Internationalized Domain Names Technical Review and Policy Implications

John C Klensin
APTLD Manila
23 February 2009

Today's Discussion

◆ Quick review of IDNA
- ◆ How it works
- ◆ Proposed changes to protocol

◆ Not a substitute for in-depth tutorial

◆ Or a review of ICANN policy machinations

◆ Focus policy-important issues

IDNA Goals

- Provide mechanism for non-ASCII domain name labels
- Without disrupting
  - DNS operations
  - Existing applications
- Still obey other DNS constraints
  - Length limits on labels and FQDNs
  - Severe limits on punctuation and special characters
  - Language-based mnemonics, not "words"

Status

- ◆ Original standards published in 2003
- ◆ Fairly widely implemented and deployed in many zones (registries)
- ◆ Clear by early 2005 that there were some issues, both conceptual and protocol
  - ◆ Upgrade path for new versions of Unicode and new scripts
  - ◆ Confusing terminology
  - ◆ Opportunities for mischief
  - ◆ Misunderstandings of DNS
- ◆ This presentation talks about 2003 standards using 2008/9 terminology

Basic Approach

- ◆ No change to DNS
  - ◆ Too disruptive of applications
  - ◆ Likely long and difficult transition
- ◆ Applications use native/local characters but...
  - ◆ Convert to Unicode (if needed)
  - ◆ Validate and process ("U-label")
  - ◆ Convert to a special, ASCII-compatible, form "A-label". (incorrectly "punycode")
- ◆ Only A-labels stored in DNS
- ◆ Applications finding A-labels convert back
  - *Usually*

The Mechanism

- ◆ Only A-labels go into DNS zones
- ◆ Interpretation, coding, and decoding are application matters, not DNS ones
- ◆ Non-IDNA-aware applications or implementations
  - ◆ Just see the A-labels
  - ◆ Understand them as ordinary hostname-format ("LDH") labels

Subject to All DNS Restrictions

- Fully-qualified Domain Names ("FQDN") made up of labels
- Up to 63 characters per label (for IDNs, fewer in practice)
- Exact-match lookups
- Hierarchy-independent matching
- No language information, just strings
- About mnemonics, not "words"
  - No expectation of being able to write all words of any language
  - Many non-words in use
- Weak alias capability

IDNs and "Multilingualism"

- ◆ A tool for navigation
- ◆ Worse for finding information by name-guessing than traditional DNS names
- ◆ Will not solve any content problems

- ◆ Real language-dependent issues
  - ◆ Cannot really be addressed in DNS
  - ◆ Need to be closer to application, user, and context.

IDNs and Other Issues

◆ Many internationalization problems, even with naming
  ◆ IDNs are the best solution to only a few of them.
  ◆ May be terrible solutions for others
◆ Sometimes best solution is determined by who benefits
  ◆ Best solution for profit-making registry may not be best one for users.
  ◆ Near impossibility of maintaining parallel DNS trees may eliminate some options.
◆ Three major alternatives for top-level IDNs
  ◆ ICANN processes ignore two of them

DNS Internationalization

- Not just IDNA
- Many areas of responsibility to make things work well
  - Standard Protocol
  - Registries and Registry Restrictions
  - Registrants – Increasing or decreasing confusion
  - Lookup implementors and Applications
  - Common sense by users and others
- Still mnemonics, not literature
- Can't *solve* confusion, but can provide better tools

Why a Revision

- A number of issues with IDNA2003
- Some very important to specific communities even though less so to others
- Others have general impact
- Goal is to keep and improve balance between
  - Mnemonics for as many languages/ scripts as possible
  - Identifiers that are
    - Safe and reliable
    - Easy to use
    - Well-behaved and more predictable

Issues with IDNA2003

- Structure of definition
  - Unicode version dependency
    - Applications don't know, regardless of what standard says
    - Looking up code points that aren't defined
  - Poor ability to understand
    - What is permitted
    - What happens to it.
  - Poor extensibility and ability to check
    - In some ways, any change at all is incompatible
    - List of label separators

IDNA2008 Still Under Discussion

- Discussion is very difficult at times
  - Several issues for which there are no clear answers
  - Tradeoffs among many positions, including
    - Including strict compatibility with some abuses of IDNA2003
    - Maximum flexibility for content designers versus reliability for users.
- Some things in this presentation may change
- A conservative registry working with labels in a script with it is familiar should see little change

Some Key IDNA2008 Goals

- Get a standard that is
  - Unicode version agnostic
  - Easier to understand
  - More predictable with regard to what happens
  - More adaptable to local conditions – realistic interoperability, not just theory
  - Deals with issues discussed earlier to extent possible
- Last big revision
  - After this, fully backward-compatible

Current IDNA2008 Status

◆ Revision effort underway
  ◆ Mostly tuning
  ◆ Largely invisible to careful existing applications
  ◆ Some changes still being debated
    ◆ Especially IDNA2003 compatibility for strings in files
    ◆ Nature of conversion/ migration advice
    ◆ Lookup restrictions (unassigned code points

Implications for Registries

- More explicit responsibility to validate
  - No more "guess what I mean"
  - Many doing that already
  - Changed "enforcement" model
- New characters
  - May require special consideration relative to existing labels and/or confusing label pairs
  - Potential for much better handling of a few scripts, including some uses of Arabic and most Indic scripts (but contextual rules)
- Inclusion list
  - No more punctuation, symbols, compatibility characters

More Reading -- IDNA2008

- New Definitions
  - https://datatracker.ietf.org/drafts/draft-ietf-idnabis-defs/
- Rationale, context, registry advice
  - https://datatracker.ietf.org/drafts/draft-ietf-idnabis-rationale/
- Actual protocol definition
  - https://datatracker.ietf.org/drafts/draft-ietf-idnabis-protocol/
- Handling of right-to-left characters
  - https://datatracker.ietf.org/drafts/draft-ietf-idnabis-bidi/
- Permitted character definitions and tables
  - https://datatracker.ietf.org/drafts/draft-ietf-idnabis-tables/

More Reading

- Radical alternative to IDNs at top level
  - http://www.isoc.org/briefings/018/
- Multilingual Internet and IDNs
  - http://www.isoc.org/educpillar/resources/docs/mul
- In-depth tutorial on DNS and balanced consideration of policy issues
  - http://books.nap.edu/catalog.php?record_id=11258