Personalized Password Guessing: a New Security Threat

Phuong Cao, Hongyang Li
Klara Nahrstedt, Zbigniew Kalbarczyk,
Ravishankar Iyer
Coordinated Science Laboratory
University of Illinois at Urbana Champaign
{pcao3,hli52,klara,kalbarcz,iyer}@illinois.edu

Adam J. Slagell
National Center for Supercomputing Applications
University of Illinois at Urbana Champaign
slagell@illinois.edu

ABSTRACT

This paper presents a model for generating personalized passwords (i.e., passwords based on user and service profile). A user's password is generated from a list of personalized words, each word is drawn from a topic relating to a user and the service in use. The proposed model can be applied to: (i) assess the strength of a password (i.e., determine how many guesses are used to crack the password), and (ii) generate secure (i.e., contains digits, special characters, or capitalized characters) yet easy to memorize passwords.

Categories and Subject Descriptors

D.4.6 [**Operating Systems**]: Security and Protection

General Terms

security, algorithm, graphical model

Keywords

guessing, password, personalized, suggestion

1. INTRODUCTION

Password is the most common security mechanism for authenticating users in critical systems and online services. Personal emails, banking accounts, mission critical infrastructures such as super computers, power grids or military networks are protected from unauthorized access by password and other mechanisms. However, users often choose dictionary-based, common, or reuse passwords making them easy to guess and putting their accounts at risks [1].

We address a new targeted attack that aims to guess an individual's passwords from collected information of a user (i.e., *user profile*) and characteristics of the service (i.e., *service profile*) that the user uses. A user profile may include user name, hobbies (e.g., football), and occupation (e.g., student); a service profile may include the type of service (e.g., university email) and the name of the service provider (e.g. University of Illinois). User profile can be retrieved from social media (e.g., Twitter or Newsle) or by social engineering techniques [3]. Current password crackers rely on dictionaries or rules to guess passwords without taking into account the user profile and the service profile. In addition, organizations require strict policies and periodically changing passwords, that makes choosing a good password a frustrating experience for the users. Our study aims to measure the potential of guessing passwords using user and service profile. We examine whether by taking into account the user and the service profile, a password cracker can significantly reduce the number of password guesses needed to reveal the password. As a side effect, we attempt to suggest secure (i.e., contains digits, special characters, or capitalized characters), personalized (i.e., based on user and service profile) yet easy to memorize passwords for users.

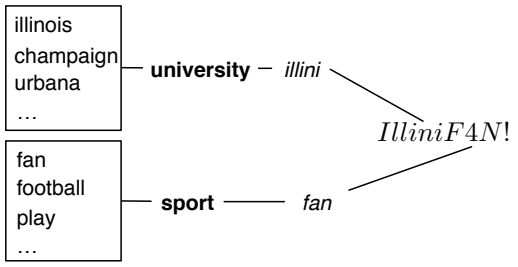

Figure 1: An example process of generating a password. A user who likes sport and studies at University of Illinois chooses two topics for the password: *study* and *university*. Two personalized words *illini* and *fan* are selected from the topics. The words are transformed using grammatical and mangling rules to the final password *IlliniF4N!*

This paper presents a personalized password generation model for a user. User profiles and service profile are examined to find the link between them and the chosen passwords. We assume that a user generates a password based on a set of personalized words. Each word is selected from a list of topics based on the user profile and the service profile. Figure 1 shows an example password: $IlliniF4N!$ [1], which consists of two personalized words $illini$ and fan. The user profile topic is $sport$, which contains several words such as fan, $football$ or $play$. The service profile topic is $university$. Assuming we know the password is for the University of Illinois, the $university$ topic can contain several words such

[1]Illinois football team

HotSoS '14, April 08 - 09 2014, Raleigh, NC, USA

http://dx.doi.org/10.1145/2600176.2600198.

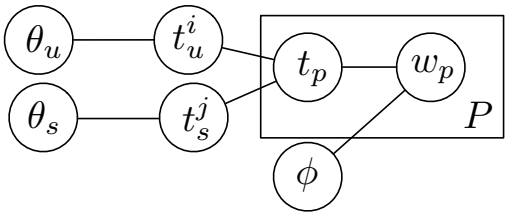

Figure 2: Generating P personalized words. A word w_p is selected from a topic t_p. The topic is based on a list of user profile topics t_u and a list of service topics t_s. A topic distribution is parameterized by θ and a word-topic distribution is parameterized by ϕ.

as *illinois*, *urbana*, or *champaign*. The personalized words are combined using *grammatical* and *mangling* rules (e.g., l33t rule or adding special characters) to devise the final passwords. *Topic modeling* is used to capture probabilistic relationships between the password and the topics [2].

2. APPROACH

In this paper, we consider following problems:

Password guessing: given a user profile and a service profile, guess the chosen password using our personalized password generation model.

Passwords suggestion: given a user profile and a service profile, suggest secure yet easy to memorize passwords.

Given a user profile and a service profile, we extract the list of topics and its related words. Each topic is associated with a weight determining the importance of the topic. A topic importance can be identified by surveying the user or by comparing the amount of information collected about the topic to other topics. This is topic distribution. Related words of a topic is collected from books, news or Wikipedia articles about the topic. Each word is associated with a frequency count, determining the importance of the word to the topic. This is word-topic distribution. Topic modeling is used to model the password generation process as follows [2].

Personalized words generation. Figure 2 illustrates the procedure of selecting P personalized words w_p relating to the user profile u and the service profile s. A word w_p is selected from a topic t_p, which is a user profile topics t_u^i or a service topic t_s^j. Each topic has a multinomial topic distribution θ_u and θ_s defining topic weights. Given the topic t_p, the word w_p is selected according to the multinomial word-topic distribution ϕ. The topics t_p and the words w_p are enclosed in a box with a notation P at the bottom right. The notation specifies the number of repetitions for the topics and the words, i.e., P words are selected from P topics. For example, a user can select two words *illini* and *fan* based on two topics: university and sport ($P = 2$) (see Figure 1).

Personalized password generation. Figure 3 illustrates the procedure of generating passwords from a list of P personalized words. The words are transformed and combined using a number of rules to create a personalized password that is relevant to the user profile and the service profile. For example, a word can be capitalized (e.g., illini → Illini) or can be written in a *l33t* style and capitalized (e.g., fan → F4N) (see Figure 1).

In summary, a personalized password can be generated as follows: select P topics from the user profile topics and the service topics, draw P words from the P topics according to the word-topic distribution ϕ, and apply grammatical/mangling rules.

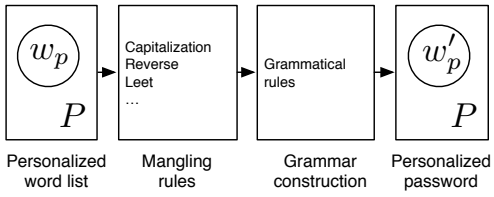

Figure 3: Generating a personalized password from P personalized words based on grammatical and mangling rules.

Using the described procedures to generate passwords, we expect that a less number of guesses will be required to obtain an exact match with the user password compared with brute-force approaches. Furthermore, the same procedure can be used to generate *personalized* passwords which are more secure while relatively easy to memorize.

3. EXPERIMENTAL DESIGN

We have designed and plan to conduct two user studies.

Personalized password guessing. We design a questionnaire form to: (i) collect basic user information (e.g., name, speaking language, or hobbies), (ii) ask a user for his/her public social media profile (e.g., Twitter, LinkedIn, or Facebook), and (iii) ask the user to generate passwords for a set of web services (e.g., email, online shopping, or social network). Our approach uses the information collected at step (i) and (ii) to guess the password at the step (iii). We measure accuracy in terms of the number of correct password guesses. In addition, the average edit distance can be quantified to measure how close the guessed password to the password provided by the user.

Personalized password generation. We design an enhanced sign up form of web services that: suggests and measures strength of the chosen password based on the user profile and the service profile. The sign up form can suggest a password and explain the connections of the password to the user. In case the user enters a password manually, we can measure strength of the chosen password by comparing the words in the password with the words in the user profile topics or the service profile topics.

In both case, a user survey will be performed using Amazon Mechanical Turk or using a Facebook application.

4. CONCLUSION

We presented a model to automate generation of personalized passwords based on user and service profiles. The proposed approach has a potential to provide means for assessing how easy it is to guess a password, and can be used to generate stronger yet easy to memorize password.

5. REFERENCES

[1] Das, A., Bonneau, J., Caesar, M., Borisov, N., and Wang, X. The Tangled Web of Password Reuse. In *NDSS* (2014).

[2] Koller, D., and Friedman, N. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[3] Yan, J. J., Blackwell, A. F., Anderson, R. J., and Grant, A. Password memorability and security: Empirical results. *IEEE Security & privacy 2*, 5 (2004), 25–31.