

Chen Lai, Zhongrong Jian, J. Sidrach

University of California San Diego

CSE 227: Computer Security - Spring 2017

# Internationalized Domain Name Homograph Attacks

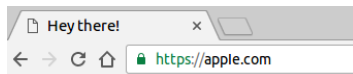
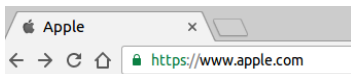
- ▶ Domains registered with Punycode (xn- prefix)
- ▶ Displayed in Unicode in browsers' navigation bar
- ▶ What could go wrong?
- ▶ *Some* of the Unicode homographs of a, b, c:

|                               |   |  |                                       |   |                                 |                                |  |
|-------------------------------|---|--|---------------------------------------|---|---------------------------------|--------------------------------|--|
| <b>a</b>                      | <b>α</b>                                  | <b>α</b>                                   | <b>а</b>                              | <b>α</b>                                    | <b>a</b>                        | <i>a</i>                       | <b><i>a</i></b>                        |
| LATIN<br>SMALL<br>LETTER<br>A | LATIN<br>SMALL<br>LETTER<br>ALPHA         | GREEK<br>SMALL<br>LETTER<br>ALPHA          | CYRILLIC<br>SMALL<br>LETTER A         | APL<br>FUNCTIONAL<br>SYMBOL<br>ALPHA        | MATHEMATICAL<br>BOLD SMALL A    | MATHEMATICAL<br>ITALIC SMALL A | MATHEMATICAL<br>BOLD ITALIC<br>SMALL A |
| <b>b</b>                      | <b>Ḃ</b>                                  | <b>Ბ</b>                                   | <b>Ბ</b>                              | <b>b</b>                                    | <b>b</b>                        | <i>b</i>                       | <b><i>b</i></b>                        |
| LATIN<br>SMALL<br>LETTER<br>B | LATIN<br>CAPITAL<br>LETTER<br>TONE<br>SIX | CYRILLIC<br>CAPITAL<br>LETTER<br>SOFT SIGN | CHEROKEE<br>LETTER SI                 | CANADIAN<br>SYLLABICS<br>AIVILIK B          | MATHEMATICAL<br>BOLD SMALL B    | MATHEMATICAL<br>ITALIC SMALL B | MATHEMATICAL<br>BOLD ITALIC<br>SMALL B |
| <b>с</b>                      | <b>с</b>                                  | <b>с</b>                                   | <b>с</b>                              | <b>с</b>                                    | <b>с</b>                        | <b>с</b>                       | <i>с</i>                               |
| LATIN<br>SMALL<br>LETTER<br>C | GREEK<br>LUNATE<br>SIGMA<br>SYMBOL        | CYRILLIC<br>SMALL<br>LETTER ES             | LATIN<br>LETTER<br>SMALL<br>CAPITAL C | SMALL<br>ROMAN<br>NUMERAL<br>ONE<br>HUNDRED | DESERET<br>SMALL LETTER<br>CHEE | MATHEMATICAL<br>BOLD SMALL C   | MATHEMATICAL<br>ITALIC SMALL C         |

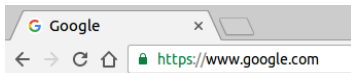
(Image provided by Google)

# Defense Mechanisms - Browsers

- ▶ URLs displayed in Punycode if certain checks fail
- ▶ No common policy across major browsers
- ▶ Successful attack



- ▶ Unsuccessful attack



# Defense Mechanisms - ICANN and TLD Registrars

## ICANN

- ▶ Rejects ccTLD applications that look similar to existing ones
- ▶ Does not enforce restrictions to second-level domains

## TLD Registrars

- ▶ No common public policy to deal with homograph domains
- ▶ Notable exception: Chinese TLDs (simplified/traditional)

# Methodology - Clustering

## Data sources

- ▶ *.com* zone snapshot
- ▶ Alexa Top 1 million web sites ranking

## Clustering Process

- ▶ Filter IDNs from snapshot
- ▶ Filter non-IDNs from Alexa ranking
- ▶ Cluster all homograph IDNs, using a non-IDN homograph domain from the Alexa ranking as the representative

# Methodology - Classification

Manual classification using

- ▶ WHOIS records
- ▶ HTTP/HTML responses

Categories

- ▶ Canonical - Parking
- ▶ Canonical - Redirect
- ▶ Third Party - Redirect to Canonical
- ▶ Third Party - Unrelated
- ▶ Third Party - Parking
- ▶ Third Party - Scam

## Results (1)

| Domains                               | #              | %             |
|---------------------------------------|----------------|---------------|
| <i>Canonical domain names</i>         | <i>458731</i>  | <i>8.31%</i>  |
| With IDN homographs                   | 825            | 6.04%         |
| Without IDN homographs                | 457906         | 2.27%         |
| <i>Internationalized Domain Names</i> | <i>1045400</i> | <i>91.69%</i> |
| With canonical homograph              | 1099           | 3.68%         |
| Without canonical homograph           | 1044301        | 2.74%         |

Table 1: Overview of the clustering results.

## Results (2)

| Domain       | # of IDN homographs |
|--------------|---------------------|
| google.com   | 24                  |
| youtube.com  | 3                   |
| facebook.com | 9                   |
| baidu.com    | 3                   |
| yahoo.com    | 4                   |
| reddit.com   | 1                   |
| qq.com       | 2                   |
| taobao.com   | 1                   |
| live.com     | 1                   |
| vk.com       | 6                   |

Table 2: Top ten .com domains in the Alexa ranking with IDN homographs.



## Results (3)

| Status                | #   | %      |
|-----------------------|-----|--------|
| <i>Canonical</i>      | 88  | 8.31%  |
| Parking               | 64  | 6.04%  |
| Redirect              | 24  | 2.27%  |
| <i>Third Party</i>    | 971 | 91.69% |
| Redirect to Canonical | 39  | 3.68%  |
| Unrelated             | 29  | 2.74%  |
| Parking               | 872 | 82.34% |
| Scam                  | 31  | 2.93%  |

Table 3: Breakdown of the manually classified homograph IDNs.

## Results (4)

| Registrant organization              | Registrant email           | # of Homograph IDNs |
|--------------------------------------|----------------------------|---------------------|
| Domains By Proxy, LLC                | –                          | 89                  |
| Super Privacy Service c/o Dynadot    | privacy@dynadot.com        | 23                  |
| Domain Registries Foundation         | –                          | 22                  |
| Duong Thien                          | thiendv@outlook.com        | 18                  |
| Syngenuity Limited                   | manager@syngenuity.com     | 12                  |
| Helpnet: Brand Development & Sales   | help@strongestbrands.com   | 12                  |
| ONUNO L.L.C.                         | corucas@gmail.com          | 11                  |
| Privacy Protection Service INC d/b/a | contact@privacyprotect.org | 10                  |
| Hubertus Henz                        | hu.h5@yahoo.de             | 9                   |
| wuyu                                 | wy65535@126.com            | 7                   |

Table 4: Top ten registrants with the most homograph IDNs.

# Conclusions

## Current state

- ▶ No common TLD policies in place
- ▶ 1000+ homograph IDNs detected
- ▶ Most domains inactives (parking)
- ▶ No guarantees they will stay parked in the future

## Future work

- ▶ Analyze other TLDs
- ▶ Improve homograph matching algorithm (OCR?)
- ▶ Automatic classification of homograph IDNs

Q & A