Computers and Electrical Engineering xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Multi-modal decision fusion for continuous authentication ☆

Lex Fridman [a,*], Ariel Stolerman [b], Sayandeep Acharya [a], Patrick Brennan [c], Patrick Juola [c], Rachel Greenstadt [b], Moshe Kam [d]

[a] Department of Electrical and Computer Engineering, Drexel University, 3141 Chestnut St, Philadelphia, PA 19104, USA
[b] Department of Computer Science, Drexel University, 3141 Chestnut St, Philadelphia, PA 19104, USA
[c] Juola & Associates, 301 Grant St, Suite 4300, Pittsburgh, PA 15219, USA
[d] Newark College of Engineering, New Jersey Institute of Technology, 323 Dr Martin Luther King Jr Blvd, Newark, NJ 07102, USA

ARTICLE INFO

Article history:
Received 16 December 2013
Received in revised form 26 October 2014
Accepted 27 October 2014
Available online xxxx

Keywords:
Multimodal biometric systems
Distributed communication
Security and privacy
Behavioral biometrics
Decision fusion
Active authentication

ABSTRACT

Active authentication is the process of continuously verifying a user based on their ongoing interaction with a computer. In this study, we consider a representative collection of behavioral biometrics: two low-level modalities of keystroke dynamics and mouse movement, and a high-level modality of stylometry. We develop a sensor for each modality and organize the sensors as a parallel binary decision fusion architecture. We consider several applications for this authentication system, with a particular focus on secure distributed communication. We test our approach on a dataset collected from 67 users, each working individually in an office environment for a period of approximately one week. We are able to characterize the performance of the system with respect to intruder detection time and robustness to adversarial attacks, and to quantify the contribution of each modality to the overall performance.

1. Introduction

The challenge of identity verification for the purpose of access control in distributed communication systems is the trade-off between maximizing the probability of intruder detection, and minimizing the cost for the legitimate user in time, distractions, and extra hardware and computer requirements. In recent years, behavioral biometric systems have been explored extensively in addressing this challenge [1].

Behavioral biometric systems rely on computer interface devices such as the keyboard and mouse that are already commonly available with most computers, and are thus low cost in terms of having no extra equipment requirements. However, their performance in terms of detecting intruders, and maintaining a low-distraction human–computer interaction (HCI) experience has been mixed [2], showing error rates ranging from 0% [3] to 30% [4] depending on context, variability in task selection, and various other dataset characteristics.

The bulk of biometric-based authentication work focused on verifying a user based on a static set of data. This type of one-time authentication is not sufficiently applicable to a live multi-user environment, where a person may leave the computer for an arbitrary period of time without logging off. This context necessitates continuous authentication when a computer is in a non-idle state. Validated access is important on two levels: (1) locally, to protect the offline data on the computer being used, and (2) globally, to protect the data traveling on a secured distributed network of which the computer is a part of [5]. To represent a real-world scenario where such an authentication system may be used, we created a simulated

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. Felix Gomez.
* Corresponding author.

http://dx.doi.org/10.1016/j.compeleceng.2014.10.018
0045-7906/

office environment in order to collect behavioral biometrics associated with typical human–computer interaction (HCI) by an office worker over a typical work week.

Using the data collected in an office environment, we consider a representative selection of behavioral biometrics, and show that through a process of fusing the individual decisions of sensors based on those metrics, we can achieve better performance than that of the best sensor from our sensor set. Due to their heterogeneous nature, it stands to reason that a properly designed set of good sensors would outperform a single sensor which is "best" under specific circumstances. Moreover, given the low cost of installing these application-level sensors, this approach may prove to be a cost-effective alternative to sensors based on physiological biometrics [6]. We consider twelve sensors, each falling in one of three biometrics categories: keystroke dynamics, mouse movement, and stylometry.

We propose to use decision fusion in order to integrate the sensor bank and make serial authentication decisions. While we consider here specific twelve sensors, the strength of our decision-level approach is that additional sensors can be added to the sensor bank without having to change the basic fusion rule, and with only minimal performance information required about the added sensors. Moreover, it is easy to evaluate the marginal improvement of any added sensor to the overall performance of the system.

We evaluate the multimodal continuous authentication system on a large real-world dataset. We consider several parameters and metrics in presenting the system's performance. First, we look at the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) when the decisions from each of the twelve sensors are combined in the decision fusion center (DFC). Second, we assess the relative contribution of each individual sensor to the performance of the overall decision. Third, we observe the tradeoff between the time to first authentication decision and the error rates. Fourth, we consider adversarial attacks on the system in the form of sensor "spoofing," and show that the system is robust to partial spoofing.

The remainder of the paper is structured as follows. In Section 2, we discuss the related work on behavioral biometrics via the modalities considered in this paper and multimodal fusion. In Section 3, we discuss the simulated work environment dataset used for training and testing. In Section 4, we discuss twelve behavioral biometrics sensors. In Section 5, we detail the fusion algorithm. In Section 6, we present the performance of this system on a 67 user dataset.

2. Related work

2.1. Keystroke dynamics and mouse movement

Keystroke dynamics is one of the most extensively studied topics in behavioral biometrics [7]. The feature space that has been investigated ranges from the simple metrics of key press interval [8] and dwell [9] times to multi-key features such as trigraph duration with an allowance for typing errors [2]. Furthermore, a large amount of classification methods have been studied for mapping these features into authentication decisions. Broadly, these approaches fall in one of two categories: statistical methods [10] and neural networks [11], with the latter generally showing higher FAR and FRR rates, but better able to train and make predictions on high-dimensional feature space.

While keyboard and mouse have been the dominant forms of HCI since the advent of the personal computer, mouse movement dynamics has not received nearly as much attention in the biometrics community in the last two decades as keystroke dynamics have. Most studies on mouse movement were either inconclusive due to small number of users [12] or required an excessively large static corpus of mouse movement data to achieve good results [1], where an FAR and FRR of 0.0246 is achieved from a testing window of 2000 mouse actions. The work in [13] drastically reduces the size of the testing window to 20 mouse clicks. We base our selection of the three mouse metrics on their work but with more emphasis on mouse movement and not the mouse button presses.

One of the benefits of the mouse as behavioral biometric sensor is that it has a much simpler physical structure than a keyboard. Therefore, it is less dependent on the type of mouse and the environment in which the mouse is used. Keyboards, on the other hand, can vary drastically in size, response, and layout, potentially providing different biometric profiles for the same user. The simulated environment dataset we consider utilizes identical computer and working environment, so in our case, this particular robustness benefit is not important to authentication based on this data.

2.2. Stylometry

Authorship attribution based on linguistic style, or stylometry, is a well-researched field [14]. The main domain it is applied on is written language – identifying an anonymous author of a text by mining it for linguistic features. The theory behind stylometry is that everyone has a unique linguistic style ("stylome" [15]) that can be quantified and measured in order to distinguish between different authors. The feature space is potentially endless, with frequency measurements or numeric evaluations based on features across different levels of the text, including function words, grammar, character *n*-grams [16] and more.

The most common practice of stylometry is in supervised learning, where a classifier is trained on texts of candidate authors, and used to attribute the stylistically closest candidate author to unknown writings. In an unsupervised setting, a set of writings whose authorship is unknown are classified into style-based clusters, each representing texts of some unique author.

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

In an active authentication setting, authorship verification is applied, where unknown text is classified by a unary author-specific classifier. The text is attributed to an author if and only if it is stylistically close enough to that author. Although pure verification is the ultimate goal, standard authorship attribution as a closed-world problem is an easier (and sometimes sufficient) goal. In either case, classifiers are trained in advance, and used for real-time classification of processed sliding windows of input keystrokes. If enough windows are recognized as an author other than the real user, it should be considered as an intruder.

In a pure authorship attribution setting, where classification is done off-line, on complete texts (rather than sequences of input keystrokes) and in a supervised setting where all candidate authors are known, state-of-the-art stylometry techniques perform very well. For instance, at PAN-2012,[1] some methods achieved more than 80% accuracy on a set of 241 documents, sometimes with added distractor authors.

In an active authentication setting, a few challenges arise. First, open-world stylometry is a much harder problem, with a tendency to high false-negative (false reject) rates. The unmasking technique [17] has been shown effective on a dataset of 21 books of 10 different 19th-century authors, obtaining 95.7% accuracy. However, the amount of data collected by sliding windows of sufficiently small durations required for an efficient authentication system, along with the lack of quality coherent literary writings make this method perform insufficiently for our goal. Second, the inconsistent frequency nature of keyboard input along with the relatively large amount of data required for good performance of stylometric techniques make a large portion of the input windows unusable for learning writing style.

On the other hand, this type of setting allows some advantages in potential features and analysis method. Since the raw data consists of all keystrokes, some linguistic and technical idiosyncratic features can be extracted, like misspellings caught prior to being potentially auto-corrected and vanished from the dataset, or patterns of deletions (selecting a sentence and hitting delete versus repeatedly hitting backspace deleting character at-a-time). In addition, it is more intuitive in this kind of setting to consider overlap between consecutive windows, resulting with a large dataset, grounds for local voting based on a set of windows and control of the frequency in which decisions are outputted by the system.

2.3. Multimodal biometric systems

A defining problem of active authentication arises from the fact that a verification of identity must be carried out continuously on a sample of sensor data that varies drastically with time. The classification therefore has to be made based on a "window" of recent data, dismissing or heavily discounting the value of older data outside that window. Depending on what task the user is engaged in, some of the biometric sensors may provide more data than others. For example, as the user browses the web, the mouse-related sensors will be actively flooded with data, while the keystroke dynamics and stylometry sensors may only get a few infrequent key press events. This motivates the recent work on multimodal authentication systems where the decisions of multiple classifiers are fused together [18]. In this way, the verification process is more robust to the dynamic mode of real-time HCI. The current approaches to the fusion of classifiers center around max, min, median, or majority vote combinations [19]. When neural networks are used as classifiers, an ensemble of sensors is constructed and fused based on different initialization of the neural network [20].

Several active authentication studies have utilized multimodal biometric systems but have all, to the best of our knowledge: (1) considered a smaller pool of subjects, (2) have not characterized the temporal performance of intruder detection, and (3) have shown overall significantly worse performance than that achieved in our study. In particular, [21] have looked at similar classes of biometrics: keyboard dynamics, mouse movement, and stylometry. They used different features and classifiers, and did not propose a fusion scheme, but rather investigated each modality separately. The overall performance achieved ranged approximately from error rates of 0.1–0.4, which are significantly worse than the error rates achieved using the approach proposed in this paper. Two fusion methods and a rich portfolio of features similar to the ones in this paper were considered in [22] to achieve multi-modal authentication performance of 0.021 FAR and 0.024 FRR on a subject pool of 31 users. These error rates are an order of magnitude worse than those achieved in our work, and use a larger time window of 10 minutes.

Our approach in this paper is to apply the Chair–Varshney optimal fusion rule [23] for the combination of available multimodal decisions. Furthermore, we are motivated by the work in [24] that greater reduction in error rates is achieved when the classifiers are distinctly different (i.e. using different behavioral biometrics). The strength of the decision-level fusion approach is that an arbitrary number of sensors can be added without re-training the sensors already in the system. This modular design allows for multiple groups to contribute drastically different classification schemes, each lowering the error rate of the global decision.

3. Dataset

The source of behavioral biometrics data we utilized for testing multi-modal fusion for the task of active authentication comes from a simulated work environment. In particular, we put together an office space, organized and supervised by a subset of the authors. We placed five desks in this space with a laptop, mouse, and headphones on each desk. This equipment

[1] http://pan.webis.de.

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

and supplies were chosen to be representative of a standard office workplace. One of the important properties of this dataset is that of uniformity. Due to the fact that the computers and input devices in the simulated office environment were identical, the variation in behavioral biometrics data can be more confidently attributed to variation in characteristics of the users.

During each of the sixteen weeks of the data collection we hired 5 temporary employees for 40 h of work. Each day they were assigned two tasks. The first was an open-ended blogging task, where they were instructed to write blog-style articles related in some way to the city in which the testing was carried out. This task was allocated 6 h of the 8 h workday. The second task was less open-ended. Each employee was given a list of topic or web articles to write a summary of. The articles were from a variety of reputable news sources, and were kept consistent between users except for a few broken links due to the expired lifetime of the linked pages. This second task was allocated 2 h of the 8 h workday.

Both tasks encouraged the workers to do extensive online research by using the web browser. They were allowed to copy and paste content, but they were instructed that the final work they produced was to be of their own authorship. As expected, the workers almost exclusively used two applications: Microsoft Word 2010 for word processing and Internet Explorer for browsing the web.

While the tasks were specified and suggested a combination of online research and word processing, the resulting behavior patterns were quite different. The productivity of workers, as measured by the number of words typed, varied drastically. They were purposefully not graded nor encouraged to be more productive, and therefore, tended to spend a large amount of their time browsing the web like they would outside of work: pursuing various interests, writing emails, commenting and chatting on Facebook and other social networks. In this way, the data we collected is representative of broader computer use than simply writing a blog on a particular subject. Each subject's interests and concerns outside of work had significant impact on their interaction with the computer.

Some of the users did not show up for work on one or more days. There were also several days on which the tracking software was shutdown prematurely for a user. Therefore, there were a few users for who the amount of data collected was significantly lower than the median. Therefore, we only used data from users who had over 54,000 s (15 h) of *active* interaction with the computer. Before filtering out users in this way, we removed idle period in the data stream, where "idle" is defined as a period where neither the mouse nor keyboard were used for longer than 2 min. All such periods were shrunk down to 2 min. Therefore, due to such a temporal compression of the data, the 54,000 s threshold is based on active interaction with the computer. In this way we reduced the number of users in the dataset under consideration in this work from 80 down to 67.

Three data files produced by two tracking applications. They contain the following data:

- Mouse movement, mouse click, and mouse scroll wheel events at a granularity of 5 ms.
- Keystroke dynamics (include press, hold, release durations) for all keyboard keys including special keys at a granularity of 5 ms.
- Mapping of keys pressed to the application in focus at the time of the keyboard's use as input. The granularity for this data is 1 s but by synchronizing with the data from the first two streams, higher resolution timing information can be inferred.

Table 1 shows statistics on the biometric data in the corpus. The table contains data aggregated over all 67 users. It also shows the average amount of data available per user. The keystroke events include both the alpha-numeric keys and also the special keys such as `shift`, `backspace`, `ctrl` and `alt`. In counting the key presses and the mouse clicks for Table 1, we count just the down press and not the release.

As an example of the variation in the dataset, Fig. 1 shows a heat map visualization of the aggregate first-day mouse movements for 14 of the 67 users. It provides an intuition that the users have unique behavioral profiles of interaction with the computer via the mouse to a degree that distinct patterns emerge even in heat maps that aggregate a full day's worth of data. Some users spend a lot of time on the scroll bar, some users focus their attention to the top left of the screen, and some users frequently move their mouse big distances across the screen.

4. Behavioral biometric modalities

The sets of features we consider in this paper are linguistic style (stylometry), mouse movement patterns, and keystroke dynamics. We construct sensors (or classifiers) from these features differently depending on the feature. For keystroke

Table 1
Statistics on the 67-user subset of the biometric data contained in the dataset.

Metric	Total	Per user
Mouse move events	34,626,337	516,811
Mouse clicks	628,862	9386
Scroll wheel events	404,531	4397
Keystroke events	1,243,286	13,514

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

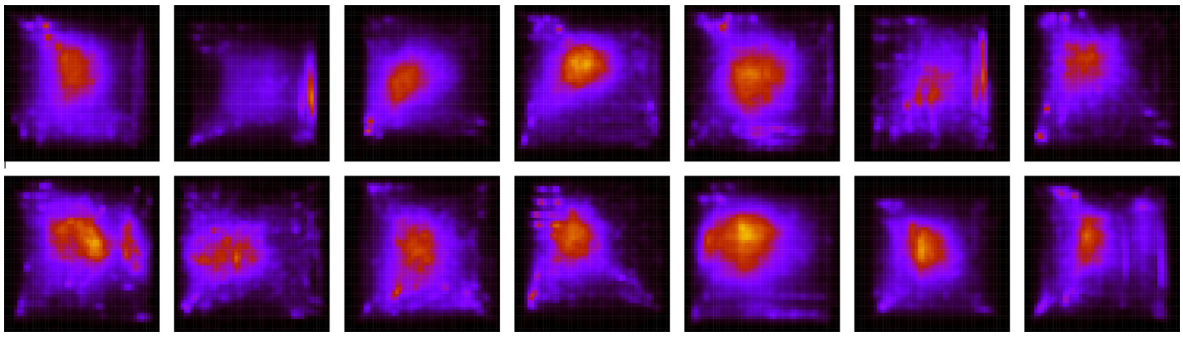

Fig. 1. Each of the above subfigures is a visualization of aggregate mouse movement for one of the 67 users on their first day. We are only presenting 14 of the 67 users. This heat map is constructed by mapping the mouse movement data from the associated user to a 50 by 50 cell square image. The brighter the intensity of the cell, the more visits are recorded in that area of the screen. These figures visualize the intuition that there are distinct differences in the way each individual user interacts with the computer via the mouse.

dynamics and mouse movement features, each individual feature is tracked by one sensor that uses a Naive Bayes classifier [25]. For stylometry, the portfolio of features is combined into one sensor using support vector machines (SVMs) [26]. Each of these types of sensors work differently in terms of required amount of input data, type of collected data (mouse events, keystroke event) and performance.

We broadly categorize the sensors in this paper according to the degree of conscious cognitive involvement measured by the sensors. The distinction can be thought of as that between "how" and "what". We refer to the mouse movement and keystroke dynamics sensors as "low-level", since they measure *how* we use the mouse and *how* we type. On the other hand, the website domain frequency and stylometry sensors are "high-level" because they track *what* we click on with the mouse and *what* we type. Table 2 shows the twelve sensors under consideration in this paper. The frequency listed is an upper-bound on frequency that a sensor produces a classification. The actual frequency depends on the time-based windows size that the sensors is configured to use in training and testing phases.

4.1. Keystroke dynamics and mouse movement

For any change in the position of the mouse, the raw data received from the mouse tracker are (1) the pixel coordinates of the new position and (2) the delay in milliseconds between the recording of this new position and the previously recorded action. Usually that delay is 5 ms, but sometimes the sampling frequency degrades for short periods of time. This tuplet gives us the basic data element based on which all the mouse movement metrics are computed (given an initial position on the screen).

In this paper, we consider nine mouse-based metrics as listed in Table 2, and illustrated in Fig. 3. A "mouse curve" is an uninterrupted sequence of three mouse move events. A "mouse path" is an uninterrupted sequence of mouse move events with other type of events before and after it. A "click path" is a mouse path that ends in a mouse button click. Conversely, a "nonclick path" is a mouse path that ends in an event other than a mouse button click. The mouse sensors are based on features of these sequences of mouse events.

We chose two of the simplest and most frequently occurring keystroke dynamics features as illustrated in Fig. 2: (K1) the interval between the release of one key and the press of another and (K2) the dwell time between the press of a key and its

Table 2
The sensors whose performance is investigated in this paper. These include 1 stylometry, 2 keystroke, and 9 mouse sensors. For each sensor, listed is the average frequency across all 67 users that an event associated with that sensor is observed during active interaction with the computer.

Metric	Frequency (Hz)
1. Key press duration	0.1295
2. Key interval	0.1248
3. Mouse curve distance	1.0271
4. Mouse curve curvature	0.7153
5. Mouse button press duration	0.0423
6. Mouse click-path speed	0.0385
7. Mouse click-path wandering	0.0385
8. Mouse click-path angle	0.0385
9. Mouse nonclick-path speed	0.0201
10. Mouse nonclick-path wandering	0.0201
11. Mouse nonclick-path angle	0.0201
12. Stylometry	0.1295

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

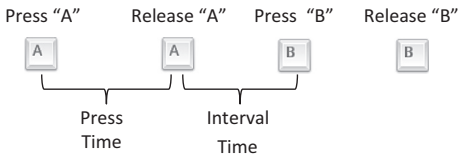

Fig. 2. The keystroke dynamics metrics are computed from the time between the press and release event and visa versa.

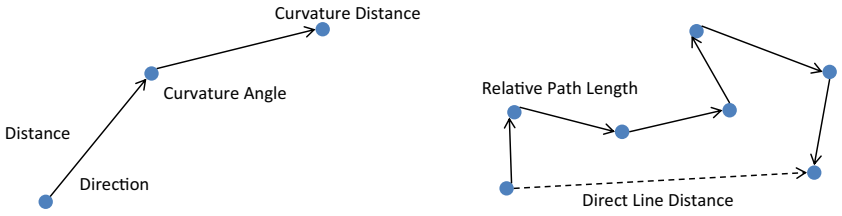

Fig. 3. The mouse movement metrics are computed from a set of continuous move events (defined by positions on the virtual screen). On the left are three points that define a "mouse curve" and based on which the mouse curve distance and curvature metrics are computed. On the right are 3 or more points that define a "mouse path" and based on which the mouse path speed, angle, and "wandering" metrics are based.

release. While the dwell time K2 is a strictly positive number, the interval K1 can be negative if another key is pressed before a prior one is released.

4.2. Stylometry

We chose the setting of closed-world stylometry: we developed classifiers trained on the closed set of users. The classifier's output is the author to which the text is attributed.

In the preprocessing phase, we parsed the keystrokes log files to produce a list of documents consisting of overlapping windows for each user, with the following time-based sizes (in seconds): 10, 30, 60, 300, 600 and 1200. For the first 3 settings we advanced the sliding window with steps of 10 s, and for the last 3 – steps of 60 s. The step size determines how often a decision can be made by the sensor.

During preprocessing, only keystrokes were considered and all special keys were converted to unique single-character placeholders. For instance `BACKSPACE` was converted to β and `PRINTSCREEN` was converted to π. Any representable special keys like `|t` and `|n` were taken as is (i.e. tab and newline, respectively).

The constructed feature set, denoted the *AA* feature set hereinafter, is a variation of the *Writeprints* [27] feature set, which includes a vast range of linguistic features across different levels of text. A summarized description of the features is presented in Table 3. By using a rich linguistic feature set we hope to capture the user's writing style. With the special-character placeholders, some features capture aspects of the user's style usually not found in standard authorship problem settings. For instance, frequencies of backspaces and deletes provide some evaluation of the user's typo-rate.

The features were extracted using the *JStylo* framework[2] [28], an open-source authorship attribution platform. JStylo was chosen since it is equipped with fine feature definition capabilities. Each feature is uniquely defined by a set of its own document preprocessing tools, one unique feature extractor (the core of the feature), feature postprocessing tools, and normalization/factoring options. The features available in JStylo are either frequencies of a class of related features (e.g., frequencies of "*a*", "*b*", ..., "*z*" for the "letters" feature class) or some numeric evaluation of the input document (e.g., average word length, or Yule's Characteristic *K*). Its output is compatible with the data mining and machine learning platform Weka [29], which we used for the classification process.

Two important processing procedures were applied in the feature extraction phase. First, every word-based feature (e.g., the function words class, or different word-grams) was applied a tailor-made preprocessing tool developed for this unique dataset, that applies the relevant special characters on the text. For instance, the character sequence `ch`$\beta\beta$`Cch`$\beta\beta$`hicago` becomes `Chicago`, where β represents backspace. Second, since the windows are determined by time and not amount of collected data, normalization is crucial for all frequency-based features (which consist the majority of the features).

For classification, we used sequential minimal optimization (SMO) support vector machines [30] with polynomial kernel, available in Weka. Support vector machines are commonly used for authorship attribution [31] and known to achieve high performance and accuracy.

Finally, the data was analyzed with the stylometry sensors using a varying threshold for minimum characters-per-window to consider, spanning from 100 to 1000 with steps of 100. For every threshold set, all windows with less than that

[2] http://psal.cs.drexel.edu/.

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

Table 3
The *AA* feature set. Inspired by the *Writeprints* [27] feature set, includes features across different levels of the text. Some features are normalized frequencies of feature classes; others are numerical evaluations of the input text.

Group	Features
Lexical	Avg. word-length Characters Most common character bigrams Most common character trigrams Percentage of letters Percentage of uppercase letters Percentage of digits Digits 2-Digit numbers 3-Digit numbers Word length distribution
Syntactic	Function words Part-of-speech (POS) tags Most common POS bigrams Most common POS trigrams
Content	Words Word bigrams Word trigrams

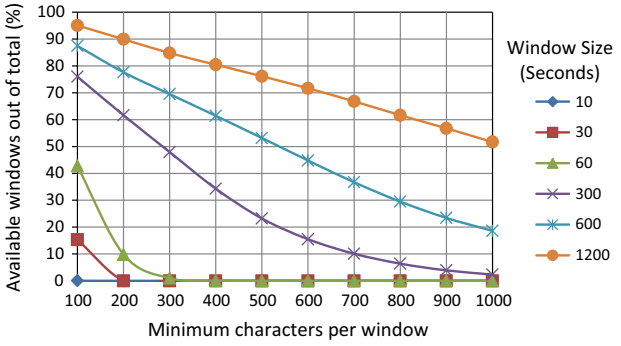

Fig. 4. Percentage of remaining windows out of the total windows after filtering by the minimum characters-per-window threshold.

amount of characters were thrown away, and for those windows the sensor output was "no decision". The different thresholds allow us to assess the tradeoff in the sensor's performance in terms of accuracy and availability: as the threshold increases, the window is richer with data and will potentially be classified with higher accuracy, but the portion of total windows that pass the threshold decreases, making the sensor less available. Fig. 4 illustrates the average percentage of usable windows, after removing all those that do not pass the minimum characters-per-window threshold.

5. Decision fusion

The motivation for the use of multiple sensors to detect an event is to harness the power of the sensors to provide an accurate assessment of a studied phenomenon, which a single sensor may not be able to provide. In centralized architectures, raw data from all sensors monitoring the same space are communicated to a central point for integration, the fusion center. However quite often the use of a centralized architecture is not desirable or practical. The factor weighing against centralization is the need to transfer large volumes of data between local detector and fusion center. Another is the fact that in many systems specialized local detectors already exist, and it is more convenient to fuse their decisions rather than re-create the detection algorithms at the fusion center. In the distributed architectures, some processing of data is performed at each sensor, and the resulting information is sent out from each sensor to a central processor for subsequent processing and final decision making. On most scenarios significant reduction in required bandwidth for data transfer and modularity are the main advantages of this approach. The price is sub-optimality of the decision/detection scheme.

Decision fusion with distributed sensors is described by Tenney et al. in [32] who studied a parallel decision architecture. As described in [33], the system comprises of n local detectors, each making a decision about a binary hypothesis (H_0, H_1), and a decision fusion center (DFC) that uses these local decisions $\{u_1, u_2, \ldots, u_n\}$ for a global decision about the hypothesis. The ith detector collects K observations before it makes its decision, u_i. The decision is $u_i = 1$ if the detector decides in favor

of H_1 (decision D_1), and $u_i = -1$ if it decides in favor of H_0 (decision D_0). The DFC collects the n decisions of the local detectors through ideal communication channels and uses them in order to decide in favor of $H_0 (u = -1)$ or in favor of $H_1 (u = 1)$. Fig. 5 shows the architecture and the associated symbols. Tenney et al. [32] and Reibman and Nolte [34] studied the design of the local detectors and the DFC with respect to a Bayesian cost, assuming the observations are independent conditioned on the hypothesis. The ensuing formulation derived the local and DFC decision rules to be used by the system components for optimizing the system-wide cost. The resulting design requires the use of likelihood ratio tests by the decision makers (local detectors and DFC) in the system. However the thresholds used by these tests require the solution of a set of nonlinear coupled differential equations. In other words, the design of the local decision makers and the DFC are co-dependent. In most scenarios the resulting complexity renders the quest for an optimal design impractical.

Chair and Varshney in [23] developed the optimal fusion rule when the local detectors are fixed and local observations are statistically independent conditioned on the hypothesis. Data Fusion Center is optimal with respect to a Bayesian cost, given the performance characteristics of the local fixed decision makers. The result is a suboptimal (since local detectors are fixed) but computationally efficient and scalable design. In this study we use the Chair–Varshney formulation. As described in [33], the *Bayesian risk* $\beta^{(k)}(C_{00}, C_{01}, C_{10}, C_{11})$ is defined for the kth decision maker in the system as

$$\beta^{(k)}(C_{00}, C_{01}, C_{10}, C_{11}) = C_{00}^{(k)} Pr(H_0, D_0) + C_{10}^{(k)} Pr(H_0, D_1) + C_{01}^{(k)} Pr(H_1, D_0) + C_{11}^{(k)} Pr(H_1, D_1) \tag{1}$$

where $C_{00}^{(k)},\ C_{01}^{(k)},\ C_{10}^{(k)},\ C_{11}^{(k)}$ are the prespecified cost coefficients of the kth decision maker for each combination of hypothesis and detector decision: $C_{ij}^{(k)}$ is the cost incurred when the k^{th} decision maker decides D_i when H_j is true. For the cost combination $C_{00}^{(k)} = C_{11}^{(k)} = 0$ and $C_{01}^{(k)} = C_{10}^{(k)} = 1$, the Bayesian cost becomes the *probability of error*. We consider a suboptimal system where each detector $k = 1, 2, \ldots, n$ minimizes locally a Bayesian risk $\beta^{(k)}$ and the DFC ($k = 0$) is optimal with respect to $\beta^{(0)}$, given the local detector design. In the subsequent work, we assume $\beta^{(k)} = \beta^{(0)},\ k = 1, 2, \ldots, n$ (all local detectors minimize the same Bayesian risk) and the superscript k is therefore omitted. Specifically we use throughout the paper

$$\begin{aligned} C_{00}^{(k)} &= C_{11}^{(k)} = 0,\ k = 1, 2, \ldots, n \\ C_{10}^{(k)} &= C_{01}^{(k)} = 1,\ k = 1, 2, \ldots, n \end{aligned} \tag{2}$$

namely the local detectors and the DFC each minimizes the probability of error.

5.1. Fusion rule

The parallel distributed fusion scheme (see Fig. 5) allows each sensor to observe an event, minimize the local risk and make a local decision over the set of hypothesis, based on only its own observations. Each sensor sends out a decision of the form:

$$u_i = \begin{cases} 1, & \text{if } H_1 \text{ is decided} \\ -1, & \text{if } H_0 \text{ is decided} \end{cases} \tag{3}$$

The fusion center combines these local decisions by minimizing the global Bayes' risk. The optimum decision rule performs the following likelihood ratio test

$$\frac{P(u_1, \ldots, u_n | H_1)}{P(u_1, \ldots, u_n | H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} = \tau \tag{4}$$

where the a priori probabilities of the binary hypotheses H_1 and H_0 are P_1 and P_0 respectively and C_{ij} are the costs as defined previously. For costs as defined in (2), the Bayes' risk becomes total probability of error and the right hand side of (4) becomes $\frac{P_0}{P_1}$. In this case the general fusion rule proposed in [23] is

$$f(u_1, \ldots, u_n) = \begin{cases} 1, & \text{if } a_0 + \sum_{i=0}^{n} a_i u_i > 0 \\ -1, & \text{otherwise} \end{cases} \tag{5}$$

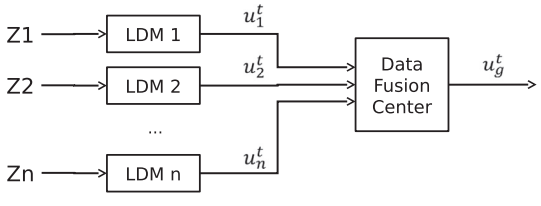

Fig. 5. Architecture for the fusion of decentralized detectors.

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

with P_i^M, P_i^F representing the *False Rejection Rate* (FRR) and *False Acceptance Rate* (FAR) of the ith sensor respectively. The optimum weights minimizing the global probability of error are given by

$$a_0 = \log\frac{P_1}{P_0} \tag{6}$$

$$a_i = \begin{cases} \log\frac{1-P_i^M}{P_i^F}, & \text{if } u_i = 1 \\ \log\frac{1-P_i^F}{P_i^M}, & \text{if } u_i = -1 \end{cases} \tag{7}$$

Kam et al. in [33] developed expressions for the global performance (global FAR and FRR) of the distributed system described above. Exact expressions for global error rates are given in [33].

The threshold in (4) requires knowledge of the a priori probabilities of the hypotheses. In practice, these probabilities are not available, and the threshold τ is determined using different considerations (such as fixing the probability of false alarm of the DFC).

5.2. Extendable fusion framework

As is explain in Section 5.1, the performance of the fused global detector improves as the number of local sensors increases. Furthermore, it is shown in [24] that fusion of classifiers trained on distinct feature sets leads to greatest reduction in system error. In our context, the ideal active authentication system gathers input from as many different behavioral biometric sensors as possible. In designing the fusion system one of our goals was to provide a straightforward way of adding sensors to the system without having to change algorithms and with simple and uniform characterization of each sensor. In fact our formulation requires only that the FAR and FRR be supplied, so that they can be incorporated in (6) and (7).

6. Continuous authentication on a 67 user dataset

6.1. Training, characterization, testing

The data of each of the 67 users' active interaction with the computer was divided into 5 equal-size folds (each containing 20% time span of the full set). We performed training of each classifier on the first three folds (60%). We then tested their performance on the fourth fold. This phase is referred to as "characterization", because its sole purpose is to form estimates of FAR and FRR for use by the fusion algorithm. We then tested the performance of the classifiers, individually and as part of the fusion system, on the fifth fold. This phase is referred to as "testing" since this is the part that is used for evaluation the performance of the individual sensors and the fusion system. The three phases of training, characterization, and testing as they relate to the data folds are shown in Fig. 6.

- Training on folds 1, 2, 3. Characterization on fold 4. Testing on fold 5.
- Training on folds 2, 3, 4. Characterization on fold 5. Testing on fold 1.
- Training on folds 3, 4, 5. Characterization on fold 1. Testing on fold 2.
- Training on folds 4, 5, 1. Characterization on fold 2. Testing on fold 3.
- Training on folds 5, 1, 2. Characterization on fold 3. Testing on fold 4.

The common evaluation method used with each sensor for data fusion was measuring the averaged error rates across five experiments; In each experiment, data of 3 folds was taken for training, 1 fold for characterization, and 1 for testing. The FAR and FRR computed during characterization were taken as input for the fusion system as a measurement of the expected

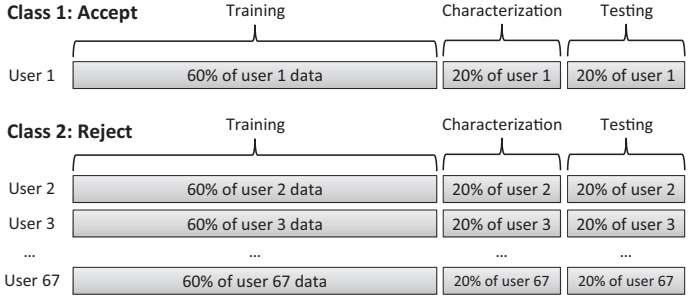

Fig. 6. The three phases of processing the data to determine the individual performance of each sensors and the performance of the fusion system that combines some subset of these sensors.

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

performance of the sensors. Therefore each experiment consisted of three phases: (1) train the classifier(s) using the training set, (2) determine FAR and FRR based on the training set, and (3) classify the windows in the test set.

Unless otherwise specified, the experiments we ran were using the fusion system on the full 67 user set with the 2 keystroke dynamics sensors, 9 mouse sensors, and the stylometry sensor.

6.2. Contribution of individual sensors

For each low-level sensor, we used the Naive Bayes classifier [35] for mapping from the feature space to the decision space. For the stylometry sensor, we used an SVM [36] as described in Section 2. In the training phase for low-level sensors, the empirical distribution for feature probabilities were constructed from the frequency of each feature in the training segment of each user's data. Two such histograms were constructed for each user j. The first histogram was constructed from the training segment of the data of that user. The second histogram was constructed from all the training segments of the other users. These two histograms are the empirical feature distributions associated with each user.

In the characterization and testing phases, for each user and each metric, the Naive Bayes Classifier considered a collection of events $\Omega = \{x_t | T_{\text{current}} - T(x_t) \leqslant \omega\}$ where ω is a fixed window size in seconds, $T(x_t)$ is the timestamp of event x_t, and T_{current} is the current timestamp. The maximum a posteriori (MAP) rule was then used to pick the most likely hypothesis:

$$H^* = \underset{i\in\{0,1\}}{\text{argmax}} P(H_i) \prod_{x_t\in\Omega} P(x_t|H_i), \tag{8}$$

where H_1 is the "authentic" class, H_0 is the "non-authentic" class, as discussed in Section 5, and H^* is the most likely class associated with the observed biometric data. Unless otherwise stated we assume $P(H_0) = P(H_1) = 0.5$. The feature probability $P(x_t|H_i)$ is estimated by a non-parametric distribution formed in the training phase.

Fig. 7 shows the FAR and FRR rates respectively for the 11 keystroke and mouse movement sensors. For all four figures, the performance is averaged over 67 users and characterized with respect to the time-window size used by each of the sensors. Any data older than the duration of the window is discarded. The sensor only provides a decision when the time-window includes a minimum amount of events. For both mouse and keyboard that threshold was set to 5 events. As the size of the decision window increases, the FAR and FRR rates generally decrease for all sensors. The performance of the individual sensors varies from error rates as low as 0.01 to above 0.3.

The absolute performance of the fusion system is presented Section 6.3, but first we look at the contribution of each of the 11 low level sensors of keystroke dynamics and mouse movement to the overall performance of the fusion system (see

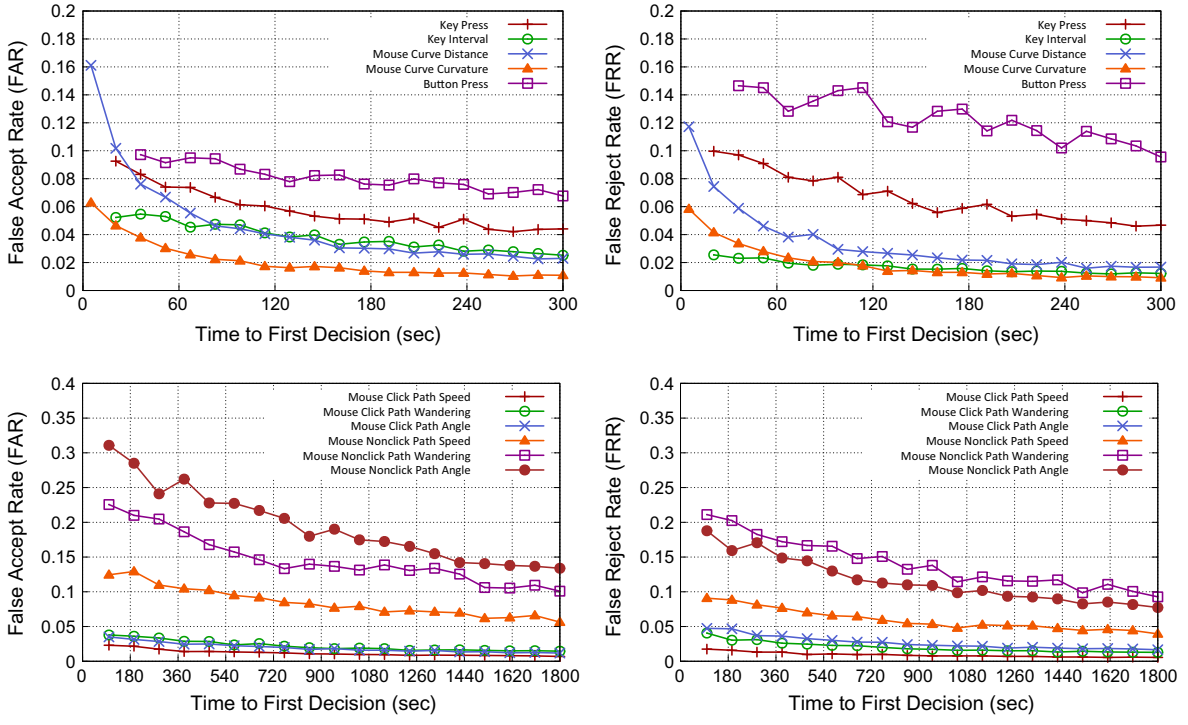

Fig. 7. FAR and FRR performance of the 11 keystroke dynamics and mouse movement sensors. Note that the range of the plots for the first set of sensors is shorter (300 s) than for the second set of sensors (1800 s).

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

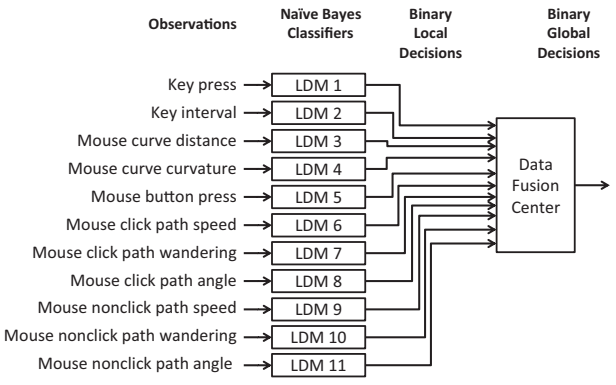

Fig. 8. The portfolio of 11 low-level sensors based on keystroke dynamics and mouse movement that forms the basis of our evaluations in Section 6.

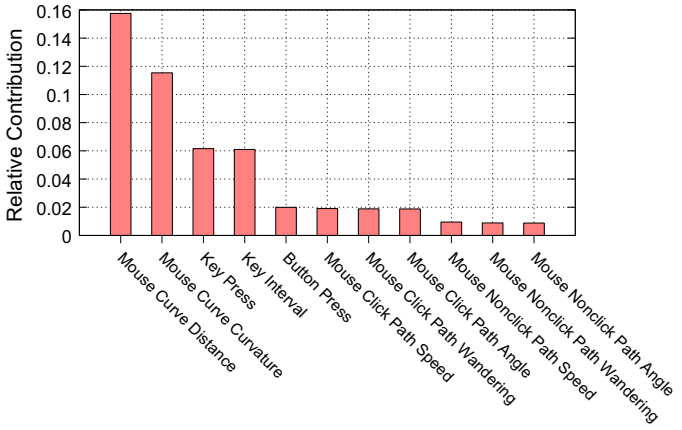

Fig. 9. Relative contribution of each of the 11 low-level sensors of keystroke dynamics and mouse movement to the fused decision. The contribution is computed according to (9).

Fig. 8). We measure this relative contribution C_i by evaluating the performance of the system with and without the sensor, and computing the contribution by:

$$C_i = \frac{E_i - E}{E_i} \tag{9}$$

where E is the error rate computed by averaging FAR and FRR of the fusion system using the full portfolio of 11 low-level sensors, E_i is the error rate of the fusion system using all but the i-th sensor, and C_i is the relative contribution of the i-th sensor as shown in Fig. 9.

Sensors based on the features of mouse curve distance, mouse curve curvature, key press duration, and key interval contributed the most to the fused decision. This can be explained by the fact that these four metrics are also those that appear with the highest frequency. Therefore, while their error rates individually are not always the lowest, the frequency of their "firing" makes up for a higher error rate when backed by the portfolio of the other sensors. On a time scale of 60–120 s where the low-level sensors excel, the stylometry sensor performed poorly and contributed almost zero to the overall decision, and thus was not included in the figure. The stylometry sensor begins contributing considerably on a longer time scale of 10–30 min.

6.3. Time to first decision

Two conflicting metrics of an active authentication system are response-time and performance. The less the system waits before making an authentication decision, the higher the expected rate of error. As more keystroke and mouse events trickle in, the system can refine its classification decision from an initial "neutral" stance of FAR = FRR = 0.5. In Fig. 10, we show the tradeoff between the decision time and performance.

The "time to first decision" is the time between the first keyboard or mouse event and the first decision produced by the fusion system. This metric can be thought of as "decision window size". Events older than the time range covered by the

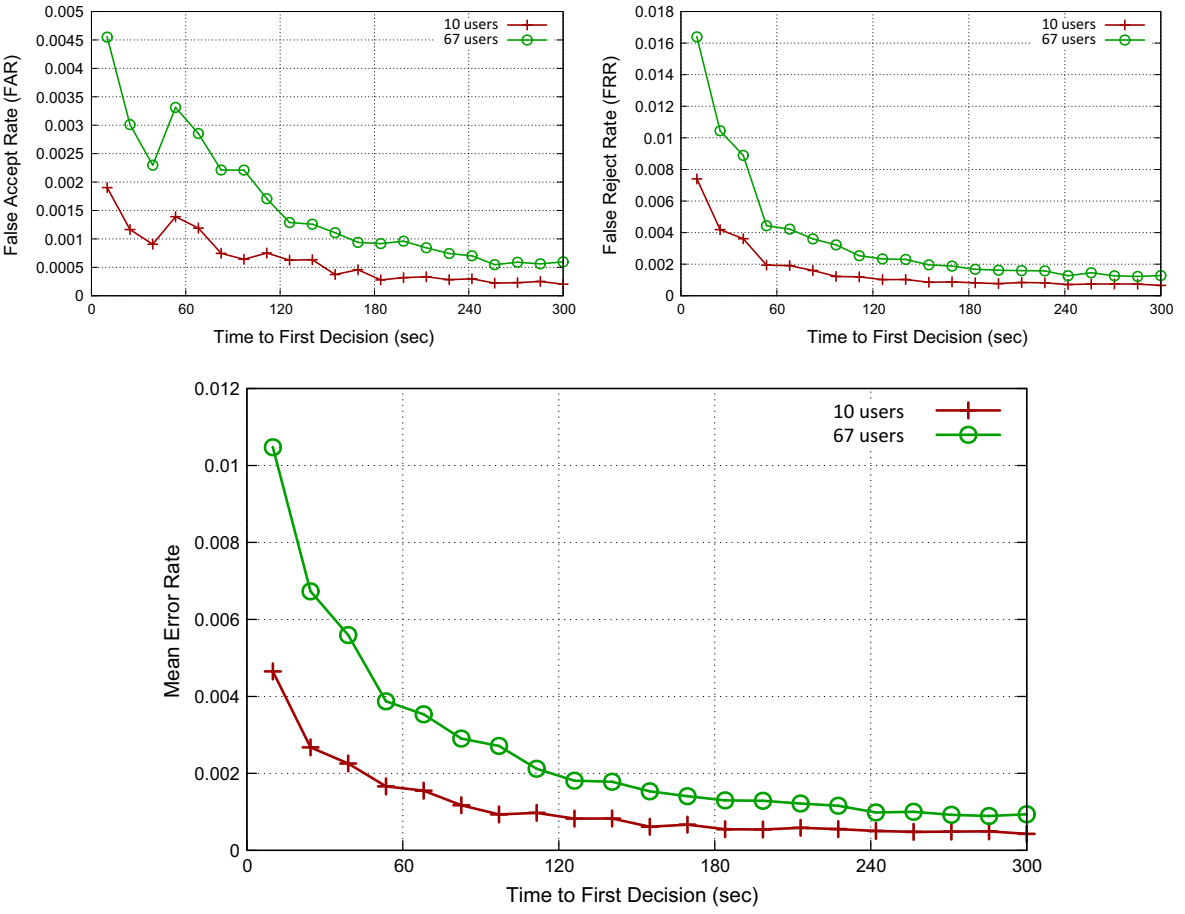

Fig. 10. The performance of the fusion system of 11 low-level sensors on the 10 users and 67 users. The standard deviation of each data point is small, with the coefficient of variation less than 0.5 for each point.

time-window are disregarded in the fused decision. As describe in Section 6.2 when a decision window contains less than 5 events, no decision is produced by the fusion system.

As the size of the decision window increases, the performance of the system improves, dropping below 0.01 FAR and FRR in 30 s as shown in Fig. 10. These plots also compare the performance of the fusion system on a 10 user subset and the full 67 user dataset. Performance degrades but not significantly and gives promise to the scalability of the system in the closed world environment.

When the user of the system changes, a decision window will contain a mix of events from two different users. In Fig. 11 the second user is an "intruder". The decision value "+1" corresponds to a valid user. The decision value "−1" corresponds to an intruder. The figure shows the real-time detection of an intruder based on two different decision windows of 10 s and 100 s. The complete detection period in this case is approximately equal to twice the decision window because both the individual sensors and the fusion system are using the same size window. For example, for a 100 s window, it is not until 100 s after the intruder enters that sensors are operating purely on the data received from the intruder and not on the previous user. It is not until 200 s after the intruder enters that the fusion system integrates sensor data based purely on the intruder interaction with the computer.

6.4. Robustness to partial spoofing

"Partial spoofing" is the successful mimicking of a valid user by an adversary on a subset of sensors contributing to the fused decision. The result is that the spoofed sensors incorrectly classify the current user as the valid user. We emulate this form of perfect spoofing by feeding valid user data to the sensors marked as "spoofed". Fig. 12 shows how the performance of the system degrades with an increasing number of spoofed sensors, in order from highest-contributing to lowest as shown in Fig. 9. In other words, *mouse curve distance* was spoofed first, *mouse curve curvature* was spoofed second, and so on. The performance of the partially-spoofed fusion system is evaluated using the FAR metric, since what is being measured is the rate at which the system incorrectly identifies an intruder as a valid user. The same sensors and fusion system described in Section 6.3 were used to generate the results in this section.

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

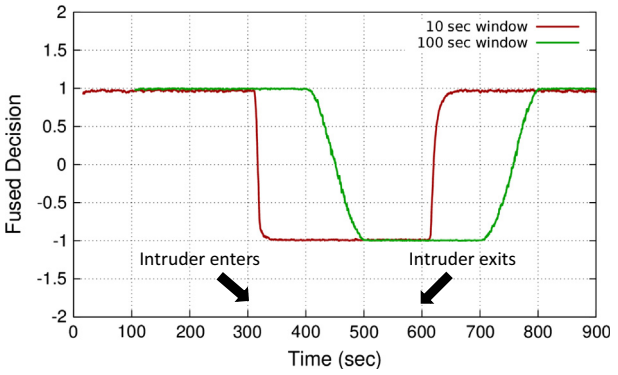

Fig. 11. Visualization of the real-time detection of an intruder averaged over 10,000 random samples of data from the 67 user dataset. A decision value of 1 indicates that the system believe the user to be authentic, and −1 otherwise. Due to the low error rates of the fusion system, an intruder is successfully detected even with small time-window of 10 s.

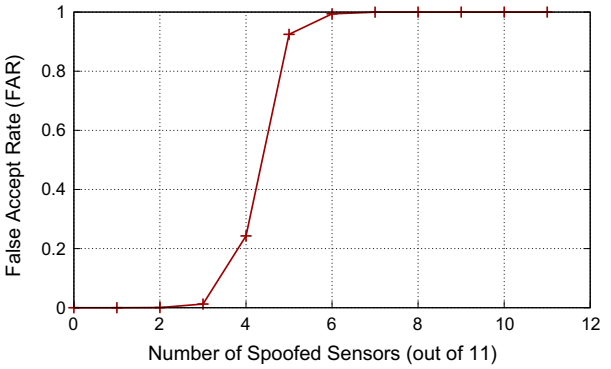

Fig. 12. FAR of a partially-spoofed fusion system. The sensors are compromised in the order of decreasing contribution as shown in Fig. 9. As the number of spoofed sensors increases from 0 to 11, the performance of the system degrades from nearly 0 to nearly 1 FAR. The standard deviation of each data point is small, with the coefficient of variation less than 0.5 for each point.

7. Conclusion

In this work, we proposed a parallel binary decision-level fusion architecture for a representative collection of behavioral biometric sensors: keystroke dynamics, mouse movement, and stylometry. Using this fusion method we addressed the problem of active authentication and characterized its performance on a dataset from a real-world office environment of 67 subjects each working on desktop computers for a period of one week. On the full dataset, the authentication system achieved false accept rates (FAR) of 0.004 and false reject rates (FRR) of 0.01 after 30 s of user interaction with the device. These error rates decreased to below 0.001 and 0.002, respectively, after 5 min. We showed that the performance of the system degrades, but not significantly with the size of the dataset. Further, we showed the contribution of each of the 11 sensors to the performance of the global decision maker. Of these 11 sensors, on average, the highest contributor was found to be "mouse curve distance", and the lowest contributor was found to be "mouse nonclick path angle." Lastly, we demonstrated that the system is robust to partial spoofing of the sensors. For the case of 11 sensors, the fusion system was robust to spoofing of 3 of the 11, beginning to degrade significantly in performance when 4 of the 11 sensors were compromised.

Acknowledgements

This work is supported by DARPA under BAA-12-06. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA.

References

[1] Ahmed A, Traore I. A new biometric technology based on mouse dynamics. IEEE Trans Depend Secure Comput 2007;4(3):165–79. http://dx.doi.org/10.1109/TDSC.2007.70207.
[2] Bergadano F, Gunetti D, Picardi C. User authentication through keystroke dynamics. ACM Trans Inf Syst Secur 2002;5(4):367–97. http://dx.doi.org/10.1145/581271.581272.

[3] Obaidat M, Sadoun B. Verification of computer users using keystroke dynamics. IEEE Trans Syst Man Cybernet Part B: Cybernet 1997;27(2):261–9. http://dx.doi.org/10.1109/3477.558812.
[4] Ord T, Furnell S. User authentication for keypad-based devices using keystroke analysis. In: Proceedings of the second international network conference (INC-2000); 2000. p. 263–72.
[5] Catania CA, Garino CG. Automatic network intrusion detection: current techniques and open issues. Comput Electr Eng 2012;38(5):1062–72. http://dx.doi.org/10.1016/j.compeleceng.2012.05.013. special issue on Recent advances in security and privacy in distributed communications and image processing. <http://www.sciencedirect.com/science/article/pii/S0045790612001073>.
[6] JAMES L. Fundamentals of biometric authentication technologies. Int J Image Graphics 2001;1(01):93–113.
[7] Karnan M, Akila M, Krishnaraj N. Biometric personal authentication using keystroke dynamics: a review. Appl Soft Comput 2011;11(2):1565–73.
[8] Bartlow N, Cukic B. Evaluating the reliability of credential hardening through keystroke dynamics. In: 17th International symposium on software reliability engineering, 2006. ISSRE'06. IEEE; 2006. p. 117–26.
[9] Giot R, El-Abed M, Rosenberger C. Keystroke dynamics authentication for collaborative systems. In: International symposium on collaborative technologies and systems, 2009. CTS'09. IEEE; 2009. p. 172–9.
[10] Umphress D, Williams G. Identity verification through keyboard characteristics. Int J Man–Machine Stud 1985;23(3):263–73.
[11] Bleha S, Knopp J, Obaidat M. Performance of the perceptron algorithm for the classification of computer users. In: Proceedings of the 1992 ACM/SIGAPP symposium on applied computing: technological challenges of the 1990's. ACM; 1992. p. 863–6.
[12] Pusara M, Brodley C. User re-authentication via mouse movements. In: Proceedings of the 2004 ACM workshop on visualization and data mining for computer security. ACM; 2004. p. 1–8.
[13] Zheng N, Paloski A, Wang H. An efficient user verification system via mouse movements. In: Proceedings of the 18th ACM conference on computer and communications security, CCS '11. New York (NY, USA): ACM; 2011. p. 139–50. http://dx.doi.org/10.1145/2046707.2046725. <http://doi.acm.org/10.1145/2046707.2046725>.
[14] Jockers ML, Witten D. A comparative study of machine learning methods for authorship attribution. LLC 2010;25(2):215–23.
[15] van Halteren H, Baayen RH, Tweedie F, Haverkort M, Neijt A. New machine learning methods demonstrate the existence of a human stylome. J Quant Linguist 2005;12(1):65–77.
[16] Stamatatos E. On the robustness of authorship attribution based on character n-gram features. J Law Policy 2013;21(2).
[17] Koppel M, Schler J. Authorship verification as a one-class classification problem. In: Proceedings of the twenty-first international conference on machine learning, ICML '04. New York (NY, USA): ACM; 2004. p. 62. http://dx.doi.org/10.1145/1015330.1015448. <http://doi.acm.org/10.1145/1015330.1015448>.
[18] Sim T, Zhang S, Janakiraman R, Kumar S. Continuous verification using multimodal biometrics. IEEE Trans Pattern Anal Mach Intell 2007;29(4):687–700.
[19] Kittler J, Hatef M, Duin R, Matas J. On combining classifiers. IEEE Trans Pattern Anal Mach Intell 1998;20(3):226–39.
[20] Chen C-H, Chen C-Y. Optimal fusion of multimodal biometric authentication using wavelet probabilistic neural network. In: 2013 IEEE 17th international symposium on consumer electronics (ISCE). IEEE; 2013. p. 55–6.
[21] Eusebi C, Gilga C, John D, Maisonave A. A data mining study of mouse movement, stylometry, and keystroke biometric data. In: Proc CSIS research day, Pace Univ.
[22] Bailey KO, Okolica JS, Peterson GL. User identification and authentication using multi-modal behavioral biometrics. Comput Secur 2014;43(0):77–89.
[23] Chair Z, Varshney P. Optimal data fusion in multiple sensor detection systems. IEEE Trans Aero Elec Syst 1986;AES-22(1):98–101. http://dx.doi.org/10.1109/TAES.1986.310699.
[24] Ali K, Pazzani M. On the link between error correlation and error reduction in decision tree ensembles; 1995.
[25] Soria D, Garibaldi JM, Ambrogi F, Biganzoli EM, Ellis IO. A non-parametric version of the naive bayes classifier. Knowl-Based Syst 2011;24(6):775–84.
[26] Brocardo ML, Traore I, Woungang I. Toward a framework for continuous authentication using stylometry. In: 2014 IEEE 28th international conference on advanced information networking and applications (AINA). IEEE; 2014. p. 106–15.
[27] Abbasi A, Chen H. Writeprints: a stylometric approach to identity-level identification and similarity detection in cyberspace. ACM Trans Inf Syst 2008;26(2):7:1–7:29. http://dx.doi.org/10.1145/1344411.1344413. <http://doi.acm.org/10.1145/1344411.1344413>.
[28] McDonald AWE, Afroz S, Caliskan A, Stolerman A, Greenstadt R. Use fewer instances of the letter i: Toward writing style anonymization. Lecture notes in computer science, vol. 7384. Springer; 2012. p. 299–318.
[29] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The weka data mining software: an update. SIGKDD Explor Newsl 2009;11(1):10–8. http://dx.doi.org/10.1145/1656274.1656278. <http://doi.acm.org/10.1145/1656274.1656278>.
[30] Platt J. Fast training of support vector machines using sequential minimal optimization. In: Schoelkopf B, Burges C, Smola A, editors. Advances in Kernel methods – support vector learning. MIT Press; 1998. <http://research.microsoft.com/~jplatt/smo.html>.
[31] Koppel M, Schler J. Ad-hoc authorship attribution competition approach outline. In: Juola P, editor. Ad-hoc authorship attribution contest, ACH/ALLC 2004; 2004.
[32] Tenney RR, Nils J, Sandell R. Decision with distributed sensors. IEEE Trans Aero Elec Syst 1981;AES-17:501–10.
[33] Kam M, Chang W, Zhu Q. Hardware complexity of binary distributed detection systems with isolated local bayesian detectors. IEEE Trans Syst Man Cybernet 1991;21:565–71.
[34] Reibman AR, Nolte L. Optimal detection and performance of distributed sensor systems. IEEE Trans Aero Elec Syst 1987;AES-23:24–30.
[35] Jordan A. On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. Adv Neural Inform Process Syst 2002;14:841.
[36] Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng 2014;40(1):16–28. http://dx.doi.org/10.1016/j.compeleceng.2013.11.024. 40th-year commemorative issue. <http://www.sciencedirect.com/science/article/pii/S0045790613003066>.

Lex Fridman is a PhD candidate at Drexel University. His research interests include machine learning, numerical optimization, and information fusion as applied to active authentication, computer vision, and mobile ad hoc wireless networks.

Ariel Stolerman is a PhD candidate in Computer Science at Drexel University. His research interests include applications in security and privacy, applied machine learning and text analysis.

Sayandeep Acharya is a PhD student at the Data Fusion Laboratory in Drexel University. He received the MS degree in Electrical Engineering from Drexel University in 2009. His current research interests include multi-sensor data fusion, hard and soft fusion, Kalman filtering, statistical analysis, machine learning and cognitive decision making.

Patrick Brennan is the COO and co-founder of Juola & Associates, a text analysis startup. He has a decade of experience in the field of software engineering. He has spearheaded many software development projects in a wide variety of fields. He holds both a BS in Computer Science from Duquesne University and an MBA/MIS dual degree from the Katz School of Business at the University of Pittsburgh.

Patrick Juola is the CEO and co-founder of Juola & Associates, a text analysis startup. He also holds the post of professor of Computer Science at Duquesne University, in Pittsburgh, PA. His expertise includes text-based analysis and profiling for authorial information such as identity, demographic information, or psychometric attributes. He is one of the two analysts who identified J.K. Rowling as the author behind the pen name Robert Galbraith. He received his

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018

PhD in computer science in 1995 from the University of Colorado and worked for three years as a postdoc at Oxford University, before joining the Duquesne faculty. Juola & Associates was founded in 2010.

Rachel Greenstadt is an Associate Professor of Computer Science at Drexel University, where she research the privacy and security properties of intelligent systems and the economics of electronic privacy and information security. Her work is at "layer 8" of the network – analyzing the content. She is a member of the DARPA Computer Science Study Group and she runs the Privacy, Security, and Automation Laboratory (PSAL) which is a vibrant group of ten researchers. The privacy research community has recognized her scholarship with the PET Award for Outstanding Research in Privacy Enhancing Technologies, the NSF CAREER Award, and the Andreas Pfitzmann Best Student Paper Award.

Moshe Kam received his BS in electrical engineering from Tel Aviv University in 1976 and MSc and PhD from Drexel University in 1985 and 1987, respectively. He is the Dean of the Newark College of Engineering at the New Jersey Institute of Technology, and had served earlier (2007–2014) as the Robert Quinn professor and Department Head of Electrical and Computer Engineering at Drexel University. His professional interests are in system theory, detection and estimation, information assurance, robotics, navigation, and engineering education. In 2011 he served as President and CEO of IEEE, and at present he is member of the Boards of Directors of ABET and the United Engineering Foundation (UEF). He is a fellow of the IEEE "for contributions to the theory of decision fusion and distributed detection."

Please cite this article in press as: Fridman L et al. Multi-modal decision fusion for continuous authentication. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.10.018