1-1-2012

The 2011 IDN Homograph Attack Mitigation Survey

Peter Hannay

Gregory Baatard
Edith Cowan University

The 2011 IDN Homograph Attack Mitigation Survey

P. Hannay[1] **and G. Baatard**[1]
[1]ECUSRI, Edith Cowan University, Perth, WA, Australia

Abstract - *The advent of internationalized domain names (IDNs) has introduced a new threat, with the non-English character sets allowing for visual mimicry of domain names. Whilst this potential for this form of attack has been well recognized, many applications such as Internet browsers and e-mail clients have been slow to adopt successful mitigation strategies and countermeasures. This research examines those strategies and countermeasures, identifying areas of weakness that allow for homograph attacks. As well as examining the presentation of IDNs in e-mail clients and Internet browser URL bars, this year's study examines the presentation of IDNs in browser-based security certificates and requests for locational data access.*

Keywords: IDN, homograph, network security, internationalized domain names, computer security

1 Introduction

Internationalized Domain Names (IDN) allow for non-latin characters to be present in domain names. There are a number of security issues associated with this. Primarily this comes from the potential impersonation of domains by an attacker. This attack is achieved through the use of non-latin characters, which are visually indistinguishable from their latin counterparts. The aforementioned attack is known as an IDN homograph attack. This paper aims to investigate the strategies utilized by current web browsers to mitigate the impact of these attacks.

1.1 Instructions for authors

Domain names have been with us for a long time, first introduced in 1983 they provided a centralized means of abstraction for IP addresses (Mockapetris 1983, Mockapetris 1983). Since their inception domain names have become a key player in the information security arena, a known domain name inspires trust on behalf of the average user and as such is a high value item for would be attackers. Internationalized Domain Name Homograph attacks represent one attack vector that such an attacker could leverage for his/her advantage.

The initial implementation of domain names allowed only for alphanumeric characters and hyphens encoded as ASCII (Mockapetris 1983). In subsequent years it become apparent that this was an unacceptable limitation as audiences that make use of non-latin character sets were not able to have domains in their respective languages. In 1998 the initial work on Internationalized Domain Names (IDN) began. This work and subsequent work cumulated in 2003 with the publication of RFC3454, RFC3490, RFC3491 and RFC3492, a set of documents outlining the function and proposed implementation of IDN (Bell-ATL 2011).

The proposed IDN solution made use of UTF-8 character encoding to allow for non-latin characters to be displayed. In order to enable existing DNS infrastructure to handle UTF-8 domains a system known as Punycode was developed (Faltstrom, Hoffman et al. 2003). Punycode provides facility to represent IDNs as regular ASCII domain names, as such no changes are required for the majority of infrastructure (Costello 2003). An example of an IDN would be the domain name ☃.com, which would be represented as xn--n3h.com when converted to punycode.

1.2 Attacks

A number of visually indistinguishable glyhps (known as homoglyphs) exist within the Unicode character space. An example pair of glyphs are Unicode 0067 "Latin Small Letter G" and its counterpart Unicode 0261 "Latin Small Letter Script G" which are visually indistinguishable from one another. The aforementioned glyphs can be seen below in Figure 1.

Figure 1 – Example of Homoglyph for "g"

Homoglyphs can be combined with characters from other scripts to form a series of glyphs, which as a whole are visually indistinguishable from their English counterpart. When a client or server interprets these homographs however they are treated in a distinct manner. Through the use of this trait attackers are able to craft domain names, which look familiar but are hostile in intent. These attacks can be deployed in the same manner as regular email phishing attacks, aiming to entice a user into accessing a hostile website in the belief that it is the genuine site being imitated. These attacks have been employed to steal financial data, passwords and corporate information.

Traditionally phishing attacks are mitigated through user education, encouraging users to check the legitimacy of links before clicking them, looking for unrelated URLs, not replying to emails asking for information if they are from an external domain name, etc. However when phishing campaigns are modified to make use of homograph domain names the ability for user education to provide mitigate is eliminated, as there is no way to make a visual identification of a fraudulent domain name. Figure 2 shows two domain names, both visually identical, however they lead to separate websites, with the one to the right making use of U+0261 rather than U+0047 for the second G.

Figure 2 – A pair of Homograph Domains

1.3 Mitigation

A number of countermeasures have been implemented in order to mitigate the effectiveness of this attack. The majority of these involve displaying punycode in place of the actual UTF-8 text. Punycode is an ASCII representation of a Unicode domain name, originally implemented as the domain name service infrastructure did not support Unicode (Costello 2003). The punycode alternative is commonly displayed in both the address bar and the status bar on hover for a particular link.

When identifying domain names to display in punycode, there are two main methods used. The first (used by internet explorer 7 and above) is to use punycode only when a domain using mixed-script is detected (Fu, Deng et al. 2006). The implications of this are that any domain, which is intended to be spoofed via the replacement of one or more characters, will be detected, however in the event that the entire domain name is made from a single script it will be presented as intended by the attacker.

The other method employed by Mozilla Firefox and Safari both utilizes a whitelist in which all IDNs are presented as punycode unless they belong to a top level domain (TLD) that has policy in place preventing the spoofing of domain names in this manner. The policies employed via TLDs to prevent this attack often require that prior to registering a domain name containing homoglyphs, the registerer must own the domain name containing the western variant of those homoglyphs. In implementing this policy the IDN homograph attack is eliminated, however a number of TLDs have failed to implement this policy (Mozilla 2005).

A final strategy involves the color coding of various scripts in URLs (Krammer 2006). In this method Cryllic scripts are highlighted one color, while western scripts are left uncolored. In this situation mixed script URLs become immediately visible to the user, even though the characters themselves are visibly identical.

2 Testing strategy

For testing purposes we developed a virtual environment comprised of Windows 7 installation, which at the time of writing at all current updates applied. A snapshot was taken prior to the installation of web browsers or email clients.

Four primary attack vectors were identified with regards to IDN homograph attacks in web browsers, corresponding to the four most prominent locations in which an IDN may be shown to the user. If an IDN is shown in Unicode, a homograph attack could result in the user being tricked into believing that a URL is that of a legitimate website. The four attack vectors, in order of prevalence, are:

- The text shown in the browser's address bar, after the "Go" (or equivalent) button has been pressed.
- The text shown in the browser's status bar while the mouse is over a hyperlink.
- The text shown when viewing prominent information about a website's SSL certificate. As most users do not examine the details of a certificate, this attack vector relies upon the presentation of IDNs in immediately visible or accessible information.
- The text shown when the user is prompted to share their location using geolocation services.

In order to summarize the findings, an overall "Mitigation Rating" was calculated for each version of each browser tested. A value of zero is awarded if the browser does not support a particular attack vector, for example a lack of support for IDNs or geolocation services. A value of negative one is awarded if the browser supports an attack vector without mitigation against IDN homograph attacks. A value of positive one is awarded if the browser supports an attack vector and does mitigate against IDN homograph attacks, for example by presenting IDNs in Punycode. As the presentation of IDNs in the browser's address bar is by far the most prominent and influential vector of attack, values of positive and negative *two* are awarded for this vector. These values are shown in Figure 3 below.

Address Bar	Status Bar	SSL Certificate	Location Request
-2 (Unmitigated)	-1 (Unmitigated)	-1 (Unmitigated)	-1 (Unmitigated)
0 (No Support)	0 (No Support)	0 (No Support)	0 (No Support)
+2 (Mitigated)	+1 (Mitigated)	+1 (Mitigated)	+1 (Mitigated)

Figure 3 – Mitigation Ratings

By applying this metric, each browser version tested can be awarded a Mitigation Rating between positive five and negative five, representing a browser that supports and mitigates all attack vectors and a browser that supports but does not mitigate any of the attack vectors respectively.

Numerous versions of five web browsers were tested, based on averaged current market share data from a number of sources (Clicky 2011). Tested browsers were Internet Explorer (Microsoft), Firefox (Mozilla), Chrome (Google), Opera (Opera Software) and Safari (Apple). The authors attempted to test the initial release of each major version of the browsers since 2003, when RFC3454, RFC3490, RFC3491 and RFC3492 and ICANN's "Guidelines for the Implementation of Internationalized Domain Names" were published. All browsers were tested in a Windows 7 environment. The results of the testing are presented below.

3 Results

Internet Explorer					
Version & Release Date	Address Bar Mitigation	Status Bar Mitigation	SSL Certificate Mitigation	Location Request Mitigation	Mitigation Rating
7.0 (2006-10)	Punycode	Punycode	No Mitigation	No Support	+2
8.0 (2009-03)	Punycode	Punycode	No Mitigation	No Support	+2
9.0.8 (2011-03)	Punycode	Punycode	No Mitigation	No Mitigation	+1

Figure 4 – Results for Internet Explorer

Support for IDNs was added to Microsoft Internet Explorer in version 7, released in late 2006. IDNs in the address and status bars were shown in Punycode, and an icon providing further information about IDNs appears next to the address bar when one is used. Support for geolocation services was implemented in the latest major version of the browser, version 9, released in March of 2011. Internet Explorer currently offers no mitigation against IDNs in SSL certificate information or geolocation requests, showing them in Unicode. While Internet Explorer has protected itself against the most significant vector of IDN homograph attacks since support for IDNs was implemented, SSL certificate information and geolocation requests are presented without any mitigating features.

Firefox					
Version & Release Date	Address Bar Mitigation	Status Bar Mitigation	SSL Certificate Mitigation	Location Request Mitigation	Mitigation Rating
1.0 (2004-11)	None	None	None	No support	-4
1.5 (2005-11)	Punycode	Punycode	None	No support	+2
2.0 (2006-10)	Punycode	Punycode	None	No support	+2
3.0 (2008-06)	Punycode	Punycode	Punycode	No support	+4
3.5 (2009-06)	Punycode	Punycode	Punycode	Punycode	+5
3.6 (2010-01)	Punycode	Punycode	Punycode	Punycode	+5
4.0 (2011-03)	Punycode	Punycode	Punycode	Punycode	+5
5.0 (2011-06)	Punycode	Punycode	Punycode	Punycode	+5
6.0 (2011-08)	Punycode	Punycode	Punycode	Punycode	+5
7.0 (2011-09)	Punycode	Punycode	Punycode	Punycode	+5

Figure 5 – Results for Firefox

The first version of Mozilla Firefox was released in late 2004, and supported IDNs without any features to mitigate against homograph attacks. From version 1.5, released approximately a year later, IDNs in the address and status bars were shown in Punycode. From version 3.0, released in mid 2008, IDNs were shown in Punycode for SSL certificate information and were also placed more prominently in the interface. When support for geolocation services was implemented in version 3.5, mid 2009, requests were shown in Punycode. Firefox incorporated features that mitigate IDN homograph attacks fairly quickly, limiting its exposure in the two main vectors to a single major release.

Google Chrome					
Version & Release Date	Address Bar Mitigation	Status Bar Mitigation	SSL Certificate Mitigation	Location Request Mitigation	Mitigation Rating
1.0.154.59 (2009-04)	Punycode	Punycode	Punycode	No support	+4
2.0.172.27 (2009-05)	Punycode	Punycode	Punycode	Punycode	+5
3.0.197.11 (2009-08)	Punycode	Punycode	Punycode	Punycode	+5
4.0.302.3 (2010-01)	Punycode	Punycode	Punycode	Punycode	+5
5.0.396.0 (2010-05)	Punycode	Punycode	Punycode	Punycode	+5
6.0.495.0 (2010-08)	Punycode	Punycode	Punycode	Punycode	+5
7.0.544.0 (2010-10)	Punycode	Punycode	Punycode	Punycode	+5
8.0.552.224 (2010-12)	Punycode	Punycode	Punycode	Punycode	+5
9.0.597.16 (2011-02)	Punycode	Punycode	Punycode	Punycode	+5
10.0.648.205 (2011-03)	Punycode	Punycode	Punycode	Punycode	+5
11.0.696.77 (2011-04)	Punycode	Punycode	Punycode	Punycode	+5
12.0.742.112 (2011-06)	Punycode	Punycode	Punycode	Punycode	+5
13.0.782.218 (2011-08)	Punycode	Punycode	Punycode	Punycode	+5
14.0.835.202 (2011-09)	Punycode	Punycode	Punycode	Punycode	+5
15.0.874.21 (2011-09)	Punycode	Punycode	Punycode	Punycode	+5
16.0.904.0 (2011-10)	Punycode	Punycode	Punycode	Punycode	+5

Figure 6 – Results for Google Chrome

Despite only being released a few years ago, Google has released sixteen versions of the Chrome web browser. As the browser is in beta, the release cycle and version numbers are not as predictable as other browsers. All versions present IDNs in the address bar, status bar, SSL certificate

information and geolocation requests in Punycode. Support for geolocation services was added in version 2 of the browser, released in mid 2009. All versions of Chrome have included defences against IDN homograph attacks, however the fact that it was first released much later than any of the other major browsers must be taken into account.

Opera					
Version & Release Date	**Address Bar Mitigation**	**Status Bar Mitigation**	**SSL Certificate Mitigation**	**Location Request Mitigation**	**Mitigation Rating**
7.00 (2003-01)	No support	No support	No support	No Support	0
8.00 (2005-04)	No Mitigation	No Mitigation	No Mitigation	No Support	-4
9.00 (2006-06)	No Mitigation	No Mitigation	No Mitigation	No Support	-4
10.00 (2009-09)	No Mitigation	No Mitigation	No Mitigation	No Support	-4
11.00 (2010-12)	Punycode	Punycode	Punycode	Punycode	+5
11.51 (2011-08)	Punycode	Punycode	Punycode	Punycode	+5

Figure 7 – Results for Opera

Version 7 of the Opera web browser, released early in 2003, did not support IDNs. The next three major releases (in 2005, 2006 and 2009) supported IDNs but offered no mitigation to IDN homograph attacks. Support for geolocation services was added in version 11, late 2010, at which point IDNs in all attack vectors started to be shown in Punycode. While all vectors are not mitigated against IDN homograph attacks, the browser was vulnerable to the attacks for approximately five years.

Safari					
Version & Release Date	**Address Bar Mitigation**	**Status Bar Mitigation**	**SSL Certificate Mitigation**	**Location Request Mitigation**	**Mitigation Rating**
3.1 (2008-03)	No Mitigation	No Mitigation	No Mitigation	No Support	-4
3.2 (2008-11)	No Mitigation	No Mitigation	No Mitigation	No Support	-4
4.0 (2009-06)	No Mitigation	No Mitigation	No Mitigation	No Support	-4
5.0.1 (2010-07)	Punycode	Punycode	No Mitigation	Punycode	+3
5.1 (2011-07)	Punycode	Punycode	No Mitigation	Punycode	+3

Figure 8 – Results for Safari

Apple's Safari browser began showing IDNs in Punycode in the address bar, status bar and geolocation requests from version 5, released in mid 2010. Prior to that version, IDNs in the address and status bar were shown in Unicode and geolocation services were unsupported. It is worthwhile noting that the default settings for Safari hide the status bar, nullifying the mitigation possible when hovering over a hyperlink. No mitigation exists for SSL certificate information. The authors also noted that the first HTTP URL to be entered into the address bar upon launching the latest version of the browser was shown in Unicode. Successive URLs were presented in Punycode. Safari was vulnerable to IDN homograph attacks for a number of years, and remains vulnerable in small areas.

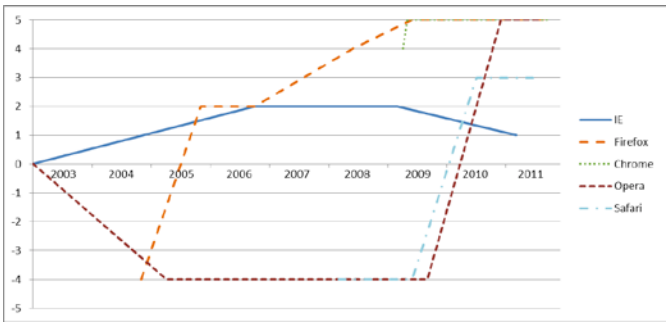

Figure 9 – Results Summary

4 Conclusion

The longitudinal data present from the nine years of software releases sampled provides interesting data. We can see that there is a strong trend towards effectively mitigating IDN homograph attacks in all products tested. However there still exists a need to ensure that location services and other potential areas of web browsers are secured in the same manner as the rest of the URL parsers & display mechanisms in the software. The lack of mitigation in some areas but not others in the same browsers suggests significant duplication of functionality in code, which is resulting in an increased attack surface. In order to better mitigate this issue it would be advantageous to consolidate these functions into single libraries which perform URL parsing, display and IDN homograph attack mitigation.

5 References

[1] Bell-ATL (2011). "Timeline of IDN." Retrieved 9th June, 2011, from http://www.bell-atl.net/articles/53357/Timeline-of-IDN.

[2] Clicky (2011, June). "Web browsers (Global marketshare)." Retrieved 8th June, 2011, from http://www.getclicky.com/marketshare/global/web-browsers/.

[3] Costello, A. (2003, March). "RFC3492 - Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)." Retrieved July 1st, 2011, from http://www.ietf.org/rfc/rfc3492.txt.

[4] Faltstrom, P., et al. (2003, March). "RFC3490 - Internationalizing Domain Names in Applications (IDNA)." Retrieved July 1st, 2011, from http://www.rfc-editor.org/rfc/rfc3490.txt.

[5] Fu, A. Y., et al. (2006). The methodology and an application to fight against unicode attacks, ACM.

[6] Krammer, V. (2006). Phishing defense against IDN address spoofing attacks, ACM.

[7] Mockapetris, P. (1983, November). "RFC882 - DOMAIN NAMES - CONCEPTS and FACILITIES."

Retrieved 9th June, 2011, from http://tools.ietf.org/html/rfc882.

[8] Mockapetris, P. (1983, November). "RFC883 - DOMAIN NAMES - IMPLEMENTATION and SPECIFICATION." Retrieved 9th June, 2011, from http://tools.ietf.org/html/rfc883.

[9] Mozilla (2005). "MFSA 2005-29: Internationalized Domain Name (IDN) homograph spoofing." from http://www.mozilla.org/security/announce/2005/mfsa2005-29.html.