12-1-2009

Assessment of Internationalised Domain Name Homograph Attack Mitigation

Peter Hannay
Edith Cowan University

Christopher Bolan
Edith Cowan University

DOI: 10.4225/75/57b405aa30dee
7th Australian Information Security Management Conference, Perth, Western Australia, 1st to 3rd December 2009

An Assessment of Internationalised Domain Name Homograph Attack Mitigation Implementations

Peter Hannay & Christopher Bolan
secau - Security Research Centre
School of Computer and Security Science
Edith Cowan University

Abstract

With the advent of internationalised domains the threat posed by non-english character sets has eventuated. Whilst this phenomenon remains well known in the development and internet industry the actual implementations of popular applications have been tested to determine their resilience to homograph based attack. The research found that most provided features that overcome such attacks, but there remain a few notable exceptions. Should an attacker take advantage of such oversights a victim would likely not be able to spot a fraudulent site or email and thus provide a perfect platform for subsequent attack.

Keywords

Homoglyphs, homographs, domain names, email, security, phishing

INTRODUCTION

With the introduction of internationalised domain names (IDN) in 2007, came the ability to use any Unicode character within a domain name (ICANN, 2007). As an unintended by-product of this measure, a new threat arose from the use of characters which are visually indistinguishable from western characters but belong to a non western script (ICANN, 2005). Such characters, known as homoglyphs, are treated as being distinct from their western counterparts when interpreted by computer software but might easily mislead a user. For example the western 'a' character when compared to the Cyrillic 'a' glyph where these are two distinct and separate characters when interpreted by a computer system, however when examined visually there exists no discernable difference between the two (Weber, 2008b). Thus after the implementation of Unicode support within operating systems and applications, the availability of homographs has become widespread (Weber, 2008a). The grouping of homoglyphs (in possible combination with western characters) to form a word is known as a homograph. These words can be comprised of multiple or single character sets, also known as scripts, from this the terms single-script and multi-script homograph are formed. The introduction of Unicode (specifically UTF-8) support in operating systems and applications has lead to a vastly increased number of available homographs. Specifically between western, Cyrillic and Greek character sets (Potter, 2005).

The availability of homographs led to a spate of published threat warnings against web based services utilising the character sets to deceive users (Holgers, Watson, & Gribble, 2006). Attacks were proposed that used homograph domain names to trick users into visiting and trusting a website based on the visually identified domain names, however there has been no documented reports of this attack being used other than as a proof of concept. Regardless of the fact that these attacks have not been seen in the wild, various web browsers & software products claim to have implemented safeguards against IDN homograph attacks (Milletary, 2005). The research presented in this paper aims to provide an overview of the effectiveness of web browsers and email clients in mitigating a multi-script IDN homograph attack.

BASICS OF THE ATTACK

The execution of an IDN homograph attack generally involves the acquisition of a domain name that is visually indistinguishable from or extremely similar to another. An example of this would be .com as an imitation of google.com. In this case, there is a minor visible difference between the second 'g', however in a number of fonts there is no way to separate these. An example of this type of variance in a number of fonts is shown in Table 1 below. Once a suitable domain name is acquired some method of convincing users to navigate to the site is required (Mozilla, 2005). It is often theorised that phishing schemes either via email, message boards, wikis or any other service that allows user contributed content are ideal for this purpose. The goal is that a user would be able to view the link, visually assess the link's legitimacy, find no sign of illegitimacy and then navigate to the page.

Table 1 - Homograph Samples showing visual homoglyph variances in common fonts

Font	Text Sample
Arial	http://www.google.com
Times New Roman	http://www.google.com
Georgia	http://www.google.com
Cambria	http://www.google.com
Calibri	http://www.google.com
Veranda	http://www.google.com
Lucida Console	http://www.google.com

A user making an assessment of the site would see that the address bar of the browser contains a URL which appeared to be correct, the content of victim site could be replicated completely. Whilst it might be assuemed that certificates would provide a measure of defence against this scenario, it has been demonstrated that certificates are obtainable for web sites containing IDN homographs and as such the "lock" symbol that many users look for when determining authenticity would be present and appear accurate (Shmoo, 2005).

CURRENT MITIGATION STRATEGIES

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 (IETF, 2003). The punycode alternative is commonly displayed in both the address bar and the status bar on hover for a particular link. An example of this may be seen in Figure 1 below.

Figure 1: An example of Punycode in the Firefox Browser

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, Wenyin, & Little, 2006). The implications of this are that any domain which is intended to be spoofed via the replacement of only 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 utilises 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 secondary mitigation strategy involves the colour coding of various scripts in URLs (Krammer, 2006). In this method Cryllic scripts are highlighted one colour, while western scripts are left uncoloured. In this situation mixed script URLs become immediately visible to the user, even though the characters themselves are visibly identical.

TESTING METHODOLOGY

A demonstration of a domain using homographs was configured to facilitate the testing of operating systems and applications in order to determine how they interact with domain names making use of mixed character sets (Hannay, 2009). The chosen domain name "google.com" replaced the second 'g' in the well know google.com domain with UTF-8 character U 0261. The combination of western and Cyrillic scripts leads to the domain name falling into the mixed-script category and thus it was expected that it would be treated with suspicion by the majority of applications.

Web Browsers

Each web browser was installed with default settings. For each browser the aforementioned "http://www.google.com" URL was placed into the address bar and the "Go" (or equivalent) button pressed. After the domain was selected a number of factors were investigated:

1. Was it possible to view the page?
2. Were any additional alerts given by the browser?
3. Did a visual inspection of the URL show any discernable differences between the attack URL and that of the original?

The first this criteria determined the browser support for internationalised domain names, whilst the second allowed the discovery of inbuilt detection features and user alerts. Finally the a visual comparison allowed for determination on the likelihood of user based detection.

Email Clients

In addition to the web browsers investigated, a number of email clients, both web and application based, were examined in order to determine any security features that may be provided by the client in question. In each instance the ability to send to IDN homograph address of "admin@google.com" was tested and a record kept of the following: errors, bounces, contents of the "To:" field on receiving the email (if received). The capability of each client to receive email from the IDN homograph address of "admin@google.com" was also evaluated. In these cases the presentation of the address in the from field was evaluated to see if a user would be able to differentiate an email from an actual address from its homograph counterpart.

RESULTS

Having completed the testing for each web browser and email client, it became apparent that the IDN homograph attack is still viable despite the awareness of such attacks. The results of testing with web browsers (shown in table 2) found that the majority of web browsers tested are not vulnerable to the attack, converting addresses to punycode as was shown previously in figure 1. The conversion to punycode allowed the tester to visually identify the web site as illegitimate. The results for Opera 10 are of note, as it did not address the multi-script IDN in any way, instead showing the domain in the address bar in its raw UTF-8 form, this result is illustrated in figure 2. Such findings contradict existing claims that the issue had been addressed in the Opera browser (Opera, 2009). A number of less common browsers also appeared to be vulnerable to the attack, however no documentation on wether the attack had been discovered or addressed was found within the literature.

Figure 2 - The IDN Homograph Attack shown in Opera 10.00 1750

Table 2 - Web Browser Results

	Mitigating Features			
Web Browser	**Converts address to Punycode**	**No conversion, with no visual distinction possible**	**No conversion, with visual distinction possible**	**Color coding or other mitigating feature**
Chrome 4.0	**X**			
Konqueror 4.2.4	**X**			
Firefox 3.5.3	**X**			
Internet Explorer 7	**X**			
Internet Explorer 8	**X**			
Opera 10.00 1750		**X**		
Maxthon 2		**X**		
Avant 11.7		**X**		
Flock 2.5.2	**X**			

When sending email the majority of email clients were unable to send email to mixed-script IDN addresses, reporting that the recipient is invalid (Shown in table 3). However the researchers were unable to determine if this result was due to a security feature or if the clients simply did not support Unicode encoded recipient addresses. Of the major applications tested, both Outlook and Yahoo web mail converted the address into punycode, thus informing the user of the possibly unintended recipient. Of note was Apple's Mail.app as it was the only email client tested that was able to successfully send email to a multi-script IDN without converting it into punycode.

Table 3 - Email Clients - Sending Mail Results

Client	**UTF-8 Supported**	**Mail Sent**	**Mitigating Features**
Mail.app 4.1	Yes	Yes	None
Alpine 1.10	Yes	No	Displays "Invalid Recipient"
MS Outlook 2007	Yes	Yes	Converts address to punycode
Thunderbird 2.0.0.23	Yes	No	Displays "Invalid Address"
Gmail (10/19/2009)	Yes	No	Displays "Invalid Address"
Hotmail (10/19/2009)	Yes	No	Displays "Invalid Address"
Yahoo (10/19/2009)	Yes	Yes	Converts address to punycode

Table 4: Email Clients – Receiving Mail Results

Client	**UTF-8 Supported**	**Mail Received**	**Mitigating Features**
Mail.app 4.1	Yes	Yes	None
Alpine 1.10	Yes	Yes	Displays punycode
MS Outlook 2007	Yes	Yes	Displays ASCII: "admin@goo??le.com"
Thunderbird 2.0.0.23	Yes	Yes	List View: Shows slight variance in font of second 'g' Message View: None
Gmail (10/19/2009)	Yes	Yes	Displays punycode
Hotmail (10/19/2009)	Yes	Yes	List View: None Message View: Displays ASCII "admin@goo??le.com"
Yahoo (10/19/2009)	Yes	Yes	Shows slight variance in font of second 'g'

The results of each email client's ability to receive email from multi-script IDN addresses are shown in Table 4. From the results we can see that Mail.app had no mitigating features, this is interesting as it also has the ability to send email to such addresses. The implication of this is that it would be possible to impersonate a specific email address and receive replies without the user ever being aware of a possible attack. Examination of Alpine & Gmail browsers revealed that the addresses were in both cases converted to punycode, in this case users would be aware that the email did not originate from the intended domain, even if they are not aware of IDN. In the case of Outlook and Hotmail it appears that non-ascii is not supported in the address field, thus question mark characters were shown in place of the non-western script

character. Finally examination of the Thunderbird and Yahoo clients show the homoglyph character in a slightly different manner, an example of this is shown in figure 3.

Figure 3 - Inbox from Yahoo mail, showing variance in homoglyph characters.

CONCLUSION

The results of testing show that it is still possible to conduct phishing attacks using IDN homographs. However in order for these attacks to provide no visual distinction to the victim a very specific subset of email clients and web browsers would be required, thus limiting but not eliminating the likelihood of success. Adding to the threat posed by such attacks is the apparent vendor claims of mitigation where none in fact existed, as well as the susceptibility of a major operating systems default mail client. With the increasing internationalisation of internet protocols, it has never been more important that homograph based attacks are considered and measures implemented to ensure their mitigation. As the rise in both internet based attacks and the potential for harm increases, so do the complexities in successfully mitigating exploit of this service. The move towards a standard methodology for addressing Unicode based attacks is therefore essential if users are to be able to identify potentially fraudulent activities in support of automated methods. Without the implementation of a standardised and verifiable approach all users will remain highly vulnerable as no degree of user education would lessen the actual risk with solutions necessary at a technical level.

REFERENCES

Fu, A., Deng, X., Wenyin, L., & Little, G. (2006). *The methodology and an application to fight against unicode attacks*.

Hannay, P. (2009). Google: Awesome Edition. Retrieved 30th June, 2009, from http://www.google.com

Holgers, T., Watson, D., & Gribble, S. (2006). Cutting through the confusion: A measurement study of homograph attacks.

ICANN. (2005). ICANN | ICANN Statement on IDN Homograph Attacks and Request for Public Comment. Retrieved 22nd March, 2009, from http://www.icann.org/en/announcements/announcement-23feb05.htm

ICANN. (2007). ICANN | On Its Way: One of the Biggest Changes to the Internet. Retrieved 22nd March, 2009, from http://www.icann.org/en/announcements/announcement-2-09oct07.htm

IETF. (2003). RFC 3492 - Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA). Retrieved 22nd March, 2009, from http://tools.ietf.org/html/rfc3492

Krammer, V. (2006). *Phishing defense against IDN address spoofing attacks*.

Milletary, J. (2005). Technical trends in phishing attacks. Retrieved 22nd March, 2009, from http://www.uscert.gov/reading_room/phishing_trends0511.pdf

Mozilla. (2005). MFSA 2005-29: Internationalized Domain Name (IDN) homograph spoofing. Retrieved 22nd March, 2009, from http://www.mozilla.org/security/announce/2005/mfsa2005-29.html

Opera. (2009). Advisory: Internationalized domain names (IDN) can be used for spoofing. - Opera Knowledge Base. Retrieved 22nd March, 2009, from http://www.opera.com/support/kb/view/788/

Potter, B. (2005). Dangerous urls: Unicode & IDN. [doi: DOI: 10.1016/S1353-4858(05)00210-2]. *Network Security, 2005*(3), 5-6.

Shmoo. (2005). The state of homograph attacks. Retrieved 22nd March, 2009, from http://www.shmoo.com/idn/homograph.txt

Weber, C. (2008a). The Lookout : Unicode security attacks and test cases Visual Spoofing, IDN homograph attacks, and the Mixed Script Confusables. Retrieved 22nd March, 2009, from http://www.lookout.net/2008/12/09/unicode-attacks-and-test-cases-visual-spoofing-idn-homograph-attacks-and-the-mixed-script-confusables/

Weber, C. (2008b,). The Lookout : Unicode security attacks and test cases Visual Spoofing, IDN homograph attacks, and the Single Script Confusables. Retrieved 22nd March, 2009, from http://www.lookout.net/2008/12/03/unicode-idn-homograph-attacks-and-test-cases-visual-spoofing-and-the-single-script-confusables/

COPYRIGHT