Internationalized Domain Names

Tina Dam
Director, IDN Program

10 September 2007

ICANN

8th International Conference

Vantutinki, Moscow, Russia

What is an Internationalized Domain Name

- An Internationalized domain name is a domain name with labels that
  - contain characters other than (a,b,…,z), (0,…9), (-)
  - is valid per the IDNA protocol
    - with a revision currently under consideration

- The domain name you register is, obviously, also the domain name that is stored in the DNS...

- ...with introduction of IDNs this is no longer as obvious:
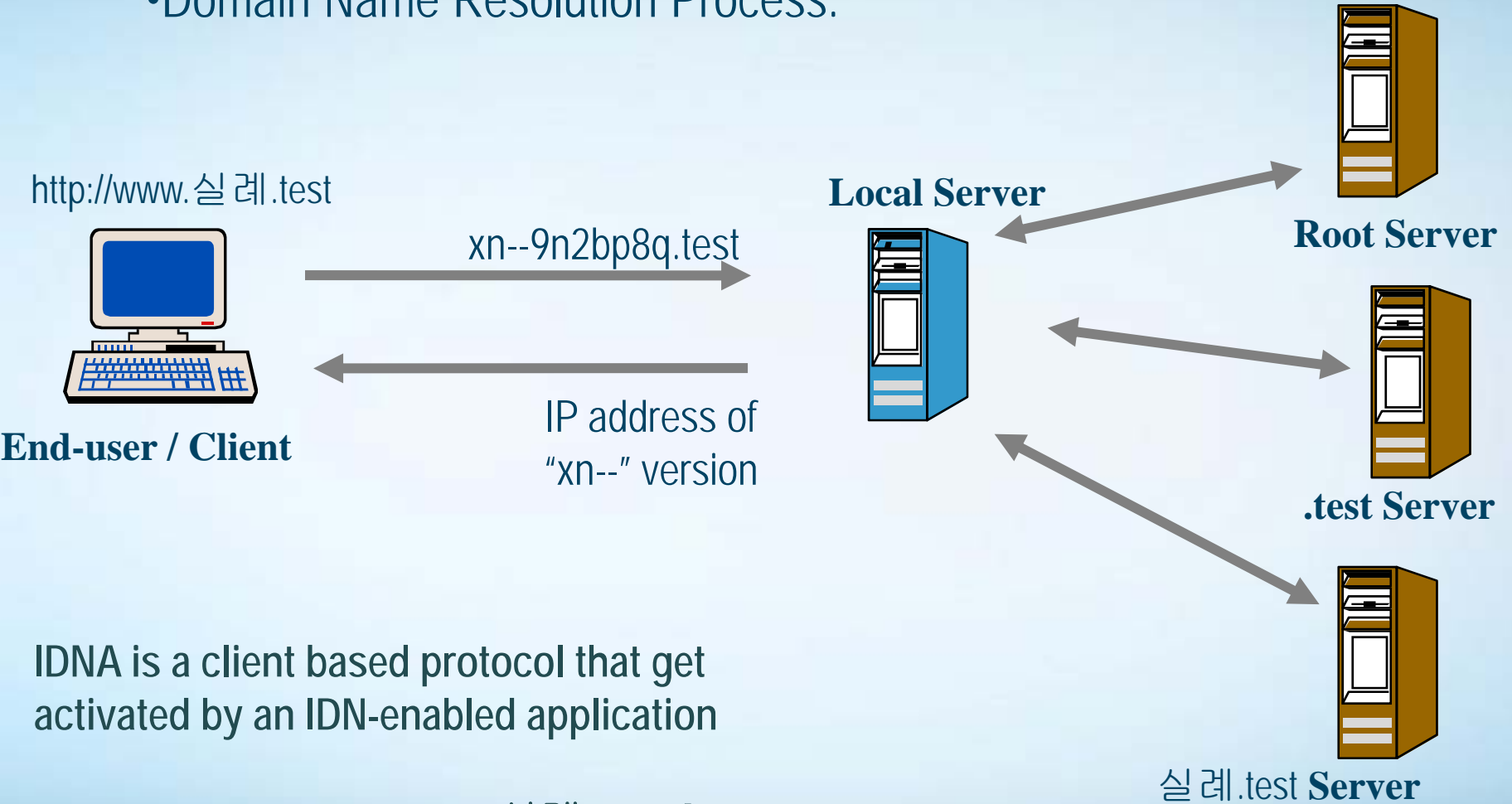 - A-labels
  - U-labels

Definitions

- The **U-label** is what the user expects to be displayed – the representation of the Internationalized Domain Name (IDN) in Unicode; for example " परीका " ("test" version in Hindi, Devanagari script)**.**

- The **A-label** is the ASCII-compatible encoding (ACE) of the same string; for example "xn—11b5bs1di" and is the form recognized by the DNS protocol.

- An **LDH-label** is a conventional all-ASCII label that obeys the "hostname" (LDH) conventions and is not transforned by IDN encoding; for example "icann" in the domain name "icann.org"

IDNA – Protocol Functionality

•Domain Name Resolution Process:

http://www.실례.test

xn--9n2bp8q.test

IP address of "xn--" version

End-user / Client

Local Server

Root Server

.test Server

실례.test **Server**

IDNA is a client based protocol that get activated by an IDN-enabled application

The result is generation of 실례.test → xn--9n2bp8q.test, which the DNS can handle

ICANN

Characters in the DNS

- The DNS can handle all US-ASCII characters
  - Examples:
    - (a…z), (0…9), (-)
    - ( ) SPACE
    - (!) EXCLAMATION MARK
    - (") QUOTATION MARK
    - ($) DOLLAR SIGN
    - (%) PERCENT SIGN
    - (&) AMPERSAND
    - etc…
- TLD registries have implemented the hostname rule allowing only (a…z), (0…9), (-) from the ASCII character set in domain names

Character set and the IDNA

- Character set: A standardized ordered list of characters, for example:

  - Unicode
  - US-ASCII

- The IDNA protocol operates on the Unicode character set

- The initial 2003 version of IDNA is linked to Unicode version 3.2

- The revised version of IDNA will not be dependant on a specific Unicode version

Characters, the DNS, and domain names

- Different languages that share the same script can easily differ in the way its individual elements are treated
  - In Czech, <ch> is a single character whereas in English it is two
  - Chinese 酒 - how do you count this? Single char? (Chinese "Jui")

- Reserved names working group recommendation is: "Single & Two Character U-labels not restricted at the top level; requested strings must be analyzed on a case by case basis depending on the script & language"

- In other cases languages that don't share a script are confusingly similar, for example Latin and Cyrillic

- IDN guidelines require that scripts are not mixed within labels unless in cases with established orthographies and conventions that require the commingled script usage

Why are we not there yet?

- Initial registration availability resulted in
  - visual confusion issues (example paypal.com)
  - damaging uniqueness principle of the DNS

- Different implementation in applications
  - security issues with IDNA2003
  - confusion of how to implement IDNA2003
  - different user experience per application

- Lack of allocation process and policies
  - Existing processes for new TLD delegation and re-delegation was not developed to work for IDNs

ICANN

Why are we not there yet?

- display of **xn--mgbh0fb** instead of مثال
- display of **xn--mgb0dgl27d** instead of ايكوم
- display of **xn--1lqs71d** instead of 東京
- display of **xn--1lq90i** instead of 北京

→ Results in trademarks being displayed where the U-label version may be a different trademark

- more user confusion and fraud opportunity
  – Registration of mïcrosoft.<tld> ?

- Protocol implementation experience and review showed other problems…

How do we solve these problems?

Proposed Revisions to IDNA Protocol

- Revising the IDNA protocol will among other things

  – build an "inclusion" based model for determining what scripts may be used for IDNs

  – increase available blocks of characters, via process

  – Non-unicode version dependant

  – fixing R-to-L error in Stringprep

- The revision effort is being managed through the IAB/IETF
  – Attempting finalization this calendar year

- The Basic Framework was published Sept-06
  – RFC4690

Evaluation of IDN TLD Capability

- Laboratory test of DNS resolver and root-server software (Autonomica, ICANN)
  - Feb07 report showed no negative effect in laboratory environment

- IANA procedure for inserting and managing top-level labels, including:
  - emergency removal procedure;
  - tolerance measure for activating emergency removal

    – Public comment period (2-22June07)
    – Procedure approved by ICANN Board in San Juan (June07)
    – Finalization of tolerance measure with RSSAC at IETF69/Chicago
    – Subsequently implemented at IANA

Technical Evaluations

- IDN TLD Application Evaluation Facilities

- The facilities are running in parallel and includes

  – Root zone insertion of '.test' (translated)

  – .test zones with second level labels 'example' (translated)

    - Arabic, Persian, 2xChinese, Russian, Hindi, Greek, Korean, Yiddish, Japanese, Tamil

Technical Evaluations

- Facility I (DNS focus):
  - a live replication of the earlier laboratory test

- Facility II (Client/Application focus):
  - an online usability with wiki's for each 'example.test' combination providing
    - Fully localized URL's - http://example.test/myname
    - Email reflector
- Anticipate live launch in Q4-2007
- Live-time of the wiki's:
  - will differ per language
  - depends on local community wishes

Live demo of .test evaluations
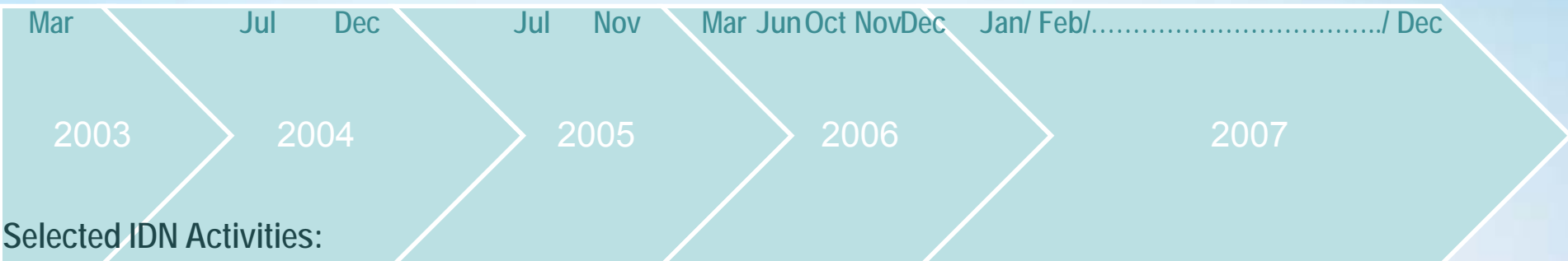
Policy and Processes

- Existing processes are not build to work for IDN TLDs:

    - Previous introductions of new gTLDs did not include IDNs
    - Process for delegation and re-delegation of ccTLDs functioning under IANA and based on ISO3166 list of 2-char codes

- IDN activity focused on

    - supplying recommendations to the ccNSO-GAC joint issues paper
        - http://ccnso.icann.org/announcements/announcement-09jul07.htm
    - GNSO: including IDN TLDs new gTLD process
    - ccNSO & GAC: two parallel tracks under consideration and discussion
        - Policy development process (PDP)
        - Limited initial launch (faster than PDP)

IDN Policy Issue Examples

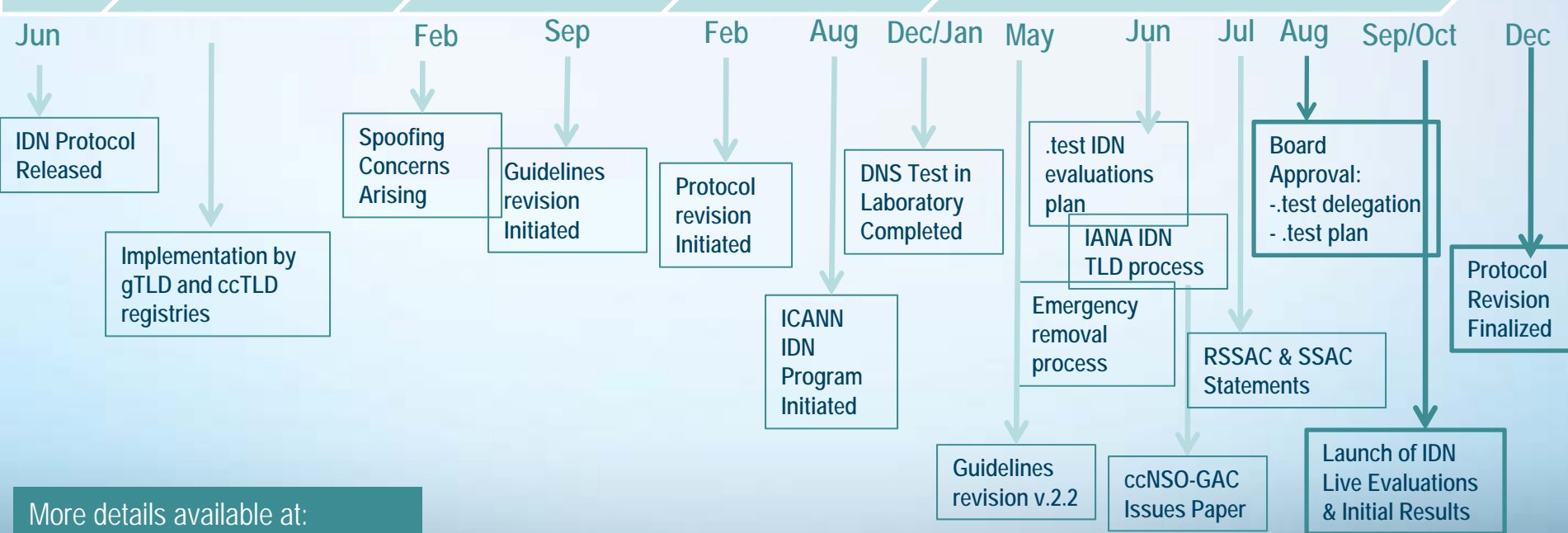- Examples of IDN policy issues
  - under consideration primarily by GNSO, ccNSO, GAC, ALAC

- **Which 'territories' are eligible for an IDN ccTLD?**

- **How many IDN ccTLDs per script per 'territory'?**

- **Are there any 'rights' attached to a given script?**

- **Should a list of IDN ccTLD strings be mandated?**

- **Do existing ccTLD delegation policies apply to the delegation of IDN ccTLDs?**

In Summary

Main IDN Related Conferences, Meetings and Events:

Mar	Jul	Dec	Jul	Nov	Mar	Jun	Oct	Nov	Dec	Jan/	Feb/…………………………../ Dec

2003	2004	2005	2006	2007

Selected IDN Activities:

| Jun | | Feb | Sep | Feb | Aug | Dec/Jan | May | Jun | Jul | Aug | Sep/Oct | Dec |

IDN Protocol Released

Implementation by gTLD and ccTLD registries

Spoofing Concerns Arising

Guidelines revision Initiated

Protocol revision Initiated

DNS Test in Laboratory Completed

ICANN IDN Program Initiated

.test IDN evaluations plan

IANA IDN TLD process

Emergency removal process

Board Approval:
-.test delegation
- .test plan

RSSAC & SSAC Statements

Protocol Revision Finalized

Guidelines revision v.2.2

ccNSO-GAC Issues Paper

Launch of IDN Live Evaluations & Initial Results

More details available at:

http://icann.org/topics/idn

18

Thank You

Tina Dam
tina.dam@icann.org