Deep Confusables

Mejorando la detección de ataques basados en codificación Unicode.

Dr. Alfonso Muñoz José Ignacio Escribano **Miguel Hernández**

Security Lab – BBVA Next Technologies

Valencia, 27 de noviembre de 2019

Índice

Introducción

- Unicode: estándar de codificación de caracteres.
- Organizado en bloques.
- La última versión es la 12.0, con 137 928 caracteres.
- **Confusables**: caracteres visualmente similares a otros. Fácilmente confundibles con otros caracteres.
  - Unicode proporciona una lista de confusables.
- Problema de seguridad recurrente: phishing.
- Medidas de seguridad: `http://www.unicode.org/reports/tr39/tr39-19.html`.
  - Punycode: definido en RFC3492. Representación de Unicode con el subconjunto de caracteres ASCII usado para representar IDNA.

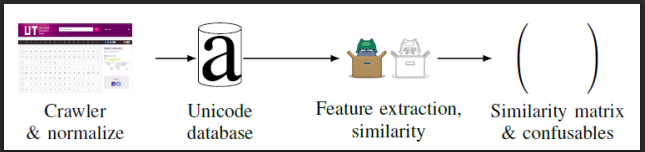
Trabajo relacionado

- Adrian Crenshaw. "*Out of Character: Use of Punycode and Homoglyph Attacks to Obfuscate URLs for Phishing*". Irongeek (2017).
- The Tarquin. "*Weaponizing Unicode Homographs Beyond IDNs*" . DEFCON 26 (2018).

- EvilURL: `https://github.com/UndeadSec/EvilURL`.
- Squatm3: `https://github.com/david3107/squatm3`
- Samesame: `https://github.com/TheTarquin/samesame`

Deep Confusables

- Sistema de generación de *confusables* usando **deep learning**.
- **Objetivo**: mejorar la lista de *confusables* proporcionada por Unicode Consortium.
- Compuesto de 3 componentes:
  - Base de datos de imágenes con caracteres Unicode.
  - Extractor de características y comparador de similitud.
  - Línea de comandos.



Base de datos de imágenes de caracteres Unicode

- Imágenes extraídas de `https://unicode-table.com`.
- Imágenes normalizadas a 34x34 píxeles.
- 38 800 imágenes de 266 bloques.

`https://github.com/next-security-lab/unicode-images-database/releases`

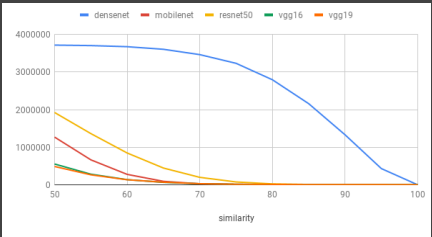
Extractor de características y comparador de similitud

1. Matriz de similaridad de cada caracter Unicode.
   - Para cada caracter Latín y no Latín:
     - Extraer *features* usando un modelo preentrenado.
     - Comparar *features* con función de similaridad.
2. Obtener *confusables*.
   - Fijar umbral θ entre $0 \leq \theta \leq 1$.
   - Obtener caracteres cuya similaridad es mayor que θ.

`https://github.com/next-security-lab/deep-confusables-similarity/releases`

Extractor de características y comparador de similitud – Modelos testados

- **VGG16**
- VGG19
- ResNet 50
- DenseNet
- MobileNet



Línea de comandos

- Crea dominios usando confusables, fijando un umbral.
- Sólo soporta caracteres Latin-1.
- Algunas características.
  - Comprueba si los dominios están activos.
  - Comprueba Whois.
  - Comprueba dominio en la API de VirusTotal.

`https://github.com/next-security-lab/deep-confusables-cli`

Diccionario de *confusables*

- Combinación de otros diccionarios de otras herramientas y nuestros confusables con umbral $0{,}75$.

`https://github.com/next-security-lab/deep-confusables-cli/blob/master/deep_confusables_lite/confusables.txt`

UriDeep

- Basado en el diccionario de confusables.
- Incluye funcionalidad adicional.
  - Substitution attack.
  - Flipping attack.
- Análisis de dominios del Top 10000 de Alexa, PYMEs españolas y del IBEX 35.

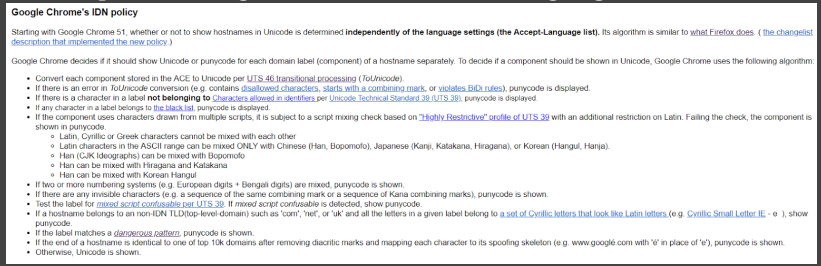
https://github.com/mindcrypt/uriDeep
https://github.com/mindcrypt/uriDeep/blob/master/data/deepDiccConfusables.txt

Algunos ejemplos

amazón.es, góogle.es skypê.net, skýpe.net, skÿpe.net, skypè.net, skypé.net, fàcebook.net, fâcebook.net, facêbook.net, facëbook.net, mınecraft.net, twīt-ter.com, t-mobìle.com, alıexpress.com, applē.com, îkea.com, braźzers.com, īns-tagram.com, netflìx.com, faceb̲ook.com, the uardian.com, ebáy.com, amerıca-nexpress.com, adïdas.com, sėx.com, whatsàpp.com, àirbnb.com, nytímes.com, baīdu.com, office.com, mìcrosoft.com, wikipédia.com, disneylandpaṛis.com, xviďeos.com, amazoŋ.com, goog e.com.ph, microsoft.com, dropḅox.com, ýou-porn.com, vodafoṇe.com, icƚoud.com, poŕnhub.com, netfl'ix.com ...

Evasión de Punycode

- Política de IDN de Google Chrome: `https://www.chromium.org/developers/design-documents/idn-in-google-chrome`



Google Chrome's IDN policy

Starting with Google Chrome 51, whether or not to show hostnames in Unicode is determined **independently of the language settings (the Accept-Language list).** Its algorithm is similar to what Firefox does. (the changelist description that implemented the new policy)

Google Chrome decides if it should show Unicode or punycode for each domain label (component) of a hostname separately. To decide if a component should be shown in Unicode, Google Chrome uses the following algorithm:

- Convert each component stored in the ACE to Unicode per UTS 46 transitional processing (*ToUnicode*).
- If there is an error in *ToUnicode* conversion (e.g. contains disallowed characters, starts with a combining mark, or violates BiDi rules), punycode is displayed.
- If there is a character in a label **not belonging to** Characters allowed in identifiers per Unicode Technical Standard 39 (UTS 39), punycode is displayed.
- If any character in a label belongs to the black list, punycode is displayed.
- If the component uses characters drawn from multiple scripts, it is subject to a script mixing check based on "Highly Restrictive" profile of UTS 39 with an additional restriction on Latin. Failing the check, the component is shown in punycode.
  - Latin, Cyrillic or Greek characters cannot be mixed with each other.
  - Latin characters in the ASCII range can be mixed ONLY with Chinese (Han, Bopomofo), Japanese (Kanji, Katakana, Hiragana), or Korean (Hangul, Hanja).
  - Han (CJK Ideographs) can be mixed with Bopomofo
  - Han can be mixed with Hiragana and Katakana
  - Han can be mixed with Korean Hangul
- If two or more numbering systems (e.g. European digits + Bengali digits) are mixed, punycode is shown.
- If there are any invisible characters (e.g. a sequence of the same combining mark or a sequence of Kana combining marks), punycode is shown.
- Test the label for *mixed script confusable* per UTS 39. If *mixed script confusable* is detected, show punycode.
- If a hostname belongs to an non-IDN TLD(top-level-domain) such as 'com', 'net', or 'uk' and all the letters in a given label belong to a set of Cyrillic letters that look like Latin letters (e.g. Cyrillic Small Letter IE - е), show punycode.
- If the label matches a *dangerous pattern*, punycode is shown.
- If the end of a hostname is identical to one of top 10k domains after removing diacritic marks and mapping each character to its spoofing skeleton (e.g. www.google.com with 'é' in place of 'e'), punycode is shown.
- Otherwise, Unicode is shown.

- Caracteres potenciales: ą ę ǵ ġ į ķ ļ ņ ŗ ş ţ ų

Problemas en distintas aplicaciones

Software	Problema	Respuesta del proveedor
Skype Desktop[a]	No se provee información para detectar un dominio falso	Reconoce el error y trabajan para solucionarlo
Foxit Reader[b]	No se provee información para detectar un dominio falso	Reconoce el error y trabajan para solucionarlo
Telegram	No se provee información para detectar un dominio falso	Reconoce el error y trabajan para solucionarlo
ProtonMail	No convierte el dominio o deshabilita el enlace	Sin respuesta
LinkedIn	No convierte el dominio o deshabilita el enlace en artículos	Sin respuesta
Redes sociales libres	No convierte el dominio o deshabilita el enlace	Sin respuesta
OpenOffice	No se provee información para detectar un dominio falso	Sin respuesta
Signal	No se provee información para detectar un dominio falso	Sin respuesta
WhatsApp	No se provee información para detectar un dominio falso	Sin solución. Considera un problema de UX o de ingeniería social
GMail	No convierte el dominio o deshabilita el enlace	Sin solución. Considera un problema de UX o de ingeniería social

[a] https://portal.msrc.microsoft.com/en-us/security-guidance/researcher-acknowledgments-online-services

[b] https://www.foxitsoftware.com/support/security-bulletins.php

Conclusiones

- Los *confusables* son caracteres visualmente similares a otros caracteres.
- El consorcio de Unicode provee una lista de *confusables*.
- Deep Confusables mejora la generación de *confusables* usando deep learning.
- Existen problemas con la codificación Unicode en distintas aplicaciones.

Deep Confusables

Mejorando la detección de ataques basados en codificación Unicode.

Dr. Alfonso Muñoz José Ignacio Escribano **Miguel Hernández**

Security Lab – BBVA Next Technologies

Valencia, 27 de noviembre de 2019