Red Panda Security

Nate Bossingham

CYSE 494

June 20, 2023

Introduction:

For the average user, cybersecurity is rarely a thought in their mind. Even those that are reasonably well informed will often adhere to the basics and feel secure. Unfortunately, cyberattacks are constantly evolving and becoming more difficult to prevent. Using an antivirus and avoiding blatant scam emails and websites is no longer enough to protect yourself from a cyberattack. Social engineering is a practice in which attackers utilize psychological manipulation tactics to convince victims to do what the attacker wants (ENISA, 2022). As technical protections become better and better, attackers have focused their efforts on improving social engineering tactics to trick and manipulate their victims. The primary social engineering-based attack vector that this paper will focus on is masquerading.

Masquerading:

"Masquerading occurs when the name or location of an object, legitimate or malicious, is manipulated or abused for the sake of evading defenses and observation." (Jerzman et al., 2023)

Masquerading has been the baseline for many attacks through the years. The phishing attack is one of the most effective and widespread attacks there is. Phishing is an attack that is accomplished in many different ways. One of the most common methods of carrying out this attack is via email. An attacker will send out an email that attempts to fool the victim into thinking the email is from a legitimate source. The email will typically contain a link or a file that contains malicious content (Itkin et al., 2023). There are many ways that an attacker can make an email and attachments appear more legitimate. This can be accomplished by modifying the header of an email or the metadata of an attachment (Itkin et al., 2023). An attacker can also hijack code from an original source to convince the user that the email is coming from an official

source. Still, most users know to check the email address that sent an email before trusting the email. Unfortunately, attackers have utilized another strategy in the past to combat this.

Typo-squatting:

"Also known as URL hijacking, [typo-squatting] is a form of cybersquatting (sitting on sites under someone else's brand or copyright) that targets Internet users who incorrectly type a website address into their web browser" (McAfee, 2022)

Simple Typo-Squatting URLs		
google.com	→	**gooogle.com**
citationmachine.net	→	**citationsmachine.net**
attack.mitre.org	→	**attackmitre.org**
support.microsoft.com	→	**supportmicrosoft.com**

Figure 1 (Bossingham, 2023)

Typo-squatting is a simple but extremely effective way to trick users. While the original intention of typo-squatting was to catch those that made a mistake when typing a URL (See Figure 1), the use case has expanded greatly. Today, attackers use typo-squatting to trick victims into believing that an email is from a legitimate source. As seen in Figure 2, without paying close attention to the email, a victim would easily fall for many of the permutations shown.

Original Email
NoReply@expediamail.com

Simple Typo-Squatting Permutations		
NoReply@expediamail.com	→	**NoReply@expedaimail.com**
NoReply@expediamail.com	→	**NoReply@expedaimial.com**
NoReply@expediamail.com	→	**NoReply@expediamal.com**
NoReply@expediamail.com	→	**NoReply@expedlamail.com**

Figure 2 (Bossingham, 2023)

Typo-squatting has still become more advanced since this use case. The primary target of this product is the Internationalized Domain Name (IDN) Homograph Attack. This attack is a specialized version of typo-squatting that takes advantage of international alphabets in domain names. As seen in Figure 3, attackers are able to create visually identical domain names that are capable of tricking even the most vigilant potential victims.

GENUINE URL								
a	p	p	l	e	.	c	o	m
U+0061	U+0070	U+0070	U+006C	U+0065	U+002E	U+0063	U+006F	U+006D
UNICODE CHARACTER CODES								
FAKE URL								
a	p	p	l	e	.	c	o	m
U+0430	U+0440	U+0440	U+006C	U+0435	U+002E	U+0063	U+006F	U+006D
Cyrillic Small Letter A	Cyrillic Small Letter Er	Cyrillic Small Letter Er	Original Latin	Cyrillic Small Letter Ie	Original Latin	Original Latin	Original Latin	Original Latin
UNICODE CHARACTER CODES WITH ORIGINS								

Figure 3 (Bossingham, 2023)

The secondary target of this product is another common attack used to trick users regarding filenames. The Right-to-Left Override attack is another masquerading attack that utilizes the Unicode character U+202E to maliciously swap file extensions. As seen in Figure 4, this is accomplished by embedding the character within the filename in a way that completely fools the victim. In the example, an attacker is able to leverage the privileges of the **.scr** file extension, while still appearing as a **.pdf** file to the victim.

RTLO Attack Example	
Filename:	Display Filename:
PotentialCharacterA\u202Efdp.scr	*PotentialCharacterArcs.pdf*
U+202E is the Right-to-Left Override Character	

Figure 4 (Bossingham, 2023)

With the groundwork now set, the goals for this product can be defined. The product will consist of a browser extension with the following features:

- The browser extension will locally scan all browser content for any international Unicode characters and notify the user of their presence. *(On Release)*
- The browser extension will scan all browser content for any malicious Unicode characters, such as the Right-to-Left Override character, and notify the user of their presence. *(On Release)*
- The browser extension will utilize public databases to flag malicious links that appear in the browser. *(Future Expansion)*
- The browser extension will utilize community reporting to build a database of malicious links. *(Future Expansion)*

The primary goal of developing this product is to create a positive and safe experience for users. Standout difference for this product is the focus on an unobtrusive experience. Many software products outright block the use of internationalized domains and improperly render the RTLO character without awareness of the content or use case. This product aims to not outright block content from the user, but to give the user the information and allow them to make a decision on the safety of the content. This will be accomplished with different layers of protection based on user choices. Each protection level will be fully customizable to the preference of the user.

The initial product will be made under the name “RPS UniProtect” created by the newly created Red Panda Security firm. Based on our research, the initial product will address a much-needed sector and will give the team the needed establishment to expand into the

additional features. This establishment of the baseline product will be accomplished by targeting high risk business sectors for large-scale contracts. These contracts will be competitively priced due to the low-cost maintenance of the baseline product. This low cost is made possible due to the entirely local nature of the product. This business model will initially be similar to other trialware companies. For the average consumer, the browser extension will be available at no cost with the option to purchase a license. For the enterprise customers, perpetual licenses will be required after the trial period. Once the product has been established, the development of the supplementary software-as-a-service features can begin.

As mentioned in the primary product goals, the extension will be iterated on with expanded security features. The primary feature that will be part of the new service provided within the extension is a malicious link detection service. This service will use the same local scan of each page to detect for malicious links. This will be accomplished by isolating any links and hashing those links. After hashing the links, the hashes will be compared against a database of malicious link hashes. This is done to prevent any data leaks or data collection. Customers will be able to trust that any content on their browser will remain local at all times. A diagram of this system can be found in Figure 5. The database of malicious links will be a combination of existing databases and user contributed links.

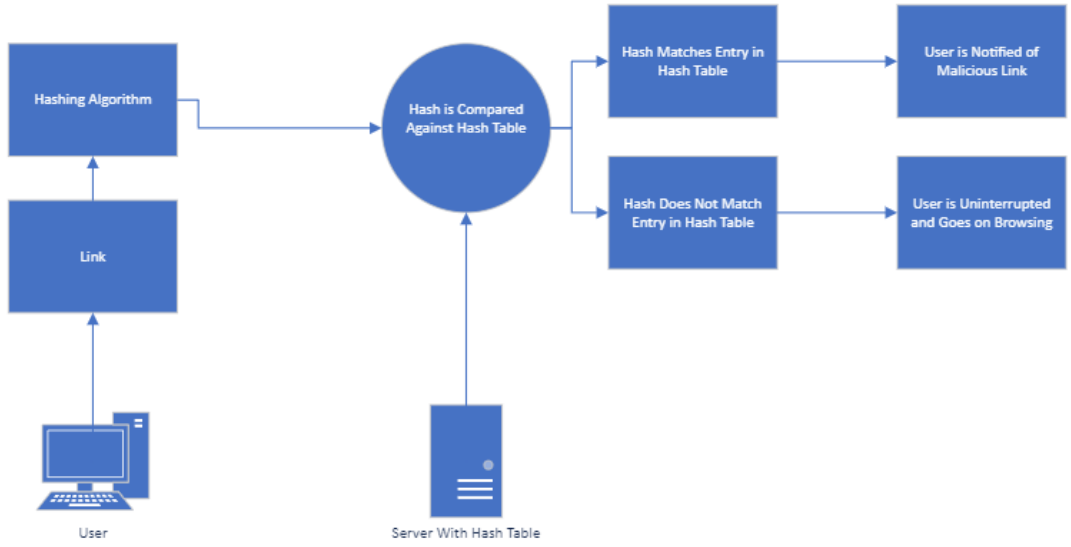

Figure 5 (Bossingham, 2023)

Research:

In the research for this product, a clear need for a solution became clear. According to IBM, a data breach cost an average of 4.35 million USD globally (IBM, 2022). Going further into this report, phishing attacks were the second most common and cost the most on average at 4.91 million USD per attack, as seen in Figure 6 (IBM, 2022). These two statistics enforce that phishing is one of the most dangerous attacks that can affect a business. The low effort of phishing combined with its impact creates one of the most effective attack vectors in the current cyberspace. As seen in Figure 7, IBM also reported that phishing is the third most difficult attack to detect (IBM, 2022). The more time an attack takes to detect, the more impactful the attack can be. In the case of phishing attacks, the average time to detect the attack is 219 days, and it takes another 76 days to then contain that attack, making it the second most time consuming attack to contain (IBM, 2022).

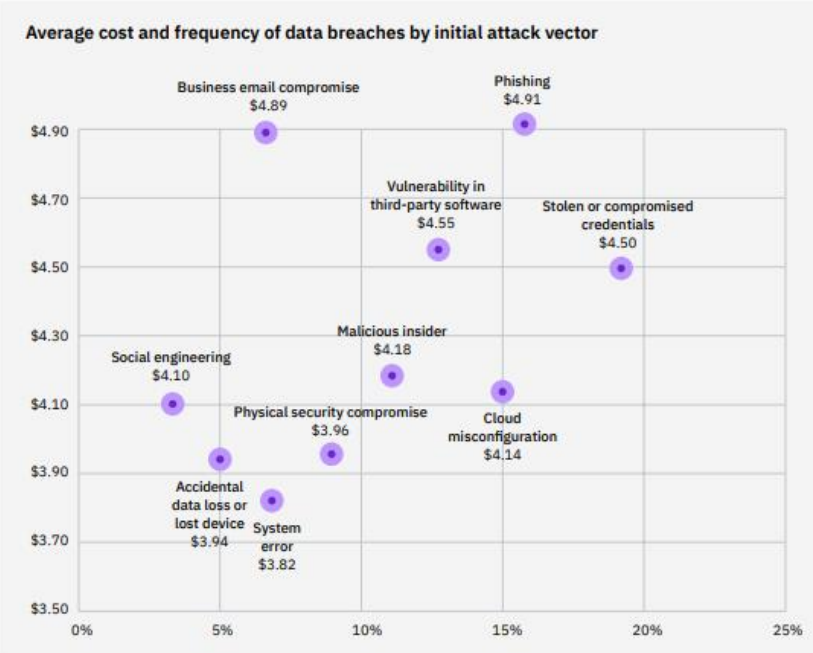

Figure 6 (IBM, 2022)

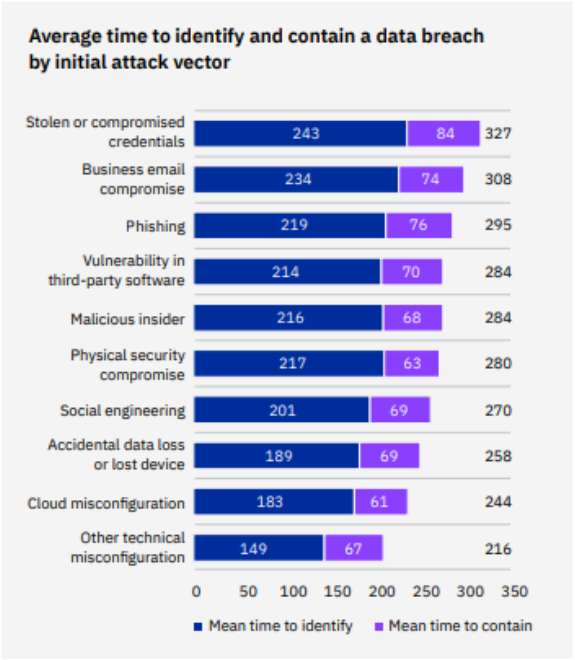

Figure 7 (IBM, 2022)

From this data, it can be understood that the phishing attack is one of the most dangerous attacks that can occur. The financial impact of a data breach resulting from a phishing attack is the greatest compared to any other attack vector. Pairing the financial impact of the attack with the time that the average phishing attack takes to detect and contain, it can reasonably be determined that the phishing attack is a major threat to business and consumers alike. Phishing as a whole is a massive sector and has plenty of room for competition within the subsectors. The product at hand aims to address a few specific attack vectors to perpetrate the phishing attack. This specific attack type is the IDN Homograph Attack. As explained above, this attack is a specialized application of typo-squatting in which an attacker uses international alphabets in Unicode to squat on URLs that are visually identical to the URL that the attacker is attempting to masquerade under.

The IDN Homograph Attack utilizes domain names that are encoded with Unicode instead of the standard ASCII encoding (Yazdani et al., 2020). On the surface, the concept of an internationalized domain name is a great idea. The DNS system has long been limited to the languages that utilize the Latin alphabet making it less accessible to those that speak a language not based on that alphabet. Opening up the DNS system to internationalized alphabets will make the internet an overall more accessible place for people around the globe. With this accessibility, the vulnerability of the IDN Homograph attack opens up. Due to the expansive character library that Unicode covers, there are many characters that are visually similar or identical to each other. These characters are called homoglyphs (Yazdani et al., 2020). Homoglyphs on their own pose no inherent threat to users, but attackers use these homoglyphs to create homographs (Yazdani et al., 2020). A homograph is a collection of characters that appears visually similar or identical to a different set of characters (Yazdani et al., 2020). An example of a homoglyph can be seen above in Figure 3.

Homoglyph Example													
Original Letter:	A	U+0041											
Homoglyph:	A	А	Α	A	ᴀ	A	Ａ	A	𝐀	𝐴	𝑨	𝒜	𝓐
Unicode ID:	U+0041	U+0410	U+13AA	U+15C5	U+1D00	U+A4EE	U+FF21	U+102A0	U+1D400	U+1D434	U+1D468	U+1D49C	U+1D4D0
Homoglyph:	𝔄	𝔸	𝕬	A	𝗔	𝘈	𝘼	Ａ	𝙰	𝘈	𝘼	A	𝘼
Unicode ID:	U+1D504	U+1D538	U+1D56C	U+1D5A0	U+1D5D4	U+1D608	U+1D63C	U+1D670	U+1D6A8	U+1D6E2	U+1D71C	U+1D756	U+1D790

Figure 8 (Bossingham, 2023)

With both phishing and homographs understood, the results of a study conducted by Florin Ilca and Titus Balan can show the true threat of the IDN Homograph Attack. Over the course of 3 days, 16,185 email were sent out using a phishing software called Gophish (Ilca & Balan, 2021). These emails were sent under the domain *Unit-bv.ro* which is a homograph for *Unitbv.ro* (Ilca & Balan, 2021). The fake attack was set up to trick victims into following a link

within the email to a spoofed intranet page. Once on this page, the victim is prompted to enter personal information to gain access to the intranet page, at which point the attack is considered successful (Ilca & Balan, 2021). From the 16,185 emails sent, 456 of those emails were opened by the recipient, and of those that opened the email, 314 clicked the link inside the email (Ilca & Balan, 2021). This means that the phishing email was opened about 3% of the time, a relatively low and manageable number. However, the important metric is the success rate of the email within the subsection of those that open the email. From this it can be seen that the phishing email had about a 69% clickthrough rate. From that clickthrough, 272 victims then entered personal information giving the phishing attack nearly a 60% success rate (Ilca & Balan, 2021). While this attack is not specifically an IDN-based attack, it gives valuable insight into the effectiveness of homograph attacks. This attack utilized a convincing phishing email and a single dash in a domain name to steal the personal information of over 270 individual victims in the span of 3 days. At scale, this attack affects millions every single day, and IDN is the newest tool in the box for these attackers.

There are a number of defense methods for the IDN attack. The primary defense system already in place is the Punycode conversion method built into most common browsers (Elsayed & Shosha, 2018). This method is implemented differently across many browsers, but in general the system looks at the browser URL bar and if foreign Unicode characters or mixed alphabets are present, the URL is converted to its Punycode equivalent. Punycode is an encoding format for URLs that allow Unicode encoding to be used with the DNS system instead of the standard ASCII encoding (Elsayed & Shosha, 2018). The issue with this protection is that it is extremely reliant on the user noticing that the URL has changed after a hyperlink is clicked. This system will not protect a user from clicking on the link in the first place. This system is also highly

reliant on the implementation by the browser. Overall, this defense method is great as a supplementary defense, but it falls short when used in isolation.

There are other defense strategies available to combat the IDN Homograph Attack. Firstly, there is a method that utilizes a database of all registered domains and filters the results down to all homographs of the most common websites (Quinkert et al., 2019). This method takes those homograph domains and inspects them for their apparent use case (Quinkert et al., 2019). If a homograph domain is found to be malicious, the domain can then be added to a database of malicious homographs. The issue with this approach is that it requires a large amount of data collection. Collecting newly registered domains on a regular basis could prove to be a resource intensive process without even processing those domains for homographs. Another suggested solution to the attack is an algorithm based on the Punycode conversion called PunyVis (Fouss et al., 2019). PunyVis aims to aid international users by assessing whether a Punycode URL is capable of being a homograph to a popular domain (Fouss et al., 2019). PunyVis then assigns a risk score based on how likely the given URL is a homograph attack (Fouss et al., 2019). PunyVis is a fantastic tool, however it does fall to some drawbacks. The primary issue with PunyVis is the lack of protection for the user before the user interacts with a link. PunyVis does improve on the browser protections by notifying the user of the risk level. The final suggested solution to this problem is a deep learning model (Ravi et al., 2023). This method utilizes deep learning models to methodically go on the counter-offensive by evaluating every domain a user visits and scoring it against the possibility that the URL is a homograph (Ravi et al., 2023). If that URL is determined to be sufficiently likely to be a homograph, the domain could then be blacklisted and no longer accessible by the user. This method is promising but is still in the

infancy stages. Deep learning is still resource intensive and relatively underdeveloped making it an ill fit for a consumer cybersecurity option.

At Red Panda Security, it has been a top priority to build an ethical product. The commitment to create a platform that collects as little information as possible with absolutely zero personal information collection is imperative to the company. Outside of the cybersecurity space, UniProtect will just be a background process without any thoughts attributed to it unless prompted. Throughout Ethics, Philosophy, and Interdisciplinary studies courses, Red Panda Security has gained the insights and integrity to create this product. This product will benefit from an interdisciplinary approach due to the mixture of cybersecurity and a proper user experience.

The Red Panda Security team is prepared to create this product, but a few necessary steps must be taken initially. The first step is development. The initial product will take the Red Panda Security team around 6 months to fully roll out. In this roll out period, the team will start to engage with enterprise customers directly via live demos. This sales process paired with the free consumer product will quickly build a customer base. Once a sizeable customer base is established, the development team will then pivot towards the development of the malicious software service. This service will take a much longer time to develop, but once released will provide a consistent income based on a subscription base. The initial product offered will consistently be updated as changes to Unicode and DNS occur but will take significantly less maintenance due to the locally run nature of the product.

There are several roadblocks that our team will encounter. Firstly, the initial cost of development will be substantial with several developers needed to ensure a working product by the end of the development period. Initial investment will be absolutely necessary to ensure that

the product will be released on schedule. The second and greatest roadblock the team will encounter is the sales process. The competition in the enterprise space is heavy, but the product being offered is sufficiently different from the other options available on the market. Regardless of this, the team will have a high barrier to overcome in the pitch process for enterprise customers. The primary solution to this problem will be the trial system and marketing the product as a very low-cost solution. Allowing the enterprise customers to use the product for a trial period will greatly lower the burden on the customer when it comes to assessing the value of the software. Offering the software at a low cost will emphasize volume customers but will overall make it easier for customers to utilize the product.

While the competition in the antivirus and anti-phishing software space is large, the success of this product will be measured on a few key metrics. The first and most important metric will be the effectiveness. Transparency and customer experience are the pillars of Red Panda Security and therefore the product must be effective. This will be ensured by exhaustive testing measures and a customer reporting system. This system will be used to allow customers to report any lapse in security that occurs to prevent it in the future. The second most important metric will be the userbase. Due to the marketing strategy of the company, the userbase is the core of the business. Without a large free userbase, the enterprise customers are less likely to trust the product. The final important metric for the success of the product is the enterprise licensing. With the trialware business model, the enterprise customers will be the primary income for the product. This makes the licensing income the most vital piece of the product for the first part of the product lifespan.

This product has a fantastic future and has taught the team a number of valuable lessons. As stated in the introduction, most informed users know the general idea of what a cyber attack is and how phishing can occur. Throughout the research presented in this paper, it cannot be understated how dangerous phishing attacks can be to even an individual. Anybody can be a victim of a phishing attack and as these attacks get more advanced, being informed becomes less reliable. The cyberspace changes too quickly to expect that an average user will be cognizant of every bleeding edge attack. If the project were to have been done differently, it would be beneficial to broaden the scope of the project slightly. The attacks targeted by the extension are large issues on their own but do little to prompt a user or business to utilize a standalone product to target such a specific attack vector. Instead, the better way to handle this is to offer a more complete antivirus program with those features as headline features. This was accomplished in some ways but falls short of a complete product by only targeting the malicious link blocking features. The next steps for this product are simple. The product needs to be developed and tested. Setting sights low for the initial product will handicap the sales and adoption of the product, but it will ensure that the product's core features receive the attention needed to create a quality user experience. As long as the product sticks to the core values, the future of the product is bright and promising.

References

Al Helou, J., & Tilley, S. (2010). Multilingual web sites: Internationalized domain name homograph attacks. 2010 12th IEEE International Symposium on Web Systems Evolution (WSE). https://doi.org/10.1109/wse.2010.5623562

Elsayed, Y., & Shosha, A. (2018). Large scale detection of IDN domain name masquerading. 2018 APWG Symposium on Electronic Crime Research (eCrime). https://doi.org/10.1109/ecrime.2018.8376212

ENISA. (2022, November 29). What is “Social Engineering”? ENISA. https://www.enisa.europa.eu/topics/incident-response/glossary/what-is-social-engineering

Fouss, B., Ross, D. M., Wollaber, A. B., & Gomez, S. R. (2019). Punyvis: A visual analytics approach for identifying homograph phishing attacks. 2019 IEEE Symposium on Visualization for Cyber Security (VizSec). https://doi.org/10.1109/vizsec48167.2019.9161590

IBM. (2022). Cost of a data breach 2022. IBM. https://www.ibm.com/reports/data-breach

Ilca, F., & Balan, T. (2021). Phishing as a service campaign using IDN homograph attack. 2021 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) & 2021 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM). https://doi.org/10.1109/optim-acemp50812.2021.9590028

Itkin, L., Ravich, L., Zaidenberg, O., Winther, P., & Cook, S. (2023, April 14). Phishing. Phishing, Technique T1566 - Enterprise | MITRE ATT&CK®. https://attack.mitre.org/techniques/T1566/

Jerzman, B., Lu, D., Elastic, Espósito, F., @Pr0teus, Carr, N., & Kolesnikov, O. (2023, April 7). Masquerading. Masquerading, Technique T1036 - Enterprise | MITRE ATT&CK®. https://attack.mitre.org/techniques/T1036/

McAfee. (2022, December 13). What is typosquatting?. McAfee. https://www.mcafee.com/learn/what-is-typosquatting/

Quinkert, F., Lauinger, T., Robertson, W., Kirda, E., & Holz, T. (2019). It's not what it looks like: Measuring attacks and defensive registrations of homograph domains. 2019 IEEE Conference on Communications and Network Security (CNS). https://doi.org/10.1109/cns.2019.8802671

Ravi, V., Alazab, M., Srinivasan, S., Arunachalam, A., & Soman, K. P. (2023). Adversarial defense: DGA-based botnets and DNS homographs detection through Integrated Deep Learning. IEEE Transactions on Engineering Management, 70(1), 249–266. https://doi.org/10.1109/tem.2021.3059664

Yazdani, R., van der Toorn, O., & Sperotto, A. (2020). A case of identity: Detection of suspicious IDN homograph domains using active DNS measurements. 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). https://doi.org/10.1109/eurospw51379.2020.00082