

GAC-ccNSO Joint IDN Session



Tina Dam
Director, IDN Program

25 June 2007

San Juan, Puerto Rico



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



Some 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 transformed by IDN encoding; for example "icann" in the domain name "icann.org“

Source: extraction from: <http://www.ietf.org/internet-drafts/draft-klensin-idnabis-issues-01.txt>

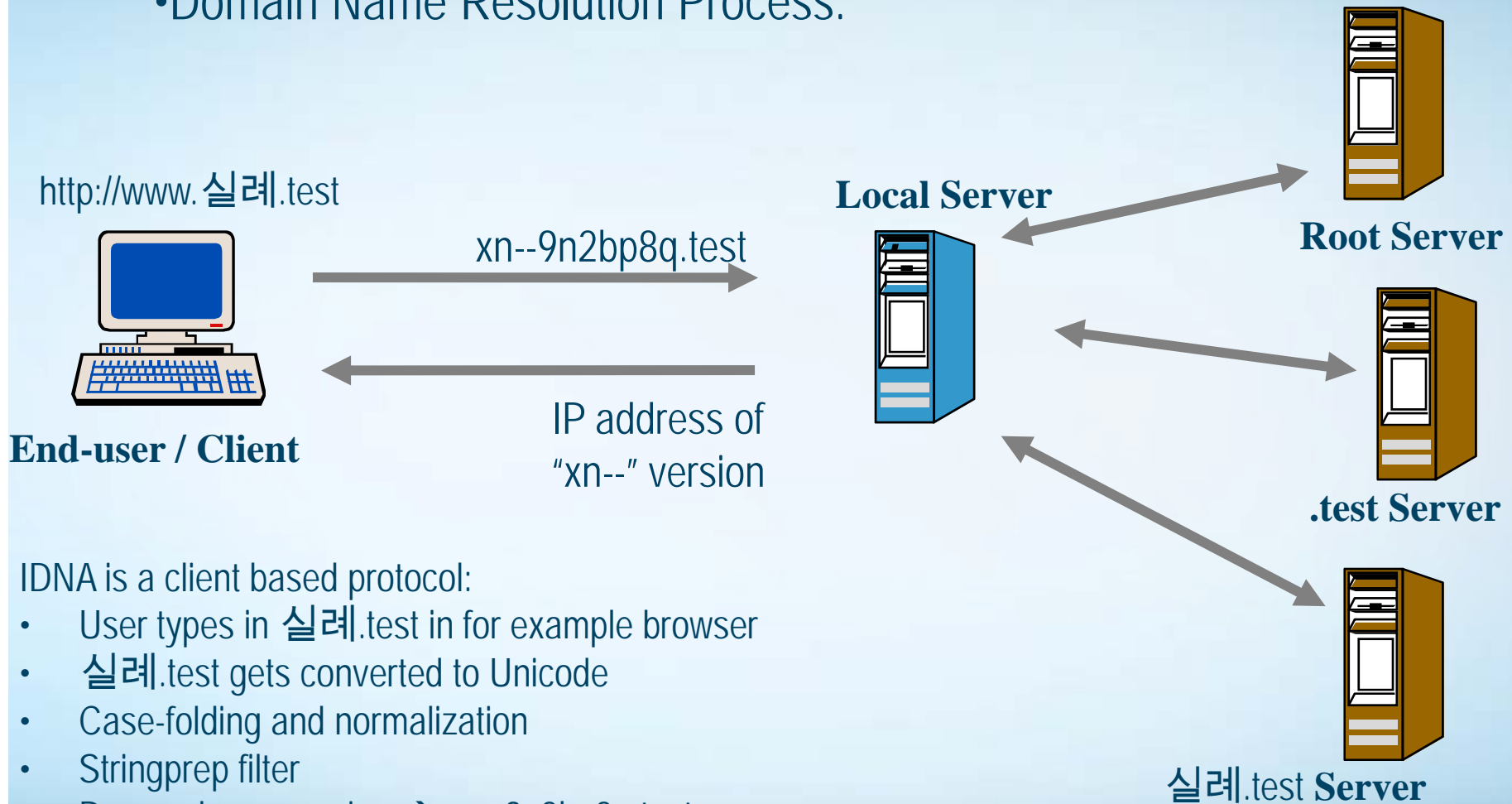
- **LDH (Letter, Digit, Hyphen):** The hostname convention defined in RFC 952 (later modified by RFC 1123) restricts domain names to the letters a-z, digits 0-9 and the hyphen "-" (despite the DNS protocol permitting all other printable ASCII characters).
 - The term "LDH code points" refers to this subset.
 - With the introduction of IDNA this rule is no longer relevant for all domain names as they are displayed, although what is actually stored in the DNS remains LDH.



IDNA – Protocol Functionality



•Domain Name Resolution Process:



IDNA is a client based protocol:

- User types in 실례.test in for example browser
- 실례.test gets converted to Unicode
- Case-folding and normalization
- Stringprep filter
- Punycode conversion → xn--9n2bp8q.test



Characters in the DNS

- Search on “*US-ASCII character set*”
- The DNS can handle all US-ASCII characters
 - Examples:
 - (a...z), (0...9), (-)
 - () SPACE
 - (!) EXCLAMATION MARK
 - (") QUOTATION MARK
 - (#) NUMBER SIGN
 - (\$) DOLLAR SIGN
 - (%) PERCENT SIGN
 - (&) AMPERSAND
 - etc...

Character set and the IDNA



- Character set: A standardized ordered list of characters, for example:
- Unicode is a commonly used encoding scheme that
 - provides a unique number for each character across a wide range of scripts that are used for writing a large number of languages
 - entabulates "code points" (unique numbers) for each of the individual characters
 - the tables continues to expand as more and more characters are encoded
 - the code points are commonly represented in a hexadecimal notation
 - for example, the word "Hello" is written U+0048 U+0065 U+006C U+006C U+006F
- 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
- Examples:
 - In Czech, <ch> is a single character whereas in English it is two
 - In Danish, <æ> is the 27th letter of the alphabet. It is a single character and does not decompose to <a e>
 - In Turkish, there is a difference between a dotted <i> and a dotless <ı>. In English there is no such distinction. Is the dot to be counted as a character in its own right, or is it not?

Localization vs Internationalization

- Localization refers to the adaption of a product, application or document content to meet language, cultural and other requirements of a specific target market
- Internationalization is the design and development of a product, application or document content that enables easy localization for target audiences that vary in culture, region, or language

Source: <http://www.w3.org>

- Labels needs to be localized
- The DNS need to be internationalized

Principles

- **Overarching principles to ensure the stability and security of the Internet**
 - Global uniqueness and interoperability of the DNS
 - Unique and unambiguous domain names with the same functionality regardless of the geographic point of access
 - Promote “Future-Proof” solutions
 - Define characters that are allowed, and provide for the addition of new ones
 - Not all characters used in the worlds’ languages can be available for use
- **Principles related to operation**
 - Diminish user confusion as much as possible
 - via technical development and implementation specifications, registry policies, and user education
- **Principles related to PDP process**
 - Promote multi-stakeholder involvement in policy development
 - ICANN supporting organizations and advisory committees are core for policy development

Thank You

<http://www.icann.org/topics/idn>