USER IMPERSONATION DETECTION AND AUTHENTICATION USING IMAGE BASED OTP

A RAJA RAJENDRAN, MUKESHVAR A, GOKUL RAJ, SHARATH KUMAR, RADHIKA G

Department of Computer Science and Engineering
Amrita School of Computing, Coimbatore
Amrita Vishwa Vidyapeetham, India

Abstract:

In this age of internet and technology, it is very important to be vigilant of one's devices and have safe and secure passwords to avoid various types of crimes. In this paper, we concentrate on two types of attacks – User impersonation and shoulder surfing attack. User impersonation is the act where one person pretends to be you/someone else to create confusion or obtain information private to you. To prevent this, we discuss a method to classify a person as authorized/unauthorized, utilizing data from the user's mobile device's accelerometer and gyroscope sensors.One of the simplest methods scammers use to acquire your personal information is by shoulder surfing, which may be as easy as simply looking over your shoulder. It is extremely common and hard to prevent in crowded areas.We proposed a method of authentication which uses OTPs and graphical password techniques, which would prove to be effective against shoulder surfing attacks.

Introduction:

Mobile devices are an integral part of everyone's life as they are being used for nearly everything – from having basic communication with one's friends and family, to using phones for all cashless transactions. The amount of personal data that is being stored in devices is increasing exponentially every year. And being careless in such a world where online presence is everything, poses a huge risk for anyone as they can be an easy target for many crimes. User impersonation is one such crime. The thing is, the attacker does not even have to use complicated techniques for impersonating someone. All they need is our phone. If a person uses your phone, where all the social media apps, mailing apps and websites have been logged in as you, then whatever the person does with the phone will adversely affect you. Many apps and online transactions deploy various techniques like password protection and OTP messages to be more secure. But even password-based authentication systems have loopholes that lead to data breaches and hacks. Hackers can exploit to intercept calls and SMS to gain OTP. Examples of attacks on password are brute force,

phishing, dictionary attack, shoulder surfing etc. One of the most common ways of stealing passwords is a shoulder surfing attack where the attacker can gain information about the victim simply by looking at the screen of their device. It may be as simple as just looking over the victim's shoulders. However, the attackers may go as far as using optical devices like cameras, binoculars etc to peek at the screen of the victim to gain personal information about them. This paper discusses a way of identifying whether the intended user is using the phone or not, and also proposes a method of verification which would solve the issue of shoulder-surfing by verification of OTP which has been generated inside the device itself, thereby reducing risk of call and SMS interception. The proposed method uses graphical password techniques for the verification of OTP.

Literature Survey:

Survey of Various Graphical Password Techniques and Their Schemes proposed in the year 2022, This paper presents an overview of authentication procedures and a survey of multiple graphical password systems, which are divided into different groups.. One of the best methods that can be implemented for authorization in our study's case, is a variation of a recall-based technique called image-pass scheme

Insights from BB-MAS - A Large Dataset for Typing, Gait and Swipes of the Same Person on Desktop, Tablet and Phone proposed in the year of 2022, this paper presents an enormous data collection that includes typing, gait and swiping activities of the same user on desktop, tablet and mobile. in first step of our research for identifying the user as a authenticated user we need behavioural data like and BB-MAS contains many types of data like keystroke features, swipe features and gait feature but we have considered only gyroscope and accelerometer from the data set

Classification of Soft Keyboard Typing Behaviours Using Mobile Device Sensors with Machine Learning proposed in the year 2019, In this paper, a system is proposed for categorising users' typing habits using information generated by built-in sensors. (accelerometer and gyroscope). Various ML algorithms like SVM, k-NN, RFC, ANN etc were applied for classification, and KNN algorithm turned out to be the best to use in this case.

Graphical password: Prevent shoulder-surfing attack using digraph substitution rules proposed in the year 2017, A new technique that hides the mechanism needed to generate password-images is proposed that makes use of digraph-substitution principles. The proposed system uses a combination of pass image technique and play fair cipher encryption technique. This method although a bit complex initially, is very effective against shoulder surfing attacks. A successful login takes an average of 9.67 seconds for participants.and also the login time required by the users decreases rapidly once they learn how to use the system.

Password Generation Based on Song Lyrics and Its Management proposed in the year 2022 showed a new way to devise and use password for verification. These passwords are different from traditional passwords because these passwords can be remembered very easily.

3C-Auth: A New Scheme for Enhancing Security published in 2014 suggests a fool-proof authentication technique using smart card, secret pin, registered fingerprint and registered mobile number of the user.

A Hybrid Scheme for Detecting Fake Accounts in Facebook published in 2019 talks about analysing Facebook accounts to identify fake and impersonating accounts. A hybrid of skin detection and machine learning algorithms has been devised to detect the existence of fake accounts.

A Machine Learning Approach to Cluster the Users of Stack Overflow Forum paper shows how to cluster users of a social forum into 4 different categories. This clustering of users is based on various metrics.

Enhancing Security of One Time Passwords in Online Banking Systems published in 2019 shows ways to make OTPs more secure. It employs a slight modification of the existing OTP scheme by adding fingerprint to it.

Analysis of student satisfaction on virtual learning platforms during covid-19 published in 2022 has details about various ML algorithms, data analysis and data visualization that can be used to classify users.

User behaviour analysis:

In this section we discuss our approaches of making the system decide if a particular user is the authorized user of that device or not. For our experiments, we have used the BB-MAS dataset. This dataset consists of data collected from 117 subjects for typing, gait and touch on various devices, with a total of about 1.7 million data-points for swipes, 3.5 million keystroke events, 57.1 million data-points for accelerometer and gyroscope each. Out of these, we are interested in the accelerometer and gyroscope values collected from phone while the users are typing.

We employed various methods as a part of research to find out which approach and which model would be the best.

Method 1: In this method, we use the data collected from accelerometer and gyroscope separately to come up with a way to classify the user. Since accelerometer and gyroscope contain 3 columns namely – x,y and z-axis value. Just 3 columns with a large number of rows might be insufficient and could lead to overfitting. Therefore, we increase the amount of data by adding some more features, which have been computed by existing features. These

new features were calculated by iterating the data using a fixed window. These new features include min, max, mean auto correlation, standard deviation auto correlation, mean auto covariance, standard deviation auto covariance, skew,mean, standard deviation, variance, kurtosis, squared mean error. After all the necessary pre-processing, the data was then labelled according to the user names. We tried predicting the user name using Logistic regression, Naïve Bayes, K-Nearest Neighbours, Multi-Layer Perceptron. However the performance of the models in this method were not satisfying. Logistic regression and Naïve Bayes gave 40% accurate results, KNN gave 60% accurate results, and the highest accuracy was given by MLP – 66%, which still is too low. Therefore, we devise a new way to classify the users.

Method 2: After all the necessary pre-processing techniques were applied, the data was labelled with binary target variables 0 and 1. The final dataset to be used consisted of 7 columns consisting of x, y and z axis values of accelerometer and gyroscope, and another column having the target variable. One particular user's data was labelled with '1', and all the other user's data were labelled as '0'. The user who is labelled as '1' is the intended user of the device. 4 machine learning models were applied on the data, namely – Logistic regression, Naïve Bayes, KNN and MLP. It was noted that KNN performed the best. This is in agreement with the findings from another paper published in 2019.

After training the models, arbitrary numbers acting as x, y and z axis values of accelerometer and gyroscope were fed as input to the model. Two outputs were possible – 1 or 0. If the output is 1, then we can say that the input values correspond to the behaviour of the intended user of the phone, and hence there is no security risk. If the output is 0, then we can say that the behaviour of the person using the phone is not similar to that of the intended user, and hence security measures are needed.

	Accuracy	Precision	F1score	Recall
lr	0.774295	0.776039	0.771282	0.770053
nb	0.818293	0.817404	0.817700	0.818185
knn	0.917143	0.916470	0.916870	0.917469
mlp	0.907316	0.909729	0.906414	0.904820

Table depicting the performance of the ML models

Using method 2, KNN and MLP give the highest performance. However, we prefer KNN over MLP because MLP is too time consuming. MLP took 8mins to give the results, whereas KNN took mere seconds.

Authentication (Proposed method):

Once we suspect that the person using the phone is not the intended user, we begin the verification stage. In this paper, we suggest a method which would prove effective against shoulder surfing attacks. Shoulder surfing attack is a situation where the attacker can gain information about the victim simply by looking at the screen of their device. It may be as simple as just looking over the victim's shoulders. However, the attackers may go as far as using optical devices like cameras, binoculars etc to peek at the screen of the victim to gain personal information about them.

This verification technique proposed here is an image assisted OTP entering mechanism. For a bystander, the pin entered for access to the phone appears to be completely random and only accessible by having complete knowledge of the registered password and the activity rules.

Registration: The user must choose two photos from a pool of images during registration. These images are important as we will use these images to verify the OTP that we will be generating.

Authentication phase: As soon as the person using the phone is suspected to be an unauthorized user, the following steps take place:

1) A random 4-digit pin is generated by the system. This 4-digit number is split into two parts with 2-digit each.
2) The user is prompted with a 3x3 matrix consisting of 9 images with a randomly generated 2-digit number under each one of the images. Out of these 9 images, one image is an image which the user selected during the registration phase. The number beneath this particular image would contain the first half of the pin generated by the system. The user must remember this number, and click on next
3) A new screen is prompted with a numeric keypad. The user must enter the 2-digit number noted down from the previous step. The numbers are masked as a dots/star as the user types each digit. Now the user must click on next
4) A 3x3 matrix is generated again, this time with another set of 9 images, with a randomly generated 2-digit number under each one of the images. One of the 9 images is the second image that the user selected during registration phase. The number beneath this particular image would contain the second half of the pin generated by the system. The user must remember this number, and click on next
5) A new screen is prompted with a numeric keypad. The user must enter the 2-digit number noted down from the previous step. The numbers are masked as a dots/star as the user types each digit. Now the user must click on next.
6) If and only if the resulting 4-digit number is equal to the 4-dig pin generated initially, the user would be allowed to continue using the device.
7) If the resulting 4-digit number doesn't match the 4-digit pin generated initially, the device is locked, and would have to opened by biometrics.

References:

Belman, Amith K., et al. "Insights from BB-MAS--A Large Dataset for Typing, Gait and Swipes of the Same Person on Desktop, Tablet and Phone." arXiv preprint arXiv:1912.02736 (2019).

Yuksel, Asim Sinan, Fatih Ahmet Senel, and Ibrahim Arda Cankaya. "Classification of soft keyboard typing behaviors using Mobile device sensors with machine learning." Arabian Journal for Science and Engineering 44.4 (2019): 3929-3942.

Yee, Lip, et al. "Graphical password: prevent shoulder-surfing attack using digraph substitution rules." Frontiers of Computer Science 11.6 (2017): 1098-1108.

Abraheem, Amna, Kenz Bozed, and Wafa Eltarhouni. "Survey of Various Graphical Password Techniques and Their Schemes." 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA). IEEE, 2022.

Gutierrez, Edgar, et al. "Analysis of human behavior by mining textual data: current research topics and analytical techniques." Symmetry 13.7 (2021): 1276.

Hari, S. S., Kavinkumar, C., Niketh, G. K., & Harini, N. Enhancing Security of One Time Passwords in Online Banking Systems.

Sai Kishan, E., Hemchudaesh, M., Gowri Shankar, B., Sai Brahadeesh, B., & Jevitha, K. P. (2022). Password Generation Based on Song Lyrics and Its Management. In International Conference on Advances in Computing and Data Sciences (pp. 278-290). Springer, Cham.

Anusha, J., Rekha, V. S., & Sivakumar, P. B. (2015). A machine learning approach to cluster the users of stack overflow forum. In Artificial intelligence and evolutionary algorithms in engineering systems (pp. 411-418). Springer, New Delhi.

Harini, N., & Padmanabhan, T. R. (2016). 3c-auth: A new scheme for enhancing security. Int. J. Netw. Secur, 18(1), 143-150.

Smruthi, M., & Harini, N. (2019). A hybrid scheme for detecting fake accounts in facebook. International Journal of Recent Technology and Engineering (IJRTE), 7(5S3).

Abirami, K. & Radhika, G.. (2022). Analysis of Student Satisfaction on Virtual Learning Platforms During COVID-19. 10.1007/978-981-19-2821-5_47.