Online Journal of Applied Knowledge Management
A Publication of the International Institute for Applied Knowledge Management

Volume 5, Issue 1, 2017

Strong password? Not with your social network data!

Ruti Gafni, The Academic College of Tel Aviv Yaffo, Israel, rutigafn@mta.ac.il

Tal Pavel, The Academic College of Tel Aviv Yaffo, Israel, tal@middleeasternet.com

Raz Margolin, The Academic College of Tel Aviv Yaffo, Israel, razmargolin@gmail.com

Ben Weiss, The Academic College of Tel Aviv Yaffo, Israel, benweiss114@gmail.com

Abstract

Passwords are the standard means of registration and access to Websites, information systems, online services and various social networks. Databases are increasingly breached and social engineering is employed to obtain usernames and passwords for online fraud, therefore, there is a need to secure existing passwords, and to create ones that will be more crack-resistant. This study addresses the issue of personal data, which users enter on social networks, and incorporate in passwords, as well as how tracking and identifying this data assists hackers in cracking these passwords. The study focuses on Facebook, conducting an online anonymous questionnaire among 195 respondents, and an experiment among a voluntary response sample of 72 participants, in which passwords were tried to been deciphered by a custom dictionary attack. The findings confirm a link between the use of accessible online personal data and success rates of password deciphering. The findings underscore the grave threat to users' information security - not only as a result of their voluntary exposure of personal data on social networks, but also due to the integration of this data into their passwords. The study argues the need to emphasize users' awareness to their password strength, with this vulnerability in mind.

Keywords: Passwords, password guessability, social networks, privacy, personal information in passwords, dictionary attack, cybersecurity.

Introduction

Despite advanced options of biometric identification or graphic passwords, text-based passwords remain the most common mechanism of identification for accessing information systems, online services, and social networks. However, the vulnerability of these passwords lies mostly in the relative ease of cracking them. Users are required to manage an increasing number of passwords. Some sites offer the ability to generate passwords that are stronger and harder to crack. However, these are often more difficult to remember, especially when users have to keep track of multiple passwords. For this reason, users tend to select passwords that are relatively easy to manage, and frequently embed personal data into their passwords to make them more easily memorable (Al-Wehaibi, Storer, & Glisson, 2011; Shen, Yu, Xu, Yang, & Guan, 2016). Social networks are fast becoming all-encompassing means of communication, on which the users share increasingly more information about themselves and their environment, including names, occupation and year of birth, as well as data of various related circles such as family, friends, colleagues, and classmates. This information is readily available, particularly if users have not used privacy settings to restrict access to their profile or misconfigured such settings. Malicious users have at their fingertips a wide range of tools for deciphering passwords, which vary depending on their

purpose, specifications, and application. This study examines the connection between the personal data available on social networks and the personal details embedded into passwords, facilitating password cracking, which in turn is used to log into the users' accounts. Furthermore, this study demonstrates the greater risk posed by constructing passwords that incorporate personal details that are accessible on social networks.

Literature Review

Fixed Patterns in Text-Based Password Construction

Passwords are the key mechanism enabling a system to verify a user's identity before granting access to its services. Text-based passwords are a chain of characters made of different components, which are mostly letters, numbers, and other characters. The scientific community invests tremendous effort in seeking alternatives means such as biometric recognition, graphic passwords and more. However, text-based passwords remain the most common identity verification mechanism (Mazurek et al., 2013). Sahin, Lychev, and Wagner (2015) addressed the fact of password-based verification being so common a method that switching to a different mechanism would be a long and complicated process. Passwords are created by users, and most tend to follow the same guidelines and mental authority while performing tasks. Creating a password is inherently associated with one of its deepest flaws: they are easy to guess. An analysis of six million real passwords leaked to the web carried out by Shen et al. (2016) found that easy to guess passwords such as "123456789" were extremely popular, exceeding 10% of the total of passwords analyzed. Jakobsson and Dhiman (2013) examined how users create passwords and found that users employed a fixed sit of rules and a small number of details:

- Concatenation - the most common rule, in which the user welds together words and characters to form a password. For example, "1passbay" is comprised of pass, bay and 1.
- Replacement - in which numbers are used to replace letters in a word. Substituting the letter "e" with the number 3, for example, turns the word "seventy" into "s3v3nty".
- Insertion - in which a user incorporates an additional component to create the password. For example, inserting"77" into the name "Christina" produces "Christi77na".

Voyiatzis, Fidas, Serpanos, and Avouris (2011) analyzed encrypted and unencrypted passwords to accounts on Greek websites that do not enforce a specific policy, meaning the passwords reflect the personal preferences of the users. They found the average password was seven characters long, and that the majority (99.6%) contained only Latin letters and numerals. De Carné, de Carnavalet, and Mannan (2014) studied password construction, focusing on websites with mechanisms assessing password strength. They pointed to inconsistencies between the mechanisms on various assessment sites, which often assessed weak passwords as strong or even very strong, giving the user a false feedback about the password's strength. A stricter password selection policy can lead to stronger passwords that are more guess-proof. Users choose relatively easy passwords for accounts of lesser importance. However, users do not modify their password choices in a manner corresponding to account-importance as might have been expected, indicating a fundamental problem: users are either unable or unwilling to manage multiple guess-resistant passwords (Bonneau, 2012). Furthermore, more than 65% use the same

password to gain access to most of their accounts (Al-Wehaibi et al., 2011), or have difficulty remembering their password. Haque, Wright, and Scielzo (2013) studied how passwords ranked high-strength (such as to bank accounts) could become a vulnerability due to reuse for other accounts. They proved that knowing the password to a low-risk account increases the chance of breaking into higher-risk accounts, based on the similarity between the various passwords used.

Password Management

Most often, password management guidelines require a periodic password replacement, use of a different designated password for every account, and refraining from sharing the password with others or entering it on or even around a computer. In practice, most users do not comply with these (Dunphy et al., 2015). The difficulty in remember and managing the increasing number of passwords, leads users to take various strategies, which undermine the security of their own information. Komanduri et al. (2011) found that the stronger the passwords, the more difficult it is to remember them, prompting users to enter them in writing. Social login is another means of managing identical passwords to multiple sites by using the social network password to register to multiple sites (Gafni & Nissim, 2014), establishing a single sign-on. In this way, users can remember fewer passwords while gaining access to a wider range of services. They note the "password fatigue" phenomenon, which causes users to reuse the same password across multiple platforms, increasing their susceptibility. Single sign-on offers time and effort saving advantages, however, this method poses a considerable challenge to security information, and poses an additional challenge to user privacy.

Incorporating Personal Information into Passwords

Social networks allow users to present themselves through a profile, connect with others, and publish social content (Franchi, Poggi, & Tomaiuolo, 2014). Due to the vast amount of information users publish on social networks, these platforms both aggregate and display a wealth of valuable information about users and their activity including location, social circles, political positions, and even sexual preferences that can be exploited by various hostile and malicious players (Beato, De Cristofaro, & Rasmussen, 2014). Seemingly, trivial information such as name, location, and age are enough for a hacker to locate the personal profile of more than half the American population (Franchi et al., 2014). They found, that although users are aware of the fact that their profiles and information are public, most tighten their security settings only after a problem occurs, often ignoring the potential future damage that might be caused by their published information. Madejski, Johnson, and Bellovin (2012) painted a worrisome picture, as users experienced at least once a discrepancy between their intentions of sharing, and their privacy settings in practice. However, they choose to take no action, suggesting that the privacy settings mechanism of social networks puts their users at risk and exposes their information to malicious users, but that this threat has not prompted a change in user behavior. Halevi, Lewis, and Memon, (2013) found that users sharing more personal information on Facebook have more lax privacy settings, putting them at greater risk of being hacked. Studies show that users tend to create passwords that incorporate their personal details (Al-Wehaibi et al., 2011; Shen et al., 2016), using their first names, birthday, or favorite football club in a password. The researchers assumed this tendency is driven mostly by fear of forgetting the password, but noted that such passwords containing valuable personal details and are more easily

guessed by hackers. Dürmuth, Chaabane, Perito, and Castelluccia (2013), addressed that passwords based on personal attributes are weaker, and should be avoided.

Difficulty in Obtaining Passwords for Comprehensive Research

Studies aim to analyze and understand the different characteristics of the passwords. Nevertheless, they are hampered by difficulty in collecting data on real passwords, using passwords created in experiment conditions rather than real life. Difficulty in obtaining the plaintext of passwords (unencrypted passwords), data based on self-reports that are not necessarily accurate, as well as real passwords leaked from low-level sites, which pre-suggests a weakness. Consequently, the key question of whether the results of these studies are valid remains unanswered (Komanduri et al., 2011; Mazurek et al., 2013). Thus, even the best guidelines on password composition policy are based on theory rather than empirical data.

Password Guessability

There are several ways to define password guessability, but all essentially refer to a term describing the strength of passwords based on the time needed to hack it by using a cracking algorithm, or alternatively, by the predicted number of guesses it will take to identify the password (Jakobsson, & Dhiman, 2013; Ur et al., 2012). Mazurek et al. (2013) showed that passwords composed by students of computer sciences, were 1.8 times stronger than those created by students of business administration students. Bonneau (2012) found that men composed passwords somewhat stronger than women, and that passwords composed by older adults and lower-frequency users were found to be slightly stronger. Shay et al. (2014) examined password security vs. effectiveness and found that policies requiring longer passwords but fewer particular requirements can be more practical as well as more secure than common strong-password policies. They found that while basic policies were easier to comply with and use than long-password policies they were more vulnerable to a relatively low number of crack attempts.

Password Reconstruction Tools and Access

A wide range of tools for reconstructing and decrypting passwords exist, differing in their functionality, purpose, specifications, and applications (Al-Wehaibi et al., 2011). While an online attack is limited in the number of attempts before a user is blocked, offline attacks usually steal a database of passwords encrypted by hash functions, and try to guess the passwords offline, which allows an indefinite number of attempts (Ur et al., 2015). In a brute-force attack, hackers try to guess passwords systematically until succeeding in cracking them. Early knowledge of the victim is not a prerequisite in this approach. Brute-force attacks are conceptually simple and effective only when attempting to crack very short passwords or those created by password generator systems (Ur et al., 2015). Brute-force attackers do not assume all guesses are equal and, therefore, do not assume all possible passwords are equal either (Sahin et al., 2015). Thus, some guesses are better than others since human guessing of passwords is not random.

A dictionary attack is one that targets systems with entry permit mechanisms by using more advanced and more sophisticated brute-force techniques, compiling a comprehensive list of words to form a dictionary, which is used to decipher usernames and passwords. During the course of the attack, every word in the dictionary is tried, or word combinations combined with

characters, in order to gain unauthorized access to a system. Systems, which employ advanced techniques to identify and verify their users, are immune to dictionary and related attacks (Soroka & Iracleous, 2011). The tendency to choose short words for passwords makes the passwords more susceptible to this kind of attack (Haque et al., 2013). Distributed dictionary attacks are carried out in the same manner as dictionary attacks, with the same predictable results. The difference is that the amount of work is divided between several computers, allowing for a much faster attack. Another advantage of distribution is that it overcomes defense mechanisms systems, which have in place to deflect this sort of attack. Dictionary attacks usually occur online; however, they can also be carried out offline, depending on the target and objective (Soroka & Iracleous, 2011). The effectiveness of the dictionary check depends on the dictionary chosen (Kelley et al., 2012). For example, a large blacklist compiled through password guessing is far more effective than a regular dictionary in preventing users from selecting highly guessable passwords.

Social engineering has always been a security threat, employing psychological manipulation to drive victims to take actions that are harmful to themselves. This is very much like traditional fraud, only in this case attackers exploit the virtual space to gain access to confidential information (Franchi et al., 2014). The proliferation of social networks and the tremendous amount of users' personal information found there makes these an attractive target for malicious users interested in this sort of attack. Different studies showed that the black market of stolen credit card information is giving way to trading in stolen social network accounts, sold to the highest bidder (Polakis et al., 2012). Several methods of attacks come under the umbrella of social engineering, including phishing, customized dictionary attacks, and others. Al-Wehaibi et al. (2011) examined password reconstruction by carrying out dictionary attacks with a customized dictionary tailored to particular users. The objective of their study was to examine whether it was possible to shorten the time needed to crack a password by making educated use of the intended victim's online information. They developed a web crawler and used to scan profiles of students at Glasgow University. The crawler was designed to extract text from the scanned profiles before significant words registered in the printout documents. These documents later served the researchers to compile a personalized diction specific to each participant in the study that was used to reconstruct their passwords in a dictionary attack using a Password Recovery Toolkit (PRTK). The participants were asked to enter a password and use it to lock a file. In order to study the impact of online information on the password reconstruction, PRTK was used to carry out three dictionary attacks using (1) the standard data PRTK dictionary; (2) the personalized dictionaries compiled for each participant; (3) combined use of both PRTK and customized dictionaries. The results of this study indicated the existence of a weak connection between victims' online information and the time it takes to recover their passwords.

Research Questions

The objective of this study was to examine the use of personal information embedded in password creation and how this information, which part of it is exposed by the users in the social networks, cause the passwords to be weak and easy to crack (decipher).

RQ1 – To what extent do users incorporate personal information in their passwords?

H1 - Because of the need to remember the passwords, users tend to embed in their password personal information, which they know and it is easy to recall (Al-Wehaibi et al., 2011; Franchi et al., 2014; Madejski, Johnson, & Bellovin, 2012).

RQ2 – How does the use of personal details available on Facebook, in creation of passwords, enhance the success rate for password recovery?

H2 - People expose a lot of personal information in their social network account, especially on Facebook (Halevi, Lewis, & Memon, 2013). The information can be misused, in order to decipher the passwords using custom dictionaries attacks (Haque et al., 2013). Thus, the use of custom dictionaries based on personal information collected from social networks will enhance the success rate of the password deciphering.

RQ3 – What are the main characteristics of the users with high risk to social hacking because of their weak passwords? Are there differences on the use of strong passwords between users according to gender, age, frequency of social networks use and technology background?

H3 - There are differences in demographic and behavioral characteristics between users who create strong or weak passwords. Mazurek et al. (2013) found that computer science students created passwords stronger than business administration students, and male students created slightly stronger passwords than female.

Methodology

In order to collect the relevant data, two research methods were combined:

1. An anonymous questionnaire was created, and distributed to users having an account on Facebook. The main goal of this questionnaire was to find out to what extent the users define their passwords based on personal information that is available on their Facebook account. Due to the need of being part of Facebook and using it, the questionnaire was distributed, using a snowball sampling (Baltar & Brunet, 2012; Noy, 2008), to social network users, who define text-based passwords for the use in different websites. The questionnaire, elaborated using Google Docs, enabled keeping the anonymity of the participants. The first questions were addressed to segment the sample population by: Gender, Age, Employment, Education and the frequency use of Facebook. This segmentation was performed in order to categorize groups and try to check whether there are groups that are at greater risk of a "dictionary attack". The following questions were related to the availability of personal information on Facebook and the way respondents compose their passwords.

   Which information appears about you on Facebook?

   Whom do you usually give permission to see your information in Facebook?

   Do you create your password using a password generator or manually?

   Do you tend to integrate information related to your personal life in your password?

   What kind of information do you use on your password?

The aim of the next set of questions was to investigate the reasons of combining personal information into the passwords, like: (1) easier to remember them, (2) the password that is harder to break, (3) these details make positive emotions, (4) it requires less effort.

The last questions tried to understand if the participant assumes that the passwords created are strong enough.

2. With the understanding that the participants in the questionnaire may not be willing to reveal their passwords, an experiment was conducted, in which the use of passwords that people actually have created were checked to find weather they use personal information available in their account on the social networks. The applications for participation in the experiment were distributed almost exclusively amongst friends and volunteers on Facebook, who agreed to participate and to allow the access to their personal details published on Facebook. This data helped to build the specific data dictionary. The participants did not know the purpose of the study. Each participant was asked to create a password. In order to simulate password policies on the network, it was determined that the password must contain eight characters, at least an English letter and one number. In addition, participants were required to remember the password they created for a period of two days, without writing it down on a paper, with the understanding that when it is generated, the password should be easily recalled by its owners. After the password was created, the participants were asked to give their Facebook ID and to answer a few questions, including the first questions of the questionnaire defined in prior section. Thus, the demographic data and information whether the password combines personal data were collected. For each respondent, their Facebook was investigated and assembled a data dictionary containing the relevant details, as following:

   (1) **Personal details**: Full name, high-school name, current city, hometown, home address, the place where served in the army, military role, dog name, nicknames, college/ university name, preferred animal, phone number.

   (2) **Names** of family members and close friends: mother, father, sister, brother, uncle, cousin, husband, daughter, son and more, close friends, boyfriend/ girlfriend name.

   (3) **Details about work**: firm, address.

   (4) **Dates**: birth date, date of marriage, birthday of relatives.

   (5) **Hobbies**: favorite singer, favorite song, name of a favorite TV series, favorite hobby (dancing, football, basketball…), favorite book, favorite hang-out place.

   (6) **Political data**: political preferences, favorite politician.

   The aim of this experiment was to try cracking the passwords provided. This was performed using a password recovery software, PRTK version-7.6.0 (password recovery toolkit). This is a standard software in the industry, which allows attacks using both a standard dictionary and a custom dictionary (Al-Wehaibi et al., 2011). Using the PRTK software, two types of runs were performed on each password:

   (1) A **Dictionary Attack** - carried out using a Standard English dictionary, which is the default dictionary in the PRTK software.

(2) A **Custom Dictionary Attack** - integrating both the Standard English dictionary, and a custom dictionary, which was created by investigating the personal information on Facebook, as mentioned.

The experiment was conducted on a two processors' server (Intel® Xeon® CPU E7- 4870 @ 2.40GHz 2.39 RDP- GHz), during one month. For each password, the maximum time allocated in order to decipher was determined to 12 hours. The allocation of time is substantial to the deciphering algorithm. The more time allocated the more successful rate. Deciphering the password in such way can take some days long. The aim was not to crack the password, but to see how easy is to do so. If the algorithm did not succeed in 12 hours, it was defined as a failure, and as a strong password. By measuring the success rates of two types of attacks on these passwords, it was possible to examine whether personal information on social networks that are integrated in passwords, affect the ability to decipher passwords and weakens them. All data were analyzed using IBM® SPSS® Statistics, version 20.

Results

The questionnaire was answered by 212 respondents, of which 17 participants indicated that they had no account on Facebook were omitted. The demographic descriptive statistics of the relevant 195 respondents of the questionnaire is shown in Table 1. The distribution between men and women was almost equal. The vast majority of respondents belonged to the age group 22-30 and had a Bachelor's degree. About two-thirds of the respondents indicated a daily use of Facebook and twenty percent of them indicated a higher usage of more than one hour a day.

The respondents were asked if they use a password generator in order to create their passwords, or they use to create them manually. Only 3% use a password generator. Those who compose the password manually were asked what kind of information they expose on Facebook, and what kind of information they use in their passwords. Each participant who did not check the "No personal data at all', could check some of the other options. This is the reason that the totals exceed 100%. Table 2 display the results. Moreover, most of the respondents (65.1%) think that using personal data in the password weakens the password allowing easy deciphering, although 71.3% actually use personal data to create their passwords. It can be seen that the fear of password cracking does not prevent people from using their personal information. The main reasons for using personal data in passwords were "Easy to remember" (92%) and "Less effort to create" (69%). Moreover, most of the respondents (74%) agree that including personal data in the password does not make it harder to crack, so they understand this is a drawback of their passwords. It seems that people are trying to prevent cognitive efforts, but are not aware that they actually expose, in social networks, the data that they use in the passwords. In order to reveal if people use the same data, the experiment was conducted.

Table 1. Demographics of the survey participants

Gender	Male		Female		
	(96) – 49.23%		(99) – 50.77%		
Age	18-21	22-30	31-40	41-55	56 and up
	(9) – 4.62%	(99) – 50.77%	(27) – 13.84%	(44) – 22.56%	(16) – 8.21%
Employment	Technological professions		Non Technological professions		
	103		92		
Education	High school education	Professional	Bachelor's degree	Master's degree	Ph.D.
	(24) – 12.32%	(20) – 10.25%	(123) – 63.07%	(25) – 12.83%	(3) – 1.53%
Frequency of using Facebook	Once a month	Once a week	Once a day	Once an hour	More than once in hour
	(10) – 5.12%	(11) – 5.67%	(38) – 19.49%	(98) – 50.25%	(38) – 19.49%

Table 2. Information exposed on Facebook vs. information used in passwords

Information type	**Exposed on Facebook**	**Part of password**
No personal data at all	12.8%	28.7%
Personal data	63.3%	49.7%
Family data	30.8%	27.2%
Work data	37.4%	0%
Hobbies data	43.1%	6.2%

There were 84 people that agreed to participate in the experiment, from which two participants were omitted because they entered invalid passwords (one containing a margin and the other was an e-mail address). Ten other participants were excluded because they did not approve the friendship request on Facebook, leaving a total of 72 relevant participants.

Table 3 summarizes the demographic descriptive statistics of the participants of the experiment. The distribution between men and women was quite similar, and the vast majority of the examined participants belonged to the age group of 22-30, apparently due to the form of distribution used.

Table 3. Demographics of the experiment participants

Gender	Male			Female		
	(29) - 41%			(43) - 59%		
Age	18-21	22-30	31-40	41-55	56 and up	
	(2) – 2%	(48) – 67%	(6) – 8%	(12) – 17%	(4) – 6%	
Employment	Technological professions		Non Technological professions			
	(24) – 33.3%		(48) – 66.6%			
Education	High school education	Professional	Under graduate students	Bachelor's degree	Master's degree	Ph.D
	(17) – 25%	(4) – 5%	(11) – 15%	(34) – 47%	(6) – 8%	(0) - 0%
Frequency of use Facebook	Once a month	Once a week	Once a day	Once an hour	More than once an hour	
	(3) – 4%	(6) – 8%	(24) – 33.5%	(24) – 33.5%	(15) – 21%	

Roughly half of the participants are Bachelor graduates and quite one-third have only a high school education. A third of the participants in the experiment classified technology as a part of their profession. Out of 72 relevant respondents, 61 actually combined personal information within the passwords they created. These passwords and the personal data accumulated from the participants exposed in their Facebook's accounts, were used in the experiment. For each password, two attempts for deciphering were conducted: in the first one, the password was attempted to be cracked using the standard dictionary, and the second time using the customized dictionary, which included the personal data found in Facebook. It is important to emphasize that this was not hidden data and the Facebook account was not attacked. The data used was the information the person exposes voluntary in the Facebook account.

The results of the attacks, for those participants who admitted that they use personal information in their passwords are presented in Table 4. As seen, most of the successful deciphering was performed using the customized dictionary (55.7% + 14.8%). This shows that people use personal data in order to create their passwords, and at the same time, they expose this data in their Facebook's account.

Table 4. Results of standard and customized dictionaries attacks

	Total		**Female**		**Non- Technical Profession**	
	Number	Percent	Number	Percent	Number	Percent
Non deciphered	13	21.3%	7	53.8%	5	38.5%
Standard dictionary success	5	8.2%	4	80%	4	80%
Customize dictionary success	34	55.7%	22	64.7%	25	73.5%
Both Standard and Customized success	9	14.8%	6	66.7%	7	77.8%
Total deciphered	48	78.7%				

Moreover, as shown in Table 4, the passwords that female participants created were easier to decipher, both by the standard dictionary and by the customized one. Additionally, participants with a non-technical profession created simpler passwords, which were very easily deciphered.

Discussion

This study examined the use of personal information in creating passwords. Especially the lack of awareness of members of social networks, who expose their personal information, the same one that they embed into their passwords. Most of the respondents of the questionnaire and the participants of the experiment create their passwords using personal data that they can easily remember (92%) as well as easier to create (74%). Only 3% use a password generator. According to the questionnaire, 28.7% do not use personal data for creating a password. However, when the participants of the experiment had to actually create a password and remember it, only 15.3% did not combine personal data in the password. The use of personal information, therefore, is very frequent where most of the people use their own information (49.7%) and others combine family information (27.2%). Information about work or hobbies are rarely used. According to the results, H1 has been accepted, meaning that because of the need to remember the passwords, users tend to embed in their password personal information, which they know and it is easy to recall. These results fit previous research, where it was found that people prefer passwords they can easily remember (Al-Wehaibi et al., 2011; Komanduri, et al., 2011), and use personal information (Al-Wehaibi et al., 2011; Halevi, Lewis, & Memon, 2013; Shen et al., 2016). It is understandable that people try to create passwords, which are easy to remember. According to Florencio and Herley (2007), the average Internet user had about 25 different accounts requiring passwords. Therefore, they needed to enter about eight passwords a day. Each account requires password changing at different times, twice or thrice a year. These requirements make it difficult for individuals to manage and remember all their passwords, leading to "password fatigue" (Gafni & Nissim, 2014). Thus, motivating the user to create rememberable passwords, most of them with personal information.

People understand that embedding personal information into their passwords is dangerous (74%). However, people are not aware of the fact that they expose the same personal information in their social networks accounts, which is open, often, to everybody, without any barriers. As it was demonstrated in the experiment, the use of personal information collected from the Facebook accounts of the participants, allowed to increase highly the possibility to decipher the passwords, accepting H2, and saying that the use of custom dictionaries based on personal information collected from social networks enhance the success rate of the password deciphering. These results fit with previous research (Franchi et al., 2014; Halevi et al., 2013; Haque et al., 2013; Madejski et al., 2012).

The main characteristics of the users with high risk to social hacking, because of their weak passwords, was found mainly in correlation of the profession. People working in a technological environment were found more aware to the risks, therefore, their passwords were more difficult to decipher. These findings accept H3, which claims that there are differences in demographic and behavioral characteristics between users who create strong or weak passwords. These findings can be correlated to those of Mazurek et al. (2013) where the students of computer science created stronger passwords than the business management ones. Moreover, females were found to have weaker passwords than males, fitting both Bonneau's (2012) and Mazurek et al. (2013) findings. Women tend to expose more information that is personal in their Facebook's account, thus, making the passwords weaker.

Limitations

Studies dealing with passwords are limited because of the difficulties in gathering real passwords data and mostly are based on experimental passwords (Mazurek et al., 2013). In this research, the passwords were collected during an experiment, in which the participants had to define a password, meet the standards criteria of most websites, and which they could remember for two days without writing it down. Nevertheless, it is possible, that the provided passwords might be different from those the same participants would define for a real website. Moreover, the sample was collected in a Snowball sampling. The survey and experimental results might be influenced by the composition of the respondents' sample, which has been performed in one country, with similar ages and education status. Perhaps, people in different countries, different cultures, ages, or education status would have created different passwords: maybe stronger or weaker.

Conclusion

In the modern era, the computerized world, including websites, applications, emails, and social networks became a significant part of our life. In order to protect these assets, people need to be authenticated and authorized to enter, usually by using user names and passwords. The need to manage simultaneously a number of passwords with different rules for creating them, different times for changing and updating them, alongside with the need to remember all these passwords, bring about the users to define passwords they can easily remember. People find it simple to remember personal data, therefore, the usage personal data in passwords is very vast. However, this personal data, which is embedded in the users' passwords, is often overt in the social network accounts, like Facebook, Tweeter, LinkedIn, and others. Due to the vast amount of information users publish on social networks, these platforms both aggregate and display a

wealth of valuable information about users and their activity, that can be exploited by various hostile and malicious players. The more information users make public, the more vulnerable they are to this information being used against them in various ways, including increasingly sophisticated and grave attacks. Seemingly, trivial information such as name, location, and age are enough for a hacker to decipher the user's passwords, when it contains personal data.

This study proved that including personal data in password creation, when the user possess a social media account, in which trivial personal data is exposed, allows cracking the passwords in an easier and simpler way than those without personal data. The awareness among the users must be emphasized, in order to educate the population to use better secured passwords, that do not include personal information, in order to diminish their risks. If users will not include personal data in their passwords, the remembrance of the passwords will be difficult, so mechanisms to secure manage of passwords must be defined. Moreover, it will be less risky for users to avoid the publication of personal data in their social networks accounts, although, at these days, it is a little too late.

References

Al-Wehaibi, K., Storer, T., & Glisson, W. B. (2011). Augmenting password recovery with online profiling. *Digital Investigation, 8*, S25-S33.

Baltar, F., & Brunet, I.(2012). Social research 2.0: virtual snowball sampling method using Facebook, *Internet Research*, *22*(1), 57-74.

Beato, F., De Cristofaro, E., & Rasmussen, K. B. (2014). Undetectable communication: The online social networks case. In *Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on*, pp. 19-26. IEEE.

Bonneau, J. (2012). The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *2012 IEEE Symposium on Security and Privacy,* pp. 538-552. IEEE.

De Carné de Carnavalet, X., & Mannan, M. (2014). From very weak to very strong: Analyzing password-strength meters. In *Network and Distributed System Security Symposium. Internet Society.*

Dunphy, P., Vlachokyriakos, V., Thieme, A., Nicholson, J., McCarthy, J., & Olivier, P. (2015). Social media as a resource for understanding security experiences: A qualitative analysis of# password tweets. In *Eleventh Symposium on Usable Privacy and Security (SOUPS 2015)*, pp. 141-150.

Dürmuth, M., Chaabane, A., Perito, D., & Castelluccia, C. (2013). When privacy meets security: Leveraging personal information for password cracking. CoRR abs/1304.6584.

Florencio, D., & Herley, C. (2007). A large-scale study of web password habits. *In Proceedings of the 16th international conference on World Wide Web*, pp. 657-666.

Franchi, E., Poggi, A., & Tomaiuolo, M. (2014). Information attacks on online social networks. *Journal of Information Technology Research, 7*(3), 54-71.

Gafni, R., & Nisim, D. (2014). To social login or not login? - Exploring factors affecting the decision. *Issues in Informing Science and Information Technology, 11*, 57-72.

Halevi, T., Lewis, J., & Memon, N. (2013). Phishing, personality traits and Facebook. *arXiv preprint* arXiv:1301.7643.

Haque, S. M., Wright, M., & Scielzo, S. (2013). A study of user password strategy for multiple accounts. In *Proceedings of the third ACM conference on Data and application security and privacy*, pp. 173-176.

Jakobsson, M., & Dhiman, M. (2013). The benefits of understanding passwords. In *Mobile Authenticatio*n, pp. 5-24. New York, NY: Springer.

Kelley, P. G., Komanduri, S., Mazurek, M. L., Shay, R., Vidas, T., Bauer, L., Christin, N.,, Cranor, L. F., & Lopez, J. (2012). Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *2012 IEEE Symposium on Security and Privacy*, pp. 523-537.

Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., ... & Egelman, S. (2011). Of passwords and people: Measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2595-2604.

Madejski, M., Johnson, M., & Bellovin, S. M. (2012). A study of privacy settings errors in an online social network. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pp. 340-345.

Mazurek, M. L., Komanduri, S., Vidas, T., Bauer, L., Christin, N., Cranor, L. F., ... & Ur, B. (2013). Measuring password guessability for an entire university. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pp. 173-186.

Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International Journal of Social Research Methodology*, *11*(4), 327-344.

Polakis, I., Lancini, M., Kontaxis, G., Maggi, F., Ioannidis, S., Keromytis, A. D., & Zanero, S. (2012). All your face are belong to us: Breaking Facebook's social authentication. In *Proceedings of the 28th Annual Computer Security Applications Conference*, pp. 399-408.

Sahin, C. S., Lychev, R., & Wagner, N. (2015). General framework for evaluating password complexity and strength. *arXiv preprint* arXiv:1512.05814.

Shay, R., Komanduri, S., Durity, A. L., Huh, P. S., Mazurek, M. L., Segreti, S. M., ... & Cranor, L. F. (2014). Can long passwords be secure and usable? In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pp. 2927-2936.

Shen, C., Yu, T., Xu, H., Yang, G., & Guan, X. (2016). User practice in password security: An empirical study of real-life passwords in the wild. *Computers & Security*, *6*, 130-141.

Soroka, E. V., & Iracleous, D. P. (2011). Social networks as a platform for distributed dictionary attack. In *Proceedings of the 5th WSEAS International Conference on Communications and Information Technology, 11*, 101-106.

Ur, B., Kelley, P. G., Komanduri, S., Lee, J., Maass, M., Mazurek, M. L., ... & Christin, N. (2012). How does your password measure up? The effect of strength meters on password creation. In *21st USENIX Security Symposium (USENIX Security 12)*, pp. 65-80.

Ur, B., Segreti, S. M., Bauer, L., Christin, N., Cranor, L. F., Komanduri, S., ... & Shay, R. (2015). Measuring real-world accuracies and biases in modeling password guessability. In *24th USENIX Security Symposium (USENIX Security 15)*, pp. 463-481.

Voyiatzis, A. G., Fidas, C. A., Serpanos, D. N., & Avouris, N. M. (2011). An empirical study on the web password strength in Greece. In *Informatics (PCI), 2011 15th Panhellenic Conference on*, pp. 212-216. IEEE.

Authors' Biographies

Ruti Gafni is the Head of the Information Systems B.Sc. program at The Academic College of Tel Aviv Yaffo. She holds a Ph.D. from Bar-Ilan University, Israel (in the Business Administration School), focusing on Information Systems, an M.Sc. from Tel Aviv University and a BA (Cum Laude) in Economics and Computer Science from Bar-Ilan University. She has more than 30 years of practical experience as Project Manager and Analyst of information systems. She also teaches in the Management and Economics MBA program at the Open University of Israel.

Tal Pavel is in charge of the Cyber studies in the Information Systems Program at The Academic College of Tel Aviv Yaffo**.** He holds a Ph.D. from Bar-Ilan University, Israel (in the Department of Middle Eastern Studies). He is a member and researcher in the Blavatnik Interdisciplinary Cyber Research Center at Tel Aviv University and in the Institute for Counter-Terrorism (ICT) at the IDC Herzliya Center. He is the editor of Middleeastermet.com, which includes information, research and opinions about internet and cyber threats and security.

Raz Margolin is a third year student in the Information Systems Program at The Academic College of Tel Aviv Yaffo. She received the Dean's Excellence Award. She worked as a software quality assurance during her studies.

Ben Weiss is a third year student in the Information Systems Program at The Academic College of Tel Aviv Yaffo.