Potential Reuse of University Credentials

Ashley Bochner, Jacob Abbott, L. Jean Camp
Luddy School of Informatics, Computing, and Engineering
Indiana University Bloomington

Introduction

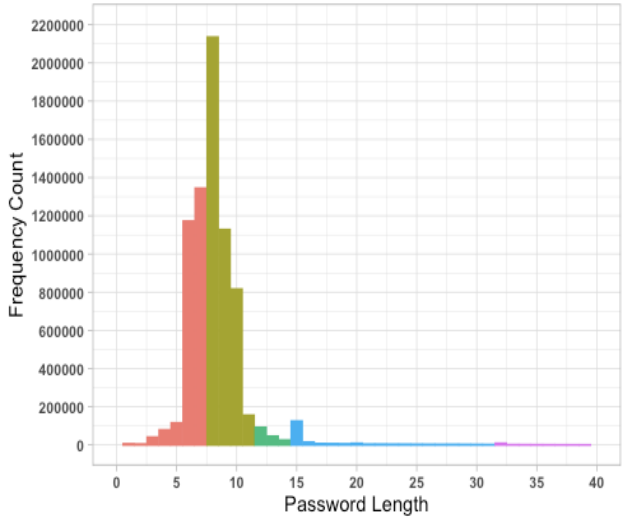
This project looked at the characteristics of compromised login credential records, specifically focusing on passwords. We looked into the re-use of login credentials, the passwords that were likely machine generated, password length, and complexity.

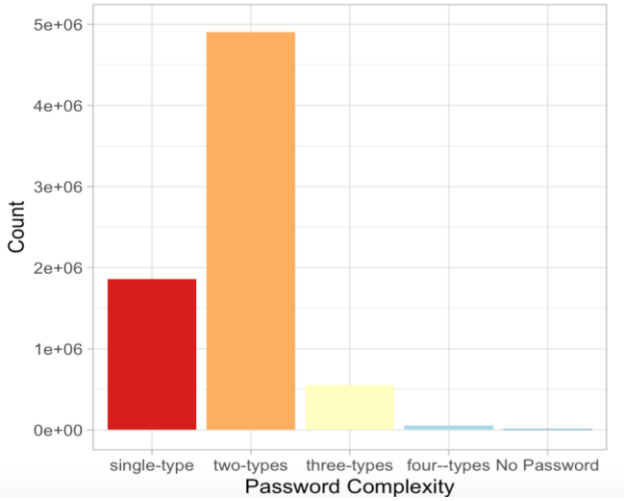
# Characters	Assumed Human	Assumed Machine
<8	2,511,249	259,883
8 - 11	4,120,957	111,415
12 - 14	161,512	1,555
14 - 31	85,835	115,189
32+	8,463	7,223

	Usernames	Passwords	Domains
Unique	3,566,905	3,193,985	61,452
Duplicate	3,817,376	4,190,296	7,322,829

Objective

Analyze a set of 7.38 million username, email, and password combinations from users in the U.S. higher education system to find potential risks of re-used credentials.





Method

The data was pulled from public disclosure of two large datasets, Exploit.in and Anti-Public. Any duplicates between email addresses, password, and username combinations were found. The information was analyzed to find if the passwords contained machine generated characteristics (number, special character, only alphas a-f) to predict potentially machine generated credentials. Password complexity was based on the amount of types of characters in a password: lowercase alphas, uppercase alphas, numbers, and special characters. If a password contained only one type it was categorized as "single-type".

Conclusion

Over 33% of passwords in the data-set would not pass the weakest university's login credential policies. Login credentials have also been re-used multiple times causing security concerns. There was a correlation between password lengths being greater than or equal to 32 characters and likely machine generated passwords, indicating that even strong passwords from users cannot protect from system weaknesses and exploits.

This research was supported in part by the National Science Foundation awards CNS 1565375 and CNS 1814518, the National Security Agency #H8230-19-1-0310, Cisco Research Support, Google Research, and the Comcast Innovation Fund. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US Government, the National Science Foundation, the National Security Agency, Cisco, Comcast, Google, nor Indiana University.