The 2010 IDN Homograph Attack Mitigation Survey

Peter Hannay
secau – Security Research Centre
School of Computer & Security Science
Edith Cowan University
Perth, Western Australia

Christopher Bolan
secau – Security Research Centre
School of Computer & Security Science
Edith Cowan University
Perth, Western Australia

The advent of internationalised domains has introduced a new threat with the non-english character sets allowing visual mimicry of common domain names. Whilst this phenomenon remains well known in the development and internet industry the actual implementations of popular applications have been previously shown to lack successful mitigation strategies and countermeasures. The research found that in the current versions of most internet browsers and email clients, some form of homograph identification or blocking exists. However, some notable and popular applications include either flawed implementations or miss key features and thus allow for IDN based attacks.

Keywords: homoglyphs, homographs, domain names, email, security, phishing

I. INTRODUCTION

With the acceptance internationalised domain names (IDN) by the Internet Corporation for Assigned Names and Numbers users of the internet where able to utilize any Unicode character in constructing a domain name [1]. Whilst on the surface the change heralded a shift from the dominance the English language had on internet infrastructure an unintended by-product of this measure has meant that domain names may be spoofed using characters which are visually indistinguishable from western characters but belong to a non western script [1].

Such characters, known as homoglyphs, are despite visual similarities, treated as distinct from their visual twins [2]. An example of such an occurrence may be seen when comparing the western ‘a’ character to the Cyrillic glyph ‘а’; whilst, interpreted by computer based systems as distinct and separate characters there exists no discernable difference between the two [3]. Despite such concerns, the addition of Unicode support has become widespread amongst internet enabled software and operating systems [2][4].

The grouping single character homoglyphs with either other homoglyphs or other characters to form a visual twin of a known word is referred to as a homograph. Homographs may be constructed across a single or multiple character set or script and thus the differential terms single-script and multi-script homograph are found throughout the literature [2].

A range of attacks have been documented that utilize homographs to trick users into visiting and trusting a website based on the visually identified domain names [5]. In direct response various web browsers & software products now claim to have implemented safeguards against IDN homograph attacks [6]. The research presented in this paper thus aims to test the effectiveness of mitigation and prevention strategies implemented across popular web browsers and email.

II. SETTING UP A HOMOGRAPH DOMAIN

The first step in constructing an IDN based homograph attack is the creation of a suitable domain. An ideal homograph requires a domain name that is visually indistinguishable from or extremely similar to another. For testing purposes the authors registered two homographs domains goоgle.com and sec.com. The selection of these homographs illustrate the two subtypes of possible construction – whilst goоgle.com maybe clearly seen to have subtle differences the homograph sec.com consists of entirely visually indistinguishable homoglyphs for the ‘sec’ domain. The table below illustrates the two constructed and registered domains demonstrating the variation against typical fonts.

TABLE I. COMPARISON OF HOMOGRAPH ACROSS FONTS

FONT	HOMOGRAPH	ORIGINAL
Arial	sec.com	sec.com
	goo□le.com	google.com
Times New Roman	sec.com	sec.com
	goo□le.com	google.com
Georgia	sec.com	sec.com
	goo□le.com	google.com
Cambria	sec.com	sec.com
	goo□le.com	google.com
Calibri	sec.com	sec.com
	goo□le.com	google.com
Veranda	sec.com	sec.com
	goo□le.com	google.com
Lucida Console	sec.com	sec.com
	goo□le.com	google.com

It is important to note that there is no barrier to the registration of such homoglyphs domains, and thus all the domains presented in this paper were registered with the required organizations.

III. 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 [7]. 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 [8]. 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 [9].

A final strategy involves the colour coding of various scripts in URLs [10]. 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.

IV. METHOD OF TESTING

As previously discussed the first step in validating the existence of controls or mitigation strategies for IDN Homograph attacks was to register two domains. The first selected domain name “goo□le.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. The second domain sec.com used a mixed script approach also but in this case the entire domain set consisted of homoglyphs rather than just a single character.

A. Web Browsers

Each web browser was installed with default settings into a clean virtual machine setup running Windows 7. For browser test the URLs were placed directly into the address bar and the “Go” (or equivalent) button pressed. After the domain was selected a number of factors were evaluated:

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

The first criterion determined the browser support for internationalised domain names, whilst the allowance of IDNs is a criteria for compliance to current internet standards it was not assumed. The second factor was used to highlight inbuilt detection features and user alerts, as whilst many browsers claim to provide a solution the details are often lacking in the literature [2]. The final test called for a visual comparison which was carried out to provide the likelihood of user based visual detection.

B. Email Clients

For the testing of email clients there were two facets for investigation:

- The sending of an email to a homoglyph domain
- The receipt of an email from a homoglyph domain

For each test of particular note were the errors, bounces (if any) and the visual inspection of sender and recipient fields. If an email client was unable to send to a homoglyph domain it would be unlikely that a user would inadvertently respond to a IDN phishing email. Likewise the receipt or inability to receive such emails would also provide crucial insights into the likelihood of an email based IDN approach.

V. RESULTS

A. Web Browsers

The first set of results from web browsers is tabulated below. Across both registered IDNs the Chrome, Konqueror, Firefox, Internet Explorer, Flock and Safari browsers all were able to differentiate the homograph and converted the URL to punycode for display. It is important to note this conversion happened after the address was entered and thus whilst visual distinctive after submitting the address, prior to submission the address appeared authentic.

TABLE II. IDN HOMOGRAPH RESULTS FROM BROWSER TESTING

Web Browser	Version	Punycode Conversion	Visual Distinction
Chrome	4.1.249.1064	X	
Firefox	3.6.3	X	
Internet Explorer 7	7	X	
Internet Explorer 8	8	X	
Opera	10.53		X
Maxthon 2	2.5.12.4586		X
Avant	11.7		X
Flock	2.5.6	X	
Safari	4.0.5	X	X

The Opera, Maxthon and Avant browsers failed to allow for punycode conversion and thus displayed the URLs in Unicode. This meant that users of these browsers would not be able to visually identify the use of an IDN homoglyph and thus would be more likely to fall victim to an attack of this nature.

Finally, the Apple browser Safari provide interesting results as it differs behaviour depending on when the IDN address is accessed. With this browser, if the first URL visited following execution is an IDN homograph it displays without conversion, if another URL is loaded prior to the IDN address then punycode conversion is shown.

B. Mail Clients

The first table in this section details the results for sending emails to the sec.com and goo**g**le.com email addresses. As may be seen two out of the three online clients were able to send/reply with no mitigating features. For the installed clients the two most popular Outlook and Mail.app both allowed emails to be sent, however Mail.app did prompt the user with a warning.

TABLE III. SENDING MAIL TO IDN HOMOGRAPH ADDRESSES

Mail Client	Version	Mail Sent	Mitigating Features
Mail.app	4.2	Yes	Prompt user stating warning
Alpine	2.0.0	No	Invalid recipient error
Outlook	12.0.6316	Yes	None
Thunderbird	3.0.4	Yes	None
Gmail	12/04/10	No	Reports that characters in addresses are not properly formed
Hotmail	12/04/10	No	Will not send due to non standard characters Does not support unicode
Yahoo	12/04/10	Yes	None

Table IV below details the email clients ability to receive emails from the homoglyph domain accounts. Every mail client was able to receive from the homograph email accounts with only Alpine, Gmail and Hotmail providing any mitigation.

TABLE IV. RECEIVING MAIL FROM IDN HOMOGRAPH ADDRESSES

Mail Client	Version	Mail Received	Mitigating Features
Mail.app	4.2	Yes	None
Alpine	2.0.0	Yes	Displays punycode
Outlook	12.0.6316	Yes	None
Thunderbird	3.0.4	Yes	None
Gmail	12/04/10	Yes	Displays punycode
Hotmail	12/04/10	Yes	Shows address domain as ?????
Yahoo	12/04/10	Yes	None

VI. CONCLUSION

The results from this research clearly show that whilst a majority of web browsers do provide mitigation against homographs and homoglyphs used in IDNs. There are however a small number that fail to provide basic support in the form of punycode conversion. In the case of Maxtor this may be a byproduct of its large non-english usage with most of its 300-million downloads occurring in China. The Avant browsers coverage is significantly smaller with only 22 million downloads but is offered in the European economic zone as one of the twelve recommended alternatives to Internet Explorer.

Perhaps of greatest concern in the web browser results is for the Opera browser as it is commonly used in devices such as mobile phones and gaming systems as the only browsing option available. With this in mind, it would appear that the lack of punycode mitigation against homograph IDNs will mean that users of portable and alternative devices are at a greater risk of falling for attacks.

The mail client results in this study were also of marked interest. Whilst all the clients claimed to support Unicode characters 4 of the 7 clients managed to send an email to an address at a Unicode domain. Of the 7 clients none appeared to carry out punycode conversion prior to sending the email and thus provided no visual mitigation to users. Several of the remaining clients would not accept the addresses and displayed invalid address errors when an attempt was made to send to a Unicode address.

The reception of emails from homograph domains is perhaps the most critical of the three factors investigated in this study. Phising emails are often spotted by users based on a visual inspection of the senders address. The results in this study show that of the clients tested, only hotmail, gmail and alpine would provide any feedback as to the use of homographs in the email domain.

Overall the findings of this research demonstrate that homograph IDN based attacks are still possible despite clear documentation of the problem stemming back to the drafting of the IDN standards. To highlight the impact of these findings a possible attack may be considered. Through the use of a homograph domain a target would receive an email that would look visually identical to the correct domain. This email could even be crafted with identical looking web links which if clicked would display sites both visually indistinguishable in content and in the domain address bar. Trusting the email the victim could then reply and thus send the response to the attacker potentially leaking sensitive information.

However, in a climate of security threats against home users, often with little or no domain expertise, the fact that any browser or email client allow this to occur is disappointing. At a minimum those with interests in protect users from fraudulent emails and sites, should insist that the use of punycode enabled applications is mandatory.

REFERENCES

[1] ICANN, "ICANN Statement on IDN Homograph Attacks and Request for Public Comment," 2005.

[2] P. Hannay and C. Bolan, "An Assessment of Internationalised Domain Name Homograph Attack Mitigation Implementations," in *7th Australian Information Security Management Conference* Perth, Western Australia, 2009.

[3] C. Weber, "The Lookout : Unicode security attacks and test cases Visual Spoofing, IDN homograph attacks, and the Mixed Script Confusables," 2008.

[4] C. Weber, "The Lookout : Unicode security attacks and test cases Visual Spoofing, IDN homograph attacks, and the Single Script Confusables," 2008.

[5] T. Holgers, D. Watson, and S. Gribble, "Cutting through the confusion: A measurement study of homograph attacks," 2006.

[6] J. Milletary, "Technical trends in phishing attacks," *Retrieved December*, vol. 1, p. 2007, 2005.

[7] IETF, "RFC 3492 - Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)," 2003.

[8] A. Fu, X. Deng, L. Wenyin, and G. Little, "The methodology and an application to fight against unicode attacks," 2006, p. 101.

[9] Mozilla, "MFSA 2005-29: Internationalized Domain Name (IDN) homograph spoofing," 2005.

[10] V. Krammer, "Phishing defense against IDN address spoofing attacks," 2006, p. 32.