Journal of Pattern Recognition Research 7 (2012) 116-139

Received April 28, 2012. Revised July 10, 2012. Accepted July 1, 2012.

Biometric Authentication and Identification using Keystroke Dynamics: A Survey

Salil P. Banerjee *salilb@g.clemson.edu*
Electrical and Computer Engineering Dept., Clemson University,
105 Riggs Hall, S.C. 29634, USA

Damon L. Woodard *woodard@clemson.edu*
Human-Centered Computing Division,
School of Computing, Clemson University,
100 McAdams Hall, S.C. 29634, USA

Abstract

Dependence on computers to store and process sensitive information has made it necessary to secure them from intruders. A behavioral biometric such as keystroke dynamics which makes use of the typing cadence of an individual can be used to strengthen existing security techniques effectively and cheaply. Due to the ballistic (semi-autonomous) nature of the typing behavior it is difficult to impersonate, making it useful as a biometric. Therefore in this paper, we provide a basic background of the psychological basis behind the use of keystroke dynamics. We also discuss the data acquisition methods, approaches and the performance of the methods used by researchers on standard computer keyboards. In this survey, we find that the use and acceptance of this biometric could be increased by development of standardized databases, assignment of nomenclature for features, development of common data interchange formats, establishment of protocols for evaluating methods, and resolution of privacy issues.

Keywords: Authentication, biometrics, cyber-security, identification, keystroke dynamics, psychology, remote monitoring, typing

1. Introduction

Computers have become an ubiquitous part of the modern society. In early 2011, online attacks on companies resulted in the shutdown of their networks and compromised the passwords and personal information of millions of users. Since we depend so much on computers to store and process sensitive information, it has become all the more necessary to secure them from intruders. For user authentication and identification in computer based applications, there is a need for simple, low-cost and unobtrusive device. A user can be defined as a person who attempts to access information stored on the computer or online using standard input device such as the keyboard. Use of biometrics such as face, fingerprints and signature requires additional tools to acquire the biometric which leads to an increase in costs. Use of a behavioral biometric which makes use of the typing pattern of an individual can be obtained using existing systems such as the standard keyboard, making it an inexpensive and extremely attractive technique. One of the major advantages of this biometric is that it is non-intrusive and can be applied covertly to augment existing cyber-security systems.

In the mid-19th century when the telegraph was used extensively it was observed that telegraph operators could identify other operators based on their typing rhythm. During World War II, a methodology known as the 'Fist of the Sender' was used to identify the sender of the telegraph by using the rhythm, pace and syncopation of the telegraph keys [43, 153, 52]. Military troop movements could be tracked using this technology to determine the operators [18]. In the early '80s the National Science Foundation and the National Bureau of Standards in the United States conducted studies establishing that typing patterns contain unique characteristics that can be identified [42]. Shaffer [139] showed that typing is a motor programmed skill and that movements are organized prior to their actual execution. Therefore, a person's typing pattern is a behavioral characteristic that develops over a period of time and therefore cannot be shared, lost or forgotten. In any behavioral biometric one may expect to observe large variations in features. However, they provide sufficiently distinct information that can be used for identification and authentication [1].

A biometric system can be divided into two categories based on the type of application – authentication and identification. Authentication is the process of determining whether someone is, in fact, who they claim to be. This authentication process is often categorized by the number of factors that they incorporate [25]: 1) something you know (e.g., a password), 2) something you have (e.g., token, certificate, ID badge etc.), and/or 3) something you are (e.g., biometrics finger print, iris scan, etc.). A strong authentication is referred to a combination of two or three of these processes. On the other hand identification is the process of associating the person with an identity [76]. In this process, the system seeks to gain knowledge about the subject and associate it with either a set of pre-defined or unknown identities. Therefore, in this paper we have clearly identified the works based on the type of application.

In recent years there have been a few surveys on keystroke dynamics [119, 140, 88, 44]. However, our paper differs from these papers in that it is more comprehensive and up-to-date than any of the existing surveys. We list all the major research in this field and also research published in lesser known journals and conferences. In this paper we have attempted to explain the psychological basis behind the use of keystroke dynamics so as to provide researchers a basic understanding of the various processes involved in typing. This might enable researchers develop better typing models. We have listed all the publicly available databases to aid future researchers interested in research on keystroke dynamics so that accurate comparison of methods and results can be performed. In addition, this is the only paper which has listed currently available commercial solutions related to keystroke dynamics. Even though mobile and touchscreen devices are growing in popularity and research is being conducted on these devices [39, 27, 101], physical keyboards still remain the primary device for data entry in many organizations. Hence, in this paper we focus our attention to keystroke dynamics applied to only physical keyboards.

The following sections describe the psychological basis, data acquisition parameters, available databases, algorithms and commercial solutions available in the field of keystroke biometrics.

2. Psychology

The advantage of using behavioral biometrics such as keystroke dynamics is that it can be collected even without the knowledge of the user. Human-computer interactions play an important role in keystroke dynamics [160]. Hence, in this section we attempt to provide a brief background on the low level human processes involved in typing behavior.

Psychological experiments conducted over the last century have demonstrated that repetitive, routine tasks such as speaking, writing, playing the piano, walking, dancing, and typing are governed by a set of actions. These actions can be predicted by developing a model which describes the series of steps undertaken to achieve the task. The motor systems plan and control the movement based on that information. Motor systems can be

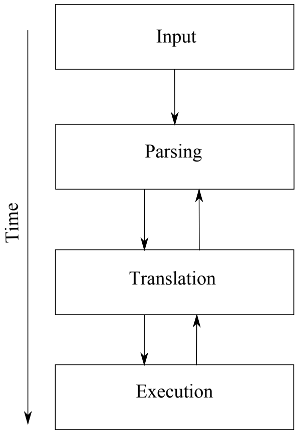

Fig. 1: Composite model of major stages in transcription typing [134].

thought of as special cases of self-organizing systems [138]. During the late 20^{th} century, studies were undertaken to develop an understanding in the physiology and psychology of motor skill with an emphasis on telegraphic keying. Bryan and Harter [24] conducted a series of experiments on thirty-seven telegraphic operators having varying degrees of skill. It was observed that telegraphic operators could recognize other operators with whom they worked by listening to their characteristic typing patterns. Many operators claimed that they could also determine the gender of the operator by the style of his/her sending a message (page 35). The authors observed that learning to receive in telegraphic language is based on developing a hierarchy of psycho-physical habits [156].

Cooper [41] presented the first general model information-flow flow identifying the major stages during transcription typewriting. Salthouse [134] proposed a composite model of the general stages in transcription typing based on earlier works by theorists as shown in Figure 1. The first stage involved the perception or recognition of the characters. In this stage it was observed that typists tend to look some words ahead of the material that they are typing. Butsch [26] found that the eye span depended on the skill of the typist. He demonstrated that skilled typists have a higher eye span than unskilled typists.

The second stage is the parsing stage. In this stage the words perceived are stored in memory for a short time before they are typed out by the hand. The use of memory as a short-term buffer before typing was supported by experiments conducted by Thomas and Jones [150]. They postulated that the conversion of text to type involves a combination of serial and parallel information processing. Cooper demonstrated that typists break text in small predictable groups, due to the limitations on memory buffer size. Experiments conducted by Verwey and Dronkert [154] suggested that the motor chunking and simultaneous processing occur during a continuous key-pressing task when a timing structure is imposed.

The third stage is the translation of the discrete characters into commands. It comprises of the muscle movements which execute the actual motion of the hands and fingers. In the research conducted by Shaffer [137], he speculated that the response latencies between typing consecutive symbols were paced by an internal, regular rhythm. He suggested that a motor program is not a fixed but a continually changing entity. He also observed that the subject makes use of the knowledge of movement transitions and that keystroke movements are organized prior to their actual execution [139]. The interval between consecutive keystrokes for expert typists was observed to be lesser than unskilled typists. The speed with which the finger moves was also twice as fast for an expert typist than in a novice [57]. Ostry [116] in his work on execution time in movement controls showed that organization of movement is associated with on-going behavior. In order to better understand the user-computer interaction performance i.e., time required to perform a task, Card et al. [29] developed a Keystroke-level model.

The fourth stage is the actual execution of the text followed by the feedback system. After execution of the keystroke, a feedback is provided to ensure the accuracy of the system. There are three types of feedback, *visual, auditory* and *kinesthetic.* Work to show that editing and execution can occur continuously accordingly to predetermined schedules was conducted by Rosenbaum et al. [128]. They conducted experiments to examine the sequence of execution and programming in human motor programming. John [81] presented an engineering model of the typing phenomena based on Salthouse's work and created a computational theory that could explain and predict it using the Model Human Processor [30]. Rumelhart and Norman [130] performed a computer simulation of the skilled typist to show how a motor system can produce such overlaps in the finger movements. They represented some of the major factors involved in typing, including inter-keystroke interval times, the switching and doubling errors found in skilled typists. Although the model was not an accurate one, this was the first step towards developing a model to understand the working of a skilled typist. Although the model presented by Salthouse [134] is not a complete model, it encompasses some of the basic processes in typing. The paper highlights the ballistic nature of typing execution. Ballistic movements are, generally, semi-autonomous (that is, once they are initiated, they can't stop) and thus are behaviors that are difficult to fake. Each of these stages indicate the various processes involved that lead to the development of a unique typing behavior. This makes the case even stronger that an understanding of such information is useful in a biometric sense.

3. Data Acquisition

It was Spillane [143] who first suggested the use of keyboards to measure the keystroke dynamics of individuals for identification. During enrollment the timing pattern and the key pressure of the user would be stored along with an unique phrase (password). Thereafter, the user could access the system by entering the password and other identification. This information would be compared by the system with the stored password and keystroke dynamics, which would be used to identify the user. In the following paragraphs we describe some of the approaches taken by researchers to acquire data from users and the features that can be extracted from the raw data.

3.1 Text entry

Analysis of keystroke dynamics can be broadly classified into two types – *static or structured text* and *dynamic or free text.* Static analysis involves analyzing keystroke behavior of an individual on predetermined phrase(s) at certain points in the system. For example, when logging in a system the user's typing pattern is analyzed when he/she types the user-id and password. It can also involve the use of a particular phrase which is common for all the users of the system. Static text entry can be deployed in systems where there is no scope for further text entry. For example, when a user logs in to check his bank accounts online there is usually no further scope of text entry. Dynamic analysis involves continuous or periodic monitoring of keystroke behavior. It is first checked when a user logs in the system and continues thereafter. For example, if a person is browsing the web, certain websites maybe frequented by the user. A list of the commonly occurring websites and the typing behavior of the user while entering the string can be stored. In this case, a training phase would be needed where the user types a particular string several times so that a model can be built for that string. During the test phase, as the user types, the string is recorded along with its timing info which can then used for authentication. However, dynamic monitoring may lead to privacy issues due to its intrusive nature. Marsters [103] proposed a solution in his thesis where he collected only quadgraphs and stored the data in a matrix instead of an ordered log. This method discourages recovery of the keystroke log thereby improving the privacy of the data.

3.2 Environment

Environment plays an important role in determining the typing behavior of an individual. Depending on the type of keyboard, the rhythm or keystroke dynamics of the user may be affected. In many of the experiments conducted to examine classify users based on their typing behavior, they were asked to type on a particular machine. The machine had a software installed to record the keystrokes and timing information. The room and lighting conditions were relatively the same for all users. Such an environment where the data was collected is known as a *controlled environment*. In such an environment, the subjects may or may not be habituated to type using the specified keyboard. Therefore in some cases, the subjects were asked to practice on the keyboards before samples were collected. The data collected in a controlled environment may not be representative of the actual conditions in which a user types. An *uncontrolled environment* can be defined as an environment where researchers have only partial or no control over the way the subject enters information. In such an environment, subjects are asked to either download the program on their personal machines (desktops or laptops) or fill out a form online to provide keystroke information. In the former case, the subjects typically typed the data and were required to send the data to the researcher. In the latter case, the data was collected dynamically with the data being stored automatically. Uncontrolled environments provide a realistic situation where these systems can be tested and deployed. However, they can be harder to analyze since there are a lot of variables to consider. For example, the way the user types - on a table, on the bed or places a laptop on his lap and timing issues with online data collection. In addition, different brands of computers have differences in the spacing, size and sensitivity of the keys. This might play an important role while evaluating results in an uncontrolled environment.

3.3 Features

Before we discuss the approaches taken by researchers in keystroke dynamics, we describe in detail the features that can be extracted from the raw typing data. While typing, the computer can record the key pressed and the time at which it was pressed. In addition to this, it can also record the time at which the same key was released and for how long it was pressed. All this information can be stored while a user is typing the data and certain features can be extracted from them as shown in Figure 2. Here we present commonly used features in keystroke dynamics research.

Latency is one of the most commonly used feature by many early researchers. There are three types of latencies as defined by [9] - *press-to-press* (**PP**), *release-to-release* (**RR**) and *release-to-press* (**RP**) latencies. The *digraph* is also defined as the **PP** latency by most researchers in literature [96]. Recently, researchers have also called the *release-to-press time* as *flight time* [144]. These features are easy to extract from the raw information since the system can log the time at which each key press was made. Gaines et al. [56] conducted the first feasibility study on the use of timing pattern of keystrokes as an authentication

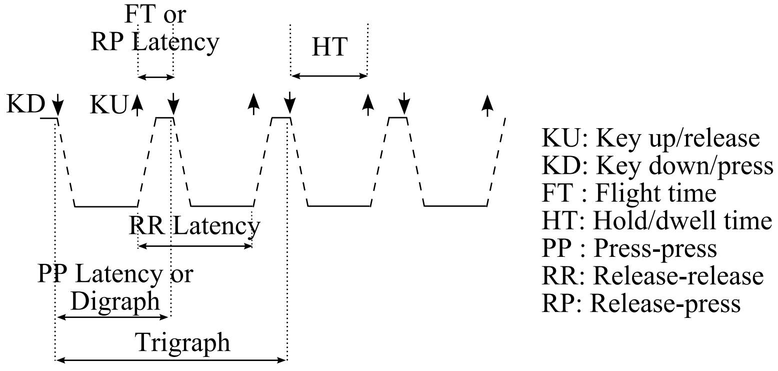

Fig. 2: Keystroke Timing Information.

method at Rand Corporation. They observed that the average digraph times ranged about 125 ms. *Trigraph* is the time interval between the presses or releases of alternate keystrokes. Some researchers have also used other N-graph features for identification. This feature was first used by Bergadano et al. [16]. It was observed that using trigraph features gave better classification results than using digraphs or higher order n-graphs. *Key hold time* or *dwell time* is defined as the time for which each keystroke was pressed. The keystroke latency is the combination of the hold and flight times. In all the previous works typing by subjects was carried out in a controlled environment. Monrose and Rubin [107] were the first to conduct experiments to check the viability of authentication users based on keystroke dynamics in an uncontrolled environment. Robinson et al. [127] concluded that hold times are much more important than inter-key times. Certain types of keyboards are available in the market which can measure the *pressure* applied to a key while typing. Attempt to recognize emotion from users typing pattern using pressure sensitive keyboards was carried out by Lv et al. [98]. Similar work was also conducted by Allen [5] as part of his thesis. In addition to these features, the *total duration* taken to type a certain string can also be measured. Using one or more of these features, sub features such as the *force* applied to a key and *speed* of typing can be derived [88]. The speed is usually measured in words per minute (wpm) or the average time to type each character. Other secondary features that can be derived are the minimum/maximum speed of typing, the mean and standard deviations of the features and the entropy [106].

3.4 Error metrics

In the authentication/verification stage, the raw data is acquired and processed to extract the biometric. This biometric is then compared with the existing template in the database. A matching algorithm is then used to determine how closely the biometric matches with an existing template in the database. On the other hand, identification is the one-to-many process of comparing a submitted biometric sample against all of the biometric reference templates [158].

Two important error rates are used to determine the performance of a biometric authentication system – False Acceptance Rate (FAR) and False Rejection Rate (FRR). FAR is the percentage of impostors inaccurately allowed as genuine users. It is defined as

$$FAR = \frac{Number\ of\ false\ matches}{Total\ number\ of\ impostor\ match\ attempts}$$

FRR is the number of genuine users rejected from using the system. It is defined as

$$FRR = \frac{Number\ of\ false\ rejections}{Total\ number\ of\ genuine\ match\ attempts}$$

Some researchers report the equal error rate (EER) instead of FAR and FRR. EER is defined as the value of FAR/FRR at an operating point on ROC where FAR equals FRR [77]. Higher FAR is generally preferred in systems where security is not of prime importance, whereas higher FRR is preferred in high security applications [1]. The lower the value of EER, the better the system is. Many researchers report the identification as a percentage of attempts correctly identified to the total number of attempts made.

Table 1: Publicly available keystroke dynamics databases

Database	Features	TT	Subjects	Samples	Keyboard Layout
Jugurta and Freire [84]	Latency	S D	32 15	320 150	Brazilian
Killourhy and Maxion [91]	Latency, hold and flight times	S	51	20400	U.S.
Giot et al. [58]	Latency between two presses, Latency between two releases, hold and flight times and, concatenation of previous vectors	S	133	7555	AZERTY
Allen [5]	Hold time, flight time & normalized key pressure	S	104	2379	U.S.
Bello et al. [14]	Key up and down times	S	54	282020*	-

TT - Text type, S - Static, D - Dynamic

(*) Keystroke events consisting of key presses and releases from 58 unique sessions

3.5 Databases

In the booming market of biometrics, a cheap and effective biometric such as keystroke dynamics has not increased with the same rate. It is difficult to compare the works of different researchers due to lack of standards for data collection and benchmarking. Reynolds from AdmitOne Security (formerly BioPassword) submitted a proposal to develop a keystroke dynamics format for data interchange [126]. Adoption of standards such as these should help facilitate the exchange of information amongst researchers and provide a better way to compare different algorithms. This would certainly reduce duplication and multiplicity of efforts [122]. Creation of large, standardized databases, such as those available in other biometric areas, which are publicly available will only further the development of this field and allow accurate comparisons of algorithms and approaches.

Table 1 lists some of the databases which have been made available by some researchers. Of all these databases, only the database by Jugurta and Freire [84] consists of static and dynamic text. This is important since it allows testing in a real world situation. However, the number of subjects and samples in this database are small as compared to other databases. The database collected by Giot et al. [58] is one of the largest databases publicly available. However, this is still small when compared to practical situations where companies may contain data from thousands of users. A joint effort needs to be undertaken by researchers to collect a large database which resembles a real world situation. The database compiled by Allen [5] is the only database which consists of the keystroke pressure information. Commercially available pressure sensitive keyboards can provide additional information without the need of any additional hardware. In addition, it would be helpful to researchers if the databases would contain soft biometric information such as age, gender and ethnicity. This would open up new vistas of research in soft biometric classification.

4. Approaches

Once the features have been extracted and templates created, classification of users is performed based on the similarities and dissimilarities among the templates. Researchers have used simple patterns derived from the statistics of the features such as the mean and standard deviations to complex pattern recognition algorithms to classify typists. In some cases, a combination of methods have also been used. The variety of algorithms used by researchers makes categorizing them a challenging task. In this paper we have categorized classification algorithms into four major categories which are described below. These four categories have been based on the type of algorithms used popularly by researchers.

4.1 Statistical Algorithms

The simplest *statistical method* consists of computing the mean and standard deviation of the features in the template. These can then be used for comparison using hypothesis testing, t-tests and distance measures such as absolute distance, weighted absolute distance, Euclidean distance and so on. Joyce and Gupta [83] were one of the early researchers who used absolute distances for authentication. Using only absolute distances they achieved a false acceptance rate (FAR) of 0.25% and a false reject rate (FRR) of 16.36%. Recently, Guven and Sogukpinar [66] have used vector analysis to classify users with a 95% accuracy. Table 2 lists all the major work undertaken by researchers towards developing authentication and identification systems using statistical algorithms. Although some researchers have reported impressive results, it should be noted that the total number of samples in these experiments are low. Since keystroke is dependent on the behavior of the subject, the features appear to be non-linear in nature. Hence, using linear, statistical approaches may not provide good results. Another disadvantage of using statistical algorithms is the lack of a training stage which can be useful to identify patterns in the keystroke data.

4.2 Neural Networks

Neural networks or artificial neural networks are adaptive non-linear statistical data modeling tools which have been inspired by biological interconnection of neurons. There are two ways in which the weights can be assigned (or learned) - *supervised learning* and *unsupervised learning*. One of the most popular methods in supervised learning is called the *backpropagation*. One of the popular methods in unsupervised learning is the Hopfield neural network. Other algorithms such as perceptron, Sum of Products (SOP), Adaline and weightless neural networks have been used to classify users based on their keystroke dynamics. Obaidat and Macchiarolo [113] presented a way to classify inter-character times using an artificial neural network. During the investigation phase, three different neural network architectures were tested - backpropagation, sum-of-products and hybrid sum-of-products. From experiments, hybrid sum-of-products was found to perform better than other architectures and achieved an identification rate of 97.8%. Yong et al. [161] suggested use of weightless neural networks for classifying users. They scaled the data before discretizing it into linear and non-linear intervals. They also observed that the non-linear intervals gave better results than linear intervals. Table 3 lists all the major work undertaken by researchers towards developing authentication and identification systems using neural networks. Many researchers have used neural networks successfully with good results. Neural networks have an advantage that they can handle many parameters. However, they can be slow not only during the training but also in the application phase. In neural networks it is difficult to decide which features are important for classification due to its "black box" mode of operation. This could be a problem for continuous keystroke authentication where results are typically desired in real time.

Table 2: Authentication and identification using statistical algorithms

Study	Features	Classification	TT	Env.	Subjects	Samples	Error Rates (%)		
							FAR	FRR	EER
Gaines et al. [56]	1	Statistical	S	C	6	36	-	-	-
Umphress & Williams [151]	1	Statistical	S	C	17	34	6	12	-
Leggett and Williams [95]	1	Digraph Test	S,D	C	36	72	5.5	5	-
Bleha et al. [20]	1	Min. dist. classifier	S	C	39	171	2.8	8.1	-
Joyce & Gupta [83]	1	Abs. distance	S	C	33	975	0.25	16.67	-
Napier et al. [109]	1,3	Statistical	S	C	24	-	-	-	3.8
Mahar et al. [100]	1	Statistical	S	C	67	-	-	-	17.6
Furnell et al. [55]	1	Statistical	D	C	30	60	15	0	-
Tapiador & Sigüenza [146]	1	Statistical	S	U	9	1620	-	-	-
Coltell et al. [40]	1	Statistical	S	C	10	-	>70	≈0	-
Bergadano et al. [16]	1,2	Statistical	S	C	44	220	0	2.3	-
Monrose et al. [106]	1,3	Statistical	S	C	20	481	20	20	-
Araújo [7]	1,3	Statistical	S	C	30	553	1.89	1.45	-
Gunetti et al. [62]	1,2	Distance measure	D	U	30	-	8.33	3.33	-
Gunetti et al. [63]	1,2	Distance measure	D	U	31	-	≈0	≈ 2.0	-
Gunetti & Picardi [61]	1,2	Relative, Abs. distance	S, D	C	205	765	0.005	5	0.5
Revett et al. [124]	1	Statistical	S	U	43	≈688	-	-	5.58
de Magalhaes [48]	1,3	Statistical	S	U	43	≈688	-	-	≈ 5.0
Lv & Wang [99]	3,4	Statistical classifiers	S	C	100	5000	1.4	1.4	1.41
Modi & Elliott [105]	1	Statistical classifiers	S,D	C	42	6300	0.33	94.87	-
Montalváo et al. [108]	1	Statistical	S D	C	-	-	- -	- -	6.2 12.7
Boechat et al. [21]	1,3	Statistical	S	C	-	-	0	3.83	-
Lee et al. [94]	1,3	Hypothesis	S	U	16	3200	≈4.5	≈5.5	-
Choraś & Mroczokwski [36] Choraś & Mroczokwski [37]	1,2	Degree of disorder	S	U	18	≈810	0	0-55	-
Davoudi & Kabir [47]	1,2,3	Statistical	D	C	21	-	9	5	-
Giroux et al. [60]	1,3	Mean	S	U	11	880	0	-	-
Killourhy & Maxion [91]	1,3	Manhattan(‡)	S	C	51	20400	-	-	9.6
Bours & Barghouthi [22]	1,3	Statistical	D	U	25	1620	-	-	-
Douhou & Magnus [49]	1,3	Statistical	S	U	1254	-	16	1	-
Chudá & Ďurfina [38]	1,3	Angle b/w latencies	S	-	15	-	8.4 3.6	2.5 4.7	- -
Xi et al. [159]	1,2	nGdv-V nGdv-C	D	U	205	765	9.43 1.65	24.7 2.75	- -

Study	Features	Classification	TT	Env.	Subjects	Samples	Identification Rate (%)
Shepherd [142]	1,3	Statistical	S	-	4	-	99
Monrose & Rubin [107]	1,3	Statistical	S,D	U	31	-	90
DSouza [51]	1	Statistical	S	C	11	-	76
Guven & Sogukpinar [66]	1	Vector Analysis	S	C	-	-	89.3 / 95†
Bergadano et al. [17]	1,2	Distance measure	S, D	C	40	364	90
Gunetti et al. [62]	1,2	Distance measure	D	U	30	-	≈90
Gunetti et al. [63]	1,2	Distance measure	D	U	31	-	90
Villani et al. [155]	1,3	Euclidean dist.	D	C U	118	-	98.3 99.4
Janakiraman & Sim [78]	1,3	Histogram	D	U	22	-	9.91 – 100*
Teh et al. [148]	1,3	DSM△	S	C	50	-	93.64
Lv et al. [98]	4	Statistical classifiers	S	C	50	3000	6.6
Rybnik et al. [131]	1,3	Statistical	S	C	37	-	72.97
Samura & Nishimura [135]	1,3	Euclidean dist.	S	C	112	-	90.7 – 100

1 - Latency, 2 - Trigraph/N-graph, 3 - Key hold time, 4 - Key Pressure, 5 - Rhythms/acoustic cues

TT - Testing type, S - Static, D - Dynamic, C - Controlled, U - Uncontrolled.

(†) Using data from experienced users a higher accuracy level was achieved; (*) For English words the highest identification rate was 9.91%, while for non-English words the highest accuracy was 100%; (‡) - In this paper, comparison of 14 algorithms were performed, of which the Manhattan distance performed the best; (△) DSM - Direction Similarity Measure

Table 3: Authentication and identification using neural networks

Study	**Features**	**Classification**	**TT**	**Env.**	**Subjects**	**Samples**	**Error Rates (%)**		
							FAR	**FRR**	**EER**
Bleha & Obaidat [19]	1	Perceptron (NN)**	S	C	24	1400	8	9	-
Brown & Rogers [23]	1,3	Distance Adaline (NN)** BP (NN)**	S	C	46	1867	14.9 17.4 12	0 0 0	- - -
Furnell et al. [55]	1	NN**	S	C	15	2100	8	7	-
Lin [97]	1	BP-NN$^{\perp}$	S	C	151	-	1.11	0	-
Obaidat & Sadoun [114]	1	potential function Neural Networks	S	C	30	6750	2.2 0	4.7 0	- -
Bechtel et al. [13]	1	ART-2 NN**	S	U	50	-	-	∼10	-
Dowland & Furnell [50]	1,2,3	NN**	S	U	35	-	4.9	0	-
Yong et al. [161]	1	Weightless NN**	S	C	15	-	-	-	-
Revett et al. [125]	1,2	Specht Probabilistic NN**	S	C	50	-	-	-	≈4
Pavaday & Soyjaudah [117]	1,3	NN**	S	C	>100	5440	1	8	-
Ahmed et al. [3]	1	NN**	D	U	22	-	0.0152	4.82	-
Sulong et al. [145]	1,3,4	RBFN*	S	C	30	180	2	-	-
Harun et al. [68]	1	NN**, dist. classifier	S	C	32 15	320 150	- -	- -	3 22.9

Study	**Features**	**Classification**	**TT**	**Env.**	**Subjects**	**Samples**	**Identification Rate (%)**
Lammers & Langenfeld [92]	1	NN**	S	C	-	-	75
Obaidat & Macchiarolo [113] Obaidat & Macchiarolo [112]	1	NN**	S	C	6	3600	97.8
Capuano et al. [28]	1,3	RBFN*	S	-	10	-	≈95
Ahmad & Abdullah [2]	1,3	ADALINE,BP-NN$^{\perp}$	S	-	30	-	≈0
Cho et al. [35]	1,3	Perceptron (NN)**	S	U	21	-	99 – 100***
Saevanee & Bhattarakosol [132]	1,3,4	Probabilistic NN**	S	C	10	300	99

1 - Latency, 2 - Trigraph/N-graph, 3 - Key hold time, 4 - Key Pressure, 5 - Rhythms/acoustic cues

TT - Testing type, S - Static, D - Dynamic, C - Controlled, U - Uncontrolled.

(**) NN - Neural Network; ($^{\perp}$) BP-NN - Backpropagation neural network; (*) RBFN - Radial Basis Function Network; (***) For 13 users the classification error was 0% and for remaining users the average error was 1%

4.3 Pattern Recognition and learning based algorithms

Pattern recognition is the act of using patterns or objects and classifying them into different categories based on certain algorithms [149]. It contains simple machine learning algorithms such as the nearest neighbor algorithms and clustering to much more complex algorithms such as data mining, Bayes classifier, Fishers linear discriminant (FLD), support vector machine (SVM) and graph theory. Yu and Cho [162], used a three step approach to improve the performance of keystroke identification. The SVM novelty detector achieved an average error rate of 0.81%. Giot et al. [58] proposed a method to identify computer users by using a support vector machine (SVM) and achieved an identification rate of 95%. SVM is a supervised learning algorithm which shows promising results for both authentication and identification. This seems to be an important algorithm against which future algorithms should be benchmarked. Table 4 lists all the major work undertaken by researchers towards developing authentication and identification systems using pattern recognition and learning based algorithms. One of the biggest advantage of using probabilistic learning algorithms is that they provide a confidence value associated with the decision made. Probabilistic learning algorithms can also reduce the problem of error propagation by ignoring outputs with low confidence values. Moreover, unsupervised learning techniques can identify patterns in the data automatically.

Table 4: Authentication and identification using pattern recognition and learning based algorithms

Study	**Features**	**Classification**	**TT**	**Env.**	**Subjects**	**Samples**	**Error Rates (%)**		
							FAR	**FRR**	**EER**
Robinson et al. [127]	1,3	MICD$^{\mathsf{T}}$, non-linear, Inductive learning	D	U	20	-	9$^{\oslash}$	10$^{\oslash}$	-
Haider et al. [67]	1	Fuzzy logic, NN** & statistical techniques	S	C	-	-	6	2	-
Gutirrez et al. [65]	1	Naïve Bayesian, statistical Decision Trees, Instance based	S	C	-	-	1.6$^{\circledast}$	14.3$^{\circledast}$	-
Kacholia & Pandit [85]	1	Random dist. function	S	-	20	440	3.4	2.9	-
Araújo et al. [6]	1,3	Fuzzy Logic	S	C	10	200	3.4	2.9	-
Eltahir et al. [53]	1,4	Autoregressive	S	C	22	22	5	-	-
Sang et al. [136]	1	SVM‡	S	-	10	-	0.02	0.1	-
Joshi & Phoha [82]	1	CSOMA$^{\triangleleft}$	S	C	43	873	0.88	3.55	-
Chang [32]	1,3	HMM$^{\vee}$	S	C	20	600	0	-	-
Sheng et al. [141]	1,3	Parallel decision trees	S	C	43	387	0.88	9.62	-
Bartlow & Cukic [10]	1,3	Random Forest	S	U	41	8775	3.2	5.5	-
Chang [33]	1,3	Hierarchical Tree-Based	S	C	-	-	0	3.47	-
Hosseinzadeh et al. [73]	1,3	GMM$^{\sqcup}$	S	U	8	80	2.1	2.4	-

Lee & Cho [93]	1,3	Gauss,Parzen,AAMLP,1-SVM, SVDD$_1$,SVDD$_2$,KMC,LVQ-ND*	S	C	21(Set A)
					25(Set B)
									0.43
									0.4–1.34***

Meszaros et al. [104]	1	PCA & Euclidean	S	C	25	2400	-	-	1
Pavaday & Soyjaudah [118]	1,3	k-nearest neighbor	S	C	60	600	-	-	35
Jiang et al. [79]	1,2,3	HMM$^{\vee}$	S	U	315	2577	-	-	≈2
Hocquet et al. [71]	1,3	SVM‡	S	U	38	-	-	-	4.5
Bartlow & Cukic [11]	1,3	Random Forest with user--specific voting threshold	S	U	53	>10000	-	-	2
Hosseinzadeh & Krishnan [72]	1,3	GMM$^{\sqcup}$	S	U	41	1230	4.3	4.8	4.4
Hu et al. [74]	1,2	k-nearest neighbor	S	C	36	-	0.045	0	-
Hempstalk et al. [69]	1,2	One-class Gaussian	D	U	19	-	11.3	20.4	-
Jin et al. [80]	1,3	Fuzzy Logic	D	C	10	-	-	-	20
Giot et al. [58]	1,3	SVM‡	S	C	133	7555	-	-	13.45
Hwang et al. [75]	1,3,5	1-SVM‡	S	C	25	4500	-	-	1
Giot & Rosenberger [59]	1,3	SVM‡	S	U	100	-	-	-	15.28
Saggio et al. [133]	1,3	multi-SVM‡	S	C	16	9600	0.91–6.73 0.69–5.10	2.31–11.69 0.0–9.92	- -
Study	**Features**	**Classification**	**TT**	**Env.**	**Subjects**	**Samples**	**Identification Rate (%)**		
Ru & Eloff [129]	1	Fuzzy logic	S	C	29	725		92.62	
Gunetti & Ruffo [64]	1,2	Relational Decision Trees	D	C	10	-		89.3	
Ong & Lai [115]	1	k-means	S	C	20	1200		85	
Changshui & Yanhua [34]	5	Autoregressive	S	C	24	432		65	
Wong et al. [157]	1	k-nearest neighbor, Neural Network	S	C	10	-		84.63 99	
Nisenson et al. [110]	1	Lempel-Ziv	S,D	C	35	-		93.57 / 96.42†	
Mandujano & Soto [102]	1	Fuzzy logic, c-means	S	C	15	3375		98	
Yu & Cho [162]	1,3	SVM‡	S	C	-	-		99.19	
Revett et al. [123]	1	Rough Sets	S	U	100	-		95	
Curtin et al. [45]	1,3	Nearest neighbor w/ Euclidean	S	U	10	-		97-100	
Tappert et al. [147]					30		94.7		
Benitez et al. [15]	1	k-means	S	U	54	-		≈85**	
Bazrafshan et al. [12]	1,3	Genetic Fuzzy classifier	S	U	8	1509		81.21	
Balagani et al. [9]	1,3	Näive Bayes TAN$^{\diamond}$ k-nearest ridge-LR$^{\triangleright}$	S	U	33	873		95.87 95.87 87.06 93.47	

1 - Latency, 2 - Trigraph/N-graph, 3 - Key hold time, 4 - Key Pressure, 5 - Rhythms/acoustic cues

TT - Testing type, S - Static, D - Dynamic, C - Controlled, U - Uncontrolled.

($^{\mathsf{T}}$) MICD - Minimum intra-class distance; ($^{\oslash}$) Error rates for the Inductive learning classifier which outperformed other techniques; ($^{\circledast}$) Error rates for the Näive Bayesian classifier which outperformed other techniques; ($^{\vee}$) HMM - Hidden Markov Models; ($^{\sqcup}$) GMM - Gaussian Mixture Models; ($^{\curlyvee}$) Support vector machine with evolutionary genetic algorithm; (*) SVD - Support vector description, LVQ-ND - linear vector quantization for novelty detection, AAMLP - auto-associative multilayer perceptron, KMC - K-means clustering; (***) The EER is for LVQ-ND in set B as newer or older passwords are used (‡) SVM -Support Vector Machine; (†) The accuracies are for single-class and two-class classifications respectively; (**)F-measure was used instead of percentage; ($^{\diamond}$) TAN - Tree augmented naive Bayes; ($^{\triangleright}$) ridge-LR - ridge-logistic regression; ($^{\triangleleft}$) CSOMA - Competition between Self Organizing Maps for Authentication

Table 5: Authentication and identification using search heuristics and combination of algorithms

Study	**Features**	**Classification**	**TT**	**Env.**	**Subjects**	**Samples**	**Error Rates (%)**		
							FAR	**FRR**	**EER**
Yu & Cho [162]	1,3	GA$^{\ominus}$ F.S. ensemble	S	C	21	-	- -	3.54 3.69	- -
Hocquet et al. [70]	1,3	Statistical, degree of disorder & time discretization	S	U	13	-	1.81	1.69	1.75
Azevedo et al. [8]	1,3	SVM with GA$^{\curlyvee}$ PSO$^{\rtimes}$	S	C	24	4200	0.43 0.41	4.75 2.07	- -
Revett [121]	1	Bioinformatics	S	C	20	40000	0.1	0.1	-
Killourhy & Maxion [90]	1,3	Mean nearest-neighbor multi-layer perceptron	S	C	51	20400	- - -	- - -	11 9.96 16.24
Ali et al. [4]	1,3	NN** and ANFIS$^{\diamond}$	S	C	5	1000	0	0	-
Study	**Features**	**Classification**	**TT**	**Env.**	**Subjects**	**Samples**	**Identification Rate (%)**		
Dahalan et al. [46]	4	ANFIS$^{\diamond}$	S	C	10	-	80		
Karnan & Akila [87]	1,3	GA$^{\ominus}$ PSO$^{\rtimes}$ Ant colony	S	C	27	2700	88.9 86.6 92.8		
Pedernera et al. [120]	1,4	AKM & ASC$^{\circ}$	S	U	-	-	82.6		

1 - Latency, 2 - Trigraph/N-graph, 3 - Key hold time, 4 - Key Pressure, 5 - Rhythms/acoustic cues

TT - Testing type, S - Static, D - Dynamic, C - Controlled, U - Uncontrolled.

(**) NN - Neural network; ($^{\ominus}$) GA - Genetic algorithm; ($^{\curlyvee}$) Support vector machine with evolutionary genetic algorithm; ($^{\rtimes}$) PSO - Particle swarm optimization; ($^{\diamond}$) ANFIS - Adaptive Neural-fuzzy Interface System; ($^{\circ}$) AKM - Adaptive k-means, ASC - Adaptive subtractive clustering

4.4 Search Heuristics and combination of algorithms

Search heuristics such as *genetic algorithms* are used to find an optimum solution. They are a part of evolutionary algorithms. Ant colony optimization (ACO) is an example where genetic algorithm is used. They also find application is areas such as bioinformatics. Bioinformatics is the application of computational technology to molecular biology. Keystroke feature selection using a hybrid system based on support vector machines (SVM) and stochastic optimization algorithms such as Genetic algorithm (GA) and particle swarm optimization (PSO) was developed by Azevedo et al. [8]. The SVM verifier which uses a GA as the evolutionary algorithm for feature selection gave a minimum error of 5.18% at a FAR of 0.43% and FRR of 4.75%. Using PSO with a personal and global acceleration of 1.5, the minimum total error was 2.21% with a FAR of 0.41% and FRR of 2.07%. Revett et al. [121] used a bioinformatics approach to achieve a FAR of 0.1% and FRR of 0.1%. This algorithm can handle 40000 samples and provide promising results. This algorithm shows promising results for large number of samples and should be considered while benchmarking future algorithms. The advantage of using genetic algorithms is that they can easily handle large databases. It also provides multiple solutions and can handle multi-dimensional, non-differential, non-continuous, and even non-parametrical problems.

Sometimes a combination of these classification techniques have been used. A combination of neuro-fuzzy algorithms such as Fuzzy-ARTMAP have been used to classify subjects based on their keystroke dynamics. Montalvao et al. [108] tested the effect of histogram equalization of time intervals on the performance of keystroke based identification algorithms. Four algorithms were used for analysis - on static text using the algorithm proposed by [20], on static and free text by using algorithm proposed by Monrose and Rubin [107], on free text by using the algorithm proposed by Gunetti and Picardi [61] and on free text using a Markov chain algorithm where the prior probability vectors are replaced by 2D and

1D histograms. From all the experiments it was observed that histogram equalization of keystroke timing data led to an improvement in equal error rate (EER). Table 5 lists all the major work undertaken by researchers towards developing authentication and identification systems using search heuristics and combination of algorithms. A recent trend seems to be using search heuristic based approaches and mixture of algorithms for keystroke evaluations and to determine the best features and patterns in the data.

5. Factors affecting performance

From the experiments conducted by Joyce and Gupta [83], they observed that shorter and easy to type login strings were easier to impersonate. They suggested use of other approaches and extraction of other features to improve the performance of the verifier. They also emphasized the importance of the timing accuracy of the machine. Killourhy and Maxion [90] conducted experiments to investigate the effect of clock resolution on keystroke dynamics. They observed that the EER increased by approximately 4.2% when using a 15 ms resolution clock instead of a 1 ms resolution clock. This is significant since the European standards [31] for access control mandates a FAR of 1% and a miss-rate of 0.001%. System load cause problems leading to timing inaccuracies. Hence, systems need to be made robust to timing inaccuracies.

Monrose and Rubin [107] observed from experiments that there was a large variation in comparing the same user performing structured and unstructured texts. They hypothesized that the reason for this could be that the users were not sure about what to type, which led to distinctive pauses during typing. However, for consistent typists it was possible to identify users from structured and unstructured text. They also conducted experiments to identify subjects based on the use of their dominant hand. However, no conclusion could be made due to insufficient number of left-handed subjects.

Ru and Eloff [129] observed that normal "English" like text seemed less discernible from each other rather than those passwords which contain special characters. A new measure known as the Goodness measure was used by Janakiraman and Sim [78] to compute the quality of a word used in keystroke dynamics. Non-English words were identified with higher accuracy than English words. Also the goodness measure provides a reference of good non-English words that can be used for testing purposes. Mandujano and Soto [102] confirmed that having a mix of alphanumeric characters in the password increased its chances of being identified with the valid user. Longer passwords provided a better means to identify individuals accurately.

The effective size, type of passwords and the number of samples needed for a person to enroll and authenticate has an impact on the error rate of the system. Peacock et al. [119] provided a graph containing the cost to enroll and cost to authenticate using works of various researchers. Chang [33] presented a way to increase the size of the training data set using resampling techniques in time and wavelet domains. They tested their technique on a database obtained from Yu and Cho [162]. Although their results were not better than Yu and Cho's, the computational cost was reduced from two hours to 3 seconds.

The emotional state of the user also has an impact on the typing speed. [89] found that a negative state led to a 70% reduction in typing speed compared to the 83% increase in the typing speed when they are in a positive state. The effect of emotions on typing was affirmed by Epp et al. [54]. Similarly, health conditions, place where a person types such as on the bed, on the table, the type and brand of computer used could also have an impact on the efficiency with which a person can be accurately classified.

Artificial rhythms and cues can also be presented to users while typing to improve the quality of keystroke data [86]. A rhythm is the natural pause which a typist includes while typing, while cues are signals presented to a user in a periodic manner. Six different strategies were examined which included various combinations of rhythms and cues. The number of pauses in the rhythms was fixed to 2 while the tempo of the cues was fixed to 400 ms. From experiments on uniqueness and discriminability of rhythm, and consistency and discriminability of cues, they concluded that rhythms and cues are unique and consistent respectively even for a low skilled typist. They observed that typing patterns from artificial rhythms were much more unique than those from a natural rhythm. Use of auditory cues also led to a reduction in the equal error rates.

6. Applications

Although this biometric has been around for a quarter of a century, it has not been a wide part of the security landscape. Some of the possible areas where this biometric could provide security are to deter practice by impostors, prevent spyware attacks and to provide cyber-security against online hacks [140]. Using surreptitious surveillance techniques impostors can attempt to impersonate user behavior. Araújo et al. [7] have shown that using impostor data which practiced user behavior led to an increase in the false acceptance rate (FAR). Lee and Cho [93] proposed a method of retraining the authenticator algorithm using impostor patterns. Monrose [106] suggested a method of password hardening using keystroke dynamics information. Spywares are one of the ways in which an impostor can collect keystroke information such as keys typed, keystroke timing durations and latencies. Until now no major investigation into the possible synthetic or bot attacks on keystroke dynamics has been undertaken barring a recent work by Stephan and Yao [144]. In the paper they showed the robustness of keystroke dynamics to synthetic forgery attacks by two simulated bots using a support vector machine (SVM) approach. Sheng et al. [141] in their work on authentication using parallel decision trees, generated simulated data from existing database. Keystroke dynamics can also be used as a verification step to recover user ids when users forget their passwords [36]. Use of keystrokes to remember and authenticate returning users surfing websites or chat rooms was suggested by Gunetti and Picardi [62]. More recently, researchers have begun looking at the possibility of detecting gender of the user based on keystroke dynamics [59]. This is one of the first steps towards using this biometric for soft biometric recognition. In order to increase the robustness of the identifying and authenticating, it could be combined with other methods such as motion of fingers [111]. Developing countries with their virtually untapped market provide a huge potential for companies to offer security and biometric services. However, use of biometric in developing countries will depend on the size and nature of the organization, in addition to the ease of use [152]. Keystroke biometrics could play an important part in this field since it is cheap and requires no additional hardware.

6.1 Commercial Solutions

Based on existing patents, companies have developed commercial solutions which are being offered to verify user identity. One of its applications includes recognizing users to prevent multiple usage of paid services such as internet or pay-by-view televisions. Information about some of the major commercial solutions and the companies offering them are provided below.

AdmitOneSecurity Inc., formerly BioPassword Inc.,was founded in 2002 and is based in Issaquah, Washington. It is one of the first patented commercial systems providing risk-based authentication solutions for preventing fraudulent use of digital identities and a robust, layered approach to security. This software can continuously assess the risk involved in operation by using a a number of factors such as keystroke dynamics, device characteristics, IP address, etc.. Rules are used to score that information, create a confidence rating, and then provide access based on the confidence or lack of risk.

Authenware Corp. was founded in 2006 and is headquartered in Miami, Florida (USA). It provides identity authentication solution based on a combination of user id and password, keystroke dynamics and other behavioral biometrics. Their product is being used by industries such as financial services, government, transportation and logistics, manufacturing, telecommunications and retail. It is one of the largest keystroke dynamics based company.

BehavioSec, based in Sweden offers Behaviometric solutions for risk based authentication of end users and is designed to tackle the security demands of both IT organizations and online services. It validates the identity of an individual continuously while the user is using the computer.

BioChec is a company based in Stony Point, NY. Keystroke Biometrics is the award-winning patented software-only seamless online authentication solution for FFIEC-compliant secondary authentication. It uses two-factor biometrics, a combination of user id and password, and keystroke dynamics to provide authentication.

Deepnet Security, a company based in London, UK makes TypeSense a keystroke dynamics based solution. They claim to employ flexible enrollment, auto-correlative training and adaptive learning which leads to better results in terms of very low false accept rate (FAR).

Delfigo Security, a company based in Boston, MA is one of the pioneers in the field of keystroke biometrics. Their core product is DSGatewayTM, a versatile authentication platform that utilizes multiple authentication factors, including keystroke and device identification, to validate the credentials of each user and transparently provide the appropriate level of system access.

ID Control, a company based in the Netherlands offers KeystrokeID which monitors and analyses the user's keyboard behavior during the time of his/her access. KeystrokeID operates without any user interference. The user does not have to be trained to use it, as they log on using their user id and password while the process runs in the background or on the network. Their products are being used in e-government, e-business, e-health and e-finance fields.

iMagicSoftware, a company based in Solvang, California makes Trustable Passwords, a patented commercial system based on keystroke dynamics. It is compatible with all web browsers and across all platforms. It works in enterprise networks and over the internet. Trustable Passwords Enterprise Suite, is being used by organizations such as healthcare, financial service, and oil and gas to strengthen their digital security and meet regulatory requirements while maintaining user convenience. It helps websites to provide authentication, fraud prevention, and identity protection.

Probayes is a French based biometric solution provider. It uses Bayesian computing to develop stronger keystroke dynamics based biometric for web applications. It consists of a set of patented software components ready to be integrated in any application. Their core technology ProBT extends the Bayesian Networks framework by providing a structured programming language, allowing the developers to increase their applications capabilities and robustness by easily integrating Bayesian models.

Psylock GmbH, a company based in Regensburg, Germany provides technology based on user's keystroke dynamics. Psylock Authentication Server was awarded the TV Certificate for Software quality functional safety and data security. It is the only keystroke biometrics-based method that requires no password.

Scout Analytics is another behavioral analytics company based in based in Issaquah, Washington which focuses on helping digital publishers maximize the value of their visitors. It helps detect clients based on their keystroke dynamics and prevent users from sharing their same account with multiple partners. They combine typing rhythms with IP addresses and browser information such as cookies to provide a strong authentication.

7. Conclusions and Recommendations

In this survey, we have presented an extensive survey of research conducted in the field of keystroke dynamics over the past three decades. However, there are a few challenges and open areas of research that should be addressed in order to make this an effective biometric.

Keystroke dynamics has a strong psychological basis which should be explored to gain deeper understanding of the motor behavior during typing. Using these concepts, models could be built to better understand the processes involved in typing. An understanding of how different people or groups of people type may provide insight into patterns in soft biometric features such as age and gender. This might help in the development of better classifiers which could improve the accuracies of existing systems.

From existing literature, we recognize that certain features tend to provide more useful than others. Future work on the impact of features or their combination will be helpful and may increase the accuracies. Development of a common nomenclature for the features will possibly clear the ambiguity present in describing features and help in accurate comparison of features.

Research on the effective size and type of passwords and the number of samples needed for a person to enroll and authenticate should be conducted so that users can be enrolled and authenticated as quickly as possible. Although larger passwords provide unique behavior, they are difficult to remember and may inhibit users from using them. Research into type of passwords and their impact on typing is an open area which we believe should be investigated. Most research on keystroke dynamics ignore the time required for training and execution of the algorithms. Any system which uses this biometric to authenticate and identify people should be able to generate a template and results fairly quickly. If the time required to authenticate or identify is too long then it may cause disinterest in the acceptance of this biometric. There are some questions regarding the effectiveness of using a static or dynamic approach to authenticate users. However, this problem depends on the type of application.

Majority of the work on keystroke dynamics involves English as the primary language of communication. However, differences in language can lead to drastically different results even with the same algorithm. This maybe due to layout of keyboards for different languages and the differences in the frequency of characters used in the language. Considering the fact that an algorithm may not provide uniform results in all languages this is an area of research which maybe worthwhile investigating.

There are some instances in existing research where it is unclear how error rates have been achieved. Some researchers report only the lowest error rates. Development of standardized protocols for keystroke system evaluation would help in providing accurate comparisons. The evaluation of the methods used is further complicated by the lack of standardized databases. Availability of large databases such as those in the field of facial recognition and fingerprints will foster research and make it easier to compare future research. Development of large databases containing data from thousands of people might help in determining whether this biometric can be deployed in commercial systems successfully.

Current keystroke data used for research has been collected over relatively short periods of time. The typing behavior of an individual changes with time due to age or health conditions, leading to a large variability in the stored template and current template. Therefore, development of adaptive templates will reduce instances of false rejection as the users age. A longitudinal study of keystroke dynamics could help in determining the problems associated with this biometric.

With mobile computing gaining popularity through the use of smartphones, tablets and other touchscreen devices, it might be worthwhile to consider the application of keystroke dynamics on these devices. It would be interesting to compare results on physical and virtual keyboards since there is a change in typing behavior. In virtual keyboards, the typing is predominantly hunt-and-peck as compared to physical keyboards where the behavior can be hunt-and-peck or all finger usage.

The field of keystroke dynamics is still in a nascent stage and there are a number of challenges that need to be overcome in order for it to become an effective biometric. However, it has tremendous potential to grow in the area of cyber-security and remote monitoring since it is non-intrusive and a cost-effective biometric.

References

[1] A. K. Jain, A. Ross and S. Prabhakar. An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, 14(1):4–20, 2004.

[2] A. M. Ahmad and N. N. Abdullah. User Authentication via Neural Network. In *Proceedings of the 9th International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, AIMSA '00, pages 310–320, 2000.

[3] A. Ahmed, I. Traore, and A. Almulhem. Digital Fingerprinting based on Keystroke Dynamics. In *Proceedings of 2nd International Symposium on Human Aspects of Information Security and Assurance (HAISA)*, 2008.

[4] H. Ali, Wahyudi, and M. Salami. Keystroke pressure based typing biometrics authentication system by combining ANN and ANFIS-based classifiers. pages 198–203, Mar. 2009.

[5] J. D. Allen. An Analysis of Pressure-Based Keystroke Dynamics Algorithms. Master's thesis, Southern Methodist University, Dallas, TX, U.S.A., May 2010.

[6] F. Araújo, L.C., M. Gustavo Liz'arraga, L. Rabelo Sucupira, J. Tadanobu Yabu-uti, and L. L. Lee. Typing Biometrics User Authentication based on Fuzzy Logic. *IEEE Latin America Transactions*, 2(1):69–74, march 2004.

[7] L. Araújo, L. S. Jr., M. Lizarraga, L. Ling, and J. Yabu-Uti. User authentication through typing biometrics features. *IEEE Transactions on Signal Processing*, 53(2):851 – 855, Feb. 2005.

[8] G. Azevedo, G. Cavalcanti, and E. C. Filho. Hybrid Solution for the Feature Selection in Personal Identification Problems through Keystroke Dynamics. In *International Joint Conference on Neural Networks*, pages 1947 –1952, Aug. 2007.

[9] K. S. Balagani, V. V. Phoha, A. Ray, and S. Phoha. On the Discriminability of Keystroke Feature Vectors Used in Fixed Text Keystroke Authentication. *Pattern Recognition Letters*, 32:10701080, 2011.

[10] N. Bartlow and B. Cukic. Evaluating the Reliability of Credential Hardening through Keystroke Dynamics. In *17th International Symposium on Software Reliability Engineering*, pages 117 –126, Nov. 2006.

[11] N. Bartlow and B. Cukic. Evaluating the reliability of credential hardening through keystroke dynamics. In *17th International Symposium on Software Reliability Engineering*, pages 117 –126, nov. 2006.

[12] F. Bazrafshan, A. Javanbakht, and H. Mojallali. Keystroke identification with a genetic fuzzy classifier. In *2nd International Conference on Computer Engineering and Technology (ICCET).*, volume 4, pages V4–136–V4–140, April 2010.

[13] J. Bechtel, G. Serpen, and M. Brown. Passphrase Authentication Based on Typing Style Through an Art 2 Neural Network. *International Journal of Computational Intelligence and Applications*, pages 131–152, 2002.

[14] L. Bello, M. Bertacchini, C. Benitez, J. C. Pizzoni, and M. Cipriano. Collection and Publication of a Fixed Text Keystroke Dynamics Dataset. In *CACIC'10*, October 2010.

[15] C. E. Benitez, M. Bertacchini, and P. I. Fierens. User Clustering Based on Keystroke Dynamics. In *CACIC'10*, October 2010.

[16] F. Bergadano, D. Gunetti, and C. Picardi. User authentication through keystroke dynamics. *ACM Transactions on Information and System Security*, 5(4):367–397, 2002.

[17] F. Bergadano, D. Gunetti, and C. Picardi. Identity verification through dynamic keystroke analysis. *Journal of Intelligent Data Analysis*, 7(5):469–496, 2003.

[18] BioPassword. Authentication Solutions Through Keystroke Dynamics. http://www.infosecurityproductsguide.com/technology/2007
BioPassword_Authentication_Solutions_Whitepaper_FINAL.pdf, 2007.

[19] S. Bleha and M. Obaidat. Computer users verification using the perceptron algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3):900 –902, May. 1993.

[20] S. Bleha, C. Slivinsky, and B. Hussein. Computer-Access Security Systems Using Keystroke Dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(12):1217–1222, Dec. 1990.

[21] G. C. Boechat, J. C. Ferreira, and E. C. Filho. Authentication personal. In *Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on*, pages 254 –256, Nov. 2007.

[22] P. Bours and H. Barghouthi. Continuous authentication using biometric keystroke dynamics. In *The Norwegian Information Security Conference (NISK)*, pages 1–12, 2009.

[23] M. Brown and S. Rogers. User Identification via keystroke characteristics of typed names using neural networks. *International Journal of Man-Machine Studies*, 39:999 – 1014, 1993.

[24] W. L. Bryan and N. Harter. Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, 4(1):27 – 53, 1897.

[25] W. E. Burr, D. F. Dodson, and W. T. Polk. Electronic Authentication Guideline: Recommendations of the National Institute of Standards and Technology. Technical Report 800-63, National Institute of Standards and Technology (NIST), Apr. 2006.

[26] R. Butsch. Eye movements and the eye-hand span in typewriting. *Journal of Educational Psychology*, 23(2):104 – 121, 1932.

[27] P. Campisi, E. Maiorana, M. Lo Bosco, and A. Neri. User authentication using keystroke dynamics for cellular phones. *IET Signal Processing*, 3(4):333–341, july 2009.

[28] N. Capuano, M. Marsella, S. Miranda, and S. Salerno. User Authentication with Neural Networks. In *Proceedings of the V International Conference on Engineering Applications of Neural Networks*, EANN 99, pages 13–15, 1999.

[29] S. K. Card, T. P. Moran, and A. Newell. The Keystroke-Level Model for User Performance Time with Interactive Systems. *Communications of the ACM*, 23:396–410, July 1980.

[30] S. K. Card, T. P. Moran, and A. Newell. *The Psychology of Human-Computer Interaction.* Lawrence Erlbaum Associates, 1983.

[31] CENELEC. *Alarm systems - Access control systems for use in security applications – Part 1: System requirements*, EN 50133-1 edition, 1996.

[32] W. Chang. Improving hidden Markov models with a similarity histogram for typing pattern biometrics. In *IEEE International Conference on Information Reuse and Integration*, pages 487 – 493, Aug. 2005.

[33] W. Chang. Reliable Keystroke Biometric System Based on a Small Number of Keystroke Samples. In G. Mller, editor, *Emerging Trends in Information and Communication Security*, volume 3995 of *Lecture Notes in Computer Science*, pages 312–320. 2006.

[34] Z. Changshui and S. Yanhua. AR model for keystroker verification. In *IEEE International Conference on Systems, Man, and Cybernetics*, volume 4, pages 2887–2890, 2000.

[35] S. Cho, C. Han, D. H. Han, and H. il Kim. Web Based Keystroke Dynamics Identity Verification using Neural Network. *Journal of Organizational Computing and Electronic Commerce*, 10:295–307, 2000.

[36] M. Choraś and P. Mroczkowski. Keystroke Dynamics for Biometrics Identification. In B. Beliczynski, A. Dzielinski, M. Iwanowski, and B. Ribeiro, editors, *Adaptive and Natural Computing Algorithms*, volume 4432 of *Lecture Notes in Computer Science*, pages 424–431. 2007.

[37] M. Choraś and P. Mroczkowski. Recognizing Individual Typing Patterns. In J. Mart, J. Bened, A. Mendona, and J. Serrat, editors, *Pattern Recognition and Image Analysis*, volume 4478 of *Lecture Notes in Computer Science*, pages 323–330. 2007.

[38] D. Chudá and M. Ďurfina. Multifactor authentication based on keystroke dynamics. In *Proceedings of the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, pages 89:1–89:6, 2009.

[39] N. Clarke, S. Furnell, B. Lines, and P. Reynolds. Keystroke dynamics on a mobile handset: a feasibility study. *Information Management & Computer Security*, 11(4):161–166, 2003.

[40] O. Coltell, J. Badfa, and G. Torres. Biometric identification system based on keyboard filtering. In *33rd International Carnahan Conference on Security Technology*, pages 203 –209, 1999.

[41] W. E. Cooper, editor. *Cognitive aspects of skilled typewriting*. Springer-Verlag, 1982.

[42] L. F. Coppenrath and Associates. Biometric Solutions By Classification. http://www.lfca.net/Reference%20Documents/Biometric%20Solutions %20By%20Classification.pdf, 2001.

[43] L. F. Coppenrath and Associates. Biopassword Technology Overview. http://www.lfca.net/Reference%20Documents/Biometric%20Technology %20Overview.pdf, 2001.

[44] H. Crawford. Keystroke dynamics: Characteristics and opportunities. In *Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on*, pages 205 –212, Aug. 2010.

[45] M. Curtin, C. Tappert, M. Villani, G. Ngo, J. Simone, H. S. Fort, and S. Cha. Keystroke biometric recognition on long-text input: A feasibility study. In *Proc. Int. Workshop Sci Comp/Comp Stat (IWSCCS 2006), Hong Kong*, June 2006.

[46] A. Dahalan, M. J. E. Salami, W. K. Lai, and A. F. Ismail. Intelligent Pressure-Based Typing Biometrics System. In *Knowledge-Based Intelligent Information and Engineering Systems*, volume 3214 of *Lecture Notes in Computer Science*, pages 294–304. 2004.

[47] H. Davoudi and E. Kabir. A new distance measure for free text keystroke authentication. In *14th International CSI Computer Conference*, pages 570 –575, Oct. 2009.

[48] S. de Magalhaes, K. Revett, and H. Santos. Password secured sites - stepping forward with keystroke dynamics. In *International Conference on Next Generation Web Services Practices.*, Aug. 2005.

[49] S. Douhou and J. R. Magnus. The reliability of user authentication through keystroke dynamics. *Statistica Neerlandica*, 63(4):432–449, 2009.

[50] P. Dowland and S. Furnell. A Long-Term Trial of Keystroke Profiling Using Digraph, Trigraph and Keyword Latencies. In Deswarte, Yves and Cuppens, Frdric and Jajodia, Sushil and Wang, Lingyu, editor, *Security and Protection in Information Processing Systems*, volume 147 of *International Federation for Information Processing*, pages 275–289. Springer - Boston, 2004.

[51] D. C. D'Souza. Typing Dynamics Biometric Authentication. Bachelor's thesis, Department of Information Technology and Electrical Engineering, University of Queensland, Queensland, Australia, October 2002.

[52] T. Dunstone and N. Yager. *Biometric System and Data Analysis: Design, Evaluation, and Data Mining*. Springer, 1 edition, 2008.

[53] W. Eltahir, M. Salami, A. Ismail, and W. Lai. Dynamic keystroke analysis using AR model. In *IEEE International Conference on Industrial Technology*, volume 3, pages 1555 – 1560, Dec. 2004.

[54] C. Epp, M. Lippold, and R. L. Mandryk. Identifying emotional states using keystroke dynamics. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, pages 715–724, 2011.

[55] S. M. Furnell, J. P. Morrissey, P. W. Sanders, and C. T. Stockel. Applications of keystroke analysis for improved login security and continuous user authentication. In *Information systems security*, pages 283–294. 1996.

[56] R. S. Gaines, W. Lisowski, S. J. Press, and N. Shapiro. Authentication by Keystroke Timing: Some Preliminary Results. Technical Report R-2526-NSF, Rand Corporation, May 1980.

[57] D. R. Gentner. *Cognitive aspects of skilled typewriting*, chapter Keystroke Timing in Transcription Typing, pages 95–120. Springer-Verlag, 1982.

[58] R. Giot, M. El-Abed, and C. Rosenberger. GREYC keystroke: A benchmark for keystroke dynamics biometric systems. In *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1 –6, Sept. 2009.

[59] R. Giot and C. Rosenberger. A New Soft Biometric Approach For Keystroke Dynamics Based On Gender Recognition. *International Journal of Information Technology and Management (IJITM) Special Issue on : Advances and Trends in Biometrics*, pages 1–16, 2011.

[60] S. Giroux, R. Wachowiak-Smolikova, and M. P. Wachowiak. Keystroke-based authentication by key press intervals as a complementary behavioral biometric. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 80–85, Oct. 2009.

[61] D. Gunetti and C. Picardi. Keystroke analysis of free text. *ACM Transactions on Information and System Security*, 8(3):312–347, 2005.

[62] D. Gunetti, C. Picardi, and G. Ruffo. Dealing with Different Languages and Old Profiles in Keystroke Analysis of Free Text. In *Proc. of the Nineth Congress of the Italian Association for Artificial Intelligence (AI*IA)*, pages 347–358, 2005.

[63] D. Gunetti, C. Picardi, and G. Ruffo. Keystroke Analysis of Different Languages: A Case Study. In A. F. Famili, J. N. Kok, J. M. Pea, A. Siebes, and A. J. Feelders, editors, *IDA*, volume 3646 of *Lecture Notes in Computer Science*, pages 133–144, 2005.

[64] D. Gunetti and G. Ruffo. Intrusion Detection through Behavioral Data. In *Proceedings of the Third International Symposium on Advances in Intelligent Data Analysis*, IDA '99, pages 383–394, 1999.

[65] F. Gutirrez, M. Lerma-Rascn, L. Salgado-Garza, and F. Cant. Biometrics and Data Mining: Comparison of Data Mining-Based Keystroke Dynamics Methods for Identity Verification. In C. C. Coello, A. de Albornoz, L. Sucar, and O. Battistutti, editors, *MICAI 2002: Advances in Artificial Intelligence*, volume 2313 of *Lecture Notes in Computer Science*, pages 221–245. 2002.

[66] A. Guven and I. Sogukpinar. Understanding users' keystroke patterns for computer access security. *Computers & Security*, 22(8):695 – 706, 2003.

[67] S. Haider, A. Abbas, and A. K. Zaidi. A multi-technique approach for user identification through keystroke dynamics. volume 2, pages 1336–1341, 2000.

[68] N. Harun, W. L. Woo, and S. S. Dlay. Performance of keystroke biometrics authentication system using artificial neural network (ann) and distance classifier method. In *Computer and Communication Engineering (ICCCE), 2010 International Conference on*, pages 1–6, May 2010.

[69] K. Hempstalk, E. Frank, and I. Witten. One-Class Classification by Combining Density and Class Probability Estimation. In W. Daelemans, B. Goethals, and K. Morik, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 5211 of *Lecture Notes in Computer Science*, pages 505–519. 2008.

[70] S. Hocquet, J.-Y. Ramel, and H. Cardot. Fusion of methods for keystroke dynamic authentication. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pages 224 – 229, oct. 2005.

[71] S. Hocquet, J.-Y. Ramel, and H. Cardot. User Classification for Keystroke Dynamics Authentication. In S.-W. Lee and S. Li, editors, *Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 531–539. 2007.

[72] D. Hosseinzadeh and S. Krishnan. Gaussian Mixture Modeling of Keystroke Patterns for Biometric Applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(6):816 –826, Nov. 2008.

[73] D. Hosseinzadeh, S. Krishnan, and A. Khademi. Keystroke Identification Based on Gaussian Mixture Models. In *IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings*, volume 3, pages III–1144 – III–1146, May 2006.

[74] J. Hu, D. Gingrich, and A. Sentosa. A k-Nearest Neighbor Approach for User Authentication through Biometric Keystroke Dynamics. In *IEEE International Conference on Communications*, pages 1556 –1560, May 2008.

[75] S. Hwang, H. joo Lee, and S. Cho. Improving authentication accuracy using artificial rhythms and cues for keystroke dynamics-based authentication. *Expert Systems with Applications*, 36(7):10649 – 10656, 2009.

[76] A. K. Jain, R. Bolle, and S. Pankanti. *Biometrics: Personal Identification in Networked Society*, chapter Introduction To Biometrics. Springer, 1 edition, January 1999.

[77] A. K. Jain, R. Bolle, and S. Pankanti. Introduction to Biometrics, 2002.

[78] R. Janakiraman and T. Sim. Keystroke Dynamics in a General Setting. In *Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 584–593. 2007.

[79] C.-H. Jiang, S. Shieh, and J.-C. Liu. Keystroke statistical learning model for web authentication. In *Proceedings of the 2nd ACM symposium on Information, computer and communications security*, ASIACCS '07, pages 359–361, 2007.

[80] Z. Jin, A. Teoh, T. S. Ong, and C. Tee. Typing dynamics biometric authentication through fuzzy logic. In *International Symposium on Information Technology*, volume 3, pages 1 –6, Aug. 2008.

[81] B. E. John. TYPIST: A Theory of Performance in Skilled Typing. *Human-Computer Interaction*, 11:321–355, December 1996.

[82] S. S. Joshi and V. V. Phoha. Investigating hidden markov models capabilities in anomaly detection. In *Proceedings of the 43rd annual Southeast regional conference - Volume 1*, pages 98–103, 2005.

[83] R. Joyce and G. Gupta. Identity authentication based on keystroke latencies. *Communications of the ACM*, 33(2):168–176, 1990.

[84] Jugurta R.M.F. and Freire O.E. On the equalization of keystroke timing histograms. *Pattern Recognition Letters*, 27:1440–1446, October 2006.

[85] V. Kacholia and S. Pandit. Biometric Authentication using Random distributions (BioART). In *Canadian IT Security Symposium*, 2003.

[86] P. Kang, S. Park, S. seob Hwang, H. joo Lee, and S. Cho. Improvement of keystroke data quality through artificial rhythms and cues. *Computers & Security*, 27(1-2):3 – 11, 2008.

[87] M. Karnan and M. Akila. Personal Authentication Based on Keystroke Dynamics Using Soft Computing Techniques. In *Second International Conference on Communication Software and Networks*, pages 334 –338, Feb. 2010.

[88] M. Karnan, M. Akila, and N. Krishnaraj. Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing*, 11(2), March 2011.

[89] P. Khanna and M. Sasikumar. Recognising Emotions from Keyboard Stroke Pattern. *International Journal of Computer Applications*, 11(9), December 2010.

[90] K. Killourhy and R. Maxion. The Effect of Clock Resolution on Keystroke Dynamics. In R. Lippmann, E. Kirda, and A. Trachtenberg, editors, *Recent Advances in Intrusion Detection*, volume 5230 of *Lecture Notes in Computer Science*, pages 331–350. 2008.

[91] K. S. Killourhy and R. A. Maxion. Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. In *IEEE/IFIP International Conference on Dependable Systems Networks*, pages 125–134, July 2009.

[92] A. Lammers and S. Langenfeld. Identity authentication based on keystroke latencies using neural networks. *Journal of Computing Sciences in Colleges*, 6:48–51, April 1991.

[93] H. Lee and S. Cho. Retraining a keystroke dynamics-based authenticator with impostor patterns. *Computers & Security*, 26(4):300 – 310, 2007.

[94] J.-W. Lee, S.-S. Choi, and B.-R. Moon. An evolutionary keystroke authentication based on ellipsoidal hypothesis space. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, GECCO '07, pages 2090–2097, 2007.

[95] J. Leggett and G. Williams. Verifying identity via keystroke characteristics. *Int. J. Man-Mach. Stud.*, 28(1):67–76, 1988.

[96] J. Leggett, G. Williams, M. Usnick, and M. Longnecker. Dynamic identity verification via keystroke characteristics. *International Journal of Man-Machine Studies*, 35(6):859 – 870, 1991.

[97] D.-T. Lin. Computer-access authentication with neural network based keystroke identity verification. In *International Conference on Neural Networks*, volume 1, pages 174 –178, June 1997.

[98] H.-R. Lv, Z.-L. Lin, W.-J. Yin, and J. Dong. Emotion recognition based on pressure sensor keyboards. pages 1089 –1092, Apr. 2008.

[99] H.-R. Lv and W.-Y. Wang. Biologic verification based on pressure sensor keyboards and classifier fusion techniques. *IEEE Transactions on Consumer Electronics*, 52(3):1057 –1063, Aug. 2006.

[100] D. Mahar, R. Napier, M. Wagner, W. Laverty, R. Henderson, and M. Hiron. Optimizing digraph-latency based biometric typist verification systems: inter and intra typist differences in digraph latency distributions. *International Journal of Human-Computer Studies*, 43(4):579 – 592, 1995.

[101] E. Maiorana, P. Campisi, N. González-Carballo, and A. Neri. Keystroke dynamics authentication for mobile phones. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC '11, pages 21–26, 2011.

[102] S. Mandujano and R. Soto. Deterring password sharing: user authentication via fuzzy c-means clustering applied to keystroke biometric data. In *Proceedings of the Fifth Mexican International Conference in Computer Science*, pages 181 – 187, Sept. 2004.

[103] J.-D. Marsters. *Keystroke Dynamics as a Biometric*. PhD thesis, University of Southampton, June 2009.

[104] A. Meszaros, Z. Banko, and L. Czuni. Strengthening Passwords by Keystroke Dynamics. In *4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, pages 574 –577, sept. 2007.

[105] S. Modi and S. J. Elliott. Keystroke Dynamics Verification Using a Spontaneously Generated Password. In *40th Annual IEEE International Carnahan Conferences Security Technology Proceedings*, pages 116 –121, Oct. 2006.

[106] F. Monrose, M. K. Reiter, and S. Wetzel. Password hardening based on keystroke dynamics. *International Journal of Information Security*, 1:69–83, 2002.

[107] F. Monrose and A. Rubin. Authentication via keystroke dynamics. In *Proceedings of the 4th ACM Conference on Computer and Communications Security*, pages 48–56, 1997.

[108] J. Montalváo, C. Almeida, and E. Freire. Equalization of keystroke timing histograms for improved identification performance. In *International Telecommunications Symposium*, pages 560 –565, Sept. 2006.

[109] R. Napier, W. Laverty, D. Mahar, R. Henderson, M. Hiron, and M. Wagner. Keyboard user verification: toward an accurate, efficient, and ecologically valid algorithm. *International Journal of Human-Computer Studies*, 43(2):213 – 222, 1995.

[110] M. Nisenson, I. Yariv, R. El-Yaniv, and R. Meir. Towards Behaviometric Security Systems: Learning to Identify a Typist. In N. Lavrac, D. Gamberger, L. Todorovski, and H. Blockeel, editors, *Knowledge Discovery in Databases: PKDD 2003*, volume 2838 of *Lecture Notes in Computer Science*, pages 363–374. 2003.

[111] N. Nishiuchi, S. Komatsu, and K. Yamanaka. Biometric verification using the motion of fingers: a combination of physical and behavioural biometrics. *International Journal of Biometrics*, 2(3):222 – 235, 2010.

[112] M. Obaidat and D. Macchairolo. A multilayer neural network system for computer access security. *IEEE Transactions on Systems, Man and Cybernetics*, 24(5):806 –813, May. 1994.

[113] M. Obaidat and D. Macchiarolo. An online neural network system for computer access security. *IEEE Transactions on Industrial Electronics*, 40(2):235 –242, Apr. 1993.

[114] M. S. Obaidat and B. Sadoun. Verification of Computer Users Using Keystroke Dynamics. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 27(2):261–269, Apr. 1997.

[115] C. S. Ong and W. K. Lai. Enhanced password authentication through typing biometrics with k-means clustering algorithm. In *World Automation Congress*, 2000.

[116] D. J. Ostry. *Tutorials in Motor Behavior*, chapter Execution-Time Movement Control, pages 457–468. Elsevier Science Publishers B. V., 2 edition, 1985.

[117] N. Pavaday and K. Soyjaudah. Investigating performance of neural networks in authentication using keystroke dynamics. In *AFRICON 2007*, pages 1–8, sept. 2007.

[118] N. Pavaday and K. M. S. Soyjaudah. Performance of the K Nearest Neighbor in Keyboard Dynamic Authentication. In *Proceedings of the 2007 Computer Science and IT Education Conference*, pages 599–604, 2007.

[119] A. Peacock, X. Ke, and M. Wilkerson. Typing patterns: a key to user identification. *IEEE, Security Privacy*, 2(5):40–47, Sep. 2004.

[120] G. Z. Pedernera, S. Sznur, G. S. Ovando, S. Garcia, and G. Meschino. Revisiting clustering methods to their application on keystroke dynamics for intruder classification. In *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, pages 36 –40, Sept. 2010.

[121] K. Revett. A Bioinformatics Based Approach to Behavioural Biometrics. In *Frontiers in the Convergence of Bioscience and Information Technologies*, pages 665–670, Oct. 2007.

[122] K. Revett. *Behavioral Biometrics: A Remote Access Approach*, chapter Keystroke Dynamics, pages 73–136. John Wiley & Sons, Ltd, 2008.

[123] K. Revett, S. T. de Magalhaes, and H. Santos. Data Mining a Keystroke Dynamics Based Biometrics Database Using Rough Sets. In *Portuguese Conference on Artificial Intelligence.*, pages 188 –191, Dec. 2005.

[124] K. Revett, S. de Magalhes, and H. Santos. Enhancing login security through the use of keystroke input dynamics. In D. Zhang and A. Jain, editors, *Advances in Biometrics*, volume 3832 of *Lecture Notes in Computer Science*, pages 661–667. 2005.

[125] K. Revett, F. Gorunescu, M. Gorunescu, M. Ene, S. T. d. Magalhaes, and H. M. D. Santos. A machine learning approach to keystroke dynamics based user authentication. *Int. J. Electron. Secur. Digit. Forensic*, 1:55–70, May 2007.

[126] S. Reynolds. Keystroke Dynamics Format for Data Interchange. Technical Report M1/05-0303, InterNational Committee for Information Technology Standards, May 2005.

[127] J. A. Robinson, V. M. Liang, J. A. M. Chambers, and C. L. MacKenzie. Computer User Verification Using Login String Keystroke Dynamics. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 28(2):236–241, Mar. 1998.

[128] D. A. Rosenbaum, V. Hindorff, and E. M. Munro. Scheduling and Programming of Rapid Finger Sequences: Tests and Elaborations of the Hierarchical Editor Model. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2):193 – 203, 1987.

[129] W. G. Ru and J. H. P. Eloff. Enhanced password authentication through fuzzy logic. *IEEE Expert*, 12(6):38–45, Nov. 1997.

[130] D. E. Rumelhart and D. A. Norman. Simulating a skilled typist: a study of skilled cognitive-motor performance. *Cognitive Science*, 6(1):1 – 36, 1982.

[131] M. Rybnik, M. Tabedzki, and K. Saeed. A Keystroke Dynamics Based System for User Identification. pages 225 – 230, Jun. 2008.

[132] H. Saevanee and P. Bhattarakosol. Authenticating User Using Keystroke Dynamics and Finger Pressure. In *6th IEEE Conference on Consumer Communications and Networking Conference*, pages 1–2, Jan. 2009.

[133] G. Saggio, G. Costantini, and M. Todisco. Cumulative and Ratio Time Evaluations in Keystroke Dynamics To Improve the Password Security Mechanism. *Journal of Computer and Information Technology*, 1:2–11, Nov 2011.

[134] T. A. Salthouse. Perceptual, Cognitive, and Motoric Aspects of Transcription Typing. *Psychological Bulletin*, 99(3):303 – 319, 1986.

[135] T. Samura and H. Nishimura. Keystroke timing analysis for individual identification in Japanese free text typing. In *ICCAS-SICE*, pages 3166 –3170, Aug. 2009.

[136] Y. Sang, H. Shen, and P. Fan. Novel Impostors Detection in Keystroke Dynamics by Support Vector Machine. In K.-M. Liew, H. Shen, S. See, W. Cai, P. Fan, and S. Horiguchi, editors, *Parallel and Distributed Computing: Applications and Technologies*, volume 3320 of *Lecture Notes in Computer Science*, pages 37–38. 2005.

[137] L. Shaffer. *Attention and Performance Vol. IV*, chapter Latency Mechanisms in Transcription. Academic Press, 1973.

[138] L. Shaffer. *Tutorials in Motor Neuroscience*, chapter Cognition and Motor Programming. Kluwer Academic Publishers, 1991.

[139] L. H. Shaffer. Timing in the motor programming of typing. *Quarterly Journal of Experimental Psychology*, 30(2):333–345, 1978.

[140] D. Shanmugapriya and G. Padmavathi. A survey of biometric keystroke dynamics: Approaches, security and challenges. *International Journal of Computer Science and Information Security*, 5, 2009.

[141] Y. Sheng, V. Phoha, and S. Rovnyak. A parallel decision tree-based method for user authentication based on keystroke patterns. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(4):826 –833, Aug. 2005.

[142] S. J. Shepherd. Continuous authentication by analysis of keyboard typing characteristics. In *Security and Detection, 1995., European Convention on*, pages 111 –114, May 1995.

[143] R. J. Spillane. Keyboard Apparatus for Personal Identification. Technical Disclosure Bulletin 17, 3346, IBM, 1975.

[144] D. Stefan and D. Yao. Keystroke-Dynamics Authentication Against Synthetic Forgeries. In *International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2010.

[145] A. Sulong, Wahyudi, and M. U. Siddiqi. Intelligent keystroke pressure-based typing biometrics authentication system using radial basis function network. In *5th International Colloquium on Signal Processing Its Applications*, pages 151 –155, Mar. 2009.

[146] M. Tapiador and J. A. Sigüenza. Fuzzy Keystroke Biometrics On Web Security. In *AutoID, Proceedings Workshop on Automatic Identification Advanced Technologies IEEE*, 1999.

[147] C. Tappert, M. Curtin, M. Villani, G. Ngo, J. Simone, H. S. Fort, and S. Cha. Keystroke biometric recognition on long-text input: A feasibility study. In *Proceedings of the 23rd International Biometric Conference*, July 2006.

[148] P. S. Teh, A. Teoh, T. S. Ong, and H. F. Neo. Statistical Fusion Approach on Keystroke Dynamics. In *Third International IEEE Conference on Signal-Image Technologies and Internet-Based System*, pages 918 –923, Dec. 2007.

[149] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Elsevier, 2009.

[150] E. A. C. Thomas and R. G. Jones. A model for subjective grouping in typewriting. *Quarterly Journal of Experimental Psychology*, 22(3):353–367, 1970.

[151] D. Umphress and G. Williams. Identity verification through keyboard characteristics. *International Journal of Man-Machine Studies*, 23(3):263 – 273, 1985.

[152] F.-M. E. Uzoka and T. Ndzinge. An investigation of factors affecting biometric technology adoption in a developing country context. *International Journal of Biometrics*, 1(3):307 – 328, 2009.

[153] J. R. Vacca. *Biometric Technologies and Verification Systems*. Butterworth-Heinemann, 1 edition, 2007.

[154] W. B. Verwey and Y. Dronkert. Practicing a Structured Continuous Key-Pressing Task: Motor Chunking or Rhythm Consolidation? *Journal of Motor Behavior*, 28(1):71–79, 1996.

[155] M. Villani, C. Tappert, N. Giang, J. Simone, H. S. Fort, and S.-H. Cha. Keystroke Biometric Recognition Studies on Long-Text Input under Ideal and Application-Oriented Conditions. In *Conference on Computer Vision and Pattern Recognition Workshop*, pages 39 – 39, June 2006.

[156] L. B. William and N. Harter. Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, 6(4):345 – 375, July 1899.

[157] F. Wong, M. H. Wong, A. S. M. Supian, A. F. Ismail, L. W. Kin, and O. C. Soon. Enhanced user authentication through typing biometrics with artificial neural networks and k-nearest neighbor algorithm. In *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 911–915, 2001.

[158] D. Woodard. *Exploiting Finger Surface as a Biometric Identifier*. PhD thesis, Notre Dame University, December 2004.

[159] K. Xi, Y. Tang, and J. Hu. Correlation Keystroke Verification Scheme for User Access Control in Cloud Computing Environment. *The Computer Journal*, 11:1632–1644, July 2011.

[160] R. V. Yampolskiy and V. Govindaraju. Behavioural biometrics: a survey and classification. *International Journal of Biometrics*, 1(1):81 – 113, 2008.

[161] S. Yong, W. K. Lai, and G. Goghill. Weightless Neural Networks for Typing Biometrics Authentication . In *Knowledge-Based Intelligent Information and Engineering Systems*, volume 3214 of *Lecture Notes in Computer Science*, pages 284–293. 2004.

[162] E. Yu and S. Cho. Keystroke dynamics identity verification–its problems and practical solutions. *Computers & Security*, 23(5):428–440, 2004.