Poster: The Light Will Be with You. Always – A Novel Continuous Mobile Authentication with the Light Sensor

Mohsen A. Alawami
Sungkyunkwan University
mohsencomm@skku.edu

William Aiken
Sungkyunkwan University
billzo@skku.edu

Hyoungshick Kim
Sungkyunkwan University
hyoung@skku.edu

1 INTRODUCTION

Existing continuous authentication proposals tend to have two major drawbacks. First, touch-based smartphone authentication approaches [1, 2] typically require explicit user interactions with the smartphone to collect sufficient touch data. These approaches may provide an attacker the opportunity to steal a victim's sensitive data before the system detects the attacker's intrusion. Likewise, an attacker may disable the continuous authentication scheme itself before detection. Second, sensor-based continuous authentication approaches [3, 4] inherently suffer from high energy consumption due to the constant usage of multiple sensors. In this paper, we present a novel continuous authentication system that collects light sensor data from a user's smartphone and analyzes them to authenticate users using support vector machines. We focus on the possibility of collecting light sensor data from users' smartphones while they are conducting daily behaviors to develop an anomaly detection system.

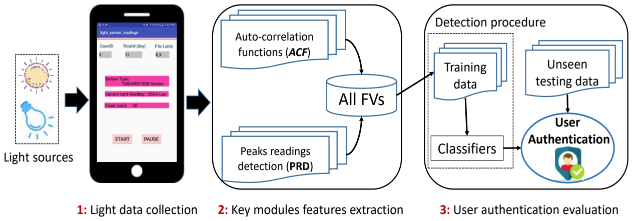


Figure 1: User authentication system design.

Figure 1 shows our system's three major steps: (1) the process of collecting light data via our Android application by utilizing the device's default light sensors, (2) the analysis of the extracted features using two different light modules, and (3) the detection procedure to evaluate the system using machine learning algorithms.

2 KEY MODULES FEATURES

We leverage two key feature vectors (FVs) extracted from auto-correlation function (ACF) and peak readings detection (PRD) modules. The auto-correlation function (ACF) is used to measure how dependent and periodic the light measurements are over time. Measuring the time dependency of the measured light readings vector and the shifted (lagged) copies as a function of the lag is a mathematical process that produces auto-correlation coefficients scaled between -1 and +1. Coefficients near zero indicate nearly all of the intensity observations are mostly random while coefficients close to -1 or +1 indicate nearly all observations are repeatable. For each user behavior, the real discrete light readings vector ($R(t) \in \mathbb{R}^{m \times 1}$), where m denotes vector length, is used to calculate the auto-correlation function (ACF) as follows:

$$ACF(k) = \frac{Cov(R(t), R(t-k))}{Var(R(t))}, \tag{1}$$

$$Cov(R(t), R(t-k)) = \frac{1}{m-1} \sum_{t=k+1}^{m} (R(t) - \bar{R})(R(t-k) - \bar{R}), \tag{2}$$

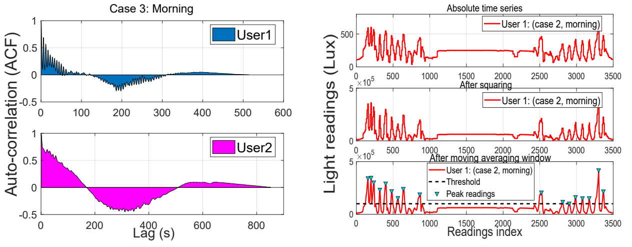
$$Var(R(t)) = \frac{1}{m-1} \sum_{t=1}^{m} (R(t) - \bar{R})^2. \tag{3}$$

The parameter $ACF(k)$ denotes the normalized auto-correlation as a function of the k^{th} lag value, for ($k = 1, 2, 3, \dots$). If we have a sample vector $R(t)$, ($t = 1, 2, 3, \dots, m$), of the collected light readings, then the k^{th} order auto-correlation $ACF(k)$ can be estimated using Eq. 1 by first calculating the covariance and variance variables from Eq. 2 and Eq. 3 respectively. $\bar{R}$ is the mean of the $R(t)$ vector and Eq. 3 is the special case of Eq. 2 in cases where $k = 0$. To show the effectiveness of the ACF module in distinguishing user behaviors, we showcase the results of *case*3 behavior (*put the phone on the user's desk, wait for 1 minute, stand up and move to left lounge, sit down and put the phone on the desk for 1 minute, pick up the phone and move back to the user's seat, put the phone on the user's desk for 1 minute*) as an illustrative example. We plotted the ACF output signatures in Figure 2a for the morning time profile where the patterns of both users have large differences in the normalized ACF values such that they can be accurately distinguished from one another. Ultimately, ACF could be used as a convenient mathematical process to distinguish user behavior signatures to model light intensity data for user authentication.

Peak readings detection (PRD) is employed as a means to focus solely on the peaks of light waveforms that aid in distinguishing subjects from one another as they move under or near light sources in indoor environments. The PRD module contains three main steps: (1) squaring, (2) smoothing, and (3) peak-finding. Together these steps perform the operation of extracting peak values from the absolute intensities. Squaring the absolute light readings vector V_{Light} obtains $V_{Sqr} = (V_{Light})^2$, which enhances large peak values more than noisy values. To achieve smoothing, we used the *MovingAverage* algorithm as a low-pass filter to smooth out

MobiSys '19, June 17–21, 2019, Seoul, Republic of Korea

ACM ISBN 978-1-4503-6661-8/19/06.
https://doi.org/10.1145/3307334.3328625



(a)

(b)

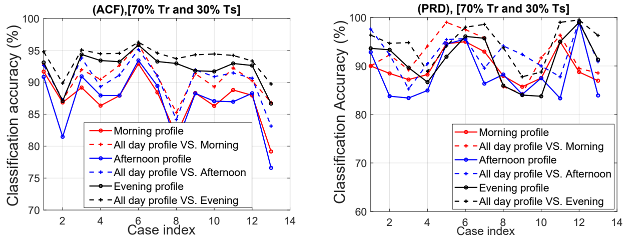
Figure 2: Illustration of the light intensity signatures for both users using: (a) ACF; (b) PRD.

noisy readings and short-term fluctuations to return a smoothed data vector V_{MAV} by computing the average over W samples of light readings. Finally, the function ($FindPeaks$) takes the smoothed vector V_{MAV} as an input, finds peaks that are greater than a predefined threshold ($P_{thr} \geq 0$), and outputs the (V_{PRD}) vector which denotes the peak values. Figure 2b show graphically the steps of implementing the PRD module and the ability to detect peaks on sample *case*2 morning behavior (*start from the elevator, move to the user's seat and put the phone on the user's desk, wait for 1 minute, move back to the elevator*) conducted by user 1.

3 PERFORMANCE EVALUATION

We evaluated our system on a real-world dataset representing 13 diverse behaviors collected by two users inside our university building over the course of 20 days at three different time profiles as follows: [Morning: (9:00 ∼ 11:00), Afternoon: (13:00 ∼ 15:00), Evening: (17:00 ∼ 19:00)]. We chose the third floor as our testing environment to collect light measurements in specific location points including the users' work cubicles, the restroom, the third floor elevator, and public lounges. The participants conducted the designated 13 case behaviors in specific postures (e.g. sitting, standing), phone placements and/or orientations (pocket, hand, desk), and phone facing directions (front, back) using our Android application installed on the same smartphone (Galaxy S9 and Android version 8.0.0).

The evaluation process was conducted in two main scenarios: the first scenario consisted of an evaluation of the three individual profiles (Morning, Afternoon, Evening) separately, and the second scenario consisted of an evaluation of the Allday profile versus each of the three individual profiles (Allday vs Morning, Allday vs Afternoon, Allday vs Evening). We individually evaluated the feature vectors (FVs) that were extracted from each of the two modules (ACF and PRD) based on experiments conducted over the course of one day. We take advantage of a supervised machine learning technique, support vector machines (SVMs) with the RBF kernel function, to implement the authentication process. For each case behavior, the four time profiles were evaluated separately by training the SVM classifier on 70% of the feature vectors for each user. Finally, we evaluated the trained classifiers with the remaining 30% of the data to compute the accuracy values. In total, each module has 78 SVM classifiers corresponding to the 13 case behaviors multiplied by the 6 time profiles, (M, A, E, Allday-M, Allday-A, Allday-E).



(a)

(b)

Figure 3: Performance results from a one-day dataset for all 13 case behaviors for both modules: (a) ACF; (b) PRD.

The ACF module results shown in Figure 3a provide classification accuracy that varies from (76.6%) to (96.26%). Both the upper and lower boundaries of the classification accuracy are enhanced using the PRD module shown in Figure 3b, from 83.33% to 99.5%. The *Allday* profile achieves classification accuracy higher than individual profiles for both modules.

Table 1: Overall accuracy [%] at all time profiles

Module	Morning	Allday vs Morning	Afternoon	Allday vs Afternoon	Evening	Allday vs Evening
ACF	87.35	90.55	86.87	90.18	92.22	93.78
PRD	89.87	91.85	88.27	92.37	91.23	94.66
ACF+PRD	99.79	99.81	99.84	99.86	99.56	99.78

In order to show the ability of each module to generalize over a set of daily behaviors instead of a single behavior, we computed the overall classification accuracy by averaging the accuracy values of all conducted 13 behaviors from a one-day dataset for individual modules. However, because our system depends on light measurements that vary from one day to another, our final evaluation on unseen datasets consisted of readings collected over the course of 20 days using a combination of the two modules feature vectors (ACF+PRD). We trained the final SVM classifiers on a dataset of 17 days, and we used the remaining 3 days (day 1, day 8, day 15) as unseen data for testing. All evaluation final results are shown in Table 1. We found that the performance of the (ACF+PRD) combination is sufficient to detect user behavior and provide high classification results even on the unseen dataset. Finally, we conclude that our method provides an accurate and low-cost alternative solution for authenticating a specific user in indoor environments.

ACKNOWLEDGMENTS

This work was supported in part by NRF of Korea (NRF-2019R1C1C1007118) and the ICT Consilience Creative program (IITP-2019-2015-0-00742).

REFERENCES

[1] Chao Shen, Yuanxun Li, Yufei Chen, Xiaohong Guan, and Roy A Maxion. Performance analysis of multi-motion sensor behavior for active smartphone authentication. *IEEE Transactions on Information Forensics and Security*, 13(1):48–62, 2018.

[2] Bin Zou and Yantao Li. Touch-based smartphone authentication using import vector domain description. In *2018 IEEE 29th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, pages 1–4. IEEE, 2018.

[3] Y. Li, H. Hu, G. Zhou, and S. Deng. Sensor-based continuous authentication using cost-effective kernel ridge regression. *IEEE Access*, 6:32554–32565, 2018.

[4] Y. Li, H. Hu, and G. Zhou. Using data augmentation in continuous authentication on smartphones. *IEEE Internet of Things Journal*, 6(1):628–640, Feb 2019.