ISPCP Notes – Marrakech ICANN Meeting
IDN Technologies Workshop

Tina Dam – Status Update on ICANN IDN Activities

In May, ICANN staff discussed the input they recieved on the TLD label experiments and started to generate a new plan for testing IDNs in the root. The staff went to other, external experts to help develop a revised plan. In June, ICANN staff revised the proposed plan for presentation to stakeholders and developed a process for finalizing the revised test plan.

The IDN program plan has several projects that are planned separately but that have several interdependencies:

- Technical and operational test
- Policy development
- IDN guidelines
- IANA processes
- Outreach planning
- Communication planning

The overall goal of the technical and operational test plans is to demonstrate that the insertion of IDN strings into the root has no appreciable negative impact on existing resolutions. Further consulting with IETF and SSAC will be done to make sure the proposed test plan meets this goal.

The proposed test plan includes interrelated milestones where some of the activities run in parallel:

- NS records based on Punycode
    - Perform tests in laboratory setting
    - Perform operational process test
    - DNS Root Name Server test
- DNAME Resource Record testing
    - Analysis of functional and practical implications

There is a proposed process for finalization of the new plan for testing. This process will not be made public until later in this meeting.

John Klensin – On the Current State of IDNs

What people started to realize when initial implementations took place is that the original technology is not completely correct. This is true from both a technical and policy point of view. The state of IDN deployment is such that we probably have once chance to get it right.

Klensin focused on issue identification rather than resolution during his talk.

The IAB did a report to identify issues and, in some cases, propose possible solutions. Some issues do not have solutions inside the DNS.

Klensin said that an IDN is one label in an domain name that consists of multiple labels. The original character set is built on the elderly ISO 646.

IDNs are solution to the problem of better mnemonic value for names in non-Latin scripts. This does not address content availability, connectivity and access, user friendly URLs or the ability to understand each other's languages. The DNS only allows for exact matching -- not "close enough" or "do you mean" options. The DNS is case sensitive in what it stores, but the matching and queries are case-insensitive.

There is a strict administrative hierarchy in the DNS. The aliasing system is very inflexible (for instance, it cannot do: "see this and see also"). Names in the real world are made up of languages, dialects and scripts. With regard to name and character matching, humans are far better than the DNS. The DNS doesn't have enough information to even try most typical approaches.

IDNA encodes IDNs into DNS. The first step is to take a Unicode string and make it into another Unicode string using a technology called nameprep. The nameprepped unicode is then made into Punycode (which looks like a classical ASCII name).

There are few outstanding technical problems. Only one major browser does NOT support IDNA. Other applications do not support IDNA (mail clients, etc.).

Using IDNA is a problem:
- There is the problem with character spoffing and similiarities. This is something that cannot be fixed techically and it is difficult to design policies that help for a great number of cases.
- Transcription from written form is difficult to do in an unambiguous way.
- Human and DNS expectations do not match.
- When characters get more complicated than ISO 646IRV the solution requires the use of tables. The character list inevitably expands over time. Matching new and old characters, and new and old tables, is going to be version sensitive.
- There are global issues related to transcribing URLs – a rule that says there should be one script per label does not fix this.

The proposed "variant" model works like this: within a given domain, collect the labels that contain similar characters, register one and then block all the others: all of them

must be registered by the same organization. This is happening in China, Japan, and Korea. Note that the "variant" proposal only has impact on storage, not on queries.

There is a proposal to have separate matching trees for different languages. This is unlikely to work at lower levels in the zone because there will be differences in the names stored in each zone. This will, possibly, pose problems for interoperability.

Making nameprep interoperable across unicode versions is also difficult. If nameprep is not stable then it is not strictly upward compatible. Migrating from one version of Unicode to another is hard because the mapped names will be different.

According to Klensin, the IETF needs to do a full IDNA review. This will include a more restrictive nameprep and a mechanism for backward compatibility with older versions of Unicode.

Klensin said that several changes have to be made. These changes may invalidate now-valid names. Any prefix change would be radical and would require software changes and careful study. He also noted that there will be new kinds disputes and dispute resolution issues. Decisions by registries imply registry responsibility. Technically, each registry can have different policies about permitted names in the IDN space.

IDNs in the TLDs
- Naming and Delegating Decisions in IDN TLDs
- Multiple Labels for the "same" TLD
- Coding and Presentation questions

Klensin claimed that we need to reduce the permitted character list in the future. Also, we need to update to Unicode 5.0 and do this in a very general way. Also, Klensin asks that there be analysis of non-DNS and above-DNS solution.

Thomas Narten – IDNs from the IETF Point of View

The IAB has published a "Review and Recommendations for Internationalized Domain Names (IDN)." This was finished last week. (approved by the IAB on June 23)

At the upcoming IETF meetings in Montreal, this will be a topic in the applications area meeting. In those meetings the IETF will discuss what it will do with regard to questions in the IDN document.

The DNAME specification is in RFC 2672. There is some deployment experience with this in the DNS, but nothing depends on it operationally. Simultaneously, much work has been done on DNSSEC during the same period. During discussions, many "what happens when DNAME is in use" came up. As a result, the IETF is reopening work on DNAME. The general IETF observation, however, is that it may not be broken.

Michel Suignard – Microsoft and Implementation Notes

IDNA status at Microsoft includes appropriate drivers being provided by platform services in Windows XP and Vista. There is support for the IDNA RFCs in lower level software – this will be used in IE7 and in the forthcoming version of Outlook.

Microsoft has worked hard to allow a user to specify a locale of their choosing and then support characters for that country. Microsoft also supports the concept of mixed scripts. This is very common in Japan. The problem is preventing homograph spoofing attacks.

Microsoft says that IDNA cannot support improvement beyond Unicode 3.2 – this means that certain languages and scripts are not supported. There are also scripts that are going through a major revision since Unicode 3.2. Microsoft also notes that there is no serious security threat mitigation.

Microsoft suggests that the community should:
- Extend support to Unicode 5.0 or even future versions of Unicode
- De-emphasize the role of the complex IDN nameprep process
- Focus on the output list instead
- Restrict problematic characters from the IDN namespace
- Standardize the IDN namespace as an ISO 10646 character collection
- Establish script based guidelines for constituencies with worldwide reach

The guidelines for success, according to Microsoft, are a worldwide name space, multi-script environments, and a secure environment.

Ming-Cheng Liang – TWNIC – EAI (Email Address Internationalization)

The format of email addresses is **local-part**@domain-part. The domain-part is handled by IDNA. Liang notes that there is no standard yet for the local-part of the domain name. There needs to be support for IDNs mixed with ASCII in both the local-part and the domain-part.

The problems include mis-interpretation in the store-and-forward model. Many mail servers also look at the local-part to make decisions about forwarding.

It looks like Korea, China, and Japan are taking the lead on this. In March of 2006 the IETF started the EAI Working Group. The EAI WG has defined a Framework, a SMTP Extension and some new UTF formats for the header of email. SMTP clients can handshake with the server to see if the server supports the extension for EAI. If not, it will be downgraded to a ASCII address.

TWNIC is working on the test plan for EAI and developing a plug-in for some famous email clients.