Dynamic Keystroke for Authentication with Machine Learning Algorithms

By Yusef Yassin, Tawab Attaie, Trenton Ward, and John Caldwell
Cyber Security Center of Excellence, Norfolk State University, Department of Computer Science Norfolk, VA 23504
Mentor: Kwesi Elliot
Advisor: Dr. Jonathan Graham

ABSTRACT

Biometrics are unique traits that distinguish every individual from each other. One biometric technique is keystroke dynamics which is an authentication method based on a user's typing rhythm on a keyboard. These rhythm patterns are based on digraphs or the timing between two successive key presses. Our goal was to record the keystroke patterns of each user to determine if the users can be authenticated accurately by this method using machine learning. We recorded the latencies between keystrokes and taught machine learning algorithms the datasets to determine if the pattern can be learnable by the machine. We tested four different machine learning algorithms: Decision Trees, Random Forest, Support Vector Machines, and Neural Networks, to determine which is most effective on accuracy. We also tested three text sizes to compare each algorithm's prediction rate based on input size.

INTRODUCTION

- Usernames and passwords are vulnerable to cyber-attacks and need a secondary authentication for extra protection
- Keystroke authentication is a biometric system that analyzes a user's typing rhythm as an identifier.
- Unobtrusive as it runs in the background when the user logs in
- With machine learning, keystroke authentication can predict the rhythm as the respective user or as an imposter

METHODOLOGY

- Create a python program that records users' timings of key presses and releases in three measurements: held time, down-down time, and up-down time.
- 80% of the data was split into training the machine learning algorithms.
- 23 testers volunteered to type the predefined texts 20 times in our research for data collection.
- Four machine learning algorithms were used: Decision Tree, Random Forest, Support Vector Machines, and Neural Network
- Compare the accuracy of each machine learning algorithm to each other with three types of criteria each differing in the amount of texts

ILLUSTRATIONS

Final Results (20 Inputs)			
Machine Learning Algorithms	Strong Password (Percentage)	One Sentence (Percentage)	Two Sentences (Percentage)
Random Forest	68%	99.8%	100%
Neural Network	40%	86%	95%
Decision Tree	41%	71%	71%
Support Vector Machines	35%	84%	97%

Table of Final Results of 20 Inputs from Each User

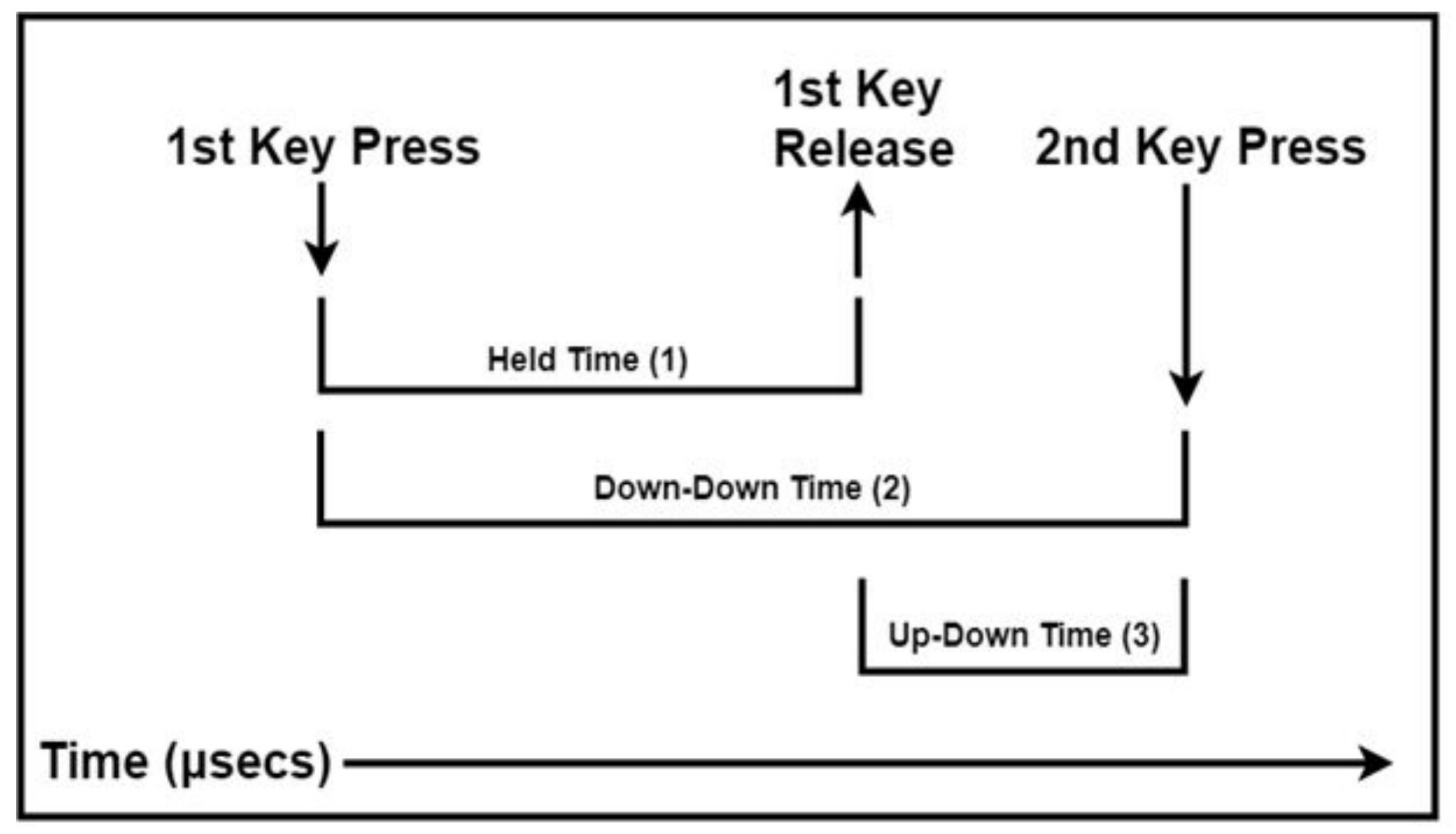

Diagram of the Types of Key Timing Measurements Collected Throughout the Experiment

Strong Password: t7Bn9=3eF-

One sentence: In our world today, there are many persons with ulterior motives.

Two sentences: In our world today, there are many persons with ulterior motives. Let us all be security conscious and not make ourselves easy targets for such persons.

Three Types of Predefined Text in Experiment

RESULTS

- The Random forest algorithm performed the best with a 100% average accuracy when using two sentences data and 99.8% for one sentence, and the Decision Tree had the worst performance at 71% accuracy for two sentences. .
- The Support Vector Machine and Neural Network algorithms performed decently with accuracies above 95% for two sentences, but not as strong accuracies for the one sentence and strong password.

CONCLUSION

The Random Forest Algorithm is most effective in terms of accuracy with keystroke authentication because it mitigates the possibility of overfitting. Strong passwords are not optimal authentication inputs for keystroke authentication as they are irrelevant to the user which prevents acquiring a clear typing pattern.

FUTURE WORK

- Implement the experiment on mobile devices
- Evaluate additional criteria that could affect user's rhythm
- Test in different situations such as at night versus in the morning
- Use more users to build a larger dataset

REFERENCES

[1] Gaines, R. Stockton, William Lisowski, S. James Press, and Norman Shapiro. Authentication by Keystroke Timing: Some Preliminary Results . Santa Monica, CA: RAND Corporation, 1980.

[2] Joyce, R., & Gupta, G. (1990). Identity authentication based on keystroke latencies. Communications of the ACM, 33(2), 168-176. doi:10.1145/75577.75582

[3] Monrose, F., & Rubin, A. D. (2000). Keystroke Dynamics as a Biometric for Authentication. Future Generation Computer Systems, 16(4), 351-359. doi:10.1016/s0167-739x(99)00059-x

[4] Killourhy, K. S., & Maxion, R. A. (2009). Comparing anomaly-detection algorithms for keystroke dynamics. 2009 IEEE/IFIP International Conference on Dependable Systems & Networks. doi:10.1109/dsn.2009.5270346

[5] Shanmugapriya, D., & Padmavathi, G. (2009). A Survey of Biometric keystroke Dynamics: Approaches, Security and Challenges. *(IJCSIS) International Journal of Computer Science and Information Security, 5*(1), 115-119. Retrieved June 5, 2019

[6] Romain Giot, Mohamad El-Abed, Christophe Rosenberger. Keystroke Dynamics Authentication. Biometrics, InTech, chapitre 8, 2011, 978-953-307-618-8. ff10.5772/17064ff. Ffhal-00990373ff

[7] Bours, P., Rosenberger, C., Idrus, S., & Cherrier, E. (2014, June 10). Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords.

[8] Krishnamoorthy, S., Rueda, L., Saad, S., & Elmiligi, H. (2018). Identification of User Behavioral Biometrics for Authentication Using Keystroke Dynamics and Machine Learning. Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications - ICBEA 18. doi:10.1145/3230820.3230829

ACKNOWLEDGEMENTS

The researchers would like to thank Norfolk State University, Kwesi Elliot, Dr. Jonathan Graham, Dr. George Hsieh, and Federal Agency Air Force Research Laboratory COE (Federal Grant Number FA8750-15-2-0120).