

CJK Local mapping in IDNA2008

draft-jet-idnabis-cjk-localmapping-01

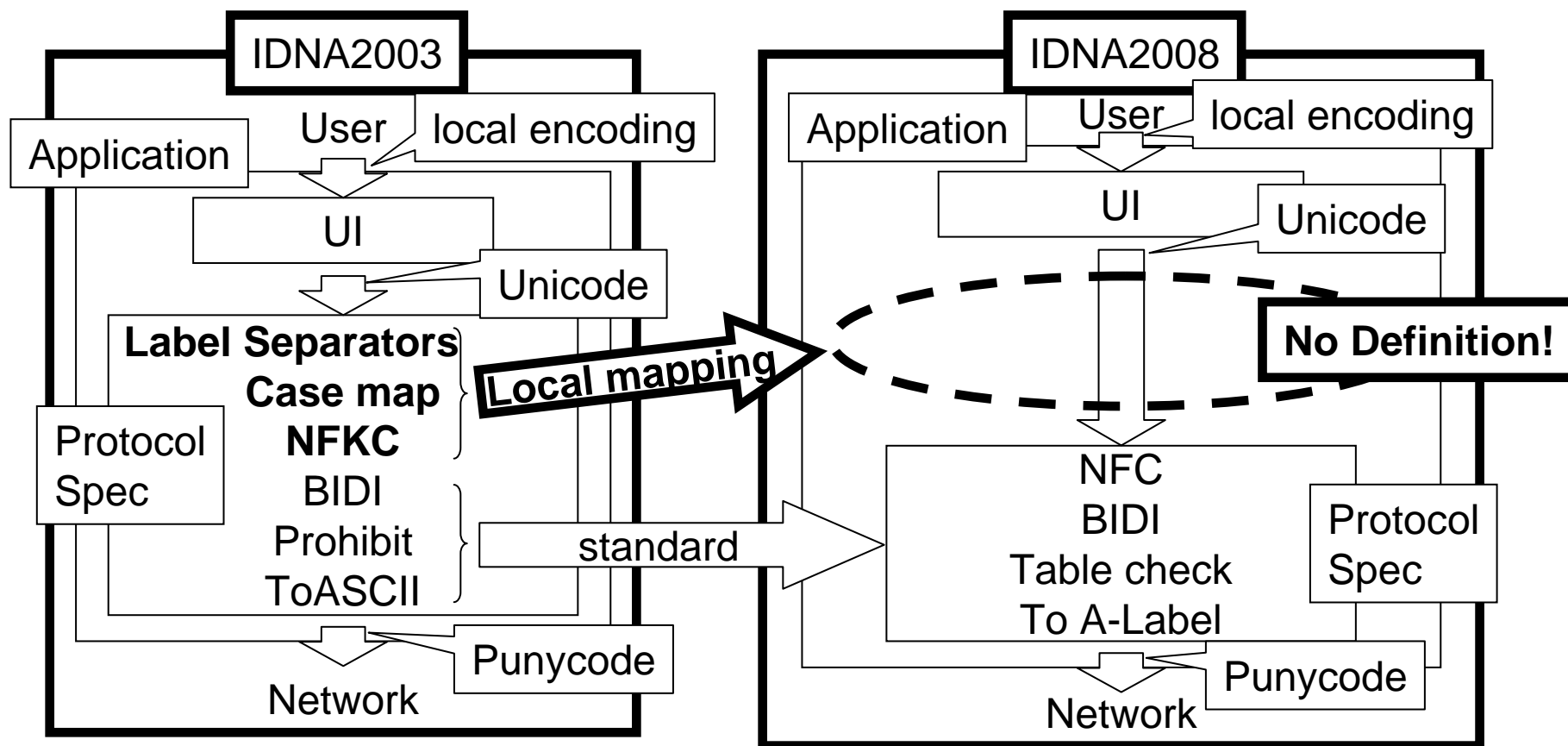
23 Mar 2009

IDNABIS WG @ IETF74

CJKT Joint Engineering Team



Background

- IDNA2008 has incompatibility with IDNA2003 which will damage existing IDNs in CJK



Examples of incompatibilities in CJK

① Normalization of string

| Input | IDNA2003 | IDNA2008 |
|---|--|--|
| ABCストアー.JP  | abcストアー.jp  | NG (Full Width and Half Width compatible chars are prohibited) |

② Label Separators

| Input | IDNA2003 | IDNA2008 |
|------------------|------------------|---|
| 日本語。jp 日本語.jp | 日本語.jp 日本語.jp | NG (Ideographic Full Stop and Full Width period are prohibited) |

③ Quasi Ideographic marks

| Input | IDNA2003 | IDNA2008 |
|--------------------|--------------------|--|
| さ々き.jp 中・日・韓.jp | さ々き.jp 中・日・韓.jp | NG (“々” have to be next to Han (Kanji), “・” have to be between Kana) |

Local Mapping Required

Motivation of CJK Local mapping

- Keep compatibility between IDNA2003 and IDNA2008 – intended to complement IDNA2008
- Definition of local mapping for CJK scripts
 - Mapping of label separators
 - Mapping of compatible characters
 - Treat quasi ideographic marks as a PVALID character
 - Apply if CJK script is used
- Aiming to be a BCP – a guideline for application developers and registries comply with IDNA2008

CJK scripts

- Chinese, Japanese and Korean IDNs are consist from one or more combination of following scripts

| Script Name | Code (Range) | C | J | K |
|---|-----------------|---|---|---|
| CJK Symbols and Punctuation | U+3000-U+3007 | Y | Y | |
| Hiragana | U+3040-U+309F | | Y | |
| Katakana | U+30A0-U+30FF | | Y | |
| CJK Unified Ideographs Extension A | U+3400-U+4DFF | Y | | |
| CJK Unified Ideographs | U+4E00-U+9FFF | Y | Y | |
| Hangul Syllables | U+AC00-U+D7A3 | | | Y |
| CJK Compatibility Ideographs | U+F900-U+FAFF | Y | | |
| Halfwidth and Fullwidth Forms | U+FF00-U+FFEF | Y | Y | |
| CJK Unified Ideographs Extension B | U+20000-U+2A6D6 | Y | | |
| CJK Compatibility Ideographs Supplement | U+2F800-U+2FA1F | Y | | |

CJK Local mapping image

- CJK local mapping should be performed between UI and IDNA protocol layer.

