Decision Fusion for Multi-Modal Active Authentication

Alex Fridman*, Ariel Stolerman*, Sayandeep Acharya*,
Patrick Brennan†, Patrick Juola†,
Rachel Greenstadt* and Moshe Kam*
*Drexel University
Philadelphia, PA
Email: {af59, ams573, sa427, greenie, moshe.kam}@drexel.edu
†Juola & Associates
Munhall, PA 15120
Email: {pjuola, pbrennan}@juolaassociates.com

***Abstract*—We consider a representative collection of behavioral biometrics: two low-level modalities of keystroke dynamics and mouse movement, and two high-level modalities of stylometry and web browsing behavior. To the best of our knowledge, the application of the latter two in the continuous authentication context has not been studied before. We develop a sensor for each modality and organize the sensors as a parallel binary detection decision fusion architecture. The decisions of each sensor are fed into a Decision Fusion Center (DFC) which applies the Chair-Varshney fusion algorithm to generate a global decision. We test our approach on a dataset collected from 19 users in a simulated work environment. We show that the fusion algorithm achieves lower probability of error than that of the best individual sensor in the fused set, and we are able to quantify the contribution of each modality to the overall performance.**

***Keywords*-Biometrics; authentication; fusion; stylometry.**

I. Introduction

Identity verification for the purpose of access control is the tradeoff between maximizing the probability of intruder detection, and minimizing the cost for the legitimate user in distractions and extra hardware requirements. In recent years, behavioral biometric systems have been explored extensively in addressing this challenge [1]. These systems rely on input 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].

We consider the real-time application of this technology for active authentication. As a user begins interacting with the machine, the classification system collects behavioral biometrics from the interaction and continuously verifies that the current user has access permission on the machine. This approach adds an extra layer of distraction-less access control in environments where a computer is at a risk of being intermittently accessed by unauthorized users.

We employ four classes of biometrics: keystroke dynamics, mouse movement, web browsing behavior, and stylometry. The latter two have not been considered in literature, to the best of our knowledge, in the continuous authentication context. Stylometric analysis, in particular, is well developed (and was deemed accurate enough to be admissible as legal evidence in some courts [3]). However, its application to continuous verification of user identity is novel. The basic assumption behind stylometry is that every person has a unique linguistic style ("stylome" [4]) that can be quantified and measured in order to distinguish between different authors. Based on the success of authorship attribution in other fields, we seek to characterize its performance in this much more dynamic and time-constrained problem space.

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 and web browsing sensors will be actively flooded with data, while the keystroke dynamics and stylometry sensors may only get a few infrequent updates. This observation motivates the recent work on multimodal authentication systems where the decisions of multiple classifiers are fused together [5]. Our approach in this paper is to apply the Chair-Varshney decision fusion rule [6] for the combination of available multimodal decisions. Furthermore, we are motivated by the work in [7] that greater reduction in error rates is achieved when the classifiers are distinctly different (i.e., when using different behavioral biometrics).

II. Biometric Sensors

The sensors we consider in this paper span across different levels and directions for profiling: linguistic style (stylometry), mouse movement patterns, keystroke dynamics and web browsing behavior. Each of these types of sensory input has different required volume of input data, nature of the collected data (mouse events, keystrokes, different usage statistics) and performance.

Following the commonly used classification of biometrics we refer here to the mouse and keystroke dynamics sensors as "low-level" and to the website domain frequency and

stylometry sensors as "high-level". Specifically, we collected and used the following:

- Low-level sensors:
  - M1: mouse curvature angle
  - M2: mouse curvature distance
  - M3: mouse direction
  - K1: keystroke interval time
  - K2: keystroke dwell time
- High-level sensors:
  - W1: website domain visit frequency
  - S1: stylometry (1000 char., 30 min. window)
  - S2: stylometry (500 char., 30 min. window)
  - S3: stylometry (400 char., 10 min. window)
  - S4: stylometry (100 char., 10 min. window)

A. Simulated Work Environment Dataset

We collected behavioral biometrics data in a simulated work environment. Specifically, we put together an office space, organized and supervised by a subset of the authors. During each of the four weeks of the data collection we hired 5 temporary employees for 40 hours of work each. Each day the employees were assigned various reading, writing and browsing tasks. Data files on their interaction with the mouse and the keyboard were produced by two tracking applications. For 19 users included in this study we collected close to 1.2 million keystroke events and close to 10 million "mouse move" events.

B. Stylometry

Authorship attribution based on linguistic style, or Stylometry, is a well-researched field [8]. The main domain it is applied on is written language – identifying an anonymous author of a text by mining it for linguistic features. The feature space is potentially boundless, with frequency measurements or numeric evaluations based on features across different levels of the text, including function words, grammar, character n-grams and more.

The feature set we used, denoted the *AA* feature set hereinafter, is a variation of the *Writeprints* [9] feature set, which includes a vast range of linguistic features across different levels of text. This rich linguistic feature set is aimed at capturing 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 classification we used sequential minimal optimization (SMO) support vector machines with polynomial kernel, available in Weka [10]. Support vector machines are commonly used for authorship attribution [11] and documented to achieve high performance and accuracy.

C. Low-Level Metrics

Keystroke dynamics is one of the most extensively studied topics in behavioral biometrics [12]. The feature space that has been investigated ranges from the simple metrics of key press interval [13] and dwell [14] times to multi-key features such as trigraph duration with an allowance for typing errors [2]. Mouse movement dynamics has also recently received considerable attention [15].

The low-level metrics of keystroke and mouse dynamics detectors, along with the domain visit frequency detector, all use support vector machines (SVMs). Here we considered three metrics based on those described in [15]: (M1) curvature angle, (M2) curvature distance, and (M3) movement direction. For keyboard dynamics, we chose two of the most commonly used keystroke dynamics features: (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 release.

D. Web Browsing Behavior

Web browsing behavior has been studied extensively in literature [16] but not in the context of active authentication. We used the same SVM classifier as for low-level sensors, and the feature vector of the visit frequency to the 20 most visited websites in the dataset, the top five of which were: google.com (7.0%), bing.com (7.0%), facebook.com (5.0%), yahoo.com (4.1%), and wikipedia.org (2.9%). The visit frequency of any one of these popular websites is not a good classification feature. However, taken together, the 20-dimensional feature vector forms a sufficiently representative profile of a user to be used in continuous authentication.

III. 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 joint assessment of the environment, which a single sensor may not be able to provide. Decision fusion with distributed sensors is described by Tenney and Sandell in [17] who studied several parallel decision architectures. As described in [18], 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 $\{u_1, u_2, ..., u_n\}$ for a global decision about the hypothesis. The i^{th} 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 make the global decision (D_0 or D_1).

Chair and Varshney in [6] developed an optimal fusion rule for a parallel binary detector architecture with respect to a Bayesian cost (here we use the probability of error as the cost). They assumed that the local detectors are pre-designed and fixed (with known probability of detection and probability of false alarm) and that local observations are statistically independent conditioned on the hypothesis.

Moreover, it was assumed that the *a priori* probabilities $P_0 = P(H_0)$ and $P_1 = P(H_1) = 1 - P(H_0)$ were known. Using its own rule, the local sensor detector collects data from its environment and decides on D_0 ($u_i = -1$) or D_1 ($u_i = 1$). A decision fusion center combines these local decisions using the rule

$$\frac{P(u_1, ..., u_n|H_1)}{P(u_1, ..., u_n|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1} = \tau \tag{1}$$

where the a priori probabilities of the binary hypotheses H_1 and H_0 are P_1 and P_0 respectively. Rule (1) can be shown to be equivalent to

$$f(u_1, ..., u_n) = \begin{cases} 1, & \text{if } a_0 + \sum_{i=0}^{n} a_i u_i > 0 \\ -1, & \text{otherwise} \end{cases} \tag{2}$$

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

$$a_0 = \log \frac{P_1}{P_0} \tag{3}$$

$$a_i = \begin{cases} \log \frac{1-P_i^M}{P_i^F}, & \text{if } u_i = 1 \\ \log \frac{1-P_i^F}{P_i^M}, & \text{if } u_i = -1 \end{cases} \tag{4}$$

Kam et al. in [18] developed expressions for the global performance of the distributed system described above.

The rows of the table in Fig. 1 are four representative combinations of 10 sensors listed in §II and the FAR/FRR rates that result when these sensors are fused. A checkmark in this table designates which of the sensors is included in the fusion for that row. There are 1024 possible combinations. We selected these four to highlight the marginal contribution of stylometry and web browsing modalities when fused with the low level modalities. The plots in Fig. 1 indicate that stylometry contributes more to reducing the error rates than web browsing.

IV. Conclusion

We applied a parallel binary decision fusion architecture on a representative collection of behavioral biometric sensors using keystroke dynamics, mouse movement, stylometry, and web browsing behavior. Using this approach and a dataset collected from 19 individuals in an office environment we addressed the challenge of active authentication and characterized the authentication performance of the sensor suite. The global decision is of better quality (i.e., lower probability of error) than that of the best sensor operating by itself. We are also able to characterize the marginal contribution of each modality to the overall FAR/FRR performance. Future work will be geared toward open world authentication on a larger data set with a more expansive portfolio of metrics.

W1	S1	S2	S3	S4	M1	M2	M3	K1	K2	FAR	FRR
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.00122	0.00218
	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.0021	0.0027
✓					✓	✓	✓	✓	✓	0.00728	0.00714
					✓	✓	✓	✓	✓	0.0102	0.01016

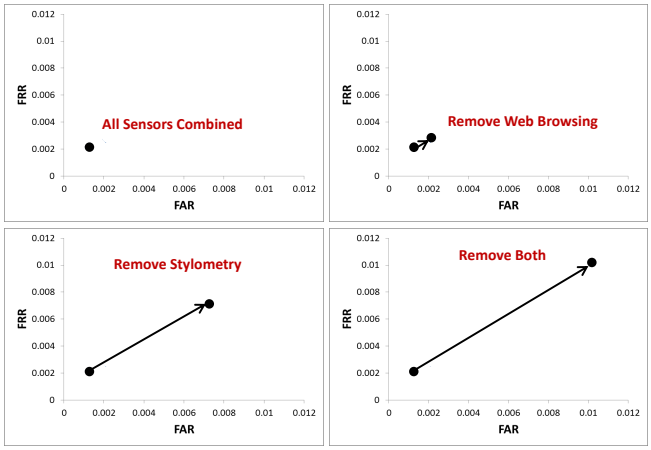

Figure 1: FAR and FRR rates for 4 representative selection of sensors of the 1024 possible combinations for fusion. These four cases are: (1) all sensors are used, (2) all sensors except for web browsing is used; (3) all sensors except for stylometric sensors are used; and (4) all sensors except for web browsing and stylometric sensors are used. (See §II for glossary).

References

[1] A. Ahmed and I. Traore, "A new biometric technology based on mouse dynamics," *IEEE Transactions on Dependable and Secure Computing*, vol. 4, no. 3, pp. 165–179, July 2007.

[2] F. Bergadano, D. Gunetti, and C. Picardi, "User authentication through keystroke dynamics," *ACM Transactions on Information System Security*, vol. 5, no. 4, pp. 367–397, November 2002.

[3] C. E. Chaski, "The keyboard dilemma and forensic authorship attribution," Advances in Digital Forensics III, Boston, pp. 133–148, 2007.

[4] H. van Halteren, R. H. Baayen, F. Tweedie, M. Haverkort, and A. Neijt, "New machine learning methods demonstrate the existence of a human stylome," *Journal of Quantitative Linguistics*, vol. 12, no. 1, pp. 65–77, 2005.

[5] T. Sim, S. Zhang, R. Janakiraman, and S. Kumar, "Continuous verification using multimodal biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 687–700, 2007.

[6] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 1, pp. 98–101, January 1986.

[7] K. Ali and M. Pazzani, *On the link between error correlation and error reduction in decision tree ensembles*. Citeseer, 1995.

[8] E. Stamatatos, "A survey of modern authorship attribution methods," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 3, pp. 538–56, 2009.

[9] A. Abbasi and H. Chen, "Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace," *ACM Trans. Inf. Syst.*, vol. 26, no. 2, pp. 7:1–7:29, April 2008.

[10] J. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods - Support Vector Learning*, B. Schoelkopf, C. Burges, and A. Smola, Eds. MIT Press, 1998.

[11] A. Abbasi and H. Chen, "Identification and comparison of extremist-group web forum messages using authorship analysis," *IEEE Intelligent Systems*, vol. 20, no. 5, pp. 67–75, 2005.

[12] M. Karnan, M. Akila, and N. Krishnaraj, "Biometric personal authentication using keystroke dynamics: A review," *Applied Soft Computing*, vol. 11, no. 2, pp. 1565–1573, 2011.

[13] N. Bartlow and B. Cukic, "Evaluating the reliability of credential hardening through keystroke dynamics," in *International Symposium on Software Reliability Engineering*. IEEE, 2006, pp. 117–126.

[14] R. Giot, M. El-Abed, and C. Rosenberger, "Keystroke dynamics authentication for collaborative systems," in *International Symposium on Collaborative Technologies and Systems*. IEEE, 2009, pp. 172–179.

[15] N. Zheng, A. Paloski, and H. Wang, "An efficient user verification system via mouse movements," in *Proceedings of the 18th ACM conference on Computer and communications security*, ser. CCS '11. New York, NY, USA: ACM, 2011, pp. 139–150.

[16] R. Yampolskiy, "Behavioral modeling: an overview," *American Journal of Applied Sciences*, vol. 5, no. 5, pp. 496–503, 2008.

[17] R. R. Tenney and J. Nils R. Sandell, "Decision with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-17, pp. 501–510, 1981.

[18] M. Kam, W. Chang, and Q. Zhu, "Hardware complexity of binary distributed detection systems with isolated local bayesian detectors," *IEEE Transactions on Systems Man and Cybernetics*, vol. 21, pp. 565–571, 1991.