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.

**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**


**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).