A REVIEW OF METHODS FOR INTERNATIONALIZED ACCESS TO DOMAIN NAMES

Sung J. Shim, Stillman School of Business, Seton Hall University
400 South Orange Avenue, South Orange, New Jersey 07079
Email: shimsung@shu.edu, Telephone: 973-761-9236

Arun Kumar, Div. of Systems & Engineering Management, Nanyang Technological University
Nanyang Avenue 50, Singapore 639798
Email: makumar@ntu.edu.sg, Telephone: 67906181

ABSTRACT

Domain names are valuable to Internet users as a main identifier of entities and resources on the Internet. This paper reviews the methods proposed for internationalized access to domain names, focusing on their technical implementation and potential problems.

INTRODUCTION

Until now, domain names have been registered and accessed only in English scripts, more specifically, a subset of ASCII characters that includes the twenty-six English alphabets (case-insensitive), ten decimal digits (0-9), and hyphen (-). This limitation may diminish the international scope of the Internet. Given the increasing number of non-English speaking users on the Internet, it is highly desirable to allow using not only domain names in ASCII scripts but also domain names in non-ASCII scripts, called internationalized domain names (IDNs), that are less universal but easier to remember and use in local contexts.

The Internet community has begun to recognize the growing need for internationalized access to domain names, and several methods have been proposed to meet the need. The Internet Corporation for Assigned Names and Numbers (ICANN) Board of Directors recognizes "that it is important that the Internet evolves to be more accessible to those who do not use the ASCII character set" [http://www.icann.org/committees/idn/]. The IDN Working Group of the Internet Engineering Task Force (IETF) has recently developed and specified a set of standards for IDNs. Several domain registries and their domain name registrars have already started to provide services for IDNs based upon the standards developed by the IETF's IDN Working Group or other methods. Several companies are also providing multilingual keyword services using their own methods and servers.

The ability to support international access to domain names on the Internet is a significant development for non-English speaking users on the Internet, but it also raises various issues from those related to technical implementation all the way to those related to languages and policies. In an effort to understand the opportunities as well as challenges involved in supporting internationalized access to domain names, this paper reviews the methods proposed for internationalized access to domain names, focusing on their technical implementation and potential problems.

DOMAIN NAME SYSTEM

On the Internet, data is transmitted through a series of protocols known as the Transmission Control Protocol/Internet Protocol (TCP/IP). Each computer connected to the Internet has a unique IP address that identifies the computer as a host on the Internet. But IP addresses, which are composed of four decimal numbers separated by dots, are not easy for users to remember and use. Thus, the DNS provides a user-friendly, mnemonic equivalent of each IP address. The DNS organizes hosts on the Internet by dividing the Internet into a series of component networks called domains, and each host belongs to one of these domains. The DNS overlays the IP address of each host with a unique string of characters called a domain name. Each domain name corresponds to exactly one IP address. By providing an efficient way of mapping each domain name to its corresponding IP address, the DNS facilitates the user's ability to navigate and communicate on the Internet. The DNS functions globally, but it still recognizes and resolves domain names only in ASCII characters.

The DNS resolves domain names into IP addresses as follows. First, the user enters a domain name into the user application. Second, the user host sends the domain name request to its local DNS server. The local DNS server keeps track of domain names and the corresponding IP addresses of its designated group of hosts. Third, the local DNS server queries the domain name in its file. If the local DNS server recognizes the domain name in its file, it sends the corresponding IP address back to the user host. If not, the local DNS server passes the request on to the nearest root DNS server. If the root DNS server recognizes the domain name in its file, it sends the corresponding IP address back to the local DNS server, which in turn sends the IP address back to the user host. Finally, using the IP address following the TCP/IP, a connection is established between the user host and the server host to which the user requested an access. This process of domain name resolution in the DNS is completely transparent to the user. Of course, the user can enter directly an IP address at the user application and get the result without any contribution by the DNS.

METHODS FOR INTERNATIONALIZED ACCESS TO DOMAIN NAMES

In order to fully support internationalized access to domain names, it is necessary to upgrade all systems from user applications to the DNS to support Unicode. Upgrading the DNS to support Unicode, however, may be very difficult, if not impossible. The current DNS is massively distributed with numerous functions, applications, and systems involved. Along with the growth of the Internet, the DNS has come to scale with numerous functional expansions to its original design and to play a number of supplemental roles beyond its original intentions. In addition to such technical involvedness of the DNS, it may be very difficult to change an existing standard or to enforce a new one on the Internet, mainly due to the lack of a central regulatory body. It may be impossible to regulate the Internet because of the complexity and openness of its structure. For these reasons as well as others, upgrading the entire DNS to support Unicode does not seem to be a plausible option in a short term, while it may be considered as a long-term solution.

In fact, most methods that have been proposed to date to support internationalized access to domain names prefer to keep the DNS with no changes to its existing infrastructure and

introduce some other ways to map Unicode scripts that are entered into user applications into unique ASCII strings that can be recognized and resolved by the DNS. These methods still allows the DNS to use domain names in ASCII characters, while they enable users to use domain names in Unicode scripts in their applications. In general, these methods can be classified into two groups based on the way to map Unicode into ASCII characters: ASCII-compatible encoding and multilingual keywords.

ASCII-Compatible Encoding

The ICANN stresses that "the internationalization of the Internet's domain name system must be accomplished through standards that are open, non-proprietary, and fully compatible with the Internet's existing end-to-end model and that preserve globally unique naming in a universally resolvable public name space" [http://www.icann.org/committees/idn/]. Following this guideline, the IETF's IDN Working Group recently developed and specified a set of three standards for IDNs including: (1) a standard for using the encoded strings in the DNS, (2) a standard for preparing IDNs, and (3) a standard for the encoding scheme. The IETF published the three standards as RFC 3490 [2], RFC 3491 [3], and RFC 3492 [1]. The ICANN already endorsed the standards and authorized registration of IDNs based upon the standards. The IETF's publication and the ICANN's endorsement of the standards would facilitate broader implementation of the standards. Several domain name registries are already working on changing their systems to meet the standards, and many domain name registrars around the world are accepting IDN registrations. Also, vendors of applications including Web browsers, electronic mail programs, and so on, are incorporating the standards into their applications.

RFC 3490 specifies the ways to use the encoded strings inside the DNS, collectively called Internationalized Domain Names in Applications (IDNA). The encoding scheme used in IDNA is an ASCII-compatible encoding (ACE), which encodes Unicode scripts into those ASCII strings, called Punycode, that are allowed in the DNS. While other ACE schemes such as UTF-7 and UTF-8 are widely used in processing free texts, Punycode is meant for encoding only IDNs. The backwards compatibility from Unicode into ASCII characters is necessary to introduce IDNs without any change to the existing infrastructure of the DNS. RFC 3491 specifies the ways to prepare and 'normalize' IDNs in order to remove ambiguities and ensure uniqueness in registering them within specific name spaces. RFC 3492 defines a general algorithm called Bootstring that allows a string of basic characters (ASCII code points) to uniquely represent any string of characters drawn from a larger set (Unicode code points). Punycode is an instance of Bootstring that uses particular parameter values for IDNA. Taken together, the IETF standards provide a technically feasible solution for IDNs with no changes to the existing infrastructure of the DNS.

RFC 3490 also recognizes that the standards do not solve all linguistic problems regarding IDNs used in different sets of scripts. For example, "names that are entered in a mix of traditional and simplified Chinese characters will not be mapped to a single canonical name" and "Scandinavian names that are entered with U+00F6 (Latin small letter o with diaeresis) will not be mapped to U+00F8 (Latin small letter o with stroke)" [2, p. 3]. In addition to such mapping problems associated with specific languages or scripts, IDNs based on the IETF standards have several limitations. First, IDNs are suffixed with generic top-level domains (gTLDs) or country code

top-level domains (ccTLDs) still in ASCII characters, as gTLDs and ccTLDs are not yet internationalized. Second, Punycode strings are just gibberish, meaningless strings of ASCII characters, while they are invisible to users most of the time. Third, Punycode strings are prefixed with a string of ASCII characters, currently 'xn--', so that they can be distinguished from ASCII domain names. Fourth, IDNs in Unicode scripts are generally shorter in length than domain names in ASCII scripts, since Punycode, like other ACE methods, uses sequences of multiple bytes to encode non-ASCII characters and the 67-character limit on domain names, including '.' (dot) and the TLD, is still stand on Punycode strings. These problems and limitations are not certainly exhaustive but rather important ones that need to be resolved for broader implementation of IDNs across languages.

The IETF standards resolve IDNs in Unicode scripts into IP addresses as follows. First, the user enters an IDN in Unicode scripts into the user application. Second, before the domain name request is sent to the local DNS server for resolution, the IDN is encoded into the corresponding Punycode string. Third, the Punycode string resolves into the corresponding IP address, following the same process of domain name resolution in the DNS as described above. Finally, using the IP address following the TCP/IP, a connection is established between the user host and the server host to which the user requested an access. For example, VeriSign (http://www.verisign.com), which operates .com, .org and .net registries, is currently testing this process of IDN resolution following the IETF standards and using those IDNs registered via its channel of registrars around the world. IDNs can be registered in over 350 languages in the VeriSign's test-bed. For now, Internet users have to download and install a special plug-in program into their applications in order to use the VeriSign's IDN service. The plug-in program encodes IDNs in Unicode scripts into the corresponding Punycode strings, which in turn are passed to the VeriSign's IDN test-bed for resolution. In the near future, Internet-capable applications will incorporate the IETF standards, eliminating the need for such plug-in program. Also, Punycode strings encoded from IDNs will be actually added to the zone files, letting IDNs function more like ASCII domain names.

Several regional domain registries that operate registries of ccTLDs, including China Internet Network Information Center (http://www.cnnic.net.cn), Japan Network Information Center (http://www.nic.ad.jp)/Japan Registry Services (http://jprs.jp), Korea Network Information Center (http://www.nic.or.kr), and Taiwan Network Information Center (http://www.twnic.net), are also testing the IETF standards and accepting IDN registrations in their respective languages and scripts. Also, several companies such as I-DNS.net International (http://www.i-dns.net), Neteka (http://www.neteka.com), and Walid (http://www.walid.com) sell and assign IDNs through their affiliated registrars. They also use ACE methods that are the same as or similar to the one used in INDA, but they provide IDN software and services without endorsement from the IETF and the ICANN. They may compete with VeriSign and regional domain name registries in the IDN market, but their IDNs may become less usable when the IETF standards go into full-fledged implementation.

Multilingual Keywords

While the IETF has made efforts to standardize the method for IDNs, several companies including Extended Name Services (http://www.xtns.net), Netpia (http://www.netpia.com), and

New.net (http://www.new.net) have offered multilingual keyword services utilizing their own servers. Multilingual keywords are common names, words and phrases in multilingual scripts. Multilingual keywords are entirely in the multilingual scripts that registrants elect, and they do not follow the hierarchical structure of domain names. Thus, they are simpler and more intuitive than IDNs. Also, unlike IDNs that are encoded into unique Punycode strings, more than one multilingual keyword can be mapped into the same domain name at the choice of the registrant. But multilingual keyword services are not endorsed by the IETF and the ICANN. Also, those multilingual keyword services currently available are not compatible with each other. Each multilingual keyword service can return a different domain name to the same multilingual keyword query. Even with such problems, multilingual keywords may play a role of complementing those standard IDNs, as English keywords provided by Netword (http://www.netword.com) complements ASCII domain names.

In general, multilingual keywords are resolve into IP addresses as follows. First, the user enters a multilingual keyword into the user application. Second, the multilingual keyword query is sent to the server of the user's Internet service provider (ISP), which in turn sends the multilingual keyword query to the server of the keyword service provider for resolution into the corresponding domain name. If the user's ISP is not supported by the keyword service provider, the user host accesses the server of the keyword service provider through a special plug-in program that the user downloads from the keyword service provider's Web site and installs into the user application. Further, the user application may directly incorporate the forwarding function of the plug-in program, eliminating the need for such plug-in program, as Internet Explorer browser had once incorporated such function for the multilingual keyword service provided by RealNames (RealNames went out of business in May 2002 because Microsoft no longer gave it the ability to offer its keyword addressing system through Internet Explorer browser). The server of the keyword service provider keeps track of multilingual keywords and their corresponding domain names. Third, the server of the keyword service provider queries the multilingual keyword in its file and sends back to the user host with the corresponding domain name that was pre-assigned by the registrant of the multilingual keyword. Fourth, the domain name resolves into the corresponding IP address, following the same process of domain name resolution in the DNS as described above. Finally, using the IP address following the TCP/IP, a connection is established between the user host and the server host to which the user requested an access. This process of multilingual keyword resolution is more like redirecting and is transparent to the user.

BENEFITS OF INTERNATIONALIZED DOMAIN NAMES

IDNs can provide various benefits to all the entities involved in the value chain of IDNs, including domain registries, domain name registrars, domain name registrants, and Internet users. A domain registry refers to an organization that has control of the database of a name spaces, including domain name dispute resolution and policy control. It also refers to the name space's database. Domain registries run the authoritative servers for their name spaces. For example, VeriSign controls the name spaces of .com, .net, and .org, and regional domain registries controls the name space of their respective ccTLD. Unlike VeriSign, most regional domain registries are affiliated with their government. Domain name registrars are agents that submit requests for domain name registration and change to their domain registries on behalf of domain name

registrants. Domain name registrants are those who register domain names through domain name registrars and make use of them.

Domain registries and domain name registrars always strive to improve the value of their domain name business. IDNs can provide new opportunities for them to expand this business and increase revenue opportunity by supplementing revenue from ASCII domain name sales that is somewhat saturated. IDNs can create a new revenue stream for domain registries and domain name registrars, with leveraging their existing infrastructure and customer base. For this reason as well as others, domain registries and domain name registrars have actively participated in the IETF's IDN Working Group and other working groups and committees on IDNs. Some of them had even rushed to sell those non-standard IDNs before the IETF set the standards, and so, confused domain name registrants as well as users. IDNs can also open new opportunities for domain name registrants who serve customers whose primary language is not English. Establishing a presence on the Internet is essential for all businesses, and a key component of establishing a presence is a domain name. IDNs can allow businesses to reach their markets more effectively by communicating in the local language of their customers and improve the customer's navigation experience. Also, IDNs enable Internet users to navigate and communicate on the Internet in their preferred local languages and scripts.

CONCLUSION

Supporting internationalized access to domain names is not a simple undertaking and involves quite a few issues from those related to technical implementation all the way to those related to policies. Some issues are specific to IDNs, while others are common to both ASCII domain names and IDNs. The ICANN established the IDN committee (http://www.icann.org/committees/idn/) to provide recommendations on various non-technical policy issues such as interoperability, cyber squatting, dispute resolution, TLDs, consumer protection, and competition. The IDN committee's perspectives, activities and recommendations on these issues are documented on its Web pages. Instead, this paper has reviewed the methods for internationalized access to domain names including ACE methods which the IETF adopted as standards and multilingual keyword methods, focusing on their technical implementation features. The kinds of technical features discussed in this paper are well worth trying to understand, since they are important for resolving existing problems and avoiding any potential problems that may affect implementation of the methods. Such understanding will prove a helpful viewpoint on the technical plausibility of the methods for all the entities involved in the value chain of IDNs, including domain registries, domain name registrars, domain name registrants, and Internet users.

REFERENCES

[1] Costello, A., "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)," RFC 3492, March 2003.

[2] Faltstrom, P., Hoffman, P., and Costello, A, "Internationalizing Domain Names in Applications (IDNA)," RFC 3490, March 2003.

[3] Hoffman, P. and Blanchet, M., "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)," RFC 3491, March 2003.