Journal of Economic Criminology 6 (2024) 100110

Contents lists available at ScienceDirect

Journal of Economic Criminology

journal homepage: www.journals.elsevier.com/journal-of-economic-criminology

Secret sharing in online communities: A comparative analysis of offender and non-offender password creation strategies

Andréanne Bergeron[a], Thomas E. Dearden[b,*]

[a] GoSecure, Montreal, QC, Canada
[b] Virginia Tech, Department of Sociology, Blacksburg, VA, USA

ARTICLE INFO

Keywords:
Cybercrime
Economic crime
Authentication
Networks
Offenders
Social identity theory

ABSTRACT

Even though several authentication methods exist, passwords remain the most common type of authentication. Researchers have demonstrated the influence of a person's environment and exposure to the Internet on their online security behavior (Bosnjak & Brumen, 2016; He et al., 2021; Juozapavičius et al., 2022). Those studies suggest that social identity seems to play a role in password choice. The objective of this study was to determine if the criminal nature of a network influences password-creation strategies. To achieve this, we utilized two databases with a substantial number of actual passwords (1485,095) that had been leaked to the Internet. One database was sourced from a non-delinquent social network, while the other was from a hacker forum. We employed logistic regression to reveal the characteristics associated with each group, ensuring a comprehensive analysis of different types of password strategies and the similarity between actors of the same network. Results show that users of the same network have passwords with characteristics that are similar to each other. Individuals with the same social interests seem more likely to use the same password-creation strategies. From a network analysis perspective, the results show that similar individuals (sharing the same interests) are similar in other aspects (password creation strategies). These findings offer valuable insights into the diverse landscape of password varieties and user behaviors, contributing to a more comprehensive understanding of internet user networks.

Introduction

Various methods have emerged for authentication, a crucial element in safeguarding data and account security in the digital realm. Yet, the use of passwords remains the most widespread. Managing a multitude of passwords, individuals often resort to strategies for enhanced memorability (Pfleeger et al., 2015; Stobert & Biddle, 2014; Ur et al., 2015). Poor passwords can significantly contribute to economic crime by making systems more vulnerable to cyber-attacks and fraud. When individuals or organizations use weak passwords or reuse passwords across multiple accounts, they create opportunities for malicious actors to gain unauthorized access to sensitive information (Nicholson et al., 2018). This is evidenced by the volume of fraud online. The FBI estimates that internet crime in 2020 resulted in $4.2 billion lost, with phishing and business email compromise schemes as the most common and costliest, respectively. (Federal Bureau of Investigation., 2020). Specifically related to passwords, Google estimates that 54.8% of breaches against Google Cloud occurred due to weak or no passwords (Office of the CISO., 2023).

Given the use of weak or stolen passwords to commit economic cybercrime, we examine password-creation strategies. Our research explores the influence of social identity and network characteristics on password creation strategies, offering a fresh perspective on internet user networks. Prior research has demonstrated the impact of an individual's environment and internet exposure on their online security practices (Bergeron, 2023; Bosnjak and Brumen, 2016; He et al., 2021; Juozapavičius et al., 2022; Van Schaik et al., 2017). These studies suggest a connection between social identity and password selection, where social identity reflects an individual's belief in belonging to a particular group, carrying emotional and value significance (Raskovic, 2021).

The objective of the present study is to consider the influence of users' environments on their password choices, examining variations across different networks. This exploratory investigation aims to identify trends in password formulation, providing a deeper understanding of the social context in which passwords are created. The impact of this research is a step towards a comprehensive comprehension of human behavior in the realm of password creation. Ultimately, it seeks to facilitate the development of

* Correspondence to: 522 Mcbryde Hall, 220 Stanger St, Blacksburg, VA 24061, USA.
E-mail address: tdearden@vt.edu (T.E. Dearden).

https://doi.org/10.1016/j.jeconc.2024.100110
Received 28 June 2024; Received in revised form 19 October 2024; Accepted 30 October 2024

targeted cybersecurity interventions conducive to positive changes in online behavior to reduce the volume and harm of economic cybercrime.

Users' password habits

Users tend to choose weak passwords, which are usually easy to remember but contribute to the fact that they are vulnerable to being guessed (Das et al., 2014; Grobler et al., 2020; Yan et al., 2004). Moreover, users tend to reuse their passwords across many different sites, which also represents a security threat to protecting users' accounts (Almehmadi and Alsolami, 2019). Users tend to use personal information when creating a password and never change it (Almehmadi and Alsolami, 2019).

System-generated passwords could be a solution to weak passwords since the system randomly creates a strong password for the user and ensures the protection of his/her account. However, strong passwords generated randomly are hard to remember. Zviran and Haga (1993) evaluated how well users could remember passwords created three months earlier. They found that only 27.2 % of the participants could remember passwords they created themselves three months earlier. However, the recall rate of system-generated passwords was even worse, with only 12.7 % of the participants remembering system-generated passwords. A solution to manage a list of strong passwords is a password manager, consisting of an application that manages all the users' passwords in one digital wallet. It allows the user to remember only one password and have different and very strong passwords for every website. However, because of the poor usability and limited user experience of password managers, users find it difficult to perform basic actions (Chaudhary et al., 2019). This might explain why so few people use this tool (Almehmadi and Alsolami, 2019). Consequently, research on password creation strategies is still needed even if tools to manage long lists of complicated passwords exist.

Understanding password strength

Passwords are stored in website administrators' databases under their hash value instead of in plain text for security reasons. A hash value is a unique value attributed to a virtual object (e.g., a picture, a video, a text file, or, in our case, a word). The hash value changes as soon as a small change appears on the virtual object. The objective of hashes is to be compared to observe if two virtual elements are identical. Attacks on passwords are only a risk when the password hash file is somehow stolen or otherwise becomes available to an attacker. Once an attacker has a list, the challenge is to uncover the plain text to which the hash value has been attributed. To do so, they must compare the hash values of the list with other hash values for which they know the plain text associated with it.[1] A common practice in password uncovering is to use dictionary attacks, which imply having a list of password possibilities with their hash value, which can be compared to the password in the stolen list. A good dictionary attack involves anticipating the users' password creation strategy and mimicking that in creating a dictionary (Kyaw et al., 2015; Narayanan & Shmatikov, 2005). To ensure that password is not uncovered in this type of attack, the most popular passwords and the passwords containing existing words or names must be avoided. If the password is in the dictionary list of the attacker, the password will be uncovered immediately.

Another common practice to uncover password hash is to use brute force, which consists of trial and error to guess the password. The attacker can program rules for said brute force attacks. These rules can follow known human behavior, like putting numbers at the end of the password or uppercase at the start (Awad et al., 2016). Hive Systems, a cybersecurity company, examined the time it takes for an attacker to brute force a password according to its characteristics. Their research suggests that weak passwords, including those with fewer than six numbers or fewer than four characters, can be brute forced instantly. However, increased complexities, including adding more characters and various types of characteristics drastically increases the time needed to brute force passwords (Hive systems., 2024). Depending on how motivated the attacker is, the password-uncovering software can function on his/her computer for as much time as desired

The rapid increase in computer performance is an element to bear in mind. A password of seven characters was considered as being extremely time-consuming for the attacker in 2018 (Tirado et al., 2018), and now, this time can be measured in minutes (Hive systems., 2024).

Influencing password choice

An individual's environment and exposure to the Internet affect their online security behavior (Bosnjak and Brumen, 2016; Van Schaik et al., 2017). Password creation strategy, defined as the conscious approach adopted by users to create a password they will remember (Zviran & Haga, 1990; Ur, et al., 2015), also seems to be influenced by a person's environment. Grobler and her colleagues (2020) observe similarities between the passwords chosen by the users according to their country. For example, 70 % of passwords in their French dataset were French words, while the top 10 passwords for the English dataset comprised seven English words, two names from the Royal family, and the name of an English football club. The Italian dataset stands out for containing eight names. Bergeron (2023) explores the macrosocial variables influencing password strength. Using a list of the 200 most common passwords in several countries, she discovered that macrosocial variables, like the level of literacy of the country or the level of data breach exposure, significantly predict users' password strength performance. Yang et al. (2012) recognize the difference in passwords across countries and discuss the cultural influence on password choice. They suggest that the rapid growth of Internet users and e-commerce markets in China has led to users creating weaker passwords. They hypothesize that providers may not have paid enough attention to security issues because of the focus on market expansion. This includes typical security strategies, such as requiring minimum password complexity requirements during account creation. The results of the aforementioned studies suggest that there is a difference in cybersecurity habits between countries and demonstrate evidence that there is a structural difference between countries and languages in password habits (Nedvěd, 2021).

Besides the macrosocial influence on users and their password choice, other factors influence password choice and strength. Males seem to have stronger passwords than females, and password complexity decreased with age for both genders equally (Juozapavičius et al., 2022). He et al. (2021) observed that the passwords of Christians, when compared with non-Christians, have different characteristics, and one of them is that they contain many words that stem from the Bible (e.g., Jesus). Other evidence points toward the informal sharing of passwords inside a network of users. Users sometimes include terms related to the semantic theme of a service or how users utilize or feel about a service in their passwords (Wei, et al., 2018). For example, in their analysis of 5 different web services, Wei and her colleagues (2018) report that the password "jobsearch" is found in Linkedin accounts while the password "freemusic" is found in the database of a music-streaming service.

From all those studies, social identity plays a role in password choice. In his reflection on the meaning of identity online, Grayson (2002) states that individual digital identity cannot be understood in isolation from the broader social identity that exists. He raises the issue that technological development created an environment that necessitates establishing a strong digital identity framework.

[1] There are many different types of hash algorithms. Some have a slightly different functioning than what is explain in this paper. The explanation concerns the high majority of hash algorithms and is simplified for the purpose of this article.

The present study

As previously noted, passwords remain the predominant form of authentication despite advancements in alternative methods. Enhancing our understanding of password dynamics empowers the ability to influence users towards creating robust passwords—an essential practice achieved through establishing written policies, using password meters, and similar measures (Wheeler, 2016).

This paper is grounded in social identity theory. Social Identity Theory (Tajfel et al., 1979) posits that individuals derive a sense of self and social identity from their membership in various social groups. According to this theory, people strive to maintain a positive self-concept by identifying with groups they perceive as favorable and distinct. This micro-sociological framework emphasizes the role of self-conception in group membership, group processes, and intergroup relations (Raskovic, 2021). It examines an individual's belief in belonging to a social group, with membership carrying emotional and value significance (Raskovic, 2021). Specifically, cognitive social identity refers to an individual's self-classification within a specific social group (Song and Phang, 2016). Certainly, this extends to online behavior, such as the formulation of online social identity. However, research often focuses on external identity creation through curation and performance on social media websites (e.g., Cover, 2012). We examine whether social identity online extends to information not publicly shared, such as passwords.

The study's objective is to investigate whether the criminal nature of a network influences password characteristics and strength. Prior research has indicated that hacker networks are common (Leukfeldt & Holt, 2019). These networks can be longstanding and highly interpersonal. For example, a thematic analysis of more than 6000 private chat messages found interpersonal dialogue between participants in this network, including discussions of the difficulty of making a profit through illegal means and the unreliability of business partners (Paquet-Clouston and García., 2022). Two hypotheses guide the analysis. First,

H_1: Passwords from different networks will exhibit distinct characteristics.

Social identity theory has been shown to affect online behavior (Lowenthal and Dennen, 2017). In this way we expect online behavior to generally adapt or align within groups. As such we expect to see differences in public (e.g., posts and comments) as well as private behavior (e.g., password choices) between networks.

H_2: Criminal networks of online offenders are more likely to choose stronger passwords compared to non-offender networks.

Although numerous reasons for this expectation can be expounded (see discussion for more possibilities), we posit that social identity theory will lead to criminal networks utilizing better password strength. When establishing that an individual is a member of the social group, distinctions are made between ingroup and outgroup. In hacking culture, ostracizing individuals due to lack of knowledge has been commonly noted in the literature. For example, in classic hacking forums, the term “script kiddie” was a derogatory term used when an individual did not have the technical sophistication to write code and instead utilized other pre-written code (Tejay and Zadig, 2012). As such, stolen accounts or weak passwords may be both a self and group indicator that an individual does not belong in an illegal network due to a lack of ability, skills, or effort to keep oneself safe.

This research contributes to a more comprehensive understanding of passwords within their contextual framework, shedding light on the influence of networks on users' choices. The exploration of discernible trends in password formulation across networks aids in a deeper comprehension of the social context of password creation. Furthermore, it adds to the evolving knowledge regarding lists of commonly used passwords that are prohibited and other related password defenses. Blacklists, in particular, play a pivotal role in thwarting users from selecting vulnerable passwords (Florêncio et al., 2014; Habib et al., 2017). This study's impact lies in advancing our understanding of human behavior in the context of password formulation, paving the way for the development of targeted cybersecurity interventions aimed at fostering positive online behavioral changes in the future.

Materials and methods

Two samples were used to test the research question. One was from a network of offenders, and the other was from a network of non-offenders. The databases were available on the Internet and are, therefore, open-source information. The databases contained usernames and passwords for both networks. They were anonymized by deleting all usernames and only keeping passwords to protect users’ identities.

Sample

The sample associated with offenders comes from a data leak on the OGUsers website. OGUsers was a hacking forum known for the sale of stolen social media accounts hacked through SIM-swapping attacks, credential stuffing attacks, and other means. In other words, offenders visit this site to sell or buy stolen accounts. They necessarily have an illegal intention. As such, the network includes individuals who commit cybercrime. To visit this site, individuals need to create an account with a username and a password. The passwords were used in the analysis of this study.

When the database was found online, the passwords were hashed. Hashing is the process of transforming any given virtual object into a unique value equivalent to a fingerprint. Documents, pictures, sentences, and passwords are all examples of objects that can be hashed. If the same word is hashed twice, it will get the same value. If one of the characters in the word changes, the hash value will be completely different. Hashes are used to store passwords securely in an administrator database. Once the password is hashed, it cannot be dehashed without substantial tools, time, and computational work. To know the string of characters corresponding to this hash value, the same string of characters must be hashed, and only then can it be compared to the hash value of the database. Comparing hashes is done through specialized software. When passwords are robust, the time to discover the matching hash increases significantly. That is why some passwords are never discovered. The database contained 200,551 hashed passwords. Our team succeeded in dehashing 62.61 % of them for a total of 125,560. There are several types of hashing. This database was using 2811 MyBB.

The second sample, associated with a network of non-offenders, was also found on the Internet but was already dehashed: the passwords were in clear text and ready to analyze. This sample comes from the leak of an online game called Grinderscape. It is a fantasy online role-playing game taking place in the medieval realm. It is a massively multiplayer game that has been running since 2008. The sample itself is for players of the game, thus, a non-offender sample. Users needed to create an account with a username and a password to connect to the game. The leak contained a total of 1358,535 accounts, and they were all dehashed. Put together, the samples allow us to analyze a total of 1484,095 passwords.

Description of the samples

The passwords were analyzed according to their characteristics. The presence of letters, numbers, and symbols was recorded. Other characteristics, such as the length of passwords, the presence of dictionary words, and the presence of profanity words, were also analyzed. In addition, we utilized the 3000 most common words that cover 95 % of everyday English conversations, English newspaper and magazine articles, and English used in the workplace (Liu and Nation, 1985; Nation,

1990).[2] The presence of profanity words was observed using an open-source database of more than 400 most popular profanity words in English[3].

Table 1 shows the differences between the two samples based on their characteristics. The samples differed on several descriptive characteristics. The range of the number of characters in the offender sample was higher, with a minimum of one and a maximum of 63 characters. The non-offender sample had a smaller range of 3–30 characters. In addition, the median and mean number of characters was higher in the offender sample.

Analysis

For the purpose of this study, a series of logistic regression analyses were conducted. Logistic regression is designed to estimate the probabilities of a binary event occurring based on a series of covariates. The estimated probabilities are statistically adjusted based on the covariates included in the regression model. The model would have been inappropriate if all the characteristics of passwords presented in Table 1 had been added as several variables overalpped (e.g., contains lower case, contains only lower cases). Having added these overlapping variables would have yielded severe multicollinearity. Two variables from each descriptive category were chosen: (1) Description of the sample (i.e., dictionary and profanity words), (2) Characteristics of weaker passwords (i.e., only lower case and only numbers), (3) Characteristics of stronger passwords (i.e., length of passwords and contains all elements). The analysis has been performed with R statistical Software version 4.2.1.

Results

Table 2 shows the results of the logistic regression. All the variables were significant. Some were associated with offenders (positive relationship), and others with non-offenders (negative relationship). Longer passwords (OR = 0.01, $p < .001$), passwords containing all the different elements like lower case, upper case and symbol (OR = 0.29, $p < .001$), the presence of dictionary words (OR = 0.03, $p < .001$), and the presence of profanity words (OR = 0.11, $p < .001$) were all associated with the offenders' network and having a password that is composed entirely of lower cases (OR = −0.12, $p < .001$) or of numbers (OR = −0.09, $p < .001$) is associated with non-offenders networks.

Discussion

Our data found considerable differences between both networks, consistent with hypothesis one. In fact, every chi-squared test indicated a difference of $p < .001$. Clearly, the password selection between networks was different. For example, both the chi-square tests and logistic regression show the offender network was more likely to contain profanity and use dictionary words. While not necessarily related to password strength without other information, these networks showed a specific preference for unique descriptive differences in password selection.

Second, our data analysis suggests that the offender network was more likely to have stronger passwords, consistent with hypothesis two. The chi-squared tests generally highlight that the offender network was more likely to contain characteristics of stronger passwords, including being more likely to have ten or more characters and contain all the different character elements. The non-offender network was more likely to contain weaker password elements, including lowercase letters, symbols, numbers, and letters. It should be noted that the offender network was more likely to contain only uppercase letters (.2% difference) and have a combination of only letters and numbers. However, given the overall differences in the chi-square tests, these differences do not negate the significantly larger volume of data suggesting that the offender network has stronger passwords.

The logistic regression supports that the offender network also has more complex passwords. Overall, the offender network had longer passwords and was more likely to have all character elements. In contrast, the non-offender network was more likely to have only lowercase letters and only numbers. The longer passwords with more variation in character types are more complicated to crack, indicating that the offender network had overall more secure passwords. The sample of non-offenders chose weaker passwords compared to the offender network. They tended to compose passwords with only lowercase or only numbers. These results allow us to conclude that online offenders utilized more secure passwords.

Several aspects can explain the result of this study. First, the offender sample consisted of an online account that organized illegal activities. Offenders who visit OGUsers website used it to sell or buy online stolen accounts. This means they knew how to steal accounts and/or what to do with them. Due to the complexities of obtaining or using accounts, they likely had higher levels of computer skills. This is generally supported by research; for example, according to the results of a survey done on security experts and non-experts, security experts had better protection strategies than non-experts (Ion et al., 2015). Their research suggests that individuals who are conscious of password risks and understand cybersecurity are more likely to choose better passwords. However, Loutfi and Jøsang (2015) suggest that the passwords of IT professionals may not be as secure as they perceive them to be when they test against actual passwords. However, their research did not compare experts' passwords with non-experts' passwords. Further research is needed to understand whether and why online offenders have stronger passwords.

A second explanation for why the offender database had better passwords is that offenders have something to hide, contrary to people who use online games for fun. Their account might be used to investigate the network, and law enforcement can use some of the information to make arrests. Online games might not produce this type of fear of being hacked, and therefore, the necessity for stronger passwords is not felt. Stobert and Biddle (2015) raise the difference between the strategies used for important accounts versus unimportant accounts. According to their results, individuals tend to use better password security strategies when they are created for important accounts (e.g., bank accounts). Their study was, however, based on the result of a survey. To test the hypothesis that offenders choose better passwords because they protect something more important, future studies should compare the passwords associated with accounts on web services that have important information, such as a bank account, and a web service on which less important information is present, like an online game.

The use of profanity words has been associated with aggressive behaviors (Gitter, 2010; Srull and Wyer, 1979) and lower self-control (Fast and Funder, 2008; Tangney, Baumeister, and Boone, 2004) among the population. Among violent offenders, the use of profanity has been associated with an increased risk of blunt force (Warren et al., 1999). Profanity is indirectly associated with offenders but is not well documented. In the present study, it was found that the network of offenders was more likely to choose passwords with profanity in them compared to the non-offender group. Literature on self-control suggests that offenders are more likely to use profanity overall (Gottfredson and Hirschi, 1984). It appears that this extends beyond the confines of communication and into passwords. While it has long been known that profanity is common in passwords (Veras et al., 2014), to our knowledge, our study is the first to examine the types of networks which are more likely to contain profanity in passwords.

[2] The list of the most common words was taken from the website of a company who offer English language learning programs: https://www.ef.com/ca/english-resources/english-vocabulary/top-3000-words/

[3] The list was downloaded from https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words

Table 1
Password description by samples.

		GrinderScape N = 1358,535	OGUsers N = 125560	Group comparisons
Descriptive characteristics				
Contains at least one lowercase letter		938,338 (69.1 %)	120,041 (95.6 %)	X^2 = 39559.42***; Phi = 0.163
Contains at least one uppercase letter		1063 (0.1 %)	53,233 (42.4 %)	X^2 = 583976.60***; Phi = 0.627
Contains at least one number		57,5029 (42.3 %)	110,729 (88.2 %)	X^2 = 97255.16***; Phi = 0.256
Contains at least one symbol		311,209 (22.9 %)	6365 (5.1 %)	X^2 = 21744.84***; Phi = −0.121
Contains at least one letter		938,858 (69.1 %)	121,750 (97.0 %)	X^2 = 43739.40***; Phi = 0.172
Contains profanity words		29,581 (2.2 %)	7986 (6.4 %)	X^2 = 8150.84***; Phi = 0.074
Contains dictionary words		58,9063 (43.4 %)	79,909 (63.6 %)	X^2 = 19097.19***; Phi = 0.113
Characteristics of weaker passwords				
Contains only uppercase letters		28 (0 %)	252 (0.2 %)	X^2 = 2404.24***; Phi = 0.04
Contains only lowercase letters		479,404 (35.3 %)	12,013 (9.6 %)	X^2 = 34331.59***; Phi = −0.152
Contains only symbol(s)		293,186 (21.6 %)	7 (0.01 %)	X^2 = 33750.10***; Phi = −0.151
Contains only number(s)		125,253 (9.2 %)	3619 (2.9 %)	X^2 = 5821.56***; Phi = −0.063
Contains only letter(s)		479,543 (35.3 %)	14,389 (11.5 %)	X^2 = 29415.42***; Phi = −0.141
Contains only letter(s) and number(s)		1047,326 (77.1 %)	119,195 (94.9 %)	X^2 = 21744.84***; Phi = 0.121
Characteristics of stronger passwords				
Contains more than nine characters		265,928 (19.6 %)	44,798 (35.7 %)	X^2 = 18006.95***; Phi = 0.110
Contains all the elements (at least one letter, one number, and one symbol)		6008 (0.4 %)	5739 (4.6 %)	X^2 = 24947.49***; Phi = 0.130
Password length	Range	1–63	3–30	T(df) = −226.74*** (1484092)
	Median	7	9	
	Mean (sd)	6.6 (3.9)	9.1 (2.0)	

***p < 0.001

Table 2
Comparison between offenders and non-offenders' password characteristics (Logistic Regression) R2 = .09 P < .001.

	b
Length of password	0.012***
Password is only lowercase letters	−0.123***
Password contains all the elements (letter, number, symbol)	0.294***
Password is only numbers	−0.094***
Password contains a dictionary word	0.034***
Password contains a profanity word	0.111***

N = 1484,095
***p < 0.001

Implications of the study

The present paper has two important contributions. The analyses of this paper was framed into social identity theory which has been shown to be vital in virtual platforms. Previous research pointed toward the fact that social identity seems to play a role in password choice (Bergeron, 2023; Bosnjak and Brumen, 2016; He et al., 2021; Juozapavičius et al., 2022; Van Schaik et al., 2017). They illustrate that digital identity cannot be understood in isolation from the broader social identity that exists (Grayson, 2002). Further, this behavior extends beyond public-facing information. Social identity appears to affect private information within networks, in this case, password selection. Although public information may be presented in a curated fashion, it appears that individuals also internalize networks in some fashion, as even non-public information is different between online networks. This paper aimed to observe whether social identification with an online criminal network influences password choice. Because both networks studied in the present research are different, the consideration of web services as networks has proven useful. Moreover, the social identity that is related to each network might explain the difference in password characteristics. The result confirms the assumption of social identity theory that focuses on the role of self-conception in group membership (Raskovic, 2021).

The second contribution of the paper is practical in the field of cybersecurity. This paper contributes to a fuller contextual understanding of password creation to participate in the prevention of ever-evolving attacks on users. Lists of frequently used passwords are often banned from user use and other related password defenses play an important role in preventing users from choosing the most vulnerable passwords (Florêncio et al., 2014; Habib et al., 2017). The impact of the present study moves toward a better understanding of human behavior in the context of password formulation specifically, to enable the future crafting of more targeted cybersecurity interventions that would lead to positive online behavioral change. These strategies can help alleviate the volume of economic or other cybercrimes.

Limits of the study

The original leak of OGUsers database contained 200,551 hashed passwords. Our team succeeded in dehashing 62.61 % of them. This means that 74,991 passwords have not been cracked and analyzed in this study. Those passwords are very good, as the usual tools to crack them were not able to uncover them. With more resources and time,

cracking additional passwords would have been possible. However, we believe that this limit in the manipulation of data does not impact the results of this study, as offenders were considered better at choosing strong passwords. The inability of our team to crack the entire list of passwords is another indicator that this network is better at choosing stronger passwords.

The primary language used on both sites of the sample is English. Therefore, the presence of profanity and dictionary words was analyzed in English. However, it is possible that some users might have chosen a password with dictionary or profanity words in another language than English, which could have impacted the results.

In this study, we categorized individuals as offenders and non-offenders based on the specific network they were associated with. However, it is plausible that some individuals within GrinderScape who were considered non-offenders in our study might engage in online offenses or criminal activities in real life. Unfortunately, our dataset doesn't provide the means to verify the real-life activities of individuals within the GrinderScape network. Despite this limitation, it is essential to acknowledge that individuals might be part of several different networks but the current network has some influence on the formulation of passwords.

There is an interesting observation in the descriptive results of networks concerning the use of symbols. According to the results, offenders are better at having a mix of all elements, including lowercase, uppercase, number, and symbols, which is associated with stronger passwords. However, non-offenders seem to score higher on having symbols in their passwords and even only symbols as passwords. This observation, which seems to be contradictory at first, is explained by the high quantity of non-offenders who choose a *space* as a password. A *space* is considered a symbol in the analysis. Having a single *space* is obviously not a good password, but it increases the presence of symbols in the results of the passwords. The symbol variable was not included in the regression model, so it did not impact the results. Future studies could complexify the model by placing the variable in interaction. This way, the passwords containing only symbols could be entered into the model, and a controlled variable of length could be added to understand the interaction between the different characteristics.

Finally, while this study examines the overall quantitative differences between passwords, it does not provide insight into users' motivations for selecting passwords. Further research is needed to examine why complex or simple passwords are chosen through qualitative strategies, such as interviews. Such studies will expand our understanding and aid in efforts to enhance cybersecurity.

Conclusion

This paper aims to delve into the impact of the environment, particularly the self-conception of group membership, on users. The study compares the passwords of a network of online offenders with those of a network of non-offenders, seeking to ascertain whether the criminal nature of a network influences password characteristics and strength. Hypotheses posited that the password characteristics of both networks would differ, with the criminal network more inclined to opt for stronger passwords than the non-offenders' network. The results substantiate these hypotheses, revealing notable distinctions in password creation strategies between the two networks. Users within the same network exhibit passwords with similar characteristics, forming a pattern distinct from the other network. The findings suggest an informal sharing of password strategies among individuals with shared social interests, emphasizing a network-centric influence on password choices.

This research significantly contributes to a more comprehensive understanding of passwords, shedding light on the influence of networks on users' choices. The exploratory investigation of discernible trends in password formulation across networks provides valuable insights into the social context of password creation. Furthermore, the study adds to the evolving knowledge surrounding banned lists of frequently used passwords and other related password defenses. Notably, the impact of this research extends toward a nuanced understanding of human behavior in the specific context of password formulation, paving the way for more targeted cybersecurity interventions designed to foster positive online behavioral changes.

Disclosures

BLINDED reports a relationship with GoSecure that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Andréanne Bergeron: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Thomas E. Dearden:** Writing – review & editing, Writing – original draft, Validation, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Almehmadi, T., Alsolami, F., 2019. Password security in organizations: User attitudes and behaviors regarding password strength. 16th International Conference on Information Technology-New Generations (ITNG 2019). Springer, Cham, pp. 9–13.

Awad, M., Al-Qudah, Z., Idwan, S., Jallad, A.H., 2016. Password security: Password behavior analysis at a small university (December). 2016 5th Int. Conf. Electron. Devices, Syst. Appl. (ICEDSA) 1–4.

Bergeron, A., 2023. Tell me where you live and I will tell your P@ssw0rd: Understanding the macrosocial variables influencing password's strength. J. Appl. Cybersecur. Internet Gov. 2 (1), 1–19. https://doi.org/10.60097/ACIG/162863

Bosnjak, L., Brumen, B., 2016. What do students do with their assigned default passwords? *39th International convention on information and communication technology*. Electron. Microelectron. 1430–1435.

Chaudhary, S., Schafeitel-Tähtinen, T., Helenius, M., Berki, E., 2019. Usability, security and trust in password managers: A quest for user-centric properties and features. Comput. Sci. Rev. 33, 69–90. https://doi.org/10.1016/j.cosrev.2019.03.002

Cover, R., 2012. Performing and undoing identity online: social networking, identity theories and the incompatibility of online profiles and friendship regimes. Converg. Int. J. Res. Into N. Media Technol. 18 (2), 177–193. https://doi.org/10.1177/1354856511433684

Das, A., Bonneau, J., Caesar, M., Borisov, N., Wang, X., 2014. The tangled web of password reuse. Proc. NDSS.

Fast, L.A. & Funder, D.C. (2008). !@#*!: Reputational and behavioral correlates of swear word usage. Poster presented at the 2008 conference of the Society for Personality and Social Psychology.

Federal Bureau of Investigation (2020). Internet crime report. Available at ⟨https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf⟩.

Florêncio, D., Herley, C., & Van Oorschot, P.C. (2014). An {Administrator's} Guide to Internet Password Research. In *28th large installation system administration conference*, pp. 44-61.

Gitter, S.A. (2010). *Shooting the shit: Profanity, self-control, and aggressive behavior*. [Doctoral dissertation]. Florida State University.

Gottfredson, M.R., Hirschi, T., 1984. The General Theory of Crime. Standford University Press.

Grayson, R.D. (2002). Philosophy of identity: Part of the identity planet series. *TRD, Grayson*. Retrieved at ⟨http://www.timothygrayson.com/PDFs/PhilosophyofID.pdf.pdf⟩.

Grobler, M., Chamikara, M.A.P., Abbott, J., Jay Jeong, J., Nepal, S., Paris, C., 2020. The importance of Social identity on password formulations. Pers. Ubiquitous Comput. 25, 813–827. https://doi.org/10.1007/s00779-020-01477-1

Habib, H., Colnago, J., Melicher, W., Ur, B., Segreti, S., Bauer, L., Cranor, L., 2017. Password creation in the presence of blacklists. Workshop on Usable Security 17 USEC.

He, D., Yu, H., Zhou, B., Zhu, S., Zhang, M., Chan, S., Guizani, M., 2021. How does social behavior affect your password? IEEE Netw. 35 (5), 284–289. https://doi.org/10.1109/MNET.101.2000762

Hive systems (2024). Are your passwords in the green? Retrieved ⟨https://www.hivesystems.com/blog/are-your-passwords-in-the-green⟩.

Ion, I., Reeder, R., Consolvo, S., 2015. … no one can hack my mind": comparing expert and non-expert security practices. Eleventh symposium on usable privacy and security. USENIX Association, pp. 327–346. Retrieved at: ⟨https://www.usenix.org/conference/soups2015/proceedings/presentation/ion⟩.

Juozapavičius, A., Brilingaitė, A., Bukauskas, L., Lugo, R.G., 2022. Age and Gender Impact on Password Hygiene. Appl. Sci. 12, 894. https://doi.org/10.3390/app12020894

Liu, N., Nation, I.S.P., 1985. Factors affecting guessing vocabulary in context. RELC J. 16 (1), 33–42.

Loutfi, I., & Jøsang, A. (2015). Passwords are not always stronger on the other side of the fence. *Proceedings of the usable security workshop*.

Lowenthal, P., Dennen, V., 2017. Social presence, identity, and online learning: research development and needs. Distance Educ. 38 (2), 137–140. https://doi.org/10.1080/01587919.2017.1335172

Nation, I.S.P., 1990. Teaching and learning vocabulary. Newbury House, New York.

Nicholson, J., Vlachokyriakos, V., Coventry, L., Briggs, P., & Olivier, P. (2018). Simple nudges for better password creation. Proceedings of the 32nd International BCS Human Computer Intreaction Confernce (HCI). http://dx.doi.org/10.14236/ewic/HCI2018.46.

Office of the CISO (2023). Threat horizons: August 2023 threat horizons report. Google Cloud. Available at ⟨https://services.google.com/fh/files/blogs/gcat_threathorizons_full_jul2023.pdf⟩.

Paquet-Clouston, M., García, S., 2022. On the motivations and challenges of affiliates involved in cybercrime. Trends Organ. Crime. 1–30. ⟨https://doi-org.ezproxy.lib.vt.edu/10.1007/s12117-022-09474-x⟩.

Raskovic, M.M., 2021. (Social) identity theory in an era of identity politics: theory and practice. AIB Insights 21 (2), 1–7.

Song, P., Phang, C.W., 2016. Promoting continuance through shaping members' social identity in knowledge-based versus support/advocacy virtual communities. IEEE Trans. Eng. Manag. 63 (1), 16–26.

Srull, T.K., Wyer, R.S., 1979. The role of category accessibility in the interpretation of information about persons: some determinants and implications. *J. Personal. Soc. Psychology* 37, 1660–1672.

Stobert, E., Biddle, R., 2015. Expert password management. International conference on passwords. Springer, pp. 3–20.

Tajfel, H., Turner, J.C., Austin, W.G., Worchel, S., 1979. An integrative theory of intergroup conflict. Organ. Identit..: A Read. 56–65.

Tangney, J.P., Baumeister, R.F., Boone, A.L., 2004. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. J. Personal. 72, 271–322.

Tejay, G. and Zadig, S.M. (2012). Investigating the effectiveness of is security countermeasures towards cyber attacker deterrence. 2012 45th Hawaii International Conference on System Sciences. https://doi.org/10.1109/hicss.2012.385.

Tirado, E., Turpin, B., Beltz, C., Roshon, P., Judge, R., Gagneja, K., 2018. A N. Distrib. Brute-Force Password Crack. Tech. Future Netw. Syst. Secur. *117–127*. https://doi.org/10.1007/978-3-319-94421-0_9

Van Schaik, P., Jeske, D., Onibokun, J., Coventry, L., Jansen, J., Kusev, P., 2017. Risk perceptions of cyber-security and precautionary behaviour. Comput. Hum. Behav. 75, 547–559.

Veras, R., Collins, C., & Thorpe, J. (2014). On the semantic patterns of passwords and their security impact. In Symposium on Network and Distributed System Security (NDSS '14). The Internet Society, San Diego, California, USA.

Warren, J., Reboussin, R., Hazelwood, R.R., Gibbs, N.A., Trumbetta, S.L., Cummings, A., 1999. Crime scene analysis and the escalation of violence in serial rape. Forensic Sci. Int. 100 (1-2), 37–56.

Wei, M., Golla, M., Ur, B., 2018. The password doesn't fall far: How service influences password choice. Who Are You 87, 108–112.

Wheeler, D.L. (2016). zxcvbn:{Low-Budget} Password Strength Estimation. In *25th USENIX Security Symposium,* pp. 157-173.

Yan, J., Blackwell, A.F., Anderson, R.J., Grant, A., 2004. Password memorability and security: empirical results. IEEE Secur. Priv. 2 (5), 25–31.