This is the published version of this conference paper:


© Copyright 2010 IEEE

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
Continuous Biometric Authentication: Can It Be More Practical?

Eesa Al Solami, Colin Boyd, Andrew Clark and Asadul K Islam
Information Security Institute, Queensland University of Technology
GPO Box 2434, Brisbane 4001, Queensland, Australia
e.alsolami@student.qut.edu.au
{c.boyd, a.clark, a.islam}@qut.edu.au

Abstract—Continuous biometric authentication schemes (CBAS) are built around the biometrics supplied by user behavioural characteristics and continuously check the identity of the user throughout the session. The current literature for CBAS primarily focuses on the accuracy of the system in order to reduce false alarms. However, these attempts do not consider various issues that might affect practicality in real world applications and continuous authentication scenarios. One of the main issues is that the presented CBAS are based on several samples of training data either of both intruder and valid users or only the valid users’ profile. This means that historical profiles for either the legitimate users or possible attackers should be available or collected before prediction time. However, in some cases it is impractical to gain the biometric data of the user in advance (before detection time). Another issue is the variability of the behaviour of the user between the registered profile obtained during enrollment, and the profile from the testing phase. The aim of this paper is to identify the limitations in current CBAS in order to make them more practical for real world applications. Also, the paper discusses a new application for CBAS not requiring any training data either from intruders or from valid users.

I. INTRODUCTION

Attacks on computer systems can be undertaken at the network, system or user levels [17]. Network-level attacks include network denial of service and probing. System-level attacks include privilege escalation, such as buffer overflow, program modification, perhaps caused by a Trojan horse or virus, and denial of service [1]. User-level attacks include masquerader and imposter attacks. Most research undertaken in recent years is concerned with system- and network-level attacks. However, there is a lack of research on attacks at the user level. This paper focuses on attacks at the user level, especially attacks during the session.

User-level attacks include the imposter or intruder who takes over from the valid user, either at the start of a computer session or during the session. Unauthorised user access was the second greatest source of financial loss according to a 2006 CSI/FBI Computer Crime and Security Survey [7]. Considering the risks in highly sensitive environments, a single, initial authentication might be insufficient to guarantee security. It may also be necessary to perform continuous authentication to prevent user substitution after the initial authentication step.

The impact of an intruder during a session is the same as any kind of false representation at the beginning of a session. Most current computer systems authorise the user at the start of a session and do not detect whether the current user is still the initial authorised user, a substitute user, or an intruder pretending to be a valid user. Therefore, a system that continuously checks the identity of the user throughout the session is necessary. Such a system is called a Continuous Authentication System (CAS).

The paper is organized as follows. The next section presents CBAS background and the motivation of the paper. In the following section we highlight the contribution of the paper. In Section II, we present the CBAS model in order to describe the characteristics and attributes of existing CBAS, and to describe the requirements of different scenarios of CBAS. Following this in Section III, we show the limitations of the existing CBAS. In Section IV, we describe a new application for CBAS without utilising training data. The final section concludes the paper and proposes future work.

A. Continuous Biometric Authentication System (CBAS)

The majority of existing CBAS are built around the biometrics supplied by user traits and characteristics [15]. There are two major forms of biometrics: those based on physiological attributes and those based on behavioural characteristics. Each one has its advantages and disadvantages. The physical type includes biometrics based on stable body traits, such as fingerprint, face, iris, and hand and are considered to be more robust and secure. However, they are also considered to be more intrusive and expensive and require regular equipment replacement. On the other hand, the behavioural type includes learned movements such as handwritten signature, keyboard dynamics (typing), mouse movements, gait and speech. They are less obtrusive and they do not require extra hardware. However, they are considered to be less accurate than physiological biometrics. This paper will consider the use of behavioural biometrics for providing continuous user authentication. Most of the literature in the CBAS is based on behavioural biometrics.

There has been an ongoing pursuit of improving CBAS. Recently, efforts have been made to improve CBAS by either embedding intrusion detection into the CBAS itself [16] or by adding a new biometric source tailored to the detection system [21]. However, these attempts do not consider the different limitations that might affect the practicability of existing CBAS in real world applications and continuous authentication.
scenarios. To our knowledge, there is no CBAS deployed in real world applications; it seems reasonable to assume that this is because existing systems lack practicality. There are a number of issues associated with existing schemes which may prevent CBAS from being applied in real world applications. These limitations include:

- the requirement for the training data to be available in advance;
- the need for too many training data samples;
- variability of the behaviour biometric between the training and testing phase; and
- variability of the behaviour biometric of the user from one context to another.

B. Contribution

A generic model is proposed for most continuous authentication (CA) scenarios and CBAS. The model of CBAS is proposed based on their detection capabilities to better identify and understand the characteristics and requirements of each type of scenario and system. This model pursues two goals: the first is to describe the characteristics and attributes of existing CBAS, and the second is to describe the requirements of different scenarios of CBAS. Also, we identify the main issues and limitations of existing CBAS, observing that all of the issues are related to the training data. Finally, we consider a new application for CBAS without requiring any training data either from intruders or from valid users in order to make the CBAS more practical.

II. CBAS MODEL

To date, CBAS have only been described by the techniques and algorithms used to detect an imposter and to decide whether or not to raise an alarm. Therefore, there is a need to build a generic model to help ensure the identification of the common characteristics and attributes of CBASs. The traditional authentication mechanisms, such as user name and password, are used to verify the user only at the start of a session and do not detect throughout the session whether or not the current user is still the initial authorised user. The CBAS is designed to overcome this limitation by continuously checking the identity of the user throughout the session based on the physical traits or behavioral characteristics of the user. A CBAS can be thought of as a kind of intrusion detection system (IDS). An IDS monitors a series of events, with the aim of detecting unauthorised activities. In the case of the CBAS, the unauthorised activity is a person acting as an imposter by taking over the authenticated session of another (valid) user.

The CBAS attacks include imposters or intruders that take over from the valid user during the session. However, the IDS attacks include both external intrusion attacks caused by an outsider and misuse attacks caused by an insider. Most IDS operate at the system or network level, and very little research performed by the intrusion detection community has focused on monitoring user-level events; notable exceptions include the recent work by Killourhy and Maxion [14]. In this paper, we use the more specific term, CBAS, to describe a scheme which aims to detect imposters through the continuous monitoring of a user session.

There are five basic components to describe a typical CBAS (see Figure II)

1) Subjects: initiators of activity on a target system, normally users, either authorised or unauthorised [5]. An authorised user is allowed to access a system by providing some form of identity and credentials. Also, they are allowed to deal with objects of the system during the session. The authorised user can be known to the system where the biometric data of that user is registered in the system as a historical profile. On the other hand, the authorised user can be unknown to the system where there is no biometric data of that user registered in the system in advance. The unauthorised user can be an adversary user acting maliciously towards the valid user or it can be a colluder invited by the valid user to complete an action on behalf of the user.

2) Sensor: a device that collects biometric data from the user, either physically or behaviourally, and translates it into a signal that can be read by an observer or an instrument such as keyboard or camera. The aim of data collection is to obtain biometric data to keep on record, and to make further analysis of that data. The sensors are based upon one or more physical traits or behavioral characteristics [15] for uniquely recognizing humans. The physical type includes biometrics based on stable body traits, such as fingerprint, face, iris, and hand [19]. The behavioural type includes learned movements such as handwritten signature, keyboard dynamics (typing), mouse movements, gait and speech [22].

3) Detector: compares profiles by analysing the biometric data and then performs measurements for errors that may detect the intruder. There are two major algorithms of detection in the surveyed systems: based on the historical profile of all the users; or only the valid user profile. The first method is a type of machine learning algorithm that requires biometric data about all users, either valid or possible imposters, to build a model
The existing CBAS schemes can be described based on the requirement of training data from user. The first class of CBAS requires the training data for both intruders and valid users. The second class requires the training data only for the valid users. The characteristics of the two classes of CBAS are described below in detail.
There are a number of examples of existing CBAS schemes that fall under Class 1. Gunetti et al. [8] created profiles for each user based upon their typing characteristics in free text. They performed a series of experiments using the degree of disorder to measure the distance between the profiles of known users to determine how well such a measure performs when assigning unidentified users to known profiles.

Pusara et al. [21] applied a similar approach using mouse movements to compare a test sample to every reference sample in the database. In that case, the learning algorithm used the training data for all users to determine decision boundaries that discriminate between these users by matching the closest user.

B. Class 2

The main difference between Class 1 and Class 2 is that in Class 2 we do not assume that a profile is available for the imposter. That is, the identity of authorised user who initiates the session is known (as is their corresponding profile which is stored in the database), but no profile is available for the imposter. The characteristics of the class can be summarised as follows.

1) This class requires the CA scenario to be in the open setting in a non-restricted environment. The environment could be in a public location where any user can use the computer system.
2) The user who is authorised to use the system at the start of the session could be a known user (as in Class 1).
3) The authorised user who is already authenticated at the start of the session should have training data registered in the database in advance (as in Class 1).
4) The training data for the unauthorised user, such as an imposter or intruder, who would try to claim the identity of the authenticated user during the session is not available or it is not possible to collect such data in advance of the prediction time.
5) The unauthorised user in this class would be an adversary user and the victim in this case would be the end user.
6) The labeled normal data from the valid users should be used in order to build the detection model. This approach is similar to the one-class classifier that learns to differentiate between one class in the training data, which is then used to predict the class of an unseen instance by making a decision about whether or not it is related or similar enough to the training class. Since systems in class 2 do not require labels for the anomaly class, they are more widely applicable than multi-class classification techniques. The approach used in such techniques is to build a model for the class corresponding to normal behaviour, and use the model to identify anomalies in the test data.

All of these characteristics of Class 2 are present in the online banking scenario described below as Scenario 2. In the online banking scenario the training data for only the valid users is available or collected before prediction time. So, this class could be suitable for some scenarios held in non-restricted environments such as public locations where the biometric data for possible imposters is not available or where it is not possible to collect such data in advance.

Scenario 2: In an online banking system, transactions are typically made in an open-setting environment and the intruder is likely to come from outside. This form of vulnerability is known as the outsider threat. The location of data collection in this scenario might be distributed as it occurs in open-setting environment and the location of data processing would be centralised on a computer server. The bank administration normally secures the communication channels between the user and server in the bank in order to avoid threats to the network traffic. However, threats can occur at the user level when the system authenticates the user at the start of the session and later if the user leaves without logging off from the session. An intruder can then take over and conduct transactions on behalf of the valid user who is already authenticated. The system accepts the whole transaction as performed by the valid user.

There are a number of examples of existing CBAS schemes that fall under Class 2. Hempstalk et al. [11] created profiles only for valid users based upon their typing characteristics in free text. They performed a series of experiments using the Gaussian density estimation techniques by applying and extending an existing classification algorithm to the one-class classification problem that describes only the valid user biometric data. They applied a density estimator algorithm in order to generate a representative density for the valid users data in the training phase, and then combined the predictions of the representative density of the valid user and the class probability model for predicting the new test cases.

Azzini et al. [3] applied a similar approach using multi-modal biometrics including face recognition and fingerprint to compare unidentified data to only the valid users data in the database. At prediction time, the learning algorithm used the training data for only the valid user to determine a decision based on comparison between the unidentified data with the class of the valid users.

C. Limitations in the current CBAS

Most previous CBAS schemes are mainly concerned with the accuracy of the system in reducing false alarms. However, these schemes do not consider issues that might affect their practicality in different real-world applications and continuous authentication scenarios. All previous schemes require training data either of both intruder and valid users like Class 1, or of valid users like Class 2. Limitations related to the training data which prevent CBAS from being applied in a practical way as a real-world application include the following.

1) The existing schemes require the historical profile either from legitimate users or from possible attackers to be available or collected before prediction time. It is impossible in some cases to gain the biometric data of the user in advance of the detection time.
2) Some schemes require many training samples of the user in the enrollment phase in order to build a profile of that user and apply it in the testing or comparison phase.
This is likely to present a severe inconvenience for some users.

3) User behaviour biometric data varies between the training and testing profile. The cause of the variability of the data could be due to the following. a) The valid user physical mood is not the same at all times and could well be different between the enrollment phase and testing phase. This could affect the user’s typing speed, for example. As a result, the false rejection rate can be increased. b) The typing speed of the valid user changes over time. This can affect the stability of the keystroke dynamics systems, for example.

4) User behaviour varies from one context to another, so every context needs a new historical profile for comparison in that context. For example, typing speed in an online exam may not be the same as the typing speed in the case of typing email.

5) Comparison between the new profile and the historical profiles in the database takes time and can delay the detection.

IV. A NEW APPLICATION FOR CBAS

In this section, we describe a new application for CBAS which, in a sense, is the most difficult case. In this application, we assume that no profile information is available for any users, authorised or not, at the beginning of the session. We do however, assume that the user who initiates the session is ‘authorised’ to do so. The challenge here is to build a profile of this ‘authorised’ user while, at the same time, trying to decide whether or not the session has been taken over by an imposter. A summary of the characteristics of the new application for CBAS follows.

1) This class requires the CA scenario to be in the open setting and in a non-restricted environment. The environment should be in a public location so that any user can use the computer system (as in Class 2).
2) While the identity of the (authorised) user who initiates the session may be known, no profile for this user is available prior to the commencement of the session.
3) Similarly, the training data for the unauthorised user is not available or cannot be collected before the prediction time.
4) The unauthorised user in this class would be an adversary user and the victim in this case of attacking would be the end user (as in Class 2).
5) There are no labels from both normal and anomalous data to be used in order to build the detection model in this class. This approach is similar to change point detection [12] that learns from the data on the fly. It considers probability distributions from which data in the past and present intervals are generated, and regards the target time point as a change point if two distributions are significantly different. Other methods may also be applicable to the new application.

In this class, the system determines whether the biometric data in the testing phase is related to one user or two users by trying to identify any significant change within the biometric data. There are three main challenges associated with this class.

1) How much biometric data should be available before it is possible to identify a significant change in the biometric data related to the imposter or intruder?
2) How much time does the system need to detect the imposter?
3) How can the system determine the start and the end activity of an imposter or intruder?

While there are likely to be a number of existing techniques which are potentially useful for the new application of CBAS, we note that most of the characteristics of this class are similar to the change point detection problem, and here we focus the discussion on assessing the applicability of change point detection techniques to the new application for CBAS. Future work will be required to perform a practical evaluation of the applicability of change point detection techniques for CBAS.

Change point detection is the problem of discovering time points at which properties of time-series data change [12]. Change point detection has been applied to a broad range of real world problems such as fraud detection in cellular systems [18], intrusion detection in computer networks [2], irregular-motion detection in vision systems [13], signal segmentation in data streams [4], and fault detection in engineering systems [6].

There is a clear need to develop some schemes for the new application for CBAS associated with change point algorithms. Various approaches to change-point detection have been investigated within this statistical framework, including the CUSUM (cumulative sum) [2] and GLR (generalized likelihood ratio) [10][9].

We now provide an example of the basic of CUSUM to show how it can be applied with CBAS. Let \( X_1, X_2, \ldots, X_{100} \) represent 100 data points. In the case of behavioural biometrics the data points could be the time between consecutive keystrokes, for example. From this, the cumulative sums \( S_0, S_1, S_2, \ldots, S_{100} \) are calculated. Notice that 100 data points leads to 101 (0 through 100) sums. The cumulative sums are calculated as follows.

1) First calculate the average:

\[
V = \frac{v_1 + v_2 + \ldots + v_{100}}{100}
\] (1)

2) Start the cumulative sum at zero by setting \( S_0 = 0 \).
3) Calculate the other cumulative sums by adding the difference between the value of current data point and the average to the previous sum

\[
S_i = S_{(i-1)} + (X_i - V)
\] (2)

for \( i = 1, 2, \ldots, 100 \).

Here \( S_i \) denotes the score of cumulative sum for the current value, \( S_{(i-1)} \) the previous score of cumulative sum, and \( X_i \) the current value. \( V \) is the average of all values in the case of offline detection, but in case of online detection it will be the average value for the previous values. As the average is subtracted from each value, the cumulative sum also ends at zero (\( S_{100} = 0 \)). After that, we compare the score of the
cumulative sum for each value with the threshold in order to identify the change in the data.

We can extend the basic technique of CUSUM to be based on a growing window. In the growing window method, the number of patterns or values is not fixed but increases. When a new pattern is added to the data set, the oldest patterns stay so that the growing window stores all the patterns that have been granted access as authenticated patterns. Another approach that might be used with CUSUM is the sliding window concept. In the sliding window method, the number of patterns is fixed. When a new pattern is added to the data set, the oldest pattern is removed. These, and other approaches need to be compared to determine the strength of each one.

The new schemes associated with change point algorithms may overcome the current limitations in the CBAS that have been identified in Section III-C. Specifically, change point analysis can be applied without the need for training data. New schemes associated with change point detection algorithms detect the attacker based on the data itself and, therefore, have the potential to be faster. However, further research is required to explore suitable behavioural biometric features for use in such change point detection algorithms.

Table II summarises the differences between the first, second and the new application.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Class 1</th>
<th>Class 2</th>
<th>New application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of environment</td>
<td>Closed</td>
<td>Open</td>
<td>Open</td>
</tr>
<tr>
<td>Training data</td>
<td>Authorised / unauthorised user</td>
<td>Authorised user</td>
<td>None</td>
</tr>
<tr>
<td>Unauthorised user</td>
<td>Adversary / colluder</td>
<td>Adversary</td>
<td>Adversary</td>
</tr>
<tr>
<td>Victim type</td>
<td>System operator / owner of the application</td>
<td>End user</td>
<td>End user</td>
</tr>
<tr>
<td>Algorithm type</td>
<td>Multi-class classification</td>
<td>One-class classification</td>
<td>Change point detection (potentially)</td>
</tr>
</tbody>
</table>

Table II: The Differences Between the First, Second Classes and the New Application.

V. Conclusion

We have analysed existing continuous biometric authentication schemes and described sample continuous authentication scenarios. We identified the common characteristics and attributes from the generic model of CBAS. To date there is no CBAS deployed in real world applications, probably because the existing systems lack practicality. We observed that the main limitations are related to the training data which prevent CBAS to be applicable in the real world applications. The problems are the requirement of the training data to be available in advance, too many training data samples required, the variability of the behaviour biometric between the training and testing phase in case of the comparison time and the variability of the behaviour biometric of the user from one context to another. Finally, the paper considered a new application for CBAS associated (potentially) with change point detection algorithms that does not require training data for both intruder and valid users which can overcome the identified limitations associated with the existing CBAS.

REFERENCES