The GCC Pilot Project for Arabic Domain Names

General Technical Introduction

(version 0.1)
Date: 9-Jun-2004
Prepared by: GCC-ADNS Taskforce

1. Introduction

Domain names are used widely by Internet users to locate resources on the Internet in a format that is easy to remember and understand. They consist of alphanumeric strings separated by dots, e.g., www.arabic-domains.org.sa. They are written using Roman characters (ASCII) particularly letters, digits, and hyphen. These names, however, are not required by the network software, but are used for human mnemonic convenience. They are used instead of the numerical addresses that are known as Internet protocol (IP) addresses, which are mainly used by machines to route data packets on the Internet. Name resolution is carried out by the Internet domain name system (DNS) in which domain names are mapped to the actual corresponding IP addresses.

The DNS is basically a distributed database of host information that is organized in a hierarchal tree structure, see Figure (1). Theoretically, there is a "root domain" at the top of the domain name tree which is usually left unnamed. Immediately underneath the root come the top-level domains (TLDs). Basically, there are tow types of TLDs. One is the generic TLDs (gTLDs) such as .com, .org, .net, and .edu. The second one is the country code TLDs (ccTLDs) such as .ae (United Arab Emirates), .bh (Bahrain), .ca (Canada), .de (Germany), .eg (Egypt), .jo (Jordan), .kw (Kuwait), .om (Oman), .qa (Qatar), .sa (Saudi Arabia), and .uk (United Kingdom). There are more than 240 ccTLDs following the two-letter country codes defined in the ISO standard number 3166.

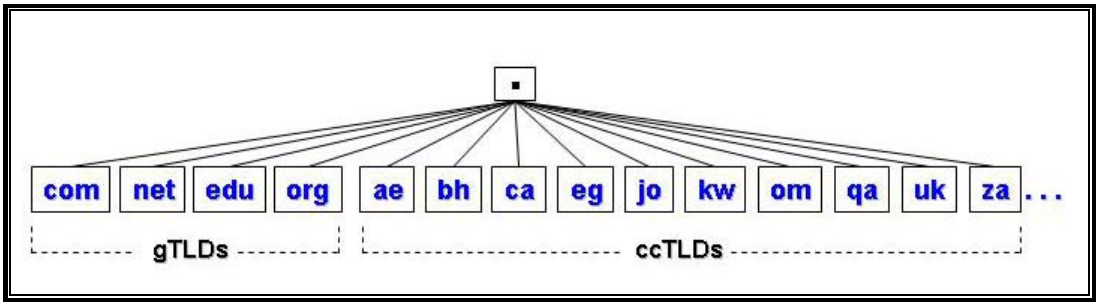

Figure 1: Domain Name Tree Structure

A domain name, whether under a gTLD or ccTLD offers a global presence, which makes sure that the corresponding web site is accessible through the Internet from anywhere. More than 170 millions of such names are estimated to be already stored in the Internet domain name system (DNS).

The Internet has become a global network of most if not all countries of the words with hundred of millions of users. Currently, it is estimated that more than 60% of the Internet contents are in languages other than English. Also, based on an estimation in year 2003 at least 30% of web users who prefer to do their on-line activities in a language other than English, and that by 2005 only one third of Internet businesses will use English for on-line communication.

Regardless of the worldwide spread of the Internet, the Internet domain name system has not supported other languages to locate resources on the Internet. Users in non-English speaking countries, such as the Arab users, are in disadvantages. Using domain names in a language that is different from the users' native language defeats the main objective of having the domain name in characters rather than just numbers.

The Internet penetration in the Arab world is estimated to be about 1.4% which is indeed very low. One of the obstacles facing the growth of using Internet in the Arab world is the language burier. Thus, many countries and nations are encouraging their people to use Internet. Therefore, it is important to make the Internet support the Arabic language not only in web contents but also in their addresses.

Multilingual domain names were first developed in Asia-Pacific countries in 1998, which led later to the creation of a number of non-for profit organizations to supervise and pursuing the deployment of multilingual domain names. Among theses organizations are: the Multilingual Internet Names Consortium (MINC), the Arabic Internet Names Consortium (AINC), the Chinese Domain Name Consortium (CDNC), the International Forum for IT in Tamil (INFITT), and the Japanese Domain Names Association (JDNA). Also, the Internet Corporation for Assigned Names and Numbers (ICANN) established an internal Internationalized Domain Name (IDN) Working Group, and the Internet Engineering Task Force (IETF) created an internationalized DNS group that have been dedicated for exploring the possibility of supporting internationalize Internet.

The IDN group of IETF has issued 3 important RFCs (3490, 3491, 3492.) for Internationalized Domain Names. These new RFC's now make it possible for domain name servers to register non-ASCII domain names and application/client vendors to implement standardized support for handling non-ASCII characters in domain names.

2. How IDN Works?

When a browser sees a host name such as http://www.arabic-domains.org.sa, it passes a request to the DNS resolver service (usually built into an OS), which in turn sends a request to a nearest domain name server to return an IP address that corresponds to the host name. This IP address is then used to connect to the web server in question.

IDN allows host/domain names with non-ASCII characters for user input into a browser's location bar or URL's embedded in web pages. At the network protocol level, there is no change in the restriction that only a subset of ASCII characters be used in URL. If end users input non-ASCII characters as part of a domain name or if a web page contains a link using a non-ASCII

domain name, the application must convert such input into a special encoded format using only the usual ASCII subset characters. RFC 3490 (Internationalizing Domain Names in Applications (IDNA)) defines characters used in IDN to be drawn from Unicode Standard 3.2. It also defines how an application should process non-ASCII characters in such a way to conform to existing host name character restrictions.

As an example, an Arabic domain name , "نطاق.السعودية", will look like the following form "xn--mgb5a8an.xn--mgberp4a5d4ar" after converting it from IDN to ASCII format (using "stringprep", "nameprep" and "punycode" operations).

3. Supporting Arabic Domain Names

Supporting the Arabic language in domain names calls for investigating and addressing a number of issues to produce a set of standards that are acceptable by the Internet community in large. These standards should cover several aspects of supporting Arabic domain names in deferent levels, such as:

1. Linguistic issues and the accepted Arabic character set.
2. The Arabic domain name tree structure, i.e., Arabic gTLDs and ccTLDs.
3. Technical solutions to Arabize the domain name system
4. The administrative and organizational issues of Arabic root servers.

The 1st and 2nd points have been addressed by the ESCWA-RFC. The 3rd point is partially addresses by the IETF 3 RFCs (3490, 3491, 3492).

Since there is no indication that ICANN is going to support full IDN in the near future, the GCC ccTLD working group agreed on implementing a pilot project among the GCC ccTLDs to address the issue of Arabic root servers (i.e., point 4). This will allow the GCC ccTLD to early experience the use of Arabic domain names, identify our needs, locate possible problems, and develop tools. In addition, we can use the outcome of the pilot project as an argument with international bodies, e.g., ICANN, to speed up the international recognition of supporting Arabic language in domain names based on our needs.

4. GCC Arabic Domain Names

The ASCWA-RFC suggests the following Arabic TLDs for the GCC countries. Thus, we have to select one for the pilot project.

س ع	السعودية	سعودية	سعودي
ق ط ر	قطر	قطر	قطري
ا م	الإمارات	إمارات	إماراتي
ب ح	البحرين	بحرين	بحريني
ع م	عمان	عمان	عماني
ك و	الكويت	كويت	كويتي

Based on each GCC ccTLD requirements here is the list of ASCII strings that correspond to the GCC Arabic TLD in this pilot project:

ASCII	IDN	Country
xn--mgberp4a5d4ar xn--ogblly9en	السعودية سعودية	Saudi Arabia
xn--wgbl6a	قطر	Qatar
xn--kgbdbap4b0ij xn--kgbeam7a8h xn--mgbaam7a8h	الإمارات إمارات امارات	United Arab Emeritus
xn--mgbcpq6gpa1a	البحرين	Bahrain
xn--mgb9awbf	عمان	Oman
xn--mgbg8edvm	الكويت	Kuwait

5. GCC Arabic Domain Names ccTLD servers

Each GCC country should setup and runs its own Arabic ccTLD server(s) for the chosen Arabic domain name that represent their country name. Although as what have agreed on each GCC ccTLD can register some test Arabic domain names such as:

Example	Arabic test Domain name	Purpose
موقع.السعودية	موقع.<**اسم الدولة**>	A site for testing Arabic domain name in each GCC ccTLD
دليل.الإمارات	دليل.<**اسم الدولة**>	A site that list all the registered Arabic test domain names in each GCC ccTLD.
مركز-التسجيل.قطر	مركز-التسجيل.<**اسم الدولة**>	A site for registering Arabic test domain names in each GCC ccTLD

In addition, each GCC ccTLD server can act as secondary for other GCC ccTLD, which will add more redundancy and localization for the DNS queries.

Note: There will be another technical document on how to setup Arabic ccTLD name server.

6. GCC Arabic Domain Names Root Servers

If each GCC country runs its own Arabic ccTLD server, users in other GCC countries would not be able to reach other GCC Arabic domain names. Therefore, we need a mechanism to address the problem.

The Task force agreed on having many Arabic root servers around the GCC countries, which will be responsible for the root zone file for all the Arabic GCC ccTLD (e.g. السعودية، قطر، البحرين، الإمارات، عمان، الكويت) and that zone file will only have the "NS" and any glue "A" record for each Arabic ccTLD zone.

These name servers will have the same copy of the simple Arabic GCC ccTLD zone files (6 zone each one of them has his own "NS" records), and all of them are authorize for these zone files but in the configuration they will act as a "slaves"(secondary) and there will be a "master" name server (which can be hidden master) that will populate all GCC zone files to them. This scenario makes the modification for any GCC Arabic zone file very easy and centralized between the root servers.

So the root servers will be one layer above the ccTLD servers and any entity that want to resolve Arabic domain name for any GCC ccTLD need to contact them first in order to reach the ccTLD (see figure 2 for more details).

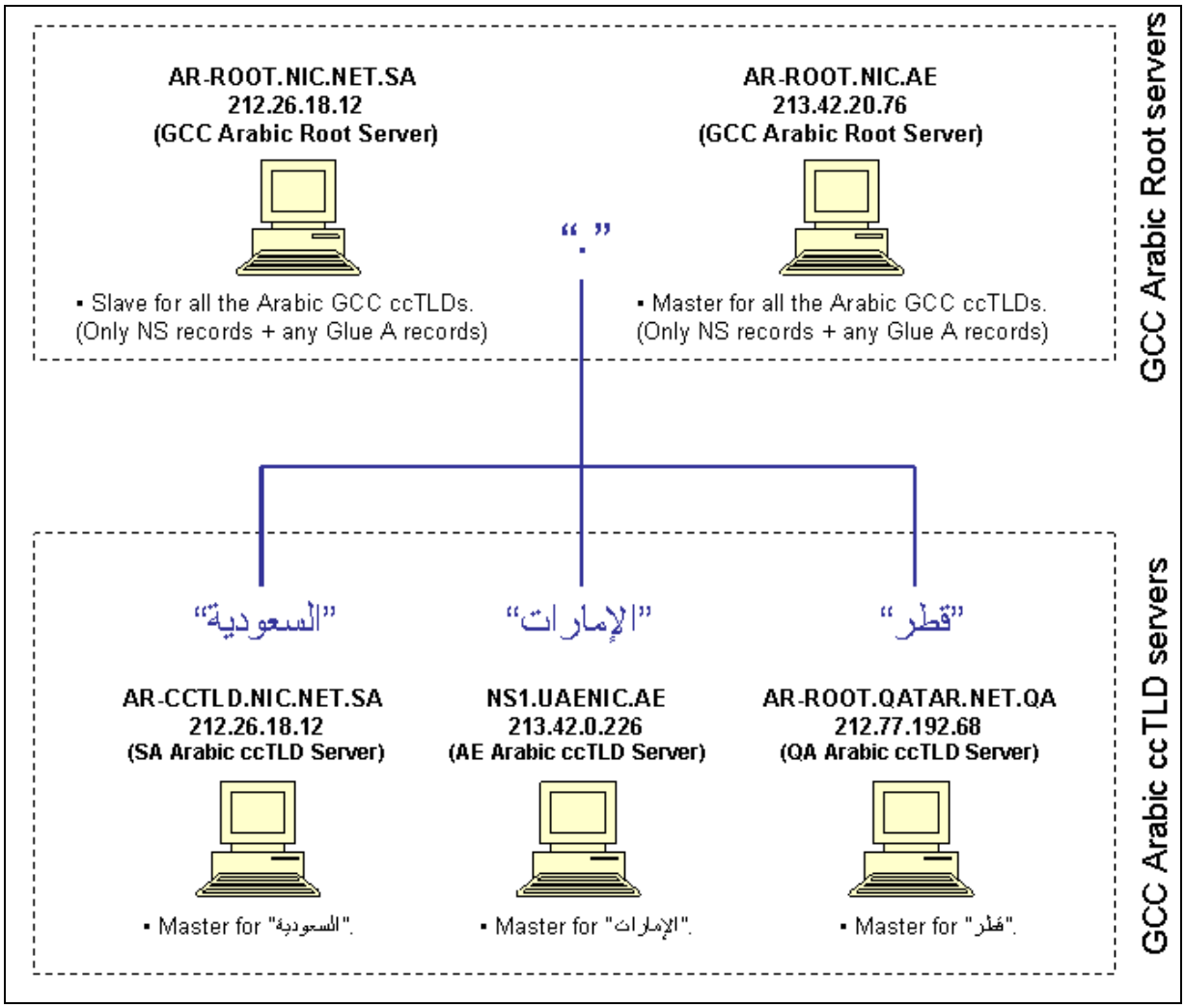

Figure 2: The Arabic Root and ccTLD servers map

Note: There is a technical document on how to setup Arabic root name server (see "How to setup Arabic root server ") and another one for how to setup Arabic ccTLD server (see "How to setup Arabic ccTLD server ").

7. How to resolve GCC Arabic Domain Names

Since ICANN has not yet supported full Arabic ccTLD any solution for supporting Arabic domain names would be only accessible locally. Therefore, there should be a hacking solution until Arabic domain names is world-wide supported. Hence, any entity needs to reach GCC Arabic Domain names must therefore do some changes in the configuration files of their resolving/caching name servers.

The Task force has agreed on using the “stub zone” configuration option for any entity that needs to resolve GCC Arabic domain names. A “stub” zone is like a “slave” zone, except that it replicates only the NS records of a master zone instead of the entire zone.

Therefore, if any entity wants to resolve Arabic GCC ccTLD domain names they should configure all the Arabic GCC ccTLD as “stub” zones in the configuration of their name server and point them to the GCC Arabic root servers.

Here is an example on how to configure stub zone using bind version 8 or higher in the file “named.conf” for all the GCC Arabic ccTLDs:

```
zone "xn--mgberp4a5d4ar" {
        type stub;
        file "xn--mgberp4a5d4ar.sa.zone" ;
        masters { 213.42.20.76; 212.26.18.12; };
};

zone "xn--kgbdbap4b0ij" {
        type stub;
        file "xn--kgbdbap4b0ij.ae.zone" ;
        masters { 213.42.20.76; 212.26.18.12; };
};
.
.
```

File: named.conf

Hint:

The above stub zones configuration can be kept in a separate file that is distributed by the admin of GCC root servers. Then it can be referenced (using the "include" statement) within the bind configuration file "named.conf" in any entity name server that want to resolve Arabic domain names.

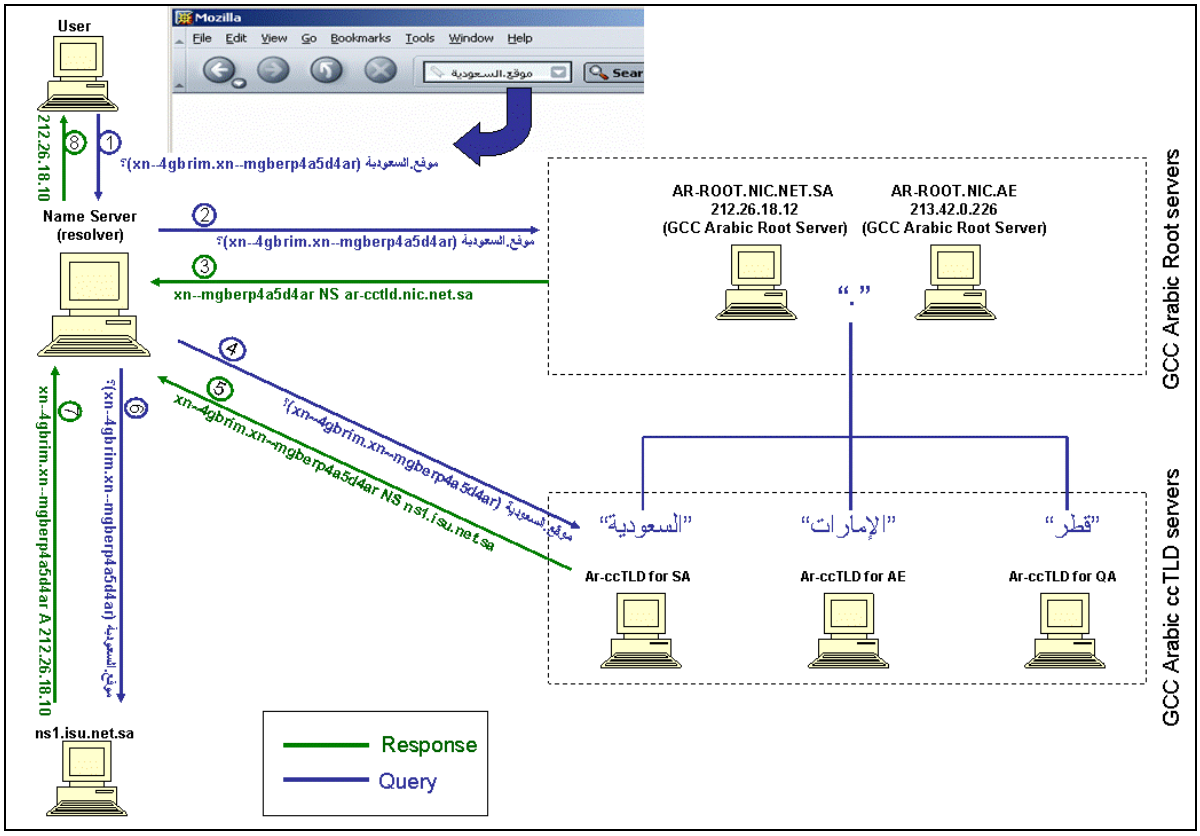

Figure 3: How to resolve Arabic domain names

Note: There is a technical document on how to setup your name server to resolve GCC ccTLD Arabic domain names (see "How to Resolve Arabic Domain Names") and another one that describes the client requirements to reach Arabic domain names (see "Requirements for Resolving Arabic Domains ").

More information

- The GCC ADNS Taskforce website:
  http://www.arabic-domains.org.sa
- IDNs RFCs:
  http://www.ietf.org/rfc/rfc3490.txt
  http://www.ietf.org/rfc/rfc3491.txt
  http://www.ietf.org/rfc/rfc3492.txt
- News on IDN:
  http://idn.isc.org/
- GNU Libidn:
  http://www.gnu.org/software/libidn/