ENGINEER - Vol. LVIII, No. 01, pp. [109-123], 2025

DOI: https://doi.org/10.4038/engineer.v58i1.7682

Non-obtrusive Imposter Detection During Online Assessments

H.A.A.K.M. Seneviratne, G.I.U.S. Perera and M.S.D. Fernando

Abstract: Cheating during online assessments is a major challenge faced by test administering institutions. Impersonation is the most common fraudulent act. Since impersonation can be resorted to at any stage of an online test, conventional access control at initial login is ineffective. Hence, the need for continuous authentication. Although technology is available to monitor candidates continuously while they take online tests, there are several constraints that limit their deployment. This paper presents a non-obtrusive, easily deployable, and effective continuous authentication system that could detect impersonation during online assessments. It uses a popular class of behavioural biometrics called 'keystroke dynamics' which are basically precision time measurements that can be made latently without the aid of any additional accessories. Authentication is performed using a novel parameter called the 'angle of departure' together with the 'pause durations', both computed from the biometrics captured during the online test and those captured during a mock test conducted as part of the candidate registration process. The underlying concept is presented in detail. Details of experiments conducted to establish proof of the concept are given. These include roleplays conducted for the acquisition of biometrics and simulations of genuine test taking as well as impersonation. The performance of the system is evaluated using both traditional metrics as well as recently established metrics that are considered as being more suitable for assessing continuous authentication systems. Analysis of the results indicates the effectiveness of the use of the angle of departure and pause durations for continuous authentication of candidates during online tests.

Keywords: Keystroke dynamics, Continuous authentication, Online tests

1. Introduction

Educational institutions are burdened with a problem when conducting tests on large numbers of students. This is due to the ever-increasing size of the off-campus student population that follows courses online. The physical resources are inadequate to concurrently accommodate all students for conventional paper and pencil tests.

Online assessments provide a convenient alternative. The institution can deploy the tests online and students can take the tests simultaneously from their comfort zones. However, the integrity of the online testing process is constantly challenged due to the possibility for candidates to engage in fraudulent practices.

The history of cheating at online assessments is as old as online assessment itself. Despite the best efforts by the academic community to counter it, cheating at online assessments remains a major challenge for test administrators [1]. This is primarily due to their inability to keep effective tab on candidates while a test is being administered remotely.

There are many forms of fraudulent practices that are prevalent at online tests. Of these, impersonation is the most common where a competent colleague poses as the registered candidate and takes the test on the latter's behalf. In the process, the candidate shares the login credentials with the imposter voluntarily. This is a scenario not encountered in any other

Eng. H.A.A.K.M. Seneviratne, *C.Eng., MIE(SL), B.Sc.Eng. Hons. (Moratuwa), M.Eng. (Moratuwa), Research Student at Department of Computer Science and Engineering, University of Moratuwa*
Email:athula@ieee.org
ORCID ID: https://orcid.org/0009-0000-8337-1912
Eng. (Prof.)G.I.U.S. Perera, *C.Eng., MIE(SL), B.Sc.Eng.Hons. (Moratuwa), M.Sc. (Moratuwa), MBA (Colombo), Ph.D. (St. Andrews), PGDBM. (Colombo), Professor in Computer Science and Engineering, Department of Computer Science and Engineering, University of Moratuwa*
Email:indika@cse.mrt.ac.lk
ORCID ID: https://orcid.org/0000-0001-5660-248X
Eng. (Prof.)M.S.D. Fernando, *C.Eng., MIE(SL), B.Sc.Eng.Hons. (Moratuwa), M.Phil. (Moratuwa), Ph.D.(TU Delft),Professor in Computer Science and Engineering, Department of Computer Science and Engineering, University of Moratuwa, Dean, Faculty of IT, Horizon Campus*
Email:shantha@cse.mrt.ac.lk
ORCID ID: https://orcid.org/0000-0002-4538-4883

type of access restricted online service. In other online services, the users would keep their access credentials confidential in their own interest, would never collaborate with an imposter, and would cooperate with the online service provider to exercise good access control. The opposite being true for online tests poses an additional challenge to the test administering institutions.

When an imposter takes a test on behalf of a candidate, the candidate could either provide the imposter with the login credentials or handover the session to the imposter after gaining access. Therefore, access control at initial login is not effective in preventing impersonation. The need arises for the 'Continuous Authentication (CA)' of the candidate.

In CA, the session is monitored continuously, watching for cues of any unauthorized access. For an online test, this would mean the continuous verification of the user to ascertain whether it is the registered candidate or someone else who is taking the test.

The most effective CA method is remote proctoring. Technology is available to monitor a candidate remotely throughout the duration of the online test so as to detect impersonation [2]. However, this is not a practicable option because the deployment of accessories required at the candidate's site for remote proctoring would require the cooperation of the candidate which would obviously not be forthcoming from a candidate with the will to cheat. Therefore, in the domain of online testing, there is an acute need for systems that could perform continuous authentication without relying on the candidate's cooperation.

What is described in this paper is a system of authentication which specifically addresses this need. It authenticates the candidate continuously during the test with the candidate not even being aware of it. It relies on 'keystroke dynamics', a biometric that is finding increased use in online access control.

2. Review of Related Work

Authentication factors commonly called 'what you are' factors in access control parlance, are generally regarded as being the most effective. They provide direct indications of 'who the user is' as opposed to 'what the user knows' or 'what the user has' and therefore cannot be easily spoofed by an imposter.

'Keystroke dynamics' is a class of such authentication factors. They refer to the mannerism and traits of a person's typing when using a keyboard. It is a class of 'behavioural biometrics'. Although not as effective as 'physiological biometrics' their use provides several advantages. They can be captured easily. They are 'non- obtrusive' in that they can be acquired without the knowledge of the person and without requiring the person's cooperation.

A very informative source on behavioural biometrics is the work of Yampolskiy and Govindaraju [3]. The authors also present a generalized algorithm that can be used for any type of behavioural biometrics. This algorithm encompasses the concept of creating a profile from the biometrics and using it as a basisof comparison when verifying identity at a later stage. This matches well with the requirements for authentication during online tests.

Hogben [4] has identified five distinct factors that are important for the successful implementation of a behavioural biometric system. The equipment used to capture biometrics is identified as the first. This is because the equipment will determine the non-obtrusiveness of the capturing process more than any other factor. El-Abed et al. have also established through studies that users are more satisfied with keystroke dynamics based systems [5].

The work of Shen Teh et al. [6] includes one of the most comprehensive and earliest descriptions of keystroke dynamics and also provides a detailed description of the 'dwell time' and the 'flight time', the two biometrics that can be captured non-obtrusively.

The work of Giotet al. [7] is one of the earliest that identifies two types of authentications using keystroke dynamics called 'static' and 'dynamic'. When authentication is based on a predetermined sequence of characters, it is 'static authentication'. When authentication is carried out regardless of what is being typed, it is 'dynamic authentication'. The latter is preferred for authentication during online tests.

Giot and Rosenberger [8] have demonstrated the use of keystroke dynamics as a 'soft biometric'. A soft biometric does not provide

absolute recognition of the subject. Instead, it classifies the user into two or more groups. In the said work, the researchers have successfully determined the gender of a user. Idrus et al. [9] have also developed a benchmark testing suite that uses keystroke dynamics in the 'soft biometrics' mode. Soft biometric classification is effective in authentication because the imposter must belong to the same groups that the user belongs in order to evade detection and imposters generally would not have any clues on that.

Keystroke dynamics parameters that are not time related are found to provide better authentication but are more difficult to capture. Nonaka and Kurihara [10] have examined the sensing of pressure exerted on keys for authentication. However, this makes the acquisition of the biometrics visible to the user and are not suitable for authentication during online tests.

Cho et al. [11] have presented a web-based system that exploits keystroke dynamics for authentication. This system includes a registration phase for the extraction of a timing vector, and the training of a machine to build a unique profile for the user.

Killourhy and Maxion have published their findings [12] on the comparative performance of 14 different algorithms that are in industrial use for keystroke dynamics based 'static' authentication. They include Manhattan, Euclidean, Mahalanobis, Neural Networks, and Support Vector Machines.

The above work indicates that the more common use of keystroke dynamics is for authentication at initial access. The use of keystroke dynamics for continuous authentication is a relatively recent research focus.

A paper published by Stylioset al. [13] includes a literature review on the general subject of 'continuous authentication'. It highlights the inadequacy of fixed authentication systems to provide effective protection against intrusion. A significant observation about this work is that there is no explicit reference to 'keystroke dynamics' which indicates that its use in continuous authentication is not common or widespread.

The work of Monrose and Rubin [14] is one of the earliest studies that focuses on the use of 'free text' as opposed to 'fixed text', as sources of keystroke dynamics for authentication. Unlike sequences like passwords that are known to the back end, free text arecontext dependent sequences of characters that are not predefined. The study also concludes that dynamic authentication based on free text is less accurate than that of static authentication using fixed text.

Tappert et al. [15] have published a research paper on the use of keystroke dynamics with long-text input for user authentication. Çeker and Upadhyaya [16] have also conducted research on long-text based authentication using keystroke dynamics. Authentication using long-text input is not the same as continuous authentication. Yet, it is a step in that direction.

The work of Bours and Barghouthi[17]which uses profiles built on the keystroke biometrics acquired from single keys as well as key combinations comes close to what is required for continuous authentication.

The doctoral research of Alsolamy[18] also deals with the use of keystroke dynamics for continuous authentication. In the literature review of the thesis, it is stated that published work on continuous authentication using 'free text' is very limited.

A recent study that has particularly focussed on the use of 'free text' as a source of keystroke dynamics for authentication is the work of Momeni and BabaAli[19]. In addition to the time-related features, they have also used spatial features based on the physical location of the keys on the keyboard. The work also uses convolutional neural networks for classification.

Traditionally, 3 metrics have been used to evaluate the performance of biometric systems, namely, the False Acceptance Rate (FAR), False Rejection Rate (FRR), and the Equal Error Rate (EER)[4]. False acceptance is the admission of an intruder while false rejection is the denial of entry to an authorized user, both being the result of the system failing to perform the authentication correctly.
An important recent realization within the access control industry is that metrics like the FAR and FRR traditionally used to evaluate the performance of access control systems are not effective in evaluating continuous authentication systems. Metrics dealing with

the time that detections are made are gaining increased attention. Senarathet al. [20], in their paper titled 'Re-evaluating Keystroke Dynamics for Continuous Authentication', describe the use of metrics called Time to Correct Reject (TCR), False Acceptance Worst Interval (FAWI), and False Rejection Worst Interval (FRWI) in their experiments. False acceptance intervals are those where a series of consecutive false acceptances are made. FAWI is the maximum of such intervals during a session. False rejection intervals are those where a series of consecutive false rejections are made. FRWI is the maximum of such intervals during a session. These indicators are considered as being more realistic and providing a more relevant indication about the effectiveness of continuous authentication systems.

Careful examination of aforesaid work indicates that continuous authentication using keystroke dynamics still has a lot of uncovered ground which makes it a ripe research area. While keystroke dynamics is having a rich application base in 'fixed authentication' it has still not found its way notably into continuous authentication systems. Also, there is no evidence of its application particularly to authentication of candidates during online tests or about coping with the unique challenge posed by collaboration between registered candidate and the imposter as found during online tests.

This paper presents a system particularly intended for the non-obtrusive continuous authentication of candidates during online tests. It uses a novel parameter based on keystroke dynamics.

3. The Concept

The system presented in this paper for the CA of candidates during online tests is based on a three-pronged concept, which could be presented as follows:

- The authentication should be continuous (i.e. candidate must be authenticated throughout the session and not only at the time of initial login)
- The authentication should be non-obtrusive, should not require the candidate's cooperation, and should not require the deployment of additional accessories at the candidate's site.
- The authentication should rely on unique profiles created for each candidate using biometrics captured during a 'mock test'.

During online tests held subsequently, CA is exercised by continuously monitoring the degree of correlation between the biometrics extracted during the test and the biometrics contained in the profiles.

The choice of 'keystroke dynamics' as the biometric to be used is primarily motivated by the requirements that the authentication should be both non-obtrusive and continuous.

As described in the work of Shen Teh et al. [6], two keystroke dynamics that could be captured non-obtrusively are the 'dwell time' and the 'flight time'. They are pure time measurements and their computation could be understood by considering the timing of the two key events associated with a keystroke, namely, the 'key down' event that occurs when a key is pressed and the 'key up' event that occurs when the key is released, as illustrated in Figure 1.

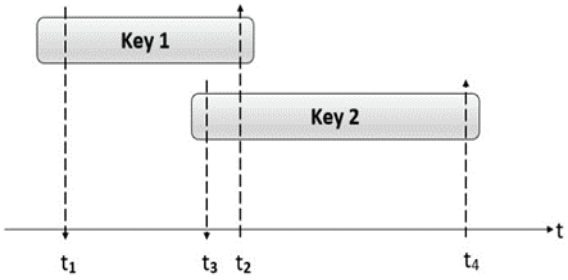

Figure 1 - Time Instances Associated with Two Consecutive Keystrokes

The 'dwell time' for a key press is the time elapsed between the instances of pressing and releasing a key. Accordingly,

Dwell time of key 1 = $t_2 - t_1$

Dwell time of key 2 = $t_4 - t_3$

There are 4 types of 'flight times' between the consecutive presses of two keys. Of these, the type of interest may be understood from Figure 1 as follows:

Flight time of type 1 between keys 1 and 2 = $t_3 - t_2$

Figure 1 also illustrates the particular case where the two keystrokes overlap, making the

flight time of type 1 negative. This is often encountered during fast typing.

Except for the keyboard which is invariably used as the input device, no other hardware or accessories are required for the capture of these two biometrics. Since the keystrokes can be monitored continuously, authentication based on them too could be continuous.

Using the 'dwell time' and the 'flight time', two parameters called the 'angle of departure' and the 'pause duration' are computed for use in the exercise of CA.

The 'angle of departure' will be presented first. It is basically the angle between two multidimensional vectors. There are many possibilities for the composition of such vectors, each of them based on either the 'dwell time' or the 'flight time' of selected key sequences. It is suggested to use as many such vectors as possible. The use of the angle of departure could be better understood in generic terms, without paying attention to the exact composition of these vectors initially.

Let $\boldsymbol{V}$ be one such N-dimensional vector computed from the keystroke dynamics of the registered candidate, extracted during the mock test. Since the keystroke dynamics are unique to the candidate, this N-D vector would also be unique to that candidate. It will map the candidate to a unique point in the N-D space.

This vector would comprise an element of the candidate's profile.

When the candidate takes an online test subsequently, the same keystroke dynamics could be extracted again and an estimate of the vector $\boldsymbol{V}$ could be computed. Let this estimate be $\hat{\boldsymbol{V}}$. Ideally, $\boldsymbol{V}$ and $\hat{\boldsymbol{V}}$ should coincide because both represent the same individual. However, due to the statistical nature of captured biometrics, the two vectors would seldom coincide and $\hat{\boldsymbol{V}}$ would lie within a small cone of uncertainty centered on $\boldsymbol{V}$ as shown in Figure 2.

The angle between the two vectors, defined as the 'angle of departure' would be an indicator of the authenticity of the candidate. If the angle of departure falls within the cone of uncertainty, it is highly probable that the online test was taken by the registered candidate. On the other hand, if it was an imposter who took the test, the poor correlation between the biometrics of the two individuals would make $\hat{\boldsymbol{V}}$ depart significantly from $\boldsymbol{V}$ and be outside the cone of uncertainty, thus indicating the possibility of impersonation having taken place.

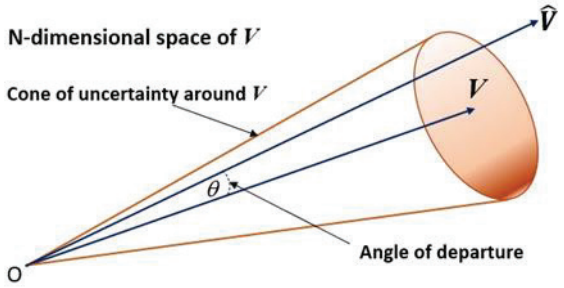

Figure 2 - Cone of Uncertainty and 'Angle of Departure'

This being the basic principle, it could be adapted for CA as follows.

Since it is required to exercise authentication throughout the online test session, the angle of departure could be computed at any chosen instances during the test, say, after every accumulation of M key events. The typing could be monitored while the candidate types in response to the questions in the test and once M key events have been accumulated, the vector $\hat{\boldsymbol{V}}_s$ could be computed, which would be a runtime estimate of vector $\boldsymbol{V}$ for those M key events.

One immediate observation that would be made about $\hat{\boldsymbol{V}}_s$ is that it may not have all the N dimensions possessed by $\boldsymbol{V}$. N would depend on the actual parameter used to compute the vector. As an example, if the vector is computed using the 'mean dwell time' of the alphabetic keys, N would be 26. In general, M key events captured during a part of the online test may not contain sufficient biometrics to define an N-D vector as in the case of $\boldsymbol{V}$, which has been computed using biometrics captured during the entire mock test. Since the spatial orientation of $\boldsymbol{V}$ is influenced by additional dimensions not common to $\hat{\boldsymbol{V}}_s$, it is irrational to use the angle between $\boldsymbol{V}$ and $\hat{\boldsymbol{V}}_s$ for authentication.

Let the dimensionality of the space of $\hat{\boldsymbol{V}}_s$ be n. Since $n \leq N$, in general, $\hat{\boldsymbol{V}}_s$ could be considered to be occupying a 'subspace' of the N-D space of $\boldsymbol{V}$. Therefore, $\boldsymbol{V}$ could be projected to this subspace of $\hat{\boldsymbol{V}}_s$. Let the projection of $\boldsymbol{V}$ onto this subspace be $\boldsymbol{V}_s$.

Since V_s and $\hat{V}_s$ are two vectors sharing a common n-D space, the angle of departure θ between them computed as follows in Equation 1 gives a parameter capable of providing rational authentication.

$$\theta = cos^{-1}\left[\frac{\mathbf{V_s}.\hat{\mathbf{V}}_\mathbf{s}}{|\mathbf{V_s}||\hat{\mathbf{V}}_\mathbf{s}|}\right] \quad \ldots(1)$$

Figure 3 illustrates this computation using the easily comprehended 3D/2D analogy where N=3 and n=2.

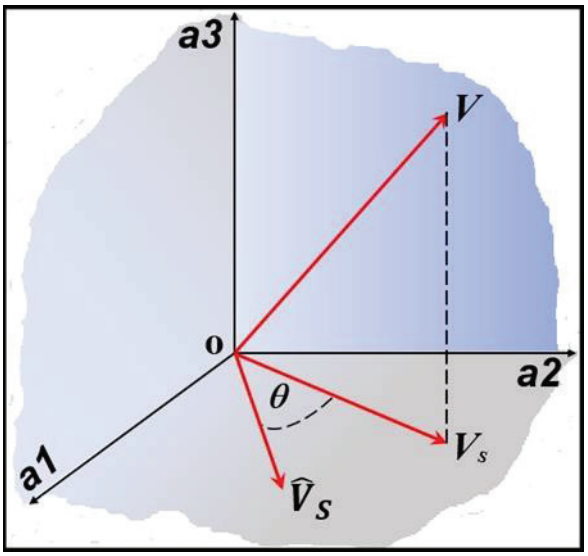

Figure 3 - 3D/2D Analogy for Runtime Computation of 'Angle of Departure'

As the candidate continues to type, there will be a stream of values of θ generated. Authentication could be exercised using θ after every accumulation of M key events, which makes the authentication continuous. Waiting for M key events, computing θ, and exercising authentication using θ would comprise a single cycle of authentication.

Since the keyboard characters that yield M key events would depend on what is being typed, it follows that V_s, $\hat{V}_s$, and the n-D space containing them would generally differ from one authentication cycle to the next. This is illustrated in Figure 4. It will also be noted that the duration of M key events would generally be different.

The other parameter, 'pause durations' is simply a representation of the candidate's pause patterns while typing. A pause is the time the candidate literally pauses while typing. Usually, a typist pauses at the end of each word or more specifically, while typing a whitespace character, namely, the SPACEBAR, the TAB key, or the ENTER key. This pause is useful for the typist to comprehend and reflect upon what has been typed.

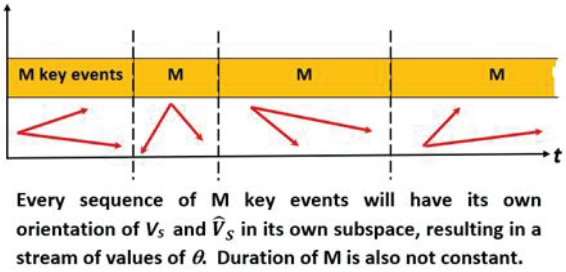

Figure 4 - Continuous Generation of Values of θ while the Candidate Types

It should be clarified that the duration of the pause is not the dwell time of the whitespace key. The pause begins as soon as the key preceding the whitespace key is released. Some typists may immediately press the whitespace key while others may press it during the pause. Therefore, the key up event of the preceding key is a more accurate mark of the beginning of the pause. Also, a typist may release the whitespace key while continuing to pause. The pause actually ends when the key following the whitespace key is pressed. Therefore, the key down event of that key marks the end of the pause.

All the pauses exhibited during the mock test may be extracted from the biometrics captured during the test and stored in the profile as a 1-dimensional array of integer values. However, this array is not an N-D vector like V and is only an ensemble. The order of its elements are of no significance. The ensemble is also not of fixed size because different candidates would exhibit different numbers of pauses during the mock test.

Since the pause patterns are unique to an individual, the ensembles of two individuals would typically exhibit a distinct clustering, as shown in Figure 5.

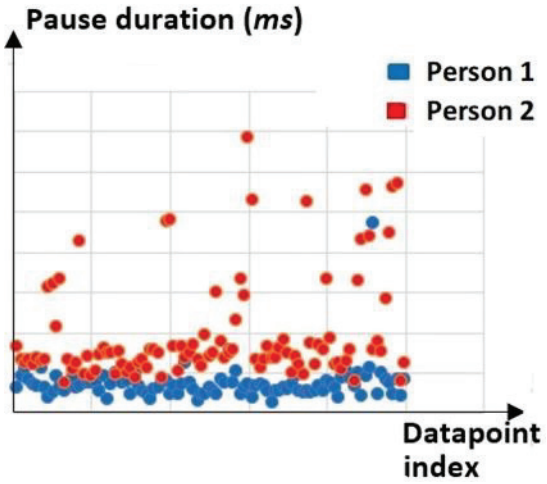

Figure 5 - Distinct Clustering of Pause Durations when Two Persons Type

Therefore, by being sensitive to this difference in clustering, an additional clue about possible impersonation could be elicited. At a subsequent online test, the mean pause duration exhibited during *M* key events may be checked to ascertain whether it falls within the clustering of the candidate's ensemble. A significant deviation from the cluster would signal the possibility of impersonation.

The choice of *M* is critical to the success of the system. Large *M* would lengthen the duration of each authentication cycle, making the authentications infrequent thus affecting its continuity. Making *M* small carries the risk of not being able to acquire sufficient biometrics to compute a vector that is adequately representative of the candidate. Hence, the choice should be based on a tradeoff between the frequency of authentication and the realistic computation of the vectors.

Use of keystroke dynamics based authentication in other online systems is often constrained by the difficulty in creating user profiles. The common practice is to use the keystroke dynamics exhibited while typing the password itself. However, for the effective initial capture of keystroke dynamics, it is necessary to make the prospective user type the password as many as 60 times during registration. In contrast, an educational institution has more opportunities to capture the keystroke dynamics for profile creation because it could use initial assessment tests often conducted at the time of registering students, for this purpose. If necessary, the 'mock test' could even be conducted solely for the acquisition of biometrics without disclosing the intention. This makes an authentication system based on the above concept more feasibly applicable for online tests. The mock test needs to be essentially conducted in a proctored environment after ascertaining the identity of the prospective student.

4. The Experiments

The experiments conducted to establish proof of the above concept comprised a data acquisition phase, a profile creation phase, a development phase and an analysis phase.

4.1 Data Acquisition Phase

During the data acquisition phase, a group of 38 volunteers experienced in taking tests online were deployed to roleplay as candidates. Each roleplayer was required to take two tests, online. Both tests covered topics of general interest, making it possible to take them without requiring any specific knowledge. The tests were deployed via the World Wide Web, which permitted all roleplayers to access them conveniently.

The roleplayers were given the following directions:

i. To take each test in a single sitting.
ii. To essentially use a physical QWERTY keyboard.
iii. To provide plausible answers in the manner they would, had the tests been actual online tests intended to evaluate them.

The first test mimicked the mock test described in the 'Concept'. Although the concept requires this mock test to be conducted physically in a proctored environment, since the test carries no stakes for the roleplayers and therefore there is no incentive for them to indulge in fraud, an online test was used as the mock test for the purpose of the experiments and for the convenience of the roleplayers. Keystroke dynamics acquired during this test were saved under the unique identity of each roleplayer.

The second test mimicked an actual 'online' test. The keystroke dynamics were again captured and saved while each roleplayer was taking this test. Although a test should ideally be concurrent for all candidates, since the tests in the experiments were intended only to acquire the biometrics of the roleplayers and

not to evaluate them, candidates were permitted to take the test asynchronously at their convenience.

The acquisition of biometrics while the candidate types was done using a daemon script made to run on the roleplayer's device. This script captured the timestamps of each keystroke with millisecond accuracy and transmitted them to the server rendering the tests where they were logged in the candidate's profile.

4.2 Profile Creation Phase

During this phase, two vectors were extracted from the keystroke dynamics of each roleplayer captured during the mock test. These two vectors could be understood by referring to the 'dwell time' and the 'flight time' described under the concept.

i. The dwell time vector

This is the vector containing the mean dwell times of the following 51 frequently used keys on the QWERTY keyboard:

All keys of the alphabetical characters	26 keys
All keys of the numeric characters	10 keys
The keys of the special characters in the following set: {SPACE, PERIOD, COMMA, [,], /, \, =, -, semicolon, reverse_accent, TAB, CR, BACKSPACE, SHIFT}	15 keys
Total number of keys =	51 keys

ii. The hot sequence vector

This is the vector containing the mean flight times of type 1 of the following 25 sequences of 2 keys that are frequently encountered in English text:

{ 'IN', 'NG', 'IE', 'ED', 'TH', 'HE', 'IT', 'EI', 'ER', 'AN', 'RE', 'ND', 'ON', 'EN', 'AT', 'OU", 'HA', 'TO', 'OR', 'IS', 'HI', 'ES', PERIOD+SPACE, COMMA+SPACE, PERIOD+CR}

Outlier elimination was performed on both vectors and the vectors were stored in the profile of the respective roleplayer.

In addition to the above two vectors, the 'pause duration ensemble' described in the concept was also extracted from the biometrics acquired during the mock test. Whenever a key down event of any whitespace key was encountered, the timestamp of the key up event of the previous key and the key down event of the key following the whitespace key were captured and their difference was computed. All the pause durations thus computed were pushed into a 1-D array. After the elimination of outliers, this array comprised the 'pause duration ensemble' of the candidate and was stored in the profile.

4.3 Development Phase

During this phase, classifier models were developed to classify the test taking as 'genuine' or as 'fraudulent'.

For each candidate, training datasets were created separately for the angle of departure of the dwell time vector, the angle of departure of the hot sequence vector, and the pause durations.

To create the dataset, an offline simulation of the roleplayers taking the mock test itself was used. This process for a single roleplayer was as follows.

The biometrics file of that roleplayer extracted during the mock test and available in the profile was opened and parsed from beginning to end. For every M key events that were encountered while parsing, the estimates of the dwell time vector and the hot sequence vector were computed from the biometrics available within the M key events. The angle of departure for the two vectors were computed for the M key events using the vectors available in the profile and employing Equation (1). These angle datapoints were stored separately in arrays. Since these angles were derived from a simulation where the biometrics and the profile were of the same roleplayer, it was reckoned as a simulation of genuine test taking and the datapoints were labeled as 'genuine'.

Since a training dataset should be balanced, fraudulent test taking was also simulated by using the profile vectors of one roleplayer and parsing the biometrics file of another roleplayer. For each roleplayer, the biometrics of each other roleplayer were taken in turn and parsed, calculating the angle of departure for the two vectors for every collection of M key

events. These data points were labeled as 'fraudulent' and stored separately, giving 37 sets of fraudulent datasets for each of the 38 roleplayers.

From the 37 fraudulent datasets, 5 sets having the smallest variance with the genuine dataset were identified. Points randomly selected from these 5 sets and the genuine dataset were used to make a balanced composite dataset having equal numbers of genuine and fraudulent datapoints, for each candidate,.

The rationale for selecting the 5 fraudulent datasets with the minimum variance from the genuine data was that a classifier that is sensitive to smaller differences between genuine and fraudulent datasets will perform better since it has learnt from the worst case data representing smaller deviations in the angle of departure.

For the pause durations, the ensembles extracted during the mock test themselves were used as datapoints. Genuine and fraudulent datasets were generated in the same manner as done for the vector based datasets.

In section 3, the criticality of selecting M was explained. Accordingly, after a series of trial runs, M was chosen to be 40 key events so that the CA would happen at the end of every 40 key events. These resulted in datasets of 200 to 300 datapoints for both θ and pause durations.

From the basic dataset, 3 more variants were generated using linear transformations.

Since the datasets were comparatively small in size, the following 4 types of classifiers were selected due to their ability to be trained and tested using relatively small datasets.

i. Naïve-Bayes classifier
ii. Random forest classifier
iii. Support Vector Machine classifier
iv. Logistic Regression classifier

Other factors considered when making this choice were the type of training data, storage, and processing overheads [21]. These factors will be critical in an actual implementation where a large number of candidates will be authenticated simultaneously. For each roleplayer, for each of the 3 features, with 4 variants of labelled datasets and 4 types of classifiers, it was possible to train 16 classifiers through supervised learning.

For a given feature, it was necessary to select the best classifier. This was done by simulating a run of genuine taking of the mock test for each roleplayer while authenticating using each of the 16 classifiers in turn. The classifier that gives the lowest False Rejection Rate (FRR) being the classifier, that is least discriminating towards the roleplayer was chosen as the best classifier.

The rationale for selecting the least discriminating classifier as the best classifier is directly related to the uniqueness associated with authenticating during online tests. Generally, access restricted online systems use a zero tolerance policy where the detection of an unauthorized access would immediately trigger denial of service. It is not possible to adopt this policy during online tests because acting merely on a single false rejection would result in irreparable damage to a genuine candidate. It is therefore necessary to give the candidate some benefit of the doubt and to favour low FRR over high rate of imposter detections.

4.4 Analysis Phase

During the analysis phase, genuine and fraudulent test taking of the second test were simulated offline as follows:

i. The profile of each roleplayer was selected in turn.
ii. Taking of the second test by all players were monitored in turn using the best machine of the player under experimentation.
iii. The performance metrics were computed and saved.

Figure 6 shows the algorithm used for the simulation of taking of the second test with CA caried out using the dwell time related features and the best classifier for that feature.

It will be noted that this algorithm is generic in that it can be used to simulate both genuine and fraudulent test taking. Genuine test taking is simulated when player1 and player2 are the same.

```
player_1 ← command line argument 1
player_2 ← command line argument 2

Retrieve dwell_time vector of player_1 as mu_dwell[]
Retrieve best classifier of player_1 for dwell time features as
model()
Open JSON file of test 02 of player_2 as dictionary array
                                                    key_events[]

genuine ← TRUE
point to first dictionary element of key_events[]
while end of key_events[] is not encountered:
    create blank matrix mu_dwell_hat[] with structure similar to
            mu_dwell[]
    read M consecutive key events from key_events[]
    compute mu_dwell_hat[] of the N keys within the M events
    determine subspace of mu_dwell_hat[]
    project mu_dwell[] to subspace of mu_dwell_hat[]
    theta ← angle between mu_dwell_hat[] and projected
            mu_dwell[]
    genuine ← model.predict(theta)
    advance pointer of key_events[] by M dictionary elements

close all files
```

Figure 6 - Algorithm for Authentication using Dwell Time Related Features

All 3 best classifiers were employed concurrently during each simulation and the overall classification was taken as the Boolean OR combination of the 3 independent classifications. With 38 roleplayers, there were simulations of 38 genuine test takes and a total of 1406 (38*37) fraudulent test takes.

The classifications made during each simulation were recorded in arrays and were analyzed using the following metrics:

i. False Rejection Rate (FRR)
ii. False Acceptance Rate (FAR)
iii. Time to First Detection (T2FD)
iv. False Acceptance Worst Interval (FAWI)
v. False Rejection Worst Interval (FRWI)

5. Interpretation of Results

For the CA system to be good, during a known case of genuine test taking, false rejection of the genuine candidate should happen very infrequently. Therefore low FRR indicates good authentication. False rejection intervals are periods during the test where the CA system makes a series of consecutive false rejections of the genuine candidate. The FRWI is the maximum of such intervals. FRWI being low indicates that the CA system is quick to revert to true detections after an occasional false rejection.

Figure 7 shows the distribution of FRR and FRWI across the 38 roleplayers when they take the second test genuinely.

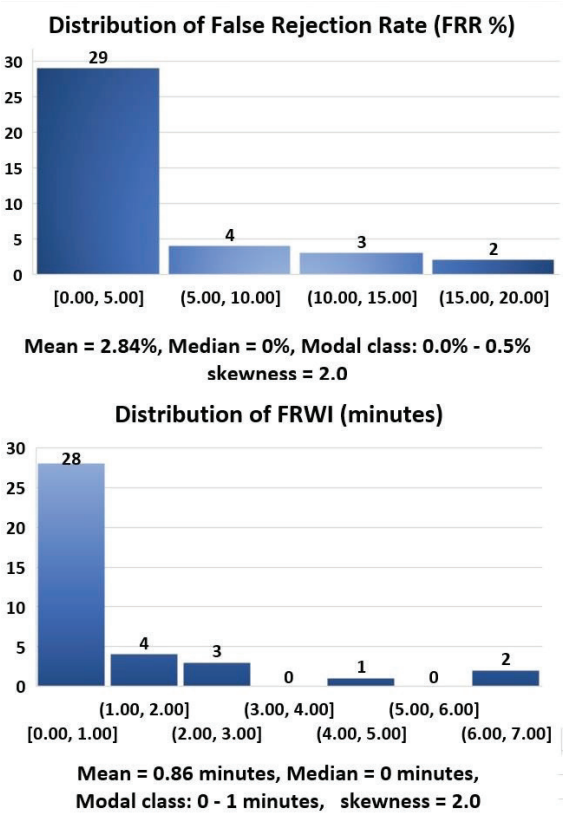

Figure 7 - Performance Metrics for Genuine Test taking by 38 roleplayers

It will be seen that the FRR is less than 5% for 29 roleplayers while all 38 roleplayers exhibit an FRR less than 20%. It will also be seen from the distribution of FRWI that in the occasional event of a false rejection, it does not persist for longer than 1 minute in the case of 28 roleplayers.

When the two metrics are considered jointly, it indicates that the CA system has performed well as far as 'non-rejection of genuine candidates' is concerned, which is important in online tests as explained in the concept.

In a known case of impersonation, how soon the CA system detects is a good indicator of the effectiveness of the authentication. Figure 8 shows the distribution of the average time taken for the first imposter detection (Time to First Detection or T2FD) across the 38 roleplayers when each of them were impersonated by the remaining 37 roleplayers.

It will be seen that in the case of 32 roleplayers, on average, the first imposter detection has

been realized within 0.62 minutes (37 seconds) of starting the test and it has been realized within 0.32 minutes (19 seconds) in the case of 16 roleplayers, This indicates that the CA system has performed well as far as its promptness in imposter detection is concerned.

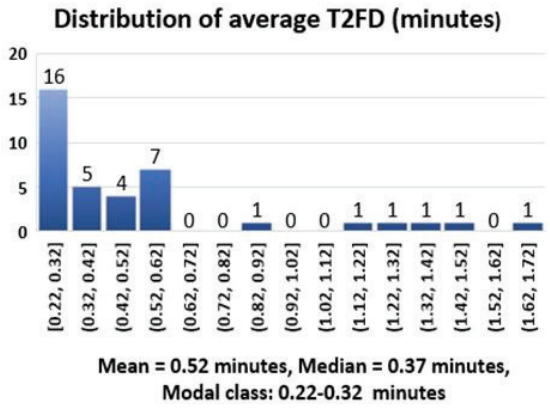

Figure 8 - Distribution of Average Time Taken for the First Imposter Detection

For the CA system to be good, during a known case of impersonation, false acceptance of the imposter as a genuine candidate should be very infrequent. Therefore low FAR indicates good authentication. False acceptance intervals are periods during the test where the CA system makes a series of consecutive false acceptances of the imposter. The FAWI is the maximum of such intervals. FAWI being low indicates that the CA system is quick to revert to true imposter detections after an occasional false acceptance.

Figure 9 shows the distribution of the average FAR and average FAWI across the 38 roleplayers when impersonated by each of the other 37 roleplayers.

It will be seen that the average FAR exhibits a normal distribution with a positive skew around a mean value of 33.0%. There are 23 roleplayers whose average imposter detection accuracy is greater than 68.0 % while there are 28 roleplayers with detection accuracy greater than 58.0 %.

The average FAWI exhibits a normal distribution with a high positive skew around a mean value of 3 minutes. On average, in the case of 30 roleplayers, imposters have not been able to evade detection for longer than 4 minutes while in the case of 35 roleplayers, imposters have not been able to evade detection for longer than 6 minutes.

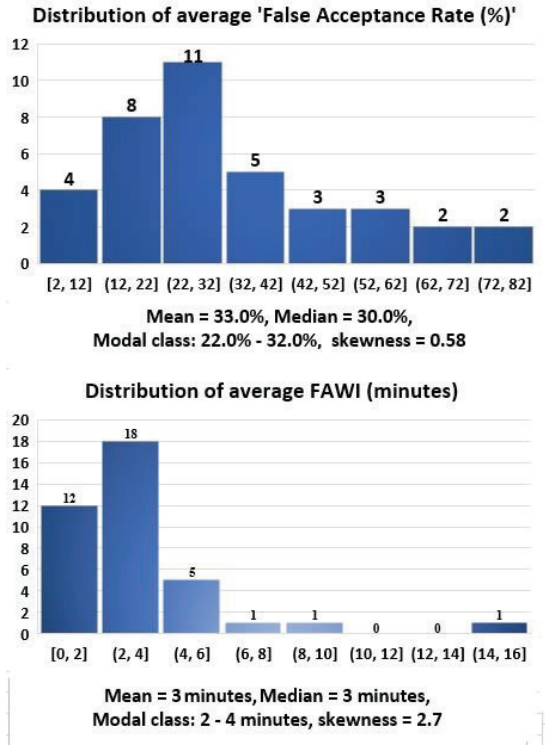

Figure 9 - Average Performance Metrics for Impersonation of 38 Roleplayers

Figure 10 shows the distribution of the minimum FAR across the 38 roleplayers when impersonated by each of the other 37 roleplayers. It will be noted that 24 roleplayers have witnessed a minimum FAR of 6% or less.

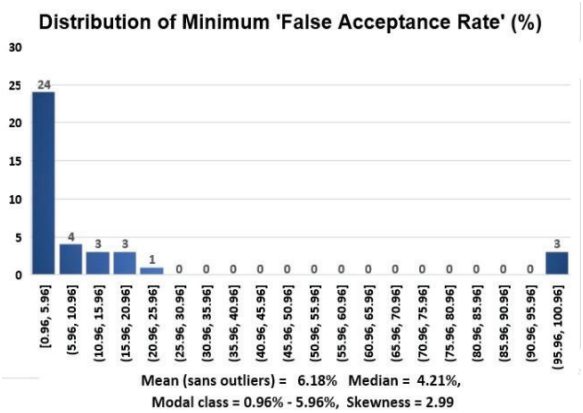

Figure 10 - Distribution of Minimum FAR During Impersonation of 38 Roleplayers

When these metrics are considered jointly, it indicates that the CA system has performed well as far as 'detection of imposters' is concerned.

Accordingly, the results of the experiments show that the CA system has performed well in all respects and therefore establishes proof of the concept presented in section'3' for a CA system for use in online tests.

6. Discussion

In the analysis, the time related metrics T2FD, FAWI, and FRWI were also used for assessment without being content with the performance indications made by the two metrics FRR and FAR, because the system is one of continuous authentication. In continuous authentication systems, since the performance metrics computed at the end of the session does not provide a complete nor useful indication of its effectiveness, the time related metrics must also be used. The mapping of performance onto the time axis is very important. The system under experimentation indicated good performance in continuous authentication.

At present no formal industry benchmarks have been established for keystroke dynamics based continuous authentication systems for both the traditional and time related metrics jointly. In systems using only fixed authentication and where collaboration between imposter and genuine user is not likely, FAR less than 9% have been achieved. One such study is the work of Piet et al. [22] which has employed 'deep learning' and achieved an FAR of 8.7% but no information is available about the time related metrics.

The work of Murphy et al. [23] could be considered as a performance benchmark for reasonable comparison for this study. Apart from the use of 'free text' and simulation of continuous authentication, the two works differ substantially in several aspects. For data collection, they have used key logging during normal computer interaction by the subjects. Subject/machine binding where the subject essentially uses the same keyboard has been enforced. Large volumes of data have been gathered from 103 subjects. Key sequences of 1000 strokes have been considered for continuous authentication. Performance has been optimized to achieve Equal Error Rate (EER). Performance assessment has been based on traditional metrics and a false acceptance rate of 10.36% has been achieved. A comparison is provided in Table 1.

The system used in the experiments while having a relatively high FAR, has a FAWI distribution of a high positive skew, indicating the system's capability to quickly regain correct imposter detection amidst the relatively frequent false acceptances. In addition, it particularly focused on CA during online tests.

Table 1 – Performance Comparison

	Benchmark system	**System used in this study**
Number of subjects	103	38
Average size of biometric data per subject	12.9 M keystrokes	233 kBytes
Sequence length	1000 keystrokes	40 key events
Acquisition environment	Uncontrolled	Controlled to mimic online tests
Biometrics	n-graphs	Monographs, di-graphs and pauses
Metrics	Traditional	Both traditional and time dependent
Optimization	EER	Low FRR
Accuracy	FAR of 10.36%	FAR of 33%

In absolute terms, an ensemble mean value of 33% for the average FAR may not be adequate for assessments involving high stakes. However, this should be considered together with the distribution of the minimum FAR where a minimum of 6% or less have been witnessed in the case of 24 roleplayers. Also the positive skew in the distribution of minimum FAR is more profound than that of the average FAR. Therefore they provide substantial proof of the concept.

The experiments used a mock test solely for the purpose of acquiring biometrics for profile creation. The typed responses to the mock test have a direct bearing on the biometrics captured. To create profiles that are highly representative of the candidates, the abundance and diversity of keystroke dynamics must be ensured. This requires the careful crafting of the test questions that induce the candidate to display particular keystroke dynamics that lend more effectively towards capturing more representative feature vectors. For example, answers requiring the frequent usage of past tense verbs would cause a high occurrence of the key sequence 'ing'. The tests used in the experiments which were constructed without prior knowledge of the nature of the resulting biometrics may have produced biometrics that yielded profiles which are somewhat less

representative. This could account for the relatively high FAR.

The crafting of test items for the creation of more representative profiles is suggested for future studies.

As already mentioned, conduction of a mock test is a feasible step within the workflow of an educational institution for candidate registration. The concept exploits this advantage which may not be available in other access restricted online systems. Since only keystroke dynamics were used, the system was non-obtrusive and was inconspicuous to the candidate. The cooperation of the candidate was neither sought nor required.

Use of the novel feature of the 'angle of departure' between profiled and runtime vectors turned out to be very effective. The experiments demonstrated the effective application of the concept using two vectors, namely, the dwell time vector and the 'hot sequence vector'. The concept can be extended to an unlimited number of vectors that could be reliably extracted from keystroke dynamics. An example would be the 'far sequences', being two-character sequences that are typed with the forefinger and the little finger of each hand. In doing so, the hand muscles will be stretched to the maximum, bringing in a trace of person dependence to the finger movements. Therefore the person dependence of the flight time of such sequences is likely to be high. There are 72 such sequences as per the standard finger assignment in touch typing. A vector of 72 dimensions could be extracted. Attempts to extract this vector during the experiments were unsuccessful because the presence of the far sequences was too scarce within the biometrics captured during the mock test. To obtain sufficient biometrics to define a vector of such dimensionality, it would be necessary to tailor the mock test as described earlier, to cause the candidate to type all the far sequences in the process of providing answers to the test. This is suggested for future work.

The experiments also demonstrated the use of pause durations for continuous authentication which is only possible with long, free text inputs like what are encountered during online tests.

Only 4 types of classifiers were used in the experiments. However the concept does not limit the use of classifiers to these 4 types. Other classifiers that perform well with small datasets and do not pose large storage or processing overheads may be tried out. Since machine learning is a fast evolving field of study, the possibility exists for trying out many other classifiers. This would provide a larger number of classifiers for each feature which would expand the scope of selection of the 'best classifier' and consequently would make the CA system more robust. This too is intended to inspire future work.

As in the case of all behavioral biometrics, keystroke dynamics are also subject to time variance. The profile of a candidate may gradually lose its validity as the candidate's typing behavior alters due to aging and experience. Therefore it is necessary to update the profiles regularly. This could be done either by repeating the mock tests regularly or by using the online tests themselves. Different update strategies are used in other access restricted online services and these could be adopted in actual implementations of the concept. The short term variance due to fatigue may also affect the keystroke dynamics. This is possible when candidates undergo a series of tests over a short period of time. Mitigation strategies available in other access restricted online services could be adopted for this as well. The study of these are also suggested as future work.

The use of keystroke dynamics as a soft biometric was noted in the review of related work. By detecting attributes like the gender, age, and dextrality of the candidate, additional layers of authentication can be established. These were not incorporated in the present study because the ethical clearance granted for the research required the anonymity of the roleplayers to be ensured and therefore did not permit the collection of additional data about the roleplayers that could lead to their identity. Therefore expanding the detection capability of the presented system through soft biometrics too is intended to inspire future work.

7. Conclusion

This paper presented a non-obtrusive continuous authentication system that uses only the 'dwell time' and the 'flight time' of keystrokes made by a candidate while taking an online test. It is intended to fulfil the need for a non-obtrusive imposter detection system during online tests. The system was centered on a profile created from the keystroke dynamics acquired during a mock test. It used a novel feature for authentication. Supervised learning was used to train the classifiers. The successful use of the system for detecting imposters was demonstrated using simulations of tests taken by 38 roleplayers. The performance of the system was assessed using traditional as well as more recently established time related metrics. The results showed that the system has good imposter detection capability which could be finetuned in various ways. These include the better crafting of the mock test, use of more features, use of soft biometrics, and trying out other types of classifiers. Several suggestions for future work were also made.

Acknowledgement

The authors wish to acknowledge with thanks, the commitment of all the roleplayers who took part in the experiments which was critical to the success of the research.

References

1. Young, J. R., "Cheating Goes High-Tech," *The Chronicle of Higher Education*, Jun. 2012.
2. Vincent, D., "Promoting Academic Integrity in Assessment in Online Distance Learning," Sep. 2013.
3. Yampolskiy, R. V. and Govindaraju, V., "Behavioural Biometrics: a Survey and Classification," *International Journal of Biometrics*, Vol. 01, No. 01, 2008.
4. Hogben, G., "ENISA Briefing: Behavioural Biometrics," European Network and Information Security Agency, Feb. 2010. Available: www.enisa.europa.eu/publications/behavioural-biometrics
5. El-Abed, M., Giot, R., Hemery, B., and Rosenberger, C., "A Study of Users' Acceptance and Satisfaction of Biometric Systems," in *2010 IEEE International Carnahan Conference on Security Technology (ICCST), 5-8 Oct. 2010*, IEEE, Nov. 2010, pp. 170–178.
6. Shen The, P., Beng Jin Teoh, A., and Yue, S., "A Survey of Keystroke Dynamics Biometrics," *The Scientific World Journal*, Vol. 2013, Nov. 2013, Available: https://onlinelibrary.wiley.com/doi/10.1155/2013/408280.
7. Giot, R., El-Abed M., and Rosenberger, C., "Keystroke Dynamics Overview," *Biometrics*, pp. 157–182, Jun. 2011, Available: https://www.researchgate.net/publication/221912833.
8. Giot, R. and Rosenberger, C., "A New Soft Biometric Approach for Keystroke Dynamics Based on Gender Recognition," *International Journal of Information Technology Management*, Jan. 2012.
9. Idrus, S. Z. S., Cherrier, E., Rosenberger, C., and Bours, P., "Soft Biometrics Database: a Benchmark for Keystroke Dynamics Biometric Systems," in *International Conference of the Biometrics Special Interest Group (BIOSIG), 2013*, Jan. 2013.
10. Nonaka, H. and Kurihara, M., "Sensing Pressure for Authentication System Using Keystroke Dynamics," in International *Conference on Computational Intelligence, ICCI 2004, December 17-19, 2004, Istanbul, Turkey*, Dec. 2004.
11. Cho, S., Han, C., Han, D. H., and Kim, H-I1., "Web based Keystroke Dynamics Identity Verification Using Neural Network," *Journal of Organizational Computing and Electronic Commerce*, Vol. 10, No. 4, pp. 295–307, 2000.
12. Killourhy, K. S. and Maxion, R. A., "Comparing Anomaly-Detection Algorithms for Keystroke Dynamics," in *IEEE/IFIP International Conference on Dependable Systems & Networks, 2009*, IEEE, Aug. 2009.
13. Stylios, I. C., Thanou, O., Androulidakis, I., and Zaitseva, E., "A Review of Continuous Authentication Using Behavioral Biometrics," in *Proceedings of the South East European Design Automation, Computer Engineering, Computer Networks and Social Media Conference, SEEDA-CECNSM 2016, Kastoria, Greece, September 25-27, 2016.*, Sep. 2016.
14. Monrose, F. and Rubin, A. D., "Keystroke Dynamics as a Biometric for Authentication," *Future Generation Computer Systems*, Vol. 16, No. 04, pp. 351–359, Feb.2000.

15. Tappert, C. C., Villani, M., and Cha, S.-H., "Keystroke Biometric Identification and Authentication on Long-Text Input", *Behavioral Biometrics for Human Identification: Intelligent Applications*, pp. 342–367, Jan. 2009, Available: https://www.researchgate.net/publication/254243762_Keystroke_Biometric_Identification_and_Authentication_on_Long-Text_Input.

16. Çeker, H. and Upadhyaya, S., "User Authentication with Keystroke Dynamics in Long-text Data," *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Niagara Falls, NY, USA, 2016, pp. 1-6.

17. Bours, P. and Barghouthi, H., "Continuous Authentication using Biometric Keystroke Dynamics," in *The Norwegian Information Security Conference (NISK) 2009*, 2009.

18. Alsolamy, E., "An examination of keystroke dynamics for continuous authentication," Doctoral Thesis, Information Security Institute, Science & Engineering Faculty, Queensland University of Technology, 2012.

19. Momeni, S.& BabaAli, b. (2024). "Free-text Keystroke Authentication using Transformers: A Comparative Study of Architectures and Loss Functions". 10.21203/rs.3.rs-3960768/v1.

20. Senarath, D., Tharinda, S., Vishvajith, M., Rasnayaka, S., Wickramanayake, S., and Meedeniya, D., "Re-evaluating Keystroke Dynamics for Continuous Authentication," *2023 3rd International Conference on Advanced Research in Computing (ICARC)*, Belihuloya, Sri Lanka, 2023, pp. 202-207.

21. *Choosing the Best Machine Learning Classification Model and Avoiding Overfitting*, eBook. Chapter 1. MathWorks. Accessed: Apr. 26, 2024. [Online]. Available: https://www.mathworks.com/campaigns/offers/next/choosing-the-best-machine-learning-classification-model-and-avoiding-overfitting.html.

22. J. Piet, A. Sharma, V. Paxson, D. Wagner, "Network Detection of Interactive SSH Impostors Using Deep Learning" *2023 32nd USENIX Security Symposium. August 2023*, CA, USA, pp. 4283-4300.

23. Murphy, C., Huang, J., Hou, D., and Schuckers, S, "Shared Dataset on Natural Human-Computer Interaction to Support Continuous Authentication Research," *2017 IEEE International Joint Conference on Biometrics (IJCB)*, Denver, CO, USA, 2017, pp. 525-530.