Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 7 – No. 36, May 2021 – www.caeaccess.org

On the Effect of Typing Hand on Mobile-based Keystroke Dynamics

Omoyele Akinsowon
Federal University of Technology, Akure
Department of Computer Science,
P.M.B. 704 Akure, Nigeria

Norman Poh
University of Surrey Guildford
Department of Computer Science
GU2 7XH Surrey, UK

ABSTRACT
With the proliferation of mobile devices, access control enabled by biometrics is going to be an indispensable function. In this study, we consider mobile biometrics based on touchscreen based keystrokes. Although the general topic of keystroke dynamics have been well studied, touchscreen based keystrokes remains challenging and relatively new. Few studies evaluate the effect of typing hand on mobile biometrics. We therefore set out to formalize the problem and explore a family of solutions based on keystrokes and typing hand, which can be left hand, right hand, or both hand. We address open questions such as: the effect of typing hand on touchscreen keystrokes, whether or not the knowledge of typing hand is important, and the effect of user-specific score normalisation on the performance.

General Terms
Security, Privacy, Algorithms, Machine Learning

Keywords
Hold Time, Flight Time, Typing Pattern

1. INTRODUCTION
Keystroke Dynamics is a behavioural biometric technology that takes advantage of the unique typing patterns of individuals for identifying them. It is usually used with passwords or PINS which can be spied or compromised. Hence, it serves as a second level authentication medium. This research work aims at using three typing methods for every subject whose keystroke dynamics are obtained in order to reduce the occurrence of False Rejection Rates (FRR) as a result of limited typing patterns being used in some previous research.

1.1. Keystroke Dynamics
Keystroke Dynamics is a biometric technology that uses the typing pattern of an individual to identify him/her. It is the process of measuring and assessing a person's typing rhythm on digital devices [9]. It is a behavioural biometric which takes advantage of the uniqueness of the typing rhythm of individuals to identify or authenticate them. It is one of the cheapest biometric technologies that can be implemented because it doesn't require purchasing or buying new hardware. It is usually incorporated into previously available devices.

The future of this biometric technology is evolving. According to GSMA real-time intelligence data, today, there are 5.26 Billion people that have a mobile device in the world. This means that 67.01% of the world's population has a mobile device. Back in 2017, the number of people with mobile devices was only 53% and breached the 5 billion mark. Statista predicts that by 2023 this number of mobile device users will increase to 7.33 billion.
(Source: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world). Hence, the need for security of these phones cannot be over-emphasized. From the security of the actual phone to the different activities that are run with the phone which range from social, personal, financial, health and educational purposes.

Therefore, in order to ensure the safety of users' lives and property, keystroke dynamics can be used with passwords that most people are already used to. One of the advantages that this technology has over other biometric technologies is its non-intrusiveness. The typing rhythms of people are captured as they type normally, so it does not require them learning a new skill or stretching any part of their body more than they normally would. The power of measuring typing biometrics is that almost every computerized device uses a keyboard input method. Therefore, this shifts the possibility of its widespread use on the software behind the technology rather than the hardware required to use it.

Despite its non-popularity among other biometric technologies because of its tendency to change due to some factors, it has great impact since it can be used in combination with passwords which most people are used to and it is readily acceptable to them. It is of great advantage because passwords, PINs or secret drawing patterns can be forgotten or easily compromised through shoulder surfing or brute force attacks. [7].

1.2. Related Works
Even though Keystroke Dynamics have not developed as maturely and quickly as other biometric traits, some appreciable has been done and many are still on-going on it. For example, Alsawwan et al (2020) studied the use of keystroke analysis to enhance password security which includes biometrics and typing patterns for everyday users of banking applications. They carried out several experiments and compared the results produced. For example, the authors used the toolkit of WEKA based on method and classifier for the first experiment. In the second experiment, the model suggested by the authors went through analysis with classifiers (such as Naïve Bayes) other than the MLP to see how the model reliability stands against classifications under more than one classifier. At the end of their work, they came to a conclusion that using keystroke dynamics analysis to enhance password security on mobile devices proved to have a great chance of success. One of the limitations of the research was that not enough experiments were carried out. Standard keyboards could have been used in order to bring

more successful research results for those keyboards, especially when it is done using other experiment configurations and topologies.

Margit Antal et al (2015) focused on how new features such as pressure or finger area can improve the accuracy of a keystroke-based authentication system. In the course of their study, they used different machine learning classification algorithms and found the best three to be Random forests, Bayesian nets and SVM. Also, EER was computed and Manhattan was the best performing distance function. At the end of their experiments, they were able to demonstrate that touchscreen-based features improve identification and verification using keystroke dynamics. However, the limitations of this work was that some of the typing patterns contained deletions; hence they had to be deleted from the dataset. Those errors could have been used for further study Also, they did not consider the possibility of impostors, making the testing phase incomplete. [2] Kang et al (2008) in their study made use of artificial rhythms and cues. They were able to show that the use of artificial rhythms increased uniqueness while cues increased consistency and both increased discriminality. However, their work was limited in their failure to build classifiers with data using artificial rhythms and cues. The small size of the dataset also limited their work [6].

Gascon et al (2014) implemented a software keyboard for Android that enabled them to collect the typing behaviour of 300 subjects. They recorded the time stamps of the keystrokes on three different events: when the user touches the screen, when the software sends the character to the underlying layer and when the user's finger leaves the screen. They designed a time-based feature extraction method and evaluated the performance of a machine learning classifier in a continuous authentication problem. Even though they captured the motion behaviour of the subjects, it turned out that some of the subjects were not distinguishable with this feature. This therefore shows that the use of those features may not have improved the efficiency of the system after all. [3]

Hwang et al (2009) made use of artificial rhythms and tempo cues and improved on the work by Kang et al (2008) by including impostor typing patterns and increasing the dataset. The future work on their study includes having a more diverse subject population. [5] In the research done by Salem et al (2016), they proposed the use of both timing and non-timing Keystroke Dynamic features with Neural Network learning algorithm for Android touch screen devices. They developed a virtual keyboard for collecting the features and used WEKA toolkit to build the MLP model based on their dataset. They were able to show that Keystroke Dynamics can be used to strengthen authentication. [8]

- typing hand
- mobile keystrokes dynamics as biometrics

Keystroke Dynamics can be implemented with two different strategies. It can either be with static text or dynamic/continuous text. Static text strategy uses single static text-based model where the same word or phrase is used to enrol every subject. This can be used for protecting smart phones and mobile devices to ensure non-intrusion by thieves or fraudsters. For the dynamic/continuous text strategy, different words or phrases are used at enrolment.

This can be implemented on platforms where different people need to gain access, for example on a Learning Management System, Bank transactions and so on.

1.3. Context and objectives of study

This research aims to develop a system that can:
(i) be trained to distinguish different typing methods for enrolled subjects,
(ii) be trained to generalize each typing method in order to recognize them when used by new subjects,
(iii) be trained to identify individuals without knowledge of which typing method has been used; and
(iv) increase in overall accuracy if more data samples are collected
(v) predict a subject's stronger typing hand based on visualizations of the typing methods.

1.4 Contributions and paper organization

In this paper, the possibility of improving passwords by including the features of individuals, i.e. typing patterns to the authentication process, was looked at. This work tried to show that an individual can be identified, no matter the hand he/she uses while attempting to gain access to his/her device. This was possible because all the possibilities of typing methods that could be required as the need may be were put into consideration. Furthermore, it was shown that an individual's typing method can be identified as well as his/her strongest or most-preferred typing method. Two different classifiers were used to train and the results show that KNN is a better classifier compared to LR. The features were also visualised to show the strength of individual subjects' typing method.

This paper began with an introduction of Keystroke Dynamics - how it is used and the challenges it has faced; then went on to review past related works. The objective behind this research was highlighted afterwards and the methodology used for reaching the expected goals our research was outlined and discussed. The different experiments that were carried out were illustrated and the conclusions reached were stated.

2. INTRODUCTION

In order to formalise the problem, a few variables were introduced. Let j denote the identity; h, the typing hand, and x, a feature vector, to be discussed in Section 2.1. The domain of these variables are shown in Table 1.
The key problem of identity inference in mobile-based keystroke dynamics is about verifying if the person is really the identity being claimed, given the observed sample, x. This inference problem can be formally described as finding the probability of having observed user j given the observation x:

$$P(j|x) \quad (1)$$

Unfortunately, in mobile-based keystroke dynamics, due to the small screen, it is fairly common for the user to type using different hands, i.e., left, right, or both hands, i.e {L,R,B}. For example, the user might use one hand to hold and type the password and another hand to hold on to a poll in a crowded, moving bus. Since different typing hands can induce different characteristics, it is reasonable to include the typing hand as a possible side information. Therefore, in order to solve the identity inference problem $P(j|x)$, one must consider the typing hand, i.e., to infer the probability of the user given the

typing method:

$$P(j|h,x) \qquad (2)$$

Table 1: Parameters

Variable	Domain	Meaning
x	$\mathbb{R}^d$	Feature vector representing keystrokes
h	$\{L,R,B\}$	Typing hand: Left hand, Right hand, or Both hands
j	J	An identity from $J \equiv \{1,...,J\}$

Table 2: Problems

No.	Problem	Meaning
1	$P(j\|x)$	Directly infer the identity without any additional knowledge about the typing hand
2	$P(j\|h,x)$	Infer the identity given the typing hand is known
3	Eqn (2)	Infer the identity indirectly using the knowledge of the typing hand, $P(j\|h,x)$
4	$P(h\|x)$	Infer the typing hand, which is needed to solve the identity inference in problem 3 above

This second inference problem is slightly different from the first one because in the second problem, the typing hand is known to the system. In many scenarios, due to the usability issue, it is not desirable or practical to request this information. If the user has to enter not only the password but also the typing hand every time a password is keyed in, this can become very annoying as this can add delay to authentication process.

Therefore, solving the first problem is revisited here. This second inference problem is slightly different from the first one because in the second problem, the typing hand is known to the system. In many scenarios, due to the usability issue, it is not desirable or practical to request this information. If the user has to enter not only the password but also the typing hand every time a password is keyed in, this can become very annoying as this can add delay to authentication process. Therefore, we shall now revisit solving the first problem, P(j | x) not directly but using the solution provided by the second problem P(j | h,x) where the typing hand is known. According to tbe Bayes rule, the inference problem should then be:

$$P(j|x) = \sum_{h\in\{L,R,B\}} P(j|h,x)\ \mathrm{P(h|x)} \qquad (3)$$

where $P(h|x)$ denotes the probability of the typing hand. Because the typing hand is not known or not observed, the Bayes rule suggests that the typing hand be inferred from the observed features x. This leads to a third sub-problem which is one of modelling $P(h|x)$.

To recap, two variations of the identity inference problem in mobile-based keystroke dynamics are discussed, namely (1) directly infer the identity without any additional knowledge about the typing hand, $P(j|x)$; and (2) infer the identity given the knowledge of the typing hand, $P(j|h,x)$. These two problems are not equivalent due to the presence or absence of the knowledge about the typing hand. Indeed, as shown by equation (2), the inference problem is not a single one but consists of several sub-problems due to different typing hands. It is, therefore, reasonable to hypothesize that the second identity inference problem, $P(j|h,x)$ is an easier problem to solve, thus should have fewer recognition errors compared to that of the first one. This is our first hypothesis.

2.1 Features

In order to create a typing template for a person's typing rhythm, different features can be extracted and analysed. This is done majorly by recording the time of pressing or releasing a key. Even though there are a number of these features, care must be taken in acquiring them because they are closely related. Hence, it is very important to choose the right combination of features while experimenting. In fact, some systems where keydown-to-keydown timings and the keyup-to-keyup timings were used together caused confusion in the feature selection methods because of their high timing correlation with each other. [4]

The features used for Keystroke Dynamics are illustrated in figure 1. They are the distinguishing features of each individual's typing pattern. The time stamps generated by each person's press down and release events are recorded in milliseconds and subsequently processed. However, for the purpose of this research, we used the hold time and down time features of users' typing rhythm patterns. Since the other features are derived from these two features, they won't have added additional (or more discriminative) information other than show different ways of presenting the keystroke information to the classifier. Hence, we have not further explored these features.

- Hold time: This is also known as dwell or down time. It is the time interval between pressing a key and releasing it; that is the length of time for which a single key is pressed down.
- Flight time: Also known as the release-to-press time is the amount of time in between releasing the first key and pressing the second, that is the time interval between a key release and the next key press.

However, while the potential for extracting many different features from a user's typing pattern is apparent, most studies focus on the hold time and flight time. [9]

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 7 – No. 36, May 2021 – www.caeaccess.org

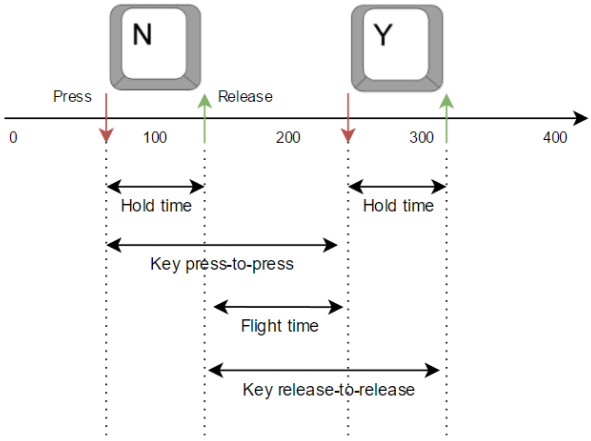


Fig. 1: The different keystroke events for characters N and Y on a timeline

2.2 Data Collection

The process of collecting Keystroke Dynamics is quite different from other biometrics because it does not require new hardware devices; and it is non-invasive. For this research work, the first step was to acquire users' data and static password strategy was used. The password used was decided based on the sparse distribution of the letters in it. Data was obtained from 20 people at the time of performing the experiments. This happened because of the time required to capture data from subjects. Considering time and the fact that Keystroke is a behavioural biometric, all the data could not be taken at once because of the changes that occur in a person's emotions which affect his/her typing dynamics. Hence, data capture was done in three different sessions of around two weeks' intervals in order to have a robust system at the end of the research.

Also, a controlled environment was used; which means that all subjects' data were captured with the same device after installing the soft keyboard prototype for Android OS on it. The Keystroke Dynamics Architecture is illustrated in Fig. 2.

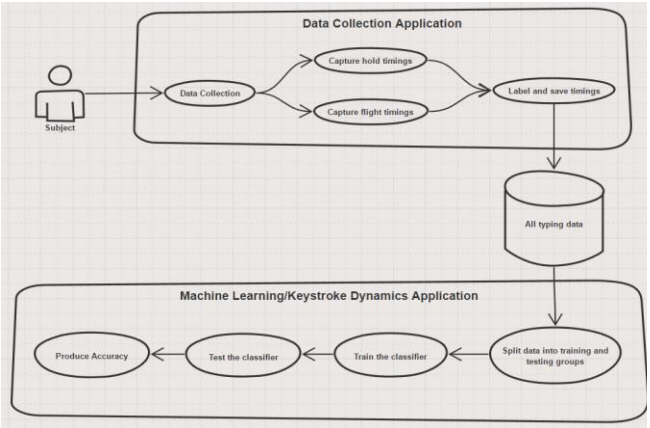


Fig. 2 Design Architecture

2. DATA EXPLORATION

The purpose of this section is to explore quickly whether or not mobile-based keystroke dynamics contain sufficient information to discriminate between different typing hands and also discriminate between subjects. To do so, two different approaches were used, computing the mean keyhold time and the mean flight time as a simple way to visualize the password features entered for each attempt. Another approach is to take the entire feature vector and then use a dimensionality reduction algorithm in order to visualize the data in two dimensions.

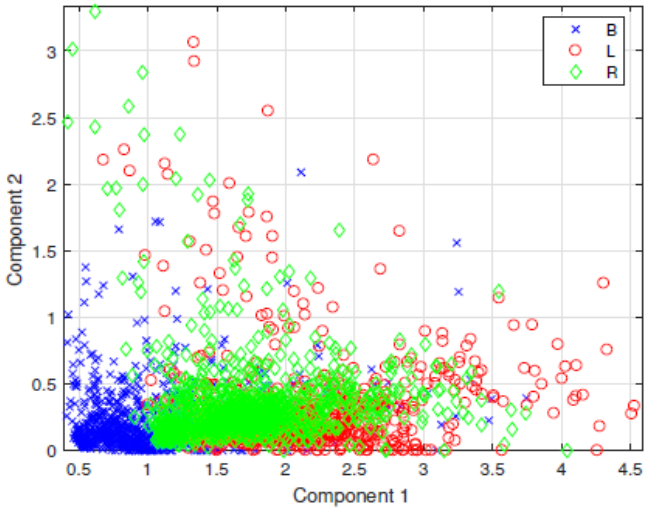
To this end, a nonnegative matrix factorisation (NNMF) which is a linear transformation was used.

3. EXPERIMENTS

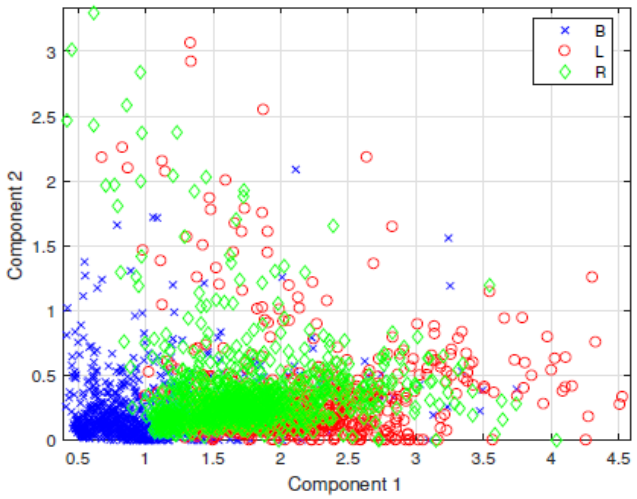
Table 3 shows the results for experiment 3. The corresponding DET curves for each system can be found here: https://goo.gl/AXkZAr.

Table 3 shows EER(%) in classifying typing hand using logistic regression and k-NN.

Hand	LR	k-NN
Left	8.4	14.1
Right	21.1	26.3
Both Hands	10.5	13.2

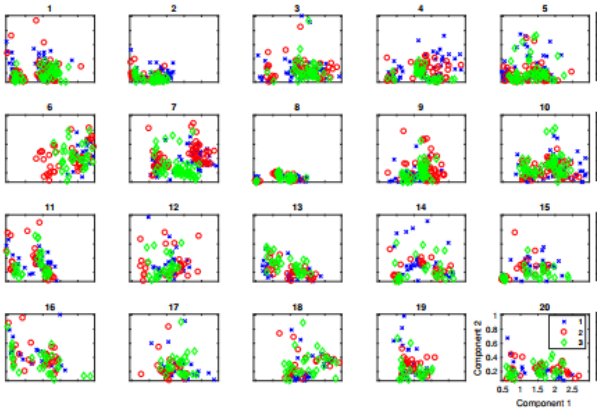


(a) NNMF

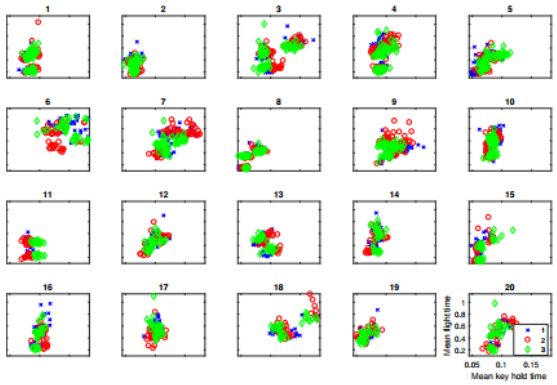


(b) Derived features

Fig 2. Visualise the features using (a) NNMF and (b) the derived features based on the mean flight time versus mean keyhold time



(a) NNMF



(b) Derived features

Figure 3. Visualise the features using (a) NNMF and (b) the derived features based on the mean flight time versus mean keyhold time

4.1 Further Analysis

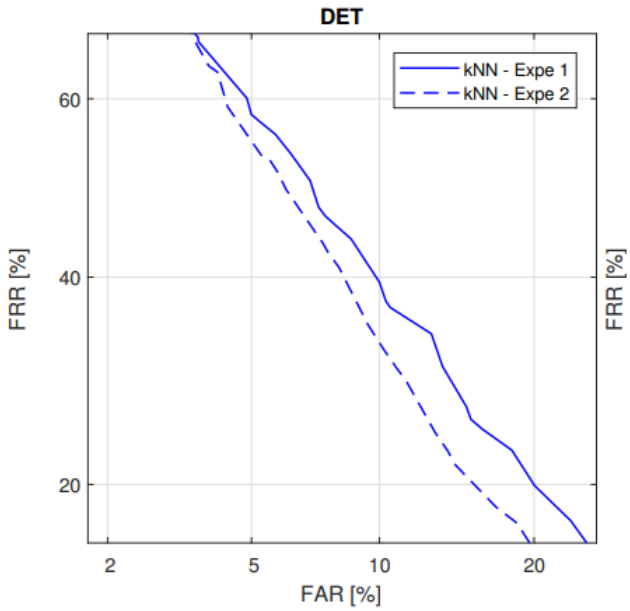
The reader can view how each pair of score distribution are overlayed one subject at a time here. https://youtu.be/jAFBeyhwScc.

4. DISCUSSIONS AND LIMITATIONS

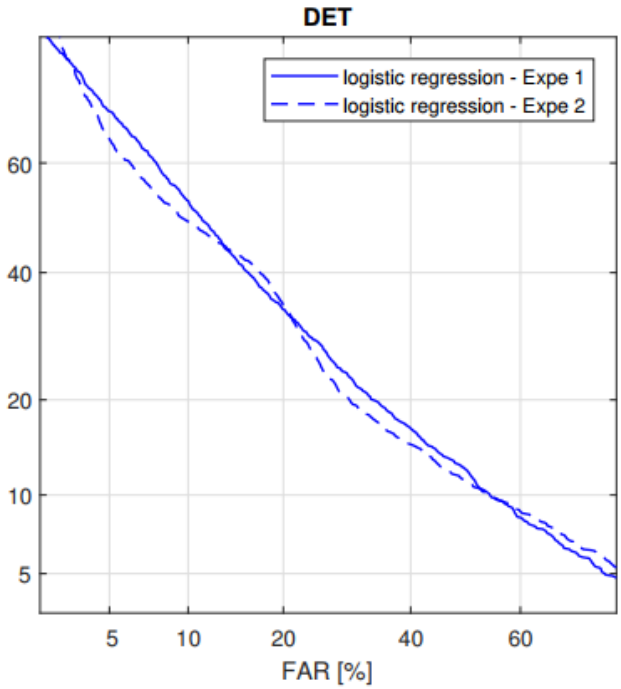
In this paper, analysis of the uniqueness in the way users type was carried out and improved the process of authentication through passwords. The limitation of this work, however is that the same device and the same password were used for every user. Even though the hypotheses that formed the motivation for this work was answered, it is believed that collecting data on different devices and using dynamic passwords would improve the Keystroke Dynamics system. Hence, a future work is proposed here.

5. CONCLUSION

In conclusion, this study showed that individual people have unique typing patterns and that a person's stronger typing hand can be identified. It was also observed that having a larger dataset would improve the accuracy of the system and reduce the occurrence of False Rejection because the emotions of people affect their typing pattern; which means that there must be sufficient typing patterns for an individual in order to reduce the false rejection rate.



(a) NNMF



(b) Derived features

Figure 4. Comparison of performance between solving inference problem $P(j|x)$ and $P(j|h,x)$and using (a) K-NN and (b) logistic regression

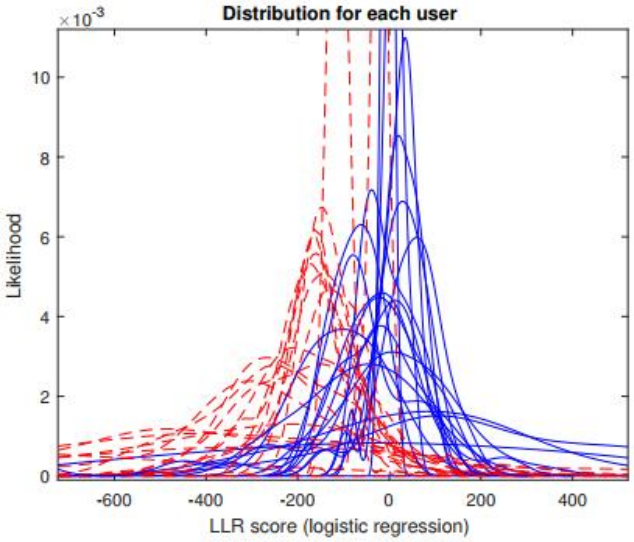


Figure 5. Caption genuine and impostor score distributions for mobile keystroke dynamics experiments. There are 20 subjects here. Blue=genuine score; Red=impostor score distribution

6. REFERENCES

[1] Alsawwan, A, Alrashdan. M, Al-Maatouk. Q, Abdulnabi. M, Indrian. V, Tubishat .M, AlRashdan M.T, Al-Othman A.Z; Keystroke Dynamics Analysis to Enhance Password Security of Mobile Banking Applications; Internationl Journal of Advanced Computer Technology, Volume-IX, Issue-V, December 2020; www.ijact.org

[2] M. Antal, L. Z. Szabo, and I. Laszlo. Keystroke dynamics on android platform. Procedia Technology, 19:820–826, 2015.

[3] H. Gascon, S. Uellenbeck, C. Wolf, and K. Rieck. Continuous authentication on mobile devices by analysis of typing motion behavior. In Sicherheit, pages 1–12, 2014.

[4] F. Halakou. Feature selection in keystroke dynamics authentication systems. In Proceedings of International Conference on Computer, Information Technology and Digital Media (CITaDIM2013), Information Technology & Digital Media Development Center, Ministry of Culture & Islamic Guidance, Tehran, Iran, 2013.

[5] S.-s. Hwang, S. Cho, and S. Park. Keystroke dynamics-based authentication for mobile devices. Computers & Security, 28(1):85-93, 2009.

[6] P. Kang, S. Park, S.-s. Hwang, H.-j. Lee, and S. Cho. Improvement of keystroke data quality through artificial rhythms and cues. computers & security, 27(1):3–11, 2008.

[7] B. S. Saini, N. Kaur, and K. S. Bhatia. Keystroke dynamics for mobile phones: A survey. Indian Journal of Science and Technology, 9(6), 2016.

[8] A. Salem, D. Zaidan, A. Swidan, and R. Saifan. Analysis of strong password using keystroke dynamics authentication in touch screen devices. In Cybersecurity and Cyberforensics Conference (CCC), 2016, pages 15–21. IEEE, 2016.

[9] P. S. Teh, A. B. J. Teoh, and S. Yue. A survey of keystroke dynamics biometrics. The Scientific World Journal, 2013, 2013.

[10] https://www.bankmycell.com/blog/how-many-phones-are-in-the-world).