Behavioural Biometrics: A Survey And Classification

Introduction

With the explosion of computers and of the internet in our everyday lives, the need for trustworthy computer security gradually increases. Biometric technologies provide user friendly and trustworthy control methodology for access to computer systems, networks and workplaces. The majority of research is expected at studying well-established physical biometrics such as fingerprint or iris scans. Behavioural biometrics systems are usually less established, and only those which are in large part based on muscle control such as keystrokes, step or mark are well analysed of behavioural data often does not require any special hardware and is so very cost effective. While most behavioural biometrics are not unique enough to provide reliable human proof of identity, they have been shown to provide sufficiently high accuracy identity proof.

In completing their everyday tasks, human beings employ different plans, use different styles and apply unique skills and knowledge. One of the defining characteristics of a behavioural biometric is the combination of time dimension as a part of the behavioural mark. The measured behaviour has a beginning, period, and an end. Behavioural biometrics researchers attempt to quantify behavioural traits revealed by users and use resulting feature profiles to successfully verify identity. In this section, authors present an overview of most recognized behavioural biometrics.

Behavioural biometrics can be classified into five groups based on the type of information about the user being collected. Group one is made up of authorship based biometrics, which is based on examining a section of text or a drawing produced by a person. Authentication is accomplished by observing style individualities typical to the author of the work being examined, such as the used terminology, punctuation or brush strokes.

Group two consists of human computer interaction (HCI)-based biometrics. In their everyday interaction with computers, human beings employ different approaches, use different styles, and apply unique abilities and knowledge. Researchers attempt to quantify such traits and use resulting feature profiles to successfully verify uniqueness. HCI-based biometrics can be further subdivided into additional groups. The first group consists of human interaction with input devices such as keyboards, computer mice, and haptics which can register inherent, distinctive and consistent muscle actions. The second group consists of HCI-based behavioural biometrics which measures advanced human behaviour such as policy, knowledge or skill exhibited by the user during interaction with different software.

The third group is closely related to the second and is the set of the indirect HCI-based biometrics which are the events that can be found by monitoring user's HCI behaviour indirectly via noticeable low-level actions of computer software. These include system call traces, audit logs, program execution suggestions, registry access, storage activity, call-stack data analysis, and system calls. Such low-level events are produced involuntarily by the user during interaction with different software.

Same HCI-based biometrics is sometimes well-known to different researchers under different names. IDS based on system calls or audit logs are often categorized as utilising program execution traces and those based on call- stack data as based on system calls. The misperception is maybe related to the fact that a lot of interdependency exists between different indirect behavioural biometrics and they are often used in combinations to improve accuracy of the system being developed. For example, system calls and program counter data may be collective in the same behavioural mark or audit logs may contain information about system calls. Also one can't forget that a human being is indirectly behind each one of those mirror image of behaviour and so a large degree of connection is to be expected.

The fourth and possibly the best researched group of behavioural biometrics relies on motor-skills of the users to accomplish confirmation. Motor- skill is an skill of a human being to utilise muscles. Muscle movements rely upon the proper working of the brain, skeleton, joints, and nervous system and so, motor skills indirectly reflect the quality of working of such systems, making person confirmation possible. Most motor skills are learned, not inherited, with incapacities having potential to affect the development of motor skills. Authors accept definition for motor- skill based behavioural biometrics, a. k. a. 'kinetics', as those biometrics which are based on distinctive, unique and stable muscle actions of the user while execution a particular task.

The fifth and final group consists of purely behavioural biometrics. Purely behavioural biometrics measures human behaviour not directly absorbed on measurements of body parts or intrinsic, inimitable and lasting muscle movements such as the way an individual walks, types, or even holds a tool. Human beings utilise different strategies, skills and knowledge during presentation of mentally demanding tasks. Purely behavioural biometrics measures such behavioural traits and makes successful identity confirmation a possibility.

Behavioural Biometrics

1. E-mail behaviour: E-mail transfer behaviour is not the same for all individuals. Some people work at night and send masses of e-mails to many different addresses; others only check mail in the morning and only relate with one or two people. All this particularities can be used to create a behavioural profile which can assist as a behavioural biometric for an individual. Length of the

e-mails, time of the day the mail is sent, how normally inbox is emptied and of course the recipients' addresses among other variables can all be collective to create a baseline feature vector for the person's e-mail behaviour. Some work in using e-mail behaviour modelling was complete by Stolfo et al. They have investigated the possibility of identifying virus propagation via e-mail by observing abnormalities in the e-mail sending behaviour, such as unusual faction of recipients for the same e-mail. For example, sending the same e-mail to your girlfriend and your boss is not an everyday existence.

Vel et al. (2001) have applied authorship identification techniques to regulate the likely author of an e-mail message. Alongside the features used in text authorship identification, authors also used some e-mail specific physical features such as: use of a salutation, farewell acknowledgment, signature, number of attachments, location of re-quoted text within the message body, HTML tag incidence distribution, and total number of HTML tags. Overall, almost 200 features are used in the experiment, but some often cited features used in text authorship determination are not appropriate in the domain of e-mail messages due to the smaller average size of such communications.

2. Audit logs. Most modern operating systems keep some records of user action and program interaction. While such audit trails can be of some attention to behavioural intrusion detection researchers, specialised audit trails specifically intended for security enforcement can be potentially much more powerful. A typical audit log may contain such evidence as CPU and I/O usage, number of associates from each location, whether a directory was accessed, a file created, another user ID changed, audit record was altered, amount of activity for the system, network and host. Experimentally, it has been shown that gathering audit events is a less disturbing technique than recording system calls. Because a massive amount of auditing data can be produced overwhelming an intrusion detection system, it has been suggested that a random sampling might be a rational approach to auditing data. Additional data might be helpful in individual suspicious activity from normal behaviour. For example, facts about changes in user status, new users being added, finished users, users on vocations, or changed job obligations might be needed to reduce the number of false positives produced by the IDS. Since so much possibly valuable information can be captured by the audit logs, a large number of investigators are concerned to this form of indirect HCI-based biometrics.

3. Biometric sketch. Al-Zubi et al. (2003) and Brömme and Al-Zubi (2003) planned a biometrics sketch confirmation method based on sketch recognition and a user's personal knowledge about the drawings content. The system directs a user to generate a simple sketch for example of three circles and each user is free to do so in any way he satisfies. Because a large number of different combinations exist for combining multiple simple physical shapes, sketches of different users are sufficiently unique to provide accurate authentication. The method measures user's knowledge about the sketch, which is only available to the before genuine user. Such

features as the breezes location and comparative position of different primitives are taken as the profile of the sketch. Alike methods are tried by Varenhorst (2004) with a system named ‘passdoodles’ and also by Jermyn et al. (1999) with a system named ‘draw-a-secret’. Finally a ‘v-go password’ requests a user to perform simulation of simple actions such as mixing a concoction using a graphical interface, with the statement that all users have a personal approach to bartending.

4. Blinking. Westeyn et al. (2005) and Westeyn and Starner (2004) have industrialized a system for identifying users by analysing voluntary song-based blink patterns. During the enrolment phase user looks at the system’s camera and blinks to the beat of a song he has formerly chosen producing a so-called ‘blinkprint’. During confirmation phase, the user’s blinking is compared to the database of the stored blinked patterns to regulate which song is being blinked and as a result user identification is possible. In addition to the blink pattern itself additional features can also be extracted such as: time between blinks, how long the eye is held closed at each blink, and other physical features the eye undergoes while blinking. Based on those additional features, it was shown to be feasible to differentiate users blinking the same exact pattern and not just a secretly-selected song.

5. Credit card use. Data mining techniques are often used in detection of credit card fraud. Looking out for statistical outliers such as unusual transactions, expenditures to far-away geographical locations, or simultaneous use of a card at manifold locations can all be signs of a stolen account. Outliers are considerably different from the residue of the data points and can be detected by using discordancy tests. Approaches for fraud related outlier detection are based on distance, density, forecast, and distribution analysis methods. A generalised approach to finding outliers is to assume a known statistical circulation for the data and to evaluate the aberration of samples from the distribution. Brause et al. (1999) have used representative and analogue number data to perceive credit card fraud. Such transaction material as account number, transaction type, credit card type, merchant ID, merchant address, etc. were used in their rule-based model. They have also shown that similarity data alone can’t serve as a satisfying source for detection of fraudulent transactions.

6. Voice/speech/singing. Speaker proof of identity is one of the best researched biometric technologies. Confirmation is based on information about the speaker’s anatomical structure conveyed in amplitude spectrum, with the location and size of haunted peaks related to the vocal tract shape and the terrain striations related to the glottal source of the user. Speaker proof of identity systems can be secret based on the freedom of what is spoken:

1. Fixed text. The speaker says a specific word selected at enrolment.
2. Text dependent. The speaker is encouraged by the system to say a particular phrase.
3. Text independent. The speaker is free to say anything he wants, confirmation accuracy

typically improves with larger amount of spoken text.

Feature abstraction is applied to the normalized amplitude of the input signal which is further decomposed into several band-pass frequency channels. A often extracted feature is a logarithm of the Fourier transform of the voice indication in each band along with pitch, tone, cadence, and form of the larynx. Correctness of voice based biometrics systems can be increased by addition of visual speech (lip dynamics) and combination of soft behavioural biometrics such as accent. Recently some research has been aimed at increasing the developed technology to singer recognition for the purposes of music database management and to hilarity recognition. Currently, the laughter-recognition software is rather crude and cannot precisely distinguish between different people

- biometric draft,
- blinking,
- calling,
- car driving,
- command line lexicon,
- credit card use,
- dynamic facial features,
- e-mail,
- step,
- game policy,
- GUI interaction,
- handgrip,
- haptic,
- keystrokes,
- lip drive,
- mouse dynamics,
- image style,
- programming style,
- signature,
- tapping,
- transcript authorship,
- voice.

Conclusion

In this survey, authors have offered only the most popular behavioural biometrics but any human behaviour can be used as a basis for personal profiling and for subsequent confirmation. Some behavioural biometrics which are rapidly gaining ground but are not a part of this survey

include profiling of errands behaviour based on market basked, web browsing and click-stream profiling.

Behavioural biometrics are mainly well suited for confirmation of users which interact with computers, cell phones, smart cars, or points of sale terminals. As the number of electronic applications used in homes and offices increases, so does the probable for utilisation of this paper and promising technology. Future research should be directed at increasing overall correctness of such systems, e.g., by looking into opportunity of developing multimodal behavioural biometrics; as people often involve in multiple behaviours at the same time, e.g., talking on a cell phone whereas driving or using keyboard and mouse at the same time.