Smarter Password Guessing Techniques Leveraging Contextual Information and OSINT

Aikaterini Kanta*[†], Iwen Coisel[†], Mark Scanlon*

*Forensics and Security Research Group, School of Computer Science, University College Dublin, Ireland
[†]European Commission, Joint Research Centre (DG JRC) - Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
Email: aikaterini.kanta@ucdconnect.ie, iwen.coisel@ec.europa.eu, mark.scanlon@ucd.ie

Abstract—In recent decades, criminals have increasingly used the web to research, assist and perpetrate criminal behaviour. One of the most important ways in which law enforcement can battle this growing trend is through accessing pertinent information about suspects in a timely manner. A significant hindrance to this is the difficulty of accessing any system a suspect uses that requires authentication via password. Password guessing techniques generally consider common user behaviour while generating their passwords, as well as the password policy in place. Such techniques can offer a modest success rate considering a large/average population. However, they tend to fail when focusing on a single target – especially when the latter is an educated user taking precautions as a savvy criminal would be expected to do. Open Source Intelligence is being increasingly leveraged by Law Enforcement in order to gain useful information about a suspect, but very little is currently being done to integrate this knowledge in an automated way within password cracking. The purpose of this research is to delve into the techniques that enable the gathering of the necessary *context* about a suspect and find ways to leverage this information within password guessing techniques.

Index Terms—Password Security, Password Guessing Techniques, Context-based Password Cracking, Open Source Intelligence (OSINT)

I. INTRODUCTION

Nowadays, criminal activity is increasingly conducted in cyberspace. Criminals take advantage of the easy access to information and the global access to victims. This leads to a scale of crime that cannot be easily achieved in the physical world. As a result, it is becoming increasingly urgent for law enforcement to be able to act swiftly in a digital forensic investigation especially in the cases where ongoing or future criminal acts must be prevented. Very often, password protected accounts or encrypted devices act as a barrier for police personnel to conduct their lawful investigations [1].

Passwords have been the go-to method of user authentication for decades – a fact that does not look like it is about to change. The difference in the last few years is the fast increase of online login systems with password policies that require passwords of different patterns, lengths and makeup [2]. This leads to users having the tendency to either reuse the same password across different systems or to create passwords that are easy to remember (i.e., weaker) to keep up with the different password policies [3], [4]. A significant portion of passwords are therefore based on dictionary words [5] or contextual information related to the user [6]. Taking this into

account, this work is exploring ways this information about a suspect as an individual or a community can be leveraged in order to better facilitate the recovery process.

II. RELATED WORK

A. Password Metrics

There have been studies that analyse the composition of passwords with focus on how people choose them, where password re-use and reliance on dictionary words can be observed [4], [5], [7]. In addition, studies have considered the demographics of participants and if they play an important role in the selection process [8], [9]. Finally, it is observed that users tend to use personal information when they create a password, as it is more easily memorable to them [6], [10].

B. Open Source Intelligence

Steele [11] defines Open Source Intelligence (OSINT) as information that is publicly available and can be used to answer a specific question. To this end, there are many OSINT tools available to the community that can aid in finding, extracting and sorting though this information[1].

C. Password Guessing Related Tools

Traditional password guessing techniques include brute force, dictionary attacks and rainbow table approaches. Lately, newer, smarter methods have been proposed with higher recovery numbers, e.g., password candidate generation tools based on Markov Chains [12], probabilistic context-free grammars [13] and variations of thereof [10], and combinator attack tools, e.g., PRINCE[2].

III. METHODOLOGY

Digital forensic investigators frequently find themselves working on a case where a password connected to a crime needs to be retrieved. Many times, it is not possible to retrieve the password of a suspect in a timely manner. This is where tying the suspect to their associated *contextual* information can prove fruitful. Context in this instance refers to contextual information about a suspect that can be harnessed for the sake of making better "educated" guesses about their password.

This information can stem from their online/offline life and can be the product of traditional means of investigation,

[1]https://osintframework.com/
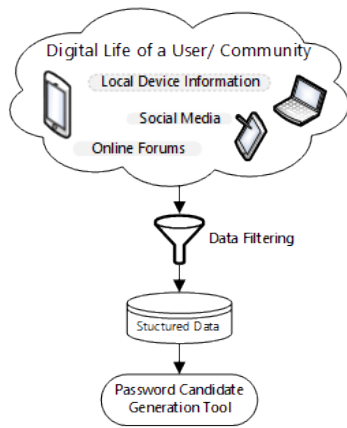[2]https://github.com/hashcat/princeprocessor

Fig. 1. Password Generation Process

e.g., forensic investigation of the suspect's residence and belongings or the digital investigation of the suspect (i.e., both their local devices and online presence). The latter type of investigation can yield results including the suspect's interests (e.g., sports, music, etc.), browsing history, other passwords, online interactions, family and pet names, etc. This is where OSINT can play a key role in an investigation. Already Social Media Intelligence (SOCMINT) is used by law enforcement in order to gain pertinent, case-progressing information [14]. It can be applied to online groups of users to detect patterns in social behaviour as well as to individual suspects [15]. OSINT tools can extract information regarding the online presence of a user or group of users from networks of acquaintances on social media. The prevalence of users increasingly living their lives online can also result in sourcing their email addresses, usernames, phone numbers, and exercise or sleeping patterns.

The information gathered can then be leveraged, filtered and translated to meaningful contextual data about the suspect. Information retrieval and machine learning techniques should be beneficial at this stage, based on the volume of available data and whether an individualised or community-based approach is chosen. For the individual-based approach, information gathered online and offline will be compiled. An analysis on the raw data and a classification into categories is the next step, in order to extract meaningful keywords that will represent user/community interests. In turn, these keywords will help law enforcement officers assemble a more useful list of password candidates to enrich and complement existing password guessing tools. A flowchart of this process can be seen in Figure 1.

IV. Conclusion and Future Work

To validate the proposed hypothesis, i.e., prove that context does play a role in the selection of passwords, our future work will firstly focus on a community-based approach. The reason for this, is the current lack of available data for individuals and the sensitivity of this data [16]. As one example, a community of users on an anime forum would be expected to have a higher percentage of passwords related to anime than a community about cooking. Current work is focused on the contextual analysis of a large corpus of passwords stemming from online leaks. Future work will include the refinement of this dataset and the expansion of the scope of this analysis to more contextual information. In addition, the manner with which OSINT can be exploited and processed towards producing meaningful data will be explored. This will subsequently be a starting point for creating a bespoke, personalised dictionary list to feed into password cracking tools.

References

[1] A. Sayakkara, N.-A. Le-Khac, and M. Scanlon, "A Survey of Electromagnetic Side-channel Attacks and Discussion on Their Case-Progressing Potential for Digital Forensics," *Digital Investigation*, vol. 29, pp. 43 – 54, 2019.

[2] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer *et al.*, "How does your password measure up? The effect of strength meters on password creation," in *Presented as part of the 21st USENIX Security Symposium USENIX Security 12)*, 2012, pp. 65–80.

[3] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman, "Of Passwords and People: Measuring the Effect of Password-Composition Policies," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 2595–2604.

[4] N. Lord, "Uncovering Password Habits: Are Users' Password Security Habits Improving?" Dec 2018. [Online]. Available: https://digitalguardian.com/blog/uncovering-password-habits-are-users-password-security-habits-improving-infographic

[5] J. Yan, A. Blackwell, R. Anderson, and A. Grant, "Password Memorability and Security: Empirical Results," *IEEE Security & Privacy*, vol. 2, no. 5, pp. 25–31, 2004.

[6] D. Wang, Z. Zhang, P. Wang, J. Yan, and X. Huang, "Targeted Online Password Guessing: An Underestimated Threat," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 1242–1254.

[7] S. Pearman, J. Thomas, P. E. Naeini, H. Habib, L. Bauer, N. Christin, L. F. Cranor, S. Egelman, and A. Forget, "Let's go in for a closer look: Observing passwords in their natural habitat," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 295–310.

[8] J. Bonneau, "The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords," in *2012 IEEE Symposium on Security and Privacy*. IEEE, 2012, pp. 538–552.

[9] M. AlSabah, G. Oligeri, and R. Riley, "Your Culture is in Your Password: An Analysis of a Demographically-Diverse Password Dataset," *Computers & Security*, vol. 77, pp. 427–441, 2018.

[10] Y. Li, H. Wang, and K. Sun, "A Study of Personal Information in Human-Chosen Passwords and Its Security Implications," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[11] R. D. Steele, "Open source intelligence," *Handbook of Intelligence Studies*, vol. 42, no. 5, pp. 129–147, 2007.

[12] M. Dürmuth, F. Angelstorf, C. Castelluccia, D. Perito, and A. Chaabane, "OMEN: Faster Password Guessing Using an Ordered Markov Enumerator," in *International Symposium on Engineering Secure Software and Systems*. Springer, 2015, pp. 119–132.

[13] M. Weir, S. Aggarwal, B. De Medeiros, and B. Glodek, "Password Cracking Using Probabilistic Context-Free Grammars," in *2009 30th IEEE Symposium on Security and Privacy*. IEEE, 2009, pp. 391–405.

[14] R. C. Van der Hulst, "Introduction to Social Network Analysis (SNA) as an Investigative Tool," *Trends in Organized Crime*, vol. 12, no. 2, pp. 101–121, 2009.

[15] A. L. Ivan, C. A. Iov, R. C. Lutai, and M. N. Grad, "Social Media Intelligence: Opportunities and Limitations," *CES Working Papers*, vol. 7, no. 2A, p. 505, 2015.

[16] D. Fleurbaaij, M. Scanlon, and N.-A. Le-Khac, "Privileged Data within Digital Evidence," in *Proceedings of the 16th IEEE International Conference On Trust, Security And Privacy In Computing And Communications (TrustCom-17)*. Sydney, Australia: IEEE, 08 2017, pp. 737–744.