Measuring the Effect of Pa$$word-Composition Policies on Security and Usability

Michelle L. Mazurek, Patrick Gage Kelley, Saranga Komanduri, Richard Shay, Lujo Bauer, Lorrie Faith Cranor, Nicolas Christin, Serge Egelman*, Julio Lopez
Carnegie Mellon University, *NIST

Motivation

- Text-based passwords remain ubiquitous
- Attackers' password-guessing capabilities are improving
- In response, sysadmins adopt complex **password-composition policies** to encourage stronger passwords
  - e.g., passwords must have at least 8 characters, including a digit and a symbol
- Little is known about the practical effect of such policies on password strength
- Little is known about how such policies affect user behavior
  - e.g., writing down passwords

Our goal: Measure the effect of policy on password strength and usability

Subgoal: Evaluate information entropy as a measure of password strength

Measuring passwords effectively

1. Data collection

- More than **12,000 passwords**, collected under controlled conditions via Mechanical Turk [2]
  - Largest study of its kind
- Part 1: Create a password for a given policy
- Part 2: Recall the password 3 days later
- Both parts: Survey about behavior, sentiment
- Participants split among 7 conditions, each with a different password-composition policy

2. Password-composition policies we tested

basic8: must have at least 8 characters
basic16: must have at least 16 characters
blEasy: must not appear in 4-million-word blacklist
blMedium: must not appear in 40-million-word blacklist
blHard: must not appear in 5-billion-word blacklist
dict8: must have at least 8 characters, no dictionary word
comp8: must have at least 8 chars including uppercase, lowercase, symbol, digit; no dictionary word

3. Measuring guessability: a new technique

- Computes # of guesses needed to crack a password
- Much more efficient than guessing directly
- Requires plaintext passwords, deterministic guess algorithm
- Tried **40+ combinations** of guess algorithm, training data
- Distributed computation using Hadoop
- Measure resistance to **50 trillion+** guesses

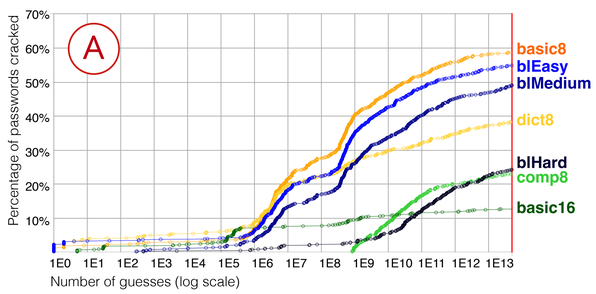
4. Comparing guessability to entropy

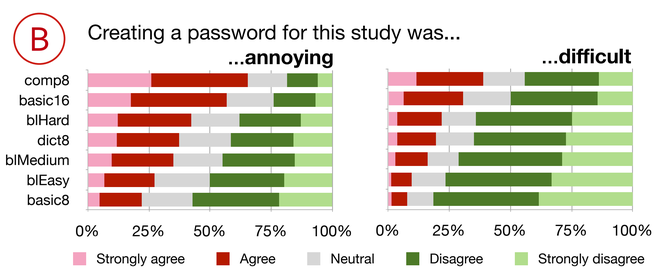
- Information entropy is a popular but controversial measure of password strength
- Empirical calculation of entropy [3]
  - Based on Shannon's formula
  - Sums entropy in letters, digits, symbols, length
- Entropy estimation using NIST guidelines [1]

Results

Longer but simpler is harder to guess, easier for users

- 16-char minimum is least guessable for powerful attackers (A)
- First empirical result under controlled study conditions
- Contradicts current popular guidelines for policy strength
- Users prefer longer and simpler to shorter with many requirements
  - Find it less annoying and difficult (B)
  - Need fewer attempts to successfully create a password that meets requirements (C)





Additional findings

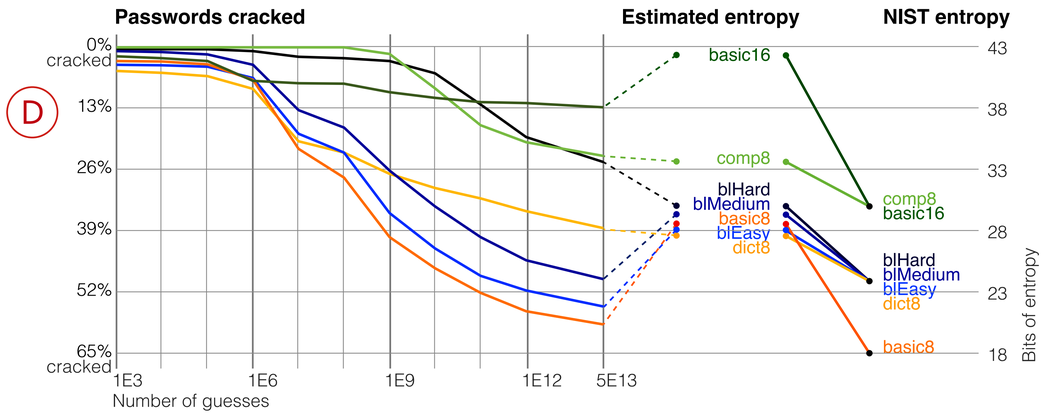
- Adding more training data improves guessing more for stronger conditions (**comp8**, **basic16**) than weaker ones (**basic8**, **blMedium**)
  - Good training data is key to evaluating strength accurately
- Subsets of large password collections that meet a policy's requirements do not represent passwords created under that policy
  - These subsets should not be used to compare policies [4]

C

Condition	Mean creation attempts
basic8	1.10
blEasy	1.17
blMedium	1.36
basic16	1.57
dict8	1.79
blHard	1.98
comp8	3.07

Entropy provides a rough approximation of guessability

- NIST entropy provides a rough ordering of policies by guessability (D)
- Empirical entropy is more accurate than NIST entropy (D)
- Entropy does not accurately reflect the magnitude of guessability differences (D)



Ongoing work

- Comparing online data with real CMU passwords
- Testing more points in the policy space

References

1. W. Burr, D. Dodson, and W. Polk. Electronic authentication guideline. Technical report, NIST, 2006.
2. S. Komanduri, R. Shay, P. Kelley, M. Mazurek, L. Bauer, N. Christin, L. Cranor, and S. Egelman. Of Passwords and people: Measuring the effect of password-composition policies. CHI, 2011.
3. R. Shay, S. Komanduri, P. Kelley, P. Leon, M. Mazurek, L. Bauer, N. Christin, and L. Cranor. Encountering stronger password requirements: user attitudes and behaviors. SOUPS, 2010.
4. M. Weir, S. Aggarwal, M. Collins, and H. Stern. Testing metrics for password creation policies by attacking large sets of revealed passwords. CCS, 2010.

Carnegie Mellon CyLab
CONFIDENCE FOR A NETWORKED WORLD