

10

Internationalization of the Domain Name System: The Next Big Step in a Multilingual Internet

Tan Tin Wee and S. Maniam

National University of Singapore, Singapore 119260
i-DNS.net International Inc, Menlo Park, California, USA
Email: s_maniam@pacific.net.sg

Abstract

Every machine on the Internet has an Internet address (IP address). The Internet domain name provides a meaningful and easy-to-remember handle for an IP address. The task of mapping domain names to IP address for all the Internet address is performed by the worlds, most extensive and scalable database system, the Domain Name System (DNS). Almost every common Internet application calls on the DNS to resolve a domain name into an IP address.

Limitations in the DNS and its operation by convention restrict the characters used in domain names to A-Z, a-z, 0-9 and - of the ASCII Latin set. Multilingual characters are not supported. Therefore, even though the content of a webpage or an email may be in a native script of a non-English language, the address cannot be rendered in that native script.

The process of supporting multilingual script and other linguistic and cultural needs on the Internet is generally known as Internationalization. The Internationalization of the DNS system is potentially a gargantuan task of upgrading the DNS protocols that use domain names, which is practically every commonly encountered Internet application, and an overhaul of the DNS servers currently installed worldwide. This process, first initiated in 1998 by a Chairman's Commission of the Asia Pacific Networking Group (APNG), led to a standards working group in the Internet Engineering Taskforce (IETF) and an international consortium comprising industry, academia and regulatory authorities called the multilingual Internet Names Consortium (MINC). Meanwhile, more than a dozen alternative multilingual DNS technologies, known generally as iDNS, have emerged and starting in late 1999, the first company, i-DNS.net International Inc., deployed in Chinese and by early 2000, Tamil was the second language to be commercially launched. Since then multi-lingual domain names, now known as IDNs or Internationalised Domain Names, have been deployed in over 60 languages with Asian, Middle Eastern, Western and Eastern Europe, Latin American origins. The advent of the mass deployment IDNs by Verisign - the company that controls the 30 million ASCII "dot com" and 'dot net' domain names amounting to more than half of all domain names in use today - in late 2000 in over 60 languages, the number of IDNs registered has exceeded 1 Million. Today, between Verisign and a number of other entities that have deployed IDNs, this number is approaching 2 million, with the majority coming from the Chinese, Japanese and Korean languages. With the recent finalisation of the IDN standard by IETF in June 2003 and the announcement since by MINC of

an Interoperability Testbed to ensure compatibility of the different deployments centered around the released IDN standard, not only is the number of IDNs in use expected to grow dramatically in the coming year but the related nascent deployments of multi-lingual email addresses is also set to explode by mid-2004.

1. Multilingual Limitations of the current DNS system

The Internet Domain Name System (DNS) was developed over time to facilitate easy recall of Internet address. It achieved this via the matching of easily remembered alphanumeric strings such as `www.yahoo.com` to the string of numbers of Internet Protocol (IP) address (version 4) eg. `137.132.19.1`. Unfortunately, in the pursuit of universality, only ASCII alphanumeric characters plus the- were acceptable characters in the domain name strings. Subsequently, because of the replacement of meaningless numbers with meaningful memorable words, this DNS convention was universally adopted as the global standard for all hostnames, email address, Web addresses, e-mail address and other Internet addressing formats. For instance, until now, no provision has ever been made to allow input domain names in Web address in a non-ASCII non - English script. This meant that any user of the Internet had to have some basic knowledge of ASCII characters (i.e. English language and languages which use basic Latin characters).

While this does not pose a problem to the scientific, technical or business user who is able to understand English as an international language of science, technology, business and politics, it is a major stumbling block especially in countries where English is not widely spoken. This hindrance contributes significantly to the digital divide as the Internet is becomes the standard way for doing business, and slows the popular adoption of the Internet in the business community, particularly in countries struggling to keep up with the pace of development in developed countries.

What has meant to be a memorable way of reaching a website to the English-speaker, has now turned out to be a linguistic barrier to the Internet. In many instances, it may even be easier for a non English-speaking person, say, in village in China, to remember a string of numbers rather than a series of unfamiliar and unintelligible Roman alphabets.

In countries such as these, understanding English has become a daunting prerequisite to performing such as basic activities as posting e-mail or accessing Web pages that are otherwise composed in the local language. The irony is that most software applications today already support a robust local language environment, but access to the Internet, whether email or websites, invariably requires use of ASCII alphabet.

The Internet already encompasses a network of communities representing a global mosaic of languages and cultures. The increasing volume of modern business, research, and interpersonal communications in non-English languages is a testament to this fact. Clearly, the existing DNS has become an anachronism in an already multilingual Internet world.

2. Internationalization of the Domain Names System

In response to the multilingual demands created by the natural evolution of the Internet, the process of internationalization of the Domain Name System (iDNS) had begun.

The task to upgrade the DNS is very complex. The DNS protocol impacts on many Internet protocols. Common protocols such as those supporting the web and email all involve the DNS. Client applications such as Web browsers and email clients have to be changed. The server software in DNS servers throughout the world has to be modified as the DNS system is arguably the world's most extensive, hierarchal, distributed and scaleable database.

In early 1998, the first author (TTW) initiated research at the Internet Research and Development Unit, National University of Singapore, into tackling this, the possibility of setting up a DNS which is as backward compatible as possible, future extensible, and does not involve too much client application or server-side modification. The iDNS proxy server was developed which intercepts multilingual characters send from client applications and converts them into an ASCII compatible encoding (ACE), and forwards them to conventional DNS servers for IP address resolution. This early experiment used the Unicode Transformation Format UTF5 by Martin D1000rst as the ACE.

Based on this success, as the Chairman of the Asia Pacific Networking Group (APNG), the first author initiated a Chairman's commission in July 1998 to look into the challenge of formally internationalizing the Internet DNS protocol.:

(<http://www.apng.org/commissiun/odns>).

An Asia Pacific development team to create a more robust multilingual domain name server- iDNS and to implement a multilingual domain name space- iDOMAIN as an international test bed beginning in the Asia-Pacific.

From August to December 1998, the chairman visited Hong Kong, India, Korea and China, demonstrating the pre-prototype system and managed to gather international interest in this project.

In December 1998, the Chairman managed to recruit the participation of an Industry partner, Bioinformatrix Pte Ltd (BIX Pte Ltd), as an early pioneer entrant, willing to risk being an early player and to sponsor servers and technical expertise.

Meanwhile, a grant application by APNG jointly with the CIR, to the international Development Research Center (IDRC), a Canadian Government funded international organization, met with success. This APNG Project is funded under the Pan Asia R&D Grant administered on behalf of IDRC by the Canadian Committee on Occupational Health and safety (CCOHS).

By the end of 1999, the early technologies created by BIX Pte Ltd, a University spin-off company was incorporated into the company, i-DNS.net International Inc.

Meanwhile, in the IETF, work started with the formation of an Internationalised Domain Name IDN Working Group, beginning with the task of identifying the basic requirements of an IDNS. After almost 3 years of effort and debate, this working group has now (June 2003) developed and published an Internet standard that is likely to be universally adopted by the Internet Architecture Board (IAB) and ICANN.

By early 2000, lead by the Chairman of Asia Pacific Top Level Domain APTLD forum, Professor Kilnam Chon, the Multilingual Internet Names Consortium (MINC) taskforce was formed to spearhead the formation of MINC. By June 2000, MINC was launched with an Interim Board and 11 Founding members and more to come. The consortium draws membership from industry, academia, regulators and government bodies. Its aims are to promote, facilitate and coordinate the development of multilingual Internet names, especially multilingual Internet domain names. At least a dozen such solutions are currently available. With the recent standardisation of IDN by IETF, MINC has announced an Interoperability Testbed for the providers of these various solutions to ensure compatibility with the announced standard. This testing will aim to unite the process of achieving a backward compatible system that preserves the global DNS so that the solution can be universally adopted without breaking down the

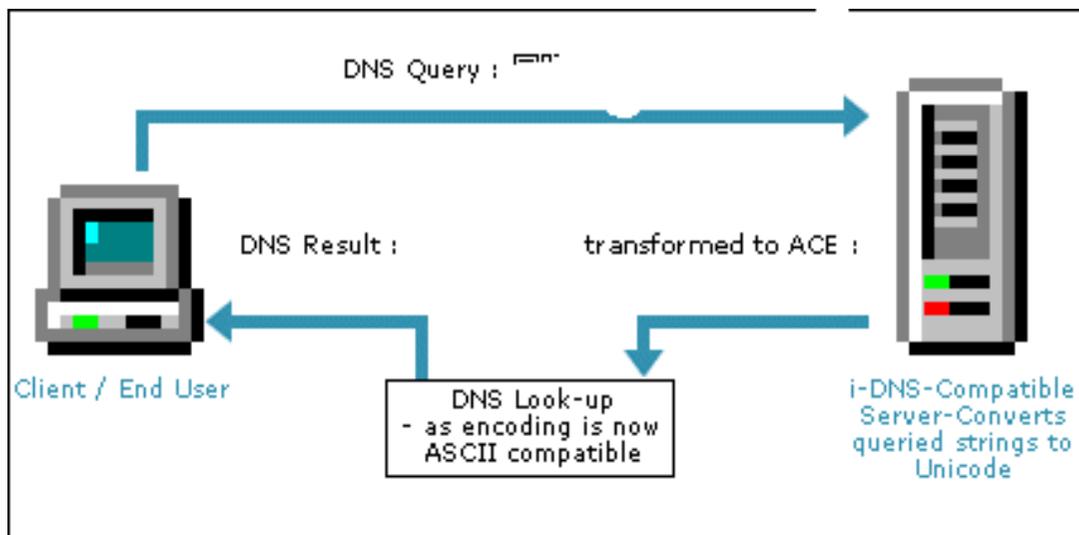
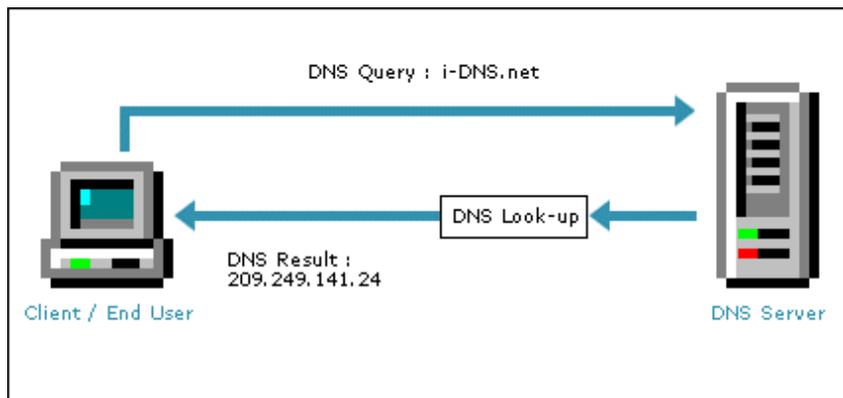
existing framework that has served the Internet so well in the past. In addition, various Language Working groups within MINC have submitted policy standards specific to these language to supplement the IETF general standard/ These groups are also preparing to participate in the Interoperability Testbed in the coming months. In fact, INFIIT via the MINC Tamil Working Group has already announced its intention to participate in the Testbed and is finalising a first draft of the Tamil Language specific policy/guideline standard.

3. Some technical considerations

How the Current DNS System Resolves Domain Name Queries

For example, we input typically `www.iDNS.net` instead of `209.249.141.24` whether it is for a website address or an email address.

Upon receipt of a query via a client/end user's browser, the DNS server checks its cache to see whether there is any way to resolve the name, and returns the IP to the client if successful.



If it cannot resolve from the data that is within its own server cache, it will send a request to a DNS root server. The latter will determine which Top Level Domain (TLD) or Country Code Top Level Domain (ccTLD) the domain name belongs to. The TLD or ccTLD domain server will then determine which server it should forward the data to for further resolving.

Once this particular domain is found, the IP is sent back via to the client (end-user browser). The client can now access the desired information from the IP of choice.

In the same way, with multilingual domain names technology and a revision of the Internet standards governing the domain name system. A Tamil language name can be supported. For Example: www.infitt.org in Tamil as உத்தமம் (202.172.239.14)

The direct implication of deploying this kind of technology is that users can finally enjoy a truly internationalized Internet experience. From input of identify, Web address or domain name, to the authoring, scripting, processing and viewing of content, users can go through the entire process seamlessly in their language of choice. On a larger scale, iDNS is an enabling technology. Vast new e-commerce markets, like the use of iDNS to enable multi-lingual email addresses for non-English speakers, will be opened as the power of the Internet is unleashed in countries where knowledge of English is not widespread.

Conclusion

The Internationalization of the Internet Domain Name System is well on the way. The protocol standardization will bring in an orderly transition from the old system to the new. Many multilingual registrars will be able to offer multilingual domain names registration to organizations, e-commerce companies and individuals. This will facilitate the adoption of the Internet in places where English is not generally understood.

In the Multilingual Internet Names Consortium (MINC) (<http://www.minc.org/>) we have a Tamil Domain Names Working Group to discuss DNS internationalization issues of Tamil (Mailing List: tamil@minc.org). and in INFITT WG3 is formed to discuss on Internet Names such as GTLDs and CCTLDs.

The INFITT WG3 has in principle agreed in late 2002, to participate in the recently announced MINC Interoperability Testbed, to ensure that the IDN technologies deployed in Tamil are compatible with the IDNA standard. To this end the groups are also working on Tamil specific supplemental standards. After such testing, the pace of adoption of Tamil IDNs and email addresses should increase rapidly.