Internationalized Domain Names

Introduction & Update

MENOG 1
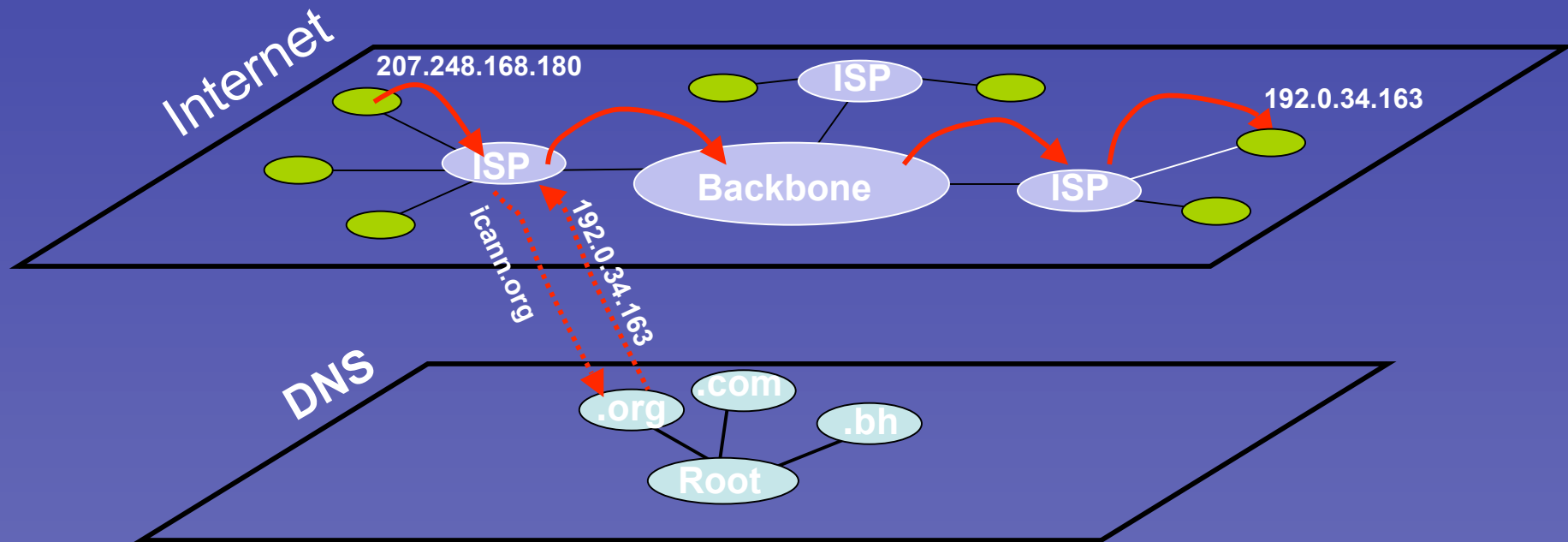Bahrain
April 3-5, 2007

By: Baher Esmat
Middle East Liaison

ICANN

IP and DNS

DNS is based on ASCII: Letters (a-z), Digits (0-9)
and Hyphen (LDH)

Need for "Internationalized" DNS

- As the Internet continues to grow, many people around the world wish to go online using their native languages
- But, is "internationalized" DNS the only solution?
  - No, it is one among others that address other areas such as e-content, application interfaces, search engines, etc…
- Does "internationalized" DNS mean a new DNS protocol?
  - Transforming the existing ASCII-based DNS into a Unicode-based DNS would have risked the stability of the entire DNS
  - The solution is to make the conversion from non-ASCII to ASCII at the user / application level (web browsers, email clients)
  - IDNA specifies how this conversion can be done

ICANN

Internationalized Domain Names (IDNs)

- "Internationalization" & "Localization"
  - Solving a "local" problem with a "global" solution or
  - Dealing with an "international" matter by adopting and implementing "local" tools
- Interoperability and consistency in "resolving" names is a prerequisite
- IDNs allow end-users to register and write down domain names and addresses using non-ASCII strings

ICANN

What is an IDN – User Perspective

- All you need is the name you want to register

- Registries will supply a list over available characters, usually in Unicode

- Registries will handle all encodings needed during registration process

- Example: ‎باهر‎.eg → xn--mgbb2a6f.eg

- Encodings tools:
  - http://josefsson.org/idn.php
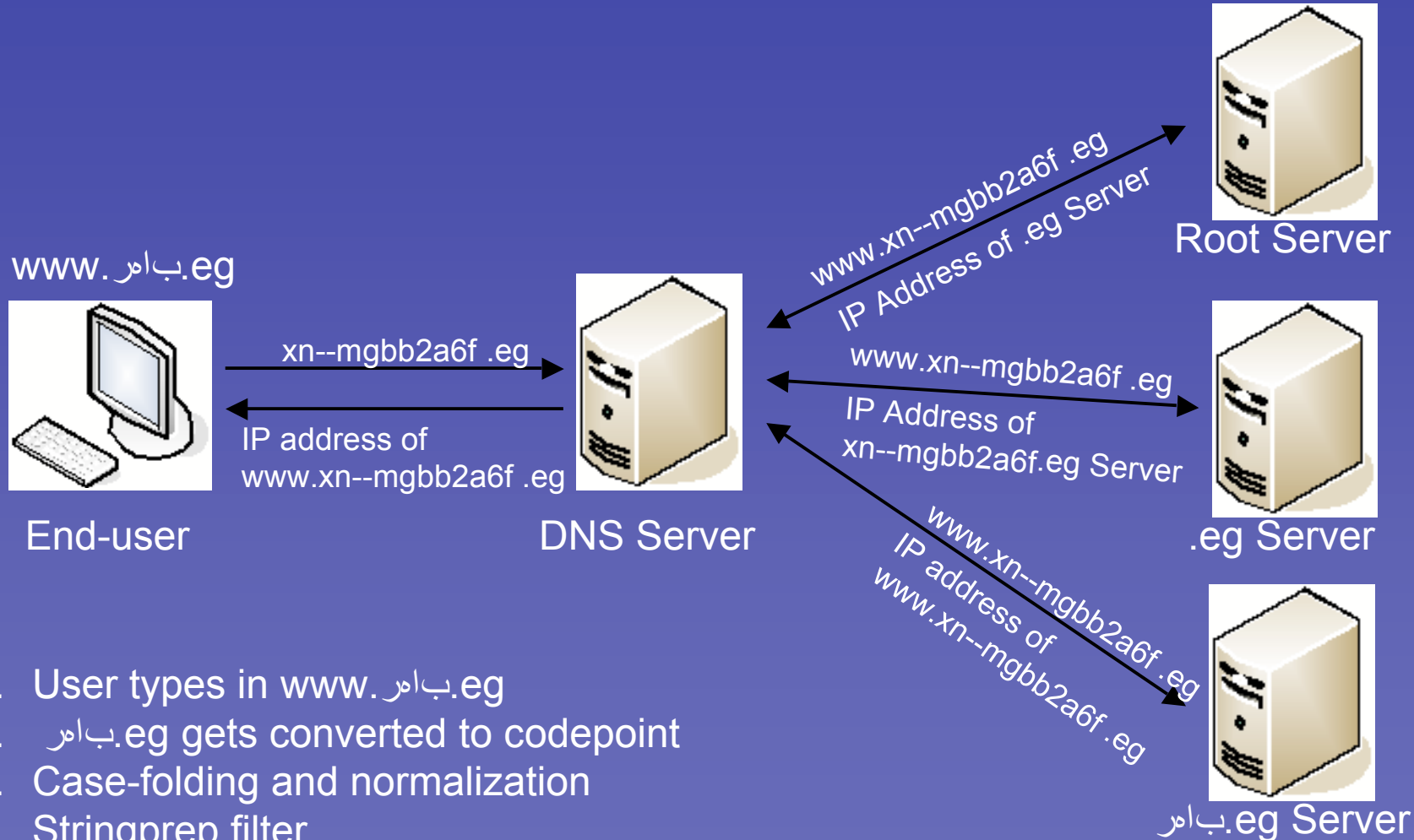 - http://mct.verisign-grs.com/index.shtml

ICANN

The Solution

- Technical specifications:

  - (RFC 3454) *Preparation of Internationalized Strings ("stringprep")* - a framework of processing rules for Unicode text

  - (RFC 3490) *Internationalizing Domain Names in Applications (IDNA)* – a mechanism for handling non-ASCII labels

  - (RFC 3491) *Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)* – processing rules that allow end users to enter IDNs into applications

  - (RFC 3492) *Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)* – an encoding algorithm that allows a string of basic code points to uniquely represent any string of code points

ICANN

Proposed Revisions to IDNA Protocol

- Effort led by the IETF

- The basic framework of the revision has already been published in RFC4690 and has addressed key issues:

  - Revision to support Unicode 5.0
  - Language specific character issues (same script, different language)
  - Multiple usage of scripts for one language
  - Bi-directional cases (right-to-left scripts)
  - Visually confusable character issues

- Three Internet drafts were published providing suggestions for solutions to the issues raised in RFC4690:

  - An overview with proposed issues and changes for IDNA
    - http://www.ietf.org/internet-drafts/draft-klensin-idnabis-issues-01.txt

  - A suggestion for solving an IDNA problem in right-to-left scripts by revising the stringprep profile
    - http://www.ietf.org/internet-drafts/draft-alvestrand-idna-bidi-00.txt

  - An overview of suggested inclusion based IDNA Unicode code points based on Unicode 5.0
    - http://www.ietf.org/internet-drafts/draft-faltstrom-idnabis-tables-01.txt

ICANN

How IDNA Works

www.بامر.eg

xn--mgbb2a6f .eg

IP address of
www.xn--mgbb2a6f .eg

End-user

DNS Server

www.xn--mgbb2a6f .eg
IP Address of .eg Server

Root Server

www.xn--mgbb2a6f .eg
IP Address of
xn--mgbb2a6f.eg Server

.eg Server

www.xn--mgbb2a6f .eg
IP address of
www.xn--mgbb2a6f .eg

بامر.eg Server

1. User types in www.بامر.eg
2. بامر.eg gets converted to codepoint
3. Case-folding and normalization
4. Stringprep filter
5. Punycode conversion → xn-- mgbb2a6f .eg
6. (Registry prohibition list)

ICANN

IDN Working Groups and Activities

- ICANN President's Advisory Committee for IDNs
  - Formed on 23 November 2005
  - Initially tasked with IDN TLD technical issues
- Supporting Organizations and Advisory Committees
  - GNSO, ccNSO, GAC, ALAC
- IDN TLD Registries
  - ccTLDs and gTLDs
- IETF and IAB
- Application Developers
- National & Regional Initiatives

ICANN

User Confusion and Spoofing Issues

- IDNs expanding risk of known problems

- Many characters can be confused with others
  - Problem exists in ASCII as well
    - Digit "1" and lower-case "l"
    - Digit "0" and upper-case "O"
  - IDNs increasing the character collection
    - From 64 in ASCII (LDH)
    - To tens of thousands in Unicode (nameprep)

- Well-known example: paypal.com
  - Second character is U+0430, Cyrillic small a
  - Looks like Roman/ASCII "a"
  - Would have been prevented by "one label, one script" rule

- This kind of confusion creates opportunities for user mistakes and frauds

ICANN

Internationalized TLD Principles

- Global uniqueness and interoperability of the DNS
  - unique and unambiguous domain names
  - URLs and emails connect as expected regardless of geographic placement of access

- Promote "Future-Proof" solutions
  - Define Unicode characters to be allowed
  - Provides ability for adding new languages, new characters far in the future

- Diminish user confusion
  - Technical limitations
  - Implementation requirements
  - Registry restricted list and policies
  - User education

- Promote multi-stakeholder involvement

ICANN

ICANN IDN Program Plan

- A program established within ICANN to achieve the possibility to insert internationalized top level labels in the root zone

- Comprised of several projects
    - Technical tests
    - IDNA protocol revision
    - IDN Guidelines
    - IDN Repository
    - Production deployment
    - Policy development
    - Outreach and Communication

ICANN

IDN Laboratory Test

- Implemented by Autonomica in coordination with ICANN IDN-PAC

- Test plans included :

  – Insertion of NS records into a copy of the root zone

  – Tests performed in closed laboratory environment with a series of systems implemented to replicate as closely as possible the server software of the various root servers:

    • Versions of BIND server software

    • Use of the most popular DNS resolver software packages

  – No end-user application testing was included

ICANN

IDN Laboratory Test Strings

- Localized labels for testing IDNs
  - http://www.icann.org/topics/idn/idn-test-labels.pdf

- Normal Unicode-Punycode conversion
  - ‏النہر8فرس1‎ → xn--18-dtd1bdi0h3ask

- Performance with a 63-character long TLD string
  - .hippo18potamushippo18potamushippo18potamushippo18po

- Right to left script

- Left to right script with sophisticated shaping properties

- Non-alphabetic script

ICANN

Test Results and Next Steps

- Laboratory test of root zone and resolver software has successfully been finalized
  - http://www.icann.org/topics/idn/idn-report-15feb07.pdf
- Controversial views from community on the steps going forward
- Technical study is shortly to be carried out by SSAC
- The goal is to insert internationalized labels in the root zoon without risking stability and security of the Internet

ICANN

IDN Policy Issues

- ccNSO, GNSO, GAC IDN working groups have been working on a number of open questions such as:
  - Should an "equivalent" ISO 3166 list be developed for IDN ccTLD strings?
  - How many IDN ccTLD can a territory have?
  - Who can apply for the IDN ccTLD?
  - Are there any ownership rights over languages?
  - How to introduce IDN to existing and new gTLDs?

ICANN

IDN Links

- IDN information area
  - http://www.icann.org/topics/idn/
- Calendar
  - http://www.icann.org/topics/idn/meetings.htm
- News feed
  - http://www.icann.org/announcements/announcement-05oct06.htm
- Mailing lists and public forums
  - http://www.icann.org/topics/idn/fora.htm

ICANN

Thank You

Baher Esmat
baher.esmat@icann.org

ICANN