

Handling of Variants

Lucy Wang
(On behalf of CDNC)
August 20, 2009

“Everyone has the right...
to seek, receive, impart
information and ideas through
any media, regardless of
frontiers”

-Universal Declaration of Human Rights

Content



The origin and facts of the variant issue



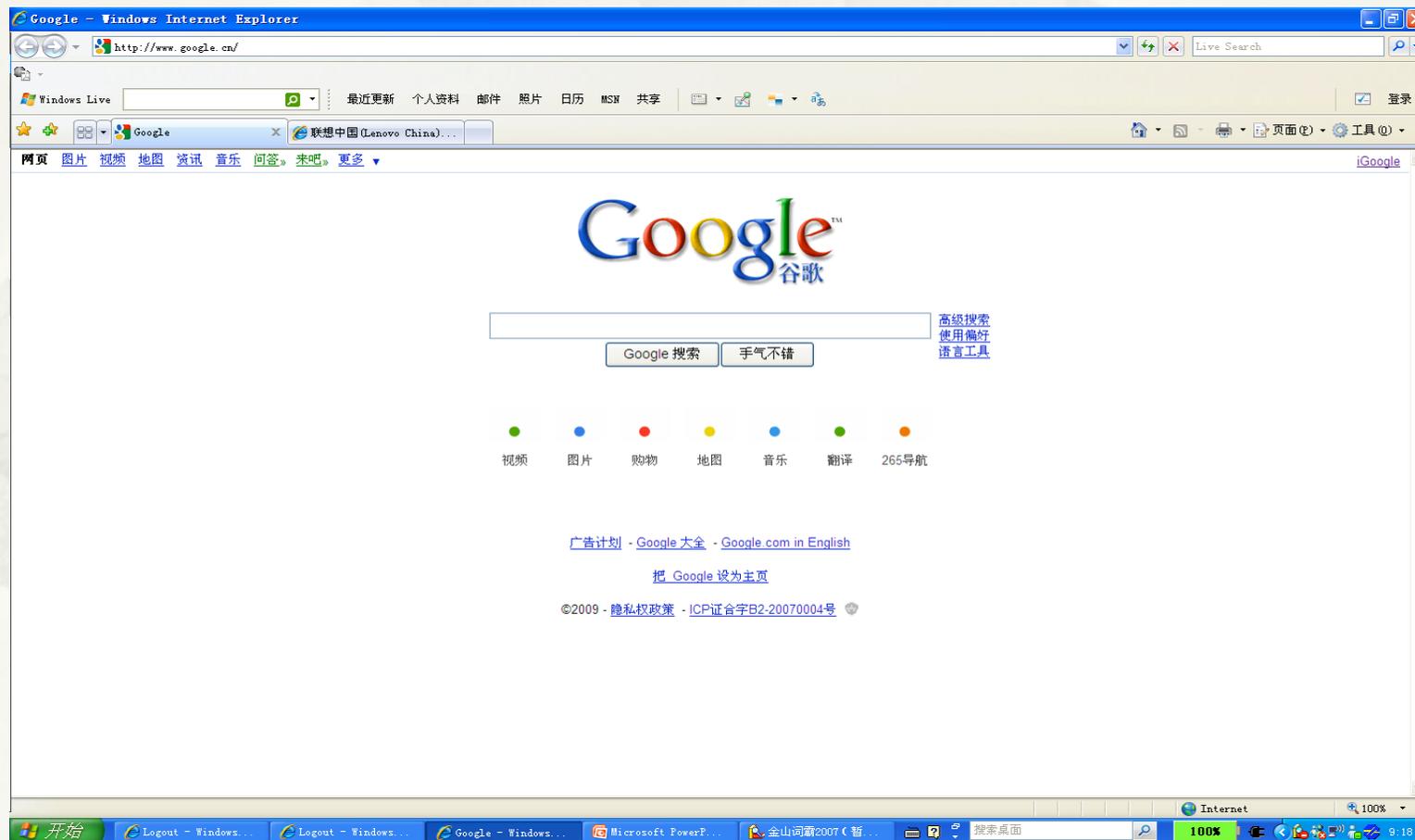
How CDNC handles the issue



CDNC Support and Petition for IDN

The origin and facts of the variant issue

www.google.com



The origin and facts of the variant issue

- **Currently Web addresses are typically expressed using Uniform Resource Identifiers or URIs. The URI syntax defined in [RFC 3986 STD 66](#) (Uniform Resource Identifier (URI): Generic Syntax) essentially restricts Web addresses to a small number of characters: basically, just upper and lower case letters of the English alphabet, European numerals and a small number of symbols.**
- **The original reason for this was to aid transcribability and usability both in computer systems and in non-computer communications, to avoid clashes with characters used conventionally as delimiters around URIs, and to facilitate entry using those input facilities available to most Internet users.**

The origin and facts of the variant issue

<http://www.mercedes-benz.com.cn;>

The screenshot shows the Mercedes-Benz China website in a Windows Internet Explorer browser window. The browser's address bar displays <http://www.mercedes-benz.com.cn>. The website header includes the Mercedes-Benz logo and navigation links for '奔驰乘用车' (Mercedes-Benz Passenger Cars), '服务与配件' (Service & Parts), '大客户销售' (Corporate Sales), '企业社会责任' (Corporate Social Responsibility), '奔驰天下' (Mercedes-Benz World), and '联系梅赛德斯-奔驰' (Contact Mercedes-Benz). The main content area features a large banner for the '全新一代梅赛德斯-奔驰E级轿车' (All-new Mercedes-Benz E-Class sedan) with the slogan '世出荣耀 耀世而出' (Worldly Honor, Shining World). Below the banner, there are sections for '快速链接' (Quick Links), '车型概览' (Model Overview), and '热点关注' (Hot Topics). The footer contains the Mercedes-Benz logo, language options (中文, English), and legal information.

Mercedes-Benz China - 梅赛德斯-奔驰(中国) - Windows Internet Explorer

http://www.mercedes-benz.com.cn

文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)

收藏夹

Mercedes-Benz China - 梅赛德斯-奔驰(中国)

奔驰乘用车 Mercedes-Benz China 经销商网络

奔驰乘用车 服务与配件 大客户销售 企业社会责任 奔驰天下 联系梅赛德斯-奔驰

世出荣耀 耀世而出
全新一代梅赛德斯-奔驰E级轿车
[了解更多](#)

精彩主题

- > 全新E级轿车
- > 驾驶奔驰穿越欧洲
- > 轮胎检测活动
- > Mercedes-Benz.tv
- > 车展在线

快速链接

- > 金融服务
- > 售后服务
- > 新闻中心
- > 联系梅赛德斯-奔驰
- > 互动社区

车型概览

浏览梅赛德斯-奔驰所有车型

[车型概览](#)

热点关注

- > 购梅赛德斯-奔驰顶级跑车 尊享“欧洲激情之旅”
- > 梅赛德斯-奔驰SUV家族“溯源中国”之旅
- > 商店大客户特惠活动
- > 售后服务调查项目
- > 梅赛德斯-奔驰“星徽服务发现之旅”
- > 梅赛德斯-奔驰与中国保护大熊猫研究中心启动合作计划

载重车

- > 乌尼莫克
- > 豪华商务车
- > 轻型商用车

Mercedes-Benz 中文 | English | 网站地图 | © 2009 Daimler AG | 隐私条款 & Cookies | 京ICP备09048804号 | 法律声明

Internet 100%

The origin and facts of the variant issue

<http://www.奔驰.cn>

The screenshot shows the Mercedes-Benz China website in a Windows Internet Explorer browser window. The address bar displays <http://www.mercedes-benz.com.cn>. The page features a navigation menu with the Mercedes-Benz logo and links for "奔驰乘用车", "Mercedes-Benz China", and "经销商网络". Below the navigation is a main banner for the "全新一代梅赛德斯-奔驰E级轿车" (All-new Mercedes-Benz E-Class sedan) with a "了解更多" (Learn more) button. To the right of the banner is a "精彩主题" (Featured Topics) section with links to "全新E级轿车", "驾乘奔驰穿越欧洲", "轮胎检测活动", and "Mercedes-Benz tv". Below the banner are three columns of links: "快速链接" (Quick Links) including "金融服务", "售后服务", "新闻中心", "联系梅赛德斯-奔驰", and "互动社区"; "车型概览" (Model Overview) with a "浏览梅赛德斯-奔驰所有车型" (Browse all Mercedes-Benz models) link and a "车型概览" (Model Overview) button; and "热点关注" (Hot Topics) including "购梅赛德斯-奔驰顶级轿车 尊享'欧洲激情之旅'", "梅赛德斯-奔驰SUV家族'溯源中国'之旅", "惠店大客户特惠活动", "售后服务调查项目", "梅赛德斯-奔驰'星徽服务发现之旅'", and "梅赛德斯-奔驰与中国保护大熊猫研究中心启动合作计划". The footer contains the Mercedes-Benz logo, language options (中文, English), a site map, copyright information (© 2009 Daimler AG), privacy policy, and a license number (京ICP备09046804号). The browser status bar shows "Internet" and "100%" zoom.

The origin and facts of the variant issue

- **User's expectations and use of the Internet have moved on since then, and there is now a growing need to enable use of characters from any language in Web addresses.**
- **A Web address in our own language and alphabet is easier to create, memorize, transcribe, interpret, guess, and relate to.**
- **It is also important for brand recognition. This, in turn, is better for business, better for finding things, and better for communicating. In short, better for the Web.**

The origin and facts of the variant issue

- **This is what we call IDN, Internationalized Domain Name, which is a mean of “localization” in someway given the global nature of the internet.**
- * **The whole systems need to be “localized”: adapted to our particular**
 - * **Language**
 - * **Writing system and character codes**
 - * **Location**
 - * **Interests**

The variant issue

- While we going through the implementation, we are facing some issues. Among them, “variant” definitely has caused some threats and concerns, and requires sophisticated technology solution and policies to deal with it



The variant issue

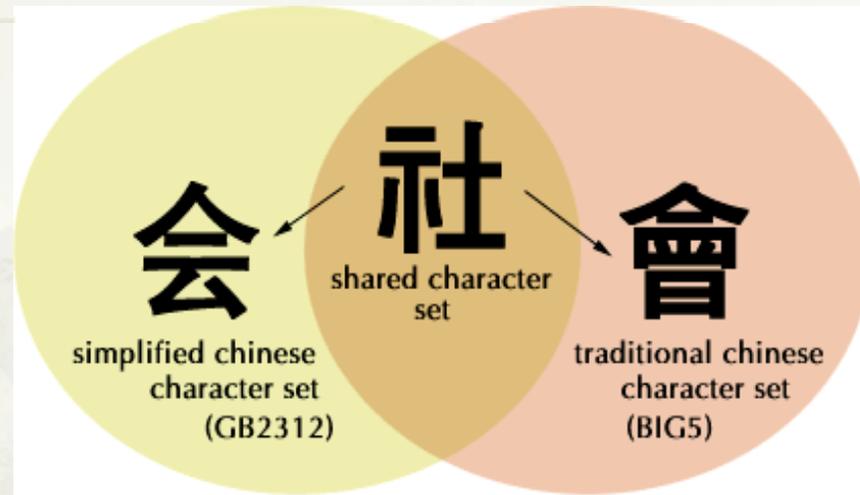
- The definition of variant:

“Variant characters are two or more characters that are similar in appearance and result in two domain names to be visually confusing.” like:

- ◆ “encyclo**p**æ**d**ia” vs “encyclo**p**a**e**dia”



The variant issue



These two versions of Chinese share many characters. Other characters specific to Simplified Chinese or Traditional Chinese scripts may represent the same meaning. These characters have the same meaning and pronunciation but they do not look the same. These characters are called character variants and have the potential to cause confusion for end users, or even to face phishing or fraud attacks when using Traditional and Simplified Chinese scripts to register domain names.

The Innovation of Chinese Characters

| | horse | cart | fish | dust | see | |
|--|-------|------|------|------|------|--|
| Oracle bone script 甲骨文 (jiǎ gǔ wén) | | | | | | The Oracle bone script was used during the Shang or Yin Dynasty (c. 1400-1200 BC) |
| Bronze script 金文 (jīn wén) | | | | | | The Bronze script was used during the Zhou Dynasty (c. 1100 - 256 BC) |
| Large Seal script 大篆 (dà zhuàn) | | | | | | The Large Seal script was used during the Zhou Dynasty (c. 1100 - 256 BC) |
| Small Seal script 小篆 (xiǎo zhuàn) | | | | | | The Small Seal script was used during the Qin Dynasty (221-207 BC) |
| Clerical script 隸書 (lì shū) | | | | | | The Clerical and Standard scripts first appeared during the Han Dynasty (207 BC - 220 AD). |
| Standard script 楷書 (kǎi shū) | | | | | | |
| Running script 行書 (xíng shū) | | | | | | The Running script has been used for handwritten Chinese since the Han Dynasty. |
| Grass script 草書 (cǎo shū) | | | | | | The Grass script is the Chinese equivalent of shorthand and has been used since the Han Dynasty. |
| Simplified script 简体字 (jiǎntǐzì) | | | | | | The Simplified script has been used in the P.R.C. since 1949. It is also used in Singapore. |
| hànyǔ pīnyīn 汉语拼音 | mǎ | chē | yú | chén | jiàn | <i>Hanyu pinyin</i> has been used in the P.R.C. since 1958. |
| zhùyīn fúhào 注音符號 | ㄇˇ | ㄔ | ㄩˊ | ㄔㄣˊ | ㄐㄧㄢˋ | <i>Zhuyin fuhao</i> was developed in China in 1913 and is still used in Taiwan. |

The history of the issue

Internationalization and the Internet

- * **Consideration given to “international characters” in the 1970s**
 - * Character set standards weren't ready
- * **Project that led to MIME**
 - * “multimedia email” capability
 - * initiated largely to standardize and permit non-ASCII characters
- * **Web**
 - * Recognized requirement early
 - * Details only for Western European languages until mid-90s
- * **All were done by “tagging”**
 - * Tagging is consistent with localization approaches

The concerns of the issue

DNS Internationalization

- * **Tension between**
 - * Network-facing identifier
 - * User-facing “name” (of a company, product, organization,...)
- * **Constraints on solutions**
 - * Short label strings – no reasonable way to tag
 - * Uniqueness of names
 - * Potential for confusion or fraud
- * **Requirement for non-ASCII names is clear but**
 - * Caution is in order – many possible traps and risks
 - * Hard to go back if too permissive

The concerns of the issue

“The worst scenarios could be one of the following two: either that IDNs will be filled with phishing attacks that IDNs will be of no use and users will be scared of using them, or restrictions in the application layer will be so strict that IDNs will for example not resolve in an adequate and at least not in a stable and secure manner.

Either way, this does not provide the community what they have asked for and what we are attempting to provide them with the implementation of IDNs, namely, equal access to the DNS by all languages and scripts. “

---Tina Dam (Director of ICANN)

Content



The origin and facts of the variant issue

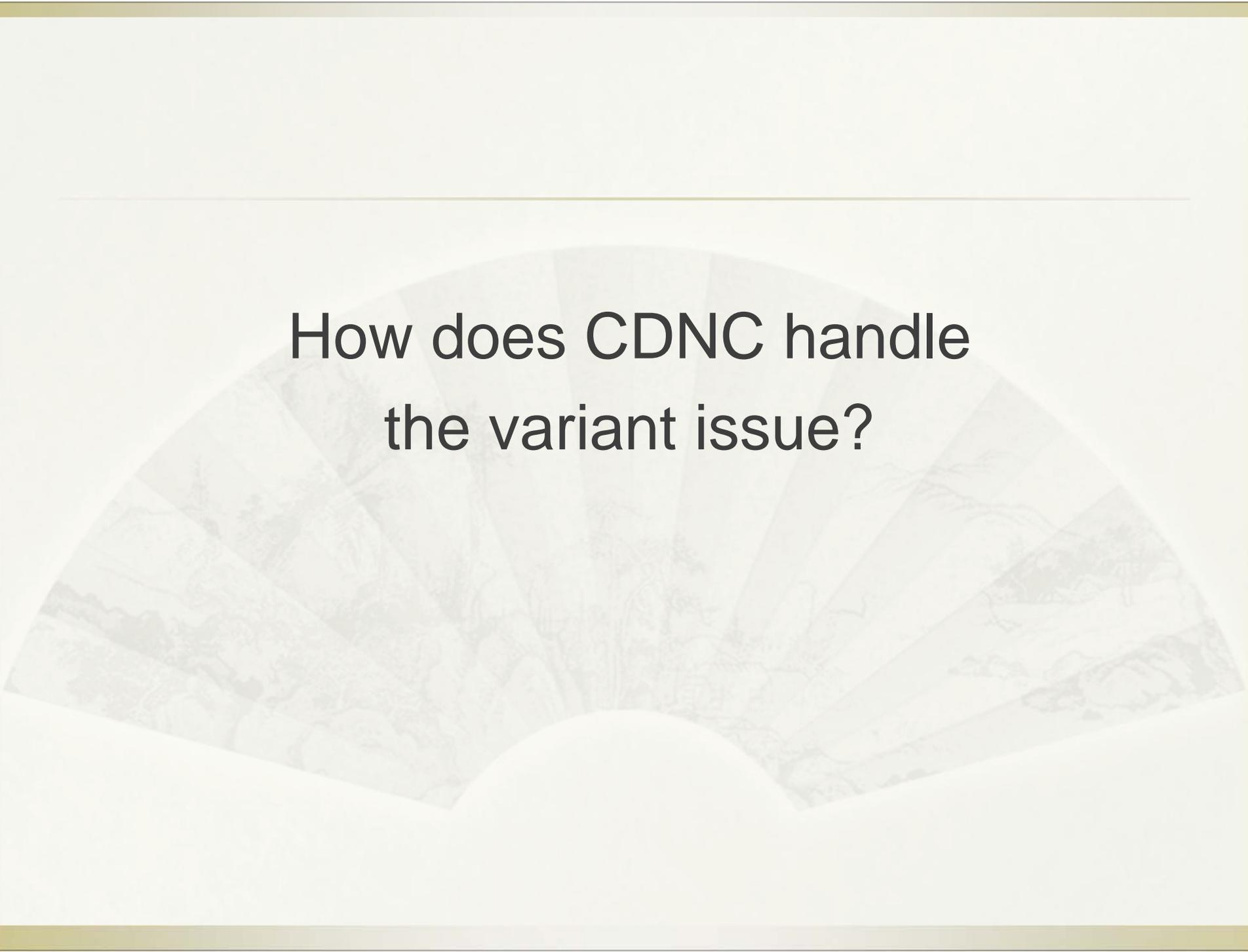


How CDNC handles the issue



CDNC Support and Petition for IDN

How does CDNC handle
the variant issue?



About CDNC

Chinese Domain Name Consortium Initiative:

- * **Founded by four NICs in 2000:**
 - * **CNNIC**
 - * **TWNIC**
 - * **HKIRC**
 - * **MONIC**
- * **Later joined by SGNIC , CONEC and a number of other world well-known registries and registrars, as well as some domain name user alliance, like CDNUA.**

Vision and Mission

- **The Vision:**

Internet and e-commerce in the digital knowledge age shall be more widely adopted by Non-English speaking communities;

- **The Mission of CDNC:**

- To harmonize, promote and self-regulate registry operations and services of Chinese domain name.
- To adopt most advanced technology available to serve the Chinese domain name users;

Mile stones

- * June. 5, 2001 - Chinese Domain Name Consortium (CDNC) Final Comments on IETF last-call of IDN WG core drafts
- * Aug 20,2003 CDNC submit "Registration and Administration Guideline for Chinese domain Names" draft to IETF(Including CDNC variant table)
- * August,2005 IANA publishes the Chinese Character Table
- * September,2005 CNNIC and TWNIC jointly submit the new version of CDN registration guideline to IETF
- * October,2005 JET IMA working group submits 4 drafts to IETF
- * IETF promulgated RFC4952 Overview and Framework for Internationalized Email Address (2007/7)
- * July.2008 IETF IESG approves 3 key drafts of EAI WG as RFC
- * IETF publishes a series of RFC to support internationalized email address(April.2009)

Handling of variants: Strategic



Strategies for adoption of internationalized domain names worldwide:

- * Minimal Changes to Current DNS system
- * Must not break existing structure and hierarchy
- * Support all languages
- * Support as many encoding as desired
- * Avoid ambiguity - uniqueness
- * Work everywhere for everyone
- * Follow IETF process
- * Seek International Consensus
- * Minimize disruption or Protocol changes
- * Harmonize solutions
- * Adopt simplest solution

Handling of variants: Principles of the solution

The Solution: RFC3743

- **Domain name string should be bundled with a specified language:** Domain name string could be bundled with many languages, but this situation should be avoided, since the consequence of bundling with many languages could result directly with impossible registration of the domain name. Therefore a domain name string should only be recognized as legitimate one within a certain language character set.
- **A sufficient Variant Table of specified language should be identified;** It is not in common practice for any of countries in the world to employ every character collected in the Unicode suite. Particularly none of countries has defined every single character in the Unicode suite to be the legal or official one. Therefore, validity of a domain name string should be verified with every language bundled.

Handling of variants: Principles of the solution

The Solution: RFC3743

- **The variants of domain name string should be reserved;** Since in a specified language, a name usually has many variants, therefore those variants of the domain name should be reserved to protect the rights of the holder. They are also entitled to be activated or deactivated at the request of the holder, e.g. the variants should be implemented in the root zone for resolution or transfers.
- **The preferred variants should be all resolved;** Domain name could have many variants, but not every variant is frequently used or formally employed. Among the most frequent used ones, there may be only a small portion of the variants which should be added into the zone documents of the DNS system for resolution.

Handling of variants: Principles of the solution

The Solution: RFC3743

- **The amount of variants should be constrained;** A name could have so many variants, some of which may not be meaningful at all. For instance, a name which has 10 Han characters could result 1024 (1K) variants if each of the character has one variant. Among these variants, some are meaningful, some don't make any sense at all. The resolution to all of the variants could be a huge burden to administration system. Therefore some reasonable methods should be deployed to reduce the amount of variants for better resolution and protection.

Handling of variants: Principles of the solution

The Solution: RFC3743

- **Name string and its variants have a Characteristic of Atom, which needs to be dealt as a package:** Once a name and its variants are created, they are relevantly compacted together. They should be dealt with as a whole package while an individual or independent handle of any of the variants in the package is strictly forbidden.



Content



The origin and facts of the variant issue



How CDNC handles the issue



CDNC Support and Petition for IDN

To the IDN Nay-Sayers

- * **Technically not possible – proven it works**
- * **No demand and no interest - plenty**
- * **No service providers – plenty came**
- * **No organisation – CDNC, ICANN, etc**
- * **No standards – RFCs of IETF**
- * **Not Interoperable – Inter-operational test with excellent results**
- * **No Language Tables – CDNC table**
- * **No legitimacy – rights of community, who gets to decide on my language? ... Language empowerment groups**

Conclusion & Petition:

- * The implementation of IDN variant is of utmost importance to our community as variants are often used interchangeably, similar although not the same, as uppercase and lowercase characters in English.
- * Members of CDNC believe that the introduction of allocation of variant strings in the root zone will also avoid visual confusability and potential phishing attacks. Such policy will also ensure the security and stability of the Internet in a multi-lingual environment.

Conclusion & Petition:

- * One of the major concerns about the variant issue is technical implementation capability, for which CDNC would like to point out that the solution of IETF standard RFC3743, and more specifically to Chinese, the RFC4713 has been in practice for nearly a decade and it has been proven to be one of the most sufficient and rigorous way of managing this matter.
- * The basic principles with the solution RFC3743 are believed in strongly and adhered to by the members of CDNC, which represents 99% of the Chinese domain name stakeholders. Such common issues shall be envisaged and dealt with across SOs and Constituencies in ICANN with respect to the stand of the majority of the CDN community.

Conclusion & Petition:

- * To help fulfill ICANN's goal of ensuring a smooth and secure launch of IDN ccTLD and gTLD in the near future, CDNC would like to recommend the ICANN community and staff to jumpstart the currently pending works on IDN guideline update.
- * CDNC strongly believes that the completion of the guideline in order to provide sufficient rules and policy on implementing IDN TLD is critical for existing registries and new gTLD applicants.
- * We strongly urge ICANN to consider our view and firm position on this particular issue.



Thanks

August 20, 2009