Design and Evaluation of a Data-Driven Password Meter

Blase Ur*, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, Noah Johnson, William Melicher
*University of Chicago, Carnegie Mellon University
blase@uchicago.edu
{fla, mza, lbauer, nicolasc, jcolnago, lorrie, hdixon, pardis, hana007, noah, billy}@cmu.edu

ABSTRACT
Despite their ubiquity, many password meters provide inaccurate strength estimates. Furthermore, they do not explain to users what is wrong with their password or how to improve it. We describe the development and evaluation of a data-driven password meter that provides accurate strength measurement and actionable, detailed feedback to users. This meter combines neural networks and numerous carefully combined heuristics to score passwords and generate data-driven text feedback about the user's password. We describe the meter's iterative development and final design. We detail the security and usability impact of the meter's design dimensions, examined through a 4,509-participant online study. Under the more common password-composition policy we tested, we found that the data-driven meter with detailed feedback led users to create more secure, and no less memorable, passwords than a meter with only a bar as a strength indicator.

ACM Classification Keywords
K.6.5 Security and Protection: Authentication; H.5.2 User Interfaces: Evaluation/methodology

Author Keywords
Passwords; usable security; data-driven; meter; feedback

INTRODUCTION
Password meters are used widely to help users create better passwords [42], yet they often provide ratings of password strength that are, at best, only weakly correlated to actual password strength [10]. Furthermore, current meters provide minimal feedback to users. They may tell a user that his or her password is "weak" or "fair" [10, 42, 52], but they do not explain what the user is doing wrong in making a password, nor do they guide the user towards a better password.

In this paper, we describe our development and evaluation of an open-source password meter that is more accurate at rating the strength of a password than other available meters and provides more useful, actionable feedback to users. Whereas most previous meters scored passwords using very basic heuristics [10, 42, 52], we use the complementary techniques of simulating adversarial guessing using artificial neural networks [32] and employing 21 heuristics to rate password strength. Our meter also gives users actionable, data-driven feedback about how to improve their specific candidate password. We provide users with up to three ways in which they could improve their password based on the characteristics of their specific password. Furthermore, we automatically propose modifications to the user's password through judicious insertions, substitutions, rearrangements, and case changes.

In this paper, we describe our meter and the results of a 4,509-participant online study of how different design decisions impacted the security and usability of passwords participants created. We tested two password-composition policies, three scoring stringencies, and six different levels of feedback, ranging from no feedback whatsoever to our full-featured meter.

Under the more common password-composition policy we tested, we found that our data-driven meter with detailed feedback led users to create more secure passwords than a meter with only a bar as a strength indicator or not having any meter, without a significant impact on any of our memorability metrics. Most participants reported that the text feedback was informative and helped them create stronger passwords.

RELATED WORK
Users sometimes make predictable passwords [22, 30, 48] even for important accounts [13, 31]. Many users base passwords around words and phrases [5, 23, 29, 45, 46]. When passwords contain uppercase letters, digits, and symbols, they are often in predictable locations [4]. Keyboard patterns like "1qaz2wsx" [46] and dates [47] are common in passwords. Passwords sometimes contain character substitutions, such as replacing "e" with "3" [26]. Furthermore, users frequently reuse passwords [9, 14, 25, 38, 48], giving the compromise of even a single account potentially far-reaching repercussions. In designing our meter, we strove to help users understand when their password exhibited these common tendencies.

Three types of interventions attempt to guide users towards strong passwords. First, password-composition policies dictate characteristics a password must include, such as particular

CHI 2017, May 06–11, 2017, Denver, CO, USA
ACM 978-1-4503-4655-9/17/05.
http://dx.doi.org/10.1145/3025453.3026050