COMPUTERS & SECURITY 59 (2016) 210–235

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

# A survey on touch dynamics authentication in mobile devices

*Pin Shen Teh* [a,*], *Ning Zhang* [a], *Andrew Beng Jin Teoh* [b], *Ke Chen* [a]

[a] *School of Computer Science, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom*
[b] *School of Electrical and Electronic Engineering, Yonsei University, Seoul 120-749, Republic of Korea*

ARTICLE INFO

*Article history:*
Received 23 September 2015
Received in revised form 8 March 2016
Accepted 14 March 2016
Available online 18 March 2016

*Keywords:*
Biometrics
Touch dynamics
Authentication
Access control
Mobile device security
Keystroke dynamics

ABSTRACT

There have been research activities in the area of keystroke dynamics biometrics on physical keyboards (desktop computers or conventional mobile phones) undertaken in the past three decades. However, in terms of touch dynamics biometrics on virtual keyboards (modern touchscreen mobile devices), there has been little published work. Particularly, there is a lack of an extensive survey and evaluation of the methodologies adopted in the area. Owing to the widespread use of touchscreen mobile devices, it is necessary for us to examine the techniques and their effectiveness in the domain of touch dynamics biometrics. The aim of this paper is to provide some insights and comparative analysis of the current state of the art in the topic area, including data acquisition protocols, feature data representations, decision making techniques, as well as experimental settings and evaluations. With such a survey, we can gain a better understanding of the current state of the art, thus identifying challenging issues and knowledge gaps for further research.



## 1. Introduction

Technological advancement in computing, communicational devices, as well as network connectivity is shifting the usage of conventional desktop computers to mobile devices. It is predicted that there will be a staggering amount of over 2 billion smartphone users worldwide by the year 2016 (emarketer, 2014). The increasing reliance on these devices inevitably implies the increase in sensitive data stored on this platform. Unfortunately, the portability of mobile devices also makes it vulnerable to theft (Raghunathan et al., 2003). Data leakage and misuse of stolen device are potentially more damaging than the cost of the device itself (Crawford et al., 2013).

Knowledge-based authentication methods, such as passwords, PINs or pattern locks (hereafter referred to as passcode), are still the primary methods used to authenticate mobile users (Khan et al., 2014). However, these methods are vulnerable to a number of security threats or attacks, including brute force attacks (Kim, 2012), shoulder surfing (Zakaria et al., 2011), and smudge attacks (Giuffrida et al., 2014). Also, the usage pattern of mobile devices is usually in short bursts of intervals (Frank et al., 2013), and this significantly increases the frequency of authentication that is required as compared to the case of desktop computers. Higher authentication frequency lowers usability (Harbach et al., 2014). To balance the conflict between security and usability, measures such as the delayed authentication, e.g. a device is set to have some idle time before reauthentication requests are required, can be used (Sen and Muralidharan, 2014). Even so, the issue of how to enhance security, while, at the same time, still be able to maintain minimal user intervention or maximum usability remains unsolved.

An alternative to the passcode approach to authentication is the biometrics authentication. The latter identifies a person based on his/her physiological or behavioral characteristics. Physiological biometrics is a relatively stable physical

* *Corresponding author.*
*E-mail address:* pinshen.teh@manchester.ac.uk (P.S. Teh).
http://dx.doi.org/10.1016/j.cose.2016.03.003
0167-4048/

feature of a human body, such as fingerprint, facial characteristics, and iris pattern. Behavioral biometrics, on the other hand, is traits that are acquired from human behavior or habits like signature, voice, gait and keystroke dynamics. Biometrics authentication methods are considered to be more secure than other authentication methods (e.g. passcode) because biometrics cannot be lost or stolen (if used securely) and is harder to be forged (Jain et al., 2008). To ensure that a biometrics authentication method is effective, the method should not only be secure, but also usable. Hanul Sieger et al. (2010) conducted a survey to study how participants ranked the perceived security protection and the usability (willingness to use the method) of different types of biometrics authentication methods on mobile devices. The survey reported that the iris and voice biometrics authentication methods were ranked the highest in terms of the perceived security protection, but the lowest in terms of usability. This shows that a biometrics authentication method, regarded as providing a stronger level of security protection, may not be the most usable method.

According to the survey results reported by De Luca et al. (2015), one of the important factors influencing the participants' choice as to which biometrics authentication method they prefer to use on a mobile device is the usability factor. They identified two primary usability issues that put off participants' adaptation of a biometrics authentication method, namely: (i) slow authentication speed and inconvenience, and (ii) social awkwardness. For the first issue, for example, in the case of face biometrics, participants felt that it was difficult and time consuming to align the face correctly in front of the device's camera. In the case of fingerprint biometrics, the participants felt that it was hard to scan a fingerprint properly when the fingers were too oily or dry, or when the device was covered with a protective casing. For the second issue (i.e. social awkwardness), for example, participants felt that it was awkward to hold a device in front of a face to perform an authentication task in a public area. This is more so the case in the context of mobile devices, where the use of these devices in a public area is very common and frequent.

Touch dynamics is a behavioral biometrics, which captures the way a person touches on a touchscreen device. Similar to other biometrics data, touch dynamics biometrics can be used to identify a person/user, and can also be used in conjunction with a passcode authentication method to achieve an enhanced level of security in user authentication and in the protection of mobile devices. This method can be implemented by employing existing sensors embedded in a mobile device, so it is comparatively cheaper than other biometrics authentication method. In addition, this method is non-intrusive and can operate in parallel with a person's normal mobile device usage activities (Shen et al., 2016). The existing passcode authentication method has a wide social acceptance, and the touch dynamics authentication method is also expected to be widely acceptable by the general public (Campisi et al., 2009).

Owing to the potential of touch dynamics biometrics, there have been increasing research efforts in this topic area, as shown in Fig. 1. This paper provides a survey of these efforts discussing their major contributions and identifying issues for further research. The main contributions of this paper are threefolds. Firstly, it presents a comprehensive survey of published works in the topic area of touch dynamics biometrics highlighting their contributions and technological advances in the topic area. Secondly, it critically analyzes these related works from a range of perspectives, leading to the identification of knowledge gaps and issues for further research. Finally, the references cited in this paper provide a useful lead into this topic area.

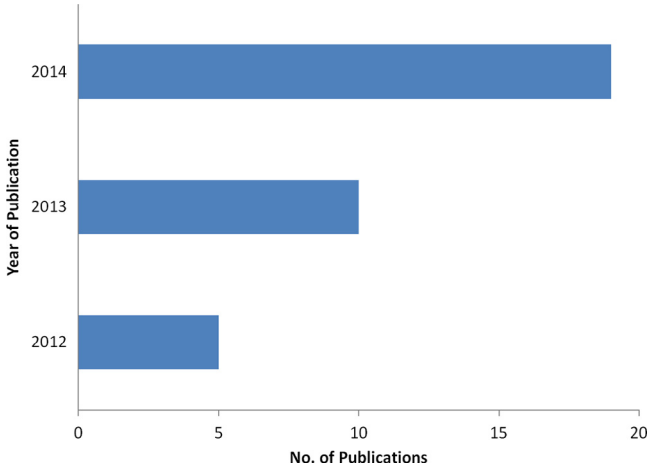


**Fig. 1 – The increasing trend of research works on touch dynamics biometrics.**

In detail, the structure of the paper is as follows. Section 2 provides an overview of touch dynamics biometrics in general. Sections 3–8, respectively, compare the related works in terms of their experimental designs, data acquisition methods, feature selection strategies, decision making techniques, fusion approaches, and data adaptation approaches. Their performances are discussed in Section 9. The identified knowledge gaps and issues for further research are outlined in Section 10. Finally, Section 11 concludes the paper. To the best of our knowledge, there has not been a similar paper published in literature at the time of this writing.

## 2. Touch dynamics biometrics

### 2.1. *Overview*

Touch dynamics biometrics refers to the process of measuring and assessing human touch rhythm on touchscreen mobile devices (e.g. smartphones and digital tablets). A form of digital signatures is generated upon human interactions with these devices. These signatures are believed to be discriminative and unique for each individual, so may be used as a personal identifier.

In the 1860s, when the telegraph was the main method for long distance communication, operators "identified" each other through ways in which they tapped on telegraph keys (Bryan and Harter, 1897). Today, telegraph keys have been replaced by computer keyboards, mobile keypads, and virtual keyboards. Computer keyboards have been the most common input devices since the late 20th century. It is well-known that human keyboard typing patterns are unique so they could be used as a personal identifier (Obaidat and Sadoun, 1996).

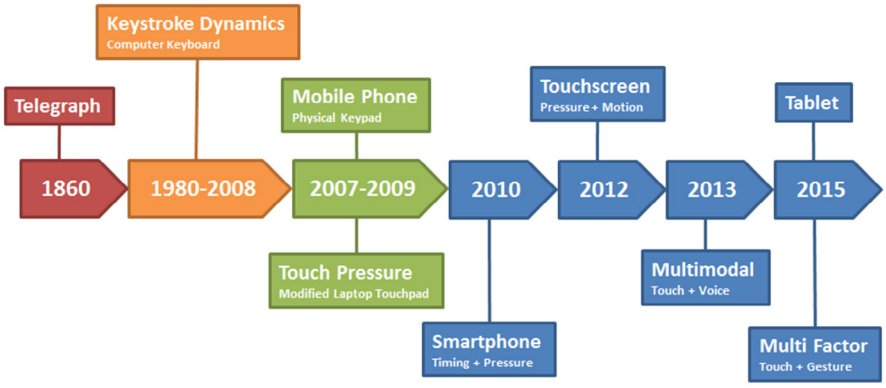


**Fig. 2 – The evolution of touch dynamics biometrics research.**

One of the earliest research works on keystroke dynamics authentication was conducted by Gaines et al. (1980). They carried out an experiment to try to recognize 6 professional secretaries by analyzing the way they typed three passages of texts consisting of 300 to 400 words each. Since then, many related efforts have been made. Crawford, Karnan et al. and Teh et al. have, independently, written surveys of the published works on keystroke dynamics authentication (Crawford, 2010; Karnan et al., 2011; Teh et al., 2013). However, these early works largely focused on keystroke dynamics authentication on computer keyboards. With the rapid development of mobile communication technologies, more recent research efforts in this area have been focused on mobile devices with physical keypads (Campisi et al., 2009; Clarke and Furnell, 2007; McLoughlin and Naidu, 2009). Most recently, research activities are largely carried out in the context of touchscreen mobile devices. Fig. 2 summarizes the timelines of the touch dynamics biometrics research as influenced by technological developments in the sector.

Touch dynamics biometrics have their unique merits or useful features, while at the same time, they also introduce challenging issues. The sections below summarize the features and the challenging issues.

### 2.2. *Features*

A touch dynamics authentication system can offer a number of useful features compared to the other types of biometrics authentication system. These are the following.

**Distinctiveness:** Touch dynamics patterns are capable of generating multi-dimensional features, such as timing, spatial and motion features. These multi-dimensional features can be measured up to a precision level that is significantly higher than human perception (Zheng et al., 2014). These unique features are hard to replicate consistently, and therefore can be used for authentication.

**Enhanced Security:** Despite its weaknesses, passcodes have been the most widely accepted and deployed authentication method (Schlöglhofer and Sametinger, 2012). By integrating touch dynamics biometrics into passcode authentication method, the overall assurance level can be increased.

**Continuous Monitoring:** Touch dynamics biometrics can be used to verify the authenticity of a user beyond the initial authentication by constantly monitoring the user touch dynamics patterns. In other words, user reauthentication can be performed easily and non-intrusively throughout an active login session. In this way, security protection goes beyond initial login without compromising usability. This is one of the most notable advantages touch dynamics biometrics have over other physiological biometrics.

**Revocability:** In an event when a passcode associated with a touch dynamics template is compromised, a new touch dynamics template can easily be generated when a new passcode is created. This is not the case for other physiological biometrics. For example, with iris or face biometrics, once they are compromised, there will be no replacement, and for fingerprints biometrics, the number of replacements is limited (humans have only 10 fingers to use after all).

**Non-dependency:** A mobile device usually operates in an on-the-go manner, so the surrounding lighting condition and background noise level are, in most cases, constantly changing. In comparison with other biometrics features, such as face and voice biometrics, the feature acquisition of touch dynamics biometrics is less sensitive to these environmental factors. Therefore, it is more suited to, and can be more easily deployed to a mobile device.

**Transparency:** Touch dynamics authentication system requires little or no additional interventions from a mobile device user. This is because the acquiring and processing of touch dynamics patterns can be carried out in the background while the user is using the device. Users may not be aware that their touch dynamics patterns are being captured, the captured data are being used for authentication, and the authentication is carried out periodically or they are protected by an extra layer of authentication. This is in a stark contrast to other biometrics authentication systems that usually require explicit alignment of a biometrics feature to a specific sensor. For example, in the case of iris authentication, a user is required to look straight into a camera to take an iris image, and in the case of fingerprint authentication, a user needs to put one of his/her fingers on the fingerprint sensor.

**Familiarity:** The touch dynamics data used for authentication is acquired during mobile users' routine input activities. This is a process which mobile users are already familiar with, so the data acquisition operation tends to have a gentler learning curve with a higher usability level than other biometrics data acquisition cases.

**Cost Effectiveness:** In contrast to other physiological biometrics authentication methods such as iris and fingerprint biometrics that typically require the use of specialized hardware, touch dynamics authentication system only uses built-in mobile sensors. This can reduce device costs and it is ideal for large-scale deployments.

### 2.3. *Challenging issues*

The design of a touch dynamics authentication system imposes a number of challenging issues as follows.

**Minimizing Computation and Communication Costs:** Computational capabilities of mobile devices are typically lower than desktop computers. This means that certain criteria such as algorithm complexity, communication cost, and authentication delay are important and should be considered in the design of touch dynamics authentication solutions. In other words, algorithm and communication costs introduced as the result of deploying this authentication means should be minimal.

**Minimizing Energy Consumption:** Mobile devices, unlike desktop computers, are operated by batteries. The less the energy an application consumes, the longer the device can operate. Though communication is the major consumer of the battery power of a mobile device (Perrucci et al., 2009), the number and usage frequencies of various sensors embedded in a mobile device, which are used to extract touch dynamics data, also have a direct impact on the mobile device battery consumption. Various measures, such as reducing the sampling rate (Niu and Chen, 2012) or performing complex computation only when a device is being recharged (Crawford et al., 2013), have been proposed to reduce power consumption of a mobile device.

**Maximizing Accuracy:** The accuracy performance of touch dynamics authentication system is relatively low in comparison to other physiological biometrics authentication system (e.g. fingerprint and iris). This is because touch dynamics biometrics features (or feature data) acquired at different occasions are likely to exhibit a certain degree of variations due to external factors such as fatigue, mood, or distraction. Therefore, consideration should be given as to how to increase the accuracy performance of a touch dynamics authentication system in the design of the method.

**Adaptation Capability:** Human behavioral characteristics typically change over time, and they usually change more frequently than physiological characteristics. A user's touch dynamics patterns can gradually change as the user gets more familiar with the passcode, input method, device, and other external factors. A touch dynamics authentication system should be capable of adapting itself to any changes in a user's touch dynamics pattern.

### 2.4. *Operational process*

A typical touch dynamics authentication system is illustrated in Fig. 3. From the figure, we can see that the operation of this system can largely be captured in three major phases: (i) User Enrollment, where touch dynamics data (or samples) are acquired, processed, and stored as a reference template; (ii) User Authentication, where a touch dynamics test sample is compared against the stored reference template(s) to determine the similarity or dissimilarity; and (iii) Data Retraining, where reference template is updated to reflect any changes in the latest touch dynamics data. The three operational phases are accomplished by a number of functional blocks (i.e. architectural components), each of which performs a well-defined function, and these components and their respective functions are described below.

#### 2.4.1. *Data acquisition*

Data acquisition is an operation by which raw touch dynamics data are acquired. This is usually carried out as the first step and during the setup stage of a touch dynamics authentication system. The acquired raw data are usually a set of repeated (multiple) input samples acquired over a specified period of time. Devices commonly used for data acquisition are commercial off-the-shelf smartphones (Buschek et al., 2015; Trojahn et al., 2013; Zheng et al., 2014) or, in some cases, digital tablets (Saravanan et al., 2014).

#### 2.4.2. *Data preprocessing*

Data preprocessing is carried out to remove outliers in the raw data, improving data quality and accuracy performance. Techniques used in this operation include outlier detection and removal (Zheng et al., 2014). A dimension reduction technique may also be used to ensure that raw data remain small yet representable, for the sake of computational efficiency on resource limited mobile devices (de Mendizabal-Vazquez et al., 2014).

#### 2.4.3. *Feature extraction*

Feature extraction is a mandatory operation that is carried out in both user enrollment and authentication phases. The main

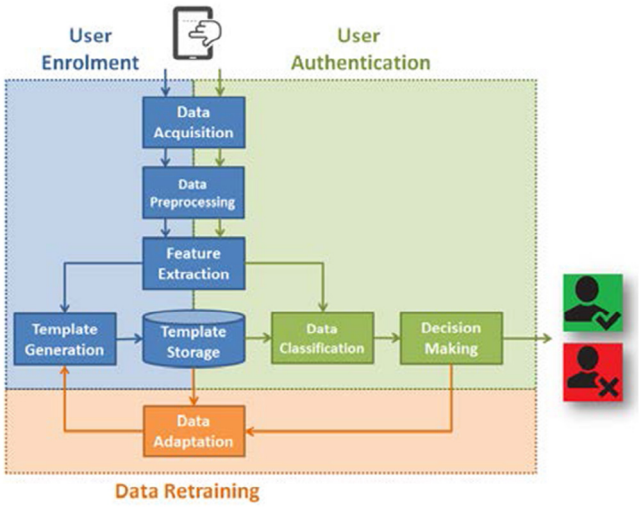


**Fig. 3 – A touch dynamics authentication framework.**

task of this operation is to identify and extract distinctive features common to a user from the acquired raw data. These features will be later used for template generation. Possible features extracted from human touch dynamics data can be categorized into three broad categories, namely timing, spatial and motion features (further discussions in Section 5).

#### 2.4.4. *Template generation*

Template generation is an operation carried out to transform the extracted feature into a compact form that uniquely represents the user's touch dynamics characteristic. Normally, several different types of features are concatenated into a sequence of n-dimensional feature vectors, where n is the number of feature elements (Cai et al., 2013; Serwadda et al., 2013). These unique reference templates are then stored for user authentication or data retraining purpose.

#### 2.4.5. *Data classification*

Data classification is the major operation for most biometrics authentication methods, where feature data are categorized and compared against reference templates. The outcome of this phase is normally associated with a matching score used for decision making. Data classification is usually carried out using machine learning techniques (also referred to as data classification algorithms or classifiers), and a number of machine learning techniques have been used for this purpose in the literature of touch dynamics biometrics (further discussions in Section 6).

#### 2.4.6. *Decision making*

Decision making is an operation carried out to determine if the touch dynamics data submitted by a user are indeed originated from the target user. This decision is made by comparing the similarity or dissimilarity score generated from a machine learning technique against a predefined threshold (Bo et al., 2014; Kolly et al., 2012). Before the final decision is made, a fusion approach may be applied to combine either the information from multiple features (Buschek et al., 2015; Jeanjaitrong and Bhattarakosol, 2013) or to combine the matching scores from different machine learning techniques (Samura et al., 2014), to increase accuracy performance.

#### 2.4.7. *Data adaptation*

Data adaptation is an operation carried out to retrain or update the reference template with the latest touch dynamics patterns from a user. This operation is required because a user's touch dynamics patterns may gradually change over time, causing the initially enrolled reference template to deviate from the most recent touch dynamics patterns from the same user. By adding an adaption component that performs the data adaptation operation after each successful authentication, these gradual changes can be captured and taken into account (Crawford et al., 2013).

### 2.5. *Evaluation criteria*

A touch dynamics authentication system can be deployed in one of the two modes, a verification (or authentication) mode and an identification (or recognition) mode. These modes serve

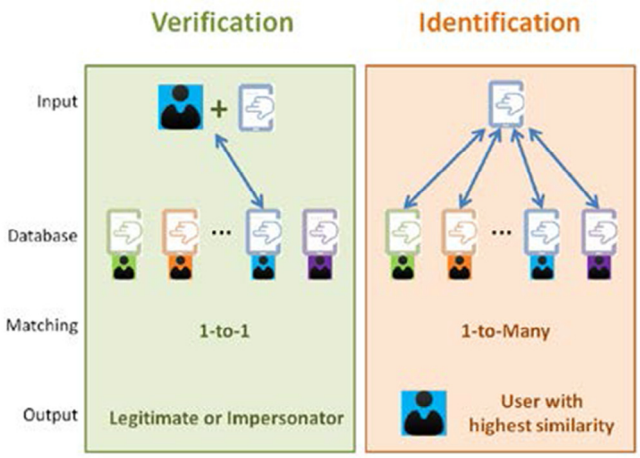


**Fig. 4 – The deployment modes of a touch dynamics authentication system.**

different purposes and usage scenarios. The verification mode is used to verify a claimed identity. It is used to answer the question "is this person whom he/she claims to be". The authentication of a mobile user or a mobile device fits into this mode. The identification mode, on the other hand, is used to classify and identify some unknown identity. It is used to answer questions such as "who is this person" or "is this person in the database". This mode is typically used for forensic investigations or intrusion detections. As shown in Fig. 4, the fundamental difference between the two modes is that, in the verification mode, the checking between the touch dynamics data submitted by a user and the reference template is 1-to-1, whereas, in the identification mode, this checking is 1-to-many. According to our literature survey, the number of papers published on the study of the verification mode (74%) is much higher than the identification mode (26%).

The focus of this paper is on authentication using touch dynamic biometrics, so hereafter our analysis is on the verification mode. To assess the suitability of a biometrics authentication method to real-world applications, three major criteria should be used to evaluate the system. These are verification accuracy, system efficiency, and system usability.

#### 2.5.1. *Verification accuracy*

The metrics that are commonly used to evaluate the verification accuracy of a biometrics authentication method are the false rejection rate (FRR), false acceptance rate (FAR) and equal error rate (EER). The relationship among these metrics is shown in Fig. 5 and their definitions are given below.

*2.5.1.1. False rejection rate (FRR).* This is the percentage ratio of the number of legitimate users who are falsely rejected against the total number of legitimate user trials. A lower FRR value indicates fewer legitimate users being falsely rejected. It also means that the system usability level is higher. FRR is also referred to as a false alarm rate, false negative rate, false non-match rate, or Type II error.

*2.5.1.2. False acceptance rate (FAR).* This is the percentage ratio of the number of illegitimate users who are falsely accepted

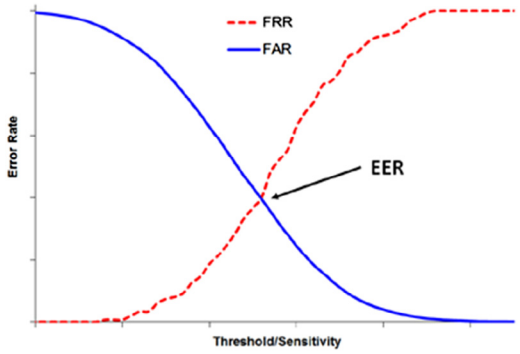


**Fig. 5 – The relationship between the FRR, FAR, and EER.**

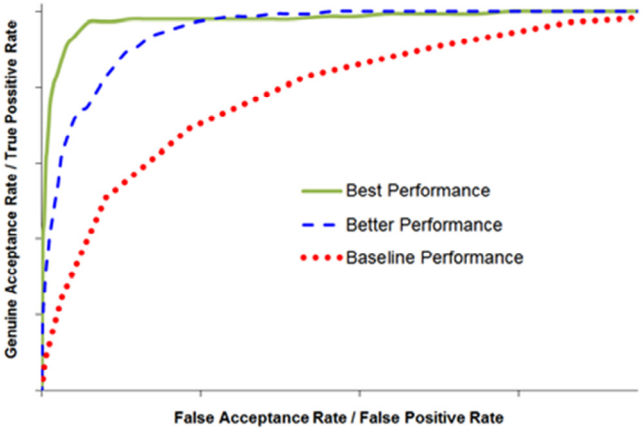


**Fig. 6 – The ROC curves of three performance scenarios.**

against the total number of illegitimate user trials. Again, a lower FAR value indicates fewer illegitimate users being falsely accepted, and this also indicates that the system has a higher security level. FAR is also referred to as miss alarm rate, false positive rate, false match rate, or Type I error.

2.5.1.3. *Equal error rate (EER)*. EER is a single-number performance metric, which is commonly used to measure and compare the overall accuracy level of different biometrics authentication method. It is sometimes also referred to as crossover error rate (CER). EER can be obtained by finding the interception point of two graphs, one for FRR and the other for FAR. Typically, the lower the FRR and the FAR values, the lower the EER value, which in turn indicates a better accuracy performance of a biometrics authentication method. However, FRR and FAR are negatively correlated, so it is not possible to lower both FRR and FAR values at the same time. Therefore, in real-life applications, FRR and FAR are usually adjusted and determined based on the security and usability requirements of the applications. In some literature, the term "accuracy", rather than EER, is used as an accuracy performance metric. It is worth noting that "accuracy" and EER are actually the same; "accuracy" is defined as the inverse of EER. In other words, a higher "accuracy" value indicates a better accuracy performance of a biometrics authentication method.

The accuracy performance can also be graphically visualized by using the receiver operating characteristic (ROC) curve as shown in Fig. 6. This graph is obtained by plotting genuine acceptance rate (GAR) against FAR at different matching threshold values. GAR is the percentage ratio between the correctly accepted legitimate users against the total number of legitimate user trials. It is also referred to as the inverse of FRR (100-FRR), true positive rate, or true match rate. A larger area under the curve (nearer the curve towards the top left corner of the graph) indicates a better performance.

2.5.2. *System efficiency*

The system efficiency refers to the computational cost or the authentication delays imposed by a biometrics authentication method. Satisfying this criterion is particularly important for computational resource-limited mobile devices. A complex authentication method may impose a higher level of computational overhead, increasing authentication delays and reducing system usability. Therefore, it is important to design authentication methods that introduce as low computational overhead as possible.

2.5.3. *System usability*

The system usability (or user acceptance) of an authentication system is also an important factor to the successful deployment of a new authentication method. Users will eventually abandon or reluctant to use any system that is tedious or slow to use, even if it can offer a higher level of security protection. Therefore, an authentication system should offer a good level of system usability and this can be achieved by (i) reducing the workload imposed on a user as much as possible, (ii) requiring users' intervention as less as possible, and (iii) making authentication delays as short as possible.

## 3. Experimental design

### 3.1. *Working mode*

The verification mode may operate in either a static or a dynamic manner. The static and dynamic working modes are complementary to each other, i.e. they may be deployed independently, or alongside with each other to enhance the security protection level afforded to the deployed mobile device. In the following, we discuss the two working modes. Hereafter, we use the term, Verification-in-Static-Mode (ViSM), to refer to the verification mode being used in the static working mode and, Verification-in-Dynamic-Mode (ViDM), to refer to the verification mode being used in the dynamic working mode.

3.1.1. *Verification-in-static-mode (ViSM)*

One application scenario of the ViSM is static authentication, which is also known as one-off authentication. In static authentication, a user attempts to authenticate himself/herself to a system at the beginning of a log-in session or at some predefined intervals during a session. For example, a touch dynamics authentication method may be integrated with an existing passcode authentication method, forming a so-called two-factor authentication system, in which the passcode

authentication method serves as the first factor and the touch dynamics authentication method serves as an additional, i.e. the second, authentication factor. This two-factor authentication system provides a stronger level of protection than any of the two authentication methods when they are used alone. In addition, the use of the second authentication factor can also prevent passcode sharing.

#### 3.1.2. *Verification-in-dynamic-mode (ViDM)*

An example of application scenario of the ViDM is dynamic authentication, also known as continuous authentication. A dynamic authentication method performs authentication checks on a user in an application or communication session (i.e. after the initial authentication is performed). The dynamics feature may be reflected by the use of information that is generated in real-time during the session to authenticate the user and/or the use of multiple instances of authentication in the session, but the intervals between the multiple authentication instances are not predefined, e.g. they may be determined by the occurrence of some touch events. A touch dynamics authentication system is particularly suited to this mode, as touch dynamics data can be acquired transparently over a period of time to revalidate the user's identity without user's intervention, and this may be done at any point during the session. Continuous authentication can reduce security risks in a number of ways such as unauthorized device sharing, device lost/theft, session hijacking, etc. Of course, as in the case of any biometrics authentication method, it is important to achieve a low FRR value to make the system more usable, as, otherwise, a legitimate user may be locked out of the service in the middle of a session. According to our literature survey, more papers have been published for static authentication (77%) than dynamic authentication (23%).

### 3.2. *Acquisition devices*

The device selection for data acquisition is also an important factor. Different devices may be equipped with different sensors that may have the ability to acquire different types of features. For example, a conventional mobile device with a physical keypad is only able to acquire timing feature. In contrast, a recently manufactured touchscreen mobile device is more likely to have multiple more powerful built-in sensors that can acquire more features (such as pressure, movement, and orientation). As the focus of this paper is on touch dynamics authentication, hereafter, we shall only center our discussions on mobile devices with touchscreens. For existing works on keystroke dynamics on mobile devices with physical keypads, readers are referred to review articles such as (Crawford, 2010; Teh et al., 2013).

The majority of the research works carried out in the domain of touch dynamics use smartphones as their data acquisition devices. The only exceptions are the work reported in Antal and Szabó (2014), Bond and Ahmed Awad (2015) and Saravanan et al. (2014), and in these cases, digital tablets were used. This is largely due to the fact that a larger population of mobile device users actually use smartphones rather than digital tablets (Taylor, 2015). Researchers have been using more recent and more powerful mobile devices to carry out their research works. Modern devices usually come with higher precision and resolution sensors that are able to capture higher quality features. Modern devices also have greater computational capabilities and resources, which can better support the use of more complex algorithms and more able sensors. Another device selection criterion is the intended development platform associated with a mobile operating system and this issue is discussed in the next section.

### 3.3. *Development platform*

To acquire touch dynamics data, we need to use some toolkit, and, for the development of the toolkit, we need to choose a development platform. Based on our literature study, Android is the most popular development platform for touch dynamics data acquisition, which is followed by iOS and then by Windows. When selecting a development platform, these four factors should be considered, i.e. its customizability, flexibility, cost and market shares.

#### 3.3.1. *Customizability*

To acquire touch dynamics data, the ability to log various touchscreen input events is essential. However, the native input methods (such as a virtual keyboard or a virtual numeric keypad) used by a mobile operating system do not provide any function calls to acquire these data. This is typically part of the security measure used to protect against the easy implementation of spyware or touch-logger applications (Kambourakis et al., 2014). Therefore, to acquire touch dynamics data, we have to create our own custom input methods with the necessary functionalities. Unlike its competitors (iOS and Windows), Android has made this easier by providing open source library functions, which allow developers to modify the application framework (Meng et al., 2013), giving them greater flexibility in their application design and customization.

#### 3.3.2. *Flexibility*

Flexibility, in terms of cross-platform development, sideloading, and file system visibility are among the criteria that contribute to the popularity of a chosen mobile development platform. Android supports cross-platform development, and this means that developers have the flexibility to develop mobile applications using any operating systems and to make use of existing resources in their developments. Also, both Android and Windows allow sideloading, which means that an application can be directly installed on a mobile device without first publishing it to the mobile application store. Application publishing involves rigid procedures and could be time-consuming. Therefore, sideloading can reduce the time and effort on application testing and developments. Furthermore, direct file explorers have been provided by both Android and Windows providers. This means that data and system files can be accessed directly without additional configurations or installations of any third party applications. This provides a convenient way for researchers to transfer acquired data files between different devices for further analysis.

#### 3.3.3. *Cost*

The cost required for acquiring a development tool and device should also be taken into account when making the

**Table 1 – The different mobile development platforms.**

| | Android | iOS | Windows |
|---|---|---|---|
| Development tool | **Free** | $99/year | **Free** |
| Publishing fee | $25 one-off | | $19/year |
| Sideloading | **Yes** | No | **Yes** |
| Programming language | Java | Swift | C# or VB |
| Open source | **Yes** | No | No |
| Cross platform development | **Yes** | No | No |
| File System visibility | **Yes** | No | **Yes** |
| Market share | **High** | **High** | Low |

selection. The official integrated development environments (IDEs) required for application developments for Android and Windows devices are Android Studio and Visual Studio Express, respectively. They are both available for download free of charge. Xcode IDE (iOS devices), on the other hand, is only available once an annual subscription fee for application publishing has been paid for, and the cost of the fee is shown in Table 1. In addition, Android has been used by a wide range of mobile devices. Among the largest manufacturers of Android-powered mobile devices are Samsung, HTC, and LG. This wider range of devices can provide us with a cheaper option to conduct our experiments.

#### 3.3.4. *Market share*

Selecting a mobile platform with a larger market share (used by more people) allows a greater accessibility to users. To date, Android mobile devices have the strongest end-user demand worldwide, followed by iOS and Windows (IDC, 2015). The similar trend is reflected in the selection of prototype development platforms by researchers, as shown in Fig. 7 (N/A indicates unspecified).

### 3.4. *Degree of control*

The degree of control refers to restrictions or constraints imposed when carrying out an experiment. It covers three different aspects, namely: (i) acquisition device selection, (ii) experimental setting control, and (iii) input string selection. Generally, the number of experiments that imposes restrictions outnumbers those that do not, and this is the case for all three aspects, as shown in Fig. 8.

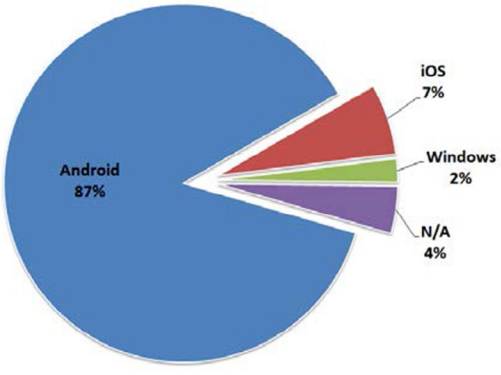


**Fig. 7 – The distribution of underlying research development platforms.**

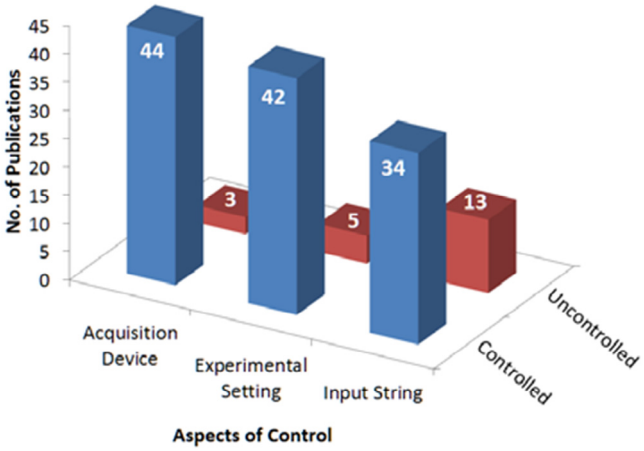


**Fig. 8 – The number of publications on reporting the degree of control of different aspects imposed in experiments.**

#### 3.4.1. *Acquisition device selection*

The devices used in experiments can be selected with one of the two approaches. One is to use a predetermined device and the other is to use a subject (hereafter refers to a mobile device user recruited for an experiment) specific device. Based on our literature survey, a majority of the experiments published in literature (94%), with an exception of the work reported in Alotaibi et al. (2014), Johansen (2012) and Samura et al. (2014), were undertaken by using the first approach. The primary reason for using a predetermined device is to prevent any inconsistencies in the features acquired. For example, the availability of hardware sensors (Alotaibi et al., 2014) and the variations in sensor resolution or sensitivity (Seo et al., 2012) between different devices may cause inconsistencies in the data acquired. Additionally, as subjects are usually more familiar with their own devices, allowing subjects to use their own devices to acquire data may introduce bias in their experimental results. This can be avoided by requiring subjects to use a predetermined device for the entire data acquisition operation (Serwadda et al., 2013).

In contrast, some experiments reported in literature were carried out without any restrictions on the types of device that should be used, so subjects can use their own devices to acquire data. In this way, more subjects from different population groups may be attracted to take part in the experiments. For example, Johansen (2012) allowed subjects to conduct data acquisition using their own mobile devices via a mobile app, so subjects do not need to be physically present or be supervised throughout the entire data acquisition operation. In other words, as this approach removes physical and geographical barriers in conducting experiments, the data acquisition can be conducted on a larger scale and can reach a wider audience. It has also been argued by Samura et al. (2014) that by using a device which the subjects are accustomed to, the experimental results obtained may better reflect their actual usage behavior. This is in line with our observation that experiments that did not

impose any restrictions on the input devices were also operated under uncontrolled environment so that a subject's touch dynamics can be acquired in a natural condition (further discussions in the following section).

#### 3.4.2. *Experimental setting control*

The experimental setting control refers to the degree of restrictions that are imposed on an experiment during a data acquisition operation. Data acquisition may be conducted under two settings: (i) supervision with a stringent protocol or (ii) unmonitored without any restrictions. The papers by Alotaibi et al. (2014), Johansen (2012), Kolly et al. (2012) and Samura et al. (2014) reported experiments that were carried out without any specific restrictions. A majority of experiments (89%) reported in literature were actually carried out under supervised and controlled environments. The primary reason to control an experiment is to reduce the level of variations in touch dynamics patterns which may be caused by external factors such as distractions, cognitive load, and sickness, etc. Controlling an experiment with a stringent protocol can prevent these external factors from inflicting noise (Trojahn et al., 2013), allowing primary experimental variables (i.e. the discriminative capabilities of features or the accuracy performance of classifiers) to be evaluated more accurately (Cai et al., 2013). Also, the unmonitored and uncontrolled experimental setting may risk the data acquired being distorted or tampered with, resulting in the reduction of data quality.

There are also views that touch dynamics data should be acquired in a natural manner without imposing any restrictions, or in an environment that can resemble a real-life touchscreen device usage scenario (Alotaibi et al., 2014). The reason given by the authors was that a subject's interactivity with his/her device may differ in different circumstances, e.g. when performing a task at hand or doing a job in a controlled environment. The authors in Rybnicek et al. (2014) suggested that experimental results obtained under a controlled environment are over optimistic compared to those acquired in uncontrolled environments.

#### 3.4.3. *Input string selection*

How to select the input strings or which input strings should be used during experiments is another factor one should consider when carrying out touch dynamics related experiments. In the majority of the experiments reported in literature (72%), the subjects were asked to provide a predefined set of input strings during a data acquisition operation. In other words, the input strings used in these experiments are identical across all the subjects. This means that the samples acquired from different subjects in the same dataset can be reused for evaluation purposes (not just for template generation), and as a result, a larger number of test samples can be acquired without acquiring them separately. However, in certain situations, this approach may not be practical. For example, the data acquisition experiments conducted by Draffin et al. (2014), Y. Meng et al. (2014) and Meng et al. (2013) were aimed at acquiring all the touch dynamics data over an entire interaction session of a device. Due to the nature and objective of these experiments, predetermining a set of input strings would not be practical.

## 4. Data acquisition

### 4.1. *Subject size*

The subject size used in an experiment is known to have an impact on the obtained results of the experiment (Maxion and Killourhy, 2010; Xu et al., 2014). The larger the subject size used in an experiment, the better the experimental results can signify the scalability of a study (Bartlow and Cukic, 2006) and reflect the true accuracy performance of a biometrics authentication method when deployed in real-world (Jagadeesan and Hsiao, 2009).

Most experiments recruited less than 50 subjects (Buschek et al., 2015; Kambourakis et al., 2014; Y. Meng et al., 2014), with some less than 5 subjects (Nixon et al., 2014; Rao et al., 2014). We were only able to find three experiments (Gascon et al., 2014; Serwadda et al., 2013; Trojahn et al., 2013), which used a large number of subjects (315, 190 and 152 subjects, respectively). The subject sizes used in the experiments published in literature are summarized in Fig. 9.

In most of the experiments (94%), subjects were recruited on a voluntary basis, i.e. without receiving any monetary benefits. Only in a few experiments, subjects were awarded cash cards (Buschek et al., 2015; Xu et al., 2014) or some form of prizes (Johansen, 2012). The awards or prizes were used to motivate the subjects to take part in the experiments, increasing the participation rates. A data acquisition operation could be a resource-intensive process that requires some dedication and efforts from the subjects. To increase the participation rate or the number of subjects taking part in an experiment, Kolly et al. (2012) suggested that the data acquisition tools can be distributed via a mobile application store. This may be a possible way of recruiting a larger number of subjects, but, by doing so, control over the data acquisition operation will certainly become limited and the risk of data being tampered with or manipulated becomes higher.

### 4.2. *Subject demography*

Subjects can be selected based on three variables, namely, by their age, affiliation, and profession. These three variables jointly

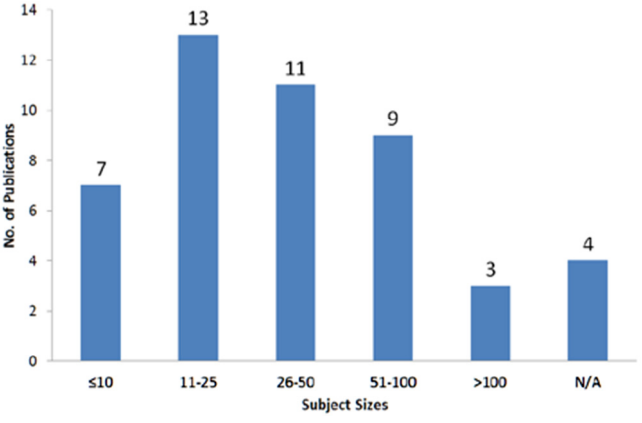


**Fig. 9 – The number of publications on reporting different subject sizes.**

correlate with a subject's device usage frequencies and familiarity. Therefore, if subjects are not selected properly, there may be unintentional biases in the experimental results.

In the published experiments, subjects were often selected from a specific group of population, e.g. from people (i) with a narrow age distribution (i.e. 19–26) (Antal and Szabó, 2014; Buschek et al., 2015; Kambourakis et al., 2014), (ii) within the same organization (i.e. within a research institute) (Alotaibi et al., 2014; Giuffrida et al., 2014; Samura et al., 2014), or (iii) with a specific profession (i.e. students) (Draffin et al., 2014; Y. Meng et al., 2014; Meng et al., 2013). This is because a certain group of population is more readily available and is cheaper for researchers who often have limited resources to conduct their experiments. However, it may be argued that the data acquired from a special group of population may not realistically represent the wider community.

Studies from Coakley et al. (2015), El-Abed et al. (2014) and Kolly et al. (2012) are among the few pieces of work we were able to find in literature, which recruited subjects from a population with wide demography, e.g. people from dissimilar age groups and/or with different professions. Their goal is to diversify the dataset's subject demography so that the obtained results of the experiments can better resemble real-world scenarios.

### 4.3. *Input string type*

The input string type is an important experimental variable in touch dynamics biometrics research, as the feature used for touch dynamics biometrics is originated from a subject's input string. Generally, subjects are required to provide character-based (i.e. alphabetic, special character, or alphanumeric input), digit-based (i.e. only numerical input), and/or other non-specific touch events. The majority of touch dynamics experiments published required a subject to either provide a character-based or a digit-based input as shown in Fig. 10.

A character-based input string can be further categorized into short and long character strings. A short input string is usually consisted of a username or a password (Huang et al., 2012; Mahnoush Babaeizadeh et al., 2014), a random character combination (Antal and Szabó, 2014; Rao et al., 2014), or a set of dictionary words (Buschek et al., 2015; Giuffrida et al., 2014). A long input string is usually a segment of texts (Gascon et al., 2014) or several paragraphs of texts (Feng et al., 2013; Samura et al., 2014). Likewise, a digit-based input string can also be classified into short and long digit strings. A short digit string is typically 4 to 6 digits long. It usually resembles a mobile device unlocking code (Zheng et al., 2014) or an ATM PIN number (Sen and Muralidharan, 2014). The length of a long digit string, on the other hand, usually exceeds 10 digits, similar to, e.g., a social security number (Johansen, 2012) or a phone number (Jain et al., 2014). Other non-specific input types include a random acquisition of some continuous touch events (Draffin et al., 2014; Y. Meng et al., 2014), a random multi-touch gesture input (Sae-Bae et al., 2012), or touch input interactions with a common user interface element (e.g. buttons, checkboxes and sliders) (Kolly et al., 2012; Saravanan et al., 2014), made over a period of time. As most of the published touch dynamics experiments focused on the character-based or digit-based input, hereafter we shall focus our discussions on these two input types.

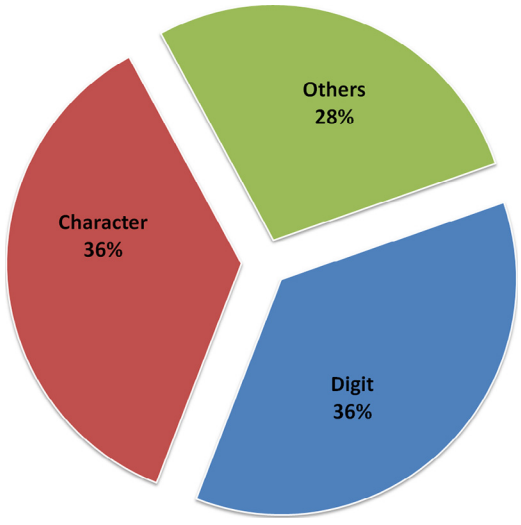


**Fig. 10 – The distribution of input string types used in touch dynamics experiments.**

### 4.4. *Input sample size*

The input sample size of a dataset is known to have an impact on the accuracy, robustness and conclusiveness of the outcome of an experiment (Sen and Muralidharan, 2014; Wolff, 2013). The larger the number of samples we use, the better the representation of a subject's touch dynamics behavior, and, as a result, the higher the accuracy performance we could achieve (Tasia et al., 2014).

There are two ways of acquiring multiple samples in one data acquisition session: (i) acquiring a fixed input string repeatedly (for character-based or digit-based input), or (ii) continuously acquire touch events over a fixed period of time (for non-specific input). According to the literature, the benchmark number of samples per session per subject is between 10 and 20 repetitions for the fixed input type (Buschek et al., 2015; Kambourakis et al., 2014; Sen and Muralidharan, 2014; Trojahn et al., 2013) and 5 to 10 minutes for non-specific input type (Meng et al., 2013; Xu et al., 2014).

Requesting a large number of samples from a subject in one data acquisition session is impractical. This is because subjects may not be available for a long stretch of time, or may feel uncomfortable with a lengthy acquisition session (Tasia et al., 2014). Therefore, selecting an optimal number of samples per data acquisition session is important. Alternatively, instead of acquiring a large number of samples in one lengthy acquisition session, we can carry out the data acquisition operation in multiple shorter sessions spread over a period of time. This approach can reduce the level of discomfort imposed on a subject and also better capture any intra-session variations (further discussions in the next section).

### 4.5. *Acquisition session and interval*

As mentioned above, a data acquisition operation for each subject can be carried out and completed either within a single session or spread over several sessions separated by a predefined time interval. In the majority of the reported experiments (64%), data from each subject were acquired within a single session.

Though the above approach is relatively cheaper, it is not always practical to request a large number of samples from each subject in a single session. More importantly, touch dynamics biometrics, like other behavioral biometrics (i.e. voice, gait and signature), is not stable over time (Xu et al., 2014). This implies that there may be intra-session variations between different input samples in a session, even if the input samples are provided by the same subject. If samples are acquired in a single session, the intra-session variations may not be captured. This is evident in the experimental work reported by Buschek et al. (2015), where the accuracy performance evaluated using data acquired in a single session is two times better than those obtained in different sessions. This means that experimental evaluation using data acquired in single sessions may be over-optimistic, and may not reflect the true accuracy of a biometrics authentication method when deployed in real-world.

Ideally, a data acquisition operation should be divided into multiple sessions separated by some intervals. In this way, intra-session variations can better be captured. This approach has been adopted in several experiments reported in literature, and, in these experiments, the selected intervals separating different data acquisition sessions vary from minutes (Johansen, 2012), days (El-Abed et al., 2014; Serwadda et al., 2013; Xu et al., 2014) to weeks (Buschek et al., 2015; Tasia et al., 2014). It is worth noting that careful considerations should be given when determining the number of samples and the lengths of intervals separating different sessions. This is because these two factors may influence the subject participation rate of an experiment. For example, a data acquisition operation that spans across a longer period of time is more likely to receive a lower subject participation rate (due to subjects' availabilities) or may result in a higher subject dropout rate (due to a greater commitment required of the subjects) (De Luca et al., 2012). Therefore, there is a balance between better capturing subjects' natural touch dynamics patterns variation, and preserving the subject participation rate and commitment to the data acquisition operation.

### 4.6. *Legitimate and illegitimate subject samples*

The degree of accuracy of a touch dynamics authentication system is measured by using the FAR and the FRR values (discussed in Section 2.5). To compute these values, two categories of samples are required, namely, legitimate and illegitimate samples. Acquiring the legitimate samples is a straightforward process that has been described above. However, for acquiring the illegitimate samples from a subject, there are three approaches: (i) partitioning the subject's samples into two subsets, one used as the legitimate samples, and the other used for the illegitimate samples; (ii) acquiring additional samples from a subset of the subjects involved in a dataset, and use these additional samples as the illegitimate samples; or (iii) recruiting additional subjects to provide the illegitimate samples.

Based on our literature research, the first approach is most frequently adopted (85%). For example, in these experiments (Giuffrida et al., 2014; Y. Meng et al., 2014; Trojahn et al., 2013; Wolff, 2013), $t$ out of $i$ samples acquired were used as the illegitimate samples, with the rest, $(t - i)$ samples, used as the legitimate ones. The subset of illegitimate samples can be selected in a randomized (Trojahn et al., 2013; Wolff, 2013) or a predefined order (Giuffrida et al., 2014; Y. Meng et al., 2014). With this approach, an equal number of the legitimate and the illegitimate samples can be obtained with minimal or no additional resources. However, this approach has a limitation when the input string is not the same across all subjects, as it is not possible to compare the touch dynamics patterns of two input strings when they are different.

The experiments that do not use the same input string across all the subjects can use the second approach to obtaining the illegitimate samples. For example, Tasia et al. (2014) recruited 100 subjects to provide the legitimate samples. Then 10 subjects were randomly selected from the 100 subjects and were given the PINs of all the other subjects. Each of these 10 subjects was given the additional task of providing 5 impersonated samples of each of the other 99 subjects. With this approach, the 10 selected subjects need to devote more time and effort to the experiment, and this may discourage voluntary participations. For this reason, the number of the illegitimate samples acquired is usually smaller than that acquired using the first approach (by partitioning the already acquired legitimate samples).

The third approach is to recruit a separate pool of subjects specifically for providing the illegitimate samples. Take Dhage et al. (2015), Gascon et al. (2014), and Sen and Muralidharan (2014) for example, instead of requesting the subjects recruited for providing the legitimate samples, the research team recruited additional subjects for providing the illegitimate samples. In this way, it is more likely to obtain a balanced number of the legitimate and the illegitimate samples. It is worth mentioning that, with this approach, for each increment of the subject size in a dataset, two subjects should be recruited. This means that the resource and the effort needed to obtain a dataset are doubled in comparison with the other two approaches.

### 4.7. *Public dataset*

The availability of a public dataset for touch dynamics biometrics research is vital. For example, with such a dataset, we could do comparisons of different algorithms on the same dataset and/or different experimental settings. The availability of a public dataset also allows researchers to focus on more challenging research issues, spending less time on data acquisitions. However, the availability of open dataset in the domain of touch dynamics biometrics is still limited. This may be due to the fact that touchscreen devices have not been with us for a very long time, and the creation or acquisition of such data is a time and resource consuming process. At the time of this writing, we are only able to find three public datasets in relation to touch dynamics. These datasets are summarized in Table 2.

**Table 2 – The three public touch dynamics biometrics datasets.**

| | Dataset 1 | Dataset 2 | Dataset 3 |
|---|---|---|---|
| Input string type | Character-based | Digit-based | Character-based |
| Subject size | 51 | 100 | 42 |
| Number of sessions | 3 | 5 | 2 |
| Session interval | 3–30 Days | 1 Week | - |
| Input sample size/session | 5 | 2 | 30 |
| Total acquisition duration/subject | - | 5 Weeks | 2 Weeks |
| Total sample size/subject | 15 | 10 | 51 |
| Separate illegitimate samples | No | Yes | No |

#### 4.7.1. *Dataset 1*

The data acquisition experiment conducted by El-Abed et al. (2014) involved 51 subjects inputting a fixed character-based string "rhu.university" on a Nokia Lumia 920 Windows smartphone. The subjects were required to attend 3 different sessions with an average inter-session interval of 5 days. The first session was used as a practice session, so the actual data acquisition commenced from the second session. A total of 15 samples were acquired from each subject over the second and third sessions.

#### 4.7.2. *Dataset 2*

Another dataset, published by Tasia et al. (2014), was based on digit strings. The device used for acquiring the data was an early generation smartphone running on Android 2.0.1 (Éclair) API level 6, which was released in December 2009. Subjects were only required to provide 2 samples per session and 5 sessions were used with an inter-session interval of at least 1 week. Data were acquired in a confined classroom environment involving mostly university students. Input strings were not predefined; subjects were allowed to freely choose their PINs, most of which have the lengths of 4 to 8 digits.

#### 4.7.3. *Dataset 3*

Another related effort to acquire and share dataset publicly was made by Antal and Szabó (2014). The number of subjects involved is the smallest among the three public datasets. The data acquisition in this experiment was done by using a Nexus 7 tablet and an LG Optimus L7 P700 smartphone. The paper did not explain why two different device types were used and whether the use of different device types would have any performance implications. The input string used for the data acquisition was predefined as ".tie5Roanl". Additionally, the touch events acquired include not only the input string but also shift key (toggle between lower and uppercase characters) and keyboard switch key (toggle between characters and digits keys). These secondary key events may capture valuable and distinctive information about a subject touch dynamics patterns. Also, in this experiment, most of the subjects provided their passwords 30 times each on 2 isolated sessions over a period of 2 weeks (the duration between the two sessions was unknown). Some invalid inputs were removed, so the dataset was unified to only 51 input samples per subject (instead of 60 from both sessions).

## 5. Feature extraction

Human touch dynamics patterns contain unique features that can be used to distinguish one another. In feature extraction phase, these features are extracted by processing the raw touch dynamics data acquired from a subject. Common features discussed in literature can be classified into three categories, namely: (i) timing, (ii) spatial, and (iii) motion. The research efforts made on these features are summarized in Fig. 11.

### 5.1. *Timing feature (TM)*

The timing feature is the most widely used feature in touch dynamic biometrics. A touch event (finger touching down or lifting up) on a virtual keyboard generates digital interrupts that can be detected by the mobile OS API function calls (Kambourakis et al., 2014). Each of these events can be coupled with a timestamp value. These timestamp values do not have semantic meaning and need to be further manipulated. Based on these timestamp values, two different types of timing feature with varied lengths can be extracted.

#### 5.1.1. *Timing feature types*

By performing mathematical operations on two touch event timestamp values, two types of timing feature types can be obtained. The first one is the Dwell Time (DT) and it refers to the time duration of a touch event with the same key. It is also known as interval, press or hold time in literature. This value can be obtained by subtracting a key release timestamp value from its key press timestamp value. The second one is the Flight Time (FT). It refers to the time interval between the touch events of two successive keys. It is also known as latency. As shown in Fig. 12, there are four variants of FT. It is worth noting that,

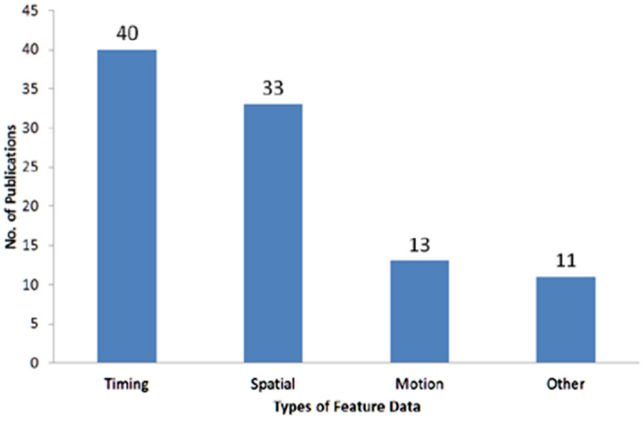


**Fig. 11 – The number of publications on reporting different touch dynamics feature types.**

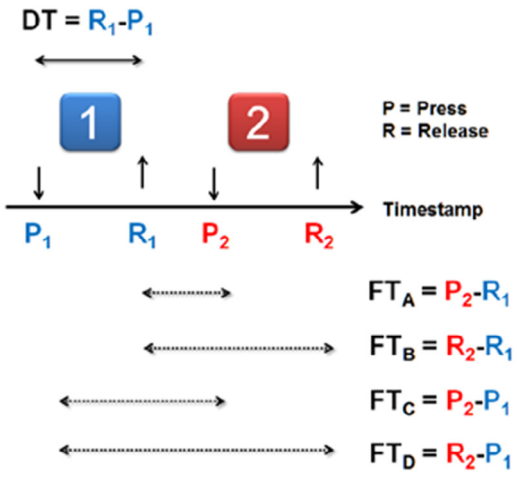


**Fig. 12 – Timing feature types.**

according to Sheng et al. (2005), $FT_A$ may have a negative value. This scenario happens when a subject presses the next key before releasing the previous one. However, this scenario is more likely to happen when acquiring the timing feature using a computer keyboard rather than using a virtual keyboard. This is due to the difference in physical and geometrical size of virtual keys against physical keys; it is very rare for a subject to use multiple fingers simultaneously when providing their input on virtual keys. As a result, the chances of pressing the next key before releasing the previous one is significantly reduced or in some cases do not exist when using a virtual keyboard.

#### 5.1.2. *Timing feature length*

A timing feature can be extracted with different feature lengths. The shortest feature length is known as uni-graph, which is the timing feature extracted by taking the touch event timestamp values of the same key. The timing features extracted from two or more keys are called di-graph and n-graph, respectively. A graphical illustration of the different n-graph lengths is shown in Fig. 13. In the majority of the experiments reported in literature, the uni-graph and di-graph are used. The only two exceptions were the experiments conducted by Giuffrida et al. (2014); Trojahn et al. (2013), where the n-graph with the size of 3 or larger were extracted. The reason why a large n-graph size is not commonly used is that a larger n-graph contains a lower feature granularity (Trojahn et al., 2013). This has been experimentally proven by Giuffrida et al. (2014). In their experiment, the authors compared the accuracy performances of different n-graph sizes. The comparison result showed that a larger n-graph size produces a lower accuracy performance.

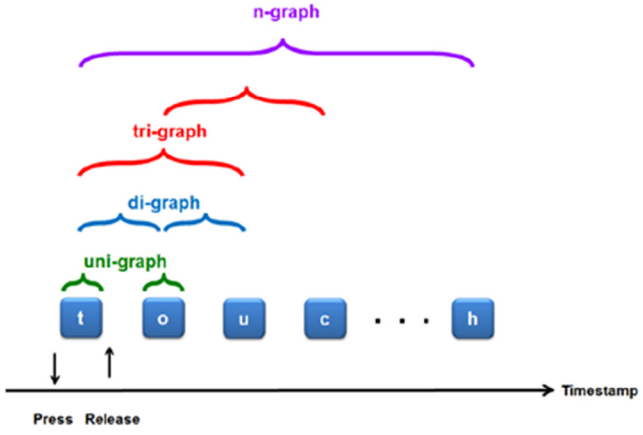


**Fig. 13 – The different timing feature lengths.**

### 5.2. *Spatial feature (SP)*

A spatial feature is a characteristic associated with physical interactions between a fingertip and a device touchscreen surface, and it can be acquired when a touch event is performed. The three most commonly reported spatial features in literature are touch size, pressure, and position. Visual examples of these three spatial features extracted using an Android mobile device are reported in Y. Meng et al. (2014).

#### 5.2.1. *Touch size*

The touch size represents an approximation of the screen area being touched in a touch event. Each touch event is associated with a touch size value. The value is typically returned from an API function and is scaled to a value in the range between 0 and 1 (Zheng et al., 2014). This value is normally used as feature data without further manipulation. The touch size value captured from a subject is determined by the subject's fingertip size. For example, Nixon et al. (2014) observed that an adult male subject usually produces a larger touch size value than a child or an adult female subject. This means that it is hard for people with different fingertip sizes to mimic each other.

#### 5.2.2. *Touch pressure*

The touch pressure is another feature that is often used along with the touch size. A touch pressure value measures the approximated force asserted on the screen upon each touch event. It is expressed in an abstract unit, with a value in the range between 0 (softer touch) and 1 (harder touch) (Zheng et al., 2014). Similar to the case for the touch size, a touch pressure value extracted by an API function can be used directly without further manipulation. A touch pressure value is linked to a subject's finger muscle that is unique to each subject. Therefore, it is hard for one subject to imitate another subject's touch pressure purely by observations, making a touch dynamics authentication system that uses touch pressure feature highly resistant to shoulder surfing attacks (Feng et al., 2013).

#### 5.2.3. *Touch position*

The touch position is a two-dimensional matrix feature that captures a fingertip landing location on a device screen (or key). Each touch event can be associated with an x and y-coordinate measured in pixel units (Kolly et al., 2012). The touch position of a key varies with a subject's fingertip size and cognitive preference. This variation allows the touch position to be used as a discriminative feature to identify a subject. This is further supported by the observations reported by Johansen (2012), where the touch positions provided by different subjects are

highly different among different subjects in their experiment. The touch position can be expressed using two different ways: (i) as the absolute coordinates of a touch event relative to the entire screen (Y. Meng et al., 2014), or (ii) as an offset to the center of a key used (Draffin et al., 2014). Also, by some mathematical manipulations, additional features can be derived. These include the distance (Buschek et al., 2015; Kambourakis et al., 2014), speed (Kambourakis et al., 2014), or angle (Serwadda et al., 2013), between two touch events. However, there is a concern with this coordinate representation of touch position values (Alotaibi et al., 2014); that is, the coordinate system of a screen is device dependent. Using different devices, the captured touch position values are not consistent. Therefore, touch position values should be normalized, unless data acquisition operation is conducted on a device with a similar model (Jain et al., 2014).

### 5.3. *Motion feature (MO)*

Modern mobile devices are embedded with two hardware motion sensors, the accelerometer and the gyroscope. These sensors have been widely used in applications, such as device pairing and sleep cycle monitoring applications, that make use of movement data or are movement dependent (Owusu et al., 2012). Each touch event usually inflicts a small amount of movement and/or rotation to the device. These motion features can be captured and used to identify a subject.

The accelerometer sensor measures the linear movement rate applied to a device over time. It is designed to detect the movement along the x, y, and z-axis in both positive and negative directions. These three values are measured in the unit of $m/s^2$ (Aviv et al., 2012). On the other hand, the gyroscope sensor measures the rotation rate applied to a device against the three axes: (i) tilt forward and backward (pitch), (ii) twist from side to side (roll), and (iii) turn from portrait to landscape (yaw). These values are measured in the unit of *rad/s* (Giuffrida et al., 2014). Fig. 14 shows a graphical representation of the different motions captured by both sensors.

Normally, raw motion data obtained from these two sensors are not readily usable as feature data. This is because each touch event generates more than one movement and rotation values. To make the data usable as feature data, we should apply some statistical computations, such as min, max, mean and variance, on the raw data, and the results of these computations can be used as meaningful feature data (de Mendizabal-Vazquez et al., 2014; Ho, 2013). Also, as Zheng et al. (2014) pointed out, both sensors are sensitive to tiny movement changes. Therefore, they chose to combine sensor values of x, y and z-axis into a vector of feature, instead of using them individually.

Researchers are divided as to whether the accelerometer sensor actually provides a better discriminative property than the gyroscope sensor. For example, the experimental results from Giuffrida et al. (2014) show that the accelerometer data can better capture a subject's touch dynamics patterns than the gyroscope data. However, the observations made by Cai and Chen (2012) show a different result, i.e. the gyroscope data provide a better accuracy, especially if a subject uses the device while moving. In literature, a majority of the touch dynamics motion feature data are from both types of sensors. This is good because data from both types of sensors may complement each other, leading to a better accuracy in identifying a subject.

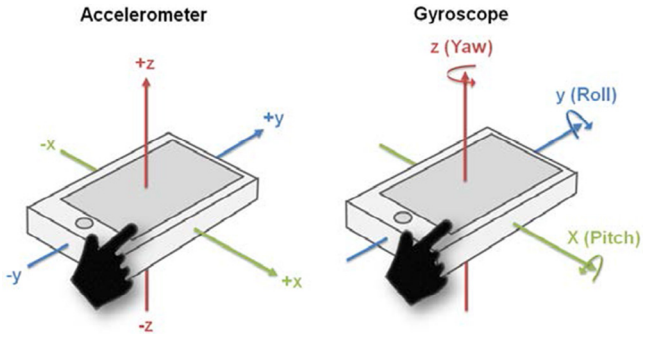


**Fig. 14 – The different motion data captured by different mobile sensors.**

## 6. Decision making

Decision making is an operation carried out to determine if the touch dynamics patterns submitted by a subject have indeed originated from the target subject. This decision is made by comparing the similarity or dissimilarity score generated from a machine learning technique against a predefined threshold. A number of such techniques have been used in a touch dynamics research reported in literature, namely: (i) Probabilistic Modeling, (ii) Cluster Analysis, (iii) Decision Tree, (iv) Support Vector Machine, (v) Neural Network, (vi) Distance Measure, and (vii) Statistical. Fig. 15 summarizes the machine learning techniques against the number of papers that adopted them in touch dynamics research.

### 6.1. *Probabilistic modeling (PM)*

The main idea behind the probabilistic modeling technique is to predict the likelihood of a given test sample belonging to a

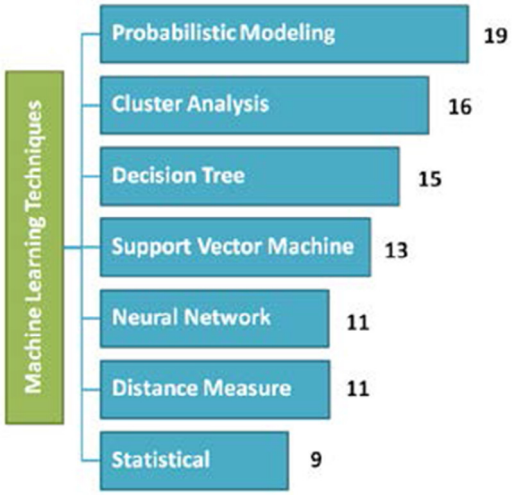


**Fig. 15 – Machine learning techniques vs the number of papers that employed them.**

particular subject using the prior probability calculated from training samples (touch dynamics data acquired during user enrollment phase). One widely used probabilistic modeling technique is the Bayesian Network (Feng et al., 2013; Saravanan et al., 2014). It uses an acyclic graph model to find the probabilistic relationship between parent and child node (Antal and Szabó, 2014). For example, feature data from a reference template will be used as the parent node and the associated subject identity as a child node. Then, given a test sample (touch dynamics data acquired during user authentication phase), the intended child node is determined by the probability of the parent node (Jeanjaitrong and Bhattarakosol, 2013). Other variants of the probabilistic modeling technique include the Naive Bayes (Buschek et al., 2015; Crawford et al., 2013) and the Gaussian Probability Density Function (Mahnoush Babaeizadeh et al., 2014).

### 6.2. *Cluster analysis (CA)*

The cluster analysis technique assumes that samples belonging to the same subject have similar properties (Meng et al., 2013). The goal is to group sample with similar properties to form a homogeneous cluster. Then the label of a test sample is decided by the degree of proximity toward a cluster (Antal and Szabó, 2014). Samples from different clusters are highly dissimilar but very similar among the samples in the same cluster. There are variants of the cluster analysis technique, including the K-means (Trojahn et al., 2013), K-Star (Sen and Muralidharan, 2014; Trojahn and Ortmeier, 2013) and k-Nearest Neighbors (k-NN) (Buschek et al., 2015; Crawford et al., 2013; Giuffrida et al., 2014).

### 6.3. *Decision tree (DT)*

The decision tree technique is popular and used in many areas. It is well known for its low computational complexity (Antal and Szabó, 2014). This technique is particularly suitable for classification problems that involve a small number of output labels. For example, in the case of touch dynamics authentication, it is often used to check whether a test sample is legitimate or not. The J48 (Y. Meng et al., 2014; Saravanan et al., 2014; Sen and Muralidharan, 2014; Trojahn and Ortmeier, 2013) and the Random Forest (RF) (Feng et al., 2013; Kambourakis et al., 2014; Saravanan et al., 2014) are the two widely used decision tree techniques in touch dynamics research. The main objective of these techniques is to create a tree-like model that predicts the class label of a given test sample based on previously known training samples. A decision tree is constructed by continuously splitting feature data into subsets so that the information gain ratio at each node of the tree is maximized. This iterative process stops when a node has only a single label, or when further splitting a tree node no longer provides additional information gain. The RF differs from the J48 in that it adds a randomized procedure in the process of splitting each tree node (Antal and Szabó, 2014). The experimental results reported in Feng et al. (2013) and Serwadda et al. (2013) show that the RF performs better than the J48 in classifying subject touch dynamics patterns. However, it requires a longer time to formulate a decision tree (Kambourakis et al., 2014). When using a decision tree technique, considerations should be given to prevent over-fitting the tree, which could result in a higher level of computational complexity and a lower level of performance.

### 6.4. *Support vector machine (SVM)*

The support vector machine is another technique commonly used in many biometrics studies (Gascon et al., 2014; Jain et al., 2014). The fundamental concept of this technique is to first determine how two classes of feature data differ from each other and then create a boundary that best separate them. Having this boundary, subsequent test samples can be classified as either legitimate or illegitimate according to which side of the boundary they are located. The search for this boundary can be performed within a 2-dimensional hyperplane using a linear kernel (separating) function. However, distinguishing the touch dynamics patterns between legitimate and illegitimate subjects are non-linear in nature (Xu et al., 2014). A non-linear kernel function such as Radial Basis Function (Antal and Szabó, 2014; Serwadda et al., 2013) can be used to map feature data onto a higher dimensional feature space to create more complex boundaries that can optimally split both classes (i.e. legitimate and illegitimate). As a result, it can more accurately determine which side of the feature space a test sample belongs.

### 6.5. *Neural network (NN)*

The neural network technique simulates the information processing structure of biological neurons. Typically, a neural network architecture consists of three interconnected layers (the input, hidden and output layer). To start with, the feature data from all subjects are fed into the input layer of the network as a set of neurons. An activation function is used to assign weights to each neuron. Then the information of the activated neurons is passed from one to another within the hidden layer. This process iterates until an output is produced. Finally, based on the output values, a learning process is used to update the weights of each neuron in the hidden layer to improve the network. Some commonly used neural network techniques are Radial Basis Function networks (RBFN) (Meng et al., 2013; Trojahn and Ortmeier, 2013) and Multi-Layer Perceptron (MLP) (Antal and Szabó, 2014; Sen and Muralidharan, 2014; Serwadda et al., 2013). A neural network generally produces a higher level of accuracy in identifying a subject but is more computationally expensive (Draffin et al., 2014) and more time consuming to be used (de Mendizabal-Vazquez et al., 2014). According to Kambourakis et al. (2014), it is impractical to run on mobile devices with less than 512MB of memory.

### 6.6. *Distance measure (DM)*

The distance measure technique calculates a dissimilarity or similarity score between a test sample and the training sample of a given subject. The score is then compared against a threshold to determine if the test sample belongs to the target subject. Most frequently used distance measure techniques include Euclidean (Crawford et al., 2013; de Mendizabal-Vazquez et al., 2014; Samura et al., 2014; Sen and Muralidharan, 2014), Manhattan (Ho, 2013; Johansen, 2012; Serwadda et al., 2013),

Mahalanobis (Antal and Szabó, 2014; Giuffrida et al., 2014) and Bhattacharyya (Wolff, 2013).

### 6.7. *Statistical (ST)*

There are several statistical techniques that have been used in biometrics research. These techniques include the mean and standard deviation (Tasia et al., 2014; Zheng et al., 2014) and the deviation tolerance (Dhage et al., 2015; Huang et al., 2012). There are a number of advantages associated with these techniques. For example, in comparison with the techniques discussed above, they are less complex and easy to implement, cost less computational time, and consume less resource such as battery power. These advantages are important for resource-limited mobile devices.

## 7. Fusion

Fusion is an approach used to combine information from multiple sources to improve the accuracy performance of a biometrics authentication method. The multiple sources may be from multiple features or by using multiple machine learning techniques. The information from these sources may be combined at three different stages, which are, respectively, referred to as (i) feature level fusion, (ii) score level fusion, and (iii) decision level fusion.

### 7.1. *Feature level fusion (FLF)*

The feature level fusion is the most used fusion approach in touch dynamics research. The fusion approach involves concatenating more than one feature data into a single feature vector and is performed before the template generation or the data classification operation. Fusion may be performed on feature data acquired from the same (Jeanjaitrong and Bhattarakosol, 2013) or different sensors (Giuffrida et al., 2014; Zheng et al., 2014). Although feature level fusion is simple to implement and it enables the utilization of additional properties of multiple feature data, it can result in an overly large joint feature vector known as the curse of dimensionality (Shi et al., 2011). Some machine learning techniques, such as decision tree, may not work well with a high dimensional feature vector (Shimshon et al., 2010). Therefore, the number of feature data fused may influence the selection of machine learning technique.

### 7.2. *Score level fusion (SLF)*

The score level fusion, unlike the feature level fusion, is performed after the data classification operation. For example, in Samura et al. (2014), two different machine learning techniques (the Weighted Euclidean Distance and the Array Disorder) were used independently on one set of feature data, resulting in two matching scores, one from each machine learning technique. The two scores are then combined into a single score for decision making. Methods such as the sum, weighted-sum, or product rules are commonly used to combine multiple scores (Dhage et al., 2015). If the scores from different machine learning techniques are not comparable, they will need to be normalized prior to fusion (Tresadern et al., 2013).

### 7.3. *Decision level fusion (DLF)*

The decision level fusion is the least complex among the three fusion approaches. It requires minimum changes being applied to the internal structure of each data classification algorithm. Fusion is performed by combining decisions (accept or reject) made by multiple machine learning techniques using voting rules, such as the AND or OR rules (Dhage et al., 2015).

## 8. Data adaptation

Human touch dynamics, unlike physiological biometrics (e.g. fingerprint or iris), are not permanent and are likely to evolve over time. After some time, a subject's reference template (generated using samples acquired during enrollment phase) may no longer reflect the subject's most recent touch dynamics patterns. One way to deal with this issue is by introducing an adaptation component. The component uses the most recent touch dynamics patterns to update the reference template of the subject, allowing gradual adjustment of the reference template based on the touch dynamics pattern changes.

To control unintended or unnecessary changes imposed on a reference template, two different policies can be used (Crawford et al., 2013), e.g., selecting samples from different input instances or mixing the most recent samples with a portion of the existing template samples. These policies can reduce the effect of short-term pattern changes of the legitimate subjects or prevent unauthorized modifications made to reference template by the illegitimate subjects.

Although the adaptation component requires additional computation time and resource, if implemented correctly, it may not degrade device performance, reduce battery life span or affect usage experience. For example, an adaptation module can be executed during the period when the execution of the component have the least effect on a device (e.g. when the device is in the standby mode or is plugged into a power source for recharging, or when the processor is idle) (Crawford et al., 2013).

## 9. Performance analysis and discussion

This section provides an overview of the performances achieved by related works, i.e. touch dynamics authentication reported in literature. For the sake of clarity, the performances are discussed based on the authentication modes that they support, static or dynamic modes, input string lengths used, feature discriminative capabilities, fusion approaches used, and the system overheads they each impose.

### 9.1. *Static mode*

In the static mode, the identity of a subject is verified based on the input provided by the subject on the first instance of accessing a system. This is the first line of protection and also

the most commonly seen security protection measure deployed on mobile devices. According to our literature research, character-based and digit-based passcodes are the most used input types in the static mode.

To test the viability of integrating touch dynamics biometrics with password, Kambourakis et al. (2014) has conducted experiments to acquire passwords using touch dynamics from 20 subjects. The password is a predefined alphanumeric character string ("7q56n5ll44"). To acquire each password string, a subject is required to move their finger from one side of the device to the other to complete the input by touching on different keys on the touchscreen keyboard. The feature extraction was performed using two different methods: individual key-based extraction and overall key-based extraction. With the first method, the feature values of every single touched key are extracted and analyzed by the classifier. With the second method, the average feature value of all the touched keys is calculated before being analyzed by the classifier. Experimental results show that regardless of the classifiers used, the first method always outperforms the second. This is because more fine-grained information can be captured in the feature value of individual touched key than the average feature value of all touched keys.

Sen and Muralidharan (2014) attempted to use touch dynamics biometrics to enhance digit-based passcode (PIN). In the experiment, 10 subjects were asked to provide 100 input samples of a predefined PIN ("1593") each. Using the Multi-Layer Perceptron classifier, legitimate subjects were able to be correctly classified up to 86% of the time (FRR 14%). The FAR test was conducted somewhat differently. Two additional subjects were recruited (acting as an attacker) to imitate the PIN input pattern of all the 10 legitimate subjects. To facilitate the impersonation attempt, the two attackers were given a visualization tool. The tool is designed to reveal the correct timing and pressure information of each digit of the PIN input of the legitimate subjects. Even by deliberately exposing the timing and pressure feature data of the PIN, the authors were still able to archive an FAR of 16%.

As can be seen from Table 3, the work by Jeanjaitrong and Bhattarakosol (2013) is the only one that uses a symbolic passcode instead of a character-based or digit-based passcode. The symbolic passcode consisted of spade, heart, diamond, and club. Each of the symbols can be represented in 4 different colors. A total of 16 symbols arranged in a $4 \times 4$ matrix block were designed as the input screen layout. To control the screen size inequality of different mobile devices, the data acquisition tool is developed as a web application so that the input screen layout can automatically scale to the screen size of different mobile devices. An accuracy performance of 82.18% was achieved by employing the Bayesian network to classify the timing and spatial feature of the 10 subjects.

### 9.2. *Dynamic mode*

In the dynamic mode, a subject's identity is continuously verified throughout the active session of a mobile device. An authentication deployed in the dynamics mode can detect unrecognized touch dynamics patterns when a subject's mobile device is used by someone else. Once any unrecognized touch dynamics patterns are detected, restriction to sensitive application can be imposed and/or additional reauthentication request can be triggered. To authenticate a subject in the dynamic mode, a longer input string is normally required than in the static mode. The longer input string is commonly acquired by researchers using three different ways: (i) requiring subject to input a long input string, (ii) accumulating touch events over a predefined period of time, or (iii) setting a predefined number of touch events to accumulate.

Feng et al. (2013) attempted to identify a subject based on the character strings commonly used in emails and chat messages. It is noted that storing feature data of each character combination of the character string may impose system

**Table 3 – The research works conducted in static mode.**

| Study | Subject size | Input type | Input length | Feature | Method | EER (%) |
|---|---|---|---|---|---|---|
| Dhage et al. (2015) | 15 | C | 10 | TM | ST | 0.806 |
| Trojahn and Ortmeier (2013) | 16 | C | 11 | TM, SP | DT | 2.03[b], 2.67[c] |
| Kambourakis et al. (2014) | 20 | C | 10 | TM, SP | DT | 26 |
| Giuffrida et al. (2014) | 20 | C | 8–9 | TM, MO | DM | 0.08 |
| Bond and Ahmed Awad (2015) | 25 | C | 34 | TM | NN | 9.3 |
| Buschek et al. (2015) | 28 | C | 6–8 | TM, SP | PM | 21.02 |
| Huang et al. (2012) | 40 | C | 11 | TM | ST | 7.5 |
| Antal and Szabó (2014) | 42 | C | 10 | TM, SP | DM | 12.9 |
| Sen and Muralidharan (2014) | 10 | D | 4 | TM, SP | NN | 15.2 |
| Jain et al. (2014) | 30 | D | 10 | TM, SP | SVM | 2.8 |
| Ho, (2013) | 55 | D | 4 | TM, SP, MO | SVM | 4.4[b], 5.3[c] |
| de Mendizabal-Vazquez et al. (2014) | 80 | D | 4 | TM, SP, MO | DM | 20 |
| Zheng et al. (2014) | 80 | D | 4 | TM, SP, MO | ST | 3.65 |
| Tasia et al. (2014) | 100 | D | 4–10 | TM, SP | ST | 8.4 |
| Wu and Chen (2015) | 100 | D | 8 | TM, SP, MO | SVM | 0.556 |
| Trojahn et al. (2013) | 152 | D | 17 | TM, SP | CA | 4.19[b], 4.59[c] |
| Jeanjaitrong and Bhattarakosol (2013) | 10 | O | 4 | TM, SP | NN | 82.18[a] |

C: Character; D: Digit; O: Other.
[a] Accuracy.
[b] FAR.
[c] FRR.

overhead on a mobile device. Therefore, only the 40 most frequently used English language character combinations were stored and used for analysis. The subjects' inputs were analyzed through a series of character strings with the lengths of 20, 40 and 60. If the touch dynamics patterns within the character string are unrecognized, the subject is declared as an impersonator and a reauthentication request will be invoked. The result shows that the best accuracy performance achieved was 1% EER by using the random forest classifier based on a 40 character string. The authors also suggested that for authentication in the dynamic mode, maintaining high usability is the highest priority. This means that achieving a low FRR value is important.

The experiment conducted by Gascon et al. (2014) is rather different from the one discussed above, as in the input strings are some predefined 160 characters pangrams (words or sentences containing every letter of the alphabet at least once). One example given was "the quick brown fox jumps over the lazy dog". The reason for using pangrams is to ensure that all characters on the virtual keyboard are used at least once. As shown in Table 4, the number of subjects recruited in the experiment is the largest (315 subjects) among the research works conducted in the dynamic mode. The FRR test was conducted somewhat differently; among all the 315 subjects' samples, only 12 subjects' samples were used to calculate the FRR value, with the remaining 303 subjects' samples used for the FAR test. An accuracy performance of 1% FAR and 18% FRR was obtained by using a support vector machine classifier with the linear separating function.

Data acquisition operation conducted by (Y. Meng et al., 2014) did not use a predefined input string. Data acquired was in the form of touch events generated by these routine activities. For example, subjects were requested to use an Android smartphone to perform their usual activity, such as text messaging and web browsing. Unlike their previous study (Meng et al., 2013), where a 10-minute session was used to acquire touch events, their latest study predefined a fixed number of touch events to be acquired in each data acquisition session. The changes were made because the number of touch events acquired by using the time-based session was unpredictable. The latter approach was able to supply a more consistent and sufficient number of touch events for analysis, and thus improves the effectiveness of the experiment. In the experiment, the performances of five different classifiers were compared. The results show that the radial basis function network achieved the best EER of 2.46%.

### 9.3. *Input string length*

Previous studies suggested that the input string length has a direct relationship with the accuracy performance of a classifier. For example, the experiment conducted by Zheng et al. (2014) compared the accuracy performance between a 4-digit and an 8-digit PIN. The dissimilarity scores between the legitimate and the illegitimate subjects have been calculated and plotted in a frequency distribution graph. The authors discovered a clear gap in the dissimilarity scores graph of the 8-digit PIN, but an overlap for the 4-digit PIN. This shows that the longer the PIN, the better it is at representing a subject touch dynamics patterns, and the higher the accuracy performance of the classifier. A likely explanation is that the 8-digit PIN is two times the length of the 4-digit PIN and has twice as many features as the 4-digit PIN, allowing it to hold more information about a subject's touch dynamics patterns.

The above study was not the only one that reported the correlation between the input length and the accuracy performance of a classifier. Draffin et al. (2014) also discovered that by increasing the number of touch events (from 5 to 15) used for data classification, the accuracy performance can be improved by 27%. Table 5 shows similar observation by other research works.

### 9.4. *Feature discriminative capability*

The timing feature has been the most frequently used feature since the early stage of keystroke dynamics research (predecessor to touch dynamics). More recent mobile devices are embedded with various sensors that are capable of providing additional features that can be used to describe a subject's touch dynamics patterns. It is interesting to compare the discriminative capabilities of these new features (i.e. the spatial feature and the motion feature) against the timing feature.

In an experiment conducted by Buschek et al. (2015), the spatial feature, such as the touch size, pressure, and position were extracted from the touchscreen sensor of a mobile device. The result shows that these spatial features always perform

**Table 4 – The research works conducted in dynamic authentication mode.**

| Study | Subject size | Input length | Input freedom | Method | EER (%) |
|---|---|---|---|---|---|
| Kolly et al. (2012) | 5 | 15 touch events | Yes | PM | 80[a] |
| Wolff (2013) | 6 | 15 min touch events | - | DM | 83[a] |
| Draffin et al. (2014) | 13 | 6000 touch events | Yes | NN | 86[a] |
| Meng et al. (2013) | 20 | 10 min touch events | Yes | NN | 2.92 |
| Xu et al. (2014) | 32 | 5 min touch events | No | SVM | <10 |
| Feng et al. (2013) | 40 | 14–53 Characters | No | DT | 1 |
| Y. Meng et al. (2014) | 50 | 120 touch events | Yes | NN | 2.46 |
| Shen et al. (2015) | 51 | 800 touch events | Yes | SVM | <8 |
| Serwadda et al. (2013) | 190 | 80 touch events | No | PM | ~13.8 |
| Gascon et al. (2014) | 315 | 160 Characters | No | SVM | 1[b], 8[c] |

[a] Accuracy.
[b] FAR.
[c] FRR.

**Table 5 – The accuracy performance of short and long input string lengths.**

| Study | Working mode | Input type | Input length | EER (%) | Improvement (%) |
|---|---|---|---|---|---|
| Kambourakis et al. (2014) | Static | C | 10 | 26 | +47.69 |
| | | | 47 | 13.6 | |
| Samura et al. (2014) | Static | C | 200 | 90.7[a] | +5.51 |
| | | | 300 | 95.7 | |
| Zheng et al. (2014) | Static | D | 4 | 5.98 | +24.75 |
| | | | 8 | 4.5 | |
| Chang et al. (2015) | Static | D | 6 | 23 | +21.74 |
| | | | 10 | 18 | |
| Feng et al. (2013) | Dynamic | C | 20 | 8.93[b], 5.6[c] | +88.8, +46.43 |
| | | | 40 | 1, 3 | |
| Draffin et al. (2014) | Dynamic | O | 5 | 67.7[a] | +27.03 |
| | | | 15 | 86 | |

C: Character; D: Digit; O: Other.
[a]Accuracy.
[b]FAR.
[c]FRR.

better than the timing feature in most of the experimental settings. The best EER was obtained using the touch location and the touch size feature, and it was approximately 14% better than the timing feature.

The motion feature extracted from the motion sensor provides additional movement information that can be used to describe a touch event performed by a subject on a mobile device. The experiment conducted by Giuffrida et al. (2014) reported that the EER of their proposed method was markedly reduced from 4.97% (by using the timing feature) to 0.08% after using the motion feature. This shows that the motion feature provides richer discriminative capability than the timing feature.

Table 6 clearly shows that some of the features perform better than the other. However, this does not mean that the lower performed features are not useful. This is because each feature data type captures a different aspect of the subject's touch dynamics patterns (Zheng et al., 2014). Studies (Buschek et al., 2015; Samura et al., 2014) suggested that by combining multiple features (using fusion approach), the accuracy performance of the classifier can be further improved (further discussions in the following section).

### 9.5. *Fusion*

To increase the overall accuracy performance, researchers have attempted to use different fusion approaches. For example, Zheng et al. (2014) suggested that different feature captures a different aspect of a touch dynamics patterns, and combining them allows a subject's touch dynamics patterns to be more precisely represented. To prove this assumption, the authors conducted four experiments using four features: (i) acceleration (motion), (ii) pressure (spatial), (iii) size (spatial) and (iii) time (timing). When the features are used individually, each obtained an EER of 19%, 12%, 25%, and 21%, respectively, whereas when all four features are used together, the EER has decreased to 4.5%. This shows that using the combination of all features produces a better result than using the features individually.

As has been discussed in the section above, there are three different types of fusion approaches. We can use more than one of the approaches simultaneously to improve the accuracy performance of a classifier. For example, Samura et al., (2014) uses both the feature level fusion and the decision level fusion approaches in their experiment. To start with, they first combined different types of feature extracted from a 300 character input text (feature level fusion). Then the individual score produced by two different data classification techniques (Weighted Euclidean Distance and Array Disorder) are combined (decision level fusion) to collectively make an authentication decision. The accuracy performance was successfully improved from 55% (before fusion) to 95.7% (after fusion).

**Table 6 – The accuracy performance of different feature data types.**

| Study | Input string | EER (%) | | | Better off by (%) |
|---|---|---|---|---|---|
| | | Timing | Spatial | Motion | |
| Samura et al. (2014) | 300 Character | 39[a] | **38** | - | +2.56 |
| Buschek et al. (2015) | 6–8 Character | 21.75 | **18.65** | - | +14.25 |
| Jain et al. (2014) | 10-Digit | 10.5 | **3.5** | - | +66.67 |
| Zheng et al. (2014) | 16-Digit | 16.5 | **11.5** | 15 | +30.03 (Timing), +30.43 (Motion) |
| Wu and Chen (2015) | 8-Digit | 71.3[a] | 69.03 | **98.26** | +37.81 (Timing), +42.34 (Spatial) |
| Giuffrida et al. (2014) | 8 Character | 4.97 | - | **0.08** | +98.39 |

*Note*: Best performed feature highlighted in bold.
[a]Accuracy.

**Table 7 – The accuracy performance before and after applying the fusion approach(es).**

| Study | Input | Fusion | EER (%) | | Improvement (%) |
|---|---|---|---|---|---|
| | | | Before fusion | After fusion | |
| Jeanjaitrong and Bhattarakosol (2013) | Symbol | FLF | 62.64[a] | 82.18 | +31.19 |
| Zheng et al. (2014) | 8-Digit | FLF | 12 | 4.5 | +62.5 |
| Trojahn et al. (2013) | 17-Digit | FLF | 9.28[b], 6.72[c] | 4.19, 4.59 | +54.8, +31.7 |
| Jain et al. (2014) | 10-Digit | FLF | 3.5 | 2.8 | +20 |
| Wu and Chen (2015) | 8-Digit | FLF | 98.26[a] | 99.17 | +0.93 |
| Buschek et al. (2015) | 6–8 Char | FLF | 18.65 | 13.74 | +26.33 |
| Antal and Szabó (2015) | 10 Char | FLF | 6.6 | 3.1 | +53.03 |
| Feng et al. (2013) | 4 Char | FLF | 17.8[b], 60[c] | 10.4, 11.1 | +41.57, +81.5 |
| Samura et al. (2014) | 300 Char | FLF, SLF | 55[a] | 95.7 | +74 |
| Dhage et al. (2015) | 10 Char | SLF, DLF | 4.032 | 0.806 | +80 |

[a]Accuracy.
[b]FAR.
[c]FRR.

Table 7 summarizes the research work in touch dynamics biometrics that uses different fusion approaches. All of the experiments show an improvement in accuracy performance after applying fusion approach.

### 9.6. *System overhead*

The accuracy performance has been the primary evaluation metric for a touch dynamics authentication. Nevertheless, system overhead in terms of computational speed and resource consumption is also an important evaluation criterion. The lower the system overhead of a proposed authentication method, the less the impact it has on the performance of a mobile device (Meng et al., 2013). The trade-off between these two metrics was normally not reported. This is because only a small number of research work (13%) conducted both the data acquisition and the data classification operation on a mobile device. Most of the research work conducted the data classification operation on a desktop computer.

To evaluate the computational speed and the resource consumption of a proposed method, one has to implement both the data acquisition and data classification operation on a mobile device. For example, Kambourakis et al. (2014) conducted their experiment entirely on a Sony Ericsson Xperia smartphone. The Random Forest (RF) and the k-Nearest Neighbor (k-NN) techniques were implemented to identify subjects based on a 47 character input string. The experimental result shows that the k-NN not only achieves a better EER but also consumes relatively the same amount of computational time, memory, and CPU resource than the RF.

The number of features used to represent a subject's touch dynamics patterns can also influence the computational time and the resource consumption of a mobile device. The more the number of features used to represent a subject's touch dynamics patterns, the more the computational time and the resource consumption incurred by a mobile device. Ideally, one would want to minimize the number of features used and yet maintaining a reasonable accuracy performance. This has been shown in an experiment conducted by Y. Meng et al. (2014), where, even when the number of features was reduced from 37 to 8, the accuracy performance was maintained at a reasonable level. By reducing the number of features used, the computational time and the resource consumption have been reduced.

## 10. Open problems and opportunities

### 10.1. *Optimal input length*

In general, the longer the input length, the better the accuracy performance of a touch dynamics authentication system. However, the increase in the length of an input reduces the usability (in the static mode) and the security (in the dynamic mode) of the authentication system in different ways. In the case of the static mode, the longer the input length (or passcode), the harder it is for the subject to remember the longer passcode, thus reducing the usability. In the case of the dynamic mode (continuous authentication), the longer the input length, the more time (to acquire more inputs) is required by the authentication system before making an authentication decision. This potentially allows an illegitimate user to use a device longer (duration) before being detected, thus reducing the security. Present continuous authentication in literature is carried out in the form of a periodic authentication mode, where the users are reauthenticated based on a block of fixed number of inputs (Feng et al., 2013) or a collection of inputs over a fixed period of time (Y. Meng et al., 2014). This method has a limitation; that is, if the system could detect an illegitimate user before the fixed number of inputs is acquired, then there is no reason to allow the illegitimate user to have continued access to the device. A realistic continuous authentication should take each input into consideration immediately when making an authentication decision or make sure that the authentication time is as short as possible. One possible way to address this problem is to use the streaming classifier such as those by Abdulsalam et al. (2011) and (Gaber et al. (2007), by which we may be able to use a lesser amount of inputs to make a quicker authentication decision with an acceptable level of accuracy performance.

### 10.2. *Feature data selection*

In general, the more the number of feature data is used to represent the subject's touch dynamics patterns, the higher the accuracy performance of the classifier, and the more effective the authentication method (Alghamdi and Elrefaei, 2015). However, it is usually undesirable to use all of the available feature data as input for a classifier because of three reasons: (i) using more feature data to train and test a classifier will result in more processing time and CPU overhead (W. Meng et al., 2014), (ii) storing more feature data will be costlier in terms of memory efficiency (Zheng et al., 2014), and (iii) not all of the feature data offer the same degree of discriminative capability; some of the feature data provide higher degree of discriminative capability, and so is more important than the other. To address this problem, a feature selection process can be used to choose a smaller subset of feature data that contains only the most important feature data. The feature selection process can be implemented using algorithms such as those by Ververidis and Kotropoulos (2009) and Peng et al. (2005). The discriminative capability of the selected feature data can be evaluated by its intra-class and inter-class variations. This can be quantitatively assessed by using the intra-class and inter-class distribution to show their degree of separations (discriminative capabilities).

### 10.3. *Unary classifier vs. binary classifier*

The primary input for a data classification operation (to build a data classification model) for a touch dynamics authentication system is the user's touch dynamics pattern samples. The samples are often associated with the user class label (i.e. legitimate or illegitimate class). Therefore, the data classification algorithms can be categorized into two types based on the class label of the samples they use to build a model, namely: (i) unary classifier, which uses only the samples from a single class (i.e. only the legitimate user samples); and (ii) binary classifier, which uses the samples from all the classes (i.e. both the legitimate and the illegitimate user samples), to build a model.

The binary classifiers have been the norm for building data classification models (Bellinger et al., 2012). However, in some practical cases, the samples of a particular class outnumbered the others or only the samples from a single class are available. In such imbalanced cases, the binary classifiers may not perform well, as they rely on samples from all classes to build a model that separates the different classes apart (Bellinger et al., 2012). This is the case for a touch dynamics authentication on a mobile device, where a mobile device is a highly personal device (rarely shared between multiple users), so obtaining the illegitimate user samples is not easy in practice. In this case, using the unary classifiers to build a model may be a better option. A unary classifier uses only a single class to build a model, so the model creation is not affected by the imbalanced dataset. Additionally, the time taken to build the model is shorter, as a lesser number of samples are used to train the classifier. Algorithms such as the one-class support vector machine (Manevitz and Yousef, 2002) and the density and probability estimates (Hempstalk et al., 2008) can be used to implement a unary classification model. The effectiveness of using the unary classifiers over the binary classifiers for touch dynamics authentication is still not clearly evaluated and is open for future investigation.

### 10.4. *Larger form factor*

A subject's touch dynamics patterns may vary when providing their input on the mobile device of different sizes. This variation can be exploited to achieve a better accuracy performance. For example, Saravanan et al. (2014) shows that the accuracy performance is higher when the inputs are acquired on a 7-inch mobile device than when they are acquired on a 4.7-inch version. The better performance can be attributed to a higher variation in touch dynamics patterns on the device with a larger screen size. If this assumption is true, and given that the average size of the mobile devices in the market is getting bigger year after year (Ben Taylor, 2014), then the viability of implementing a touch dynamics authentication system on future mobile devices is promising. The majority of the devices used in touch dynamics biometrics literature are smartphones (approximately 5-inch or less). To understand the impact of the different device sizes on the accuracy performance, devices with larger screen size, such as a 10-inch digital tablet or beyond (table-top or touchscreen panel) should be used in future studies.

### 10.5. *Continuous identification*

Although there is an increasing number of literature about the use of touch dynamics biometrics for continuous authentication, there are still limited works on the study of touch dynamics biometrics in the context of continuous identification on mobile devices. In some application scenario, e.g., a mobile device sharing scenario, the continuous identification may be a better choice than the continuous authentication. This can be explained using the case as follows. When a mobile device is shared with a guest, often, the identity of the guest is not known or predetermined. So establishing a credential (e.g. an ID and a PIN for verification) for the unknown guest could be troublesome and impractical. Without an established credential for the guest, revalidating the guest's identity could be difficult to perform in the continuous authentication. In such a scenario, the continuous identification (one-to-many matching) could be a better option.

A continuous identification can be implemented on a mobile device in such a way that when a guest is observed, a privacy protection mechanism (PPM) will be activated automatically. The PPM protects and prevents unauthorized or unintentional access or modification to the owner's data, files, and system settings. After the guest leaves and the device returns to the owner's possession, the PPM will be deactivated, returning the full access privilege of the device back to the owner. If an impersonator is detected, the mobile device will be locked and/or alert the owner through e-mail.

Implementing an effective and robust continuous identification in a mobile device is not easy. Unlike the other biometrics identification systems (e.g. forensic investigations or intrusion detections) that involve searching for a matched identity over a large number of subjects/classes, the identification task in a mobile device is usually limited to three classes,

i.e. the owner, guest, and impersonator. Despite the much smaller number of classes involved, which may seem to have reduced the complexity of the identification task, two problems still exist: (i) the touch dynamics data for the owner need to be incrementally accumulated, and (ii) the touch dynamics data for the guest and the impersonator are often very limited or even unavailable and so the owner/guest/impersonator data classification operation is rather difficult. This makes training the classifier to perform continuous identification a challenging problem. Therefore, how to devise a proper, robust, and effective continuous identification module deserves more attention in future work.

### 10.6. *Cross-session performance evaluation*

In contrast with other physiological biometrics, which produces a more stable pattern across different acquisition sessions (e.g. the iris, face or fingerprint pattern), the touch dynamics patterns acquired from different sessions have a certain degree of variations from one to the other. These variations occur because the touch dynamics patterns can be affected by behavioral variability. The source of behavioral variability comes from several factors such as the (i) cognitive factor (e.g. increasing familiarity with the operations or input methods of a device), (ii) psychological factor (e.g. tiredness, anger, or distressed), (iii) physiological factor (e.g. sickness or injury), or (iv) environmental factor (e.g. distraction or position when using the device). It is imperative to find out whether touch dynamics patterns variabilities between different sessions may have any implications to the accuracy performance of a touch dynamics authentication system. To conduct such a study, the data should be acquired from several sessions spread over a longer period of time. For example, after the first data acquisition session, subsequent sessions should be conducted at least 1 week apart from each of the other sessions. This will increase the chances that the acquired dataset captures the touch dynamics patterns' variabilities. The findings or insights from this kind of study could facilitate another stream of related work on touch dynamics authentication, called the *data adaptation*, which employ effective relearning framework (Crawford et al., 2013) or online incremental learning algorithms (He et al., 2011) to continuously learn the new incoming touch dynamics pattern samples to increasingly enhance the accuracy performance, validity, and flexibility of the authentication model.

### 10.7. *Realistic and open dataset*

Often, the subject size recruited are small in numbers (e.g. less than 30), confined to a specific age group (e.g. 19–26), or people from the nearby population (e.g. within a research institute). The dataset acquired under such conditions are often considered as the convenience samples, as it is relatively easier to acquire the data from the nearby population (Kenneth N. Ross, 1978). Sometimes, the samples analyzed or the conclusions drawn from the convenience samples could not realistically generalize to a wider population. This is more so the case when the convenience samples consist of people who are familiar with or frequently use mobile devices. To ensure that the conclusions drawn from a dataset generalized to a wider population, the best practice is to recruit a large group of subjects (e.g. at least 100 people) and the subjects have to be selected from diversified age groups or with different levels of device familiarity/usage frequency.

Besides, the availability of public datasets is still very limited in this field. Without properly documented and realistic datasets, cross-comparison between different methodologies employed may not be conclusive. Sharing datasets or data acquisition tools is a highly recommended practice as it comes with a three-fold benefit. Firstly, by using a common dataset a more conclusive cross-comparison between different methodologies can be made. Secondly, the availability of open datasets facilitates researchers whom may not have the resource to develop a proper dataset for the experiment. Thirdly, using shared data acquisition tool, other researchers from different institutes can expand a dataset with subjects across different geographical locations.

## 11. Conclusions

Touch dynamics biometrics have promising potentials to strengthen the security of mobile devices or on-line services accessible via mobile devices without additional hardware requirements. The availability of various sensors in recent mobile devices provides added opportunities for such potentials to be explored. This paper gives a comprehensive review of the research work or efforts made on touch dynamics biometrics on mobile devices. The paper first provides an overview, outlines the primary operational process and defines a set of criteria for the evaluation, of a touch dynamics authentication system. Then it presents detailed implementations, experimental settings, and approaches of each of the process, namely, data acquisition, feature extraction, and decision making. Next, the performances reported in published work have been discussed. Finally, it discusses open issues in the topic area and recommends areas for further research. The review presented in this paper may provide a roadmap and stimulate further research in this area.

## Acknowledgments

This research work was supported by the University of Manchester. We would also like to thank the anonymous reviewers, whose advice and comments substantially improved this article.

## REFERENCES

Abdulsalam H, Skillicorn DB, Martin P. Classification using streaming random forests. IEEE Trans Knowl Data Eng 2011;23:22–36. doi:10.1109/TKDE.2010.36.

Alghamdi SJ, Elrefaei LA. 2015. Dynamic user verification using touch keystroke based on medians vector proximity. Presented at the 2015 7th International Conference on

Computational Intelligence, Communication Systems and Networks (CICSyN), pp. 121–6. doi:10.1109/CICSyN.2015.31.
Alotaibi N, Bruno EP, Coakley M, Gazarov A, Monaco V, Winard S, et al., 2014. Text input biometric system design for handheld devices. Proceedings of Student-Faculty Research Day. pp. B7.1–8.
Antal M, Szabó LZ. 2014. Keystroke dynamics on Android platform. Proceedings of the 8th International Conference Interdisciplinarity in Engineering, INTER-ENG 2014, Romania, pp. 131–6.
Antal M, Szabó Z-L. 2015. An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices. doi:10.1109/CSCS.2015.16.
Aviv AJ, Sapp B, Blaze M, Smith JM. 2012. Practicality of accelerometer side channels on smartphones. Proceedings of the 28th Annual Computer Security Applications Conference, ACSAC '12. ACM, New York, NY, USA, pp. 41–50. doi:10.1145/2420950.2420957.
Babaeizadeh M, Bakhtiari M, Maarof MA. Authentication method through keystrokes measurement of mobile users in cloud environment. Int J Adv Soft Comput Appl 2014;6:94–112.
Bartlow N, Cukic B. 2006. Evaluating the reliability of credential hardening through keystroke dynamics. 17th International Symposium on Software Reliability Engineering, 2006, pp. 117–26. doi:10.1109/ISSRE.2006.25.
Bellinger C, Sharma S, Japkowicz N. 2012. One-class versus binary classification: which and When? Presented at the 2012 11th International Conference on Machine Learning and Applications (ICMLA), pp. 102–6. doi:10.1109/ICMLA.2012.212.
Bo C, Zhang L, Jung T, Han J, Li X-Y, Wang Y. 2014. Continuous user identification via touch and movement behavioral biometrics. IEEE International Performance Computing and Communications Conference (IPCCC), 2014, pp. 1–8. doi:10.1109/PCCC.2014.7017067.
Bond WF, Ahmed Awad EA. 2015. Touch-based static authentication using a virtual grid. Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec '15. ACM, New York, NY, USA, pp. 129–34. doi:10.1145/2756601.2756602.
Bryan WL, Harter N. Studies in the physiology and psychology of the telegraphic language. Psychol Rev 1897;4:27–53. doi:10.1037/h0073806.
Buschek D, De Luca A, Alt F. 2015. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15.
Cai L, Chen H. On the practicality of motion based keystroke inference attack. In: Katzenbeisser S, Weippl E, Camp LJ, Volkamer M, Reiter M, Zhang X, editors. Trust and trustworthy computing, lecture notes in computer science. Springer Berlin Heidelberg; 2012. p. 273–90.
Cai Z, Shen C, Wang M, Song Y, Wang J. Mobile authentication through touch-behavior features. In: Sun Z, Shan S, Yang G, Zhou J, Wang Y, Yin Y, editors. Biometric recognition, lecture notes in computer science. Springer International Publishing; 2013. p. 386–93.
Campisi P, Maiorana E, Lo Bosco M, Neri A. User authentication using keystroke dynamics for cellular phones. IET Signal Process 2009;3:333–41. doi:10.1049/iet-spr.2008.0171.
Chang T-Y, Tsai C-J, Tsai W-J, Peng C-C, Wu H-S. A changeable personal identification number-based keystroke dynamics authentication system on smart phones. Secur Commun Netw 2015;doi:10.1002/sec.1265.
Clarke NL, Furnell SM. Authenticating mobile phone users using keystroke analysis. Int J Inf Secur 2007;6:1–14. doi:10.1007/s10207-006-0006-6.
Coakley MJ, Monaco JV, Tappert CC. 2015. Numeric-passcode keystroke biometric studies on smartphones. Presented at the Proceedings of Student-Faculty Research Day, Pace University, pp. B4.1–6.
Crawford H. 2010. Keystroke dynamics: characteristics and opportunities. Eighth Annual International Conference on Privacy Security and Trust (PST), 2010, pp. 205–12.
Crawford H, Renaud K, Storer T. A framework for continuous, transparent mobile device authentication. Comput Secur 2013;39(Part B):127–36. doi:10.1016/j.cose.2013.05.005.
de Mendizabal-Vazquez I, de Santos-Sierra D, Guerra-Casanova J, Sanchez-Avila C. 2014. Supervised classification methods applied to keystroke dynamics through mobile devices. Presented at the 2014 International Carnahan Conference on Security Technology (ICCST), pp. 1–6. doi:10.1109/CCST.2014.6987033.
De Luca A, Hang A, Brudy F, Lindner C, Hussmann H. 2012. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12. ACM, New York, NY, USA, pp. 987–96. doi:10.1145/2207676.2208544.
De Luca A, Hang A, von Zezschwitz E, Hussmann H. 2015. I Feel like i'm taking selfies all day!: towards understanding biometric authentication on smartphones. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15. ACM, New York, NY, USA, pp. 1411–14. doi:10.1145/2702123.2702141.
Dhage S, Kundra P, Kanchan A, Kap P. 2015. Mobile authentication using keystroke dynamics. Presented at the 2015 International Conference on Communication, Information Computing Technology (ICCICT), pp. 1–5. doi:10.1109/ICCICT.2015.7045746.
Draffin B, Zhu J, Zhang J. KeySens: passive user authentication through micro-behavior modeling of soft keyboard interaction. In: Memmi G, Blanke U, editors. Mobile computing, applications, and services, lecture notes of the institute for computer sciences, social informatics and telecommunications engineering. Springer International Publishing; 2014. p. 184–201.
emarketer. 2 billion consumers worldwide to get smart(phones) by 2016 - eMarketer [WWW Document], <http://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694>; 2014 [accessed 03.23.15].
El-Abed M, Dafer M, El Khayat R. 2014. RHU keystroke: a mobile-based benchmark for keystroke dynamics systems. Presented at the 2014 International Carnahan Conference on Security Technology (ICCST), pp. 1–4. doi:10.1109/CCST.2014.6986984.
Feng T, Zhao X, Carbunar B, Shi W. 2013. Continuous mobile authentication using virtual key typing biometric. Presented at the 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 1547–52. doi:10.1109/TrustCom.2013.272.
Frank M, Biedert R, Ma E, Martinovic I, Song D. Touchalytics: on the applicability of touchscreen input as a behavioral biometric for continuous authentication. IEEE Trans Inf Forensics Secur 2013;8:136–48. doi:10.1109/TIFS.2012.2225048.
Gaber MM, Zaslavsky A, Krishnaswamy S. A survey of classification methods in data streams. In: Aggarwal CC, editor. Data streams, advances in database systems. Springer US; 2007. p. 39–59.
Gaines RS, Lisowski W, Press SJ, Shapiro N. Authentication by keystroke timing: some preliminary results (No. R-2526-NSF). Santa Monica, CA: Rand Corporation; 1980.
Gascon H, Uellenbeck S, Wolf C, Rieck K. 2014. Continuous authentication on mobile devices by analysis of

typing motion behavior. Lecture Notes in Informatics. pp. 1–12.
Giuffrida C, Majdanik K, Conti M, Bos H. I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics. In: Dietrich S, editor. Detection of intrusions and malware, and vulnerability assessment, lecture notes in computer science. Springer International Publishing; 2014. p. 92–111.
Harbach M, Zezschwitz E, von Fichtner A, Luca AD, Smith M. 2014. It's a hard lock life: a field study of smartphone (un)locking behavior and risk perception. Presented at the Symposium On Usable Privacy and Security (SOUPS 2014), pp. 213–30.
He H, Chen S, Li K, Xu X. Incremental learning from stream data. IEEE Trans Neural Netw 2011;22:1901–14. doi:10.1109/TNN.2011.2171713.
Hempstalk K, Frank E, Witten IH. One-class classification by combining density and class probability estimation. In: Daelemans W, Goethals B, Morik K, editors. Machine learning and knowledge discovery in databases, lecture notes in computer science. Springer Berlin Heidelberg; 2008. p. 505–19.
Ho G. TapDynamics: strengthening user authentication on mobile phones with keystroke dynamics. Stanford University; 2013.
Huang X, Lund G, Sapeluk A. 2012. Development of a typing behaviour recognition mechanism on Android. Presented at the 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 1342–7. doi:10.1109/TrustCom.2012.127.
IDC. Android and iOS squeeze the competition, swelling to 96.3% of the smartphone operating system market for both 4Q14 and CY14, according to IDC [WWW Document], <http://www.idc.com/getdoc.jsp?containerId=prUS25450615>; 2015 [accessed 04.10.15].
Jagadeesan H, Hsiao MS. 2009. A novel approach to design of user re-authentication systems. Proceedings of the 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems, BTAS '09. IEEE Press, Piscataway, NJ, USA, pp. 379–84.
Jain AK, Flynn P, Ross AA, editors. Handbook of biometrics. Boston, MA: Springer US; 2008.
Jain L, Monaco JV, Coakley MJ, Tappert CC. Passcode keystroke biometric performance on smartphone touchscreens is superior to that on hardware keyboards. Int J Res Comput Appl Inf Technol 2014;2:29–33.
Jeanjaitrong N, Bhattarakosol P. 2013. Feasibility study on authentication based keystroke dynamic over touch-screen devices. Presented at the 2013 13th International Symposium on Communications and Information Technologies (ISCIT), pp. 238–42. doi:10.1109/ISCIT.2013.6645856.
Johansen UA. Keystroke dynamics on a device with touch screen. Gjøvik University Colleg; 2012.
Kambourakis G, Damopoulos D, Papamartzivanos D, Pavlidakis E. Introducing touchstroke: keystroke-based authentication system for smartphones. Secur Commun Netw 2014;doi:10.1002/sec.1061.
Karnan M, Akila M, Krishnaraj N. Biometric personal authentication using keystroke dynamics: a review. Appl Soft Comput 2011;11:1565–73. doi:10.1016/j.asoc.2010.08.003.
Khan H, Atwater A, Hengartner U. 2014. Itus: an implicit authentication framework for Android. Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, MobiCom '14. ACM, New York, NY, USA, pp. 507–18. doi:10.1145/2639108.2639141.
Kim I. Keypad against brute force attacks on smartphones. IET Inf Secur 2012;6:71–6. doi:10.1049/iet-ifs.2010.0212.
Kolly SM, Wattenhofer R, Welten S. 2012. A personal touch: recognizing users based on touch screen behavior. Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones, PhoneSense '12. ACM, New York, NY, USA, pp. 1:1–5. doi:10.1145/2389148.2389149.
Manevitz LM, Yousef M. One-class SVMs for document classification. J Mach Learn Res 2002;2:139–54.
Maxion RA, Killourhy KS. 2010. Keystroke biometrics with number-pad input. IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2010, pp. 201–10.
McLoughlin IV, Naidu N. 2009. Keypress biometrics for user validation in mobile consumer devices. IEEE 13th International Symposium on Consumer Electronics, 2009, pp. 280–4.
Meng W, Wong DS, Kwok L-F. The effect of adaptive mechanism on behavioural biometric based mobile phone authentication. Inf Manag Comput Secur 2014;22:155–66. doi:10.1108/IMCS-09-2013-0062.
Meng Y, Wong DS, Schlegel R, Kwok L. Touch gestures based biometric authentication scheme for touchscreen mobile phones. In: Kutyłowski M, Yung M, editors. Information security and cryptology, lecture notes in computer science. Springer Berlin Heidelberg; 2013. p. 331–50.
Meng Y, Wong DS, Kwok L-F. 2014. Design of touch dynamics based user authentication with an adaptive mechanism on mobile phones. Proceedings of the 29th Annual ACM Symposium on Applied Computing, SAC '14. ACM, New York, NY, USA, pp. 1680–7. doi:10.1145/2554850.2554931.
Niu Y, Chen H. Gesture authentication with touch input for mobile devices. In: Prasad R, Farkas K, Schmidt AU, Lioy A, Russello G, Luccio FL, editors. Security and privacy in mobile information and communication systems, lecture notes of the institute for computer sciences, social informatics and telecommunications engineering. Springer Berlin Heidelberg; 2012. p. 13–24.
Nixon KW, Chen Y, Mao Z-H, Li K. User classification and authentication for mobile device based on gesture recognition. In: Pino RE, editor. Network science and cybersecurity, advances in information security. Springer New York; 2014. p. 125–35.
Obaidat MS, Sadoun B. Keystroke dynamics based authentication. In: Jain AK, Bolle R, Pankanti S, editors. Biometrics. Springer US; 1996. p. 213–29.
Owusu E, Han J, Das S, Perrig A, Zhang J. 2012. ACCessory: password inference using accelerometers on smartphones. Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications, HotMobile '12. ACM, New York, NY, USA, pp. 9:1–6. doi:10.1145/2162081.2162095.
Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27:1226–38. doi:10.1109/TPAMI.2005.159.
Perrucci GP, Fitzek FHP, Sasso G, Kellerer W, Widmer J. 2009. On the impact of 2G and 3G network usage for mobile phones' battery life. Presented at the European Wireless Conference, 2009, pp. 255–9. doi:10.1109/EW.2009.5357972.
Raghunathan A, Ravi S, Hattangady S, Quisquater J-J. 2003. Securing mobile appliances: new challenges for the system designer. Design, Automation and Test in Europe Conference and Exhibition, 2003. pp. 176–81.
Rao KR, Anne VPK, Chand US, Alakananda V, Rachana KN. Inclination and pressure based authentication for touch devices. In: Satapathy SC, Avadhani PS, Udgata SK, Lakshminarayana S, editors. ICT and critical infrastructure: proceedings of the 48th annual convention of computer society of india- vol i, advances in intelligent systems and computing. Springer International Publishing; 2014. p. 781–8.

Ross KN. Sample design for educational survey research, evaluation in education. Pergamon Press; 1978.
Rybnicek M, Lang-Muhr C, Haslinger D. 2014. A roadmap to continuous biometric authentication on mobile devices. Presented at the 2014 International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 122–7. doi:10.1109/IWCMC.2014.6906343.
Sae-Bae N, Ahmed K, Isbister K, Memon N. 2012. Biometric-rich gestures: a novel approach to authentication on multi-touch devices. Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI '12. ACM, New York, NY, USA, pp. 977–86. doi:10.1145/2207676.2208543.
Samura T, Izumi M, Nishimura H. 2014. Flick input authentication in Japanese free text entry on smartphones. Presented at the Proceedings of the SICE Annual Conference (SICE), 2014, pp. 1348–53. doi:10.1109/SICE.2014.6935267.
Saravanan P, Clarke S, Chau DHP, Zha H. 2014. LatentGesture: active user authentication through background touch analysis. Proceedings of the Second International Symposium of Chinese CHI, Chinese CHI '14. ACM, New York, NY, USA, pp. 110–13. doi:10.1145/2592235.2592252.
Schlöglhofer R, Sametinger J. 2012. Secure and usable authentication on mobile devices. Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia, MoMM '12. ACM, New York, NY, USA, pp. 257–62. doi:10.1145/2428955.2429004.
Sen S, Muralidharan K. 2014. Putting "pressure" on mobile authentication. Presented at the 2014 Seventh International Conference on Mobile Computing and Ubiquitous Networking (ICMU), pp. 56–61. doi:10.1109/ICMU.2014.6799058.
Seo H, Kim E, Kang H. A novel biometric identification based on a user's input pattern analysis for intelligent mobile devices. Int. J Adv Robot Syst 2012;9:1. doi:10.5772/51319.
Serwadda A, Phoha VV, Wang Z. 2013. Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms. Presented at the 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–8. doi:10.1109/BTAS.2013.6712758.
Shen C, Zhang Y, Cai Z, Yu T, Guan X. 2015. Touch-interaction behavior for continuous user authentication on smartphones. Presented at the 2015 International Conference on Biometrics (ICB), pp. 157–62. doi:10.1109/ICB.2015.7139046.
Shen C, Zhang Y, Guan X, Maxion RA. Performance analysis of touch-interaction behavior for active smartphone authentication. IEEE Trans Inf Forensics Secur 2016;11:498–513. doi:10.1109/TIFS.2015.2503258.
Sheng Y, Phoha VV, Rovnyak SM. A parallel decision tree-based method for user authentication-based on keystroke patterns. IEEE Trans Syst Man Cybern B Cybern 2005;35:826–33. doi:10.1109/tsmcb.2005.846648.
Shi W, Yang J, Jiang Y, Yang F, Xiong Y. 2011. SenGuard: passive user identification on smartphones using multiple sensors. Presented at the 2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 141–8. doi:10.1109/WiMOB.2011.6085412.
Shimshon T, Moskovitch R, Rokach L, Elovici Y. 2010. Clustering di-graphs for continuously verifying users according to their typing patterns. IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI). pp. 445–9.
Sieger H, Kirschnick N, Möller S. 2010. Poster: user preferences for biometric authentication methods and graded security on mobile phones. Presented at the Symposium on Usability, Privacy, and Security (SOUPS 2010).
Tasia C-J, Chang T-Y, Cheng P-C, Lin J-H. Two novel biometric features in keystroke dynamics authentication systems for touch screen devices. Secur Commun Netw 2014;7:750–8. doi:10.1002/sec.776.
Taylor B. Why smartphone screens are getting bigger: specs reveal a surprising story [WWW Document], <http://www.pcworld.com/article/2455169/why-smartphone-screens-are-getting-bigger-specs-reveal-a-surprising-story.html>; 2014 [accessed 02.16.16].
Taylor B. 5 ways the smartphone is conquering the tablet [WWW Document], <http://www.pcworld.com/article/2889275/5-ways-the-smartphone-is-conquering-the-tablet.html>; 2015 [accessed 04.07.15].
Teh PS, Teoh ABJ, Yue S. A survey of keystroke dynamics biometrics. Sci World J 2013;2013:e408280. doi:10.1155/2013/408280.
Tresadern P, Cootes TF, Poh N, Matejka P, Hadid A, Levy C, et al. Mobile biometrics: combined face and voice verification for a mobile platform. IEEE Pervasive Comput 2013;12:79–87. doi:10.1109/MPRV.2012.54.
Trojahn M, Ortmeier F. 2013. Toward mobile authentication with keystroke dynamics on mobile phones and tablets. Presented at the 2013 27th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 697–702. doi:10.1109/WAINA.2013.36.
Trojahn M, Arndt F, Ortmeier F. 2013. Authentication with keystroke dynamics on touchscreen keypads – effect of different N-graph combinations. Presented at the MOBILITY 2013, The Third International Conference on Mobile Services, Resources, and Users, pp. 114–19.
Ververidis D, Kotropoulos C. Information loss of the Mahalanobis distance in high dimensions: application to feature selection. IEEE Trans Pattern Anal Mach Intell 2009;31:2275–81. doi:10.1109/TPAMI.2009.84.
Wolff M. Behavioral biometric identification on mobile devices. In: Schmorrow DD, Fidopiastis CM, editors. Foundations of augmented cognition, lecture notes in computer science. Springer Berlin Heidelberg; 2013. p. 783–91.
Wu J, Chen Z. An implicit identity authentication system considering changes of gesture based on keystroke behaviors. Int J Distrib Sens Netw 2015;2015:e470274. doi:10.1155/2015/470274.
Xu H, Zhou Y, Lyu MR. 2014. Towards continuous and passive authentication via touch biometrics: an experimental study on smartphones. Symposium on Usable Privacy and Security (SOUPS 2014). USENIX Association, Menlo Park, CA, pp. 187–98.
Zakaria NH, Griffiths D, Brostoff S, Yan J. 2011. Shoulder surfing defence for recall-based graphical passwords. Proceedings of the Seventh Symposium on Usable Privacy and Security, SOUPS '11. ACM, New York, NY, USA, pp. 6:1–12. doi:10.1145/2078827.2078835.
Zheng N, Bai K, Huang H, Wang H. 2014. You are how you touch: user verification on smartphones via tapping behaviors. Presented at the 2014 IEEE 22nd International Conference on Network Protocols (ICNP), pp. 221–32. doi:10.1109/ICNP.2014.43.

Pin Shen Teh is currently a PhD student in the School of Computer Science at The University of Manchester, UK. He acquired his Bachelor of Information Technology (Hons) Software Engineering and Master of Science (IT) degree in 2006 and 2009, respectively. His research interest is in the area of biometrics, pattern recognition and mobile authentication.

Ning Zhang received the BSc degree in electronics engineering from Dalian Maritime University, Dalian, China, and the PhD degree in electronics engineering from the University of Kent, Canterbury, UK. She is currently a Senior Lecturer with the School of Computer Science, The University of Manchester, Manchester, UK. Her current research interests include mobile computing, security in

networked and distributed systems, applied cryptography, data privacy, and trust and digital right managements.

Andrew Beng Jin Teoh is an associate professor in The Electrical and Electronic Department, College Engineering of Yonsei University, South Korea. His research, for which he has received government and industry funding, focuses on biometric security and pattern recognition, specifically in biometric template protection and bio-crypto key computation. He has published more than 220 international refereed journals, conference articles, and several book chapters in the areas mainly in biometric security and biometric systems. He is a senior member of the IEEE Signal Processing Society. He serves as chair for IEEE Biometric Council Newsletter.

Ke Chen received his BSc, MSc, and PhD degrees in computer science in 1984, 1987, and 1990, respectively. He has been with The University of Manchester since 2003. He was with The University of Birmingham, Peking University, The Ohio State University, Kyushu Institute of Technology, and Tsinghua University. His main research interests include machine learning, pattern recognition, machine perception, computational cognitive systems and their applications in intelligent system development including computer games.