

國際化的網域名稱系統

(internationalization of Domain Names System, iDNS)

蔣大偉、陳文水、黃勝雄

中央研究院·計算中心·網路組

台北市南港區 11529 研究院路二段一二八號

dawei@sinica.edu.tw, chenws@sinica.edu.tw, huangk@sinica.edu.tw

摘要

目前在“網際網路”(Internet)上穩定運作的“網域名稱系統”(DNS),無法滿足“非英語系”使用者的需要,本土化網域名稱的需求呼之欲出。“國際化的網域名稱系統”(iDNS)之需求也逐漸成形中。本文分別針對網域名稱系統的國際化、區域化、和標準化提出探討。同時針對此需求,進行 iDNS 系統的開發。希望經由理論與實務架構之並行作業,能夠提升 iDNS 系統的可用性、穩定性、和安全性。

關鍵字: dns、idns、TLD、gTLD、ccTLD。

一、研究動機

西元 1969 年網際網路在美國發跡,短短三十年間網際網路開枝散葉地成為廣佈全球的資訊網絡,世界儼然成為一個地球村。隨著“非英語系”上網人口的不斷增加,許多早期制定的標準都已不符使用,這也是當初始料未及的狀況。以網際網路上最重要的“網域名稱系統”(DNS)而言,當初所制定的標準 RFC1034 (Domain Names - Implementation and Specification) 和 RFC1035 (Domain Names - Concepts and Facilities) 把網域名稱局限在英文的大小寫字母、阿拉伯數字、以及連字符號的範圍中,這完全是針對英語系使用者

所做的權宜決定,並未把其他語系的使用者納入考量。是否能夠實作出一個國際化的網域名稱系統(iDNS),以滿足“非英語系”使用者“本土化”(區域化)的需求,並進而促成“非英語系”網域名稱的標準化?

二、背景說明

“網域名稱”(Domain Name)就是一般人熟知的“網址”,它讓網路使用者(的電腦)能夠在網路上找到想要聯絡的對方。“網域名稱”好比是網路世界的門牌號碼,當然一個門牌號碼只能給一部上網的電腦使用(不過每部上網的電腦卻可以擁有多個門牌號碼),透過它網路世界的公民們可以互通有無。

根據 RFC1034、RFC1035、和 RFC1591 (Domain Name System Structure and Delegation) 等標準的定義,我們知道:每個完整的網域名稱(Full-Qualified Domain Name)是由數層網域所構成,每層網域必須以點(.)做為間隔符號。網域名稱局限在英文的大小寫字母(A-Z 或 a-z; 網域名稱大小寫不分)、阿拉伯數(0-9)、連字符號(-)等範圍中。最右邊的網域內容,代表著“最上層的網域”(Top-Level Domain name; TLD),而 TLDs 又可區分成“通用網域”(generic TLDs; gTLDs)和“國碼網域”(country code TLDs; ccTLDs)。gTLDs 目前由七個通用網域所組成:com(商業)、edu(教育學術)、gov(美國政府)、mil(美國軍方)、net(網路服務)、org(非營利組織)、和 int(國際組織)。而 ccTLDs 則是採用 ISO3166 標準的

二碼制國名代號（譬如 TW 代表臺灣，CN 代表中國）。

從 RFC1034、RFC1035、和 RFC1591 等標準的制定中，我們不難看出一些端倪。由於國際網路的發源地在美國，所以舉凡“網域名稱”和“通用網域”等標準的制定，都是以美國為中心來設計的。隨著上網人數的不斷增加，機器節點也持續地加入到網路上來，RFC1591 標準已不足以容納急劇增加的網域名稱，遂有許多組織和國家推動新的 gTLDs 標準。同時因為“非英語系”國家上網人數已成氣候，RFC1034 和 RFC1035 等標準也已不符使用，遂有網域名稱“本土化”（或區域化）的呼聲。

三、iDNS（國際化的網域名稱系統）計劃

中央研究院（以下簡稱中研院）計算中心有鑑於國內對網域名稱有“本土化”（中文網址）的需求，遂提出 iDNS（國際化的網域名稱系統）計劃，希望能提供一個安全、可靠、以及穩定的中文網域名稱系統給國內使用，而且能夠相容於舊有的網域名稱系統（也就是原有的英文網域名稱仍然能夠使用），並希望以“開放原始程式碼”（open source）的方式流通以利推廣。此計劃共有三個目標，分別是：網域名稱系統的國際化、區域化、和標準化。網域名稱系統的國際化和區域化其實是一體兩面的事情。以目前既成的事實而言，唯有國際化（去除不必要的限制）後的網域名稱系統，纔能容忍區域化（或本土化）後的網域名稱；而在網域名稱區域化的過程中，我們必須為本土化網域名稱的引用，去除掉一些不必要的限制，遂有網路應用程式（譬如 sendmail）、系統函式庫（譬如 gethostbyname）等的國際化需求。

目標一、網域名稱系統的國際化 （Internationalization, I18N）

1. 除去“網域名稱系統”對網域名稱所作，不必要的限制和轉換的動作（或稱為 8-bit clean）。
2. 讓網域名稱系統可以同時處理“英語系”和“非英語系”的網域名稱。

就第一個子目標而言，要建置一個“8-bit clean”的網域名稱系統，必須去除 RFC1034 和

RFC1035 等標準所強加的限制。可行的方法就是修改 BIND（網域名稱系統的軟體實作）套裝程式。

就第二個子目標而言，要能同時處理“英語系”和“非英語系”的網域名稱，必須讓原本處理英文網域名稱的 root name server（根名稱伺服器）也能處理非英文的網域名稱，在“非英語系”網域名稱尚未標準化之前，這個理想是不可能達成的。可行的方法就是自己設置一個可以處理“非英語系”網域名稱的新 root name server，而且為了使用原有的英文網域名稱，新的 root name server 必須到原來的 root name server 處取得一份最新的根（或最上層）網域資料（root.zone）併入“非英語系”的根（或最上層）網域資料中。值得注意的是，一般用戶端所使用的名稱伺服器，若要加入“非英語系”的網域名稱系統一起運作，則必須修改其名稱快取檔案（named.ca），只能使用“非英語系”網域名稱系統所提供的 root name server。

目標二、網域名稱系統的區域化 （Localization, L10N）

1. 讓網域名稱系統可以處理“某個語系”的網域名稱。
2. 讓網路應用程式可以直接引用“非英語系”的網域名稱。

就第一個子目標而言，有了“國際化”所發展出來的“8-bit clean”網域名稱系統，就當然可以處理“非英語系”的網域名稱。接下來的問題在於，是否有適用的作業平台。就臺灣目前的現況（以 open source）來說，CLE（以 RedHat 為基礎的中文延伸套件）算是中文作業環境最佳的選擇，因為它為我們妥善地解決了中文（BIG5/GB）的輸入和輸出問題。因此我們會在 CLE 這個中文作業平台上，發展與安裝“8-bit clean”的網域名稱系統。此外，當網域名稱本土化之後，還會引發其它的問題。以中文網域名稱來說，會有：統一命名的問題（全球資訊網、首頁、萬維網... 等都是指 www）、正簡寫的問題（『台』與『臺』、『体』與『體』... 等等）、使用習慣的問題（中文地址的習慣是由大到小，譬如說：臺灣.教育.中研院）... 等等。

就第二個子目標而言，我們發現到當網路應用程式在解析網域名稱時，會呼叫標準函式庫所提供的 `gethostbyname` 函式，因此若能修正標準函式庫裡的 `gethostbyname` 函式，則可以解決掉大部分的問題（然而，微軟中文視窗不須做任何修改，上面的網路程式便可以直接使用中文的網域名稱）。

目標三、“非英語系”網域名稱系統的標準化 (Standardization, S13N)

1. 網域名稱各自區域化後，如何相互引用對方的網域名稱。
2. 促成“非英語系”網域名稱的標準化。
3. 廢除 *RFC1034* 和 *RFC1035* 等標準對網域名稱所作不合時宜的限制。並制定新的標準。

就第一個子目標而言，“非英語系”網域名稱應該要符合本土使用者實際的需求，否則根本就無法推行，所以必須採行各自發展與區域合作的步驟，唯有如此纔能收水到渠成之效。以中文網域名稱來說，臺灣地區使用的是 **BIG5** 字集編碼的繁體字，而中國大陸使用的是 **GB-2312-80** 字集編碼的簡體字。這兩套字集編碼互有重疊，無法做自動判斷的動作，因此必須指明目前的編碼方式纔有可能做轉碼的動作。**GB-2312-80** 的編碼只有 6763 個漢字，而 **BIG5** 的編碼卻有 13053 個漢字，所以有很多字轉不出來的。同時也要注意兩地各有不同的習慣用語。目前較可行的方法，就是使用者端的 `gethostbyname` 函式，一遇到最上層網域為“中國”的網域名稱，就把它轉換成 GB 編碼的字串，當然臺灣地區的 **root name server** 必須能夠將它指向中國大陸的 **root name server**；反之亦然。

就第二個子目標而言，“非英語系”網域名稱的機制將會引發“多國語言” (multilingual) 同時使用的問題，是否有這樣的作業平臺可供使用也是個問題，**Mule** (Multilingual Emacs) 的發展似乎是一個可以借力的地方。要促成“非英語系”網域名稱系統的標準化，以及免除轉碼的不方便，到最後似乎還是得採用可以滿足“多國語言”需求的字集編碼，譬如 **Unicode**、**ISO 10646-1:1993** 等標準。當然這必須取得大家的

共識纔有意義。

四、實作方法

步驟一、“國際化”網域名稱系統的下載、安裝、設定與測試。

- 下載經由“中研院計算中心”修改過 (8-bit clean) 的 `bind` 程式套件 [12]：

1. **REDHAT 5.2 + CLE v0.7** 的使用者可以直接下載修改過之 RPM 包裝的二元碼檔。

```
bind-8.2-1iDNS.i386.rpm
bind-utils-8.2-1iDNS.i386.rpm
bind-devel-8.2-1iDNS.i386.rpm
caching-nameserver-5.2-1iDNS.noarch.rpm
```

2. **REDHAT 6.0 + CLE v0.8** 的使用者可以直接下載修改過之 RPM 包裝的二元碼檔。

```
bind-8.2-6iDNS.i386.rpm
bind-utils-8.2-6iDNS.i386.rpm
bind-devel-8.2-6iDNS.i386.rpm
caching-nameserver-6.0-2iDNS.noarch.rpm
```

3. 其它版本的使用者，下載修改過的原始碼檔 `bind-8.2iDNS-src.tar.gz` 與 `named.root`。

- 安裝“國際化”網域名稱系統：

1. **REDHAT 5.2 + CLE v0.7** 的使用者：

先檢查系統是否已經安裝過 `bind` 程式套件，請輸入

```
rpm -qa | grep bind
```

若已安裝執行的結果會像這樣

```
bind-8.2-1
```

```
bind-devel-8.2-1
```

```
bind-utils-8.2-1
```

此時我們應該把原有的 `bind` 程式套件移除，請輸入

```
rpm -e --nodeps bind-8.2-1
```

```
bind-devel-8.2-1 bind-utils-8.2-1
```

現在我們可以開始安裝修改過(國際化)之 RPM 包裝的二元碼檔，請在二元碼檔

所存放的目錄下輸入
 rpm -ivh bind*
 rpm -ivh --force
 caching-nameserver-5.2-1iDNS.noarc
 h.rpm

2. **REDHAT 6.0 + CLE v0.8** 的使用者：

請在二元碼檔所存放的目錄下輸入
 rpm -ivh --force bind*
 rpm -ivh --force
 caching-nameserver-6.0-2iDNS.noarc
 h.rpm

3. 其它版本的使用者：

請把修改過（國際化）的原始碼檔
 bind-8.2-src.tar.gz 解開，然後切換
 到 src 這個子目錄下，按照檔案
 INSTALL 的說明進行編譯與安裝的工
 作。並以 named.root 取代 /var/named
 目錄下的 named.ca 檔案。

• 設定區域化網域名稱（以中研院的中文
 網域名稱為例）：

1. 檔案 /etc/resolv.conf 的內容如下

```
search 中研院.教育.臺灣
nameserver 140.109.1.137
```

2. 檔案 /etc/named.conf 的內容如下

```
options {
    directory "/var/named";
}
zone "." {
    type hint;
    file "named.ca";
};
zone "0.0.127.in-addr.arpa" {
    type master;
    file "named.local";
};
zone "中研院.教育.臺灣"
{
    type master;
    file "db.sinica.edu.tw";
};
zone "137.1.109.140.in-addr.arpa"
{
    type master;
```

```
file "db.140.109.1.137";
};
```

3. 檔案 /var/named/db.sinica.edu.tw
 的內容如下

```
$ORIGIN 中研院.教育.臺灣.
;
@           IN      SOA      名
稱伺服器.中研院.教育.臺灣. root.名稱
伺服器.中研院.教育.臺灣.
                                199907280 ; Serial
                                36000    ; Refresh
                                7200     ; Retry
                                3600000  ; Expire
                                172800  ); Minimum
                                IN      NS
名稱伺服器.中研院.教育.臺灣.
;
名稱伺服器           IN      A
140.109.1.137
全球資訊網           IN      A
140.109.4.130
```

4. 檔案 /var/named/db.140.109.1.137
 的內容如下

```
$ORIGIN
137.1.109.140.in-addr.arpa.
@           IN      SOA
名稱伺服器.中研院.教育.臺灣. root.名
稱伺服器中研院.教育.臺灣.
                                199907270 ; Serial
                                36000    ; Refresh
                                7200     ; Retry
                                3600000  ; Expire
                                172800  ); Minimum
                                IN      NS
名稱伺服器.中研院.教育.臺灣.
;
                                IN      PTR
名稱伺服器.中研院.教育.臺灣.
```

5. 檔案 /var/named/named.ca 的內容如
 下

```
; formerly IDNS.SINICA.EDU.TW
.           3600000    NS
名稱伺服器.網路.臺灣.
```

名稱伺服器.網路.臺灣. 3600000 A
140.109.1.2

6. 檔案 `/var/named/named.local` 的內容如下

```
@      IN      SOA      本機. root.  
本機. (  
    199907270 ; Serial  
    28800     ; Refresh  
    14400     ; Retry  
    3600000   ; Expire  
    86400 )   ; Minimum  
1      IN      PTR      本機.
```

- 測試區域化網域名稱（以中研院的中文網域名稱為例）：

先以 `ndc start` 的方式啟動名稱伺服器 `named`。

1. 若輸入 `nslookup -type=soa "中研院.教育.臺灣"` 則結果如下所示：

```
Server: 名稱伺服器.中研院.教育.臺灣  
Address: 140.109.1.137  
中研院.教育.臺灣  
    origin = 名稱伺服器.中研院.  
教育.臺灣  
    mail addr = root.名稱伺服器.  
中研院.教育.臺灣  
    serial = 199907230  
    refresh = 36000 (10H)  
    retry = 7200 (2H)  
    expire = 3600000 (5w6d16h)  
    minimum ttl = 172800 (2D)  
中研院.教育.臺灣    nameserver  
= 名稱伺服器.中研院.教育.臺灣  
名稱伺服器.中研院.教育.臺灣  
internet address = 140.109.1.137
```

2. 若輸入 `nslookup -type=a "全球資訊網.中研院.教育.臺灣"` 則結果如下所示：

```
Server: 名稱伺服器.中研院.教育.臺灣  
Address: 140.109.1.137  
Name:    全球資訊網.中研院.教育.臺灣  
Address: 140.109.4.130
```

3. 若輸入 `nslookup -type=ptr`

`137.1.109.140.in-addr.arpa` 則結果如下所示：

```
Server: 名稱伺服器.中研院.教育.臺灣  
Address: 140.109.1.137  
137.1.109.140.in-addr.arpa  
name = 名稱伺服器.中研院.教育.臺灣  
137.1.109.140.in-addr.arpa  
nameserver = 名稱伺服器.中研院.教育.臺灣  
名稱伺服器.中研院.教育.臺灣  
internet address = 140.109.1.137
```

步驟二、讓網路應用程式可以直接引用區域化的網域名稱。

- 網路應用程式（`ping`, `traceroute`, `telnet`, `ftp` ...）的修正（區域化）：

修正前，這類程式一般的回應就是 `Unknown host`。

修正後，可正常引用中文（區域化）的網域名稱。

1. 若執行 `ping "全球資訊網.中研院.教育.臺灣"` 則結果如下所示：

```
PING 全球資訊網.中研院.教育.臺灣  
(140.109.4.130): 56 data bytes  
64 bytes from 140.109.4.130:  
icmp_seq=0 ttl=254 time=0.6 ms  
64 bytes from 140.109.4.130:  
icmp_seq=1 ttl=254 time=0.5 ms  
--- 全球資訊網.中研院.教育.臺灣  
ping statistics ---  
2 packets transmitted, 2 packets  
received, 0% packet loss  
round-trip min/avg/max =  
0.5/0.5/0.6 ms
```

2. 若執行 `traceroute "全球資訊網.中研院.教育.臺灣"` 則結果如下所示：

```
traceroute to 全球資訊網.中研院.教育.臺灣 (140.109.4.130), 30 hops  
max, 40 byte packets  
 1 140.109.1.1 (140.109.1.1)  
0.882 ms 0.748 ms 0.706 ms  
 2 * * gate (140.109.4.130) 0.557
```

ms

3. 若執行 `telnet "全球資訊網.中研院.教育.臺灣"` 則結果如下所示：
Trying 140.109.4.130...
Connected to 全球資訊網.中研院.教育.臺灣.
Escape character is '^]'.
UNIX(r) System V Release 4.0 (gate)
login:
4. 若執行 `ftp "全球資訊網.中研院.教育.臺灣"` 則結果如下所示：
Connected to 全球資訊網.中研院.教育.臺灣.
220 gate FTP server (UNIX(r) System V Release 4.0) ready.
Name (全球資訊網.中研院.教育.臺灣:root):
5. 文字或圖形使用者介面的網頁瀏覽器，當然也可以正常的使用：
`lynx http://"全球資訊網.中研院.教育.臺灣"`
`netscape http://"全球資訊網.中研院.教育.臺灣"`

- **注意事項**

由於網路應用程式在解析網域名稱時，會呼叫標準函式庫所提供的 `gethostbyname` 函式，因此只要修正標準函式庫即可（然而微軟中文視窗不須做任何修改，上面的網路程式可以直接引用中文的網域名稱）。

BIG5 是個多位元組 (multiple-byte) 的字集編碼，由於其第二個位元組有可能小於 128 (0x80)，因此與 ASCII 字碼共用時無法確定小於 128 的位元組值就是 ASCII 字碼，這將造成軟體在處理 ASCII 字碼時的不相容。譬如說“一”的 **BIG5** 字碼為 0xA4 0x40，其第二個的位元組值將與 @ 的 ASCII 字碼重疊，這將造成與 `bind`、`sendmail` 等程式不相容的情況發生。

五、問題討論

網路名稱系統 (DNS) 是網路基礎建設的一環，一有任何閃失都將導致多數人權益受損，所以修改網路名稱系統的原則應該兼顧系統的可用性、穩定性、和安全性。這有賴於系統開發者與使用者的不斷地測試、回饋、和修正，方能達成。

在 **RFC1034** 和 **RFC1035** 等標準修改前，爲了和舊有網路名稱系統相容，我們必須自行建置可以取代原有系統的 `root name server`。

網路名稱系統的國際化是在爲本土化的網路名稱做準備，所以發展的方式應該是先做好本土化的網路名稱系統，然後進行區域間的合作 (regional root system operation)，最後透過標準化的過程推廣到全世界。

iDNS 計劃的旨在補充原有網路名稱系統功能之不足，而非創新的技術 (enabling technology)，所以原有的運作機制仍舊不變。在目前 **DNS** 所提供資訊無法滿足多數人需求的狀況下，藉由 **iDNS** 所提供的服務，除了可支援 **DNS** 之不足外，同時也爲原本不夠用的 **gTLDs** 命名空間開闢出一條新的方向。在未來我們樂見有更多新的技術 (譬如，目錄服務) 能提供更多樣化且整合式的解決方案。

六、結論

由於網路上“非英語系”的使用者不斷增加，遂有本土化網域名稱的需求，然而因爲原有網域名稱標準 **RFC1034** 和 **RFC1035** 的限制，造成網域名稱本土化的困難重重。中研院計算中心的 **iDNS** 計劃，希望能透過網域名稱系統的“國際化” (**8-bit clean**)、“區域化”、和“標準化”等三個目標，來達成網域名稱本土化的目的。目前已經發展出了一個 **iDNS** 系統的雛形，基本上已滿足國際化和區域化的目標，未來可望透過更多人的參與以及國際間的合作來達成標準化的目標。

七、參考文獻

1. P. Mockapetris, RFC-1034, "DOMAIN NAMES - CONCEPTS AND FACILITIES", November 1987.

2. P. Mockapetris, RFC-1035, "DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION", November 1987.
3. J. Postel, RFC-1591, "Domain Name System Structure and Delegation", March 1994.
4. Cricket Liu & Paul Albitz, "DNS and BIND" Third Edition, September 1998.
5. ken Lune, "CJKV Information Processing", January 1999.
6. Joe Babcock, Linux Journal ISSUE 59, "Polyglot Emacs 20.4 - A look at multilingual Emacs", March 1999.
7. Juay-Kwang Tag, "idNS, an Experimental DNS System with Unicode Support", March 1999.
8. ISC BIND,
<http://www.isc.org/view.cgi?products/BIND/index.phtml>
9. DNS Resources Directory,
<http://www.dns.net/dnsrd/>
10. An LDAP Roadmap & FAQ,
<http://www.kingsmountain.com/ldapRoadmap.shtml>
11. Chinese Linux Extensions,
http://cle.linux.org.tw/CLE/e_index.shtml
12. Internationalization of Domain Names System,
<http://idns.sinica.edu.tw/>