

Internationalized Domain Names

The past, the present, the future

Erik Nordmark
erik.nordmark@sun.com
Sunlabs Europe

History

- IETF BoF in November 1999
- IETF IDN Working Group formed in early 2000
 - Lots of debate
 - Participants with many different backgrounds
 - Educational
 - Some folks learning about language and script issues
 - Some folks learning about DNS issues
 - Some folks learning how hard the problem is
- Slow!!!

Disclaimer

- These are personal observations and opinions
- NOT IETF consensus

Types of Requirements

- The main types that have been discussed
 - Fit the model of domain names
 - Allow for graceful migration
 - User requirements
 - The IETF skill sets
- This will turn out to be an overconstrained problem

The model of domain names

- Lookup and not search – a single exact answer or no answer
- Uniform rules for matching
 - Today case insensitive for ASCII characters
 - The secondary servers and caches know this rule
- Just a name without additional context
 - No language context
 - No cultural context (e.g., French is different in France and Quebec)

Graceful migration

- Don't break the existing DNS!!!
- Use existing DNS infrastructure
 - With an ASCII compatible encoding (ACE)
- Use existing applications
 - Be able to reply to email and click on URLs even if the characters display as bq—gobbledygook
- IDN aware applications
 - Pretty display and input of the names

Possible user requirements

- Not discussed in depth in the WG
- Assume user knows the language and script
 - Be able to type a URL or email address from print
 - Business card, advertisement
 - Be able to do this from voice communication
 - Phone call, radio
 - With reasonable probability of getting it right
- Minimize homographs
 - Such as MICROSOFT vs. MICR0S0FT

Possible user requirements (contd)

- Sensible name registration?
 - What does this mean?
 - Register räksmörgås.se and also get raksmorgas.se?
 - Generalizes to names with same “meaning”
 - Unclear to which extent this can be done in general
- Consistent delegation?
 - When multiple names have the same “meaning”

IETF expertise

- No expertise in character sets, languages, and scripts
- Even with such expertise it takes a loooong time to develop or modify character sets
 - Need input from all language communities in the world
 - And any new character set might just do different tradeoffs – not be strictly “better”
- Thus pick an existing character set

Attributes of the solution

- Picked Unicode 3.2 as the character set
- Clients prepare names
 - Case fold, map out, Unicode NFKC, and prohibit
 - For lookups and for dynamic DNS update
- Punycode is the chosen ACE
 - Reasonably simple yet efficient coding of Unicode
- DNS infrastructure is unchanged
 - New administrative tools to handle Unicode and ACE when registering names

Attributes of the solution (contd)

- DNSSEC properties hold
 - Data origin authentication i.e., can receive data from an untrusted server or cache and verify the data
- Unmodified application protocols use ACE
- Applications can be modified incrementally to parse and display the richer set of characters
 - bq—gobbledygook until they have been modified
- New application protocols might not need ACE
 - Might be defined to carry IDNs as UTF-8

Hard Issues

- Case is folded: ö = Ö, but Ø ≠ Ö
 - Different appearance
 - German ß = ss ?
- Latin “A” and Greek “Α” are different
 - But same appearance in most fonts
- “Theatre” and “Theater” are different
 - Obviously ...

Simplified and Traditional Chinese

- Simplified Chinese has evolved in China since '49
 - Making the script easier to read and write
- Hongkong and Taiwan use mostly traditional chinese characters
- Result is two different characters
 - Different Unicode code points – different appearance
 - Pronounced the same and have the same meaning
- In some cases two characters replaced by one
 - Or one character replaced by two

SC/TC equivalence?

- Would make sense for Chinese users
- Unicode has unified the Han characters that are common to Chinese, Japanese, and Korean
 - Same character and the same appearance
- But the SC/TC rules are specific to the Chinese *language*
- The domain name model has no language context

Making it easier for users? Could we do better?

- Internet Resource Name Search Service (IRNSS)
BoF at IETF in December 2001
- Example documents
 - draft-klensin-dns-search-04.txt
 - draft-mealling-sls-02.txt
- Idea is to layer one or two layers on top of the domain name model
 - With language and other context
 - With approximate matching
- Little traction in the IETF so far

Next steps

- Define local part of email addresses?
 - E.g., örjan@räksmörgås.se
- Formalize URI extensions for ACE
- Define an IRI format (Internationalized Resource Identifier) for use in new protocols?
- The future of IRNSS?
- ccTLDs deploying IDN for their language(s) make sense

Status of IDNA standardization

- Two pieces being progressed separately
 - Stringprep draft separate because useful outside of IDN
 - IDNA, nameprep, and punycode drafts handled together
- Both sets have been discussed by the IESG
 - Comments sent to the IDN WG and largely addressed
 - Minor editorial issues with documents
 - Need better problem statement and scope for IDNA
 - AD will write it