IDNA Protocol Revision

Workshop

Paris, France
26 June 2008
Workshop Agenda

• Introductions:
  - General intro/Rationale (Tina)
  - Revision Details (Cary)

• The two major revision topics:
  - Character list generation rules (Patrik)
  - R-t-L/Bidi (Harald)

• Implementation:
  - Registration including non-protocol responsibilities (Cary)
  - Resolution (Patrik)
  - Example of script coordination: ASIWG (Ram)

• Conclusion:
  - Timing scenarios (Cary)
  - Summary and questions (Tina)
Characters in the DNS

- The DNS can handle all US-ASCII characters
  - Examples:
  
  | a,b,c...,z | (!) EXCLAMATION MARK |
  | 0,...,9    |                   |
  | -         |                   |
  | (') APOSTROPHE | (&) AMPERSAND |
  | (%) PERCENT SIGN | ($) DOLLAR SIGN |
  | (#) NUMBER SIGN | (") QUOTATION MARK |

- Most TLD registries have implemented the hostname rule
  - Also known as LDH
  - Domain names can only contain:
    - a,b,c...,z
    - 0,...,9
    - (dash)

- That was before internationalization....
IDN Definitions

• Internationalized domain names are:
  – Names with characters other than the standard ASCII (a,b,…z), (0,1,…9), (-)

• IDNs are about localized solutions
  – But need to be ‘internationalized’ due to the global nature of the Internet

• IDNs have existed as second level since 2003
  – under the protocol standards (IDNA)
  – email protocol standard is underway (IETF)
  – IDN Guidelines was issued in 2003 to support the IDNA protocol
### What we have and where we are headed

<table>
<thead>
<tr>
<th>Today's Availability</th>
<th>Future Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCII domain names (a, b,...,z), (0,1,...,9), (-)</td>
<td>IDN TLDs</td>
</tr>
<tr>
<td>domainname.TLD.icann.org</td>
<td>실례.TLD -under various existing TLDs</td>
</tr>
<tr>
<td>IDN second level</td>
<td>실례. 테스트</td>
</tr>
</tbody>
</table>
IDN Definitions

• The A-label
  – is what is transmitted in the DNS protocol and this is the ASCII-compatible (ACE) form of an IDNA string; for example "xn--11b5bs1di".

• The U-label
  – is what should be displayed to the user and is the representation of the Internationalized Domain Name (IDN) in Unicode; for example "परीका" ("test" version in Hindi, Devanagari script).

• The LDH-label
  – strictly refers to an all-ASCII label that obeys the "hostname" (LDH) conventions and that is not an IDN; for example "icann" in the domain name "icann.org".

• The IDNA protocol provides the transition back and forth between A-labels and U-labels
Displayed Form vs. Stored Form

• Historically the domain name you register is also the domain names stored and usable in the DNS.

• This is changed with introduction of IDNs.

• Usually the stored form usually gives no meaning.
  – Example: ﻓﺮﺳﺎﻟﻨﻬﺮ.tld → xn--mgbtbg2evaoi.tld

• However, there are exceptions:
  – xn--gibberish - decodes into the Arabic characters ﻓﺮﺳﺎﻟﻨﻬﺮ
  – xn--trademark - with different versions of trademarks
  – This is coincidentally and hence not intentionally.

• xn-- prefix indicates to application software that the label needs to be decoded back into Unicode for proper display to the user.
Rationale for the IDNA revision

• Proposed revision at IETF
  – RFC4690 requests the revision and provides suggestions to solutions to some problems

• Reasons and results of the revision:

<table>
<thead>
<tr>
<th>Current Version</th>
<th>Revised Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicode version 3.2</td>
<td>Unicode version independent</td>
</tr>
<tr>
<td>Some/New characters excluded</td>
<td>All characters in Unicode will have a status</td>
</tr>
<tr>
<td>Not all words can be represented</td>
<td>Not all words can be represented</td>
</tr>
<tr>
<td>Exclusion Based:</td>
<td>Inclusion Based:</td>
</tr>
<tr>
<td>- Table based</td>
<td>- Property and procedure based:</td>
</tr>
<tr>
<td></td>
<td>- Protocol-valid (w/ context rules)</td>
</tr>
<tr>
<td></td>
<td>- Disallowed</td>
</tr>
<tr>
<td></td>
<td>- Unassigned</td>
</tr>
<tr>
<td>App developers have difficulty in understanding</td>
<td>Separates registration and resolution in detailed</td>
</tr>
<tr>
<td>description of standard</td>
<td>steps</td>
</tr>
</tbody>
</table>
Rationale for IDNA Protocol Revision

• Other issues was discovered during the revision process
  – For example: bidirectional problems

• Overview of documents, by Patrik Faltstrom:
  – http://stupid.domain.name/idnabis/
    – Overall rationale and explanation
    – Protocol: registration vs. resolution
    – Tables and procedures
    – Bidirectional issues solutions
Workshop Agenda

- **Introductions:**
  - General intro/Rationale (Tina)
  - Revision Details (Cary)

- **The two major revision topics:**
  - Character list generation rules (Patrik)
  - R-t-L/Bidi (Harald)

- **Implementation:**
  - Registration including non-protocol responsibilities (Cary)
  - Resolution (Patrik)
  - Example of script coordination: ASIWG (Ram)

- **Conclusion:**
  - Timing scenarios (Cary)
  - Summary and questions (Tina)