Metrics and Adversary Models for Implicit Authentication

Submitted by:
Daniel Calderon, Preethi Josephina Mudialba, Siddharth Nair

Advisor: Lujo Bauer
Client: UnifyID

2017

Executive Summary

Problem Statement

Authentication is quietly undergoing a revolution moving beyond passwords. The increasing maturity of recognition technology has encouraged the use and deployment of more usable, secure authentication systems based on biometrics that can provide a complete improvement over the existing paradigm[2, Whither Biometrics,2010]. In particular, one particular strand of authentication research called implicit authentication holds great promise for disrupting the field by greatly expanding system usability while ensuring system security [1, Shi et al, 2010]. Implicit Authentication passively analyzes users' behavioural biometrics to continuously and transparently authenticate users without the requirement of user action [1, Shi et al, 2010]. This authentication technique relies on the observation that human beings in general are creatures of habit, and more or less have a fixed routine that they follow in their daily activity. Additionally, Implicit Authentication is greatly assisted by the increasing ubiquity of wearables and mobile devices with a multitude of sensors that can easily accumulate a user's related routine data (such as location, motion, usage of application, etc.) to create unique user profiles.

Many systems have been proposed for achieving the goals of this field, but it remains unclear how to evaluate across systems since the field lacks an agreed-upon set of performance evaluation metrics. To compound this problem further, not all systems necessarily consider the same, if any, adversarial threats to their system that could compromise the security or usability of the system. In this report, we reviewed the literature on performance evaluation, and reviewed the broader computer security authentication literature, and determined a set of important criteria that a metric should have to be valuable for evaluating an implicit authentication system. We also reviewed the authentication literature to construct a comprehensive threat model. We propose a recommended suite of metrics, and a recommended threat framework, seeking to motivate the research communities to adopt these recommendations in order to improve the comparability of research results.

Criteria for Metrics

Through this research, we classified metrics by their relevance to three components of the authentication system: the enrollment phase, the authentication phase, and the overall usability of the system. We developed a set of 7 criteria for evaluating metrics that we determined were necessary for holistic evaluation of system performance, and also considered whether the metrics was popular in the recent literature by documenting the evolution of

selection of metrics by system designers over time, biasing towards more popular and more recent techniques [See Figure 1 in the section of Evolution of Metrics]. We quickly determined that no one metric satisfied all seven, so instead, we propose the adoption of a suite of metrics that together span all seven criteria. These seven criteria are defined in Table 1.

Criteria	Description
Performance On Subsets	Is the metric well-suited to distinguish performance on subsets/populations of the dataset (e.g. males vs females)?
Neyman Pearson Applicability	Can an alternative form of classifier optimization be used, which is one inspired by the Neyman-Pearson Lemma that minimizes for one type of error while setting a tolerance for the other type of error [8]?
Computational Complexity	Is the metric simple to compute, or does it requires O(n) or more additional computations beyond one Authentication event?
Positive- Negative Sensitivity	Does the metric make a distinction between Type I and Type II errors (i.e. False Positives and False Negatives), and can the performance effect of each error type be teased apart?
Class Skew Insensitivity	Is the metric insensitive (as in, unaffected by) to the distribution of true positive instances and true negative instances (i.e. the balance of the dataset)?
Worst Case Performance	Can the metric be used to identify per-class misclassifications, and thereby communicate the worst-possible misclassification by a targeted adversary (i.e the metric takes into account extreme case scenarios, where an adversary is particularly bad)?
Multi-Class Generalizability	Can the metric be used to evaluate when there are more than just two classes being considered?

Table 1: The Seven Evaluation Criteria for Performance Metrics

Evolution of Metrics

We analyzed the existing biometric literature, selecting literature based on relevancy, number of citations and time. Figure 1 contains the counts of the top 4 metrics found in literature over a period ranging from 2005 – 2017, after analysing 15 papers per year.

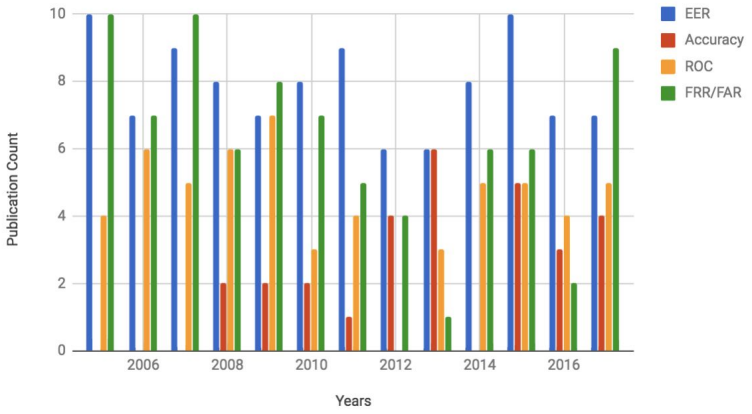

Figure 1: Evolution of Performance Metric Selection by IA System Designers Across Time

Recommendations for Metrics

Based on their superior performance across the criteria and their contribution to spanning all of the criteria, we select the metrics listed in Table 2 for each of the metric categories.

Enrollment Metrics	Authentication Metrics	Usability Metrics
Failure to Enroll	Receiver Operating Characteristic	Mean Time to Enroll
Failure to Detect	Matthews Correlation Coefficient	Mean Time to Detect
Failure to Capture	Confusion Matrix	
	False Match Rate/ False Non-Match Rate	

Table 2: Recommended Suite of Metrics

Combined, these 9 metrics check all of our evaluation criteria, provide a holistic idea of the performance of the system across phases, and could be used as a benchmark standard for comparison across future biometric authentication systems.

Recommended Threat Model Framework

We considered a number of frameworks that could be relevant for an implicit authentication system, and ultimately recommend a framework strongly influenced by the 2013 60839-11-1 European CENELEC Standard for Authentication Systems, but incorporating attacker models drawn from the Implicit Authentication literature[6, CENELEC, 2013]. We recommend that authentication systems consider an adversary that has physical access to the device and is looking to steal data from the device that can be accessed through correct authentication, similarly to the environment considered by Lee et al. [7, Lee, 2016]. The CENELEC 60839-11-1 report included a general risk-based framework for the adversaries to an authentication system with multiple grades

that describe level of security provided by the system (grade 1 being the lowest security for low-risk settings protecting "low value assets," and grade 4 being the highest security level for high-risk settings protecting "high value assets")[6, CENELEC, 2013]. For each grade, an adversary for which the grade was expected to protect against was described that incorporated the amount of information the attacker had about the system, how many resources the attacker, and the risk level associated with the attack based on the value of the resources the authentication system was protecting. Table 3 presents a simplified and modified form of this framework that has been tailored for the implicit authentication setting and for the aforementioned attacker with physical access.

Grade	1	2	3	4
Risk Level	Low	Low to Medium	Medium to High	High
Example Contexts	General-Purpose Accounts	General e-commerce, e-mail	Priority/ Primary/ Corporate Email, accounts with financial information, SSO Portals	highly-sensitive, valuable facilities (military, corporate R&D, critical infrastructure, etc.)
Adversary Skill Level	Low Information, Low Resources	Medium Information, Low to Medium Resources	Medium-High Information, Medium Resources	High Information, High Resources
Example Attacks	Brute-Force Attacks, Low-Resource Social Engineering Attacks	Denial of Service Attacks, Black Box Attacks on ML systems[3]	Replay Attacks, Grey-Box Attacks on ML Systems[4], low-resource mimicry attacks, Biometric dB attacks	Mimicry attacks, Biometric dB attacks, White-Box Attacks on ML systems[5]

Table 3: Proposed Risk-based Threat Framework, derived from CENELEC 60839-11-1 standard, tailored to attackers of Implicit Authentication systems [6, CENELEC, 2013]

Conclusion

Through our analysis, we have identified a notable omission in the literature so far regarding the consideration of the performance metrics used to analyze implicit authentication systems, and provided recommendations to fill in this gap. Going forward, we aim to encourage the community to consider the use of our recommended set of metrics, given that it is important the subset spans all 7 criteria we have noted it is essential be captured to properly evaluate an authentication system. We also encourage the community to consider the threat modelling framework for considering the capabilities of adversaries, which may assist with complying with standards for real-world deployments.

References

1. Elaine Shi, Yuan Niu, Markus Jakobsson, and Richard Chow, “Implicit Authentication Through Learning User Behavior,” International Conference on Information Security, Springer Berlin Heidelberg, 2010.
2. Whither Biometrics Committee, “Biometric Recognition: Challenges and Opportunities,” National Academies Press, 2010.
3. Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, “Explaining and Harnessing Adversarial Examples,” arXiv preprint arXiv:1412.6572, 2014.
4. Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z. Berkay Celik, and Ananthram Swami, “Practical Black-Box Attacks Against Deep Learning Systems Using Adversarial Examples,” arXiv preprint arXiv:1602.02697 2016
5. Bita Darvish Rouhani, Mohammad Samragh, Tara Javidi, and Farinaz Koushanfar, “CuRTAIL: Characterizing and Thwarting Adversarial Deep Learning,” arXiv preprint arXiv:1709.02538, 2017.
6. CENELEC. Alarm and electronic security systems -Part 11-1: Electronic access control systems -System and components requirements (IEC 60839-11-1:2013),2013.
7. Lee, Wei-Han, and Ruby Lee. "Implicit Sensor-based Authentication of Smartphone Users with Smartwatch." In *Proceedings of the Hardware and Architectural Support for Security and Privacy 2016*, p. 9. ACM, 2016.
8. Clayton Scott and Robert Nowak, “A Neyman-Pearson Approach to Statistical Learning,” IEEE Transactions on Information Theory 51, no. 11: 3806-3819, 2005.