Your PIN Sounds Good! Augmentation of PIN Guessing Strategies via Audio Leakage

Matteo Cardaioli[1,2], Mauro Conti[1,4], Kiran Balagani[3], and Paolo Gasti[3]

[1] University of Padua
[2] GFT Italy
[3] New York Institute of Technology
[4] University of Washington, Seattle

Abstract. Personal Identification Numbers (PINs) are widely used as the primary authentication method for Automated Teller Machines (ATMs) and Point of Sale (PoS). ATM and PoS typically mitigate attacks including shoulder-surfing by displaying dots on their screen rather than PIN digits, and by obstructing the view of the keypad. In this paper, we explore several sources of information leakage from common ATM and PoS installations that the adversary can leverage to reduce the number of attempts necessary to guess a PIN. Specifically, we evaluate how the adversary can leverage audio feedback generated by a standard ATM keypad to infer accurate inter-keystroke timing information, and how these timings can be used to improve attacks based on the observation of the user's typing behavior, partial PIN information, and attacks based on thermal cameras. Our results show that inter-keystroke timings can be extracted from audio feedback far more accurately than from previously explored sources (e.g., videos). In our experiments, this increase in accuracy translated to a meaningful increase in guessing performance. Further, various combinations of these sources of information allowed us to guess between 44% and 89% of the PINs within 5 attempts. Finally, we observed that based on the type of information available to the adversary, and contrary to common knowledge, uniform PIN selection is not necessarily the best strategy. We consider these results relevant and important, as they highlight a real threat to any authentication system that relies on PINs.

1 Introduction

Authentication via Personal Identification Numbers (PINs) dates back to the mid-sixties [5]. The first devices to use PINs were automatic dispensers and control systems at gas stations, while the first applications in the banking sector appeared in 1967 with cash machines [7]. PINs have found widespread use over the years in devices with numeric keypads rather than full keyboards [22].

In the context of financial services, ISO 9564-1 [10] specifies basic security principles for PINs and PIN entry devices (e.g., PIN pads). For instance, to mitigate shoulder surfing attacks [17,12,13], ISO 9564-1 indicates that PIN digits must not be displayed on a screen, or identified using different sounds or sound duration for each key.

As a compromise between security and usability, PIN entry systems display a fixed symbol (e.g., a dot) to represent a key being pressed, and provide the same audio feedback (i.e., same tone, same duration) for all keys. While previous work has demonstrated that observing the dots as they appear on screen as a result of a key press reduces the search space for a PIN [4], to our knowledge no work has targeted the use of audio feedback to recover PINs.

In this paper, we evaluate how the adversary can reduce PIN search space using audio feedback, with (and without) using observable information such as PIN typing behavior (one- or two-handed), knowledge of one digit of the PIN, and knowledge of which keys have been pressed. We compare our attacks with an attack based on the knowledge of PIN distribution.

Exploiting audio feedback has several advantages compared to observing the user or the screen during PIN entry. First, sound is typically easier to collect. The adversary might not be able to observe the ATM's screen directly, and might risk being exposed when video-recording an ATM in a public space. In contrast, it is easy to record audio *covertly*, e.g., by casually holding a smartphone while pretending to stand in a line behind other ATM users. The sound emitted by ATMs is quite distinctive and can be easily isolated even in noisy environments. Second, sound enables higher time resolution compared to video. Conventional video cameras and smartphones record video between 24 and 120 frames per second. In contrast, audio can be recorded with a sampling rate between 44.1 kHz and 192 kHz, thus potentially allowing at least two orders of magnitude higher resolution.

Contributions. In this paper, we analyze several novel side channels associated with PIN entry. In particular:

1. We show that it is possible to retrieve accurate inter-keystroke timing information from audio feedback. In our experiments, we were able to correctly detect 98% of the keystroke feedback sounds with an average error of 1.8ms. Furthermore, 75% of inter-keystroke timings extracted by the software had absolute error under 15 ms. Our experiments also demonstrate that inter-keystroke timings extracted from audio can be more accurate than the same extracted from video recordings of PIN entry as done in [3,4].
2. We analyze how the behavior of the user affects the adversary's ability to guess PINs. Our results show that users who type PINs with one finger are more vulnerable to PIN guessing from inter-keystroke timings compared to users that enter their PIN using at least two fingers. In particular, the combining inter-keystroke timing with the knowledge that the user is a single-finger typist leads to 34-fold improvement over random guessing when the adversary is allowed to perform up to 5 guessing attempts.
3. We combine inter-keystroke timing information with knowledge of one key in the PIN (i.e., the adversary was able to see either the first or the last key pressed by the user), and with knowledge of *which* keys have been pressed by the user. The latter information is available, as shown in this paper as well as in recent work [24,11,1,16] when the adversary is able to capture a thermal image of the PIN pad after the user has typed her PIN. Our experiments show that inter-keystroke timing significantly improves performance for both attacks.

For example, by combining inter-keystroke timing with a thermal attack, we were able to guess 15% of the PINs at the first attempt, reaching a four-fold improvement in performance. By combining multiple attacks, we were also able to drastically reduce the number of attempts required to guess a PIN. Specifically, we were able to guess 72% of the PINs within the first 3 attempts.

4. Finally, we show that uniform PIN selection might not be the best strategy against an adversary with access to one or more of the side-channel information discussed in this paper.

Organization. Section 2 reviews related work on password and PIN guessing. Section 3 presents our adversary model. We present our algorithms for inter-keystroke timing extraction in Section 4.1. In Section 4, we present the results of our experiments, while in Section 5 we analyze how different side-channels affect the guessing probability of individual PINs. We conclude in Section 6.

2 Related Work

Non-acoustic Side-channels. Vuagnoux and Pasini [20] demonstrated that it is possible to recover keystrokes by analyzing electromagnetic emanations from electronic components in wired and wireless keyboards. Marquardt et al. [15] showed that it is possible to recover key presses by recoding vibrations generated by a keyboard using an accelerometer. Other attacks focus on keystroke inference via motion detection from embedded sensors on wearable devices. For example, Sarkisyan et al. [18] and Wang et al. [21] infer smartphone PINs using movement data recorded by a smartwatch.

Those attacks require that the adversary is able to monitor the user's activity while the user is typing. However, there are attacks that allow the adversary to exploit information available several seconds after the user has typed her password. For instance, one such attack is based the observation that when a user presses a key, the heat from her finger is transferred to the keypad, and can be later be measured using a thermal camera [24]. Depending on the material of the keyboard, thermal residues have different dissipation rates [16], thus affecting the time window in which the attacks are effective. Abdelrahman et al. [1] evaluated how different PINs and unlock patterns on smartphones on can influence thermal attack performance. Kaczmarek et al. [11] demonstrated how a thermal attack can recover precise information about a password up to 30 seconds after it was typed, and partial information within 60 seconds.

Acoustic Side-channels. Asonov and Agrawal showed that each key on a keyboard emits a characteristic sound, and that this sound can be used to infer individual keys [2]. Subsequent work further demonstrated the effectiveness of sound emanation for text reconstruction. Berger et al. [6] combined keyboard acoustic emanation with a dictionary attack to reconstruct words, while Halevi and Saxena [9] analyzed keyboard acoustic emanations to eavesdrop over random password. Because ISO 9564-1 [10] specifications require that each key emits the same sound, those attacks do not apply to common keypads, including those on ATMs.

Another type of acoustic attack is based on time difference of arrivals (TDoA) [25,23,14]. These attacks rely on multiple microphones to triangulate the position of the keys pressed. Although this attacks typically result in good accuracies, they are difficult to instantiate in realistic environments.

Song et al. [19] presented an attack based on latency between key presses measured by snooping encrypted SSH traffic. Their experiments show that information about inter-keystroke timing can be used to narrow the password search space substantially. A similar approach was used by Balagani et al. [4], who reconstructed inter-keystroke timing from the time of appearance of the masking symbols (e.g., "dots") while a user types her password. Similarly, Balagani et al. [3] demonstrated that precise inter-keystroke timing information recovered from videos drastically reduces the number of attempts required to guess a PIN. The main limitation of [4,3] is that they require the adversary to video-record the ATM screen while the user is typing her PIN. Depending on the location and the ATM, this might not be feasible. Further, this reduces the set of vulnerable ATMs and payment systems to those that display on-screen feedback.

To our knowledge, this is the first paper to combine inter-keystroke timing information deduced from sound recording with observable information from other sources, and thereby drastically reduce the attempts to guess a PIN compared to prior work. Our attacks are applicable to a multitude of realistic scenarios. This poses an immediate and severe threat to current ATMs or PoS.

3 Adversary Model

In this section we evaluate four classes of information that the adversary can exploit to infer PINs. These classes are: (1) Key-stroke timing information extracted from audio recordings; (2) Knowledge of whether the user is a single- or multi-finger typist; (3) Information about the first or the last digit of the PIN; and (4) Information about which keys have been pressed, but not their order. Next, we briefly review how each of these classes of information can be collected by the adversary.

Class 1: Keystroke Timing. Keystroke timing measures the distance between consecutive keystroke events (e.g., the time between two key presses, or between the release of the key and the subsequent keypress). Collecting keystroke timing by compromising the software of an ATM located in a public space, or physically tampering with the ATM (e.g., by modifying the ATM's keyboard) is not practical in most cases. However, as shown in [3], the adversary can infer keystroke timings without tampering with the ATM by using video recordings of the "dots" that appear on the screens when the user types her PIN. In this paper, we leverage audio signals to infer precise inter-keystroke timings.

Class 2: Single- or Multi-finger Typists. The adversary can typically directly observe whether the user is typing with one or more fingers. While the number of fingers used to enter a PIN does not reveal information about the PIN itself, it might be a useful constraint when evaluating other sources of information leakage. Figure 1 shows users typing using a different number of fingers.

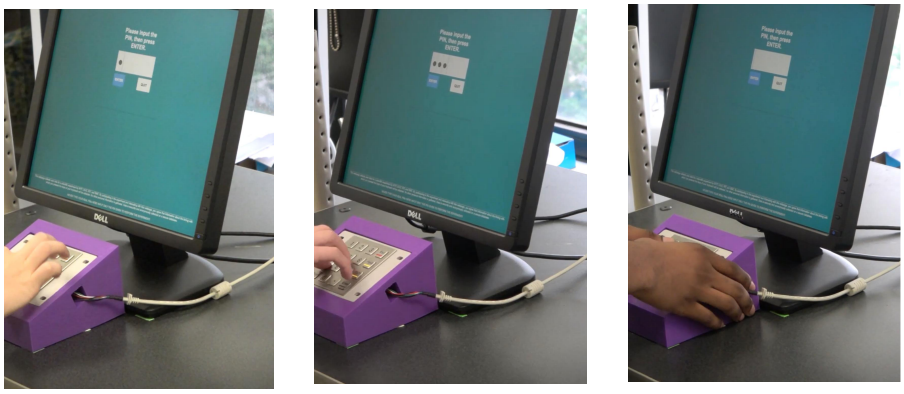


Fig. 1: Different typing strategies. Left: one finger; center: multiple fingers of one hand; right: multiple fingers of two hands.

Class 3: Information about the first or the last digit of the PIN. As users move their hands while typing their PIN, the adversary might briefly have visibility of the keypad, and might be able to see one of the keys as it is pressed (see Figure 1). We model this information by disclosing either the first or the last digit of the PIN to the adversary.

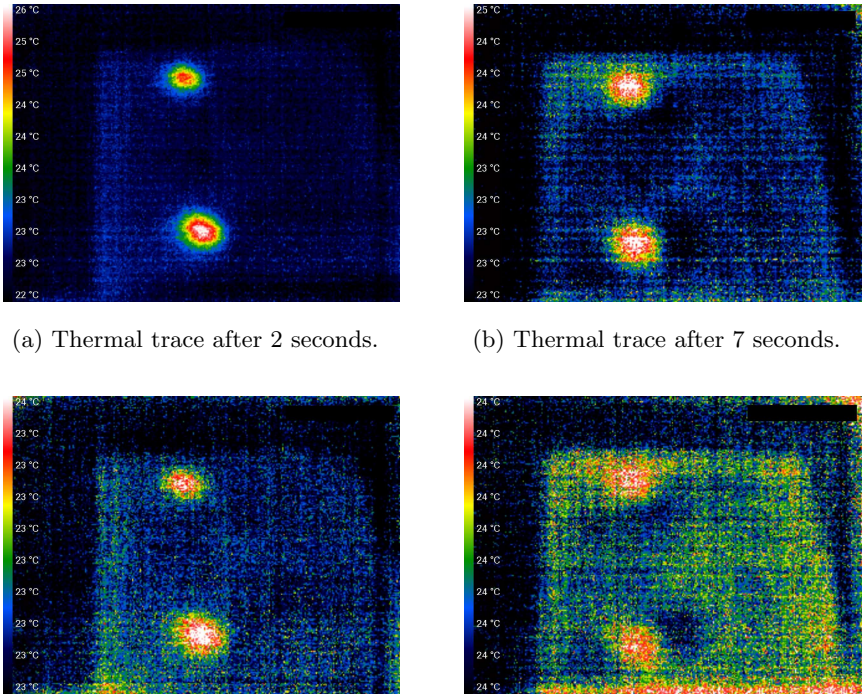
Class 4: Which Keys Have Been Pressed. This information can be collected using various techniques. For instance, the adversary can use a thermal camera to determine which keys are warmer, thus learning which digits compose the PIN (see, e.g., Figure). As an alternative, the adversary can place UV-sensitive powder on the keys before the user enters her PIN, and then check which keys had the powder removed by the users using a UV light.

While these attacks do not reveal the order in which the keys were pressed (except when the PIN is composed of one repeated digit), they significantly restrict the search space. Although this attack can be typically performed covertly, it requires specialized equipment.

4 Experiment Results

We extracted keystroke sounds using the dataset from [4]. This dataset was collected from 22 subjects, who typed several 4-digit PINS on a simulated ATM (see Figure 3). Nineteen subjects completed three data collection sessions, while three subjects completed only one session.

In each session, subjects entered a total of 180 PINs as follows: each subject was shown a 4-digit PIN. The PIN remained on the screen for 10 seconds, during which the subject was encouraged to type the PIN multiple times. After 10 seconds, the PIN disappeared from the screen. At this point, the subject was asked to type the PIN 4 times from memory. In case of incorrect entry, the PIN



(a) Thermal trace after 2 seconds.

(b) Thermal trace after 7 seconds.

(c) Thermal trace after 10 seconds.

(d) Thermal trace after 15 seconds.

Fig. 2: Thermal image of a metallic PIN pad after applying a transparent plastic cover for PIN 2200.

was briefly displayed again on the screen, and the subject was allowed to re-enter it. This procedure was repeated in three batches of 15 PINs. As a result, each PIN was typed 12 times per session.

Each time a subject pressed a key, the ATM simulator emitted an audio feedback and logged the corresponding timestamp with millisecond resolution. Users were asked to type 44 different 4-digit PINs which represented all the Euclidean distances between keys on the keypad. Sessions were recorded in a relatively noisy indoor public space (SNR -15 dB) using a Sony FDR-AX53 camera located approximately 1.5 m away from the PIN pad. The audio signal was recorded with a sampling frequency of 48 kHz.

4.1 Extraction of Keystroke Timings from Keypad Sound

To evaluate the accuracy of timing extraction from keystroke sounds, we first linearly normalized the audio recording amplitude in the interval $[-1, 1]$. We applied a 16-order Butterworth band-pass filter [8] centered at 5.6 kHz to isolate

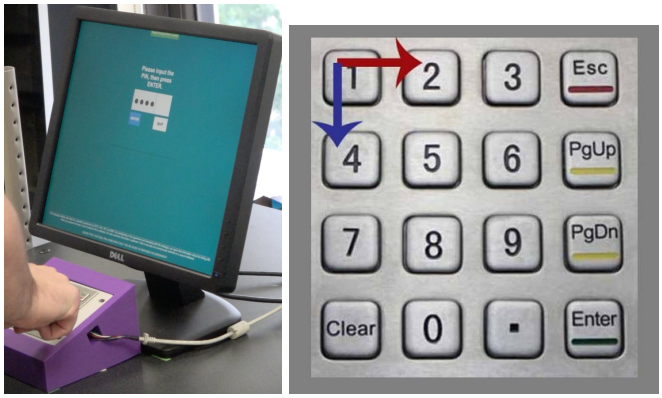


Fig. 3: Left: user typing a PIN using the ATM simulator. Right: close up view of the ATM simulator's keypad.

the characteristic frequency window of the keypad feedback sound. Finally, to isolate the signal from room noise, we processed the audio recording to "mute" all samples with an amplitude below a set threshold (0.01 in our experiments).

We then calculated the maximum amplitude across nearby values in a sliding window of 1,200 samples (consecutive windows had 1199 overlapping samples), corresponding to 25 milliseconds of audio recording. We determined the length of the window by evaluating the distance between consecutive timestamps logged by the ATM simulator (ground truth), which was at least 100 ms for 99.9% of the keypairs. Figure 4 shows the result of this process.

We then extracted the timestamps of the peaks of the processed signal and compared them to the ground truth. Our results show that this algorithm can accurately estimate inter-keystroke timing information. We were able to correctly detect 98% of feedback sound with a mean error of 1.8 ms.

Extracting timings from audio led to more accurate time estimation than using video [4]. With the latter, 75% of the extracted keystroke timings had errors of up to 37 ms. In contrast, using audio we were able to extract 75% of the keystroke with errors below 15 ms. Similarly, using video, 50% of the estimated keystrokes timings had errors of up to 22 ms, compared to less than 7 ms with audio. Figure 5 shows the errors distribution for timings extracted from video and audio recordings.

4.2 PIN Inference from Keystroke Timing (Class 1)

This attack ranks PINs based on the estimated Euclidean distance between subsequent keys in each PIN. In particular, we calculated an inter-key Euclidean distance vector from a sequence of inter-keystroke timings inferred from audio feedback. As an example, the distance vector associated with PIN `5566` is $[0, 1, 0]$, where the first '0' is the distance between keys `5` and `5`, '1' between keys `5` and `6`, and '0' between `6` and `6`. Any four-digit PIN is associated with one distance vector of size three. Each element of the distance vector can be 0, 1, 2, 3, diagonal distance

1 (e.g., 1-3), diagonal distance 2 (e.g., 3-7), short diagonal distance (e.g., 2-9), or long diagonal distance (e.g., 3-0). Different PINs might be associated with the same distance vector (e.g., 1234 and 4567). The goal of this attack is to reduce the search space by considering only PINs that match the estimated distance vector.

For evaluation, we split our keystroke dataset into two sets. The first (training set) consists of 5195 PINs, typed by 11 subjects. The second (test set) consists of 5135 PINs, typed by a separate set of 11 subjects. This models the lack of knowledge of the adversary of the specific typing patterns of the victim user.

To estimate the Euclidean distances between consequent keys, we modeled a set of gamma function on the inter-keystroke timing distribution, one for each distance. We then applied the algorithm from [3] to infer PINs from estimated

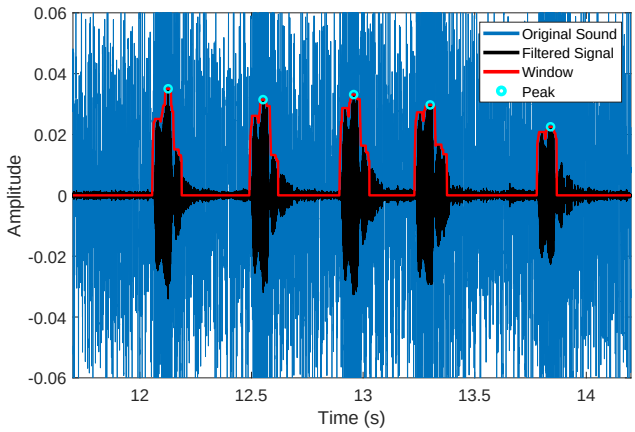


Fig. 4: Comparison between the original sound signal, filtered sound signal, windowed signal, and extracted peaks.

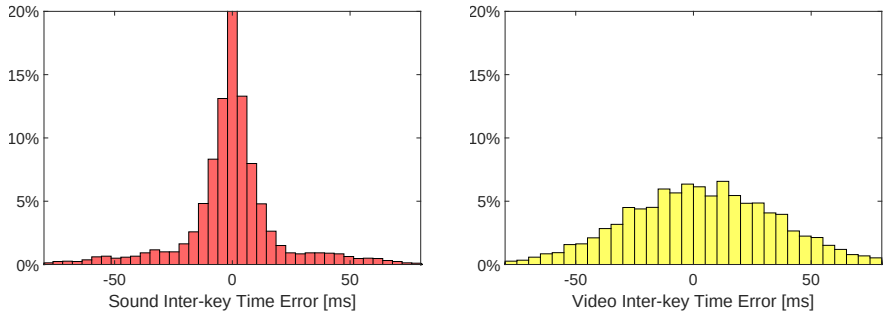


Fig. 5: Error distribution of estimated inter-keystroke timings. Left: timing errors from audio. Right: timing errors from video.

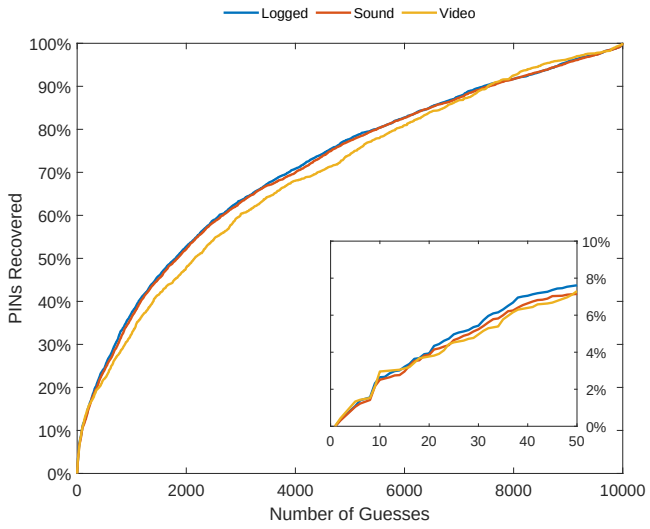


Fig. 6: CDF showing the percentage of PINs recovered using keystroke timing information derived from the ground truth (logged), sound feedback, and video.

distances. With this strategy, we were able to guess 4% of PINs within 20 attempts—a 20-fold improvement compared to random guessing.

Figure 6 shows how timings extracted from audio and video feedback affect the number of PIN guessed by the algorithm compared to ground truth. Timings extracted from audio feedback exhibit a smaller decrease in guessing performance compared to timings extracted from video.

4.3 PIN Inference from Keystroke Timing and Typing Behavior (Class 2)

This attack improves on the keystroke timing attack by leveraging knowledge of whether the user is a single- or multi-finger typist. This additional information allows the adversary to better contextualize the timings between consecutive keys. For single-finger typists, the Euclidean distance between keys `1` and `0` is the largest (see Fig 3), and therefore we expect the corresponding inter-keystroke timing to be the largest. However, if the user is a two-finger typist, then `1` might be typed with the right hand index finger, and `0` with the left hand index finger. As a result, the inter-keystroke time might not be representative of the Euclidean distance between the two keys.

To systematically study typing behavior, we analyzed 61 videos from the 22 subjects. 70% of the subject were single-finger typists; 92% of them entered PINs using the index finger, and 8% with the thumb. We divided multi-finger typists into three subclasses: (1) PINs entered using fingers from two hands (38% of the

PINs typed with more than one finger); (2) PINs entered with at least two fingers of the same hand (34% of the PINs typed with more than one finger); and (3) PINs that we were not able to classify with certainty due obfuscation of the PIN pad in the video recording (28% of the PINs typed with more than one finger).

In our experiments, subjects' typing behavior was quite consistent across PINs and sessions. Users that were predominantly single-finger typists entered 11% of their PINs using more than one finger, while multi-finger typists entered 23% of the PINs using one finger.

We evaluated guessing performance of timing information inferred from audio feedback on single-finger PINs and multi-finger PINs separately. We were able to guess a substantially higher number of PINs for each number of attempts for users single-finger typists (see Figure 7) compared to multi-finger typists. In particular, the percentage of PINs recovered within 5 attempts was twice as high for PINs entered with one finger compared to PINs entered with multiple fingers. Further, the guessing rate within the first 5 attempts was 34 times higher compared to random guessing when using timing information on single-finger PINs. However, our ability to guess multi-finger PINs using timing information was only slightly better than random. This strongly suggests that the correlation between inter-keystroke timing and Euclidean distance identified in [4] holds only quite strongly for PINs entered using a single finger, and only marginally for PINs entered with two or more fingers.

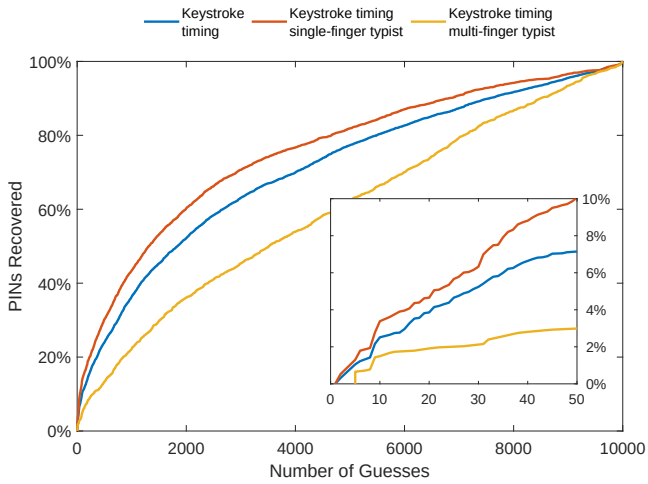


Fig. 7: CDF showing the percentage of PINs recovered using only keystroke timing information from audio feedback, compared to timing information for single- or multi-finger typists.

4.4 Knowledge of the First or the Last Digit of the PIN (Class 3)

In this section, we examine how information on the first or last digit of the PIN reduces the search space when combined with keystroke timings. Knowledge of one digit alone reduces the search space by a factor of 10 regardless of the digit's position, because the adversary needs to guess only the remaining three digits. (As a result, the expected number of attempts to guess a random PIN provided no additional information is 500.)

To determine how knowledge of the first or the last digit impacts PIN guessing based on keystroke timing, we applied the same procedure described in Section 4.2: for each PIN in the testing set, we associated a list of triplets of distances sorted by probability. We then pruned the set of PINs associated with those distance triplets to match the knowledge of the first or last PIN. For instance, given only the estimated distances 3, 0, and $\sqrt{2}$, the associated PINs are 0007, 0009, 2224, and 2226. If we know that the first digit of the correct PIN is 2, then our guesses are reduced to 2224 and 2226.

Information about the first or last digit of the PIN boosted the guessing performance of the keystroke-timing attack substantially, as shown in Figure 8. In particular, guessing accuracy increased by 15-19 times within 3 attempts (4.36% guessing rate when the first digit was known, and 5.57% when the last digit was known), 7 times within 5 attempts, and about 4 times within 10 attempts, compared to timing information alone. In all three cases, timing information substantially outperformed knowledge of one of the digits in terms of guessing rate.

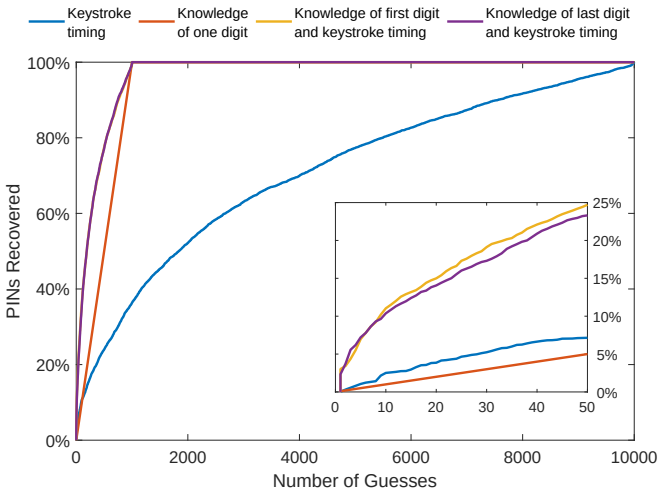


Fig. 8: CDF showing the percentage of PINs recovered using keystroke timings inferred from audio, random guessing over 3 digits of the PIN, and using inferred keystroke timings and knowledge of the first or the last digit of the PIN.

4.5 Knowledge of Which Keys Have Been Pressed (Class 4)

In this section, we evaluate how knowledge of *which digits* compose a PIN, but not *their order*, restricts the PIN search space, in conjunction with information about keystroke timings. The adversary can acquire this knowledge, for instance, by observing the keypad using a thermal camera shortly after the user has typed her PIN [11], or by placing UV-sensitive powder on the keys before the user enters her PIN, and then checking which keys were touched using a UV light.

Information on which digits compose a PIN can be divided as follows:

1. The user pressed only one key. In this case, the user must have entered the same digit 4 times. No additional information is required to recover the PIN.
2. The user pressed two distinct keys, and therefore each digit of the PIN might be repeated between one and three times, and might be in any position of the PIN. In this case, the number of possible PINs is $2^4 - 2 = 14$, i.e., the number of combinations of two values in four position, except for the combinations where only one of the two digits appears.
3. The user pressed three distinct keys. The number of possible PINs is equal to the combinations of three digits in four positions, i.e., $4 \cdot 3 \cdot 3 = 36$
4. The user pressed four distinct keys. The number of possible PINs is $4! = 24$.

We evaluated how many PINs the adversary could recover given keystroke timings and the set of keys pressed by the user while entering the PIN. Our results, presented in Figure 9, show that combining these two sources of information leads to a high PIN recovery rate. Specifically, within the first three attempts, knowing only which keys were pressed led to the recovery of about 11% of the PINs. Adding timing information increased this value to over 33%.

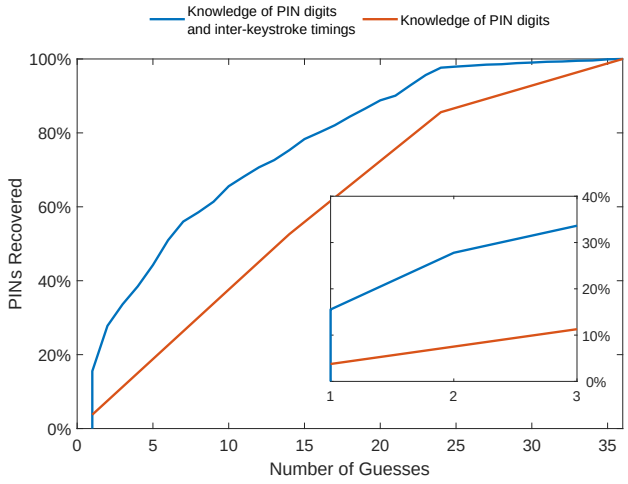


Fig. 9: CDF showing the percentage of PINs recovered with the knowledge of which keys have been pressed with and without inter-keystroke timing information.

4.6 Combining Multiple Classes of Information

In this section we examine how combining multiple classes of information leads to an improvement in the probability of correctly guessing a PIN.

First, we investigated how guessing probability increases when the adversary knows one of the digits of the PIN (first or last), the typing behavior (single-finger typist), and is able to infer inter-keystroke timing information from audio feedback. We used 3461 PINs typed by 11 subjects containing only single-finger PINs. In our experiments, we were able to guess 8.73% of the PINs within 5 attempts, compared to 6.97% with timing information and knowledge of one digit.

We then considered knowledge of the values composing the PIN, typing behavior, and inferred timing information. In this case, we successfully guessed 50.74% of the PINs within 5 attempts, and 71.39% within 10 attempts.

Finally, when we considered the values composing the PIN, one of the PIN's digits, and inferred timing information, we were able to guess 86.76% of the PINs in 5 attempts, and effectively all of them (98.99%) within 10 attempts.

All our results are summarized in Table 1.

5 PINs and Their Guessing Probability Distribution

In this section, we evaluate whether the classes of information identified in this paper make some of the PINs easier to guess than others, and thus intrinsically less secure. With respect to estimated inter-keystroke timings, different timing vectors identify a different number of PINs. For instance, vector [0,0,0] corresponds to 10 distinct PINs (`0000`, `1111`, ...), while vector [1,1,1] corresponds to 216 PINs (`0258`, `4569`, ...). This indicates that, against adversaries who are able to infer inter-keystroke timing information, choosing PINs uniformly at random from the entire PIN space is not the best strategy.

The adversary's knowledge of which digits compose the PIN has a similar effect of the guessing probability of individual PINs. In this case, PINs composed of three different digits are the hardest to guess, with a probability of 1/36, compared to PINs composed of a single digit, which can always be guessed at the first attempt.

The adversary's knowledge of one digit of the PIN and/or the typing behavior do not affect the guessing probability of individual PINs.

6 Conclusion

In this paper, we showed that inter-keystroke timing inferred from audio feedback emitted by a PIN pad compliant with ISO 9564-1 [10] can be effectively used to reduce the attempts to guess a PIN. Compared to prior sources of keystroke timing information, audio feedback is easier to collect, and leads to more accurate timing estimates (in our experiments, the average reconstruction error was 1.8 ms). Due to this increase in accuracy, we were able to reduce the number of attempts needed to guess a PIN compared to timing information extracted from videos.

Table 1: Results from all combinations of attacks considered in this paper, sorting by guessing rate after 5 attempts. Because single finger reduces the PIN search space only in conjunction with inter-keystroke timings, we do not present results for single finger alone.

Information				PINs Guessed Within Attempt				
Keystroke Timing	Single Finger	First Digit	PIN Digits	1	2	3	5	10
				0.01%	0.02%	0.03%	0.05%	0.10%
		o		0.10%	0.20%	0.30%	0.50%	1.00%
o				0.02%	0.31%	0.70%	1.05%	2.51%
o	o			0.03%	0.52%	0.91%	1.30%	3.38%
o		o		3.02%	3.72%	4.36%	6.97%	11.04%
o	o	o		3.73%	4.13%	5.43%	8.73%	14.01%
			o	3.76%	7.52%	11.28%	18.80%	37.60%
o			o	15.54%	27.79%	33.63%	44.25%	65.57%
o	o		o	19.04%	34.01%	40.60%	50.74%	71.31%
		o	o	13.27%	26.62%	39.88%	66.40%	92.80%
o		o	o	35.27%	53.46%	66.84%	86.76%	98.99%
o	o	o	o	40.86%	60.24%	71.77%	89.19%	99.28%

We then analyzed how using inter-keystroke timing increases guessing performance of other sources of information readily available to the adversary. When the adversary was able to observe the first or the last digit of a PIN, inter-keystroke timings further increased the number of PINs guessed within 5 attempts by 14 times. If the adversaries was capable of observing which keys were pressed to enter a PIN (e.g., using a thermal camera), adding inter-keystroke timing information allowed the adversary to guess 15% of the PINs with a single attempt. This corresponds to a 4 times reduction in the number of attempts compared to knowing only which keys were pressed.

We evaluated how typing behavior affects guessing probabilities. Our results show that there is a strong correlation between Euclidean distance between keys and inter-keystroke timings when the user enters her PIN using one finger. However, this correlation was substantially weaker when users typed with more than one finger.

We then showed that the combination of multiple attacks can dramatically reduce attempts to guess the PIN. In particular, we were able to guess 72% of the PINs within the first 3 attempts, and about 90% of the PINs within 5 attempts, by combining all the sources of information considered in this paper.

Finally, we observed that different adversaries require different PIN selection strategies. While normally PINs should be selected uniformly at random from the entire PIN space, this is not true when the adversary has access to inter-keystroke timings or thermal images. In this case, some classes of PINs (e.g., those composed of a single digit) are substantially easier to guess than other classes (e.g., those composed of three different digits). As a result, uniform selection from appropriate *subsets* of the entire PIN space leads to harder-to-guess PINs against those adversaries.

We believe that our results highlight a real threat to PIN authentication systems. The feasibility of these attacks and their immediate applicability in real scenarios poses a considerable security threat for ATMs, PoS-s, and similar devices.

References

1. Abdelrahman, Y., Khamis, M., Schneegass, S., Alt, F.: Stay cool! understanding thermal attacks on mobile-based user authentication. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 3751–3763. ACM (2017)
2. Asonov, D., Agrawal, R.: Keyboard acoustic emanations. In: IEEE S&P (2004)
3. Balagani, K., Cardaioli, M., Conti, M., Gasti, P., Georgiev, M., Gurtler, T., Lain, D., Miller, C., Molas, K., Samarin, N., et al.: Pilot: Password and pin information leakage from obfuscated typing videos. Journal of Computer Security **27**(4), 405–425 (2019)
4. Balagani, K.S., Conti, M., Gasti, P., Georgiev, M., Gurtler, T., Lain, D., Miller, C., Molas, K., Samarin, N., Saraci, E., et al.: Silk-tv: Secret information leakage from keystroke timing videos. In: European Symposium on Research in Computer Security. pp. 263–280. Springer (2018)
5. Bátiz-Lazo, B., Reid, R.: The development of cash-dispensing technology in the uk. IEEE Annals of the History of Computing **33**(3), 32–45 (2011)
6. Berger, Y., Wool, A., Yeredor, A.: Dictionary attacks using keyboard acoustic emanations. In: Proceedings of the 13th ACM conference on Computer and communications security. pp. 245–254. ACM (2006)
7. Bonneau, J., Preibusch, S., Anderson, R.: A birthday present every eleven wallets? the security of customer-chosen banking pins. In: International Conference on Financial Cryptography and Data Security. pp. 25–40. Springer (2012)
8. Butterworth, S.: On the theory of filter amplifiers. Wireless Engineer **7**(6), 536–541 (1930)
9. Halevi, T., Saxena, N.: A closer look at keyboard acoustic emanations: random passwords, typing styles and decoding techniques. In: Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security. pp. 89–90. ACM (2012)

10. ISO: Financial services – personal identification number (pin) management and security – part 1: Basic principles and requirements for pins in card-based systems (2017), https://www.iso.org/standard/68669.html
11. Kaczmarek, T., Ozturk, E., Tsudik, G.: Thermanator: Thermal residue-based post factum attacks on keyboard password entry. arXiv preprint arXiv:1806.10189 (2018)
12. Kumar, M., Garfinkel, T., Boneh, D., Winograd, T.: Reducing shoulder-surfing by using gaze-based password entry. In: Proceedings of the 3rd symposium on Usable privacy and security. pp. 13–19. ACM (2007)
13. Kwon, T., Hong, J.: Analysis and improvement of a pin-entry method resilient to shoulder-surfing and recording attacks. Ieee transactions on information forensics and security **10**(2), 278–292 (2015)
14. Liu, J., Wang, Y., Kar, G., Chen, Y., Yang, J., Gruteser, M.: Snooping keystrokes with mm-level audio ranging on a single phone. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. pp. 142–154. ACM (2015)
15. Marquardt, P., Verma, A., Carter, H., Traynor, P.: (sp) iphone: Decoding vibrations from nearby keyboards using mobile phone accelerometers. In: Proceedings of the 18th ACM conference on Computer and communications security. pp. 551–562. ACM (2011)
16. Mowery, K., Meiklejohn, S., Savage, S.: Heat of the moment: Characterizing the efficacy of thermal camera-based attacks. In: Proceedings of the 5th USENIX conference on Offensive technologies. pp. 6–6. USENIX Association (2011)
17. Roth, V., Richter, K., Freidinger, R.: A pin-entry method resilient against shoulder surfing. In: Proceedings of the 11th ACM conference on Computer and communications security. pp. 236–245. ACM (2004)
18. Sarkisyan, A., Debbiny, R., Nahapetian, A.: Wristsnoop: Smartphone pins prediction using smartwatch motion sensors. In: 2015 IEEE international workshop on information forensics and security (WIFS). pp. 1–6. IEEE (2015)
19. Song, D.X., Wagner, D., Tian, X.: Timing analysis of keystrokes and timing attacks on ssh. In: USENIX Security Symposium (2001)
20. Vuagnoux, M., Pasini, S.: Compromising electromagnetic emanations of wired and wireless keyboards. In: USENIX security symposium. pp. 1–16 (2009)
21. Wang, C., Guo, X., Wang, Y., Chen, Y., Liu, B.: Friend or foe?: Your wearable devices reveal your personal pin. In: Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security. pp. 189–200. ACM (2016)
22. Wang, D., Gu, Q., Huang, X., Wang, P.: Understanding human-chosen pins: characteristics, distribution and security. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. pp. 372–385. ACM (2017)
23. Wang, J., Zhao, K., Zhang, X., Peng, C.: Ubiquitous keyboard for small mobile devices: harnessing multipath fading for fine-grained keystroke localization. In: Proceedings of the 12th annual international conference on Mobile systems, applications, and services. pp. 14–27. ACM (2014)
24. Zalewski, M.: Cracking safes with thermal imaging. ser. http://lcamtuf. coredump. cx/tsafe (2005)
25. Zhu, T., Ma, Q., Zhang, S., Liu, Y.: Context-free attacks using keyboard acoustic emanations. In: ACM CCS (2014)