Poster: The Art of Password Creation

Blase Ur, Saranga Komanduri, Richard Shay, Stephanos Matsumoto, Lujo Bauer,
Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Michelle L. Mazurek, and Timothy Vidas
Carnegie Mellon University
Pittsburgh, PA, USA
{bur,sarangak,rshay,smatsumoto,lbauer,nicolasc,lorrie,pgage,mmazurek,tvidas}@cmu.edu

Offline attacks, in which an adversary acquires a hashed password database and attempts to guess its contents, remain a significant threat to password security. Recent, highly publicized examples include LinkedIn, eHarmony, Sony, and Gawker. Because many users reuse passwords (exactly or with minor adjustments), these passwords can have value beyond the source site [1].

In an effort to make passwords more resistant to guessing attacks, system administrators provide users with suggestions and/or requirements for the passwords they create. This guidance may include password-composition requirements, such as requiring the password to have a minimum length and include both letters and numbers. Other common strategies include using a meter to suggest or enforce composition rules, forbidding use of dictionary words, and forbidding common passwords. Many of these strategies were developed based on folk wisdom and educated guesses; until recently, most had not been empirically evaluated [2].

In previously published work, we tested different password-composition policies using 12,000 passwords collected in an online study [3], [4]. We found that requiring long passwords with no other restrictions provides stronger guessing resistance than other tested policies, while being more palatable to users than other relatively strong policies.

We also examined the effectiveness of password meters that nudge users toward stronger passwords without enforcing strict requirements [5]. We found that while meters with a variety of visual appearances led to longer passwords, only meters that scored passwords stringently led to significantly more guess-resistant passwords. Meters that were too stringent, however, led to increased user annoyance and in some cases to users discounting the importance of satisfying the meter.

Our recent work considers password strength from a new perspective: not only how guidance affects password strength, but why. We examine in depth how users create passwords, which words they use, and how the component pieces of passwords relate to each other. We analyze which characteristics of password creation are associated with strong and weak passwords when attacked with current crackers. We then search for patterns that current crackers do not exploit, but which could indicate vulnerability to cracking. Overall, we examine password creation more thoroughly than any previous analysis.

We find that passwords contain predictable patterns beyond well-known habits such as appending numbers to dictionary words. Our finding that chunks within patterns are related indicates that context-free analysis (e.g., [6]) discards potentially valuable information. We show that, in parts of speech and words chosen, passwords resemble each other more than natural-language English. We find that when a user's password is rejected for non-conformance with policy, users who make small edits rather than starting over tend to create stronger passwords. Counterintuitively, our password-guidance results also suggest that forcing users to comply with strict policies can sometimes result in less secure passwords.

In the following sections, we describe the datasets we analyze and highlight three of our findings in more detail.

Password datasets. We analyze 13,499 passwords collected in online studies for previously published experiments, as well as publicly available passwords leaked from the RockYou website (more than 32 million) and from Yahoo! (more than 450,000). This provides both the authenticity of real-world data and the detailed password-creation instrumentation of experimental data.

The experimental passwords were collected in the following conditions; none could be shorter than eight characters.

BASIC8: 8 or more characters.
DICTIONARY8: Not in the free Openwall dictionary.[1] (Non-alphabetic characters discarded before checking.)
BLACKLISTEASY: Not in the Unix dictionary.
BLACKLISTMEDIUM: Not in the paid Openwall dictionary.[2]
BLACKLISTHARD: Not in a set of 5×10^9 passwords generated using a probabilistic cracking algorithm [6].
COMPREHENSIVE8: dictionary8, plus an uppercase letter, lowercase letter, digit, and symbol.
BASIC16: 16 or more characters.

Adjacent words within passwords. We considered whether knowing one piece of a password provides an advantage for guessing the subsequent piece. We used Wang et al.'s Word Breaker [7] to divide passwords into chunks roughly approximating words. Then, using the Google Web N-gram Corpus[3], we considered how the probabilities of alphabetic chunks relate to each other. We found that 16% of passwords contained at least one digram AB such that the conditional probability of guessing B given A is greater than the probability of guessing B without context ($p(B|A) > p(B)$). Further, 40% of the AB digrams we examined ranked within the top 100 guesses for B given A; without context, B is found within the first 100 guesses only 11% of the time. Table I provides more details.

Words used in passwords. Prior work has established that passwords contain words [8], [9], but has not considered how

[1] http://download.openwall.net/pub/wordlists/
[2] http://www.openwall.com/wordlists/
[3] http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13

Condition	% with Digrams	Among digrams: % in corpus	Among digrams: % AB in Top 10	Among digrams: % B in Top 10	Among digrams in the corpus: % $p(B\|A) > p(B)$	Among digrams in the corpus: Median Ratio
RockYou	28	91	14	6	69	4.4
Yahoo!	31	89	13	6	66	3.1
meterStandard	45	81	16	9	65	4.4
meterStringent	48	83	17	10	65	2.9
basic8	40	90	18	7	65	2.9
dictionary8	58	86	14	7	63	1.9
blacklistEasy	43	90	15	7	64	2.1
blacklistMedium	48	88	17	6	64	1.8
blacklistHard	55	86	14	7	61	1.6
comprehensive8	59	89	14	7	61	1.7
basic16	79	80	23	8	70	5.8
Total	30	90	13	6	67	3.8

TABLE I. DIGRAM ANALYSIS. A TOP-10 RANK FOR AB REFERS TO A FREQUENCY-ORDERED LIST OF ALL DIGRAMS BEGINNING WITH A IN THE GOOGLE CORPUS. THE MEDIAN RATIO REFERS TO $\frac{p(B|A)}{p(B)}$. ROCKYOU RESULTS CALCULATED ON A RANDOM SAMPLE EQUAL IN SIZE TO YAHOO!

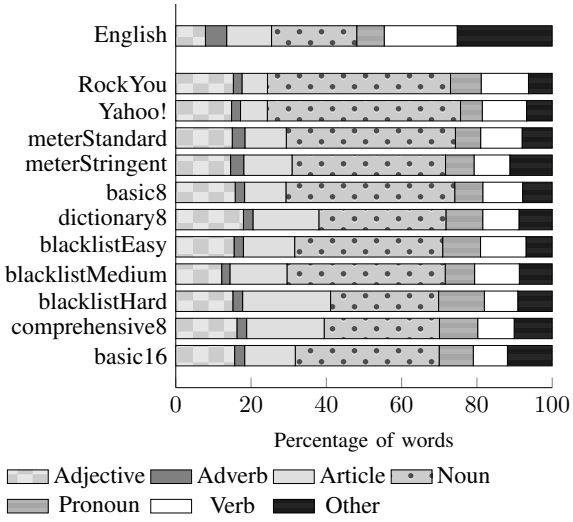

Fig. 1. The distribution of parts of speech for words in English (COCA) and in passwords.

similar these words are to natural language. We used the Corpus of Contemporary American English (COCA) [10] to assign parts of speech to password chunks and compare the distribution to natural language. Passwords were more likely than English to contain nouns and adjectives, but much less likely to contain verbs or adverbs (Figure 1).

We also considered how password chunks differ between sets of passwords, as well as between passwords and English. Using Jensen-Shannon Divergence (JSD), we found that most password sets were relatively similar to each other, with RockYou least similar to the others; none of the password sets were very similar to COCA. Combined with the part-of-speech results, this suggests passwords are more like other passwords than like natural-language English.

Condition	Required retries (%)	Mean # of retries	*Impact on guessability (%)*: Harder to guess	*Impact on guessability (%)*: Easier to guess
dictionary8	19.0	2.3	63.5	13.3
blacklistEasy	1.0	1.5	81.8	0.0
blacklistMedium	6.0	1.8	90.6	6.3
blacklistHard	20.1	2.7	55.1	11.5
comprehensive8	52.0	3.1	46.8	3.2
basic16	9.5	1.9	47.5	0.0

TABLE II. RETRIES FOR POLICY COMPLIANCE. WE COMPARE ORIGINAL, NON-COMPLIANT PASSWORDS WITH FINAL, COMPLIANT PASSWORDS. IF NEITHER WOULD HAVE BEEN GUESSED WITHIN OUR THRESHOLD, THE SECURITY IMPACT IS UNKNOWN.

Policy compliance strategies. We examined how users modified their passwords after a non-compliant password was rejected, considering only those users who submitted at least one non-compliant password of length eight or greater. Using a password-guessing calculator [4], we compared the strength of each user's initial (non-compliant) and final (compliant) password. In all conditions, more passwords became harder to guess than became easier to guess; surprisingly, however, many users in dictionary8 (13.3%) and blacklistHard(11.5%) ended up with weaker passwords. Table II presents details.

Acknowledgments. This research was supported by NSF grants DGE-0903659 and CNS-1116776, by CyLab at Carnegie Mellon under grants DAAD19-02-1-0389 and W911NF-09-1-0273 from the Army Research Office, by Air Force Research Lab Award No. FA87501220139, and by a gift from Microsoft Research.

REFERENCES

[1] D. Florêncio and C. Herley, "A large-scale study of web password habits," in *Proc. WWW*, 2007.

[2] W. E. Burr, D. F. Dodson, and W. T. Polk, "Electronic authentication guideline," NIST, Tech. Rep., 2006.

[3] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman, "Of passwords and people: Measuring the effect of password-composition policies," in *Proc. CHI*, 2011.

[4] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez, "Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms," in *Proc. IEEE Symp. Security & Privacy*, May 2012.

[5] B. Ur, P. G. Kelly, S. Komanduri, J. Lee, M. Maass, M. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor, "How does your password measure up? The effect of strength meters on password creation," in *Proc. USENIX Sec.*, Aug. 2012.

[6] M. Weir, S. Aggarwal, B. d. Medeiros, and B. Glodek, "Password cracking using probabilistic context-free grammars," in *Proc. IEEE Symposium on Security and Privacy*, 2009.

[7] K. Wang, C. Thrasher, and B. Hsu, "Web scale nlp: a case study on url word breaking," in *Proc. WWW*, 2011.

[8] J. Campbell and K. Bryant, "Password Composition and Security: An Exploratory Study of User Practice," *Proc. ACIS*, 2004.

[9] M. Zviran and W. J. Haga, "Password security: an empirical study," *J. Mgt. Info. Sys.*, vol. 15, no. 4, 1999.

[10] M. Davies, "The corpus of contemporary American English: 425 million words, 1990–present," Available online at http://corpus.byu.edu/coca/, 2008.