

draft-saintandre-xmpp-i18n-01

XMPP WG / NewPrep BoF, IETF 77
Peter Saint-Andre


Background

- XMPP (a.k.a. Jabber) is an XML streaming technology invented in 1999 and specified in RFC 3920 (published in 2004)
- Has always used UTF-8 encoding
- Currently uses three Stringprep profiles for XMPP addresses
- How proceed “post-IDNA2008”?

XMPP Addresses (JIDs)

- `<localpart@domainpart/resourcepart>`
- Localpart (“node”) can be a username, a chatroom, a bot, etc.
- Domainpart is a server or component
- Resourcepart can be a device identifier, a nickname in a chatroom, etc.

JID Examples

- <münchen.de>
- <司馬 安平@jabbercn.org>
- <jiří@jabber.cz/domů>
- <jdev@conference.jabber.org/the  >

How Common?

- An unscientific sample (a certain large buddy list, with 2500+ users) yielded no IDNs but several i18n usernames and multiple i18n device identifiers
- More thorough sampling would be good
- Seemingly used mostly by end users in parts of the address under their control

Expressivity

- Users consider it a benefit of using XMPP that they can have non-ASCII usernames, roomnicks, etc.
- Deployments might restrict allowable characters, scripts, etc. when provisioning an account or joining a chatroom (but no methods for this at present)

Domainpart

- Currently use Nameprep and IDNA2003
- Seems advisable to migrate to IDNA2008
- Support only U-labels on the wire?
- Note: U-label comparison in IDNA2008 occurs without case-folding
- Possible issues: certificates, dependency on registrars, software configuration, other?

Localpart (goals)

- Often a username, which might be my real name in my native language or script
- Slightly restrictive (no spaces, disallow characters like @ and &)
- Case insensitive (stpeter@jabber.org is the same as StPeter@jabber.org)
- Ideally, consistent with email, SASL, etc.

Localpart (RFC 3920)

- Most chars above U+0020 are allowed
- Disallowed chars: ASCII and non-ASCII spaces, controls, C.3-C.9 from stringprep, also " & ' / : < > @
- Case mapping: B.1 + B.2 from stringprep
- Normalization: NFKC

Localpart (bis)

- All “namelike” characters should be valid (but do we need symbols, stars, etc.?)
- Disallow chars that are disallowed now
- Case mapping: as in IDNA2008, or specify case-folding methods?
- Normalization: NFC?
- Harmonize with EAI and SASLprep?


Resourcepart (goals)

- Sometimes a “nick” (e.g., in chatrooms)
- Often a free-form device identifier, such as a location, a machine name, or some fanciful text such as a glyph or dingbat
- Less restrictive than localpart
- Case sensitive (but is this a good idea?)

Resourcepart (RFC 3920)

- Most chars above U+0020 are allowed
- Disallowed chars: non-ASCII spaces, controls, C.3-C.9 from stringprep
- Mapping: B.1 from stringprep
- Normalization: NFKC

Resourcepart (bis)

- All “namelike” characters should be valid
- Do we really need things like “the ”?
- Case mapping: continue to ignore case?
- Normalization: NFC?
- Perhaps use same rules as for localpart?

Migration Issues

- Lots of deployed software
- Certificates
- Tools (configuration, registration, etc.)
- User credential databases (cf. SASLprep)
- JID-based authorization
- Reserved roomnicks

XMPP Enforcement (I)

- Which entities are responsible for enforcement of internationalized JIDs on the network?
 - Only an XMPP server?
 - Also XMPP clients?
- And what do we really mean by enforcement? ...

XMPP Enforcement (2)

- Account provisioning / registration
- Login / authentication
- Upload vCard
- Add / modify roster item
- Send or receive an XMPP stanza
- In general: what are the JID “slots”?

Open Issues

- Clearly there are many open issues
- This is holding up XMPP WG milestone to submit rfc3920bis to the IESG
- How do we make forward progress?