

ICANN Rio Meeting

IDN Authorization for TLDs with ICANN agreements

26 March, 2003

Andrew McLaughlin



Internationalized Domain Names

华人.公司.cn

華人.商業.tw

高島屋.会社.jp

삼성.회사.kr

三星.회사.kr

الاهرام.م

tschüß.de

viagénie.qc.ca

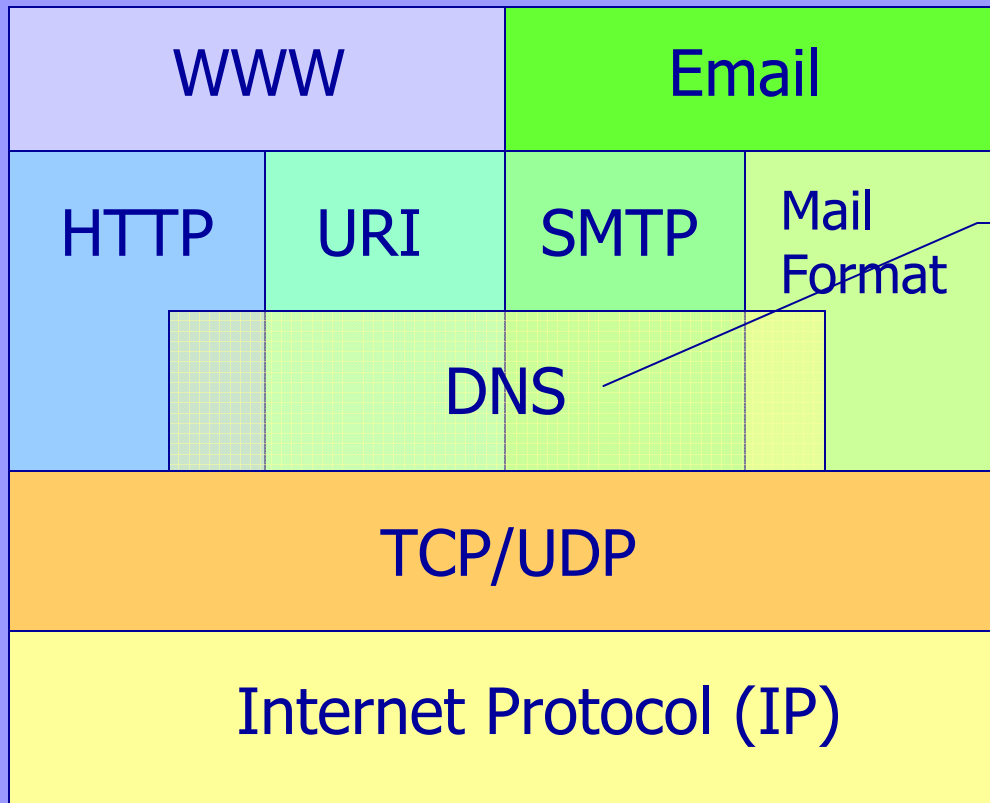
ישראל.קום

ทีเอชเน็ต.พาณิชย์.ไทย

現代.com

ヤフー.com

Where is IDN?



IDN is here

Stability and Compatibility

- From the IETF IDN-WG charter:

“A fundamental requirement in this work is to not disturb the current use and operation of the domain name system, and for the DNS to continue to allow any system anywhere to resolve any domain name.”
- IDN changes a fundamental service
 - IAB RFC2825 – “Issues of I18N, Domain Names, and the other Internet protocols”
 - IAB RFC2826 – “Unique Root”
- A lot of existing Internet Protocols utilize domain names and assume letters, digits, hyphen (LDH) only: A-Z, 0-9 and “-”

Problems

- Encoding
 - Using one universal character set or multiple characters sets?
 - What encoding to use? UTF-8? UTF-16?
- Matching / Equivalence
 - yahoo.com = YAHOO.com
 - 華人.com = 华人.com ?
 - BEHEMOTH.com = BEHEMOTH.com?
- Local issues
 - Language specific considerations

The New Standard - 1

Punycode

<draft-ietf-idn-punycode-03.txt>

- It uniquely and reversibly transforms a Unicode string into an LDH (Letter, Digit, Hyphen) string
- ASCII Compatible Encoding (ACE)
- Some compression to produce shorter string
- Example: 新加坡 = xn--3bs3aw5wpa2a

The New Standard - 2

Stringprep & Nameprep

<draft-ietf-idn-nameprep-11.txt>

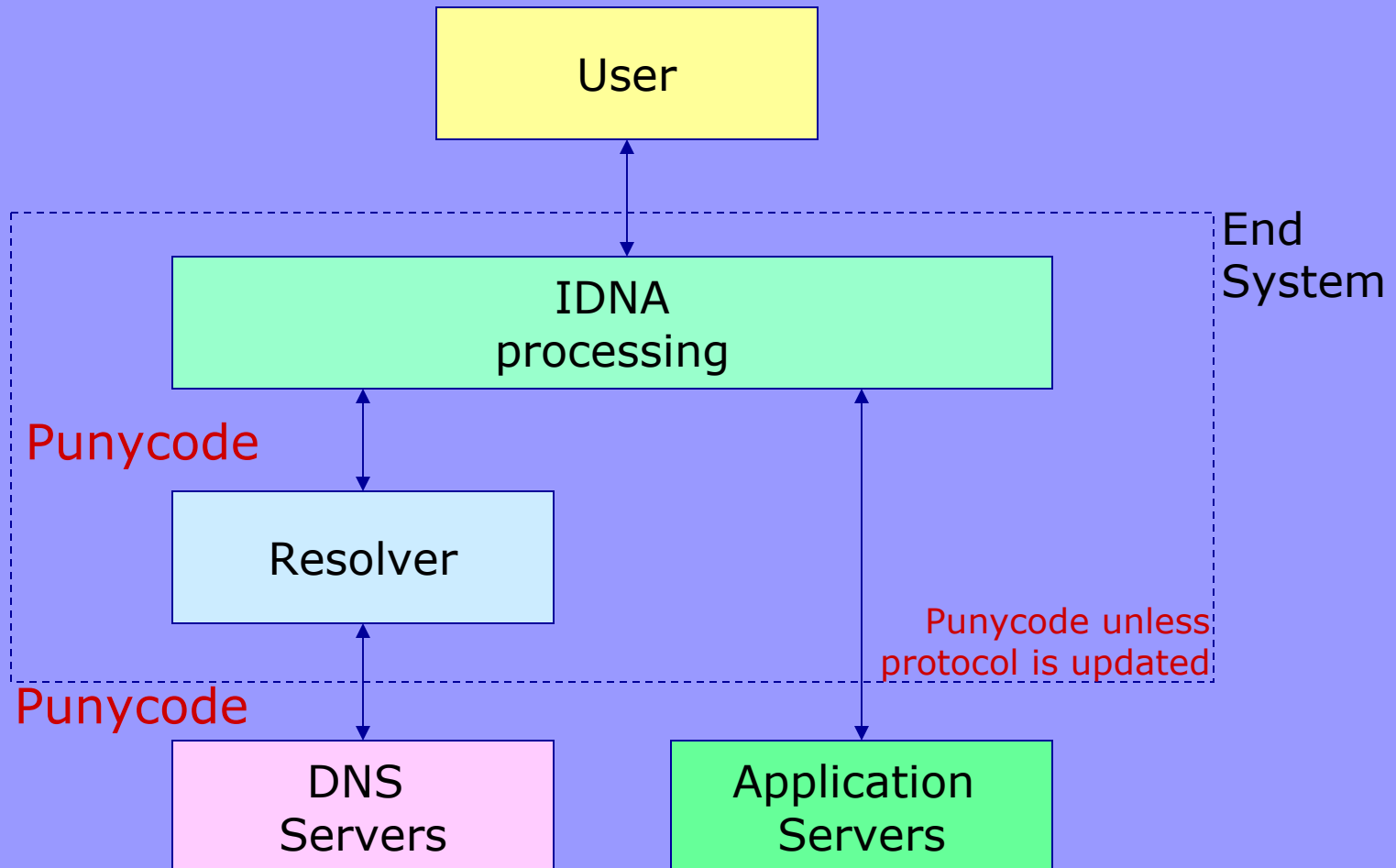
- Tables for matching and filtering
- Prepares internationalized domain name labels in order to increase the likelihood that name input and name comparison work in ways that make sense for typical users throughout the world.
- Stringprep is a generic processing method
- Nameprep is a profile of stringprep for Internationalized Domain Names

The New Standard - 3

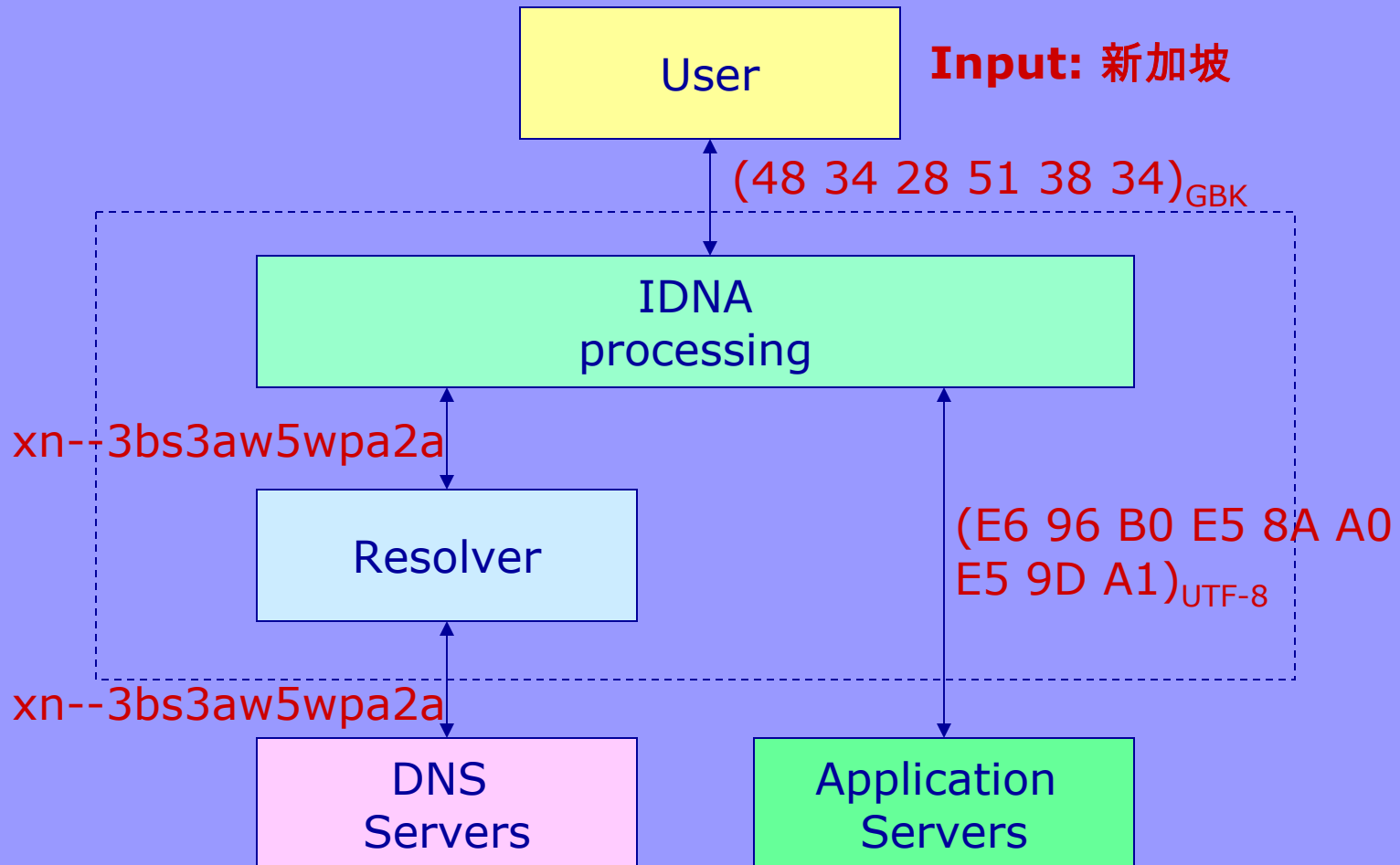
IDNA <draft-ietf-idn-idna-14.txt>

- “Internationalized Domain Names in Applications”
- Base protocol defining “ToUnicode” and “ToASCII” operations
- IDNA involves an upgrade in end-user applications to handle IDNs
- Consideration of legacy encoding and interoperability
- Requires Nameprep/Punycode in applications
- Sends Nameprep-ACE’ed IDNs over the wire

IDNA + Nameprep + Punycode



IDNA + Nameprep + Punycode



Problem of equivalents

Basic DNS/hostname standards define that upper-case and lower-case letters are treated as equivalents:

A = a

X = x

B = b

Y = y

C = c

Z = z

MICROSOFT.com = microsoft.com =
microSOFT.com = micROsoft.com, etc.

Registry-level issues

Registration policies and rules are a REGISTRY matter
(not ICANN's)

- Different registries may implement different rules & policies
- It's up to each registry to decide
- Harmonization is good, but depends upon cooperation of registries
- Consultation with affected communities is essential

For example, character-related issues:

- Confusion of names
- Alternative and variant characters
- Reserved name issues
- Non-language characters
- Mixed scripts (e.g., AHC.com = Greek? Cyrillic? Latin?)

CJK Rules

- For use of Han characters in the Chinese (Simplified and Traditional), Japanese, Korean languages
- Internet-draft on CJK now in third version: <draft-jseng-idn-admin-02.pdf>
- Great document; shows lots of hard work
- CJK has special problems:
 - *Language overlaying*
 - *Recent character reforms*
 - *Japanese and Korean are mixed-script*

Notes

- Definitions of permitted characters, preferences, and variant tables are per-zone
 - E.g., no requirement to agree on one universal table for Chinese (important, e.g., some areas have not adopted Simplified forms)
- Details will differ from TLD to TLD, but the CJK principles may be useful
 - Each zone (each TLD) bears some responsibility for overall stability of DNS, integrity of references, etc.

Real Cases: <IDN>.com

为什么 为什麼
为甚么 為什麼
為什麼 為什麼

→ **six names should
be registered as one**

schroeder.de
schröder.de

→ **two names should
be registered as one?**

[CJK Guidelines use term
“domain name package”]

TLDs w/ ICANN agreements

- ICANN authorization required to begin accepting registrations with hyphens in the third and fourth positions.
- Proposed guidelines for authorization:
 - Summary of two years of discussions among IDN-active registries (e.g., CJK registries, gTLDs), ICANN IDN committee & IDN-RIC
 - Still under discussion, wordsmithing, etc.
 - Consult with IAB as to specification of standards and compatibility of guidelines
 - Also: Analyze Nameprep exclusions for Dingbats, etc.?
 - Goal is commonly agreed statement – a documented commitment by a group of TLDs (those with ICANN agreements + a few others)



Credits

Thanks to James Seng, John Klensin, Wen-Sung Chen, Hilde Thunem for some of the points and images above.

