International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2015): 6.391

Shadow Attacks Based on Password Patterns Password Reuses

Bhavika Garase[1], N. D. Kale[2]

[1, 2]Computer Engineering Department, PVPIT, Pune, Maharashtra, India

Abstract: ***As the number of websites is increasing, the security level of password enabled accounts is no longer secured. It may possible that end users can create the many accounts on either same website or multiple websites, hence the passwords for all accounts from the same user likely to similar or same. This leads the users account can compromised by attackers and then identify the same user passwords in sensitive accounts like banking accounts. This type of attacks on password is basically known as shadow attacks on password. In this project, we are presenting the framework to study the state of art of Cross Site Password Reused (CSPR) and Intra Site Password Reuses (ISPR) based on large scale password datasets in order to improve the password reuse success rate. The novelty of the proposed approach is detailed empirical study on web password reuses of both cross site and intra site password reuses. The pre-processing step is redesigned in our proposed work in which noise removal is done in terms of invalid email addresses, spaces etc. rather removing the duplicate accounts. This will helps to improve the guessing success rate performance as compared to previous work.***

Keywords: Logo detection, Context-dependent kernel, recognition of Context-dependent kernel, CDS, SURF

1. Introduction

Password-based authentication is one in every of the foremost wide used methods to demonstrate a user before granting accesses to secured websites. The wide adoption of password-based authentication is that the results of its low value and simplicity: a user will enter his or her passwords anyplace by a keyboard or barely screen with none alternative additional devices. The popularity of passwords and therefore the proliferation of websites, however, result in a priority on password reuses between accounts on totally different websites or perhaps on identical web-sites. Moreover, the recent various high-pro le password leakage events didn't build the password scenario higher, and that we raise the questions: What do password reuses mean to accounts between web sites and even those among identical websites? What's the implication of a compromised website or account to others? However simple are shadow attacks, i.e., an someone compromises an ac-count utilizing the passwords of alternative accounts that are either on a similar web site or from alternative sites? To search out the answers, during this paper we tend to analyze password reuses and shadow attacks by trial and error.

It is well-known that passwords are usually reused by a user across different websites, yet little work has been devoted to understanding passwords being shared among multiple accounts of the same user on the same website. Since both pass-word reuses within the same website and across multiple ones can enable shadow attacks, in this paper, we analyze the both scenarios: (i) a user creates accounts with the same password on the same websites, which we term as Intra-Site Pass-word Reuses (ISPR), and (ii) a user creates accounts with the same password across different websites, which we term as Cross-Site Password Reuses (CSPR). While having the same passwords for multiple accounts is simple and convenient to users, it raises security concerns, e.g., if a password on one website is leaked, an adversary can have an enhanced chance to crack the other accounts of the same user, regardless of whether the accounts are on the same or different websites. We note that account ownership can be identified by the registered email addresses. As a result, we argue that users account with passwords of higher security level could be relatively easily compromised, given the knowledge of the passwords at a lower security level, e.g., web forums.

The continued domination of passwords over all other methods of end-user authentication is a major embarrassment to security researchers. As web technology moves ahead by leaps and bounds in other areas, passwords stubbornly survive and reproduce with every new web site. Extensive discussions of alternative authentication schemes have produced no definitive answers. Over forty years of research have demonstrated that passwords are plagued by security problems and openly hated by users. We believe that, to make progress, the community must better systematize the knowledge that we have regarding both passwords and their alternatives. However, among other challenges, unbiased evaluation of password replacement schemes is complicated by the diverse interests of various communities. In our experience, security experts focus more on security but less on usability and practical issues related to deployment; biometrics experts focus on analysis of false negatives and naturally-occurring false positives rather than on attacks by an intelligent, adaptive adversary; usability experts tend to be optimistic about security; and originators of a scheme, whatever their background, downplay or ignore benefits that their scheme doesn't attempt to provide, thus overlooking dimensions on which it fares poorly. As proponents assert the superiority of their schemes, their objective functions are often not explicitly stated and differ substantially from those of potential adopters. Targeting different authentication problems using different criteria, some address very specific environments and narrow scenarios; others silently seek generic solutions that all environments at once, assuming a single choice is mandatory. As such, consensus is unlikely. These and other factors have contributed to a longstanding lack of progress on how best to evaluate and compare authentication proposals intended for practical use. In response, we propose a standard benchmark and framework allowing schemes to be rated across a common, broad

spectrum of criteria chosen objectively for relevance in wide ranging scenarios, without hidden agenda

2. Literature Survey

R. Morris and K. Thompson, Password security: A case history, Communications of the ACM, vol. 22(11), pp. 594597, 1979: This paper describes the history of the design of the password security scheme on a remotely accessed time-sharing system. The present design was the result of countering observed attempts to penetrate the system. The result is a compromise be-tween extreme security and ease of use.

A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang, The tangled web of password reuse, in NDSS2014, 2014: We study several hundred thousand leaked passwords from eleven web sites and conduct a user survey on password reuse; we estimate that 43-51 % of users reuse the same password across multiple sites. We further identify a few simple tricks users often employ to transform a basic password between sites which can be used by an attacker to make password guessing vastly easier. We develop the first cross-site password-guessing algorithm, which is able to guess 30 % of transformed passwords within 100 attempts compared to just 14 % for a standard password-guessing algorithm without cross-site password knowledge.

J. Bonneau, The science of guessing: Analyzing an anonymized corpus of 70 million passwords, in 2012 IEEE Symposium on Security and Privacy (SP), 2012, pp. 538552.: Author report on the largest corpus of user-chosen passwords ever studied, consisting of anonymized password histograms rep-resenting almost 70 million Yahoo! users, mitigating privacy concerns while enabling analysis of dozens of subpopulations based on demographic factors and site usage characteristics. This large data set motivates a thorough statistical treatment of estimating guessing difficulty by sampling from a secret distribution. In place of previously used metrics such as Shannon entropy and guessing entropy, which cannot be estimated with any realistically sized sample, we develop partial guessing metrics including a new variant of guesswork parameterized by an attacker's desired success rate. Authors new metric is comparatively easy to approximate and directly relevant for security engineering. By comparing password distributions with a uniform distribution which would provide equivalent security against different forms of guessing attack, we estimate that passwords provide fewer than 10 bits of security against an online, trawling attack, and only about 20 bits of security against an optimal offline dictionary attack. Author and surprisingly little variation in guessing difficulty; every identifiable group of users generated a comparably weak password distribution. Security motivations such as the registration of a payment card have no greater impact than demo-graphic factors such as age and nationality. Even proactive e orts to nudge users towards better password choices with graphical feedback make little difference. More surprisingly, even seemingly distant language communities choose the same weak passwords and an attacker never gains more than a factor of 2 efficiency gain by switching from the globally optimal dictionary to a population-specific lists.

J. Ma, W. Yang, M. Luo, and N. LI, A study of probabilistic password models, in Proceedings of IEEE Symposium on Security & Privacy, 2014: A probabilistic password model assigns a probability value to each string. Such models are useful for research into understanding what makes users choose more (or less) secure passwords, and for constructing password strength meters and password cracking utilities. Guess number graphs generated from password models are a widely used method in password research. In this paper, we show that probability-threshold graphs have important advantages over guess-number graphs. They are much faster to compute, and at the same time provide information beyond what is feasible in guess-number graphs. We also observe that research in password modeling can bene t from the extensive literature in statistical language modeling. We conduct a systematic evaluation of a large number of probabilistic password models, including Markov models using different normalization and smoothing methods, and found that, among other things, Markov models, when done correctly, perform significantly better than the Probabilistic Context-Free Grammar model proposed in Weir et al., which has been used as the state-of-the-art password model in recent research

3. Proposed Approach Framework and Design

3.1 Architecture

For secured websites, the password based authentication is frequently used approach for authenticating the end user before granting the access. The growing use of password based authentication approach at increasing websites leads to the important issue of possibility of password reuses among accounts of various web-sites or similar websites. Additionally, the recent study on numerous high pro le password hacking claims that password situation is not better. Under such cases, there is huge possibility of shadow attacks in which an attacker can successfully compromise the account that reuses the password of other accounts those are from similar website or different websites. The reuse of passwords for different accounts under same website is called as Intra-Site Password Reuses (ISPR). The reuse of passwords for different accounts under different websites is called as Cross-Site Password Reuses (CSPR). Therefore in order to prevent such shadow attacks on passwords, first we need to understand and examine the both ISPR and CSPR based on publically available password datasets. However, there is no in-depth empirical study conducted in literature except the one very recently introduced on Chinese password datasets. However the problem with this method is that they are removing the duplicate pro les and passwords largely in their pre-processing step, this can reduces the scalability of password reuses.

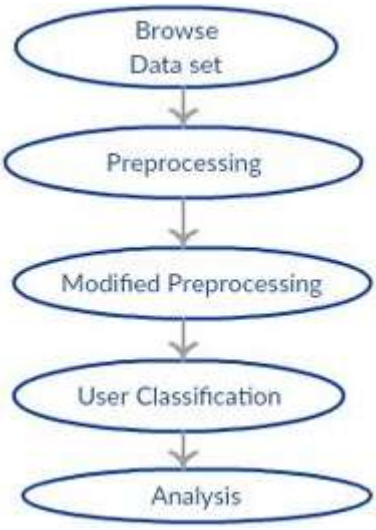

Figure 1: Proposed System Architecture

3.2 Mathematical Model:

Input Set
X- Password datasets.

NP-Complete
In computational complexity theory, a decision problem is NP-complete when it is both in NP and NP-hard. The set of NP-complete problems is often denoted by NP-C or NPC. Below figure 3.2 is indicating the exact meaning of NP-complete and NP-hard problems in this research.

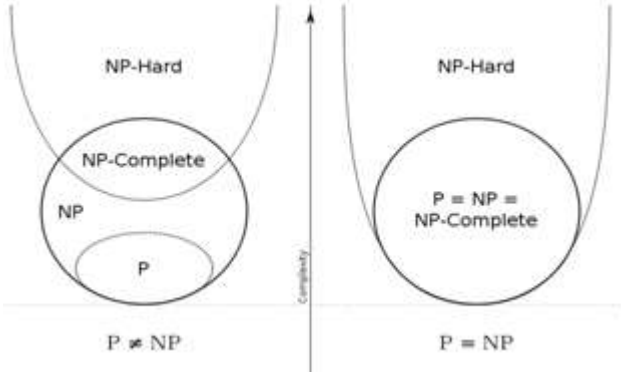

Figure 2: NP-complete and NP-hard problems

In our research, NP complete problem are presenting the framework to study the state of art of Cross Site Password Reused (CSPR) and Intra Site Password Reuses (ISPR) based on large scale password datasets in order to improve the password reuse success rate. The novelty of the proposed approach is detailed empirical study on web password reuses of both cross site and intra site password reuses. Whereas NP-hard problem is the proposed work which is proposed to extend NP-complete problem. In NP-hard problem, our proposed modified pre-processing method is replacing pre-processing method of NP-complete problem with goal of improving accuracy of detection.

Below algorithms used for NP-complete problem:

Below figure 3.3 is showing the existing feature extraction method based epileptic detection framework by considering transmitter and receiver components.

Algorithm:
Step 1: Browse and Load Password Dataset
Step 2: Remove accounts with blank passwords
Step 3: Remove accounts with invalid email addresses, Remove duplication accounts
Step 4: Generate pre-processed password data and store into mysql.
Step 5: Apply User Classification in three types and Divide passwords into six different sets
Step 6: Rate of password reuses
Step 7: Password strength analysis and Keyboard Patterns analysis
Step 8: Web password reuses in different user groups
Step 9: Measure performance metrics resistance to guessing, reuse rate, password pattern rates etc for existing work.
Min-entropy, H∞

$$H_{\infty} = -\log_2(p)$$

Marginal success rate or β-success rate,

$$\tilde{\lambda}_{\beta} = \log_2(\frac{\beta}{\lambda_{\beta}})$$

Guesswork G

$$\tilde{G} = \log_2(2.G - 1)$$

α.-guesswork

$$\tilde{G}_{\alpha} = \log_2(\frac{2.G_{\alpha} - 1}{\lambda_{\mu_{\alpha}}}) + \log_2(\frac{1}{2 - \lambda_{\mu_{\alpha}}})$$

Limitations of DWT
- Quality and number of datasets
- Mapping between persons and users

NP-Hard Problem
This section deals with proposed contribution method to overcome the limitations of existing methods of modified pre-processing.

Step 1: Escaping HTML characters
Step 2: Decoding data
Step 3: Apostrophe Lookup
Step 4: Removal of Stop-words
Step 5: Removal of Punctuations
Step 6: Removal of Expressions
Step 7: Split Attached Words
Step 8: Slangs lookup
Step 9: Standardizing words
Step 10: Removal of URLs

Output Set:
performance metrics resistance to
- guessing,
- reuse rate,
- password pattern rates, etc.
- Accuracy and parameters.

4. Work Done

In this section we are discussing the practical environment, scenarios, performance metrics used etc.

4.1 Input

In this Training and Testing Image is the input for our practical experiment.

4.2 Hardware Requirements:

Processor : Pentium IV 2.6 Ghz
Ram : 512 Mb
Hard Disk : 20 Gb

4.3 Software Requirements:

Front End : J2SE
Back End : MySql 5.1
Tools Used : Net Beans 7.2.1 or above
Operating System : Windows 7/8

4.4 Results of Practical Work**:**

Following figures are showing results for practical work which is done. Following figure showing the main screen. That takes the input data set,

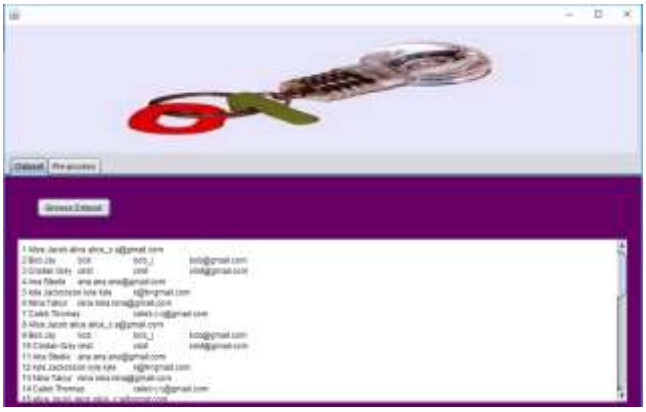

Figure 1: Browse The Input Dataset To perform analysis

Figure 2: Pre-processed Data

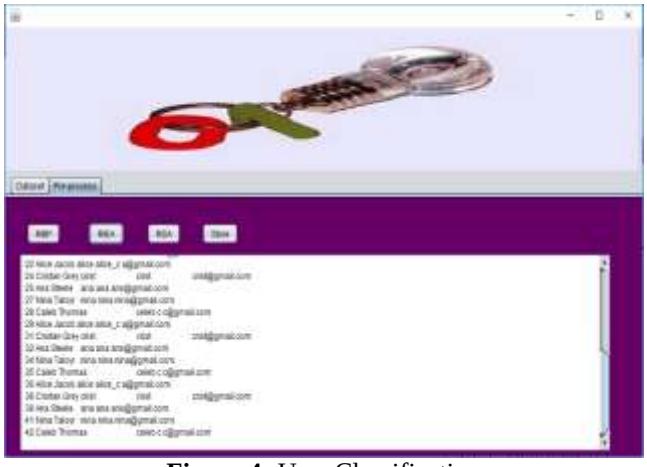

Figure 4: User Classification

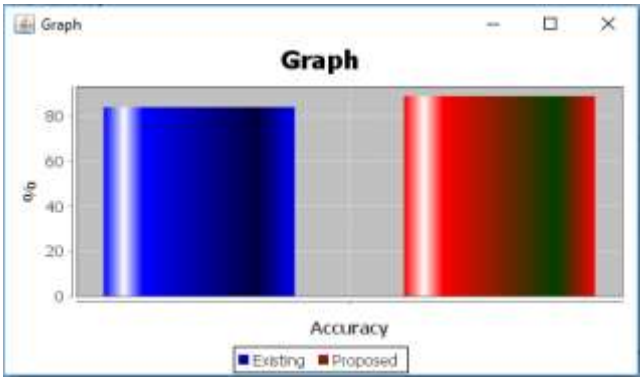

Figure 5: Accuracy Comparison Graph Between Existing and Proposed Method

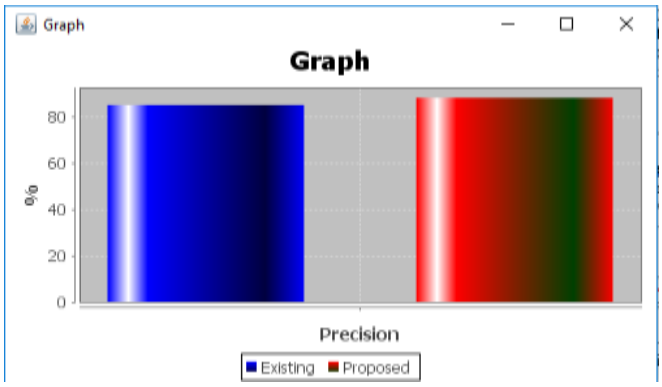

Figure 6: Precision Comparison Graph Between Existing and Proposed Method

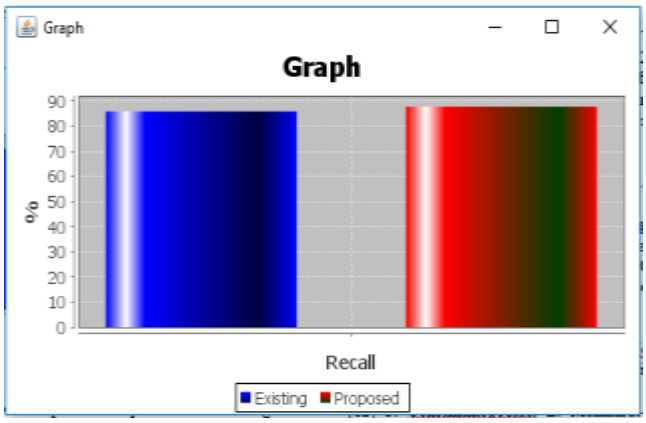

Figure 6: Recall Comparison Graph Between Existing and Proposed Method

5. Conclusion and Future Work

To the best of our knowledge, this is the first empirical study on web password reuses by analyzing a large number of sample data. Although the web password reuses are known to researchers and Internet users, it is yet to perform a large-scale empirical study. We obtained 2,671,443 distinct users each of whom has at least two accounts from the same site, and 2,306,055 distinct users each of whom had at least two accounts from different websites. We also obtained 350,849 distinct users who has at least two accounts on the same site and across sites simultaneously. The quantitative answers shed lights on the serious threat of web password reuses, i.e., password shadow attacks, where an adversary may attack an account of a user using the same or similar passwords of his/her other less sensitive accounts. As a future direction, we would study CSPR from both adversaries' and defenders' points of view, leveraging the logs or activities that are available in the public domain. In addition, we will evaluate how the password policies affect CSPR after understanding the policies of these four websites. Last but not the least, we plan to study the impact of single sign-on tools on password reuses.

References

[1] R. Morris and K. Thompson, "Password security: A case history," Communications of the ACM, vol. 22(11), pp. 594–597, 1979.
[2] A. Das, J. Bonneau, M. Caesar, N. Borisov, and X. Wang, "The tangled web of password reuse," in NDSS'2014, 2014.
[3] D. Florencio and C. Herley, "A large-scale study of web password habits," in WWW'07 roceedings of the 16th international conference on World Wide Web, 2007, pp. 657–666.
[4] CSDN, "http://www.csdn.net/company/about.html."
[5] Tianya, "http://help.tianya.cn/about/history/2011/06/02/166666.shtml."
[6] Duduniu, "http://baike.baidu.com/view/1557125.htm."
[7] 7k7k, "http://www.7k7k.com/html/about.htm."
[8] J. Bonneau, "The science of guessing: Analyzing an anonymized corpus of 70 million passwords," in 2012 IEEE Symposium on Security and Privacy (SP), 2012, pp. 538–552.
[9] J. Ma, W. Yang, M. Luo, and N. LI, "A study of probabilistic password models," in Proceedings of IEEE Symposium on Security & Privacy, 2014.
[10] Z. Li, W. Han, and W. Xu, "A large-scale empirical analysis of chinese web passwords," in 23rd Usenix Security Symposium. San Diego: USENIX, 2014.
[11] W. Han, Z. Li, L. Yuan, and W. Xu, "Regional patterns and vulnerability analysis of chinese web passwords," IEEE Transactions on Information Forensics and Security, vol. 11, no. 2, pp. 258–272, 2016.
[12] D. Wang, H. Cheng, Q. Gu, and P. Wang, Understanding passwords of chinese users:characteristics, security and implications," https://www.researchgate.net/, July 2014.
[13] D. Schweitzer, J. Boleng, C. Hughes, and L. Murphy, "Visualizing keyboard pattern passwords," in Visualization for Cyber Security, 2009. VizSec 2009. 6th International Workshop on. IEEE, 2009, pp. 69–73.
[14] Wikipedia, "Levenshtein distance," http://en.wikipedia.org/wiki/Levenshtein distance, May 2014.
[15] ——, "Longest common subsequence problem," http://en.wikipedia.org/wiki/Longest common subsequence, May 2014.