SaudiNIC:

Securing Arabic Domain Names

Raed Alfayez, SaudiNIC
APTLD74, Tashkent, Sep 2018
Agenda

➢ About SaudiNIC
➢ About Arabic Script/Language
➢ Securing Arabic domains
➢ What's Missing?
About SaudiNIC

- **Administering** the domain name space under:
  - (.sa) since 1995
  - (السعودية) Experimental since 2004, official since 2010.

- Operated by a government organization:
  - CITC (Communication and Information Technology Commission)

- Leading the local and regional communities efforts towards supporting **Arabic language** in Domain Names since **2001** (more than 15 years of experience)

- Participated in several **TF & WG** related to Arabic domains regionally and globally

- Developed: **Tools, Pilot Projects, Reports & RFC** related to Arabic domains
About SaudiNIC

55,850 Saudi domain names
2,540 Arabic domains under (.السعودية)
466 Variant

2LD/3LD Domain Names Distribution %

- .com.sa: 53%
- .sa: 28%
- .pub.sa: 0.28%
- .med.sa: 0.75%
- .sch.sa: 1.3%
- .edu.sa: 2.6%
- .net.sa: 2.9%
- .org.sa: 3.3%
- .gov.sa: 3.6%
- السعودية: 4.5%
Introduction: Arabic Script

• The 2\textsuperscript{nd} most widely used alphabetic writing system in the world
• Used by many languages such as:
  – Arabic, Urdu, Persian, Turkish, Kurdish, Pashto, ... etc
• It is widely used by more than 43 countries
  – more than one billion potential users could be concerned in using Arabic script domain names.

Source: http://en.wikipedia.org/wiki/Arabic_script
Introduction: Arabic Language

• Ranked as the 5\textsuperscript{nd} language by native speakers in the world.

• Considered as Official/Co-official language in 25 country, \(~\text{295 million}\) native speakers.

Source: https://en.wikipedia.org/wiki/Arabic
**Introduction: Arabic Language / Script**

- The Arabic **language** is a subset of the Arabic **script**
  
  **Script**: 0600..06FF, 0750..077F, 08A0..08FF  
  **Language**: 0621..063A, 0641..064A, 0660..0669

- **Table of Arabic Script and Language**

<table>
<thead>
<tr>
<th>Script</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>0600..06FF</td>
<td>0621..063A</td>
</tr>
<tr>
<td>0750..077F</td>
<td>0641..064A</td>
</tr>
<tr>
<td>08A0..08FF</td>
<td>0660..0669</td>
</tr>
</tbody>
</table>

- **Diagram of Arabic Alphabets and Script**
Securing Arabic Domains

1. Identifying allowed code points
   - Identifying allowed code points at language level and/or script level
   - Not allowing unsecure/unsafe code points that may effect the stability of the registry domain space (e.g. control code points, combining marks, diacritics, Tatweel ..etc).

2. Identifying variants
   - Variants within a language
   - Variants across the whole script
   - Reachability variants: So that a registered domain name can be accessed regardless of the input devices (language table) being used by the navigator users.

3. No script mixing
   - A label can be constructed only by Arabic code points and digits

4. No language mixing
   - (utilizing the powerful tools: Language tables)
     - control input via the user interface
     - tremendously reduce the number of unnecessary/unrealistic variants
     - Registry choice to protect the TLD-space.

5. No digits mixing
   - Simplify the need for 3 digits sets in Arabic script
     - European digits
     - Arabic-Indic digits
     - Eastern Arabic-Indic digits

6. WLE Rules (language/scripts)
   - Safeguard to handle some special cases (e.g. similarity in certain positions)
Securing Arabic Domains

(1) Identifying code points at both language and script level

Include only:

✓ Allowed code points at language level and/or script level

Not included:

- Combining Marks
- Diacritics (Tashkeel)
- Special codes (ZWNJ, ZWJ)
- Unused/Historical
Securing Arabic Domains

(2) Identifying variants (language level)
There are a number of groups of characters that have the same shapes (Homoglyph), eg.:

- Kaf group,
- Heh group,
- Yeh group,
- Alef group
- ...

Securing Arabic Domains

(2) Identifying variants (script level)
For reachability purposes, variants should be addressed to be activated by the registry, so that:

- A registered domain name is accessed regardless of the input devices (language table) being used by the navigator users.
- For example:
  - A user registered the domain “مكة” (all characters from the Arabic language)
  - If another user try to reach that domain name from an Internet café in Pakistan he/she will type “مکہ” (all characters from the Urdu language)
  - If the “activated” variants were not allocated, delegated and hosted then the domain name will not be reachable!

Hence, reachability issue (based on input devices used by other language communities) should be carefully considered when defining variants (by language communities).
Securing Arabic Domains

(3) No Script Mixing

Recommendation:

• No mixing within a label or cross labels in the same domain

<table>
<thead>
<tr>
<th>Format</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;FirstName&gt;&lt;Arabic-LastName&gt;</td>
<td>Salehفلاني</td>
</tr>
<tr>
<td>&lt;Arabic-LastName&gt;&lt;FirstName&gt;</td>
<td>الفلانيSaleh</td>
</tr>
<tr>
<td>&lt;TLD&gt;.&lt;Arabic-Domain&gt;</td>
<td>رسيلالسعودية</td>
</tr>
<tr>
<td>&lt;Domain&gt;.&lt;Arabic-TLD&gt;</td>
<td>raseelالسعودية</td>
</tr>
</tbody>
</table>
Securing Arabic Domains

(4) No language Mixing

– Not allowing code point that exists only in one language with another code points that exists only in another language in the same label
  – In the below example: ك & ﺲ cant exist in the same label,
  – same thing for: ﺖ & ﺪ & ﺲ
## Securing Arabic Domains

(4) No language Mixing

<table>
<thead>
<tr>
<th>IDN</th>
<th>Total Variants</th>
<th>Allocatable</th>
<th>Blocked</th>
<th>Blocked due to Language Mixing</th>
</tr>
</thead>
<tbody>
<tr>
<td>مكة-المكرمة</td>
<td>3239</td>
<td>34</td>
<td>3205</td>
<td>3181 (99.25%)</td>
</tr>
<tr>
<td>القرآن-الكريم</td>
<td>11999</td>
<td>111</td>
<td>11888</td>
<td>11836 (99.56%)</td>
</tr>
<tr>
<td>هيئة-الإعلام</td>
<td>47999</td>
<td>81</td>
<td>47918</td>
<td>47764 (99.68%)</td>
</tr>
<tr>
<td>كهف-الياسمين</td>
<td>28799</td>
<td>65</td>
<td>28734</td>
<td>28680 (99.81%)</td>
</tr>
<tr>
<td>كهف-اكيا</td>
<td>21599</td>
<td>47</td>
<td>21552</td>
<td>21534 (99.92%)</td>
</tr>
</tbody>
</table>
Securing Arabic Domains

(5) No Digit Mixing

- Digits in the Arabic scripts:
  - European digits: U+0030 .. U+0039 (0123456789)
  - Arabic-Indic digits: U+0660 .. U0669 (۰۱۲۳۴۵۶۷۸۹)
  - Eastern Arabic-Indic digits: U+06F0 .. U+06F9 (٠۱۲۳۴۵۶۷۸۹)

<table>
<thead>
<tr>
<th>#</th>
<th>Input string</th>
<th>Display</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>مؤتمر 2009 conference2009</td>
<td>acceptable: Pure European digits</td>
</tr>
<tr>
<td>2</td>
<td>مؤتمر ٢٠٠٩ conference٢٠٠٩</td>
<td>acceptable: Pure Arabic-Indic digits</td>
</tr>
<tr>
<td>3</td>
<td>مؤتمر ٢٠٠٩ conference٢٠٠٩</td>
<td>acceptable: Pure Eastern Arabic-Indic digits</td>
</tr>
<tr>
<td>4</td>
<td>مؤتمر ٩٠٠٢ conference٩٠٠٩</td>
<td>not-acceptable: Mix between European digits &amp; Arabic-Indic digits</td>
</tr>
<tr>
<td>5</td>
<td>مؤتمر ٢٠٠٩ conference٢٠٠٩</td>
<td>not-acceptable: Mix between European digits &amp; Arabic-Indic digits</td>
</tr>
<tr>
<td>6</td>
<td>مؤتمر ٢٠٠٩ conference٢٠٠٩</td>
<td>not-acceptable: Mix between Arabic-Indic digits &amp; Eastern Arabic-Indic digits</td>
</tr>
<tr>
<td>7</td>
<td>مؤتمر ٩٠٠٢ conference٩٠٠٩</td>
<td>not-acceptable: Mix between European digits &amp; Eastern Arabic-Indic digits</td>
</tr>
</tbody>
</table>
Securing Arabic Domains

(6) WLE Rules

– HEH and TEH MARBUTA
  – They are variants only in the final form (at the end)
    – جدة vs جدة

– ALEF MAKSURA:
  – Preventing connected ALEF MAKSURA
    – سَد vs سَد
What’s Missing?

Registry

Register and enable necessary variants:

- مكة
- مكة
- مكة

DNS Hosting

Configure DNS & add need RRs (e.g. NS & A & CNAME) for:

- xn--ogb5cf
- xn--ogb9c4p
- xn--hhb4rwc

Email Hosting

Configure Email account and email aliases:

- رائد@مكة
- رائد@مكة
- رائد@مكة

Web Hosting

Configure web-server and account and aliases:

```
<VirtualHost 10.10.10.10>
    DocumentRoot /makkah
    ServerName xn--ogb5cf
    ServerAlias xn--ogb9c4p
    ServerAlias xn--hhb4rwc
</VirtualHost>
```
Gift

• SaudiNIC’s Best Practices in Supporting and Managing Arabic Domain Names

http://www.nic.sa/docs/SaudiNIC_ADNBP.pdf
Thank you

للمزيد من المعلومات يمكنكم زيارة:
For more information you can visit:

هيئة الاتصالات وتقنية المعلومات
Communications and Information Technology Commission

سجل.السعودية
nic.sa

هيئة الاتصالات.السعودية
citic.gov.sa