Decision Fusion for Multimodal Active Authentication

Lex Fridman, Ariel Stolerman, and Sayandeep Acharya, *Drexel University*
Patrick Brennan and Patrick Juola, *Juola & Associates*
Rachel Greenstadt and Moshe Kam, *Drexel University*

The authors apply a decision fusion architecture on a collection of behavioral biometric sensors using keystroke dynamics, mouse movement, stylometry, and Web browsing behavior. They test this active authentication approach on a dataset collected from 19 individuals in an office environment.

Identity verification for access control presents a trade-off between maximizing the probability of intruder detection and minimizing the cost for the legitimate user in terms of distractions and hardware requirements. In recent years, researchers have extensively explored behavioral biometric systems to address this challenge.[1] These systems rely on input devices, such as the keyboard and mouse, that are already commonly available with most computers. 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, stylometry, and Web browsing. Depending on the task in which the user is engaged, some of the biometric sensors might provide more data than others. For example, as the user browses the Web, the mouse and Web browsing sensors will be

1520-9202/13/$31.00 © 2013 IEEE Published by the IEEE Computer Society

actively flooded with data, while the keystroke dynamics and stylometry sensors might only get a few infrequent updates. This observation motivates the recent work on multimodal authentication systems, which fuses together decisions from multiple classifiers.[3] Our approach is to apply the Chair-Varshney decision-fusion rule[4] to combine available multimodal decisions. Furthermore, we are motivated by Kamal Ali and Michael Pazzani's work,[5] which shows that using distinctly different classifiers (that is, different behavioral biometrics) helps reduce error rates.

Biometric Sensors

The sensors we consider here span different levels and directions for profiling: linguistic style (stylometry), mouse movement patterns, keystroke dynamics, and Web browsing behavior. Each type of sensory input has a different requirement in terms of the volume of input data, nature of the collected data (mouse events, keystrokes, and 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." The low-level sensors we used were

- M1: the mouse curvature angle,
- M2: the mouse curvature distance,
- M3: the mouse direction,
- K1: the keystroke interval time, and
- K2: the keystroke dwell time.

For the high-level sensors, we used

- W1: the website domain visit frequency,
- S1: stylometry with 1,000 characters and a 30-minute window,
- S2: stylometry with 500 characters and a 30-minute window,
- S3: stylometry with 400 characters and a 10-minute window, and
- S4: stylometry with 100 characters and a 10-minute window.

We collected the behavioral biometrics data in a simulated work environment. Specifically, we put together an office space. During each of the four weeks of data collection, we hired five temporary employees, each of whom worked 40 hours. 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 the 19 users included in this study, we collected close to 1.2 million keystroke events and 10 million "mouse move" events.

Low-Level Metrics

Keystroke dynamics have been extensively studied in behavioral biometrics,[6] ranging from the simple metrics of key press interval[7] and dwell times[8] to multikey features, such as trigraph duration with an allowance for typing errors.[2] Mouse movement dynamics have also recently received considerable attention.[9]

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: the curvature angle (M1), curvature distance (M2), and movement direction (M3).[9] For keyboard dynamics, we chose two of the most commonly used keystroke dynamics features: the interval between the release of one key and the press of another (K1) and the dwell time between the press of a key and its release (K2).

Stylometry

Authorship attribution based on linguistic style, or stylometry, is a well-researched field.[10] Typically, stylometry is applied to written language to identify an anonymous author by mining the text 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, and character n-grams.

The feature set we used (denoted the "AA" feature set), is a variation of the Writeprints feature set,[11] 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) SVMs with polynomial kernel, available in WEKA (the Waikato Environment for Knowledge Analysis).[12] SVMs are commonly used for authorship attribution[13] and documented to achieve high performance and accuracy.

Web Browsing Behavior

The research literature also includes many studies of Web browsing behavior,[14] 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 were google.com (7 percent), bing.com (7 percent), facebook.com (5 percent), yahoo.com (4.1 percent), and wikipedia.org (2.9 percent). The visit frequency of any one of these popular websites isn't a good classification feature. However, taken together, the 20-dimensional feature vector forms a sufficiently representative profile of a user for continuous authentication.

Decision Fusion

The motivation for using multiple sensors to detect an event is to harness the sensors' power to provide an accurate joint assessment of the environment, which a single sensor might not be able to provide. Robert Tenney and Nils Sandell have described decision fusion with distributed sensors,[15] studying several parallel decision architectures. Furthermore, Moshe Kam, Wei Chang, and Qiang Zhu have described a distributed binary detection system that comprises 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.[16] 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 to make the global decision (D_0 or D_1).

Z. Chair and P.K. Varshney developed an optimal fusion rule for a parallel binary detector architecture with respect to a Bayesian cost[4] (here we use the probability of error as the cost). They assumed that the local detectors were predesigned and fixed (with known probability of detection and probability of false alarm) and that local observations were 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 DFC combines these local decisions using the rule

$$\frac{P(u_1,\ldots,u_n \mid H_1)}{P(u_1,\ldots,u_n \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1} = \tau$$

where the a priori probabilities of the binary hypotheses H_1 and H_0 are P_1 and P_0, respectively. This can be shown to be equivalent to

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

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

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

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

Kam and his colleagues developed expressions for the global performance of the distributed system just described.[16]

Figure 1a shows the four representative combinations of 10 low- and high-level sensors described earlier and the FAR and FRR rates resulting from fusing these sensors. A checkmark designates which of the sensors is included in the fusion for that row. There are 1,024 possible combinations. We selected these

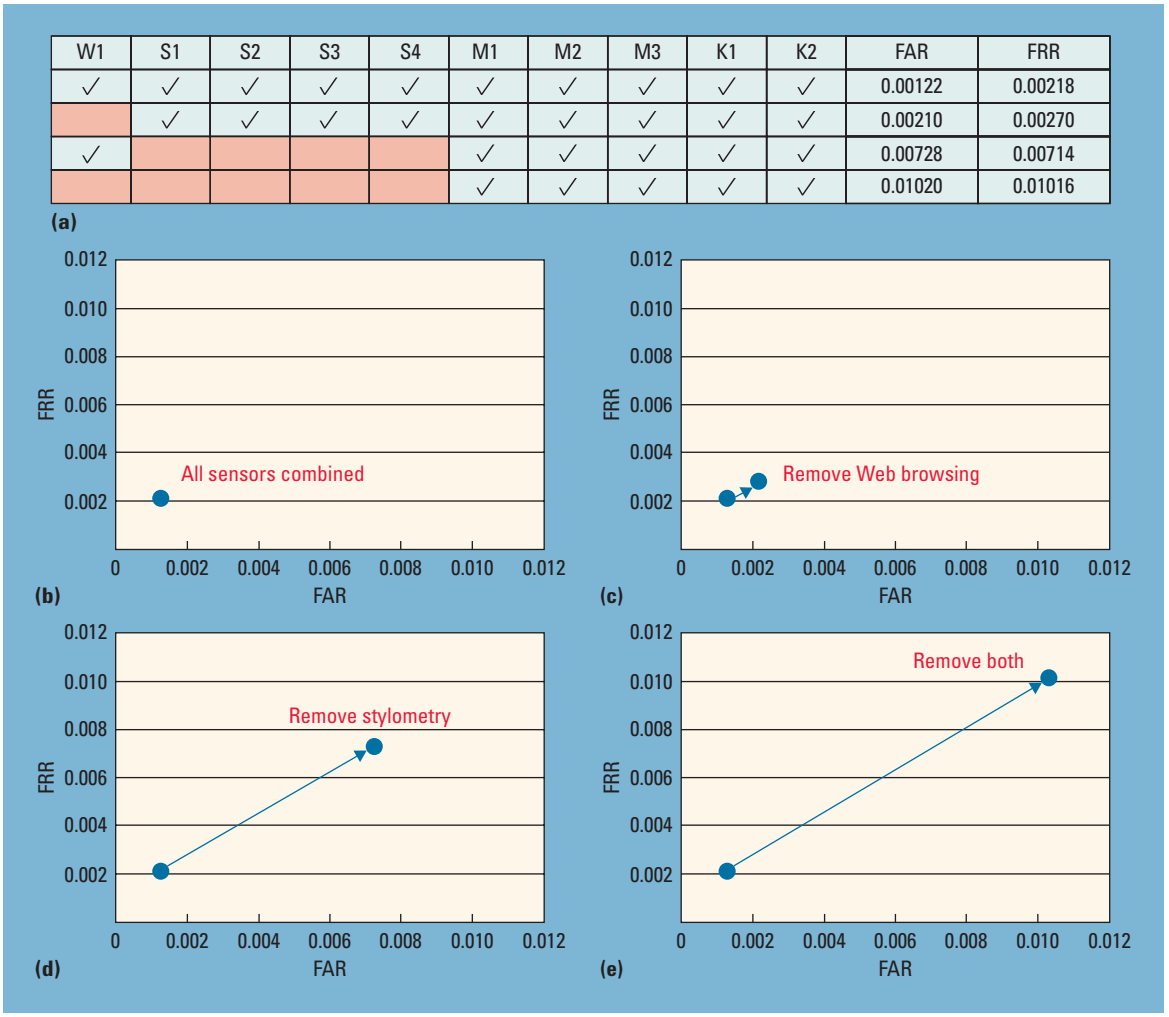

Figure 1. Four representative combinations of the 10 sensors used: (a) FAR and FRR rates for four representative selections of sensors of the 1,024 possible combinations for fusion. The four cases are (b) all sensors are used, (c) all sensors are used except for Web browsing, (d) all sensors are used except for the stylometric sensors, and (e) all sensors are used except for the Web browsing and stylometric sensors.

four to highlight the marginal contribution of stylometry and Web browsing modalities when fused with the low level modalities. The plots Figures 1b–1e indicate that stylometry contributes more to reducing the error rates than Web browsing.

In attempting to address the challenge of active authentication, we learned that the global decision has a lower probability of error than that of the best sensor operating by itself. Future work will be geared toward open world authentication on a larger data set with a more expansive portfolio of metrics.

Acknowledgements

This material is based on work supported by DARPA under BAA-12-06.

References

1. A. Ahmed and I. Traore, "A New Biometric Technology Based On Mouse Dynamics," *IEEE Trans. Dependable and Secure Computing*, vol. 4, no. 3, 2007, pp. 165–179.
2. F. Bergadano, D. Gunetti, and C. Picardi, "User Authentication through Keystroke Dynamics," *ACM Trans. Information System Security*, vol. 5, no. 4, 2002, pp. 367–397.
3. T. Sim et al., "Continuous Verification Using Multimodal Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, 2007, pp. 687–700.

4. Z. Chair and P. Varshney, "Optimal Data Fusion in Multiple Sensor Detection Systems," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-22, no. 1, 1986, pp. 98–101.
5. K. Ali and M. Pazzani, *On the Link between Error Correlation and Error Reduction in Decision Tree Ensembles*, tech. report, Information and Computer Science, Univ. of California, Irvine, 1995.
6. M. Karnan, M. Akila, and N. Krishnaraj, "Biometric Personal Authentication Using Keystroke Dynamics: A Review," *Applied Soft Computing*, vol. 11, no. 2, 2011, pp. 1565–1573.
7. N. Bartlow and B. Cukic, "Evaluating the Reliability of Credential Hardening through Keystroke Dynamics," *Proc. Int'l Symp. Software Reliability Engineering*, IEEE, 2006, pp. 117–126.
8. R. Giot, M. El-Abed, and C. Rosenberger, "Keystroke Dynamics Authentication for Collaborative Systems," *Proc. Int'l Symp. Collaborative Technologies and Systems*, IEEE, 2009, pp. 172–179.
9. N. Zheng, A. Paloski, and H. Wang, "An Efficient User Verification System via Mouse Movements," *Proc. 18th ACM Conf. Computer and Communications Security* (CCS 11), ACM, 2011, pp. 139–150.
10. E. Stamatatos, "A Survey of Modern Authorship Attribution Methods," *J. Amer. Society for Information Science and Technology*, vol. 60, no. 3, 2009, pp. 538–56.
11. A. Abbasi and H. Chen, "Writeprints: A Stylometric Approach to Identity-Level Identification and Similarity Detection in Cyberspace," *ACM Trans. Information Systems*, vol. 26, no. 2, 2008, pp. 7:1–7:29.
12. J. Platt, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization," *Advances in Kernel Methods—Support Vector Learning*, MIT Press, 1998, pp. 185–208.
13. A. Abbasi and H. Chen, "Identification and Comparison of Extremist-Group Web Forum Messages Using Authorship Analysis," *IEEE Intelligent Systems*, vol. 20, no. 5, 2005, pp. 67–75.
14. R. Yampolskiy, "Behavioral Modeling: An Overview," *American J. Applied Sciences*, vol. 5, no. 5, 2008, pp. 496–503.
15. R.R. Tenney and Nils R. Sandell, "Decision with Distributed Sensors," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-17, 1981, pp. 501–510.
16. M. Kam, W. Chang, and Q. Zhu, "Hardware Complexity of Binary Distributed Detection Systems with Isolated Local Bayesian Detectors," *IEEE Trans. Systems Man and Cybernetics*, vol. 21, no. 3, 1991, pp. 565–571.

Lex Fridman *is a PhD candidate at the Data Fusion Laboratory in the Department of Electrical and Computer Engineering at Drexel University. Contact him at af59@drexel.edu.*

Ariel Stolerman *is a PhD student and research assistant at the Privacy, Security and Automation Laboratory in the Department of Computer Science at Drexel University. Contact him at ams573@cs.drexel.edu.*

Sayandeep Acharya *is a doctoral candidate in the Electrical and Computer Engineering Department at Drexel University. Contact him at sa427@drexel.edu.*

Patrick Brennan *is the president of Juola & Associates. Contact him at pbrennan@jgaap.com.*

Patrick Juola *is the director of research, CEO, and a founder of Juola & Associates, a text analysis firm. He is also an associate professor of computer science at Duquesne University. Contact him at pjuola@juolaassociates.com.*

Rachel Greenstadt *is an assistant professor of computer science at Drexel University. Contact her at rachel.a.greenstadt@drexel.edu.*

Moshe Kam *is the department head and Robert Quinn Professor of Electrical and Computer Engineering at Drexel University. Contact him at kam@drexel.edu.*