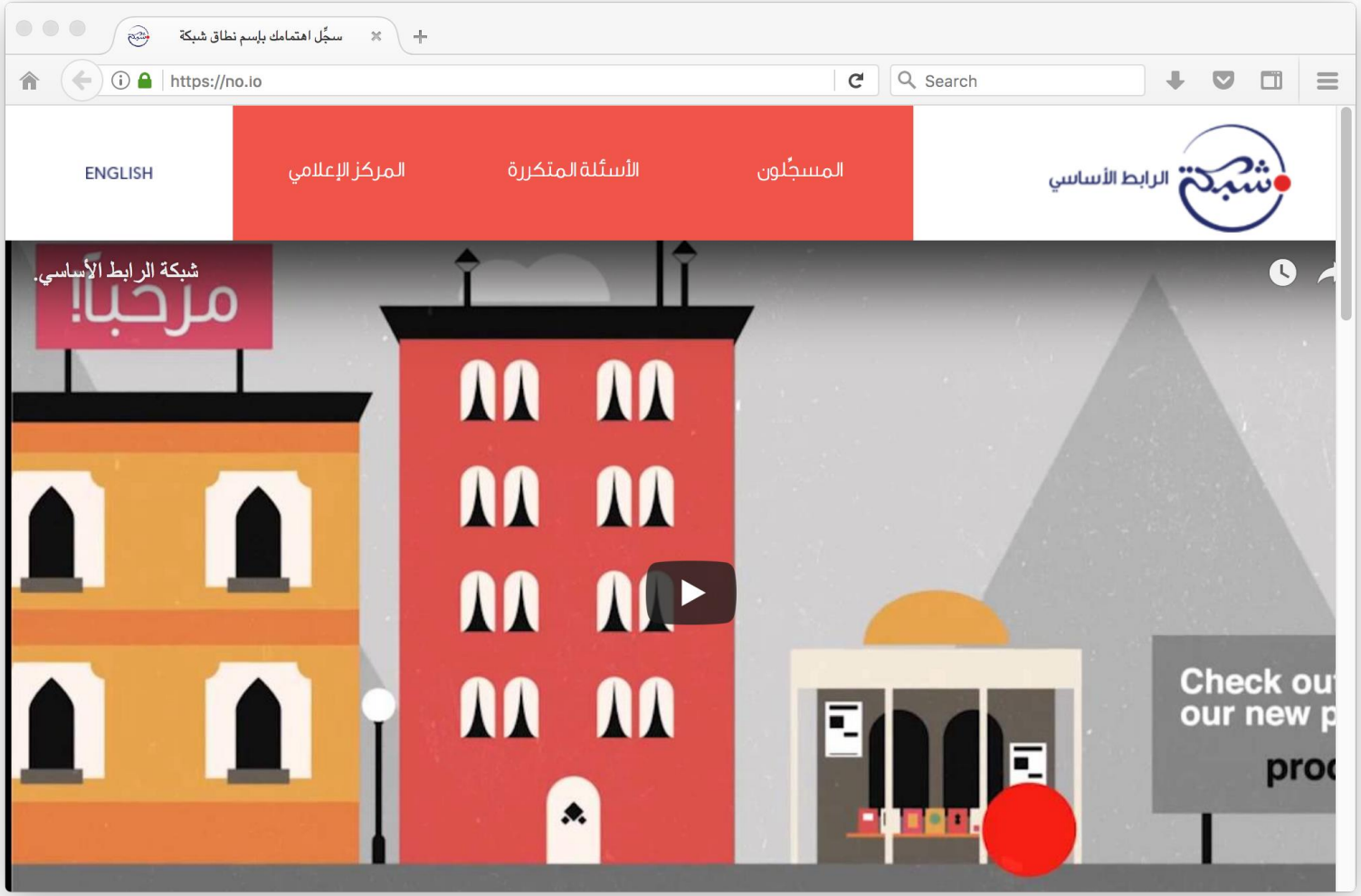
Pathname: #%20%20%20no.io
Direction: RTL

SLD: اسماء
Punycode= xn--ggbla1c4e
Direction: RTL

A lot of %20
%20=Blank character

Hebrew
Direction: RTL

https://xn--ggbla1c4e.xn--ngbc5azd/#%20%20%20סו.סה

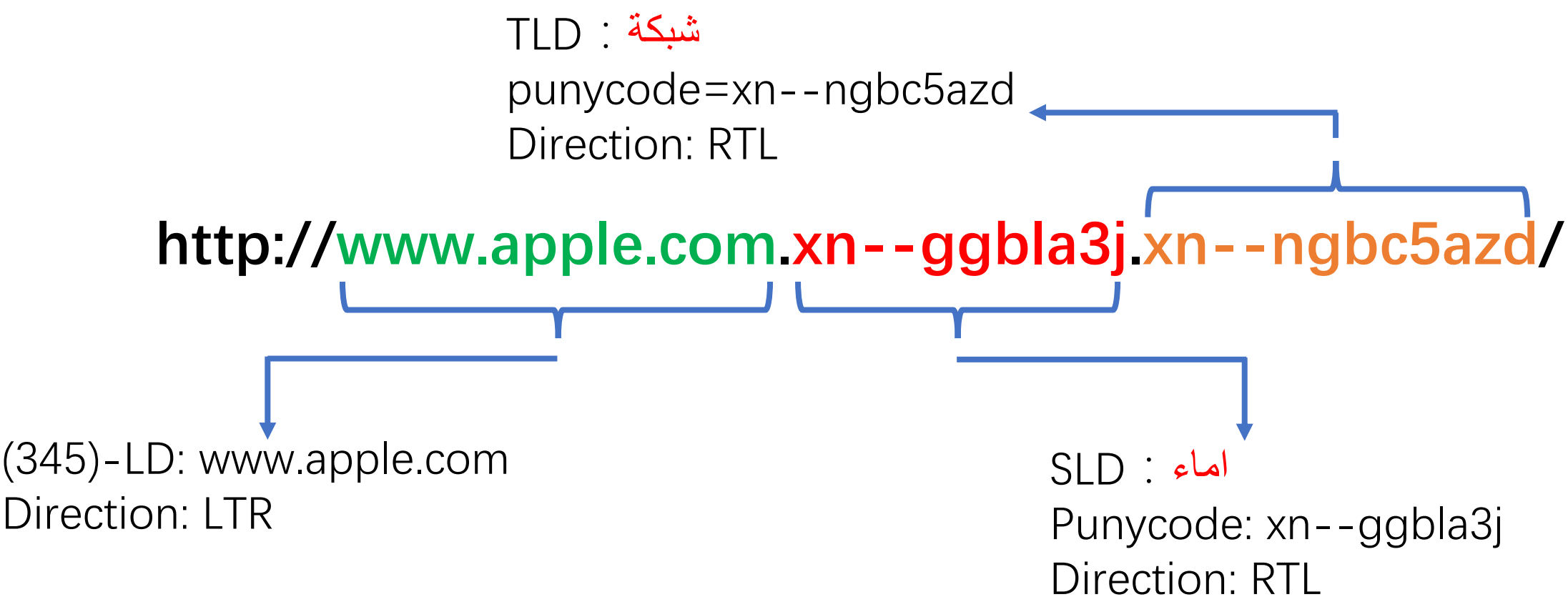


Bidirectional Text

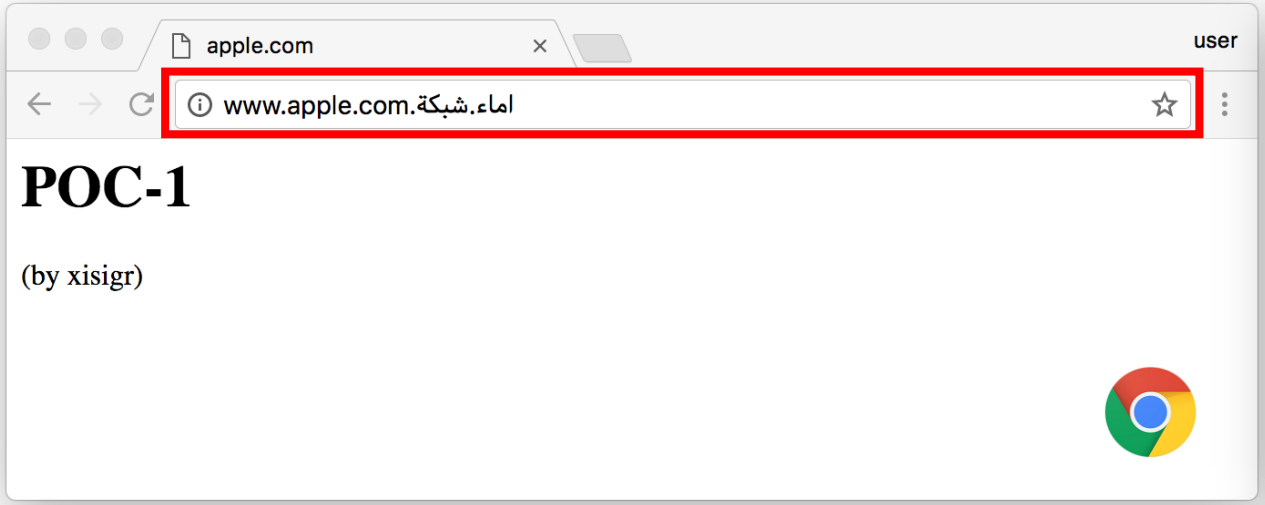
URL Spoof Using RTL IDN TLD
CVE-2018-4205

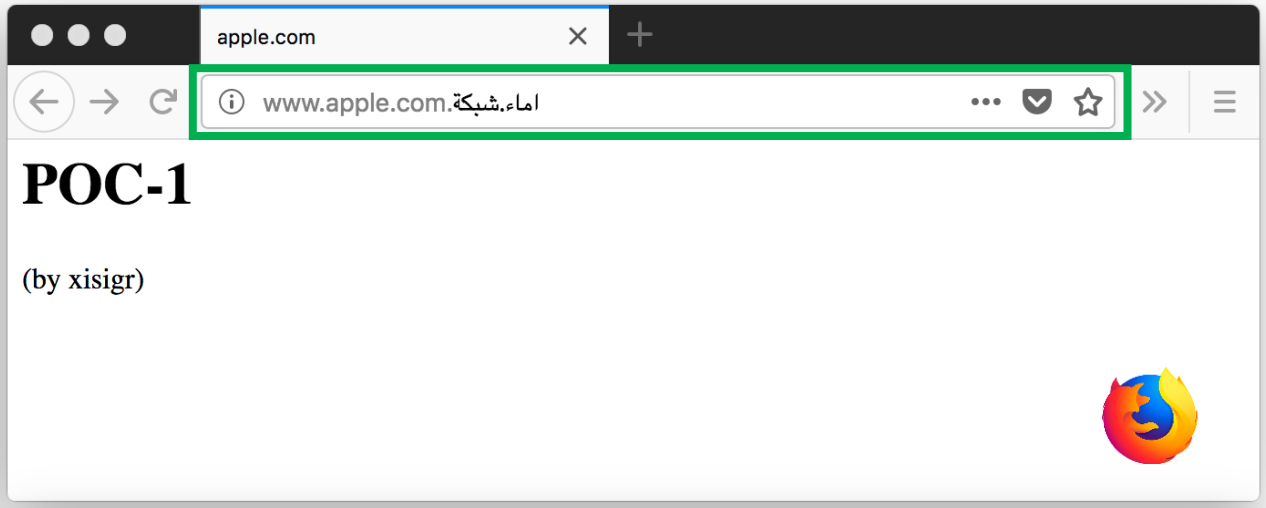
https://support.apple.com/en-us/HT208854

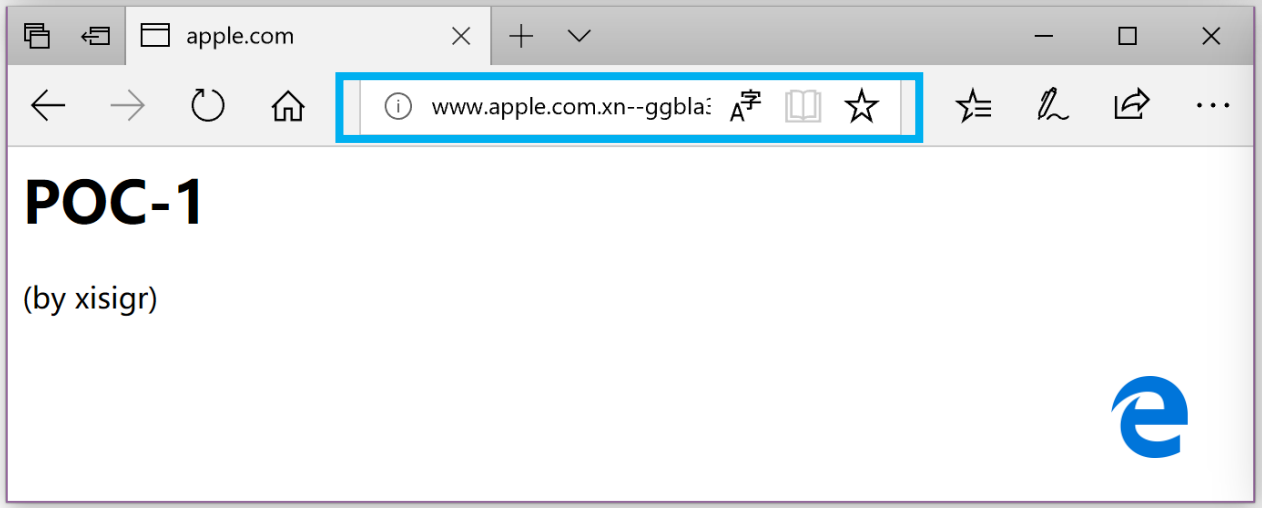
POC-1

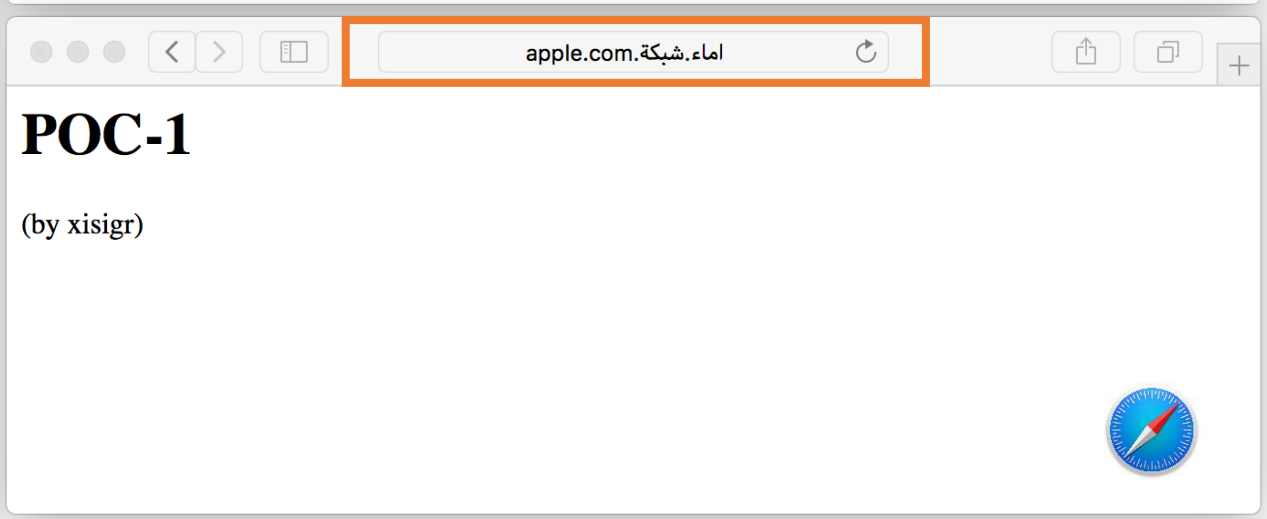

TLD：شبكة
punycode=xn--ngbc5azd
Direction: RTL
http://www.apple.com.xn--ggbla3j.xn--ngbc5azd/
(345)-LD: www.apple.com
Direction: LTR
SLD：اماء
Punycode: xn--ggbla3j
Direction: RTL

POC-1

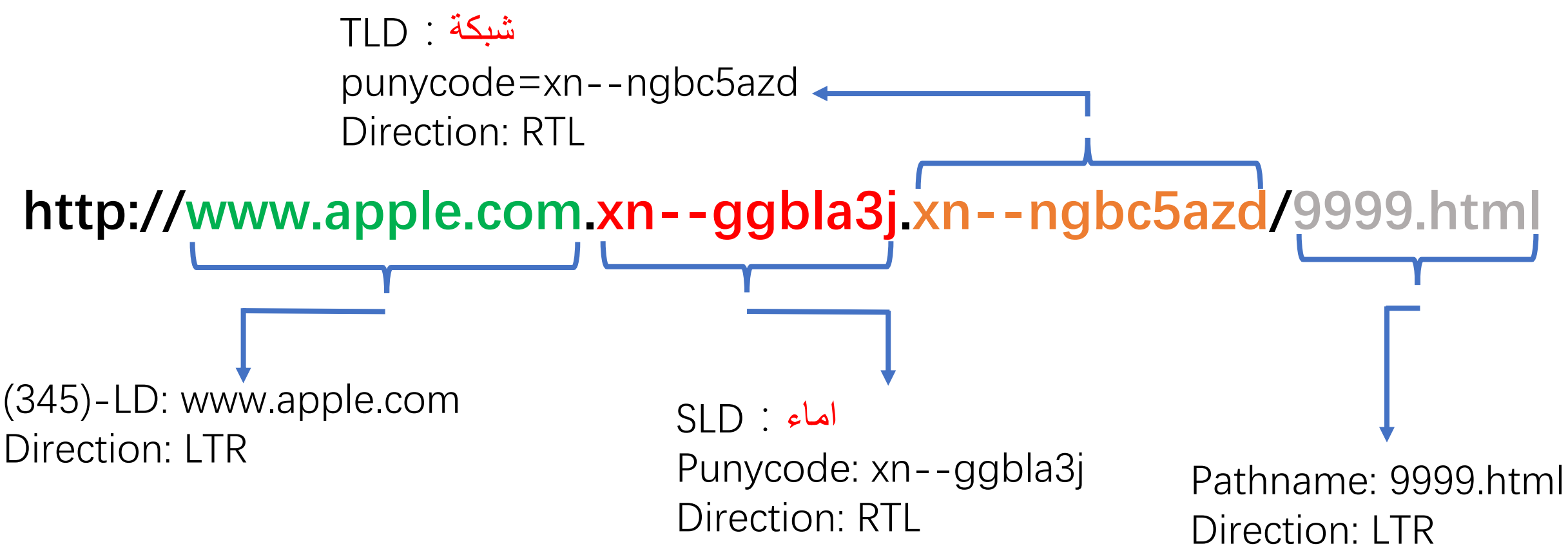

apple.com
user
www.apple.com.اماء.شبكة
POC-1
(by xisigr)


apple.com
www.apple.com.اماء.شبكة
POC-1
(by xisigr)

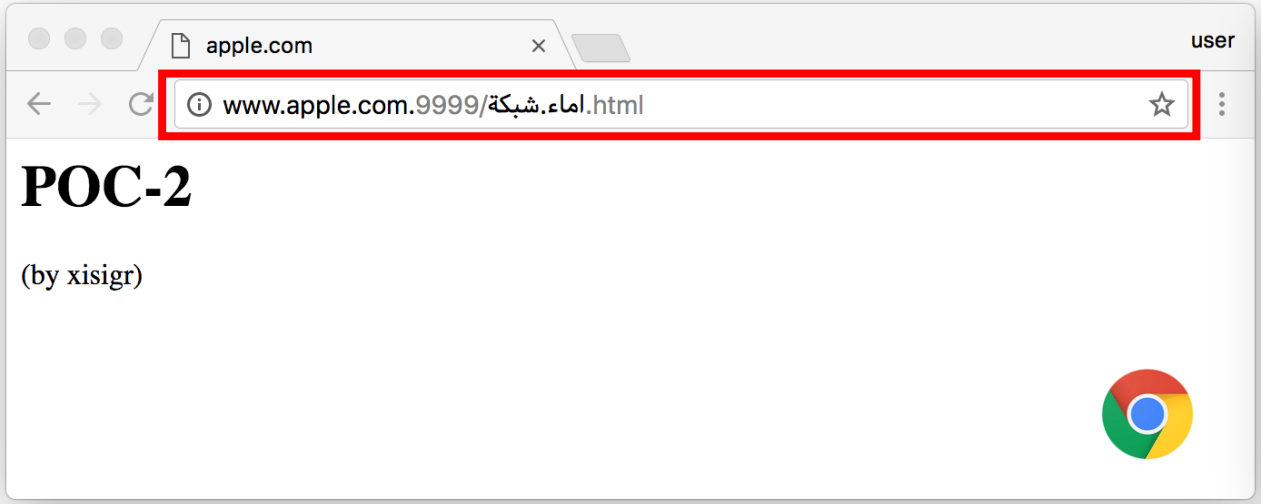

apple.com
www.apple.com.xn--ggbla
POC-1
(by xisigr)

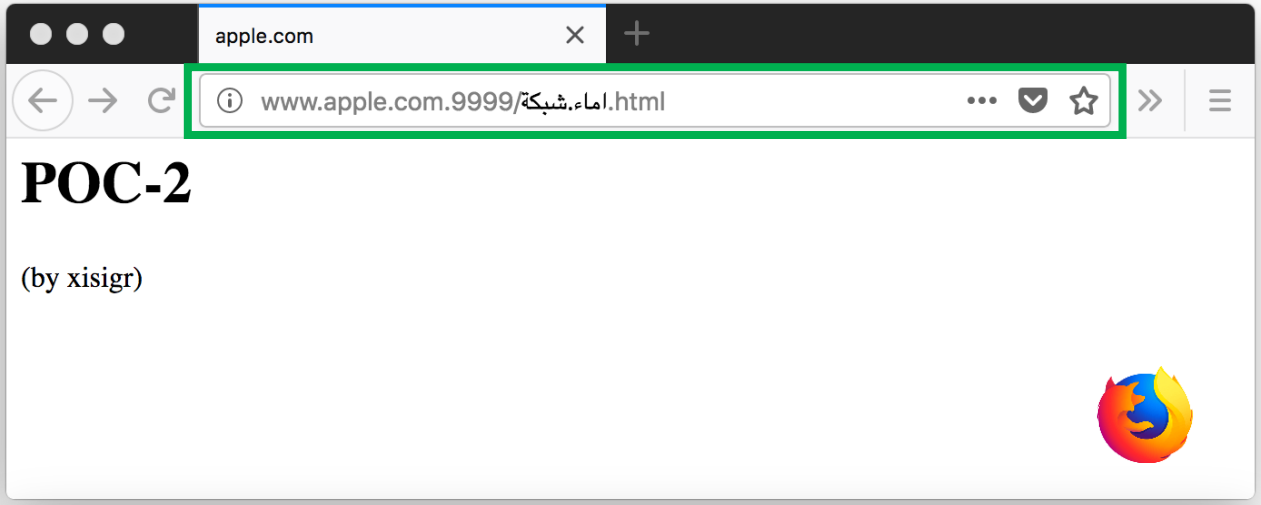

apple.com.اماء.شبكة
POC-1
(by xisigr)

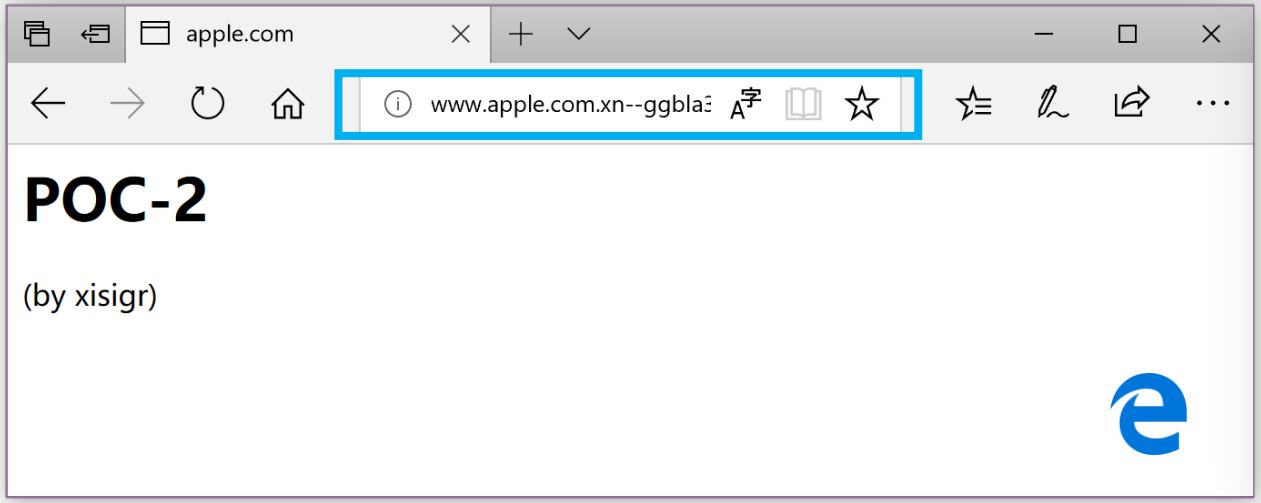
POC-2

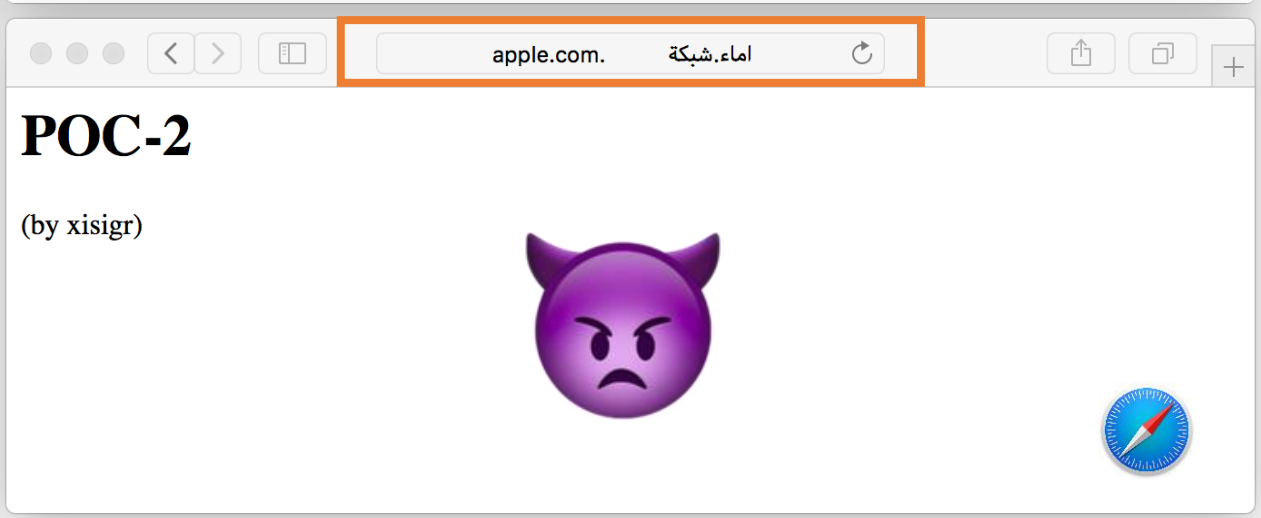

TLD：شبكة
punycode=xn--ngbc5azd
Direction: RTL
http://www.apple.com.xn--ggbla3j.xn--ngbc5azd/9999.html
(345)-LD: www.apple.com
Direction: LTR
SLD：اماء
Punycode: xn--ggbla3j
Direction: RTL
Pathname: 9999.html
Direction: LTR

POC-2


apple.com
user
www.apple.com.9999/اماء.شبكة.html
POC-2
(by xisigr)


apple.com
www.apple.com.9999/اماء.شبكة.html
POC-2
(by xisigr)


apple.com
www.apple.com.xn--ggbla
POC-2
(by xisigr)


apple.com.
اماء.شبكة
POC-2
(by xisigr)

POC-3

TLD：شبكة
punycode=xn--ngbc5azd
Direction: RTL

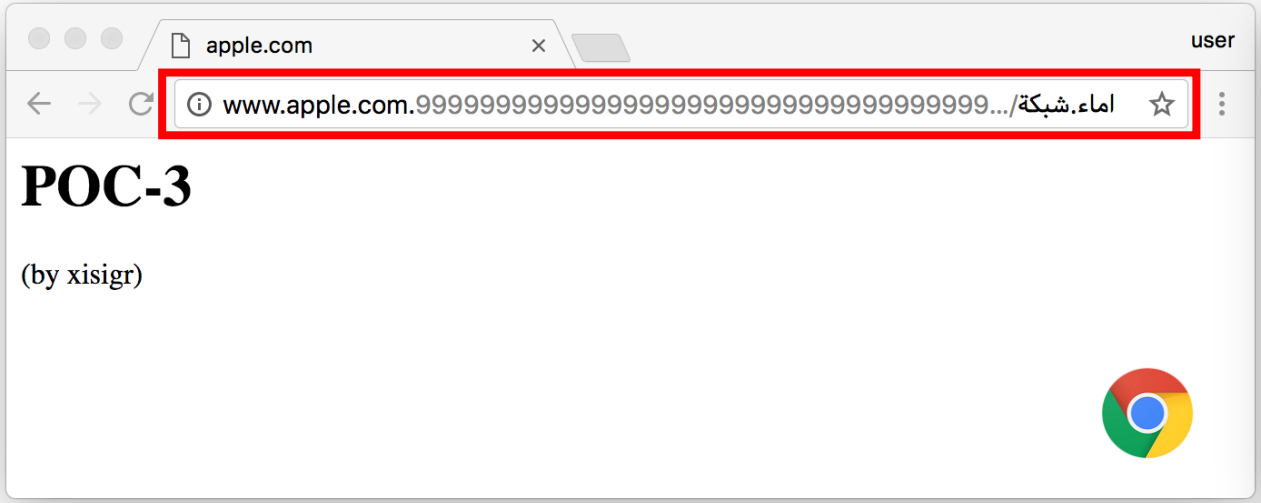
http://www.apple.com.xn--ggbla3j.xn--ngbc5azd/999...html

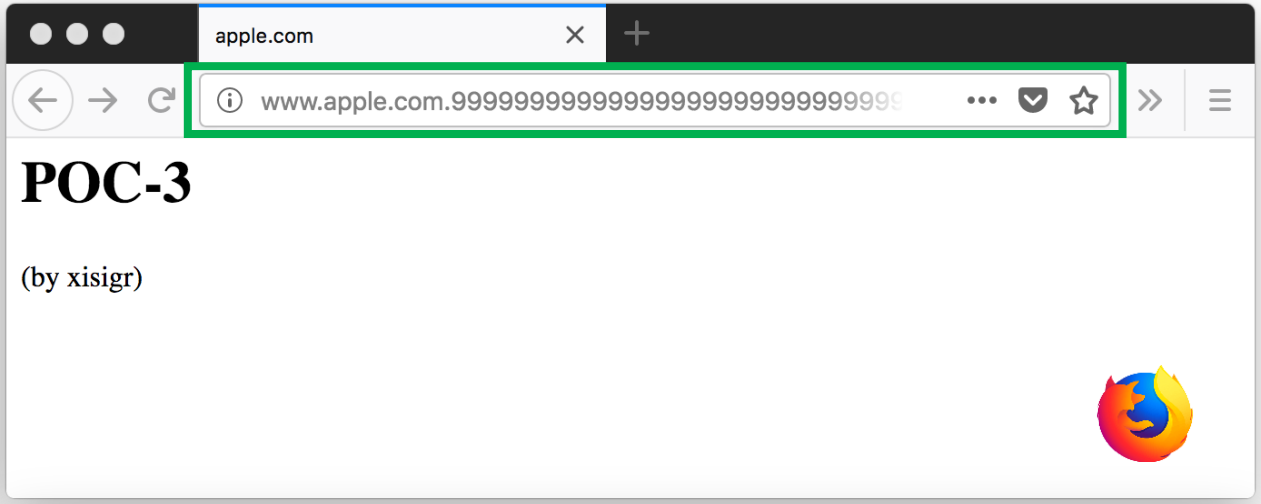
(345)-LD: www.apple.com
Direction: LTR

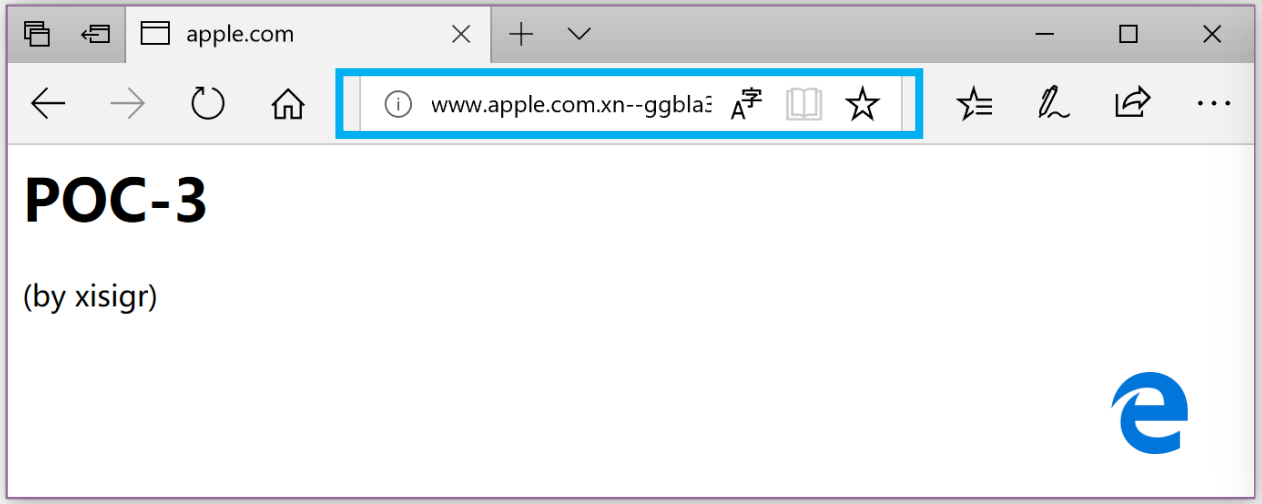
SLD：اماء
Punycode: xn--ggbla3j
Direction: RTL

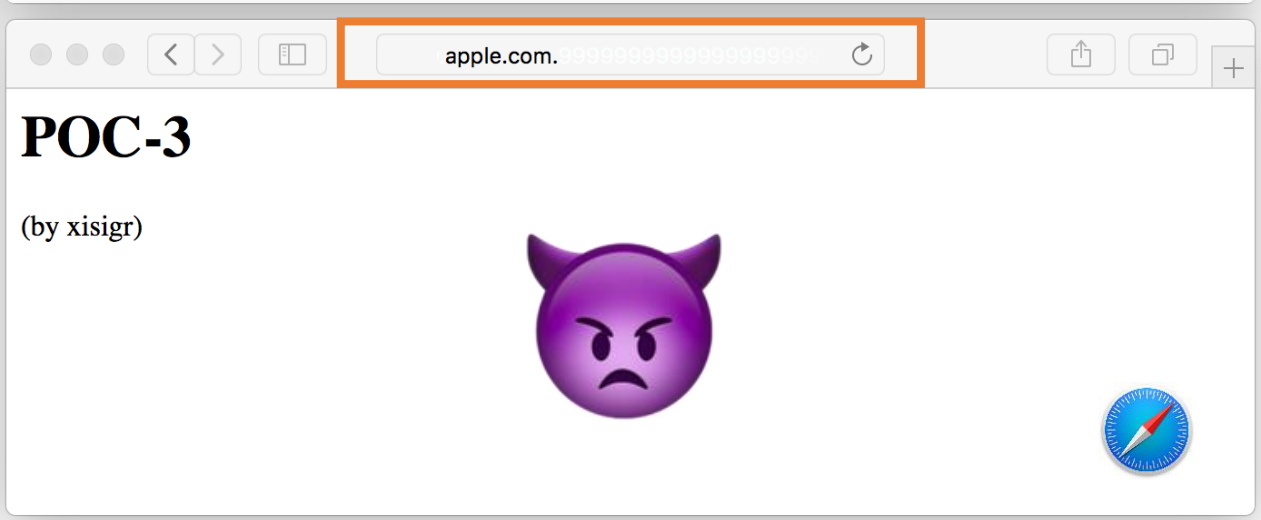
Pathname: 999999999999999999999999999999999999999999999999.html
Direction: LTR

POC-3

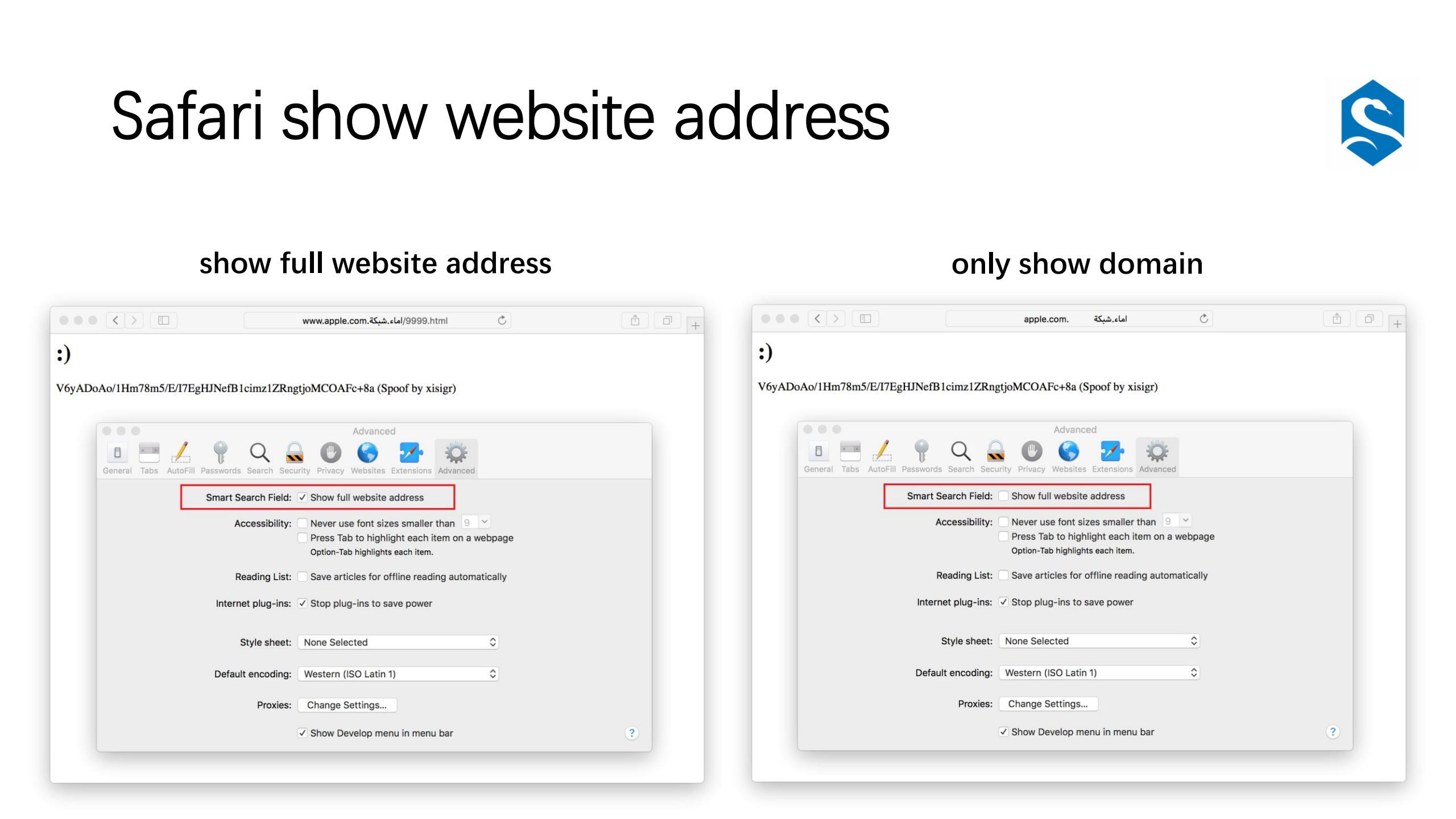

apple.com
user
www.apple.com.9999999999999999999999999999999999999999.../اماء.شبكة
POC-3
(by xisigr)


apple.com
www.apple.com.9999999999999999999999999999
POC-3
(by xisigr)


apple.com
www.apple.com.xn--ggbla
POC-3
(by xisigr)


apple.com.
POC-3
(by xisigr)

Safari show website address



show full website address

www.apple.com.اماء.شبكة/9999.html

:)

V6yADoAo/1Hm78m5/E/I7EgHJNefB1cimz1ZRngtjoMCOAFc+8a (Spoof by xisigr)

Advanced

General Tabs AutoFill Passwords Search Security Privacy Websites Extensions Advanced

Smart Search Field: ✓ Show full website address

Accessibility: Never use font sizes smaller than 9
Press Tab to highlight each item on a webpage
Option-Tab highlights each item.

Reading List: Save articles for offline reading automatically

Internet plug-ins: ✓ Stop plug-ins to save power

Style sheet: None Selected

Default encoding: Western (ISO Latin 1)

Proxies: Change Settings...

✓ Show Develop menu in menu bar

only show domain

apple.com. اماء.شبكة

:)

V6yADoAo/1Hm78m5/E/I7EgHJNefB1cimz1ZRngtjoMCOAFc+8a (Spoof by xisigr)

Advanced

General Tabs AutoFill Passwords Search Security Privacy Websites Extensions Advanced

Smart Search Field: Show full website address

Accessibility: Never use font sizes smaller than 9
Press Tab to highlight each item on a webpage
Option-Tab highlights each item.

Reading List: Save articles for offline reading automatically

Internet plug-ins: ✓ Stop plug-ins to save power

Style sheet: None Selected

Default encoding: Western (ISO Latin 1)

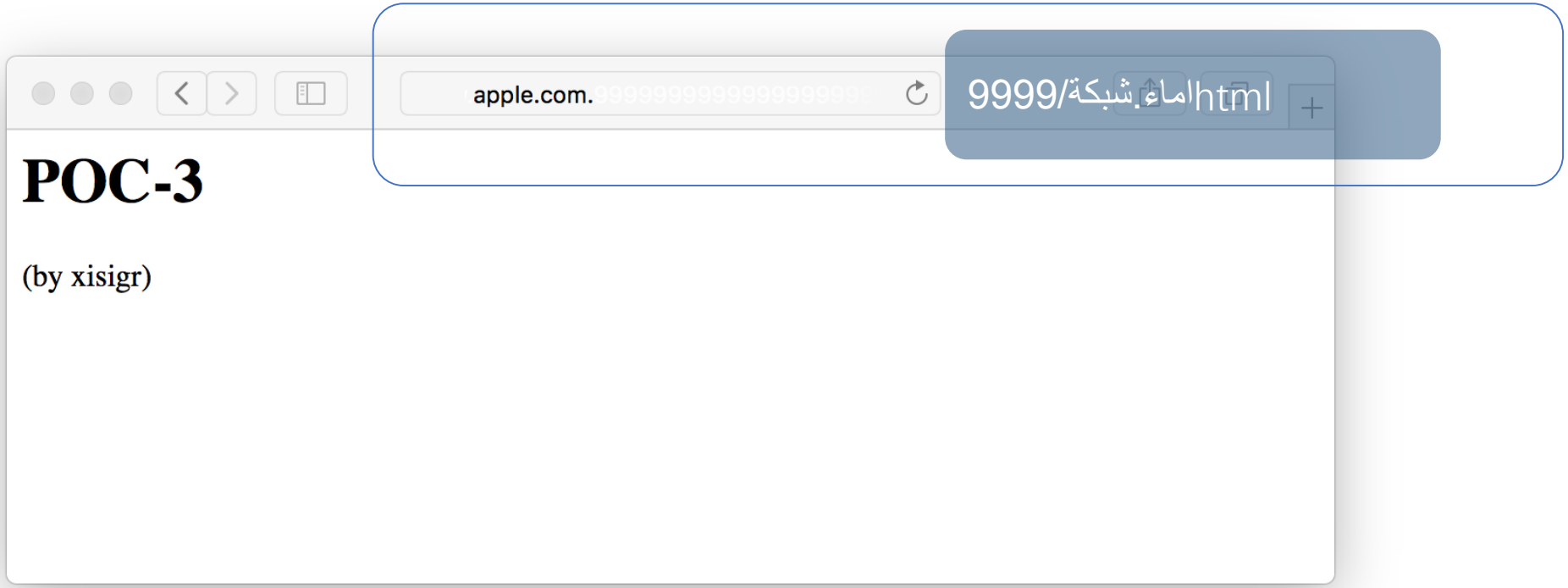
Proxies: Change Settings...

✓ Show Develop menu in menu bar

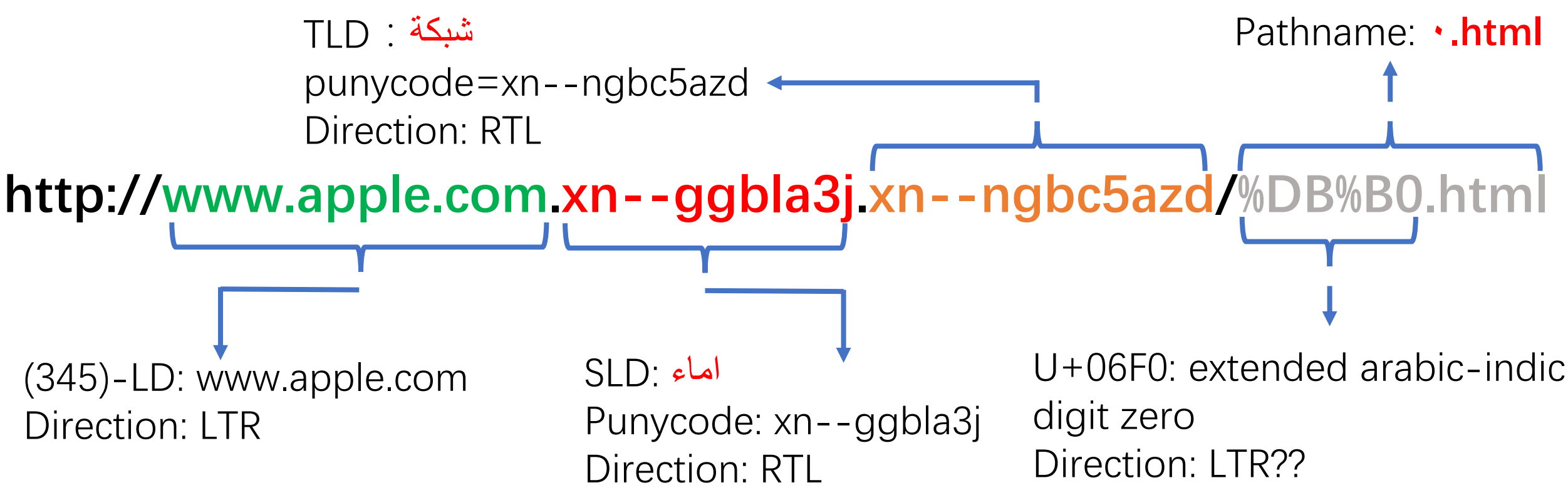
only show domain

http://www.apple.com.xn--ggbla3j.xn--ngbc5azd/999...html

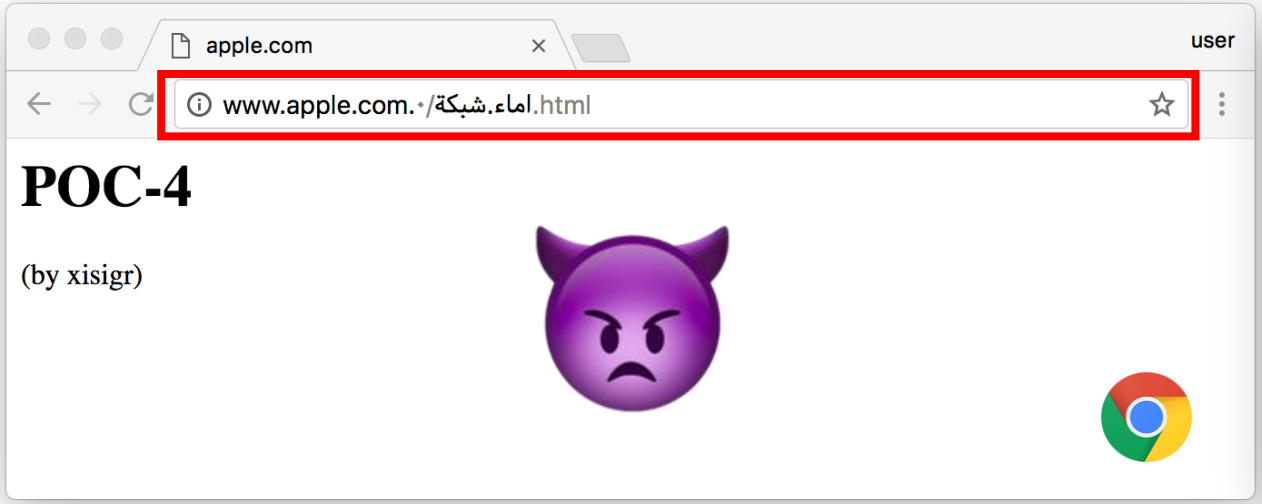
http://www.apple.com.9999/شبكة.اماء.html

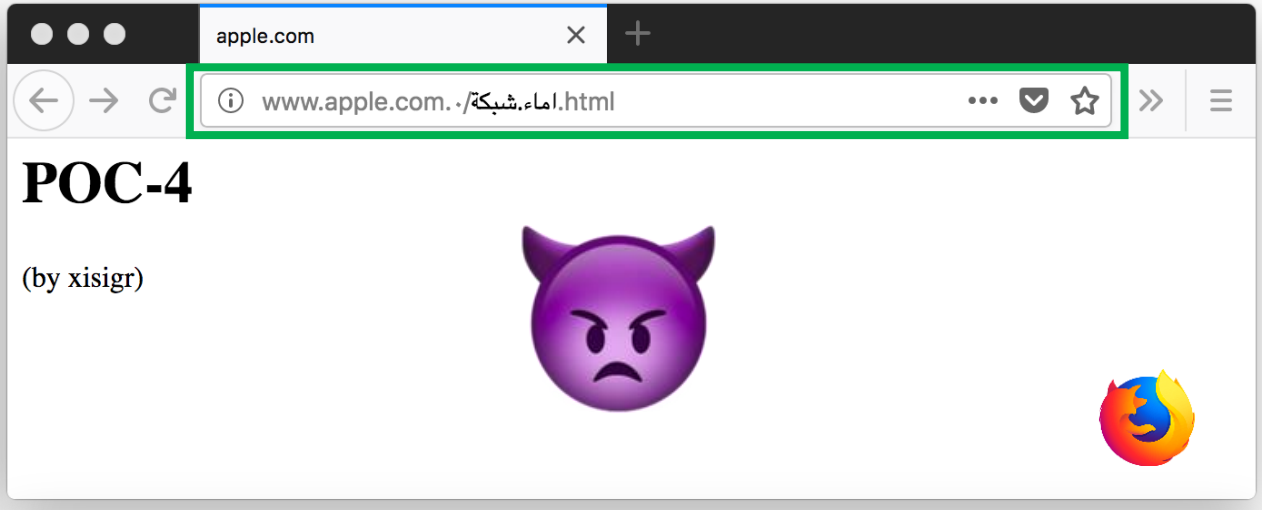


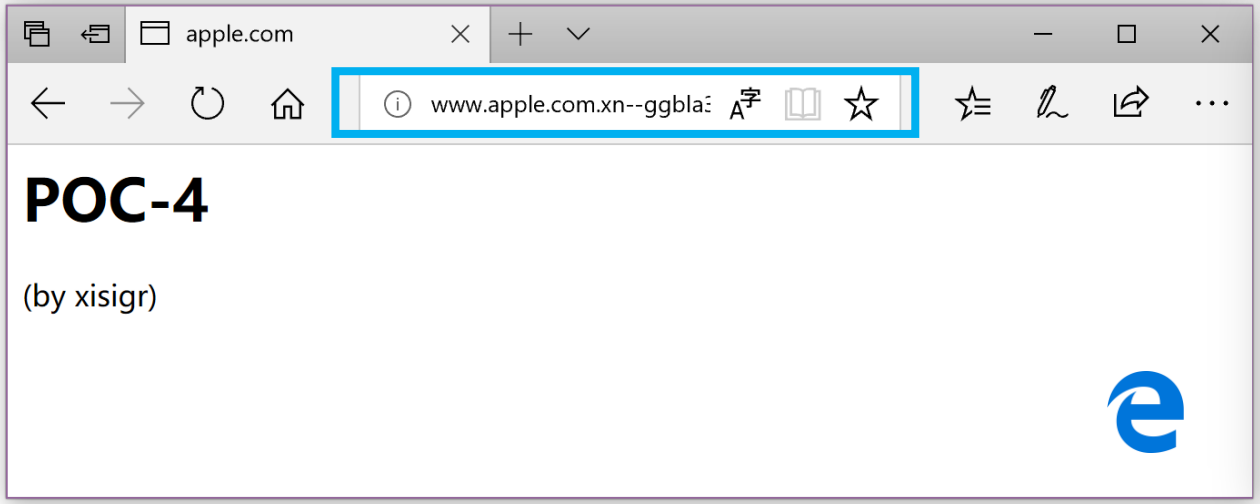
POC-4

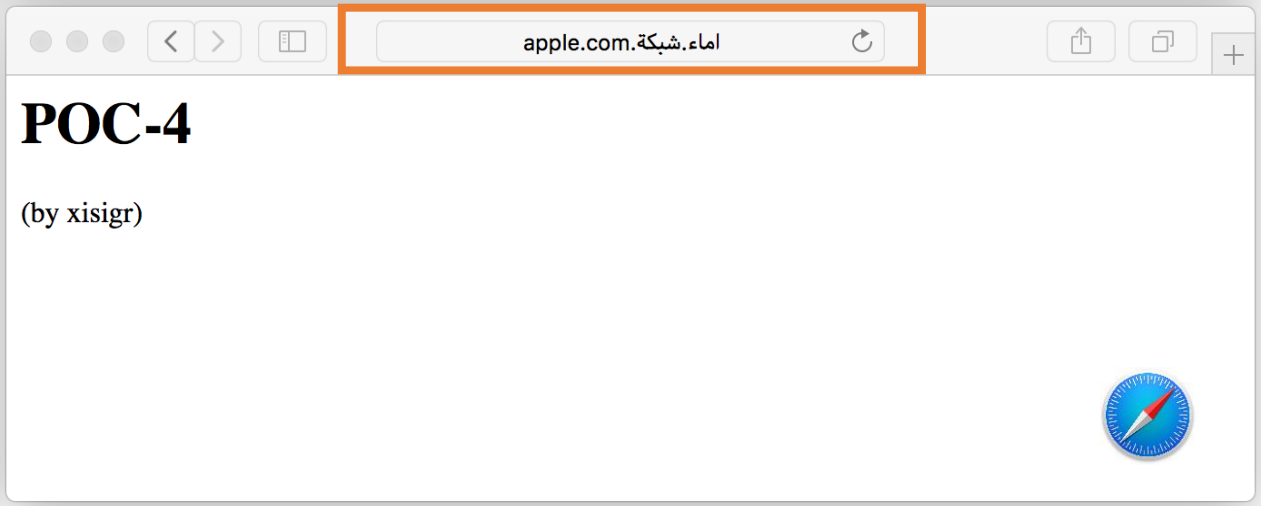

TLD：شبكة
punycode=xn--ngbc5azd
Direction: RTL
Pathname: ٠.html
http://www.apple.com.xn--ggbla3j.xn--ngbc5azd/%DB%B0.html
(345)-LD: www.apple.com
Direction: LTR
SLD: اماء
Punycode: xn--ggbla3j
Direction: RTL
U+06F0: extended arabic-indic digit zero
Direction: LTR??

POC-4


apple.com
user
www.apple.com.٠/اماء.شبكة.html
POC-4
(by xisigr)


apple.com
www.apple.com.٠/اماء.شبكة.html
POC-4
(by xisigr)

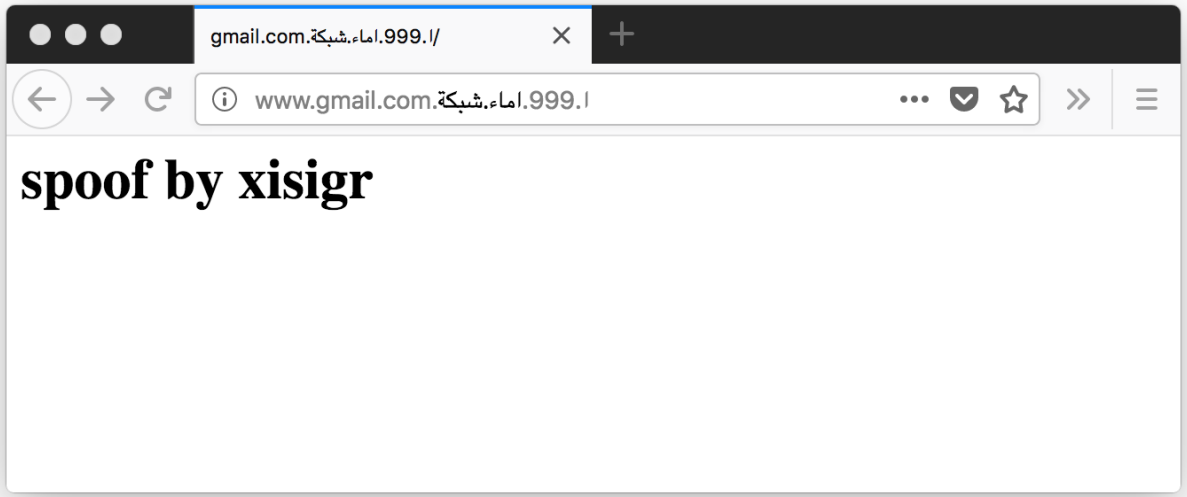

apple.com
www.apple.com.xn--ggbla
POC-4
(by xisigr)


apple.com.اماء.شبكة
POC-4
(by xisigr)

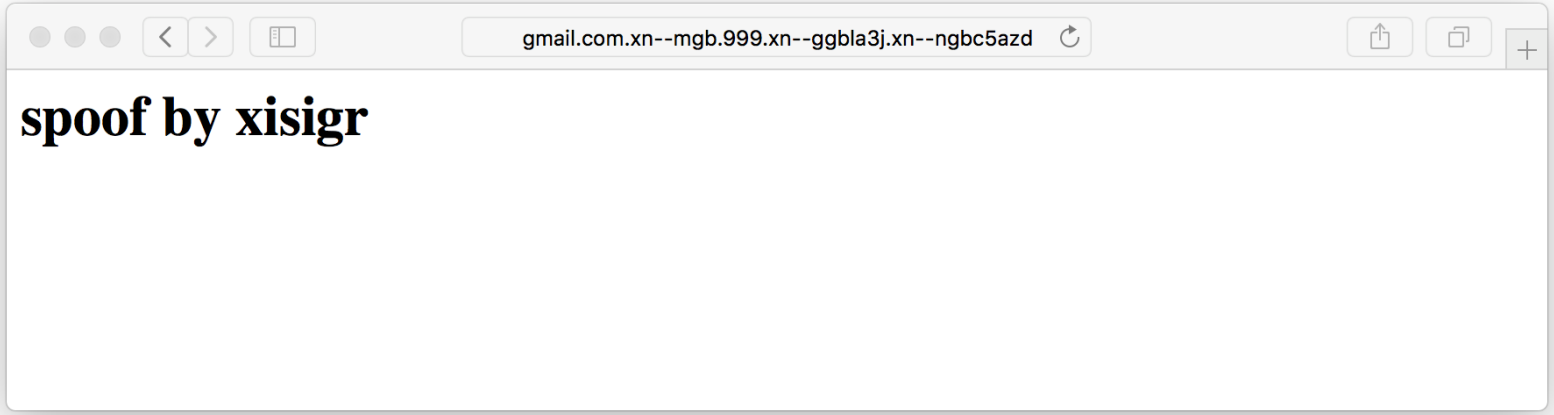
Origin out-of-order

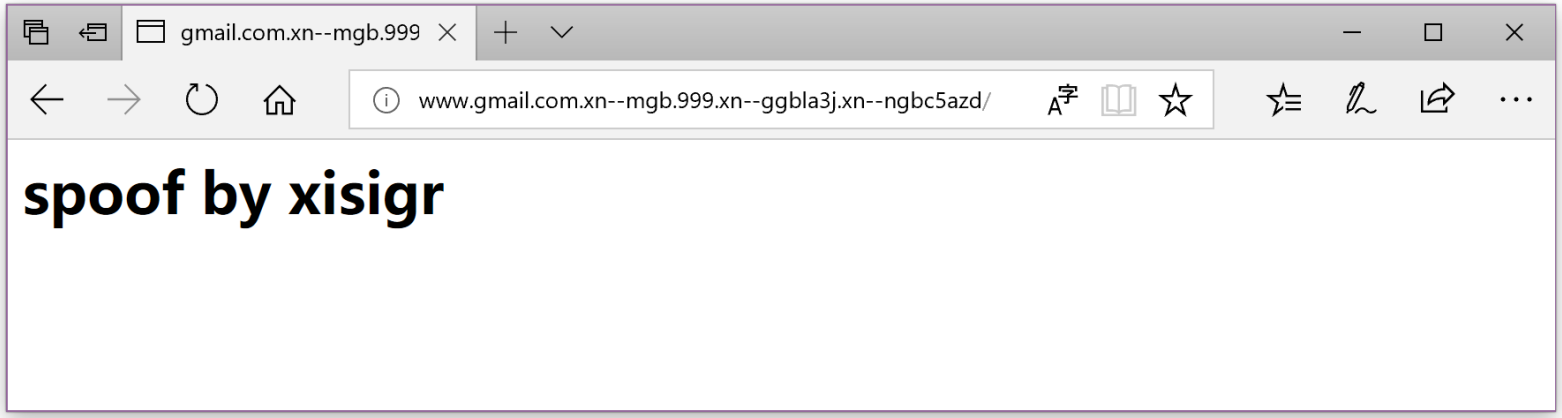
www.gmail.com.xn--mgb.999.xn--ggbla3j.xn--ngbc5azd

Firefox/Chrome



Safari/Edge


gmail.com.xn--mgb.999.xn--ggbla3j.xn--ngbc5azd
spoof by xisigr


gmail.com.xn--mgb.999
www.gmail.com.xn--mgb.999.xn--ggbla3j.xn--ngbc5azd/
spoof by xisigr

Think about it

- How to resolve RTL Spoof in address bar?
- Only show origin?
  - showing only the origin reduces utility, though. That's for the UX designers to decide.
  - Either way, it doesn't fully resolve these issues. You can still do RTL spoofs with just the origin (like having the labels shown out-of-order).
  - Some error logic.
    - Eg: "www" or "m" subdomain is removed
    - https://bugs.chromium.org/p/chromium/issues/detail?id=881694
- Spoofable RTL URLs in the UI
    - https://bugs.chromium.org/p/chromium/issues/detail?id=351639

Combining character

Safari Address Bar Spoof Using Combining character
CVE-2018-4260

https://support.apple.com/en-us/HT208938

Hebrew

	059	05A	05B	05C	05D	05E	05F
0		05A0	05B0	05C0	05D0	05E0	05F0
1	0591	05A1	05B1	05C1	05D1	05E1	05F1
2	0592	05A2	05B2	05C2	05D2	05E2	05F2
3	0593	05A3	05B3	05C3	05D3	05E3	05F3
4	0594	05A4	05B4	05C4	05D4	05E4	05F4
5	0595	05A5	05B5	05C5	05D5	05E5	
6	0596	05A6	05B6	05C6	05D6	05E6	
7	0597	05A7	05B7	05C7	05D7	05E7	
8	0598	05A8	05B8		05D8	05E8	
9	0599	05A9	05B9		05D9	05E9	
A	059A	05AA	05BA		05DA	05EA	
B	059B	05AB	05BB		05DB		
C	059C	05AC	05BC		05DC		
D	059D	05AD	05BD		05DD		
E	059E	05AE	05BE		05DE		
F	059F	05AF	05BF		05DF		

ו hebrew letter vav (U+05D5)

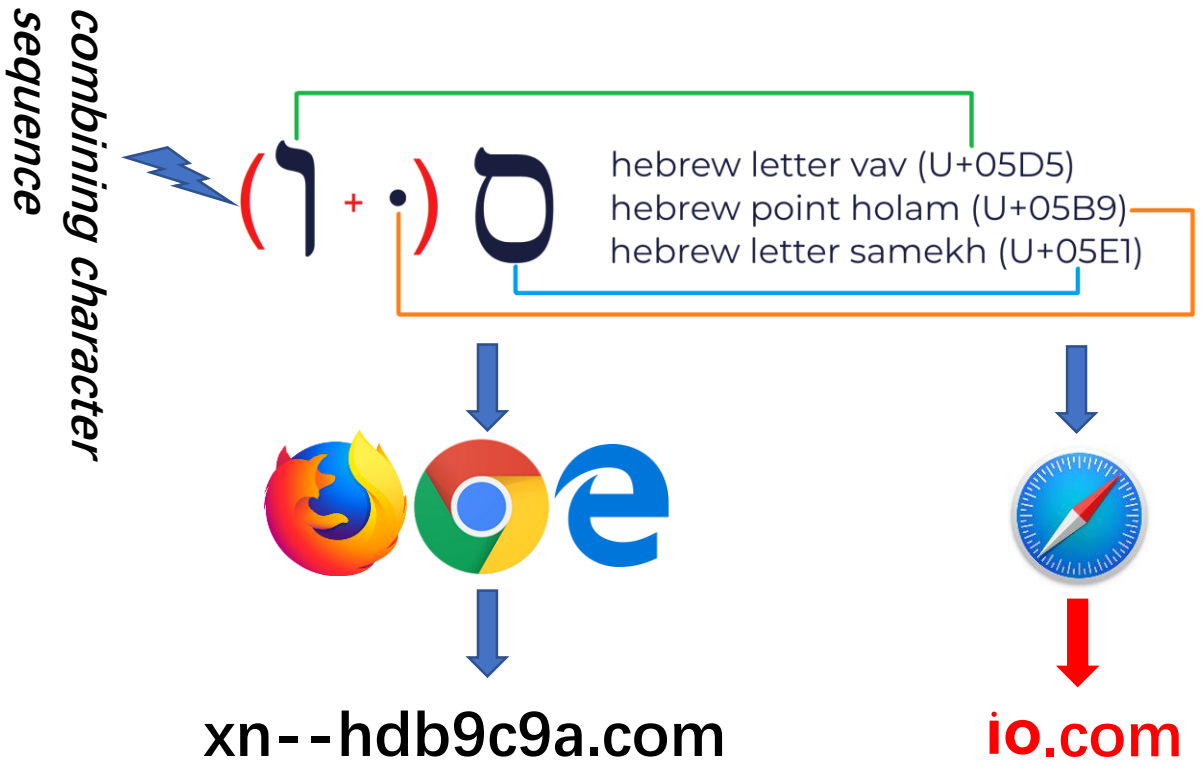
• hebrew point holam (U+05B9)

ס hebrew letter samekh (U+05E1)

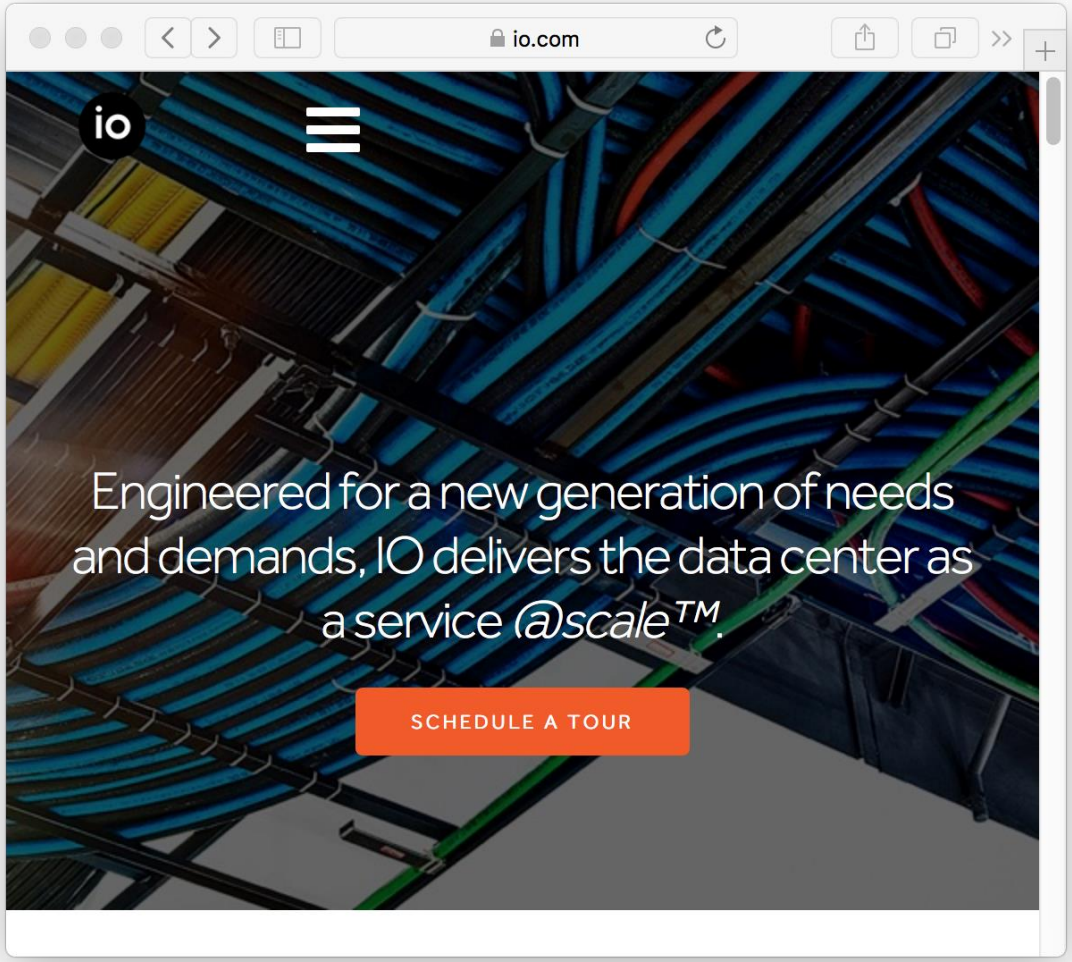
combining character sequence

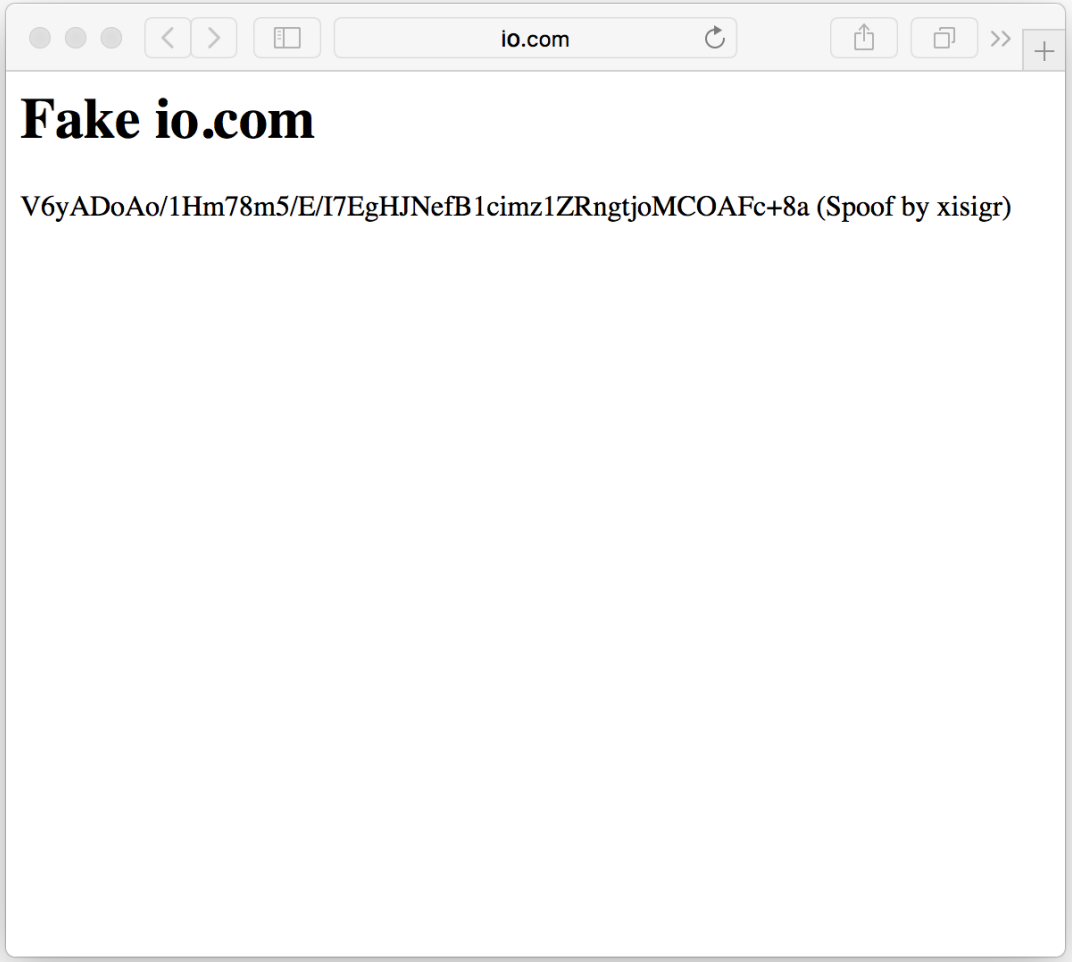
- Character + combining mark = ***combining character sequence***

Latin: io.com VS Hebrew: ׅוס.com

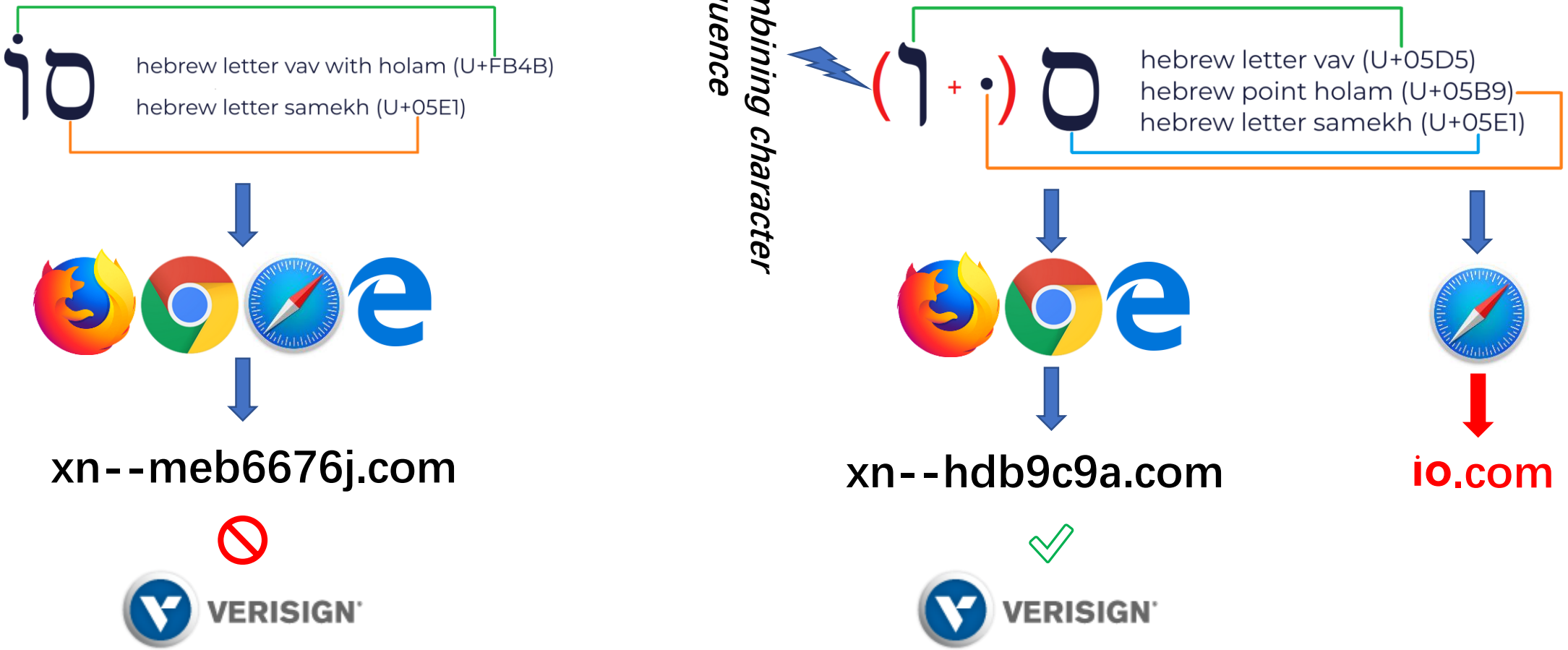


Latin: io.com VS Hebrew: ıo.com

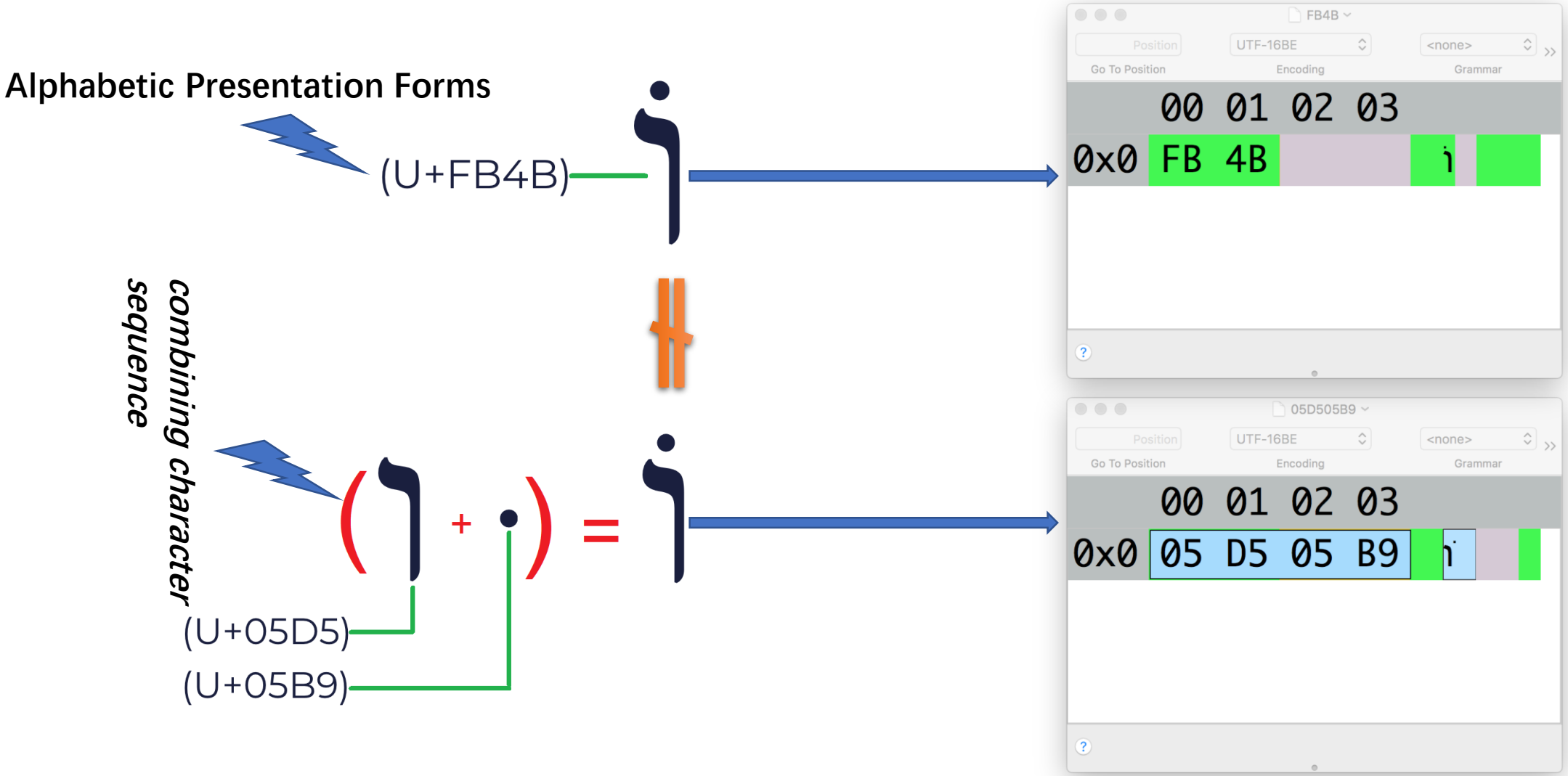




Hebrew alphabet: ׀ופ.com VS Hebrew: ׀ופ.com



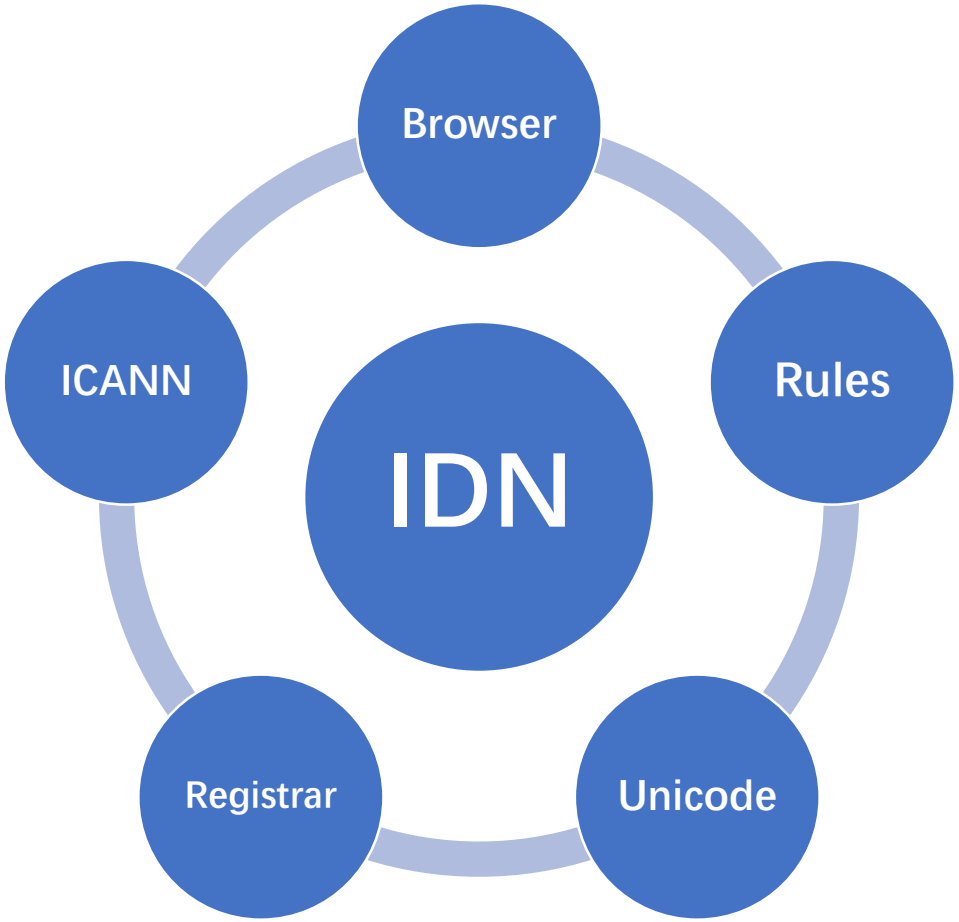
grapheme is same but meaning is difference



Think about it

- Should combination character sequence be supported to register?
- Is it possible to strictly check and prevent such malicious intentions at the Registrar stage?
- Think about these questions in the URL, you may find something new.
  - http://unicode.org/faq/char_combmark.html.

IDN Challenging




thinking of you