

IDNA

Internationalized Domain names in Applications
ICANN IDN Workshop, Malaysia
21st July 2004

James Seng

庄振宏

Assistant Director, IDA Singapore
former co-chair, IETF IDN Working Group

Problems

- How do we encode non-ASCII character in DNS?
- How do we deal with ‘equivalent’ characters?
- How to make IDN work?

Encodings

- Old protocols can only handle a subset of US-ASCII (A-Z etc)
 - domain names are not only for web surfing
- People wants to use their own *language* in domain names
 - Confusion between scripts vs language

Punycode (RFC 3492)

- Uses Unicode
 - Limitation : Script not Language
- ASCII Compatible Encoding (ACE)
 - Transcode Unicode strings to ASCII
 - e.g. 新家坡 => xn--3bs3aw5wpa2a
- Does not change DNS

'equivalent' characters

- Domain names are case-insensitive
 - AOL.com = aol.com
- Simple in English but not true with other languages
 - ä = Ä
 - i = I or i = Ĩ (U+0130)

'equivalent' characters

- Different way to represent same characters
 - compose ä (U+00E4)
 - decompose a .. (U+0061 U+0308)
- Similar looking characters
 - spot the diff: ICANN and ICANN
 - U+410A CYRILLIC CAPITAL LETTER A

‘equivalent’ characters

- “Variants” of same characters
 - 𡵓 = 𡵓 (U+8C48 = U+F900)
 - 莊 = 莊 (U+838A = U+8358)
 - 兩 = 𠂇 = 两 (U+5169 = U+4E21 = U+4E24)
 - 鬃 = 𠂇鬃宗 (U+9B03 = U+2FF1 U+9ADF U+5B97)

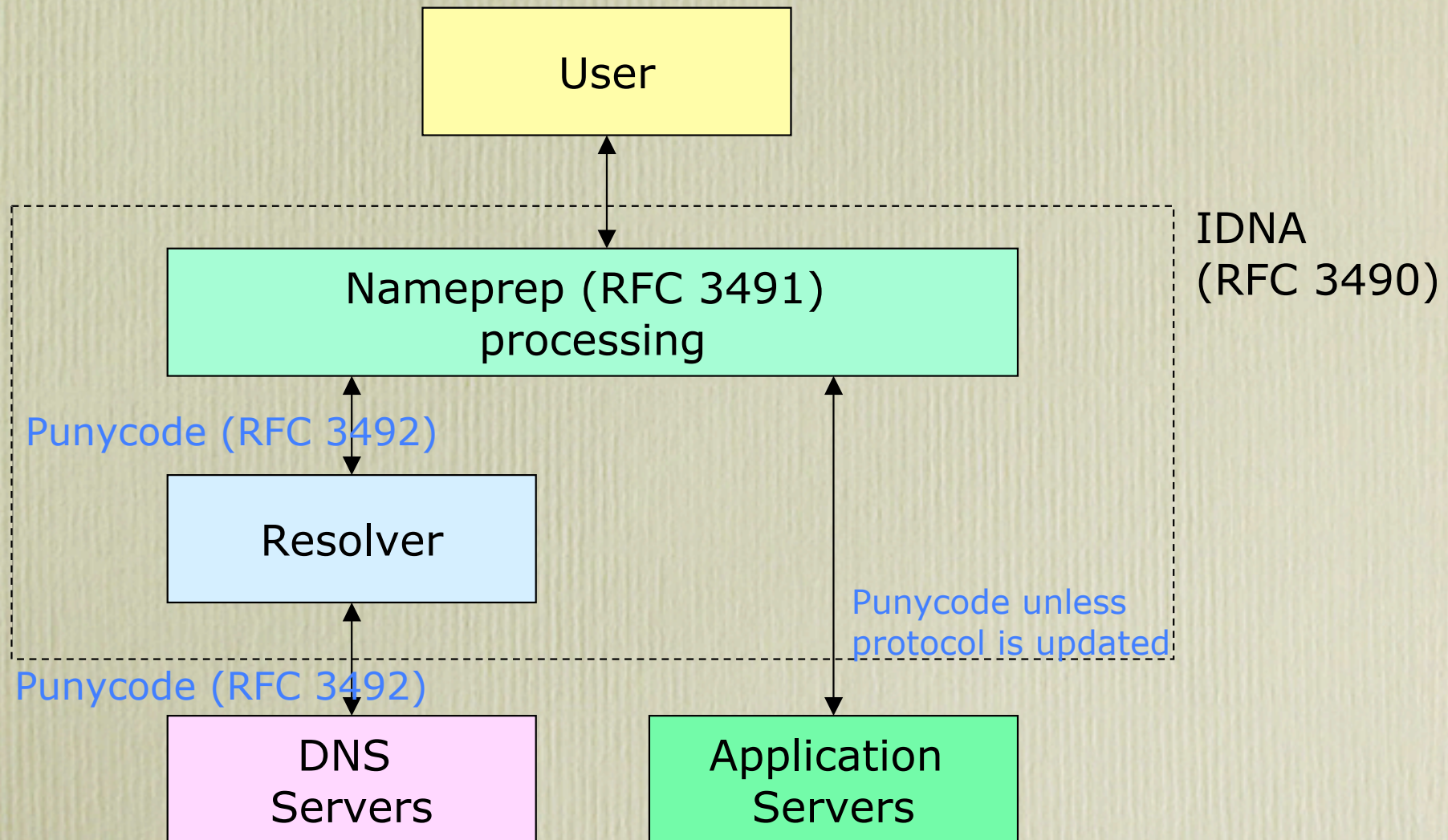
Nameprep (RFC 3491)

- Based on Unicode Normalization form KC (UTR#15) and Case folding (UTR#21)
- Goal is to reduce confusion and to have the *highest chance* of getting the domain name right
 - Law of Least Astonishment

Limitation of Nameprep

- Internationalization not Localization
- What is the domain name and what is in the zone file are two different things:
 - domain name e.g. 新家坡
 - zone file: xn--3bs3aw5wpa2a

IDNA (RFC 3490)



Software supporting IDNA

- Mozilla 1.4 and Netscape 7.1
- Konquerer 3.2
- Safari 1.2
- Opera 7.2
- Verisign i-Nav
- IDN-OSS Plugin
- etc etc

What does this means?

- There is a standard and least disruptive method of resolving IDN
- Neither the DNS nor the application level protocols have to understand Unicode
- Applications needs to be IDNA aware

But...

- IDNA deals with scripts
 - Users expect language
- IDNA is Internationalization
 - Users expect localization
- IDNA requires upgrade of applications
 - Users expect it to ‘just work’

Next step?

- Localization should be done at local registry
 - May involve linguistic issues
 - IANA registries for language tables
- Awareness and adoption of IDNA
 - Educate users of IDNA aware applications
 - Encourage developers to use IDNA

And finally...

- IDN Top Level Domain
 - Some languages don't mixed with English well
 - More intuitive to the users
- But ICANN should consider carefully how to deploy IDN TLD and esp. precedence
 - should .com get .公司 and other translation?

One day...

