precis: problem statement
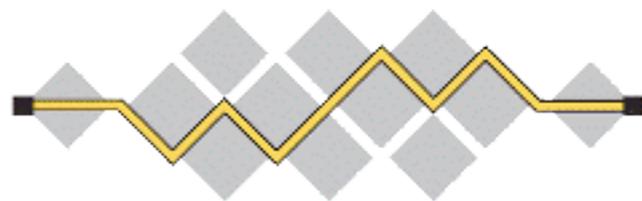
Marc Blanchet, Andrew Sullivan
Beijing, China, November 2010

I E T F®

Issue 1: all charsets or UTF-8?

Abstract says this:

Using Unicode codepoints in protocol strings that expect comparison with other strings requires preparation of the string that contains the Unicode codepoints.

I E T F

Issue 1: all charsets or UTF-8?

What are we trying to do?

- Deal with internationalized and localized input?

- Deal with UTF-8 input, and let the application cope with localization?

- Does IDNA2003 → IDNA2008 work provide analogies? Disanalogies?

I E T F

Issue 2: reviews? hello, WG?

In Maastricht, people said they'd do reviews of protocols they cared about.

As of -00 document, no such reviews done. Tickets in trac.

IF NOBODY REVIEWS THESE PROTOCOLS, THERE IS NO PROBLEM STATEMENT. WE SHOULD CLOSE WG.

I E T F

Issue 3: context rules

- IDNA2008 has complicated context rules

- Very much a product of trying to "internationalize LDH"

- Do we have (any) profiles like this?

  - ◉ Which, and what do we want to do about it? I'm taking names.

I E T F

Issue 4: Case folding, preservation, and mapping

- Case preservation is when the protocol passes case along

- Case sensitivity is when it matters whether you typed (say) "a" or "A"

- If you can do case folding, then you can be sure that "a" and "A" are "the same enough"

Issue 4: Case folding, preservation, and mapping

Not every script has cases.

At least some language speakers don't understand why you have a problem.

Issue 4: Case folding, preservation, and mapping

- Why so important in IDNA?

- Is it important in your protocol?

I E T F

Issue 5: delimiters

- This is sort of a nightmare in DNS: the "." *is only sort of* on the wire.

- For IDNA2008, not exactly a problem: up in UI, and delimiter is always DNS limit (i.e. label markers).

- Does anyone need us to cope with delimiters?

Issue 6: NFC and NFKC

- How many people understand this difference?

I E T F

Issue 6: NFC and NFKC

- How many protocols' users never see the "protocol-real" string?

- How many protocols' users have the "side-of-a-bus" problem?

- Does this distinction help?

Issue 7: Surprise! Characters moved categories!

- IDNA2008 depends on Unicode categories

- It should be *really strange* for a character to move from PVALID to DISALLOWED

- Latest Unicode standard implies this

- How do we face this?

Issue 8: IRI paper just posted to mailing list

Is this a problem for us?

I E T F