Comparative Evaluation of Security and Convenience Trade-offs in Password Generation Aiding Systems

Michael Stainbrook[1] and Nicholas Caporusso[1]

[1] Fort Hays State University, 600 Park Street,
67601 Hays, United States
mjstainbrook@mail.fhsu.edu, n_caporusso@fhsu.edu

Abstract. A strong password is considered the most important feature for the security of any account credentials. In the last decades, several organizations focused on improving its strength and produced awareness initiatives and security guidelines on how to create and maintain secure passwords. However, studies found that users perceive security and convenience as a trade-off, and they often compromise password strength in favor of a key phrase that is easier to remember and type. Therefore, nowadays websites and applications implement password generation aiding systems (PGAS) that help, and even force, users to create more secure passwords. Nowadays, several types of PGAS are available, each implementing a different strategy for stimulating users in crating stronger and more secure passwords. In this paper, we present the results of a study in which we compared six different PGAS and evaluated their performance in terms of security and convenience, with the aim of suggesting the system that has the most beneficial trade-off depending on the type of application.

Keywords: password meters · cybersecurity · credentials

1 Introduction

In the last decade, novel access systems implementing sophisticated credentials, such as, two-factor authentication and biometric identification [1], have increasingly been utilized for protecting the security and privacy of information in devices and accounts. Nevertheless, text-based passwords are still the most common authentication method for accessing websites, electronic mailboxes, and other types of accounts (e.g., wireless networks). However, the main limitation of passwords lies in the paradox of the trade-off between security and convenience: strong and secure passwords typically are inconvenient and difficult to remember [2]. As a result, research showed that users tend to utilize a total of 3-6 passwords shared between multiple accounts, even if they adopt a password manager; moreover, they prefer to create key phrases that are easy to remember by including information that is meaningful to them (such as, important dates and names) which, in turn, affects their security [3].

Password generation aiding systems (PGAS) have been developed to address the issue and help users increase the strength of their credentials by enforcing specific requirements (e.g., minimum length, presence of special symbols, or entropy) when they create a new password. Although they provide users with feedback about security score, research showed that users typically approach most PGAS as a checklist, and they do not result in any security improvement beyond the lowest mandatory level [3]. Indeed, there are several types of PGAS and they have very different strategy, design, and interface characteristics. However, only a few studies took into consideration the performance implications of the different components of user experience in password generation aiding systems, such as, interface, usability, and type and timeliness of feedback.

2 Related Work

Text-based passwords can be considered an imperfect method of authentication. For users to have strong and secure key phrases, they must create them in ways that make them inconvenient and difficult to remember [2]. Previous research studied users' behavior in password creation: users tend to reiterate the same key phrase over multiple accounts and prefer to use alphanumeric strings that are easy to remember and type whilst including something meaningful to them [3]. This challenge of generating passwords falls into the trade-off of security and convenience, where a convenient and easy to remember password is inconvenient for the user and difficult to remember [4]. Developers have attempted to influence users to create more secure passwords through the implementation of real-time feedback systems. These systems generally attempt to give users a real-time security evaluation of their password, considering the entropy, length, characters used, and looking for non-sequential letters and numbers. The calculated score of the password meter gives the user an estimate into the security of their password.

Research has attempted to determine the effectiveness of password feedback systems in aiding users to create more secure passwords. Studies have found that in all cases password feedback systems have influenced users with low password scores to create a more secure password [5]. The feedback systems studied have varied from simplistic password meters, to detailed systems and peer feedback. When users were questioned about which feedback system provided the best information and would most likely use, respondents selected a password meter which provided the most information about their password; however, the users also responded that this detailed password feedback was the most difficult to understand as well [6]. Another study of password feedback focused on a peer feedback password meter. This meter would show how the newly created password compared to other user's passwords on the site. Researchers found that the peer feedback system did not provide a significant effect on motivating users to increase the security of their password unless explicit instructions were included with the meter [6].

Password feedback mechanisms may help to improve users with low password security; however, it seems that context and account type may be a key identifying factors for if the feedback is taken into consideration by the user. A study found that password feedback systems did not increase password security for users creating unimportant accounts, even though they are commonly deployed on such sites. However, for accounts that contain sensitive information users appeared to take the password feedback into account when changing their passwords on these sites. Additionally, researchers found that providing a password meter for users when changing passwords helps to influence users to create stronger passwords, opposed to showing a password meter when first creating the account [7].

Password meters are currently active in many of Alexa's 100 most visited global sites. Out of the top 100 sites, 96 allow the ability for users to create a password and account, out of the 96, 73% gave some sort of feedback to users on their passwords, many of the sites used similar or the same password meter ranging from a bar meter to checkmarks systems after meeting requirements [8]. These systems were dynamically updated as the user types and considered the length, characters used, and occasionally blacklisted words [10]. A large-scale study on the effectiveness of password meters found that overall, password meters do change user behavior when interacting with them. Researchers found that users seeing a password meter nudged users to create longer passwords. Furthermore, findings suggested that the visual component of a password meter did not lead to significant differences. Users presented with a dancing bunny password meter reacted the same as those presented with traditional password feedback; however, the combination of text and a visual component was an important factor in the effectiveness. Additionally, the researchers found that users substantially changed their behavior when they were presented with stringent meters that would purposefully lower the security score of the password to motivate users to create a more secure password. Moreover, the stringent meters did motivate users to change their passwords; however, users also reported these meters to be annoying and frustrating to use [9].

Most research agrees that password feedback can help improve user's passwords [6] [7] [9]; however, password meters with requirements may be viewed by some users as a checklist. A study on the impact of feedback on password-creation behavior found that some users would quit adding characters and length to their passwords after fulfilling the password feedback requirements, such as: minimum of eight characters, at least one upper and lower-case letter, etc. Researchers theorized that some users may view the feedback as requirements and quit improving their password security after fulfilling them as they give the user a feeling of finality from reaching the requirements. Whereas, in situations where the password feedback did not have requirements, users may not be aware that the requirements were met and add additional security such as length and special characters. Another, theory the researchers had about why users may stop adding security after meeting requirements is that they rely on the feedback system for security. Therefore, users trust the feedback system, relying on it to help them create a secure password. Ultimately, the researchers recommend password feedback systems prompt users to continue adding security mechanisms to their passwords after fulfilling password requirements [10].

Implementing password feedback systems and meters may help to improve user passwords by giving explicit instructions and providing a visual representation, usually in a horizontal bar that increases and fills as the security increases. Moreover, this does not help improve the problem of users and poor password management. A large-scale study of user-password interaction found that the average user had 6.5 passwords shared between 3.9 sites [8]; another study found that 63.8% users reported using their password elsewhere, despite being aware of security practices [7]. This illustrates the problem of getting users to take real-time password feedback mechanisms and instructions into account when creating a new account. Moreover, it seems that younger generations may have the worst password security practices. A study found that younger users tend to ignore password feedback requirements for creating secure passwords: they may persuade themselves that the contents of their accounts are of little use to malicious users, not taking into consideration that their login credentials for secure accounts may be reused or similar [9]. In addition to password security, other studies [10] analyzed current issues and practices in enforcing username security. In conclusion, given current trends in cybercrime [12] and the rapidly changing dynamics of Internet consumption (e.g., cybermigration [13]), PGAS keep playing a fundamental role in creating awareness and fostering security.

3 Strategies in Password Generation Aiding Systems

In this Section, we review the most common strategies utilized in PGAS for helping users increase the security of their passwords: reactive techniques include suggesting guidelines, enforcing requirements, and giving feedback, whereas proactive methods automatically generate a secure key phrase for the user. They are detailed in Figure 1. Although their functionality might vary depending on the type of information in an account, their general purpose is to improve the trade-off between security, that is, effort to crack, and convenience (i.e., effort to generate and use).

3.1 Showing Password Guidelines

In its simplest form, a password generation aiding strategy would display the guidelines for creating a more secure key phrase. As discussed earlier, they might vary depending on the type of information that is maintained in the digital resource. Moreover, in its simplest form, a PGAS would list password criteria as a suggestion to improve the security of the key phrase, such as, minimum length, use of a mix of numbers and letters with mixed case, and adding some special symbols. However, it would not enforce them, enabling users to proceed with account registration even if the password does not meet the requirements. As a result, users are left with the responsibility of creating a strong password, and they can opt for a key phrase that they consider convenient, even if less secure.

3.2 Enforcing Password Requirements

The majority of PGAS implemented in today’s systems fall in this category: in addition to showing the requirements, they enforce them by preventing users from creating an account or changing their passwords if the chosen key phrase does not meet the specified criteria. The main difference with the system described in 3.1 is in that systems enforce specifications as requirements. As a result, the user must generate a password that meets the level of security of the system. Furthermore, this type of PGAS typically provide feedback, either in real-time or after form submission, about the items that have not been addressed yet. The advantage of this system is two-fold:

(1) it prevents creating passwords that are below a specific strength and (2) it educates users about security by having them practice.

3.3 Strength meter

As the level of security of a password is directly proportional to its complexity, it can be increased by manipulating two dimensions, that is, length and content diversity. Primarily, key phrase guidelines and requirements have the purpose of supporting users in generating a key phrase that is compliant with both criteria. Nevertheless, a secure password could be achieved by manipulating either length or content. As a result, instead of specifying requirements, strength meters evaluate password entropy without taking into consideration its specific components. Unfortunately, visualization in the form of labels might create some ambiguity because password meters might yield different results depending on the security requirements of the system. As a result, a password that is considered good by the PGAS of a website could be graded differently in another that has higher standards.

3.4 Time to Crack

Strength meters are a user-friendly abstraction of the concept of entropy: labels identifying password's strength are a good compromise between the need of protecting an account with a secure key phrase and the possibility of giving users the possibility of manipulating either of the two degrees of freedom (i.e., length or entropy). Moreover, they provide quick and convenient feedback. Similarly, systems using time-to-crack as a measure of password security display entropy as an estimate of the time that a brute-force attack would require to guess the password. Unfortunately, this type of PGAS suffer from the same limitations as strength meters. Moreover, time to crack depends on a variety of factors: its measurement might not be accurate and might significantly vary depending on the type of technique and resources used in an attack. Also, as cybercrime evolves, novel and sophisticated techniques are expected to be faster in breaching systems.

3.5 Comparing with User Base

As discussed in Section 2, studies showed that users tend to underestimate their own risk of being hacked. Thus, they tend to prefer convenience in favor of security though they are aware of cybersecurity practices. To this end, password requirements, strength meters, and time-to-crack estimates provide an impersonal measure of a hypothetical risk. Conversely, peer feedback has been studied as a strategy to incentivize individuals to adopt stronger key phrases by comparing them to others. As a result, this type of password meter engages users in a competitive dynamic aimed at fostering the creation of a key phrase that is stronger than the average password utilized by the members of a community. Indeed, this psychological trigger could be implemented by merely changing the interface of a standard password meter based on entropy.

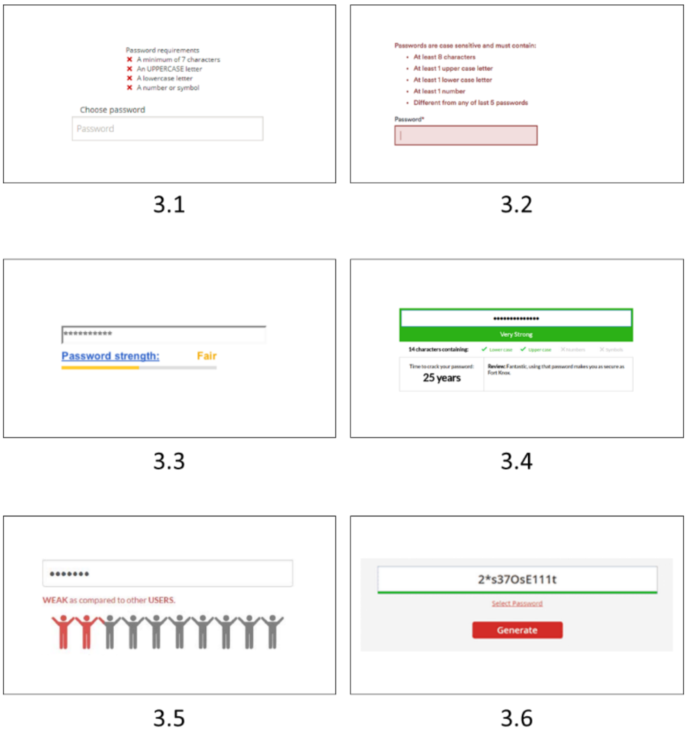
3.6 Automatic Password Assignment

Studies about passwords demonstrated that users are very prone to forgetting their passwords when it is particularly complex, and thus, secure. Typically, this happens when users do not utilize password managers, or when the password is not stored in an account manager. Moreover, considering that nowadays users have dozens of accounts, they might have to generate, store or remember, and use many different passwords for websites and digital resources that they might access very seldom. As a result, several even started questioning the need of generating a password that is easy to remember: depending on the frequency of use of an account, users might want to create a password that is very secure and hard to remember, and then, either use a password manager or rely on the procedure for recovering their authentication information. Consequently, software and websites can automatically generate a secure key phrase for the user to store in a password manager, so that they can make sure that the security requirements. PGAS using automatic password assignment increase convenience of creating account credentials but rely on a third-party software or procedure (e.g., password manager or password recovery process) to cope with higher probability of password forgetfulness.

4 Experimental Study

In this Section, we detail a study in which we compared the different strategies and types of PGAS. Specifically, we evaluated the impact of human factors on the relationship between usability and performance in terms of security and convenience trade-off. To this end, we designed a web-based account creation system that implements the six different PGAS discussed in Section 3. We recruited a total of 115 subjects (33 females and 82 males aged 28±9). Participants were sampled from a population having low to medium familiarity with IT (based on degree, background, time spent working with a desktop computer, and number of accounts) to prevent results from being biased by computer literacy.

In the experimental task, participants were directed to an Internet URL in which: (1) they were presented with a sign-up form incorporating a password generation aiding system selected at random and they were asked to create an account using a key phrase they never utilized in the past and save the information using the method of their choice; (2) they were redirected to an account activation page that had the purpose of interrupting their task; (3) they were presented with a sign-in form and they were asked to use their credentials. Finally, they were asked to fill a questionnaire in which they evaluated the convenience and usability of PGAS in the sign-up phase and in signing in. Participants were asked to realize the task six times (one for each type of system). The order of PGAS was selected at random to prevent training effect. The experimental software calculated the entropy of the password generated using each PGAS. Moreover, it acquired the time required for typing the key phrase at sign up (calculated from the first character entered until the user left the password field) and for signing in (calculated from page load until the user clicked the sign in button). The security criteria were the same for all systems: minimum 8 characters long, at least a number and an uppercase and a lowercase letter.



Password requirements
- ✗ A minimum of 7 characters
- ✗ An UPPERCASE letter
- ✗ A lowercase letter
- ✗ A number or symbol

Choose password

Password

3.1

Passwords are case sensitive and must contain:
- At least 8 characters
- At least 1 upper case letter
- At least 1 lower case letter
- At least 1 number
- Different from any of last 5 passwords

Password*

3.2

****

Password strength: Fair

3.3

••••••••••••••

Very Strong

14 characters containing: ✓ Lower case ✓ Upper case ✗ Numbers ✗ Symbols

Time to crack your password: 25 years

Review: Fantastic, using that password makes you as secure as Fort Knox.

3.4

•••••••

WEAK as compared to other USERS.

3.5

2*s37OsE111t

Select Password

Generate

3.6

Fig. 1. Different types of Password Generation Aiding Systems and strategies as described in Section 3 and utilized in the study: showing password requirements (3.1) is the most basic approach to security, which can be enforced using systems that require the specifications to be met (3.2). Strength meters (3.3) score password robustness whereas systems based on time to crack (3.4) measure the security of a password in terms of time required to guess it, which can be estimated using length and entropy measurement. Also, non-conventional meters display the strength of the chosen password as compared to other users (3.5). Finally, a few software and websites are automatically generating secure passwords for the user when they register (3.6).

5 Results and Discussion

PGAS that suggest or require specific symbols resulted in the lowest overall score, as they required 4.20 seconds on average to create and use, had the least entropy and length, and were preferred last in terms of convenience. In line with findings in the literature, subjects utilized them as a checklist. Conversely, password meters had better results, though label-based systems had different outcome than meters based on time-to crack and peer comparison. Specifically, the former resulted lasts in terms of users' perception even if they had better performance, both as calculated effort to create and use, and as resulting entropy and length. Finally, PGAS that automatically generate a password resulted in the lowest effort to create and use and in the best preference. Although they require users to store the key phrase in password manager, they also prevent reusing the same word across multiple accounts and, thus, might have an impact on the overall security of the user. Table 1 shows a summary of all the experiment data. Figure 2 reports the effort to create and use the password, calculated as seconds spent in the sign up and sign in phases. Figure 3 indicates password strength, measured in terms of key phrase entropy and length. Perceived convenience in account creation and access was recorded using a Likert scale (see Figure 4). Overall, users spent between 3 and 5 seconds on average to sign up and sign in. The entropy of password ranged from weak to strong depending on the PGAS, whereas the average length was 10 characters. Finally, systems were perceived similarly, though there is a statistical difference between them.

Table 1. Experiment data acquired from the different PAGS described in Section 3: password guidelines (3.1), password requirements (3.2), strength meter (3.3), time to crack indicator (3.4), peer strength meter (3.5), and password generator (3.6).

PGAS	**3.1**	**3.2**	**3.3**	**3.4**	**3.5**	**3.6**
Time to sign up (seconds)	4.15 ±0.85	4.24 ±0.88	4.11 ±0.88	4.36 ±1.44	4.12 ±1.42	0.25 ±0.83
Time to sign in (seconds)	4.42 ±1.44	4.26 ±1.33	3.78 ±1.17	5.08 ±1.68	4.91 ±1.74	5.75 ±3.06
Entropy (bits)	39.62 ±7.30	41.22 ±8.13	47.26 ±8.80	43.28 ±9.17	49.24 ±8.30	47.5 ±8.71
Length (characters)	9.45 ±1.05	9.47 ±1.15	11.25 ±2.01	11.15 ±1.92	11.12 ±2.03	9.11 ±0.57
Convenience in sign up (Likert)	3.38	3.47	3.41	4.36	4.29	4.85
Convenience in sign in (Likert)	3.12	3.21	3.77	2.77	3.04	2.74

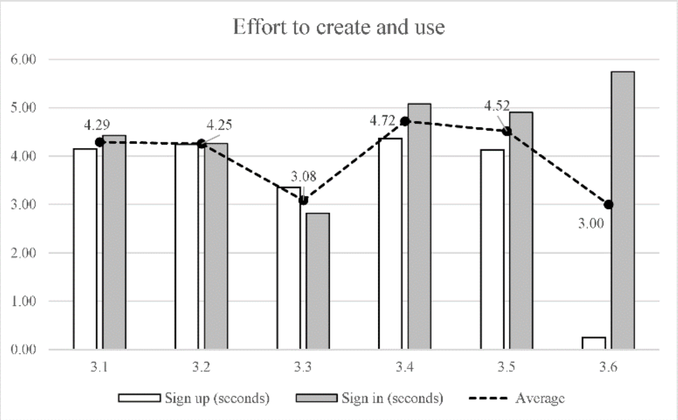


Fig. 2. Effort to create and use a password, measured in seconds required to generate it and to use it. PGAS that automatically generate a password (3.6) ranked best, though there was is a 5 second difference between sign up and sign in phases. Standard label-based password meters are the most efficient alternative.

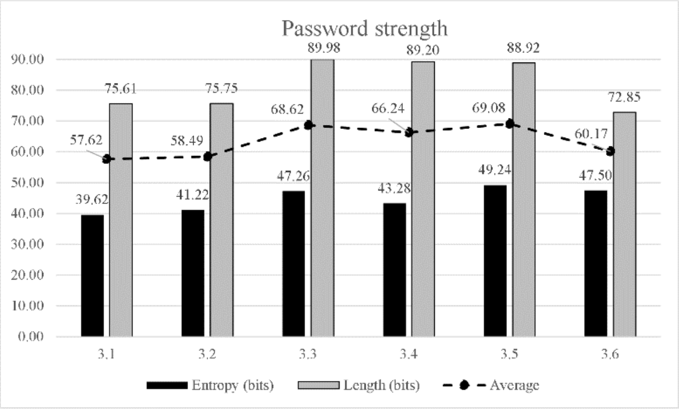


Fig. 3. Password strength calculated as entropy and length. The latter, which was initially measured in characters, was converted to bits (8 bits per character) for visualization purposes. Password meters resulted in the highest score of entropy and length. However, this was because only a few subjects changed the automatically-generated password in (3.6), which is statistically comparable to 3.3 in terms of entropy.

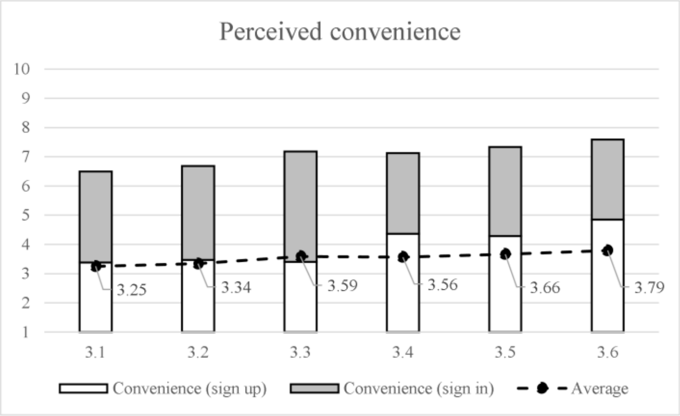


Fig. 4. Perceived convenience in the sign up and sign in phases, scored on a Likert scale. Systems that automatically generate a password for the user (3.6) ranked best, though individual steps differ significantly. Also, respondents indicated a preference for password meters over system that suggest or enforce password requirements.

6 Conclusion

In this paper, we presented an overview of the main types of currently available PGAS, we highlighted their differences, and we discussed the results of a comparative evaluation of their security performance and perceived usability. From our findings, we can conclude that interface design strategies lead to better trade-off between password robustness and convenience. Specifically, meters based on simple labels result in improved entropy, length, and usability. However, automatically-generated passwords score best in convenience and security, and they prevent users from reusing the same key phrase, though they require third-party systems for storing the password.

References

1. Bevilacqua, V., Cariello, L., Columbo, D., Daleno, D., Fabiano, M.D., Giannini, M., Mastronardi, G. and Castellano, M., 2008, September. Retinal fundus biometric analysis for personal identifications. In International Conference on Intelligent Computing (pp. 1229-1237).
2. Bonneau, J., Herley, C., Van Oorschoto, P. C., & Stajano, F. (2015). Passwords and the evolution of imperfect authentication. Communications of the ACM, 58(7), 78-87. https://doi.org/10.1145/2699390
3. Stainbrook, M., & Caporusso, N. (2018) Convenience or strength? Aiding optimal strategies in password generation. Proceedings of Advances in Human Factors in Cybersecurity. AHFE 2018. Advances in Intelligent Systems and Computing, Vol 782. https://doi.org/10.1007/978-3-319-94782-2_3
4. Tam, L., Glassman, M., & Vandenwauver, M. (2010). The psychology of password management: A trade-off between security and convenience. Behaviour & Information Technology, 29(3), 233-244. https://doi.org/10.1080/01449290903121386
5. Ciampa, M. (2013). A comparison of password feedback mechanisms and their impact on password entropy. Information Management & Computer Security Vol (21).
6. Dupuis, M., & Khan, F. (2018). Effects of peer feedback on password strength. APWG Symposium on Electronic Crime Research, San Diego, CA, pp.1-9. https://doi.org/10.1109/ECRIME.2018.8376210
7. Egelman, S., Sotirakopoulos, A., Muslukhov, I., Beznosov, K., & Herley, C. (2013). Does my password go up to eleven? The Impact of password meter on password selection. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems ACM, New York, USA, 2379-2388. doi https://doi.org/10.1145/2470654.2481329
8. Florêncio, D., & Herley, C. (2007). A large-scale study of web password habits. In Proceedings of the 16th International Conference on the World Wide Web, ACM Press (New York, NY, USA, 2007), 657–666
10. Ur, Blase & Gage Kelly, Patrick & Komanduri, Saranga & Lee, Joel & Maass, Michael & Mazurek, Michelle & Passaro, Timothy & Shay, Richard & Vidas, Timothy & Bauer, Lujo & Christin, Nicolas & Cranor, Lorrie. (2012). How does your password measure up? The effect of strength meters on password creation. Proc. Security '12, USENIX Association.
11. Shay, R., Bauer, L., Christin, N., Cranor, L. F., Forget, A., Komanduri, S., & Ur, B. (2015, April). A spoonful of sugar? The impact of guidance and feedback on password-creation behavior. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 2903-2912).
12. Caporusso, N., Chea, S. and Abukhaled, R., 2018, July. A Game-Theoretical Model of Ransomware. In International Conference on Applied Human Factors and Ergonomics (pp. 69-78). Springer, Cham. https://doi.org/ 10.1007/978-3-319-94782-2_7
13. Xiao, X. and Caporusso, N., 2018, August. Comparative Evaluation of Cyber Migration Factors in the Current Social Media Landscape. In 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW) (pp. 102-107). IEEE. https://doi.org/10.1109/W-FiCloud.2018.00022
14. Fandakly, T. and Caporusso, N., 2019, July. Beyond Passwords: Enforcing Username Security as the First Line of Defense. In International Conference on Applied Human Factors and Ergonomics. Springer. To be published.