## Investigating Feasibility of Distribution-Aware Password Lockout Mechanisms Against Online Password Guessing Attacks

Large-scale online password guessing attacks are wide-spread and continuously qualified as one of the top cyber-security risks. Currently, the common method for mitigating the risk of online cracking is to lock out the user after a fixed number of consecutive incorrect login attempts but this approach leads to user annoyance since oftentimes honest users are locked out after a few mistakes.

Motivated by the observation that honest user mistakes typically look quite different than the password guesses of an online attacker, we introduce the notion of a password distribution aware lockout mechanism to reduce user annoyance while minimizing user risk.

As the first step towards implementation of such mechanism, we have investigated its security and usability consequences on the artificial dataset: we generated a synthetic dataset to model honest user logins in the presence of an online attacker. The data for the password distribution was taken from the long-known breached databases such as RockYou and frequency dataset such as Yahoo! We maintain a ``hit count" for each user which is based on (estimates of) the cumulative probability of all login attempts for that particular account, and a user is only locked out when this hit count is too high. The preliminary results of our analysis on this synthetic dataset strongly support our hypothesis that distribution aware lockout mechanisms can reduce user annoyance and risk of being compromised.

However, popular passwords change over time, hence using a static dataset for computing password frequency might not be efficient. For example, in early 2000s, “spiderman” was in the top of most popular passwords, while in the recent years “starwars” has taken its place after the latest movie release. We use the count sketch probabilistic data structure which estimates the frequency of an item in a stream to approximate frequency of a password. To minimize user risk, we add a differential privacy component to the count sketch data structure by covering the actual frequency estimates with the Laplacian noise. This approach ensures that the system security is now decreased by introducing a new data structure.

Initial results of using of the differentially private count sketch data structure with the synthetic dataset demonstrate that there can be found a hit count parameter value which allows to have less percentage of annoyed users than in the system without attack and less compromised users than in the system under persistent online attack at the same time.

Our next step is to test this lockout mechanism on real life data. One option is to collaborate with the CoreLogic group in differentially private manner and assess the efficiency of the distribution-aware password lockout mechanism on the recent 70-million password dataset that they possess. Another alternative is to collaborate with an interested corporation which in return will benefit by decreasing the risks of compromising its users and locking them out instead of giving them one more login attempt. We plan to set up a differentially private count sketch data structure using this dataset and implement a distribution-aware lockout mechanism in the real life setting of an enterprise. The existing implementation of an enterprise and our mechanism can be run at the same time for a trial period during which the optimal hit count parameter can be determined. After the trial period the results will be evaluated and the recommendation about the adoption of the new approach can be formulated.