Acoustic Emanation of Haptics as a Side-Channel for Gesture-Typing Attacks

Jonathan Francis Roscoe, Max Smith-Creasey
Future Security and Cyber Defence,
BT Applied Research, Adastral Park, UK
{jonathan.roscoe, max.smith-creasey}@bt.com

***Abstract*—In this paper, we show that analysis of acoustic emanations recorded from haptic feedback during gesture-typing sessions is a viable side-channel for carrying out eavesdropping attacks against mobile devices. The proposed approach relies on acoustic emanation resulting from haptic events, namely the buzz of a small vibration motor as the finger initiates the gesture-typing of a work in a sentence. By analysing time between haptic feedback events, it is possible to identify the text that a user enters via the soft keyboard on their device. The attack requires no prior interaction or need to install software on the target device (unlike similar works); only the ability to record audio within the vicinity. We present an experimental framework to illustrate the feasibility of the attack. In the experiments we show that sentences can be detected with an accuracy of 70% with some sentences identified with up to 95% accuracy. The attack can be conducted with minimal computation and on non-specialist consumer equipment. The paper concludes by proposing a number of countermeasures that mitigate the ability of an attacker to successfully intercept keyboard input.**

***Index Terms*—mobile, gesture, eavesdropping, haptic, feedback, side-channel, attack, acoustic emanation, DTW**

I. Introduction

Mobile devices have become one of the most popular technologies in the world. In 2020 is it predicted that a total of 1.5 billion devices will be sold [1]. Such devices are used in both personal and professional contexts with users installing applications to access a variety of different services such as social media, email and instant messaging. The plethora of data communicated via these devices, however, is often of a personal nature. The sentences typed by users is implied to be confidential and for the eyes of the recipient only. However, in recent years a number of attacks have come to the fore which pose new challenges for security professionals.

Some attacks focus on software implants that can transparently collect sensitive data for egress. Such attacks have an advantage in that the attacker need not be present for the attack. However, these attacks are becoming increasingly more difficult due to the improved detection of malware, use of cryptography and enterprise security policies. Attacks that focus on the user's physical interaction with mobile devices have also been shown to be a successful vector for attack. Such attacks often rely on emissions of sounds or visual cues (e.g.: [2]). One of the vectors that can yield sound based data from which attacks may be realised is through the soft keyboard interface.

In this paper we demonstrate that it is possible for a third-party to capture audio signals from haptic feedback mechanisms emitted from gesture-typing and perform classification to identify text that has been entered into a device. This does not require direct access to the device in any form. Our attack requires only the ability to capture audio produced by the haptic feedback mechanism of a target device, which can be done with a second device in proximity to the target. We present this as a new and novel attack vector.

II. Background

A side-channel attack is one based upon non-conventional information leaks that occur due to the way in which a system is implemented, which is constrained by practical hardware limitations [3]. Attacks of this nature involve analysis of unconventional data such as timing [4], thermal imaging [5], acoustic emanations [2], [6], electromagnetic radiation [7], optical emission [8] and reflections [9]. These attacks typically enable eavesdropping of data previously assumed to be secured through higher level security mechanisms such as cryptographic algorithms.

A. Gesture-Typing

Keyboards on modern mobile devices are implemented through a touchscreen. These keyboards appear and disappear as required and can be typed upon through tapping the letter on the screen or through gesture-typing. Gesture-typing requires the user place their finger on the first letter of the word and then drag their finger to each subsequent letter in the word until they reach the last letter at which point the finger is removed and the gesture processed into a word. Thus far, only keystroke-based attacks have been implemented against mobile devices and no attack has been attempted through the sounds emitted by a gesture-typing keyboard.

B. Haptic feedback

Haptic refers to mechanisms that convey a sense of touch to users. Haptic technology is a common design element that provides useful feedback to users that input has been successfully entered. Haptic feedback for gesture-typing is common, and usually occurs as a small vibration when the user first places their finger on the screen to begin a word. Haptic feedback is usually implemented as motor vibration.

TABLE I

SENTENCES CHOSEN FOR A SMALL-SCALE TRIAL, SIMILAR WORD AND SENTENCE LENGTHS WERE USED, BUT CHARACTERS LIKE COMMAS AND APOSTROPHES WERE AVOIDED.

ID	**Sentence**	**Words**
1	The quick brown fox jumped over the lazy dog	9
2	The pin for my card is 1234	10
3	The temperature in the house is too hot	8
4	In London April is a spring month	7
5	Could I have chocolate on my cappuccino	7
6	Computer security conferences are the best	6
7	The lazy fox jumped over the quick brown dog	9
8	In summer I like to go strawberry picking	8
9	My car leaks so the rain gets in and makes it wet	12
10	The Caribbean has a great climate for a holiday	9

C. Related Work

Conventional keyboards have also been the subject of acoustic emanation attacks as shown in [2], [10]. Similarly, the use of multiple microphones concealed in a PIN-entry device is sufficient to recover input [6]. Similar techniques have been demonstrated against soft keyboards, although not gesture-based [11]. Eavesdropping attacks on gesture-typing has been demonstrated in prior work by Simon *et al.* [12] who utilised user-space permissions in Android to monitor interrupt counters for the soft keyboard. This attack requires the installation of a malicious application onto the target device, but demonstrates the ability to recognise text entered via gesture-typing by analysing the time between haptic events. Research works have shown that there are differences in the way different words are typed due to the points of pause and redirection in gesture-typed words, as seen in [13] and [12].

III. PROPOSED SCHEME

In this section we describe the data collection, pre-processing and classification approach.

A. Data Collection & Pre-processing

There is no publicly available dataset that contains the data that we require for our experiment. Therefore, we collect our own dataset in this paper. The devices used are a Moto G5S Plus (XT1803) and a Samsung Galaxy S10e (both owned by the study authors). The devices are placed on the same table in close proximity. One device is set up with a notepad application to allow typing and the other runs a recording application (downloaded via *Google Play*). One user writes sentences on one device flat on the table whilst the other initiates and ceases the recording of each sentence from the other device flat on the table. The sentences used are shown in Table I. For efficiency, due to WAV files having a high sampling rate, we down-sample the signal to use every 100^{th} value in the signal.

B. Classification

Dynamic Time Warping (DTW) is an algorithm that can be used to classify time series data where sequences may be of different lengths or contain unique events but at different times in the series. We use this because it is popular for finding similar samples of audio [14]. The technique works by warping the dimension of time such that each event in one sequence is mapped to an event in the other sequence that yields the shortest distance between the two sequences. This is achieved through the construction of a 2D matrix used to store the accumulated distance of the event-to-event comparisons. Each individual distance between two sequence events i and k is computed as $d_{i,k} = |i - k|$. This result in $N \times M$ distance values for two sequences s_1 and s_2 of lengths N and M. The accumulated cost for each event-to-event mapping is represented in the matrix by the minimum of $(i-1,k)+d_{i,k}$, $(i,k-1)+d_{i,k}$ and $(i-1,k-1)+d_{i,k}$. The time complexity for a DTW comparison is $O(NM)$.

IV. EXPERIMENTATION & RESULTS

In this section we perform experimentation following our proposed attack approach. The first experiment is designed to assess both the feasibility of this attack in a limited scenario and also to establish an optimal number of sentence samples to include in a dataset for accurate sentence classification. This experiment is performed using all 10 sentences. Given each user has 10 samples for each sentence, there are a maximum of 20 samples for each sentence. For each number of sentence samples, N_{samp}, involved in the experiment, N_{samp}-fold cross-validation is performed where one sample is held out for testing and the remaining samples from all sentences used for training. The test sample is then compared to all training samples using DTW. If the test sample is matched to a training sample of the same sentence then it is recorded as a match (and a non-match if otherwise). The final accuracy is computed as the portion of correctly matched sentences. The number of samples for each sentence (N_{samp}) is varied (from 2 to 20 in increments of 2) to assess the effect of a larger training set and to find an optimal number of training samples. For all comparisons for all sentence samples in the dataset, the computation time is recorded to identify any patterns.

The results for this experiment are shown in Figures 1 and 2. In Figure 1, it can be seen that overall a greater number of sentence samples used in the training set results in the most accurate sentence identification. When all 20 sentences are used in the experiment for the comparison, an accuracy of 70% is achieved. In Figure 2, the time taken to process each experiment for each variation of N_{samp} is shown when run on a single thread of a ThinkPad P52 laptop with an Intel Core i7-8750H CPU and 24GB of RAM. As noted by the figure, using all sentences also takes the most amount of time to process but the increase is time is linear.

The next experiment explores the results at a greater level of granularity. The experiment uses the approach that gave the most accurate results in the previous experiment; where all other sentence samples are used for comparison. Here, the number of matches and non-matches for each sentence are recorded and the accuracy is computed as before. The results for this experiment are shown in Figure 3. We can see that different sentences have varying degrees of accuracy. The most accurate sentence was sentence ID 6 (see Table I) at 95%. This

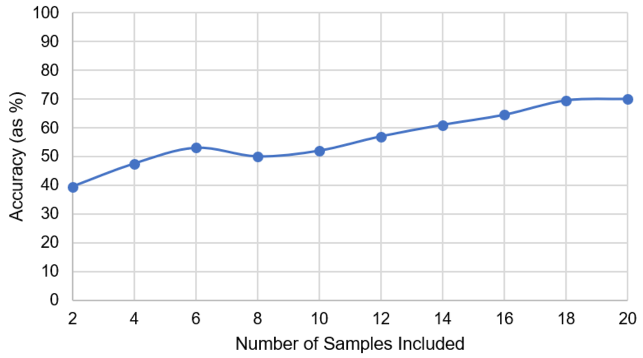

Fig. 1. The accuracy of our proposed scheme as the number of sentences included in the approach is varied. The greater the sentences in the experiment the higher the accuracy due to greater likelihood of a sentence match.

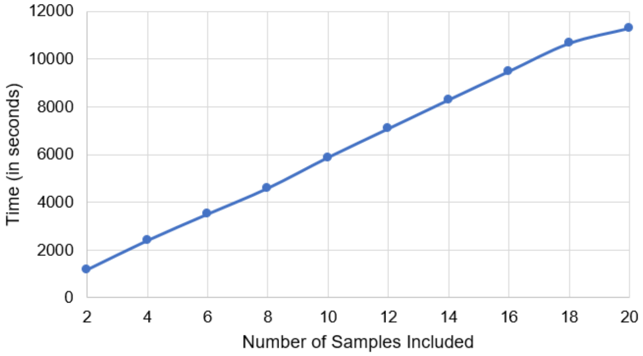

Fig. 2. The time taken to run each experiment with different numbers of sentences included. Note that although DTW is $O(NM)$, the increase in experiment complexity per sentence addition is linear due to the increase in sentences simply adding more comparisons, resulting in $O(n)$ for this aspect.

may be due to it having a unique number of very distinct words of varying length compared to other sentences.

V. Discussion

Our results demonstrate that it is possible to perform an eavesdropping attack against commonplace soft keyboards based entirely on inadvertent acoustic emanations from haptic mechanisms. This has significant consequences that must be considered by those practicing operational security.

The position of the device has an impact on the ability to detect audio emanations, we posit that if the device is laid on a flat, solid surface such as a desk, then sound resonates through the surface allowing for easier detection by an attacker device mounted on the same surface. Similarly, the typing style of the user and features like fingernail length can improve the ability to capture signals.

The most obvious cause for concern is that user input could be identified, regardless of the security (such as end-to-end encryption) of the application in use. Furthermore, the literature demonstrates the ability to identify individual authors in similar attacks [12].

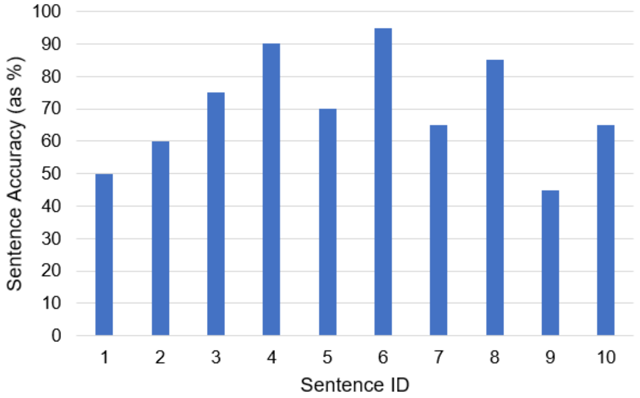

Fig. 3. The sentence accuracy for each sentence in this experiment (results generated in a setup using all sentences).

A. Scaling the Attack

Whilst our experimentation has had a number of constraints, it is possible that this is an attack that can be scaled and applied in the real-world, with significant concerns for user safety and privacy. Further research into identification of sentence fragments and high-fidelity capture of the audio emanations could enable large-scale surveillance utilising this side-channel. Classification of audio is a well understood task, driven by commercial developments [15], [16] that make the likelihood of detecting and retrieving haptic feedback in a noiser environment higher. There are a number of practical technologies easily deployed today and the state-of-the-art continues to evolve.

There is a historical record of technical ingenuity and willingness to conduct espionage. This is demonstrated by "the Great Seal Bug", a covert listening device [17] and techniques known to have been developed as part of "TEMPEST" operations based on EM side-channel attacks [18]. The ease of such attacks was revealed more publicly by van Eck [7]. Thus, the risk of such attacks being carried out by nation state actors or anyone else with the inclination, are a credible concern. Proliferation of smart speaker technology incorporating microphones provides a large potential attack surface and even equipment not traditionally intended to capture audio, such as hard drives can be reconfigured [19].

B. Counter-Measures

Haptic feedback is a useful feature of soft keyboards and it is difficult to overcome the vulnerability we have presented without getting rid of it entirely.

The most likely countermeasures involve modification of the existing haptic feedback used, which falls between being secure and being useful. The most secure, is no haptic feedback at all, which provides no useful information to the user. A less secure, but more useful approach, would be to have haptics occur at the start of word events as well as zero-motion events (such as pausing over a letter, or changing direction).

Other approaches may involve the use of additional hardware that can mitigate the ability of an attacking device to identify acoustic emanations, this may be in the form of active noise-cancelling or chaff.

More advanced electronics may reduce inadvertent audio noise though we could expect more sophisticated audio detection equipment in turn. New forms of haptic feedback with a lower audio profile might be considered, such as solid-state electrosensory feedback.

C. Future Work

The future work of this study will focus on improving the practicality of the scheme. Firstly, our work will investigate the effects on accuracy as the distance recording device is varied. We hypothesise that the further away the recording device, the lower the accuracy will be. Furthermore, we will investigate the effect of the type of different recording devices. It is possible that a high quality recording device capable of capturing fine-grained sound and subsequent pre-processing could make the attack more viable.

Research towards the identification of specific words, n-grams and sentence fragments would be necessary to utilise the attack outside of a controlled environment, where it may not be possible to ascertain the start and end of sentences. In the first stage this would involve the identification of key phrases within a larger input sample. The ability to detect sentence fragments and individual authors has already been demonstrated through analysis of software interrupt timing [12] and could be translated to the haptic side-channel.

Approaches such as Markov chains and recurrent neural networks that may provide a greater ability to predict sentence fragments than the nearest-neighbour approach carried out with DTW. This would require extraction of higher-order statistics from the original source audio.

Additional data from a wider variety of users and larger corpus of text is also required. The use of synthetic datasets generated by calculating zero-motion events for words may greatly enhance the ability to train classifiers. Of particular interest is the ability to uniquely identify individual users not present in the training data.

There is a variety of piezo, laser and bone conduction technology that may yield superior ability to capture audio in various environments. It is also known that attack performance is impacted by the properties of materials used for audio transmission [20].

VI. Conclusions

Eavesdropping attacks in the modern world traditionally focus on taking advantage of flaws in software on a target device. This paper has made a novel scientific contribution by documenting the presence of an eavesdropping risk through the analysis of acoustic emanations from haptic feedback in soft keyboards. This forgoes the need to target devices with malicious software. Our initial experimentation demonstrates there is a credible threat and we have identified key areas for further research, both to assess the scalability of the attack and to explore potential counter-measures.

References

[1] S. O'Dea, "Cell phone sales worldwide 2007-2020," Feb 2020. [Online]. Available: https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/

[2] D. Asonov and R. Agrawal, "Keyboard acoustic emanations," in *IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004*. IEEE, 2004, pp. 3–11.

[3] K. Mai, "Side channel attacks and countermeasures," in *Introduction to Hardware Security and Trust*. Springer, 2012, pp. 175–194.

[4] P. C. Kocher, "Timing attacks on implementations of Die-Hellman, RSA, DSS, and other systems," in *Advances in Cryptology— Crypto*, vol. 96, 1996, p. 104113.

[5] M. Hutter and J.-M. Schmidt, "The temperature side channel and heating fault attacks," in *International Conference on Smart Card Research and Advanced Applications*. Springer, 2013, pp. 219–235.

[6] G. de Souza Faria and H. Y. Kim, "Differential audio analysis: a new side-channel attack on PIN pads," *International Journal of Information Security*, vol. 18, no. 1, pp. 73–84, 2019.

[7] W. Van Eck, "Electromagnetic radiation from video display units: An eavesdropping risk?" *Computers & Security*, vol. 4, no. 4, pp. 269–286, 1985.

[8] J. Ferrigno and M. Hlaváč, "When AES blinks: introducing optical side channel," *IET Information Security*, vol. 2, no. 3, pp. 94–98, 2008.

[9] M. Backes, M. Dürmuth, and D. Unruh, "Compromising reflections-or-how to read LCD monitors around the corner," in *2008 IEEE Symposium on Security and Privacy (sp 2008)*. IEEE, 2008, pp. 158–169.

[10] L. Zhuang, F. Zhou, and J. D. Tygar, "Keyboard acoustic emanations revisited," *ACM Transactions on Information and System Security (TISSEC)*, vol. 13, no. 1, pp. 1–26, 2009.

[11] C. Ling, X. Hei, K. Kong, M. Peays, and M. Guizani, "You cannot sense my pins: A side-channel attack deterrent solution based on haptic feedback on touch-enabled devices," in *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016, pp. 1–7.

[12] L. Simon, W. Xu, and R. Anderson, "Don't interrupt me while I type: Inferring text entered through gesture typing on Android keyboards," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 3, pp. 136–154, 2016.

[13] M. Smith-Creasey and M. Rajarajan, "A novel word-independent gesture-typing continuous authentication scheme for mobile devices," *Computers & Security*, vol. 83, pp. 140 – 150, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404818306552

[14] A. Pikrakis, S. Theodoridis, and D. Kamarotos, "Recognition of isolated musical patterns using context dependent dynamic time warping," in *2002 11th European Signal Processing Conference*, Sep. 2002, pp. 1–4.

[15] V. Kepuska and G. Bohouta, "Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2018, pp. 99–103.

[16] A. Wang *et al.*, "An industrial strength audio search algorithm." in *Ismir*, vol. 2003. Washington, DC, 2003, pp. 7–13.

[17] G. Brooker and J. Gomez, "Lev Termen's Great Seal bug analyzed," *IEEE Aerospace and Electronic Systems Magazine*, vol. 28, no. 11, pp. 4–11, 2013.

[18] P. Rohatgi, "Electromagnetic attacks and countermeasures," in *Cryptographic Engineering*. Springer, 2009, pp. 407–430.

[19] A. Kwong, W. Xu, and K. Fu, "Hard drive of hearing: Disks that eavesdrop with a synthesized microphone," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 905–919.

[20] Q. Yan, K. Liu, Q. Zhou, H. Guo, and N. Zhang, "SurfingAttack: Interactive Hidden Attack on Voice Assistants Using Ultrasonic Guided Waves," in *NDSS*, 2020.