Elysium Journal of Engineering Research & Management

P-ISSN: 2347-4408
E-ISSN: 2347-4734

Enhancement of Security and Usability Standards with a Novel Storage Optimized Honeyword Generation Approach

Avinash Pandey[1] and Dilip motwani[2]

[1]Lecturer, Department of Information Technology, Mumbai University, VIT College, Mumbai-400037, India.
[2]Professor, Department of Information Technology, Mumbai University, VESIT College
Mumbai – 400 074, India.

Abstract— ***Among the modern security threats on password based authentication techniques, the brute force computation is the one that performs the inversion of hash values. Several technologies have been developed for the computation of brute force in the inversion attack. This threat can be mitigated by detecting the password cracking with the Honeyword based authentication protocol. Though various existing techniques have some limitations such as storage overhead, weak DoS resistivity and multiple system vulnerability. To overcome these existing drawbacks, a novel honeyword generation approach with the decoy data mechanism is proposed in this work. The Paired Distance Protocol is used in this work and implemented for evaluating the proposed technique. The performance of the proposed techniques is compared with the existing techniques and provides better results in the security with reduced storage overhead and storage cost.***

Index Terms— ***Authentication, Password, Inversion attack, Honeyword, Paired distance.***

1. INTRODUCTION

The most commonly used authentication method is the password based authentication method because of its standards in usage and security balance. Though, the password based approaches also have some challenges due to the various attacks as in any other security approaches. Inversion attack is one of the typical attack grown in recent times which are described as follows. The user must submit the username and password for registering in the website during the creation of web-account. The username is recorded by the system in the form of plain text and the password in the form of hash by converting them using hashing algorithm H. Therefore the login credential of i^{th} user that is recorded by the system is denoted using a tuple $< u_i, H(p_i) >$. An attacker can invert the hashes i.e. estimating p_i from $H(p_i)$ magnificently from the compromised password file F in the inversion attack model. During the inversion of hashes, initially the attacker determines a password string with the help of some existing methods. Then after converting it onto a hash value by H, the matching process is carried out for the password string. The attacker is succeeded in inverting the hashes if the determined hash value is matched with the recorded hash value. For cracking a password, brute force attack was initially accompanied by predicting many probable combinations. As the attacker attempts for every possible choices to break a password, this approach becomes a very high time consuming approach. To reduce the complexity in the inversion attack due to time, the most available password breaking algorithm was introduced by John and Ripper in 2008. Later in 2009, Weir et.al. Proposed an approach which has the ability to crack about 28% - 129% more than the John and Ripper's password braking approach. Ma.et.al recently introduced a password cracking method using the model of Markov chain that provides substantial improvement compared to the Wier's approach. There are many security threats that are interrupted in some of the web based organizations. To address this security issue, some security methods were developed. Some tricks are used to make the conversion of password into hash value harder that increases the login time. Also few fake login accounts were created for this by the administrator. The system detects the security break whenever an attacker succeeds in the inversion of hash values for any account. The real and system created usernames are distinguished by attackers using some careful analysis.

The possibility of providing security in contradiction of inversion attack uses honeyword based methods. A password list that contains real users' password together with the system created passwords are maintained in this method which is called as honeywords. The system detects the attack and the necessary actions are taken based on the security policy, while the password file F is compromised and the attacker enters any honeywords from the password list. Some honeyword generation algorithms like take-a-tail, modelling-syntax, etc are used for generating these honeywords by system. Among these methods, the take-a-tail algorithm provides strongest security but it affects the standard of usage to a greater extend.

In this paper a decoy mechanism is proposed for the honeyword password generation algorithm. Fake data is created for the illegal users in this proposed work. Whenever the user tries to hack anyone's account by entering false password, then after some attempt the user will be redirected to the fake account for pretending them as they are succeeded in hacking account.

The remaining sections of this paper are organized as follows: Section II reviews some of the existing works related to the password breaking algorithms. Section III provides an outline of honeyword based authentication technique and its limitations. Section IV presents a detailed description of the proposed methodology. The comparative analysis of the proposed work is described in the Section V. Finally, the paper is concluded in Section VI.

2. RELATED WORK

An extension of basic password cracking method was proposed in [1] to improve the security of the hashed password. This method uses multiple passwords for each account. The attempted honeyword sets off an alarm when an attack was detected. In [2] the breaking of several real time passwords were examined using an empirical study. The effectiveness of the password strength meters and the impacts of the security in passwords were also determined and compared to the existing methods. This proposed method gave a deep knowledge about the vulnerabilities of the real time passwords. The estimation of password guessability, password vulnerabilities by the real time attackers were discussed in [3]. The large password set were analyzed by the single cracking approach in this work and the parameters such as password strength, security impacts were analyzed and compared with the existing approaches. In [4] an effective and practical approach was proposed for detecting the password cracking. Here, a machine dependent function like Hardware Security Module (HSM) or Physically Unclonable Function (PUF) were utilized. The structure of the proposed approach was same as that of the traditional method. The fake passwords were detected when the attacker tried to hash the passwords. This proposed method was used to improve the security of the hash password that are stored. The survey about password cracking research , their technologies and their countermeasures were discussed in [5]. In this proposed work the design measures at two sections like password design state and after the password design state were described. The chief goal of this proposed survey was to provide knowledge about the password cracking and security approaches. In [6], proposed an improved dictionary-based password cracking approach. In this proposed work, the multiword patterns and the keyboard pattern were added to the probabilistic password cracking based context free grammars. This proposed work provided effective results of the password guessing mechanism. In [7], the construction of encrypted vaults were investigated for the resistance of cracking and attackers. In this proposed method a novel secure encoding algorithm called Natural Language Encoder (NLE) was used. This proposed work implemented and evaluated the NLE based cracking resistant vaults. A new multimodal strength metric was introduced in [8] which integrates the various imperfect metrics. This proposed work was implemented in order to overcome the weakness in the password. In this integrated model, the final multimodal metrics were comprised based on the heuristic and statistic method. The paper [9] presented an implementation of Rainbow tables which were the large basic table with hash values. This proposed method was used for the operation of cracking passwords under various applications and discussed the advantages of the proposed method and compared with the existing method. A study of probabilistic password model was examined in [10] in which the probability threshold graph provided more benefits compared to that of the existing guess number graphs. This proposed model was evaluated using Markov model and compared with the existing probabilistic context free grammar model. The performance of this proposed method resulted better than the probabilistic context free grammar model. In [11], an open source analysis and research of the password system was introduced. The comprehensive and uniform research of the password were described in this proposed work. It also offered the better knowledge about the effects of security in the password cracking systems. In [12] 15 password policies were examined that were focused on the length of the password. Several existing policies were not suit for the longer password in the usability standards as well as security issues. Hence this proposed work provided a brief study about the password cracking policies with longer passwords. To protect the passwords from the attacks, a platform for analyzing the password patterns were developed in [13] and their probabilities were also estimated. This proposed method was compared to the existing bruteforce and John-the – Ripper attacks. This proposed method helps to reduce the size of the search space for password. Paper [14], explained about honeyword system and an alternative approach was also proposed for selecting the hone6yword from the existing password. This proposed work provided accurate honeywords with the reduced storage cost. In [15], a Dynamic Password Policy Generator (DPPG) was introduced as an alternative to the existing password checker. The existing checks might leak the informations and enhanced the performance of the attacks. To avoid these problems, this DPPG was proposed in this work. This proposed method contributed a comprehensive security for the password system.

3. OUTLINE OF HONEYWORD BASED AUTHENTICATION APPROACH AND ITS LIMITATIONS

Initially, the working principal of the honeyword based authentication approach is described in this section. Then the limitations of several existing approaches are discussed in this section. Before that some of the symbols that are related to the honeyword approach are expressed in Table 1.

Table 1: Related Symbols

Symbols	Meaning
u_i	i^{th} user in system
p_i	Password of i^{th} user
W_i	Tuple of passwords stored for u_i
k	Number of elements in W_i
c_i	Index of correct password in W_i
$sweetword$	Each element of W_i

A list W_i is maintained in the honeyword generation approach against every username u_i. Another file is used to maintain the index of the correct password in a different system which is also referred to as "honey checker". The core idea of the honeyword generation approach is that, when W_i is compromised and each sweetword is successfully inverted, the attacker also gets confused about the real password because of the distribution of password information over two distinct systems. If the attacker chooses any sweetword and submits it to the user id u_i, then that sweetword index is focused to the "honeychecker". The honeychecker provides a positive feedback if the sweetword gets matched with the correct password index. Else the honeychecker provides a negative feedback to the system administrator. Then according the received feedback, the system administrator takes necessary actions using the security policy. Hence the honeyword based system offers disseminated security that are harder to compromise.

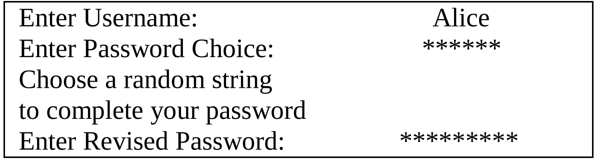
Even though the existing approaches based on the honeyword method provides better security, some limitations are also presented which are as follows:

- Storage overhead
- Co-relational hazard
- Distinct password patterns
- Resistivity of DoS
- Multiple system vulnerability
- Problems due to typo safety

4. PROPOSED METHODOLOGY

The Paired Distance Protocol (PDP) is proposed in this work which offers three significant information such as (a) Username (b) Password and (c) a Random String RS with the length of ℓ that contains alphabets and numbers. Generally the length of RS is set to 3 as default length. The user needs to remember the secret information as RS along with the password. Even though it leads to an overhead, the RS has numerous advantages. The important characteristics of RS are as follows. (i) The same RS can be used for various systems. (ii) The RS that are chosen by the user are hard to guess and doesnot has a particular pattern. (iii) Does not requires correlation between the username and password.

Using PDP the registration interface can be expressed in the Fig. 1.



Enter Username: Alice
Enter Password Choice: ******
Choose a random string
to complete your password
Enter Revised Password: *********

Fig. 1: Registration interface of PDP

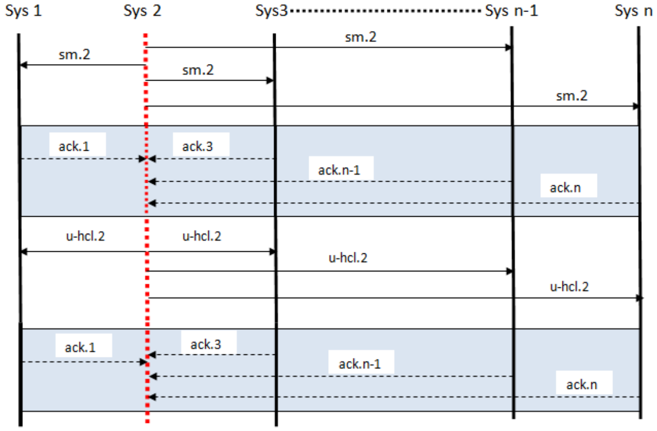


Fig. 2: PDP protocol used by n different users

According to the usability standards, the basic differences between the proposed approach and the existing take-a-tail approach is represented in Table.2

Table.2 Difference between a take-a-tail and PDP from the usability perspective.

Take-a-tail	PDP
The additional informations that are generated by system are remembered by the user	The additional information of user's own choice are remembered by them
The user must remember n information for n different accounts	The user may remember single information for n different accounts

4.1. Honey Circular List

A circular list is created which is referred to a honey circular list or hcl with the length of |hcl|. This honey circular list contains alphabets and digits randomly. Here the default

|hcl| value is considered as 36 and for that default value the |hcl| is shown in Fig. 3.

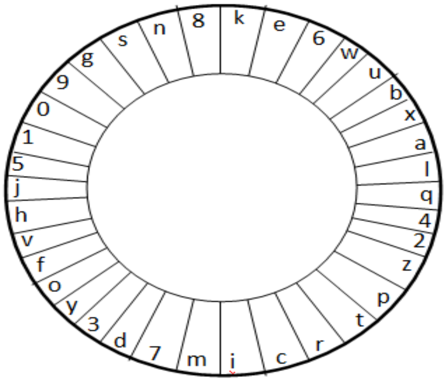


Fig. 3. Honey Circular List

Then this hcl is distributed securely to m distinct systems by PDP protocol for the creation of honeywords and maintained in the password file F.

4.2. Maintenance of user database

For maintaining the login information of the user in database the system follows some tricks. Initially the system records the username and password of the user. Then the distance between the successive elements of RS regarding the elements recorded in hcl are measured by the system. The distance between any two elements are called as paired distance. It is defined as follows.

Let e_1 and e_2 be the two elements of hcl and the paired distance between these elements is the number of cells in the hcl to be negotiated in clock wise direction from element e_1 to element e_2. This paired distance between the two elements are represented as $Pr(e_1, e_2)$.

The system maintains the distance chain in a password file F along with the username and password and that distance chain is obtained from RS. This distance chain can be defined as the set of n-1 paired distances between every two successive elements of RS with length n. A special property of distance chain is found while analyzing it and this special property is identified as uniqueness property. The uniqueness property of a distance chain can be determined by a given hcl and a specific distance chain. If the first element of RS is known, then it can be uniquely derived. By using the RS string and the hcl, the honeywords are created.

From the distance chain that are recorded in the system, the attacker can able to derive the various possible strings that also contains the RS which is chosen by user. For creating the distance chain, the used password can be used in the place of Rs which make the attacker to distinguish the original password of the user. This is because random arrangement of the characters in hcl and leads to a very less probability of obtaining the string.

4.3. Working principle of the proposed mechanism

The implementation of the proposed decoy data to the system is described as follows and the algorithm for this proposed method is also discussed here.

Administrator of the system is an authorized person to add the decoy data in system. By logging into the account, the administrator may add decoy data to be used by system in case of unauthenticated user trying to access the users' account. The administrator should upload data for all kinds of extensions which system can allow.

Steps involved in the proposed algorithm

```
Counting the attempts of client at the time of logging.
If [count of logging exceeds particular threshold value with
    false password]
Then
    Provide successful login to that user showing that
    particular users data.
        If [user tries to access that data]
        Then
            User will get data with the content uploaded
            by admin to the system. //Here the user will be
            impression of accessing particular user
            account data
Else
    User will be prohibited from login and prompted for
    wrong credentials.
```

4.4. Probability of detecting the attacks

PDP stores a single extra information as distance chain instead of storing k-1 extra information. There are |hcl| number of possible RS that corresponds to the distance chain. By storing a single information, the system confuses the attacker among various possibilities. For the default value of |hcl| as mentioned before, the probability of detecting the attcaks will be obtained as 97%.

4.5. Password meter

The password meter helps to show the randomness of RS. The password meter shows strong signal when the randomness of RS is high. Otherwise, it shows weak signal. Some examples of choices of RS for low randomness is described as follows:

- If RS makes some dictionary word, it is concatenated with user password. For e.g. password – rab, RS – bit.
- If RS the aforementioned is a dictionary word. For e.g. fox.
- RS has a particular pattern like sequential keystroke that are differentiated by attacker.

If the password shows low randomness, then the users are recommended to change their RS. Also the randomness of RS becomes when there is a co-relation between the username and password. But it is not addressed by the password meter.

5. COMPARATIVE ANALYSIS

The proposed PDP is compared with some recent honeyword generation approaches like take-a-tail, chaffing-by-tweaking-digits and modelling-syntax-approach in terms of standards of usage and security. There are three well-defined parameters like flatness, DoS resiliency and security against MSV are used for the evaluation of robustness in honeyword generation approaches. Using these security standards the strength of the proposed PDP is estimated. Similarly, for the evaluation of usability, three parameters such as system interference, memorability stress and typo safety are used for in the honeyword generation approaches. The comparison between the proposed and existing techniques are described as shown in table 3.

From the table, it is observed that the user can gain the security as same as that of the take-tail for providing security in regard to MSV and flatness. Take-a-tail has limited strength for providing security and this can be overcome by the proposed PDP model. The highest security level is ensured by this PDP approach based on the RS randomness. When compared to the take-a-tail approach, the PDP increases the bar by the terms of memorability stress and system interference and also makes this proposed PDP as an extremely practical approach for the use of common users. Also it has the benefit of storing a single information that reduces the storage overhead significantly compared to that of the existing approaches.

Table 3. Comparative usability analysis of honeyword generation methods. ⊛ represents if randomness of RS is high. $U \approx G$ represents if honeywords are distributed as user chosen password from the attacker point of view.

Honeyword based techniques	Flatness	Security against MSV	System interference	Stress on memorability	Typo safety	DoS Resiliency	Storage overhead
Take-a-tail	$1/k$	High	High	High	High	Low	k-1
CTD	$\frac{1}{k}$if $U \approx G$	Low	No	Low	Low	Low	k-1
Modelling-syntax	$\frac{1}{k}$if $U \approx G$	Low	No	Low	High	High	k-1
PDP	$1/k$ ⊛	High	Low	Low	High	High	1

6. CONCLUSION AND FUTURE WORK

The current trending popular password based authentication techniques are the Honeyword based techniques that provides various advantages over traditional techniques. But in the honeyword based techniques, the major drawback is the storage cost and overhead. To overcome these existing drawbacks, this paper introduces a novel honeyword generation approach with the new decoy mechanism. This proposed work is implemented and compared with the existing approach. From the results, it is concluded that the proposed honeyword based approach provides improvement in security with reduced storage cost and storage overhead.

REFERENCES

[1] A. Juels and R. L. Rivest, "Honeywords: Making password-cracking detectable," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 145-160.

[2] S. Ji, S. Yang, X. Hu, W. Han, Z. Li, and R. Beyah, "Zero-sum password cracking game: A large-scale empirical study on the crackability, correlation, and security of passwords," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, pp. 550-564, 2017.

[3] B. Ur, S. M. Segreti, L. Bauer, N. Christin, L. F. Cranor, S. Komanduri, *et al.*, "Measuring Real-World Accuracies and Biases in Modeling Password Guessability," in *USENIX Security Symposium*, 2015, pp. 463-481.

[4] M. H. Almeshekah, C. N. Gutierrez, M. J. Atallah, and E. H. Spafford, "Ersatzpasswords: Ending password cracking and detecting password leakage," in *Proceedings of the 31st Annual Computer Security Applications Conference*, 2015, pp. 311-320.

[5] A. L.-F. Han, D. F. Wong, and L. S. Chao, "Password cracking and countermeasures in computer security: A survey," *arXiv preprint arXiv:1411.7803*, 2014.

[6] S. Houshmand, S. Aggarwal, and R. Flood, "Next gen PCFG password cracking," *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 1776-1791, 2015.

[7] R. Chatterjee, J. Bonneau, A. Juels, and T. Ristenpart, "Cracking-resistant password vaults using natural language encoders," in *Security and Privacy (SP), 2015 IEEE Symposium on*, 2015, pp. 481-498.

[8] J. Galbally, I. Coisel, and I. Sanchez, "A New Multimodal Approach for Password Strength Estimation—Part I: Theory and Algorithms," *IEEE Transactions on Information Forensics and Security*, vol. 12, pp. 2829-2844, 2017.

[9] H. Kumar, S. Kumar, R. Joseph, D. Kumar, S. K. S. Singh, and P. Kumar, "Rainbow table to crack

password using MD5 hashing algorithm," in *Information & Communication Technologies (ICT), 2013 IEEE Conference on*, 2013, pp. 433-439.

[10] J. Ma, W. Yang, M. Luo, and N. Li, "A study of probabilistic password models," in *Security and Privacy (SP), 2014 IEEE Symposium on*, 2014, pp. 689-704.

[11] S. Ji, S. Yang, T. Wang, C. Liu, W.-H. Lee, and R. Beyah, "Pars: A uniform and open-source password analysis and research system," in *Proceedings of the 31st Annual Computer Security Applications Conference*, 2015, pp. 321-330.

[12] S. M. Segreti, B. Ur, L. Bauer, and N. Christin, "Designing Password Policies for Strength and Usability," ed: TISSEC, 2016.

[13] H.-C. Chou, H.-C. Lee, H.-J. Yu, F.-P. Lai, K.-H. Huang, and C.-W. Hsueh, "Password cracking based on learned patterns from disclosed passwords," *IJICIC*, 2013.

[14] I. Erguler, "Some Remarks on Honeyword Based Password-Cracking Detection," *IACR Cryptology ePrint Archive*, vol. 2014, p. 323, 2014.

[15] S. Yang, S. Ji, and R. Beyah, "DPPG: A Dynamic Password Policy Generation System," *IEEE Transactions on Information Forensics and Security*, 2017.