Looking at Internationalization ... Again

John C Klensin
DISPATCH – 2017-11-13
Focus Tends to be On Character Encoding and Matching

• But very broad set of issues, including
  • Date and time formats and time zones (affects calendaring, etc.)
  • Data entry conventions (affects UIs, character pickers, IMEs, OS design, etc.)
  • Presentation for particular situations (rendering, type style selection, etc.)
  • Confusable characters and deceptive strings (security issue, among other things)
  • Language-specific conventions (especially where language info is not available)
  • Transitions from ASCII-only systems (when, how, and where)
  • And that character encoding stuff

• Many suggested solutions depend on “let’s have This and That”
  • Horrible complications unless there is
    • Broad consensus on a canonical form
    • That consensus is understood and accepted by end users
Nonetheless, Focus Today Will Be On DNS and IDNA-related Issues

- Hypothesis: IDNA may have been a mistake
  - Backing out may be impossible
  - Will come back to that, but, if it is correct, question is how to make the best of a bad situation
- Many warts and drafts to address, at least partially, some of them
- IETF actions in the last three years have been:
  - Clearing up some PRECIS definitional issues
  - IAB statement to ICANN suggesting doing nothing until the IETF sorts this out
  - A BOF (LUCID) that went down a few ratholes and led to nothing
  - Some patches to allow non-ASCII domains and email addresses in certs
    Not our usual definition of progress on core issues
The Immediate Queue

• Three drafts awaiting consideration/processing
  • draft-klensin-idna-5892upd-unicode70-05
    • Updated exploration of the issues (“non-decomposing characters”) that led to the LUCID BOF at IETF 92
  • draft-klensin-idna-5891bis-01
    • Clarification of an IDNA issue, nothing really new
  • draft-freytag-troublesome-characters-01
    • One of several possible attempts to work around the issues

• Other work needed to address now-known problems
What IETF Participants Want

• Most frequent answer is
  • “Just tell me what to do “
    • so I can follow the recipe and be safe, without having to think or understand
    • No such easy answers, sorry.

• When we try to think about the issues, we keep
  • Falling into the Latin Script trap
  • Or at best the “my two languages” trap
  • In general, extrapolation from one language and script doesn’t work (and two doesn’t work much better)
What is to be done?

• If we care, “nothing” is not an option
  • Many different “interpretations” of IDNA
  • Each variation creates risk of
    • User confusion
    • Systems not working as expected (false negatives or positives)
    • Attacks (not just phishing)
  • IDNA isn’t the only problem, but DNS implies special complications

• Whatever we are doing now, it doesn’t work
  • And things are getting worse

• ART area problem to solve
  • Unless we can find another way
Possibilities That Have Been Discussed

• Create a WG
  • Problem with critical mass of expert people
  • Problem with leadership
  • Problem with resources

• Create a special review team and empower it
  • Problem with how to select membership
  • Would IESG and the community accept the conclusions

• Leave the problems to someone else (e.g., Unicode Consortium)
  • Has not served us well in the past
  • Contradictory advice already (if one believes DNS names are identifiers)
  • Apparently can’t conceive of language-independent identifiers

• Other ideas?  ????
Just Remember

- Ignoring the issues –doing nothing –
  - Creates risks
  - Makes, or lets, things get worse.

- If our job is to make the Internet work better globally
  - Ignoring internationalization is inconsistent with that goal