ISSN No. 0976-5697

Volume 8, No. 1, Jan-Feb 2017

International Journal of Advanced Research in Computer Science

RESEAECH PAPER

Available Online at www.ijarcs.info

Password Predictability Using Neural Networks

Menal Dahiya
Department of Computer Science
Maharaja Surajmal Institute
Delhi, India

Abstract: In this paper, we use Recurrent Neural Network for predicting the password characters using back propagation algorithm. For today's security system text password is an extensively used form of authentication. Evaluation of password strength and guessing attacks are some concepts that help in modelling of the password. We propose Artificial Neural Network to generate the text password on the basis of their training methods. This approach is fast and accurate and applied on text data only.

Keywords: Authentication; Back propagation; Password strength; Recurrent Neural Network; Guessing Attacks.

I. INTRODUCTION

There are a number of authentication mechanisms that are present in today's computer security system. Text based passwords are the most rigorously used authentication mechanism in the current system and seems to be continued in future. Researchers have listed some problems regarding text based passwords [1]. As users often choose light passwords and easily predictable passwords therefore, guessing of text based passwords is the simple process for the attackers/hackers. The research society has introduced significant efforts in increasing the alternatives, including the use of biometric password like iris, face, and fingerprint, etc., graphical passwords like image password or combination of image and text, hardware tokens and the use of smart cards. However text based passwords remains at the top priority of the users for authenticating their resources [2]. Continuous effort has been done in the field of password usage and integrity, but different types of obstacles are there in collecting realistic data for analyzing, collecting data from experimental studies rather than from deploying authentication systems [3].

The systems that use passwords for their security purposes, then it is essential to evaluate the strength of a password and evaluate their ability to recover from guessing attacks [4-6]. Various guessing techniques like probabilistic approach, which describe a number of ways a user build or choose a password. These probabilities are then used by attackers in guessing attacks and determine the password sequences. Advanced probabilistic guessing schemes use leaked password list for their experiments and upgrade the strategy of guessing passwords that have not even been leaked [7-9]. Password recovery tools can model the weaknesses of passwords to guessing by expert attacker's [10]. These techniques require large memory around hundreds of gigabytes of disk space, large computational power and taking days to execute. Therefore, these techniques are not very much feasible in the evaluation of password strength. At the same time, difficulty in guessing password remains important. Passive attacks, active attacks and offline attacks where attackers crack passwords up to a trillion of guesses [11]. Password security is an integral part of any organizational infrastructure and attacks become efficient, so solutions are designed in such a way that help users in choosing strong and better passwords [12]. Password strength is known as the number of times an attacker would guess the password. The guessing strategy computes the password strength and increases the computational efforts. In this paper, we first outline the related work in the field of password guessing (section II); section III describes the overview of neural network and section IV explains the proposed approach of password prediction.

II. RELATED WORK

The guessing attacks are directly related to password strength. If guessing a password is difficult, then it means gauging password strength is also difficult or it is a strong password. Password guessing attacks are a threat in many situations like in year 2014 theft of celebrity personal photos from Apple's iCloud, large scale guessing becomes possible [13]

The first demonstration on password guessing attack was in 1979, when Morris and Thompson explained that computers guess "a considerable fraction" of the passwords for unix systems through brute force and dictionary attacks [14]. Rainbow chains are a technique that memorizes computed passwords very efficiently and find password that appear in them [15]. Frequently, very offline attacks are caused when a database of hashed passwords is stolen [16-17]. An attacker takes a password, hash it and search the match in the database when match is found, try to use the same credentials on other systems [18].

More recently, probabilistic attacks are proposed to slow down the number of guesses for passwords that are not found in dictionaries. One probabilistic method applies probabilistic context free grammars (PCFGs) [19]. Here a template has built in clear text through a training set of passwords and through a stochastic process each password has a given probability of being chosen. Furthermore, semantic patterns are amended in PCFGs, natural language dictionaries improves guessing [20].

The Markov model proposed in 2005 was used to guess password. The Markov model predicts the probability of the next character based on previous characters. Using more context characters, over fitting risk is increased [21]. Smoothing and backoff methods are applied to indemnify the risk.

III. NEURAL NETWORK

Neural Networks are the machine learning network used for approximating highly dimensional functions. Neural

Network is based on the model of human neurons and solve complex problems, generate sequences, solve classification problems and applicable in many fuzzy applications. Concept of password generation is similar as generating sequences or text sequences.

Recurrent Neural Networks are a type of neural network that have been used to generate sequences. Recurrent Neural Network (RNN) can be trained by using a supervised learning algorithm known as back propagation algorithm. Back propagation is a widely used algorithm and RNNs are a collection of dynamic models that have been applied in generating sequences in many domains. Password creation and password generation are although similar concepts which are not widely be researched yet. Ten years ago, a neural network was proposed as a method for classification for passwords, but that work did not maintain the guessing attack aspect [22]. Recurrent Neural Networks can be trained by data sequences, one step at a time and predicting what comes next. Number of iterations can be performed on the network and the network output is fed in the sample as input at the next step. Recurrent Neural Networks should be able to generate sequences, but are unable to store information for a long time [23]. Long Short Term Memory (LSTM) is an RNN architecture that gives better result of storing and accessing information than the general RNNs model. LSTM shows positive response in a variety of sequence generating tasks like speech and handwriting tasks [24-26].

IV. PASSWORD TEXT PREDICTION APPROACH

Neural Networks in our system are trained to generate the next letter of a password. Predictions are based on the conditions of the characters or depend on the context of the password. For example, to calculate the probability of the password “car”, the network predicts a “r” given the context “ca”. For the password prediction experiment, the training set contains 20 passwords; we design the context size and training methodology. In this work, we use Recurrent Neural Network architecture with back propagation algorithm. Each word is separated by a token and every password is start by the null vector. Network architecture consists of a single hidden layer and 40bit inputs and outputs. We take input of same size passwords say eight characters if an input password is less than eight characters than we pad the input with zeros. The network was trained with a gradient descent method with learning rate 0.0001 and momentum of 0.99.

For Recurrent Neural Networks, the input will always be a sequence and have signals travelling in both directions. These networks are dynamic in nature and change their state until they reach an equilibrium point. Artificial Neural Networks are relying on the basic neural structure of the brain. At a time they process only one record and learn by comparing their prediction with the actual record. The error of the initial prediction of the first character is fed back to the network and used to modify the network for the second iteration. These steps are replaced multiple times using back propagation algorithm. Recurrent Neural Networks have been shown to be effective for predicting password in the context of natural language. In this network , connections can process the characters in sequences and use memory for remembering about previous character in the sequence. For e.g. the above discussed password “car” we would query the network for the probability of finding a “c”, then seeing an “a” after “c” and then seeing a “r” after “ca” then of seeing a complete password “car”. Figure 1 shows an example of using neural networks to predict the next character of a password. The probabilities of each next letter/character are the output of the network. The network can predict the passwords from testing datasets by learning passwords from training data sets.

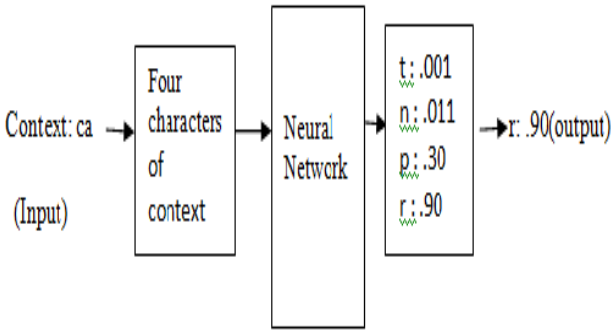

Figure 1. An Example of Using a Neural Network to Predict the Next Character of a Password Sequence.

The network learns all the passwords from training datasets and when we start with the new password and query the network for the probability of seeing any character. The network will generate the similar password and prohibit the generation of non-desirable passwords by filtering of those passwords. By using heuristic search, we can count the possible number of passwords whose probability is above a given threshold for the network.

V. CONCLUSION

This paper describes the use of Recurrent Neural Network architecture to predict the password from the testing dataset that are trained by back propagation supervised learning algorithm. Prior work has used neural network architecture for password guessing and for calculating the password strength. Recurrent Neural Network is very effectively used in generating sequences so this network solves the problem of memorizing the text passwords as it easily generate the password sequence which is already learned by the network. This approach is fast and develop accurate results.

VI. REFERENCES

[1] M. Dell’Amico and P. Michiardi, Y. Roudier, “Password Strength: An Empirical Analysis,” Proc. 29th Conference on Information Communications (INFOCOM’10), San Diego, USA, 2010, pp. 983-991, ISBN: 978-1-4244-5836-3.

[2] D. Malone and K. Maher, “Investigating the Distribution of Password Choices,” Proc. 21st International Conference on World Wide Web, Lyon, France, 16th-20th April, 2012, pp. 301-310, ISBN: 978-1-4503-1229-5.

[3] B. D. Medlin and J. A. Cazier, “An Empirical Investigation: Health Care Employee Passwords and their Crack Times in Relationship to Hipaa Security Standards,” International Journal of Healthcare Information Systems and Informatics, Volume.02, Issue.03, pp. 39–48, 2007.

[4] W. Melicher, B. Ur, S. M Segreti and et. al, “Fast, Lean, and Accurate: Modeling Password Guessability Using Neural Networks,” Proc. 25th USENIX Security Symposium, Austin, TX, 10th -12th August 2016, pp. 175-190.

[5] P. G. Kelley, S. Komanduri, M. L Mazurek and et.al, "Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms," Proc. IEEE Symposium on Security and Privacy, USA, 20th -25th May 2012, pp. 523-537, ISBN: 978-0-7695-4681-0.

[6] M. Weir, S. Aggarwal, M. Collins and H. Stren, "Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords," Proc. 17th ACM Conference on Computer and Communications Security, Chicago, Illinois, USA, 4th -8th October 2010, pp. 162-175, ISBN: 978-1-4503-0245-6.

[7] Z. Li, W. Han and W. Xu, "A Large-Scale Empirical Analysis of Chinese Web Passwords," Proc. 23rd USENIX Security Symposium, San Diego, USA, 2014, ISBN: 978-1-931971-15-7.

[8] J. Ma, W. Yang, M. Luo and N. Li, "A Study of Probabilistic Password Models," Proc. IEEE Symposium on Security and Privacy, 18th-21st May 2014, DOI: 10.1109/SP.2014.50.

[9] R. Veras, C. Collins and J. Thorpe, "On the Semantic Patterns of Passwords and their Security Impact," Proc. Network and Distributed System Security Symposium, January 2014, DOI: 10.14722/NDSS.2014.23103.

[10] B. Ur, S. M. Segreti, L. Bauer, N. Christin and et.al, "Measuring Real-World Accuracies and Biases in Modeling Password Guessability," Proc. 24th Usenix Security Symposium, Washington, DC, USA, 12th -14th August 2015.

[11] J. Bonneau, "The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords," Proc. IEEE Symposium on Security and Privacy, USA, 20th-25th May 2012, pp. 538-552, ISBN: 978-0-7695-4681-0.

[12] J. Bonneau, C. Herley, P. C. V. Oorschot and F. Stajano, "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes," Proc. IEEE Symposium on Security and Privacy, USA, 20th-25th May 2012, pp. 553-567, ISBN: 978-0-7695-4681-0.

[13] A. Greenberg, "The Police Tool that Pervs Use to Steal Nude Pics From Apple's iCloud," Wired, February 2014, https://www.wired.com/2014/09/eppb-icloud/.

[14] R. Morris and K. Thompson, "Password Security: A Case History," Communications of the ACM**,** Volume.22, Issue.11, November 1979, pp. 594-597, DOI: 10.1145/359168.359172.

[15] P. Oechslin, "Making a Faster Cryptanalytic Time-Memory Trade-off," Springer Advances in Cryptology (CRYPTO), Volume.2729, 2003, DOI: 10.1007/978-3-540-45146-4_36.

[16] J. Bonneau, "The Gawker Hack: How a Million Passwords Were Lost," Light Blue Touchpaper, University of Cambridge, December 2010, https://www.lightbluetouchpaper.org/2010/12/15/the-gawker-hack-how-a-million-passwords-were-lost/.

[17] J. Brodkin, "10 (or so) of the Worst Passwords Exposed by the Linkedin Hack," Ars Technica, June 2012, https://arstechnica.com/security/2012/06/10-or-so-of-the-worst-passwords-exposed-by-the-linkedin-hack/.

[18] A. Das, J. Bonneau, M. Caesar, N. Borisov and X. Wang, "The Tangled Web of Password Reuse," Proc. NDSS Symposium, February 2014.

[19] M. Weir, S. Aggarwal, B. De Medeiros and B. Glodek, "Password Cracking Using Probabilistic Context-Free Grammars," Proc. 30th IEEE Symposium on Security and Privacy, Oakland, USA, 17th-20th May 2009, DOI: 10.1109/SP.2009.8.

[20] R. Veras, C. Collins and J. Thorpe, "On the Semantic Patterns of Passwords and their Security Impact," Proc. NDSS Symposium, February 2014.

[21] A. Narayanan and V. Shmatikov, "Fast Dictionary Attacks on Passwords Using Time-Space Tradeoff," Proc. 12th ACM Conference on Computer and Communications Security, Alexandria, VA, USA, 7th -11th November 2005, ISBN:1-59593-226-7.

[22] A. Ciaramella, P. D'Arco, A. D. Santis, C. Galdi and R. Tagliaferri, "Neural Network Techniues for Proactive Password Checking," Proc. IEEE Transactions on Dependable and Secure Computing, Volume.03, Issue.04, October 2006, pp. 327-339, DOI: 10.1109/TDSC.2006.53.

[23] S. Hochreiter, Y. Bengio, P. Frasconi and J. Schmidhuber, "Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies," Field Guide to Dynamical Recurrent Neural Networks, IEEE Press, 2001.

[24] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, Volume.09, Issue.08, November 1997, pp. 1735-1780, DOI: 10.1162/NECO.1997.9.8.1735.

[25] A. Graves, A. Mohamed and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," Proc. International Conference on Acoustics, Speech, and Signal Processing, 26th-31st May 2013, E-ISBN: 978-1-4799-0356-6.

[26] A. Graves and J. Schmidhuber, "Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks," Proc. 21st International Conference on Neural Information Processing Systems, Canada, 8th -10th December 2008, pp. 545-552, ISBN: 978-1-6056-0-949-2.

