Assessing the Reliability of Keystroke Timing for Authentication

Priyanshu Tomar

Department of CSE&Cybersecurity

New York Institute of Technology

Manhattan, NY, USA

ptomar@nyit.edu

***Abstract*—This project explores the development and implementation of a sophisticated authentication system, leveraging the power of SQL databases and Python's Pandas library for data analysis. The primary goal is to enhance security measures by analyzing user authentication attempts and patterns.**

I. INTRODUCTION

In the digital age, the security of information systems is paramount. As cyber threats become increasingly sophisticated, the need for advanced and reliable authentication systems has never been greater. The goal of this project is to develop and analyze an enhanced authentication mechanism using SQL databases for data storage and Python's Pandas library for data analysis.

The project focuses on the critical aspect of user authentication - a fundamental component in safeguarding against unauthorized access to systems and data. We aim to investigate the patterns and behaviors in authentication attempts, which can provide insights into potential security vulnerabilities and user interaction with authentication systems.

To achieve this, we utilize an in-memory SQLite database, a lightweight and efficient solution for handling authentication data. The database stores several types of data related to user authentication, such as user input strings, subjects' information, and detailed logs of each authentication attempt. This data includes timestamps and key press details, enabling a granular analysis of the authentication process.

Python's Pandas library is employed to import this data into a flexible and powerful data analysis environment. Pandas' capabilities in data manipulation and analysis allow us to conduct a comprehensive examination of the time intervals between authentication attempts and the specific actions taken by users during these attempts. By analyzing these aspects, we can gain a deeper understanding of user behavior patterns and the efficacy of the authentication system.

Through this project, we aim to bridge the gap between raw authentication data and actionable insights that can enhance system security. By applying a data-driven approach to the study of authentication systems, we contribute to the development of more secure and user-friendly authentication methods. This project is not only a step forward in the field of cybersecurity but also sets a foundation for future research and advancements in authentication technologies.

II. RELATED WORK

We have read a few research papers related to this research. Below is the detailed summary of them:

A. Your PIN Sounds Good! Augmentation of PIN Guessing Strategies via Audio Leakage:

This research paper presents a groundbreaking study in cybersecurity, particularly focusing on the vulnerabilities associated with PIN-based authentication systems. This study delves into how audio feedback from keypad entries, such as those on ATMs and point-of-sale terminals, can be exploited to infer PIN codes. Researchers have shown that analyzing the timing of the audio signals emitted during key presses makes it possible to derive inter-keystroke timing information with a high degree of accuracy. It reveals that accurate inter-keystroke timing information can be derived from audio feedback, achieving a remarkable detection accuracy of 98% with an average error of only 1.8ms. This method significantly surpasses traditional approaches that rely on visual analysis from video recordings. This approach significantly surpasses the traditional methods that rely on visual analysis from video recordings, demonstrating a novel and potentially more effective way of predicting PIN codes.

The study further explores scenarios where this technique becomes even more potent, such as cases where partial PIN information is known or user behavior patterns, like single finger typing, can be observed. This multimodal approach led to substantial improvements in guessing performance, up to 34 times more effective than random guessing. Additionally, integrating inter-keystroke timing with thermal imaging data showed promising results, with up to 15% of PINs correctly guessed on the first attempt. These findings highlight a critical security gap in

PIN-based systems and suggest that audio leakage from keypads could be a feasible attack vector for malicious entities. The paper emphasizes the need for enhanced security protocols to mitigate the risks posed by such audio-based side-channel attacks. By bringing attention to this often-overlooked aspect of security, the research contributes valuable insights into improving the robustness of authentication systems against emerging threats.

B. PILOT: Password and PIN Information Leakage from Obfuscated Typing Videos:

This study presents an in-depth analysis of security vulnerabilities associated with typing passwords and PINs. It focuses on a novel attack method named PILOT, which stands for "Password and PIN Information Leakage from Obfuscated Typing Videos." This method is designed to exploit the vulnerability that occurs when typing passwords or PINs on devices where the characters are masked or obfuscated on-screen. The PILOT attack extracts inter-keystroke timing information from video recordings of these masked character feedback instances. This report conducted a series of experiments to test the efficacy of the PILOT attack. These experiments involved analyzing video recordings of individuals entering their passwords or PINs on various devices. The key aspect of the research was to determine how much information could be inferred about the passwords or PINs based on the timing between keystrokes, despite the characters themselves being hidden.

The results of the study showed that the PILOT method could significantly reduce the number of attempts needed to correctly guess passwords and PINs compared to random guessing. For alphanumeric passwords, PILOT could recover them in as few as 19 attempts in certain scenarios. When guessing PINs, the method improved the success rate by about 26 times compared to random guessing, being able to guess 3% of PINs within 10 attempts. These findings challenge the assumption that masked characters during password entry provide sufficient protection against visual attacks and highlight a critical vulnerability in current graphical user interfaces used for password masking. This finding is crucial as it challenges the common assumption that masked characters during password entry are sufficient to protect against such visual attacks. The research highlights a critical vulnerability in current graphical user interfaces (GUIs) used for password masking.

C. Whispering Devices: A Survey on How Side-Channels Lead to Compromised Information:

This research paper shows the realm of side-channel attacks, which have become a significant concern in cybersecurity. The paper extensively surveys various forms of side-channel attacks, categorizing them based on their nature and medium of exploitation. This includes electromagnetic, acoustic, and power consumption side-channels, among others. Each type is analyzed for how it can be exploited to extract confidential information from electronic devices. For example, electromagnetic emissions from a device can be intercepted to infer data being processed, while acoustic side-channels can involve analyzing sound patterns emitted by hardware components.

The study offers a comprehensive classification of side channels, considering factors like the criticality of the information at risk, the intentionality of the attack, the types of potential attackers, and the physical medium through which the side-channel operates. This classification helps in understanding the various dimensions and complexities of side-channel threats. The research identifies that even seemingly innocuous channels can be manipulated to glean sensitive data, underscoring the ingenious and often unexpected nature of these security breaches.

One of the key findings of the research is the effectiveness of using timing information alone in guessing PINs. The study shows that with timing information, the accuracy of correctly guessing the PIN increases significantly. Specifically, when only timing information was utilized, the guessing accuracy improved by 15-19 times within the first three attempts, 7 times within five attempts, and approximately 4 times within ten attempts, compared to a scenario where one digit of the PIN is already known. This clearly demonstrates the potential of timing information as a powerful tool in deciphering PINs, even without any additional knowledge of the PIN itself.

Furthermore, the research delves into the impact of user typing behavior on the guessing rates. It was observed that PINs entered using a single finger were much more susceptible to being guessed using timing information. In such cases, the guessing rate within the first five attempts was 34 times higher compared to random guessing. This stark difference in the ease of guessing single finger entered PINs highlights a significant security concern. In contrast, the ability to guess PINs entered using multiple fingers was only marginally better than random guessing. This suggests that the method by which users input their PINs – whether with a single finger or multiple fingers – can influence the vulnerability of their PINs to timing-based side-channel attacks.

Overall, the research underscores a critical vulnerability in PIN-based systems, where timing information can be leveraged to achieve a high success rate in guessing PINs, particularly for single-finger typists. This highlights the need for enhanced security measures in systems relying on PIN authentication to safeguard against such vulnerabilities.

D. We Can Hear Your PIN Drop: An Acoustic Side-Channel Attack on ATM PIN Pads:

This research paper shows an acoustic side-channel attack, Pindrop, targeting ATM PIN pads. It demonstrates Pindrop's ability to accurately reconstruct PINs by profiling the unique audio signatures of key presses on PIN pads. The method was validated through experiments with 58 participants and two different PIN pad models, simulating real-world conditions including various distances and noise levels. It shows that Pindrop can accurately reconstruct up to 96% of 4-digit and 94% of 5-digit PINs within three attempts in controlled settings. The results indicate a significant success rate in reducing PINs, revealing a considerable security risk in current ATM systems, and emphasizing the need for enhanced protection measures against such acoustic attacks.

This study explores a novel attack method called Pindrop. This technique uses acoustic side-channel attacks to infer ATM PINs based on the sounds of key presses on PIN pads. The researchers evaluated the performance of Logistic Regression (LR) and Support Vector Classification (SVC) classifiers in two different PIN pad models under various scenarios, including different numbers of attackers and digits entered by each attacker. The study found that accuracy in identifying keys and reconstructing PINs is influenced by the size of the training set and the distance of the recording device from the PIN pad. Additionally, the effectiveness of the attack is demonstrated under various noise conditions, highlighting the real-world applicability of this method and the need for robust countermeasures against such acoustic attacks.

III. GOAL

The project embarks on an in-depth exploration of user behavior in PIN-based authentication systems. The foundational phase of this project revolves around meticulous data collection and preparation. This involves gathering comprehensive data on user interactions with PIN-based systems, detailed in datasets, which potentially contains records of each PIN entry attempt by users. A critical part of this phase is data cleaning, ensuring the reliability and relevance of the data for subsequent analysis. This stage sets the groundwork for insightful analysis by filtering out inaccuracies and irrelevant information, thus creating a robust dataset for the project.

A key aspect of this project is the analysis of time differences in authentication attempts, which is central to understanding the dynamics of user interactions with PIN systems. By employing a Python script, the project aims to dissect the timing intervals between successive PIN entries. This analysis is crucial as it sheds light on typical user behavior patterns and highlights any deviations that might suggest unauthorized access attempts or difficulties faced by users in the authentication process. The goal here is to unravel the subtleties of user behavior that go beyond mere successful or unsuccessful authentication, delving into the 'how' and 'when' aspects of PIN entry.

In addition to the temporal analysis of individual PIN entries, the project also focuses on calculating average time differences for users. This involves processing data, computing and comparing the average time taken by each user for authentication. This step is instrumental in profiling standard user behavior, thereby enabling the identification of outliers. These outliers are crucial for pinpointing potential security threats, as they represent deviations from the norm that could be indicative of either security breaches or areas where the user interface may need improvement.

The project also encompasses a significant component of data querying and management, utilizing SQL queries. This step is essential for the efficient extraction and manipulation of data to support focused analysis. By executing tailored queries, the project can filter and sort through vast amounts of data to extract meaningful insights. This phase is about transforming raw data into actionable intelligence, which can then be used to answer specific research questions or test hypotheses about user behavior and system security.

Finally, the culmination of the project lies in its ability to draw security implications and recommendations from the data analysis. The insights gained are expected to pave the way for enhancing the security protocols of PIN-based authentication systems. This phase involves not only interpreting the data patterns and anomalies but also formulating practical recommendations to mitigate identified security risks. The project aims to propose modifications to current systems or introduce new measures that could flag unusual entry patterns or reinforce the verification process for suspicious attempts. Furthermore, the findings and recommendations will be compiled into a comprehensive report, supplemented with data visualizations to ensure clarity and impact. This report will serve as a vital tool for communicating complex data patterns in an understandable manner to both technical and non-technical stakeholders, thereby facilitating informed decision-making in enhancing the security of PIN-based systems.

IV. TOOLS AND METHODOLOGY

Tools:

A. Python:

1. Python's Functionality in the Project:

Python serves as the backbone for data processing and analysis in this project. Its rich ecosystem of libraries, such as Pandas for data manipulation and NumPy for numerical calculations, makes it an ideal choice for handling large datasets and performing complex statistical analyses. Python's versatility allows for the integration of various data sources, processing of large volumes of information, and execution of sophisticated algorithms. Its readability and ease of use also facilitate rapid development and testing of analytical models.

2. *Python's Application in the Project:*

In this project, Python is employed to parse and process data from .csv files, execute the script, and perform intricate analyses such as calculating time intervals between PIN entries. It plays a crucial role in extracting actionable insights from raw data, applying statistical methods to uncover patterns in user behavior, and identifying anomalies that could indicate potential security breaches or areas for system improvement. Database Utilization: We employed an SQL-based database, which is a structured collection of data. This database contains records of various PIN entry attempts by different users.

B. *SQL and Database Management Systems:*

1. *SQL's Functionality in the Project:*

SQL is indispensable for efficiently querying and managing large databases. It allows for precise data retrieval, manipulation, and management, making it a fundamental tool for any data-centric project. SQL's ability to handle complex queries and its compatibility with various database systems make it a versatile tool for data analysts and researchers, enabling them to extract specific information from large, intricate datasets.

2. *SQL's Application in the Project:*

In this cybersecurity project, SQL is used to extract specific data sets for analysis. It enables the researchers to perform targeted queries, such as selecting authentication attempts within a specific time limit, filtering entries based on success or failure rates, or combining data from various sources. This targeted extraction is key to the project's focus on analyzing user behavior and identifying security vulnerabilities in PIN-based systems.

C. *<u>CSV Files:</u>*

1. *CSV Files' Functionality in the Project:*

CSV files are a fundamental tool for storing and handling tabular data in a simple, efficient manner. Their compatibility with a wide range of data analysis tools, including Python and SQL, makes them a preferred choice for data storage in various research and analytics projects. The format's simplicity allows for easy importation and exportation of data, facilitating the exchange of information between different systems and software.

2. *CSV Files' Application in the Project:*

In this project, CSV files such as pin_data_NYIT.csv and average_time_diff_per_user.csv are used to store raw data collected from the authentication systems. These files serve as the primary data source for analysis, containing detailed records of user authentication attempts, time intervals, and other relevant data. The project utilizes these CSV files to aggregate, organize, and prepare data for in-depth analysis, forming the foundation upon which further analytical processes are built.

D. *Data Visualization Tools:*

1. *Data Visualization Tools' Functionality in the Project:*

Data visualization tools are crucial for translating complex data sets into understandable and insightful visual formats. Tools like Matplotlib in Python are used to create charts, graphs, and other visual representations that simplify the interpretation of data. These tools are particularly important in identifying trends, patterns, and anomalies in large data sets, making them invaluable in data analysis and reporting.

2. *Data Visualization Tools' Application in the Project:*

In this cybersecurity project, data visualization tools are employed to illustrate the findings from the data analysis. They help in visualizing complex patterns such as the distribution of time intervals in PIN entries and the identification of outliers. These visual aids are integral to the project’s reports, providing clear and concise representations of the data that can be easily understood. They play a key role in communicating the project's findings, enhancing the clarity and impact of the reported results.

Methodology:

A. *Data Collection and Preparation:*

Purpose: Gathering and preparing the right data is crucial for any data analysis project. It involves sourcing data from databases or logs and cleaning it to ensure accuracy and relevance.

Application in Project: This involves extracting and cleaning the data from CSV files, ensuring it is free from errors and in a format suitable for analysis.

B. *Time Analysis:*

Purpose: Time analysis involves studying the temporal aspects of user interactions with the system, such as how long it takes to enter a PIN.

Application in Project: The Python script is used to calculate the time intervals between PIN entries, analyze average times, and identify any unusual patterns or outliers.

C. *Behavioral Analysis:*

Purpose: Understanding user behavior is key to identifying security vulnerabilities and improving system design.

Application in Project: Statistical methods are applied to the time data to profile typical user behavior and spot anomalies that might indicate security issues or usability problems.

D. *Database Querying:*

Purpose: Efficiently extracting specific data from large datasets is crucial for focused analysis.

Application in Project: SQL queries are used to pull relevant data from databases, such as specific user groups, time periods, or transaction types.

E. *Data Visualization and Reporting:*

Purpose: Transforming data into a visual format makes it more accessible and understandable. Reporting findings in a structured manner is essential for communicating results.

Application in Project: Data visualization tools are used to create graphs and charts that illustrate the project's findings, which are then compiled into a comprehensive report.

F. *Security Analysis and Recommendations:*

Purpose: The goal of the project is to enhance the security and efficiency of PIN-based systems.

Application in Project: Analyzing the data to identify potential security flaws and making recommendations for system improvements based on the findings.

V. RESULTS

The results of the project analyzing user authentication patterns in PIN-based systems through keystroke timings reveal substantial insights into the security vulnerabilities of such systems. The histogram data, as part of the project's findings, indicates a distribution of average time differences per user that clusters around specific intervals. This suggests that most users exhibit a certain rhythm or speed when inputting their PINs, with notable peaks in the frequency distribution that could represent typical user behavior. The presence of these peaks might also suggest that users have a natural cadence which, if disrupted, could be indicative of a forced or unfamiliar entry, potentially signaling a security concern.

Outliers in the data—users with average time differences that fall outside the common intervals—could indicate exceptional cases or anomalous behavior. These could be due to distinct reasons, from individual differences in interaction with the keypad to more concerning scenarios like attempted fraud. In a security context, these outliers are particularly important as they could be indicative of sophisticated attacks where a malicious actor may be inputting a stolen or guessed PIN. The project's methodology of analyzing audio feedback proves to be a novel approach that surpasses traditional visual surveillance, offering a more discreet and potentially more accurate means of detecting fraudulent activity.

The project's findings also open discussions about the balance between security and user privacy. While acoustic analysis can enhance security, it must be implemented in a way that respects user privacy and complies with data protection regulations. Additionally, the insights from this project could lead to recommendations for redesigning the acoustic feedback of keypads to minimize the risk of eavesdropping, or to the development of noise-masking technologies that obscure the audio signature of key presses.

The approach to using acoustic signals for security purposes underscores the need for a layered security model. This model would incorporate numerous factors, including time analysis, user behavior, and acoustic signatures, into a comprehensive security strategy. The insights gained can assist in creating more resilient PIN-based authentication systems, capable of adapting to and mitigating sophisticated cyber threats while maintaining a user-friendly interface.

VI. CONCLUSION

The conclusion of this project on analyzing acoustic emanations for PIN entry times provides a nuanced understanding of user behavior and its implications for security. The data, represented by the histogram of average time differences per user, suggests there is a typical range within which most users input their PINs. This standard range, where most data points cluster, could be indicative of a normal behavioral pattern for PIN entry. Such a pattern, once established, can be a benchmark for normal user behavior, with significant deviations potentially flagging a review or additional authentication steps.

Outliers depicted in the histogram, representing users with different average entry times, present an opportunity for enhancing security measures. These could be benign, influenced by numerous factors such as user dexterity, familiarity with the device, or even the ergonomics of the PIN pad itself. On the other hand, these outliers might be indicative of fraudulent attempts, where an intruder, unfamiliar with the PIN, exhibits a distinct entry pattern that could be automatically flagged by security systems.

Expanding on these insights, the project's conclusion emphasizes the potential for dynamic security protocols. Such protocols would adapt in real-time, leveraging behavioral biometrics to provide a non-static, more challenging target for potential attackers. By analyzing not just the correctness of a PIN but also the manner of its entry, security systems can evolve to detect subtler signs of account compromise. Moreover, the project underscores the importance of designing user-friendly security solutions that minimize false positives, preventing legitimate but atypically behaving users from being unduly inconvenienced.

The project's exploration of acoustic feedback as a security vulnerability also highlights the need for countermeasures that can mitigate the risks associated with these side-channel attacks. Future work could involve developing noise-masking techniques or feedback randomization strategies that obscure the distinctive acoustic signatures of key presses, thereby enhancing user privacy and security.

VII. References

1. Cardaioli, M., Conti, M., Balagani, K. S., & Gasti, P. (2020). Your PIN sounds good! Augmentation of PIN guessing strategies via audio leakage. In *Lecture Notes in Computer Science* (pp. 720–735). https://doi.org/10.1007/978-3-030-58951-6_35
2. Balagani, K. S., Cardaioli, M., Conti, M., Gasti, P., Georgiev, M., Gurtler, T., Lain, D., Miller, C., Molas, K., Samarin, N., Saraci, E., Tsudik, G., & Wu, L. (2019). PILOT: Password and PIN Information Leakage from Obfuscated Typing Videos. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1904.00188
3. Lavaud, C., Gerzaguet, R., Gautier, M., Berder, O., Nogues, E., & Molton, S. (2021). Whispering Devices: A survey on how side-channels lead to compromised information. *Journal of Hardware and Systems Security*, *5*(2), 143–168. https://doi.org/10.1007/s41635-021-00112-6
4. Balagani, K. S., Cardaioli, M., Cecconello, S., Conti, M., & Tsudik, G. (2022). We can hear your PIN drop: an acoustic Side-Channel attack on ATM PIN pads. In *Springer eBooks* (pp. 633–652). https://doi.org/10.1007/978-3-031-17140-6_31