Poster: Inferring Mobile Payment Passcodes Leveraging Wearable Devices

Chen Wang[†], Jian Liu[†], Xiaonan Guo[§], Yan Wang[*], Yingying Chen[†]

[†]WINLAB, Rutgers University, North Brunswick, NJ 08902, USA
[§]Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202, USA
[*]Binghamton University, Binghamton, NY 13902, USA
chenwang@winlab.rutgers.edu,jianliu@winlab.rutgers.edu,xg6@iupui.edu,yanwang@binghamton.edu,
yingying.chen@rutgers.edu

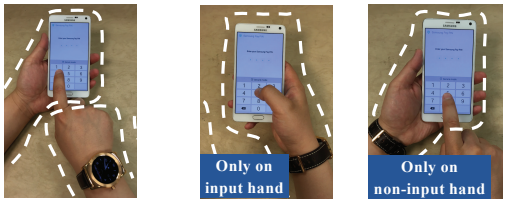
ABSTRACT

Mobile payment has drawn considerable attention due to its convenience of paying via personal mobile devices at anytime and anywhere, and passcodes (i.e., PINs) are the first choice of most consumers to authorize the payment. This work demonstrates a serious security breach and aims to raise the awareness of the public that the passcodes for authorizing transactions in mobile payments can be leaked by exploiting the embedded sensors in wearable devices (e.g., smartwatches). We present a passcode inference system, which examines to what extent the user's PIN during mobile payment could be revealed from a single wrist-worn wearable device under different input scenarios involving either two hands or a single hand. Extensive experiments with 15 volunteers demonstrate that an adversary is able to recover a user's PIN with high success rate within 5 tries under various input scenarios.

ACM Reference Format:
Chen Wang, Jian Liu, Xiaonan Guo, Yan Wang, Yingying Chen. 2018. Poster: Inferring Mobile Payment Passcodes Leveraging Wearable Devices . In *The 24th Annual International Conference on Mobile Computing and Networking (MobiCom '18), October 29-November 2, 2018, New Delhi, India.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3241539.3267742

1 INTRODUCTION

With the prevalent use of mobile devices (e.g., smartphones), mobile payments become increasingly attractive because they allow users to perform near real-time transactions anytime and anywhere conveniently. Users can easily use their digital wallets for in-store payments, make online purchases via in-app payments, and perform money transfer between two accounts using mobile money transfer. Thus, mobile payments bring users complete freedom from the shackles of currency and credit cards in transactions.



(a) Two hands: mobile device and wearable on **two hands**

(b) Single hand: both mobile device and wearable are on **the same hand**

Figure 1: Representative passcode input scenarios.

The extreme convenient utility of mobile payments also makes it an attractive target for adversaries. The passcode of the user is the first line of defense preventing the malicious use of mobile payments. Recent study demonstrates that motion sensors embedded in the increasingly popular wearable devices (e.g., smartwatches and fitness trackers) possess a more severe threat [3]. Toward this end, we propose a passcode inference system, to investigate to what extent the wearable's sensing data can reveal a user's mobile payment passcode when the user is operating on a small-sized smartphone screen with the consideration of different hand input scenarios (e.g., using two hands or single hand as illustrated in Figure 1)

The passcode input scenarios can be classified into two categories, namely *two-hand* and *one-hand*, based on which hand the user holds the mobile device and wears the wearable during the mobile payment process. In the two-hand scenarios, the user usually has the mobile device and wearable on two different hands when inputting PINs (i.e., Figure 1(a)). Whereas in the one-hand scenarios, the user uses

MobiCom '18, October 29-November 2, 2018, New Delhi, India

ACM ISBN 978-1-4503-5903-0/18/10.
https://doi.org/10.1145/3241539.3267742

the same single hand to wear the wearable and hold the mobile device (i.e., Figure 1(b)). Furthermore, when the wearable is on the input hand (i.e., dominant hand), the key tapping dynamics become weaker, which only involve thumb movements (as shown in the left figure of Figure 1(b)); When the wearable is on the non-input hand (i.e., non-dominant hand), it is even harder to capture the motion of the input hand because the wearable device is on the opposite wrist as depicted in the right figure of Figure 1(b). Recent studies [1, 2] show the possibility of classifying single keys on smartphone screen via smartwatch inertial sensing and smartphone touch events. The achieved classification accuracy is similar to those directly using smartphone sensors. However, the attacker's capability of revealing the complete passcodes under different hand-input scenarios via the wearables remains unclear. In this work, we perform a comprehensive study to explore the possibility of revealing the passcodes under all of the aforementioned hand-input scenarios. We summarize our main contributions as follows:

- We develop a system to explore the possibility of revealing the user's private information (e.g., passcode entered on smartphones) via wrist-worn wearables during the mobile payment process under various hand-input ways.
- We develop the training-free Euclidean distance-based model and the parallel PIN inference algorithms that can infer the user's PIN entries in the two-hand scenarios.
- The proposed system extracts unique features over the time duration of each tap in the one-hand scenarios. The multi-dimensional features in time series can well capture the weak wrist vibrations in response to PIN entries and classify taps accurately.

2 SYSTEM DESIGN

Figure 2 shows the flow of our system, which consists of two major building blocks: 1) *Devices on Two Hands* utilizes the sensor data to track fine-grained hand movement trajectories and infers users' passcodes when the mobile and the wearable devices are on two different hands; 2) *Devices on a Single Hand* identifies users' passcode entries when the devices are on the same hand. We note that the proposed system first determines victims' input scenarios (i.e., two-hand or one-hand) and then picks the corresponding building block to infer the victims' passcodes. Specifically, the system exploits the quaternions from the victims' mobile and wearable devices to determine the spatial relationships between the two devices and utilizes a threshold-based method to determine the input scenario.

2.1 Devices on Two Hands

After obtaining the motion sensor readings (e.g., Acceleration, Quaternion) from the wearable device, the system first performs the *Noise Reduction and Coordinate Alignment*. It

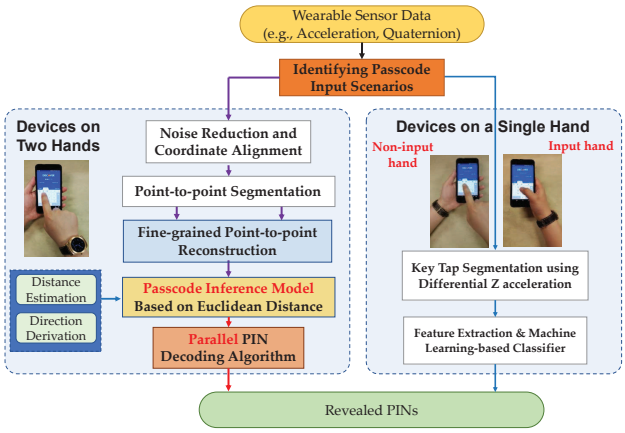


Figure 2: Mobile payment passcode inference framework.

removes high-frequency noises from the raw sensor readings and exploits quaternion measurements to align the coordinates of the two free-axis devices. Thus, the hand dynamics captured by the wearable sensors are translated to the movements on the on-screen keypad for PIN inference. Then the *Point-to-point Segmentation* examines the translated acceleration to determine the point-to-point segments by detecting the key taps of a PIN entry based on differential Z acceleration. Next, the *Fine-grained Point-to-point Reconstruction* estimates the distance and direction of the hand movement in each segment and reconstructs the point-to-point trajectory. A point-to-point trajectory reflects the hand movement between two consecutive key taps. For inferring passcode entries, the system builds a Euclidean-distance based model to describe the practical geometric relationships between real keys. The *Parallel PIN Decoding Algorithm* is designed to integrate the point-to-point trajectories in the model and search for the most likely PIN. Note that the decoding starts from all possible starting keys/dots in parallel and our algorithms only do add-and-compare operations, which greatly reduce the computational cost.

2.2 Devices on a Single Hand

Different from the two-hand scenarios, it is hard to recover the hand movement trajectory of the input hand if both the phone and the wearable are on the input hand. Moreover, it is even harder to capture the dynamics of the input hand if the wearable is on the non-input hand. We resort to capture the minute wrist movement differences that result from the various finger tapping positions on the on-screen keypad to recognize each tapped key. When the two devices are on the input hand, the movement of the thumb during tapping can be passed by the tendon to cause minute wrist movement. When the two devices are both on the non-input hand, the key tap on the phone causes vibrations on the phone, which are passed down to vibrate the wrist slightly. The

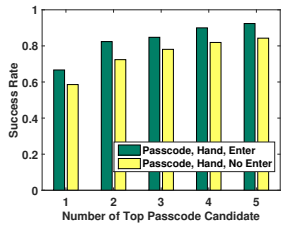


Figure 3: Performance of parallel PIN decoding in the two-hand scenarios.

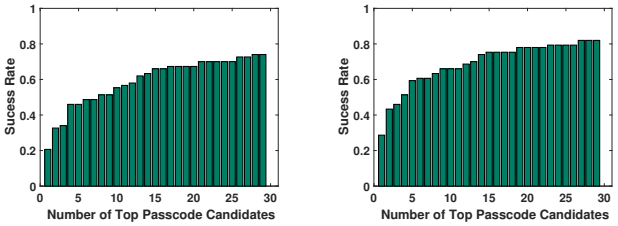
system utilizes a machine learning-based method to classify the tapping positions based on the unique vibration features. In particular, after obtaining the raw sensor data from the wearable, our system first performs *Key Tap Detection Using Differential Acceleration Z* to detect tapping actions based on differential acceleration Z and extract the data segment within a short time around each tap. Then the system further divides each tap segment into small pieces and extracts unique features in time series from both the coordinate aligned and non-aligned sensor data. The non-aligned sensor data (e.g., acceleration and gyroscope readings) describes the movement of the wearable itself and the aligned sensor data (e.g., accelerations aligned with the mobile device coordinate) shows the relative position change between the wearable and the smartphone. Based on the unique features, a machine learning-based classifier is proposed to recognize finger taps as each key to infer a complete PIN.

3 PRELIMINARY EXPERIMENTS

To evaluate the system, we ask our volunteers to enter passcodes on the on-screen keypad of multiple smartphones including Google Nexus one, Nexus 6P, Samsung Note 4 and Note 3 while wearing a smartwatch LG W150. When the volunteers enter passcodes, the smartwatch collects acceleration and quaternion readings under 100 *samples*/*sec* and sends the sensor data to a nearby server via Bluetooth.

We conduct experiments covering three passcode input scenarios as shown in Figure 1. We provide the participants with PINs from a pool, which is designed to include most difficulty levels of recovering the hand movement trajectories. The participants are asked to be familiar with their chosen ones before collecting data. Particularly, 20 distinct 4-digit PIN combinations are collected from 15 volunteers. In total, 1200 entered passcodes are collected.

Two Hands: PIN Inference. Figure 3(a) shows the top-k success rate of inferring PINs on the on-screen keypad with and without "Enter" key. We find that our system effectively reveals both types of PIN entries. In particular, by choosing the top-1 PIN candidate, our system achieves over 67% success rate for the PINs with an "Enter", while the success rate is about 60% for the PINs without an "Enter". Furthermore, the success rate to reveal the two types of PINs increases if



(a) Single input hand (b) Single non-input hand

Figure 4: 4-digit PIN decoding accuracy in the one-hand scenarios.

the adversary utilizes more candidates from the top-k candidate list. Specifically, 92% success rate is achieved to infer the PINs with an "Enter" by using the top-5 candidates. And the success rate for the PIN without an "Enter" is 84%. This indicates that the adversary can break the user's entered PINs with high probability within limited tries. Besides, we also find that the success rate of inferring the PINs with an "Enter" has higher accuracy. The reason is that the last tapped position of the PIN is fixed at the "Enter" key, which enables our parallel PIN decoding algorithm to start from one fixed key without guess.

Single Hand: Revealing PINs. We then evaluate the top-k success rate of inferring the user's complete PINs in the single-hand scenarios. Figure 4 shows the PIN inference accuracy for both single-hand scenarios. Specifically, when the adversary only tries once, the success rates are around 21% and 30% for the input hand and the non-input hand scenarios respectively. Within five tries, the attacker can achieve around 50% for the single input hand and 59% success rates for the non-input hand, which is a non-negligible security breach. Moreover, if the adversary can try 15 times, over 70% and 78% accuracies can be achieved for the single input hand and the single non-input hand scenarios, respectively. The results show that the wearable can capture the minute wrist motions in both single hand scenarios to accurately reveal a user's PIN on mobile devices.

Acknowledgments. This work was partially supported by the National Science Foundation Grants CNS-1820624, CNS-1826647 and ARO Grant W911NF-18-1-0221.

REFERENCES

[1] Anindya Maiti, Murtuza Jadliwala, Jibo He, and Igor Bilogrevic. 2015. (Smart) watch your taps: side-channel keystroke inference attacks using smartwatches. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 27–30.

[2] Sougata Sen, Karan Grover, Vigneshwaran Subbaraju, and Archan Misra. 2017. Inferring smartphone keypress via smartwatch inertial sensing. In *Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops*. IEEE, 685–690.

[3] Chen Wang, Xiaonan Guo, Yan Wang, Yingying Chen, and Bo Liu. 2016. Friend or Foe?: Your Wearable Devices Reveal Your Personal PIN. In *ACM ASIACCS*. 189–200.