Considering the Plausibility of IDN Homograph Attacks on iOS

Abusing Unicode Character "ɴ" (U+0274) to Spoof News Media Organizations

Tyler Butler

Abstract

The introduction of International Domain Names (IDNs) drastically increased the potential availability of homograph exploits- the use of Unicode characters to create misleading information, most notably domain names. In the years since, several mitigation strategies have been deployed by leading developers to protect users from risk. Despite these developments, Apple's iOS devices continue to be susceptible IDN homograph exploits. The research builds on previous work and considers the plausibility of abusing design features in Apple iOS devices to exploit IDN homographs to spread misinformation and targeted malware, specifically with the ability to spoof popular news media outlets. It finds that iOS devices continue to be susceptible to such attacks, and describes techniques used to abuse features in iMessage, Messages, and Safari Web Browser This flaw leaves iOS users at significant risk of spoofing attacks which can be used to spread misinformation, steal credentials, or deliver targeted malware. Of particular concern is the ability for hostile governments to use the described techniques to target journalists with spyware. The research was presented to Apple in December of 2020, and the vendor has identified they will not be issuing a fix.

Keywords

International Domain Names, Punycode, Spoofing, ASCII, Unicode

INTRODUCTION

The Domain Network System (DNS) was a foundational internet protocol that dictated the available characters which could be used to name websites. Its nascent form included the 256 characters in the popular character encoding set ASCII (The American Standard Code for Information Interchange). This satisfied the needs of Western societies as the American standard includes all letters in the English alphabet, numbers, control characters, and special characters like punctuation marks. Non-English speakers would not be able to register domain names using their native language until 2007 when International Domain Names (IDNs) were first introduced.[1] IDNs enabled additional characters sets to be used such as Chinese, Portuguese, and Scandinavian languages. While providing benefits the protocol brought criticism for enhancing the ability to register domains that are homographs.[2] A homograph is a word formed using homoglyphs- western characters that are identical or nearly identical to non-western characters.

The concern that homograph IDNs could be used as a method for spoofing popular domains forced developers to adopt a wide range of mitigation strategies. These strategies included representing IDNs in punycode, whitelisting select IDNs, and color-coded scripts. Despite available options, many applications have been slow to adopt a comprehensive solution.[3] For example, Mozilla's Firefox uses a combination of IDN whitelist and a custom algorithm.[4] According to the policy, certain whitelisted IDN's are always shown in punycode, the rest are determined based on a series of restrictions identified by the Unicode Technical Standard 39, such as common + inherited + a single script.[5] Apple first started to address the issue with IDNs in 2010. The first version of Safari web browser to use punycode was its Safari 5 release in mid 2010.[3] While Apple's Safari maintains a whitelist[6], the threat of unknown or unblocked IDNs still remains at large.

METHOD

To demonstrate the risks of current iOS design features related to the display of IDNs, a proof of concept was established to create false and misleading websites spoofing top news reporting agencies. The technique involves 3 main components; identification and registration of homograph domains, development of cloned websites, and smishing users with IDN exploits. Once each step was complete, the results of the iOS IDN display was then compared against Google Chrome. The POC is tested on iOS Messages and Safari for MacBook (macOS Catalina 10.15.7) and for iPhone (iPhone Xs Max 14.0.1).

---

[1] ICANN Successfully Conducts Laboratory Tests of Internationalised Domain Names. (n.d.). Retrieved October 18, 2020, from https://www.icann.org/news/announcement-4-2007-03-07-en

[2] ICANN Statement on IDN Homograph Attacks and Request for Public Comment. (n.d.). Retrieved October 18, 2020, from https://www.icann.org/news/announcement-2005-02-23-en

[3] Hannay, P. , & Baatard, G. (2012). The 2011 IDN Homograph Attack Mitigation Survey. Proceedings of International Conference on Security and Management (SAM'12) . (pp. 653-657).

[4] IDN Display Algorithm. (n.d.). Retrieved October 18, 2020, from https://wiki.mozilla.org/IDN_Display_Algorithm

[5] Unicode Security Mechanisms. (n.d.). Retrieved October 18, 2020, from http://www.unicode.org/reports/tr39/

[6] J. Al Helou and S. Tilley, "Multilingual web sites: Internationalized Domain Name homograph attacks," *2010 12th IEEE International Symposium on Web Systems Evolution (WSE)*, Timisoara, 2010, pp. 89-92, doi: 10.1109/WSE.2010.5623562.

Identification and Registration of Homograph Domains

This report does not make an assessment of the amount of available homograph domains not whitelisted and thus still in circulation, nor does it consider a wide range of homoglyphs that can be used for such attack. Because this research was limited to specifically target IDN attacks related to misinformation, characters that are most often used in domains for media websites were considered. Unicode Character "ɴ" (U+0274) was chosen to be an ideal homoglyph because it is used in many media domains. Of the top 50 newspapers by average Sunday circulation for Q3 2015, Q3 2016, Q3 2017 and Q3 2018 identified by Pew[7], 68% used the character "n". Table 1.0 Shows a small subset of such domains and a homograph exploit using Unicode Character "ɴ" (U+0274) to replace one "n". Due to its popularity in American discourse, The New York Times was chosen to be the domain for this research, and the homograph https://www.ɴytimes.com was registered on Google Domains. Another plausible homoglyph that was considered was Unicode Character "ȷ" (U+0237), which can replace "j" in sites like The Wall Street Journal (wsj.com).

Organization Name	Original Domain	IDN Homograph Domain	Unicode Character
The New York Times	https://www.nytimes.com/	https://www.ɴytimes.com	"ɴ" (U+0274)
The New York Post	https://nypost.com/	https://www.ɴypost.com	"ɴ" (U+0274)
NPR	https://nypost.org	https://www.ɴpr.com	"ɴ" (U+0274)
Fox News	https://www.foxnews.com	https://www.foxɴews.com	"ɴ" (U+0274)
ABC News	https://www.abcnews.com	https://www.abcɴews.com	"ɴ" (U+0274)
NBC News	https://www.nbcnews.com	https://www.nbcɴews.com	"ɴ" (U+0274)
CBS News	https://www.cbsnews.com	https://www.cbsɴews.com	"ɴ" (U+0274)
The Wall Street Journal	https://www.wsj.com	https://www.wsȷ.com	"ȷ" (U+0237)

Table 1.0 IDN Homograph Proof of Concept Domains

Development of Cloned Domain Websites

The website to serve as the spoofed domain was developed using Jekyll, a static website generator written in Ruby. First, a Jekyll theme was selected for its resemblance to the New York Times domain. The Aspirethemes type theme was selected for its minimalistic design, making it easily adaptable[8]. To make the site resemble the New York Times, raw html content was copied from the original article using the Google Chrome Inspector Tools. The entire header and body HTML contents were copied and pasted into the theme template. Figure 1 shows a screenshot of copying the html content.

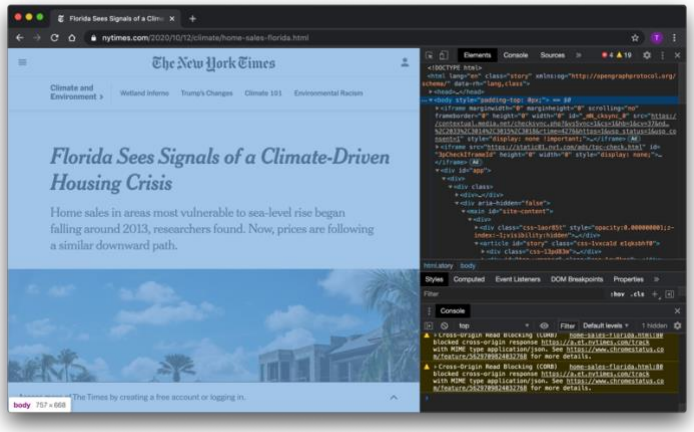


Figure 2: Copying Header and Body Source Code from The New York Times

In order to ensure that iMessage previews would be the same between the real website link and the spoofed one, it was crucial to copy meta tag headers used to generate such previews. With the same tags being used in both sites, the previews would also be the same. Generating the site preview using the local Jekyll development server revealed the resulting spoof site was a nearly identical to the original site. A small selection of JavaScript functions did not work because the original code searched for .js files from a relative path /vi-assets/static-assets/. This was fixed by replacing all instances with the original link path https://www.nytimes.com/vi-assets/static-assets/. GitHub[9] was used to store the source code for the new spoofed site, and Netlify was used to host the domain.

[7] Methodology: State of the News Media. (2020, May 30). Retrieved October 18, 2020, from https://pewresearch-org-preprod.go-vip.co/journalism/2019/07/23/state-of-the-news-media-methodology/

[8] Aspirethemes. (n.d.). Aspirethemes/type. Retrieved October 18, 2020, from https://github.com/aspirethemes/type

[9] tcbutler320 - Overview. (n.d.). Retrieved October 18, 2020, from https://github.com/tcbutler320/

Smishing Users with IDN Exploits

Delivering the IDN homograph exploit to a sample user was achieved by sending a link to the spoof domain through iMessage. It did not matter whether the link was sent in IDN or punycode form, as iMessage preview automatically converted the punycode domain back into its IDN equivalent. Figure 2 shows this conversion process. iMessage preview is pivotal to the relevance of this exploit, as additional information shown to the user such as the preview image and subject line add to the authenticity of the link.

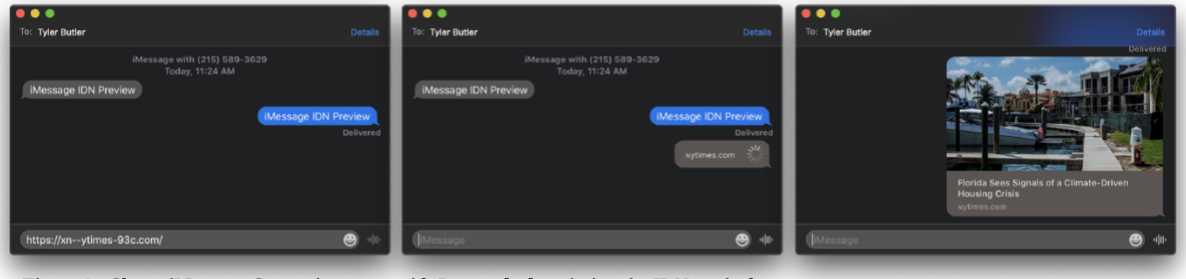


Figure 2: Shows iMessage Converting a text with Punycode domain into its IDN equivalent

RESULTS

Overall using Unicode Character "ɴ" (U+0274) as an IDN homograph exploit on iOS devices was determined to be an effective method of spoofing news media websites and fooling users into trusting an attacker owned domain. The most effective threat vector is to use iMessage for iPhone, as there is no option to "hover over" the link, which on iMessage for MacBook converts the IDN domain into its punycode form. This conversion was the only noticed defense employed by Apple to evade such confusion. In comparison to Google Chrome, Safari web browser did not protect users by using punycode in the website address bar.

iMessage IDN Preview Results

When the two links were previewed using iMessage on both MacBook and iPhone, the difference was barely noticeable. Both the preview image and link subject line were the same. The Unicode character "n" is slightly noticeable, but without a side by side comparison it would be difficult for the average user to know the difference. The only use of punycode found was that when hovering over the spoofed link on the MacBook the domain would pop-up in punycode, shown in Figure 4. This preview only appears if hovering for a few seconds, and a normal click of the link does not trigger the preview.

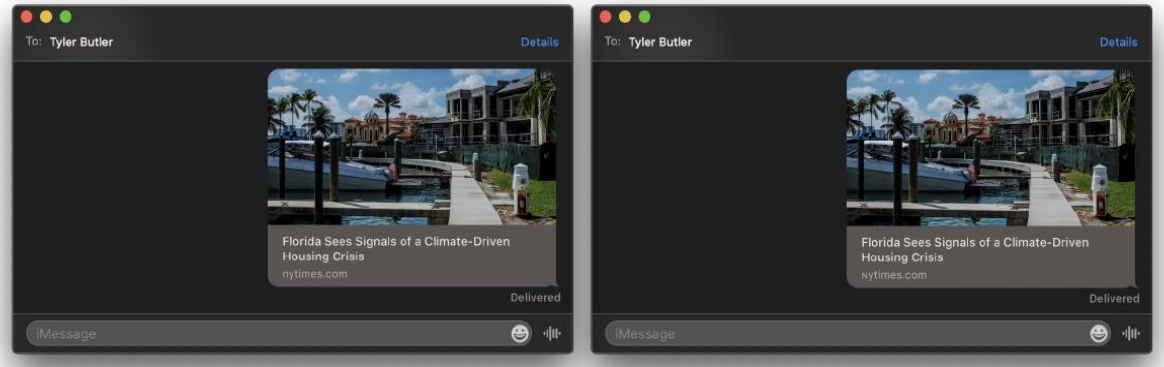
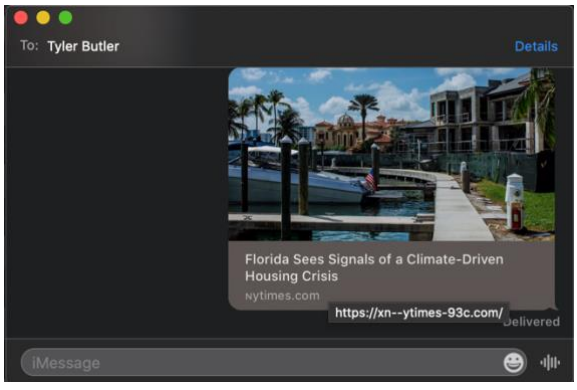


Figure 3: Previews of the real article (left) and the spoofed IDN article (right) are shown in iMessage's on MacBook



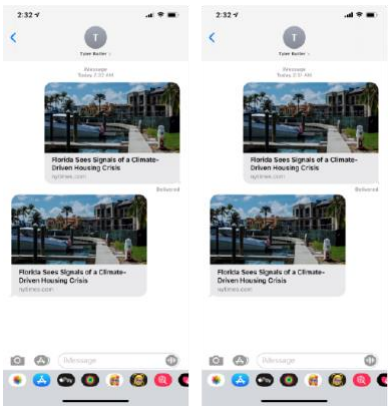


Figure 4: Hovering a cursor over iMessage Link Preview shows punycode domain

Safari IDN Preview Results

The difference in appearance between the spoofed domain and the legitimate New York Times article is barely noticeable to the naked eye on Safari. Figures 6 and 7 show that Chrome (left), shows the punycode domain name of the IDN site. Punycode is used here because this domain violates several IDN rules in use by Chrome.[10] This is in stark contrast to Safari, which does not.

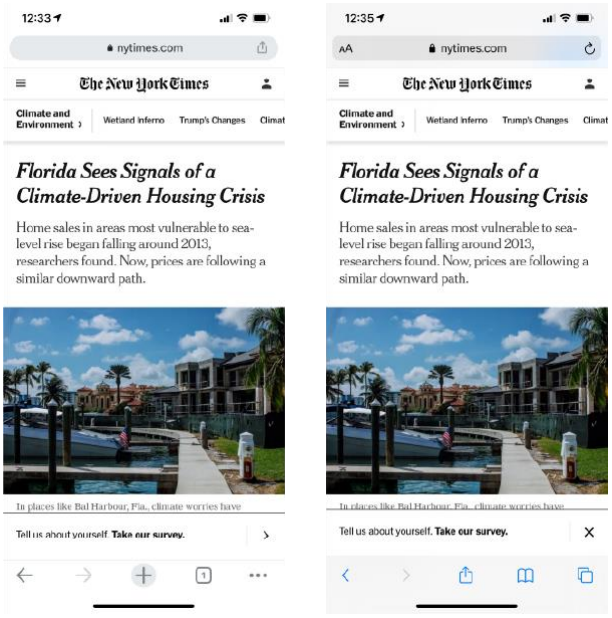


Figure 6: Chrome (left) and Safari (Right) showing the legitimate NYT article on iPhone

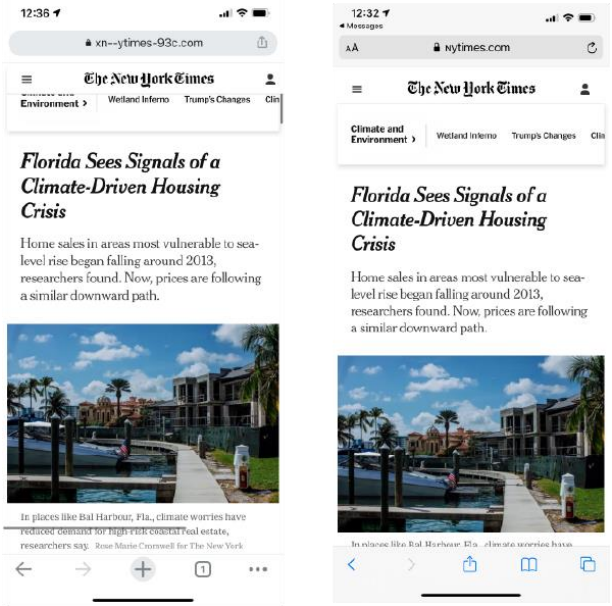


Figure 7: Chrome (left) and Safari (Right) showing Spoofed article on using the IDN Homograph Domain. Chrome shows punycode domain name

Similar results were found on MacBook. Safari shows the spoofed domain in its IDN form while Chrome shows the domain in punycode format. Other noticeable differences can be seen such as the "New York Times" Logo placement being located in different spots; however, this is due to mistakes in the cloning process.

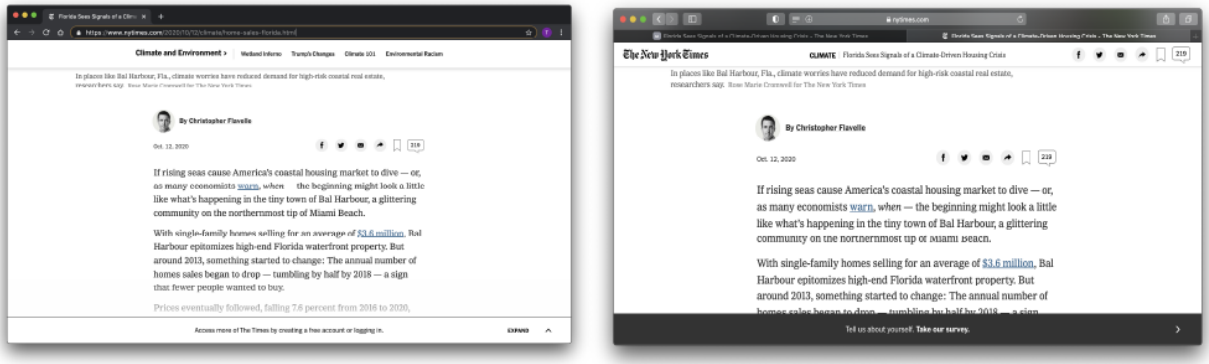


Figure 8: Chrome (left) and Safari (Right) showing the real New York Times Article on MacBook

[10] Internationalized Domain Names (IDN) in Google Chrome. (n.d.). Retrieved October 18, 2020, from https://chromium.googlesource.com/chromium/src/+/master/docs/idn.md

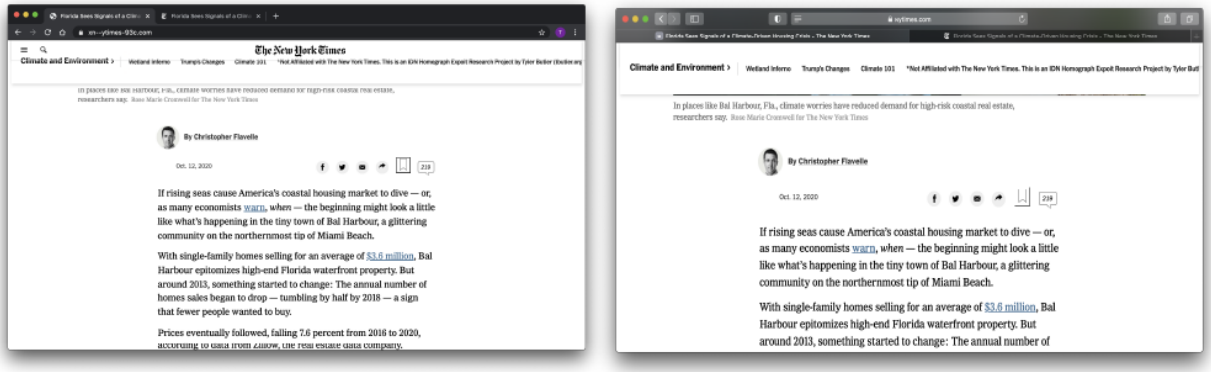


Figure 9: Chrome (left) and Safari (Right) showing the spoofed New York Times Article on MacBook

USE CASES

The research presented describes a valid social engineering method that could be employed by various threat actors to deliver targeted malware to a vulnerable user. Due to the fact that the domains used in this report already violated IDN rules on many applications, the exploit is limited for use on only iOS targets. Additionally, any target which has changed their default web browser settings to chrome or other browsers would have additional protection from this attack. For the exploit to be used successfully, research would need to be conducted to confirm the targets browser settings. Possible use cases include threat actors such as foreign intelligence services attempting to target journalists with spyware, criminal operations engaged in phishing and smishing operations, and advanced misinformation operations designed to target particular users with targeted messaging.

Example: Targeted Misinformation

An attacker can use this technique to send a target a crafted article that aims to change their views on a particular topic. An obvious pitfall to this attack is that secondary validation by the target to search for other articles on the topic would show the payload was the only source.

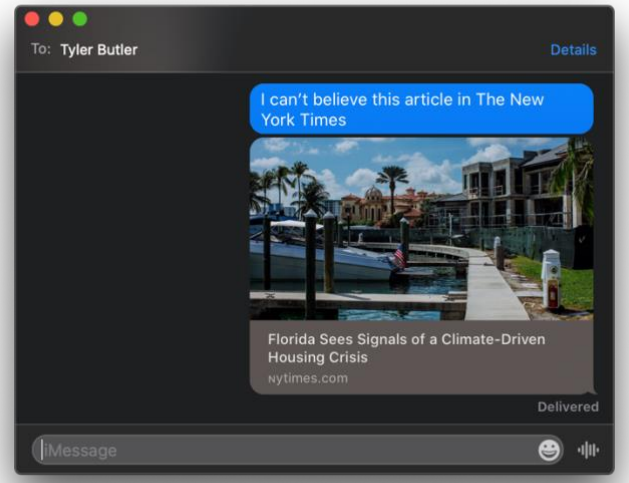


Figure 10: An example of a targeted misinformation attack

About the Researcher

Tyler Butler is a security researcher and penetration tester focusing on web application security. He has undergraduate degrees in Information Security and Criminology from The Pennsylvania State University and holds the Web Application Penetration Tester (eWPT) and Junior Penetration Tester (eJPT) certifications from eLearnSecurity. For more information, including previous research and contact information, refer to https://tbutler.org.

References

ICANN Successfully Conducts Laboratory Tests of Internationalised Domain Names. (n.d.). Retrieved October 18, 2020, from https://www.icann.org/news/announcement-4-2007-03-07-en

ICANN Statement on IDN Homograph Attacks and Request for Public Comment. (n.d.). Retrieved October 18, 2020, from https://www.icann.org/news/announcement-2005-02-23-en

Hannay, P. , & Baatard, G. (2012). The 2011 IDN Homograph Attack Mitigation Survey. Proceedings of International Conference on Security and Management (SAM'12) . (pp. 653-657).

J. Al Helou and S. Tilley, "Multilingual web sites: Internationalized Domain Name homograph attacks," *2010 12th IEEE International Symposium on Web Systems Evolution (WSE)*, Timisoara, 2010, pp. 89-92, doi: 10.1109/WSE.2010.5623562.

Methodology: State of the News Media. (2020, May 30). Retrieved October 18, 2020, from https://pewresearch-org-preprod.go-vip.co/journalism/2019/07/23/state-of-the-news-media-methodology/

Aspirethemes. (n.d.). Aspirethemes/type. Retrieved October 18, 2020, from https://github.com/aspirethemes/type

tcbutler320 - Overview. (n.d.). Retrieved October 18, 2020, from https://github.com/tcbutler320/

Internationalized Domain Names (IDN) in Google Chrome. (n.d.). Retrieved October 18, 2020, from https://chromium.googlesource.com/chromium/src/+/master/docs/idn.md