An Evaluation of the Efficacy of Password Strength Meters

CPSC 525 Project Proposal – Group 10

Mark De Castro
University of Calgary
Calgary, Alberta, Canada
mdecastr@ucalgary.ca

Carlos Hernandez Rosas
University of Calgary
Calgary, Alberta, Canada
carlos.hernandezrosa@ucalgary.ca

Masih Sadat
University of Calgary
Calgary, Alberta, Canada
masih.sadat2@ucalgary.ca

Niroojen Thambimuthu
University of Calgary
Calgary, Alberta, Canada
niroojen.thambimuthu@ucalgary.ca

Cole Towstego
University of Calgary
Calgary, Alberta, Canada
ctowsteg@ucalgary.ca

ABSTRACT

Passwords remain the most common form of user authentication that we use in our daily lives. As attackers use ever more sophisticated methods of techniques in an endeavor to crack passwords, it is imperative that users are able to select passwords which are strong and secure. The most common tool to assist with this is the strength meter, prevalent in a large number of websites. However, the exact implementation of these password meters can vary greatly, potentially confuse the user, and may not effectively evaluate the strength of a given password. We will be attempting to examine if these password meters are in fact truly effective in their goals of helping users to select a password that is secure.

1 PROBLEM

We use passwords to protect the information we store on various web services, ranging from our social media accounts to the sensitive data found in our bank accounts. However, users of such websites may not always be aware of what a strong password may be, or may have no motivation to create one as long as they are able to immediately access these services. It then becomes the task of these websites to aid such users in formulating secure passwords which will help protect the information they store from possible threats or attacks. They commonly use password strength meters to do so. These are often the coloured bars that provide visual feedback to a user where in the range of “weak” to “strong” their proposed password lies. They are often accompanied by a set of guidelines of what a password should be comprised of. A number of studies have been conducted in the past to assess the integrity and efficacy of these password meters. In 2014, de Carnavalet and Mannan conducted a large-scale analysis of the password meters of high-profile websites where they concluded that such meters are “highly inconsistent, fail to provide coherent feedback, and sometimes provide strength measurements that are blatantly misleading [1].” Wang and Wang in 2015 conducted a similar empirical analysis of the password policies of 50 leading web services, and found that they likewise provide highly inconsistent outcomes under identical testing, and that they “largely fail to withstand online guessing attacks [2].” Similarly, in a paper by Ur *et al.*, while they found that these meters do affect user behaviour, they state that “the resulting passwords were only marginally more resistant to password cracking attacks [3].” Thus, it is apparent that there is a serious problem among these password strength meters, in that they vary greatly in implementation, which can lead to incoherent or confusing feedback, and that they may not actually be correctly assessing a password’s strength. These findings thus weaken the purpose of these password meters. This poses an alarming security risk, as this can allow individuals to have the wrong perception of what a strong password is. Consequently, with the ever growing number of attacks happening among various websites today (such as the common occurrences of password leaks), individuals may then be more susceptible to password cracking and guessing attacks, and leaks of personal information may become more prevalent. Therefore, with a constantly growing online world, it is relevant to analyze and reassess these password meters to see if there have been changes made in increasing their integrity and security, and if they are indeed the effective tools they ought to be.

2 APPROACH

We will repeat and extend parts of the empirical analyses and studies on the password strength meters of many popular websites, as found in the cited papers above, and in any other studies that we may find as our project progresses. To perform our evaluation, we will first choose a number of popular, high-profile websites to work with (as based on their rankings from *Alexa Internet*, which ranks web pages according to their web traffic over three-month periods). We aim to select a diverse list of websites, encompassing different areas ranging from social networks to online retailers, so that our study will be representative enough of the large number of web services available today. We will utilize pre-existing leaked passwords to conduct our tests (such as those released by Mark Burnett from 2015 [4]). This will allow us to use real passwords from real individuals, and in doing so, we can ensure that the tests we are performing are emulating real-world scenarios and real user behaviour as much as possible. To conduct our tests and analyses, we will build upon and use parts of the techniques and tools used by de Carnavalet and Mannan, Wang and Wang, and Ur *et al.* (and of any others that we may deem appropriate for our research) to assess the password meters of our selected websites. These may include: the analysis, assessment and comparison of the password meters' notable characteristics, policies and guidelines, a test to verify if they use any blacklists, an analysis of the algorithms they use, or even emulating online guessing attacks, among several. Then, as an extension of these studies, we will propose and formulate our own technique or tool to assess the efficacy and integrity of these password strength meters. To do so, we will adhere closely to the regulations produced by the National Institute of Standards and Technology (NIST), such as those found in the *Digital Identity Guidelines* from 2017, which published recommendations for the protection and confidentiality of passwords [5], so that our tests and analyses will abide by the standards set by the security industry. Finally, we will compare our findings to those in the original studies and note if any significant changes or improvements have occurred.

3 ANTICIPATED RESULTS

By the end of our study, we expect to be able to verify if the password strength meters among various web services are indeed effective tools in providing security (i.e., if they are actually helping users in creating more secure passwords). We also expect to be able to list any vulnerabilities or threats that are present in their implementations, as well as any notable weaknesses they may possess, and the possible inconsistencies in the resulting assessments they provide of an individual's desired password. Furthermore, with our findings, we aim to produce a list of standard but minimal list of requirements of what is needed to be able to create an effective password strength meter, which these websites can use. However, it should be noted that every password meter should still have some distinction from other meters (in terms of implementation), as having one stringent set of rules or requirements of how these meters should operate will only make the lives of attackers much easier. In the end, we expect that our findings will aid in further improving the integrity and security of these existing password meters, and thus allow them to be the effective tools that they are required to be.

4 RELEVANCE TO THE COURSE

This project is relevant to the course as it falls under the scope of: (1) *software security* since, through this, we will be able to pinpoint any vulnerabilities that password meters may possess as a result of poor implementation or development choices, or if any tools or algorithms were improperly utilized, and thus be aware of any exploits attackers may use to compromise the security of online services; and (2) *authentication* and *access control*, since the use of passwords is very common in regulating the access of various resources. With this project, we expect to learn what constitutes a strong and secure password, and thus be able to properly study and critique the tools and techniques used in assessing the efficacy of password meters. In doing so, we will have the chance to use common security evaluation standards, such as those promulgated by NIST, and thus use these to produce proper reviews of the security mechanisms found in the implementation of password meters. Consequently, we will also be made aware of the current methods that attackers may be using today to compromise the security of various systems. Indeed, these aforementioned goals match those found in the learning objectives for this course.

REFERENCES

[1] Xavier de Carné de Carnavalet and Mohammad Mannan. Large-scale evaluation of high-impact password strength meters. *ACM Transactions on Information and System Security*, 18(1):1–32, June 2015.

[2] Ding Wang and Ping Wang. The emperor's new password creation policies: an evaluation of leading web services and the effect of role in resisting against online guessing. In *Proceedings of the 20th European Symposium on Research in Computer Security (ESORICS'2015)*, pages 456–477, Vienna, Austria, September 2015.

[3] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L. Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. How does your password measure up? The effect of strength meters on password creation. In *Proceedings of the 21st USENIX Security Symposium (USENIX Security'12)*, pages 65–80, Bellevue, Washington, USA, August 2012.

[4] Mark Burnett. Today, I am releasing ten million passwords. https://xato.net/today-i-am-releasing-ten-million-passwords-b6278bbe7495. Accessed February 1, 2018.

[5] Paul A. Grassi, Michael E. Garcia, and James L. Fenton. NIST Special Publication 800-63-3: Digital identity guidelines. Technical report, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, June 2017.