

# A Survey on Acoustic Side Channel Attacks on Keyboards

Alireza Taheritajar, Zahra Mahmoudpour Harris, and Reza Rahaeimehr

<sup>1</sup> Augusta University, Augusta GA, USA [ataheritajar@augusta.edu](mailto:ataheritajar@augusta.edu)

<sup>2</sup> Shariaty Technical College, Tehran, Iran [zahramahmoudpourharris@gmail.com](mailto:zahramahmoudpourharris@gmail.com)

<sup>3</sup> Augusta University, Augusta GA, USA [rרהaeimehr@augusta.edu](mailto:rרהaeimehr@augusta.edu)

**Abstract.** Most electronic devices utilize keyboards to receive inputs, including sensitive information such as authentication credentials, personal and private data, emails, plans, etc. However, these systems are susceptible to acoustic side-channel attacks. Researchers have successfully developed methods that can extract typed keystrokes from ambient noise. As the prevalence of keyboard-based input systems continues to expand across various computing platforms, and with the improvement of microphone technology, the potential vulnerability to acoustic side-channel attacks also increases. This survey paper thoroughly reviews existing research, explaining why such attacks are feasible, the applicable threat models, and the methodologies employed to launch and enhance these attacks.

**Keywords:** Acoustic Side Channel Attacks · Acoustic Signal Analysis · Keystroke Recognition · Keyboards Vulnerabilities · Keylogging · Side Channel Attacks

## 1 Introduction

With emerging cyber crimes in the 1980s, hacking and stealing data, and consequently, data privacy and confidentiality, became the new concerns for computer scientists. Since then, scientists have tried to find better mechanisms to secure systems, and hackers have attempted to break the current mechanisms or discover new vulnerabilities. Although several well-thought-out privacy and confidentiality-preserving mechanisms have been introduced in theory, achieving a perfectly secure system is impossible in practice. One reason is that every process in the world has some side effects, for instance, every process consumes some energy and may make noise, generate vibration, emit optical particles, change electromagnetic fields, produce heat, and delay the execution of other processes and all these side effects, to some extent, leak information about the process [22].

In computer science, using the side effects of a process to gain knowledge about the process is called a side-channel attack. This type of attack is applicable to any system, no matter how carefully the system is designed. In 1973, Butler W. Lampson et al. [43] discussed the possibility of abusing intermediate files and resources used or generated during a process. In 1985, a research project [73] explained the possibility of eavesdropping on display units by gathering and analyzing their electromagnetic interference. In 1996, Paul Kocher et al. [41] published the first significant attack in this area. They discovered they could break some RSA-based cryptosystems by analyzing the time it takes to do the private key operations. The attacks [4, 5, 11, 15, 21, 25, 30, 35, 36, 40–42, 44, 45, 71, 77] are just a short list of side-channel attacks that demonstrate how different side effects can be used to launch fatal attacks.

In 2004, R. Agrawal [6] successfully led an effort to use acoustic emanations of keystrokes to partially extract the data typed by mechanical keyboards and automated teller machine (ATM) pads. Since then, researchers have done tens of published works to improve the performance of

prior works or find a better way for specific scenarios. With the advancement of technology, better recording devices, and new ways to record acoustic emanations, acoustic side-channel attacks on keyboards are more feasible and accurate nowadays.

This paper results from our comprehensive literature review covering more than 170 papers. In short, we categorize attack types, briefly explain the techniques used to launch attacks, compare the success rate of each attack, and discuss the efficiency and applicability of attacks in different scenarios. This work gave us the insight to initiate two projects in this area, and we try to help researchers understand the area faster and more conveniently.

## 2 Why It Is Feasible

The phenomenon of acoustic side-channel attacks on keyboards arises from the sound and vibration that emanate from the act of pressing the keys on a keyboard. In this section, we will discuss the idea behind this kind of attack and will look into the methods that attackers use to collect acoustic sounds from users during their typing.

### 2.1 Distinguishable Sounds

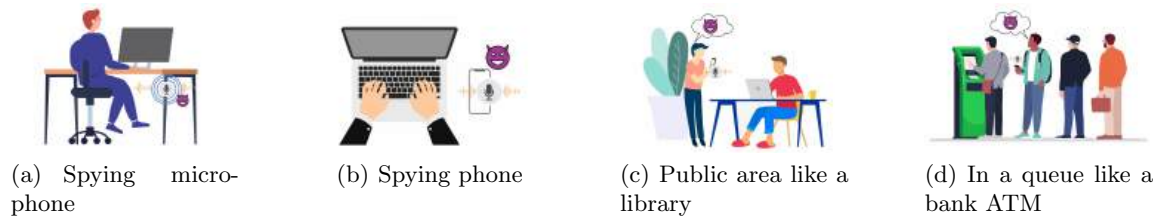
It has been observed that pressing certain keys on a keyboard, such as the Enter and Space keys, produces distinct auditory responses. Each keystroke may create some specific sounds for two main reasons: the physics of keyboards and people's typing styles. Physically there is a key plate under the keys of each keyboard. It is like playing a drum, when a person presses a key, it strikes a specific point on the key plate that generates a unique noise [6, 57]. Other reasons, like microscopic differences in keys' production and the different environments of each key, are also mentioned for the diversity of keystroke noises [6]. Type style is the other important factor. People may press different keys with different fingers, pressure, or angles, which in turn, different noises with different intensities might be generated [57]. In addition, people may type some words or syllables with certain styles that some AI techniques can exploit to launch successful attacks on specific victims. More on this and examples are explained in the following sections.

### 2.2 Practically Recordable Sounds

There are several scenarios in which attackers can record keystroke sounds. Here we categorize them into the following:

1. **Physical Proximity:** Today, mobile phones have powerful microphones that have become a good option for recording high-quality stereo sound. Some smartphones have two or three powerful microphones that are good enough to eavesdrop on keystrokes up to 1 meter [3, 8, 46, 47, 58, 74, 81].

In some papers, authors assume that the attacker is physically near the victim and, using a small microphone or a cellphone, records the victim's keystroke noises. In [81], victims work in public areas like coffee shops or libraries. In [3, 55, 74], the attacker or his allies can visit/intrude on the victim's office or leave a smartphone, a hidden microphone, or mp3 players near the victim [58]. Some authors consider a scenario where the victim and the attacker are waiting to get a service from a machine by entering a code or password on a keypad [6, 13, 27, 47, 56, 60, 68, 69]. ATM machines, point of sale (POS) devices, and digital door locks are examples of such machines.



**Fig. 1.** Examples of Physical Proximity scenarios

2. **Physical Access:** in some cases, an attacker may have physical access to the victim’s device for a limited time. For example, the victim may have a shared office with the attacker, or the device itself might be a shared device like the computers of public sites in hotels, car dealers, hospitals, banks, and governmental offices. Some special keyboards like personal identification number (PIN) pads are shared devices by nature. This scenario works for all attack methods, which only focus on the physical characteristics of audio signals and keyboards [32, 33].
3. **Indirect Physical Access:** The attacker may not gain access to the victim’s device, but he can provide a similar device to simulate the real attack and gather acoustic information produced by similar keyboard types. Although the success rate of attacks may be reduced with this approach, it is still a practical attack scenario. Note that most companies and government offices use a limited set of brands and models. It is not hard to find the common keyboard models used in a company, buy a similar keyboard, and launch experiments with them.
4. **Remote Indirect Access:** In this method, the attacker tries to record keystroke noises using a medium tool:
  - Trojans: to accomplish the attack, the attacker may create a malicious application that masquerades as a legitimate app, such as a game, virtual journal, or weather app, which is then uploaded to a typical app marketplace like Google Play and App Store [17, 18]. One can be a malicious app installed on the victim’s phone, [29, 31, 80] smartwatch [24], or computer. The adversary has different methods of installing this malicious app. Firstly, the victim may naively install the app [64, 65]. Secondly, social engineering techniques may trick the victim into installing the app. Lastly, the adversary may access the victim’s phone somewhere and download the app. Once the malicious application is installed, it requests access to microphone privileges, which users often overlook and grant to the app without question [18, 20, 79]. This allows the adversary to use the microphone to record audio without the user’s knowledge or consent. Several studies have shown that many popular apps request microphone access, and users tend to grant this access without considering the consequences. [38, 46, 53, 54, 63, 65, 66, 74].
  - Hacking third-party applications: Attackers can exploit software bugs to gain unauthorized access to a person’s computer or mobile device, allowing them to steal sensitive information or spy on the user. One way attackers can exploit software bugs is by using them to access a device’s microphone. Doing so allows them to record ambient sounds without the person’s knowledge, including keystroke noises [51]. In [2], authors can record victims’ ambient sound by using the recorder plugin for Skype named “Supertintin” [2].
  - Voice assistant devices: Like any other connected device, voice assistant devices are potentially vulnerable to spying attacks. One potential vulnerability of voice assistant devices

is that they can be hacked or compromised through vulnerabilities in their software or firmware, which could allow attackers to remotely access the device, listen to conversations, or perform malicious actions [17, 79].

5. **VoIP and Online Meetings:** Voice over Internet Protocol (VoIP) is a technology that allows voice communication to be transmitted over the Internet instead of traditional telephone lines. Recently, it has become increasingly popular as an alternative to traditional telephone services. VoIP offers a range of additional features and capabilities, such as video conferencing, screen sharing, and instant messaging. Attackers can record audio signals and keystrokes during a Voice-over-IP (VoIP) call or online meetings using software such as Zoom, Skype, or Google Hangouts without having to install malware on the victim's phone [2, 14, 19, 62]. This method differs from other methods in that the attacker can, even during a seemingly legitimate and ordinary online meeting or voice call, record the ambient noise of the victim. The attack is made possible by assuming that most users multitask during VoIP calls and online meetings and engage in sensitive activities such as sending emails, chatting, taking notes, paying financial bills, and entering PINs and random passwords. When a user presses a key, it produces a sound transmitted to the other side of the call, allowing attackers to capture the audio packets and remotely decode the keystrokes.

### 3 Attack Surface

#### 3.1 Recording Medium

Data collection is the most critical step of an acoustic side-channel attack. Without appropriate data, launching a successful attack is less feasible. In our case, the appropriate data is the clear keystroke sounds. Any other voice is considered noise. Less noisy environments like public libraries or private offices increase the chance of successful acoustic side-channel attacks. Researchers considered a variety of recording devices in the data-gathering stage:

1. **Off-the-shelf Microphones, Attacker's Cellphone or Voice Recorder:** These devices can be used when the attacker has physical access to the victim's office or when the victim works in quiet public environments like hotel lobbies or libraries. These recording devices should be placed near the victim's keyboard (less than a meter) [13, 46, 47, 49, 58, 69, 74, 81].
2. **Victim's Cellphone or Computer:** Generally, these devices are the better recording devices to gather the required data because they are the nearest thing to the victim's keyboard. People put their cell phones in close proximity to their keyboards. A laptop's internal microphone is another suitable recording device if the target is a laptop. The downside of this approach is that the attacker must compromise the victim's cellphone or laptop. But, it makes remote attacks possible and usually gathers less noisy data [8, 27, 52, 63, 66]. It should be mentioned that some researchers assume attackers capture the victim's keystroke sounds during a legitimate-looking online meeting. In this case, even though the attacker uses the victim's computer to gather the keystroke sounds, there is no need to compromise the victim's computer [14, 19].
3. **Hyperbolic or Parabolic Microphones:** These are special recording devices that capture weak sound emitting from a far source using a reflective dish (See Figure 2). The dish focuses incoming sound waves onto a small microphone to capture distant sounds with enhanced clarity. Hyperbolic microphones are commonly used in wildlife recording, surveillance, sports broadcasting, and long-range audio capture applications. They excel at picking up sounds from specific directions while reducing ambient noise and interference. In [6], the authors could launch an acoustic side-channel attack on a keyboard from a 15-meter distance.



Fig. 2. Parabolic Microphone

### 3.2 Keyboard Types

1. **Mechanical PC Keyboards:** Using a commodity mechanical keyboard to interact with PCs, laptops, and notebooks is common. This type of keyboard produces more audible noises, making them more vulnerable to acoustic side-channel attacks. Hence, most articles like [6, 10, 26, 32, 33, 38, 57, 58, 61, 74, 75, 81] have focused on them. In general, laptop keyboards have softer keys and emanate fewer noises [8, 47]. Therefore, they are less vulnerable to many attacks, and most researchers have reported lower success rates than PCs' mechanical keyboards [6, 32, 33].
2. **Virtual Keyboards:** Most modern smartphones use software-based virtual Keyboards, which can be used to enter PIN codes, authentication patterns, or text. This type of keyboard is also vulnerable to acoustic side-channel attacks. In [31, 64], the authors target the virtual keyboard of Android smartphones or tablets. Zarandy et al. implemented an attack on the virtual keyboard of an Android smartphone, which extracts the Pin codes [79]. In another work [18], the unlock pattern of Android phones was attacked. In [63], the authors designed a malicious app (like a Trojan) that legitimately grants mic access from victims, records phone calls, and extracts sensitive information like credit card numbers when victims need to enter them over a phone call. A series of works [39, 49, 54, 78] tried to extract keystrokes on touchscreen QWERTY keyboards.
3. **PIN Keypads:** PIN keypads (See Figure 3) are electronic devices used for entering personal identification numbers (PINs). They are commonly found in a wide range of security systems and payment processing applications, such as automated teller machines (ATMs), point-of-sale (POS) terminals, and door access control systems. PIN keypads typically consist of a panel of a few buttons, and people usually use only one figure to enter sensitive information one by one. Consequently, they are more attractive and vulnerable to acoustic side-channel attacks [13, 27, 56, 60].



Fig. 3. PIN Keypads

### 3.3 Target Data

Researchers have tried to increase their attack success rate by limiting the target of the attacks. Here is the list of some data types that were the target of attacks:

1. **Unlock Patterns:** A phone unlock pattern is a security feature commonly found on Android devices. It involves tracing a specific pattern on a grid of dots displayed on the phone's lock screen. This pattern serves as a passcode to unlock the device and access its contents. The papers [18] and [80] are two examples of papers that focused on Android unlocking patterns.
2. **Single Keys:** Many beginners tend to press keyboard keys individually, allowing attackers to distinguish each key's sounds clearly. Even professional typists may create a noticeable gap between keystrokes when entering one-time codes or random passwords, which enables the same thing. Assuming there are some scenarios that allow us to gather clear sound of keystrokes, several works have been focused on detecting single keys [6, 23, 26, 46, 52, 54, 61].
3. **Passwords:** Passwords, the common login credentials, are the most attractive data for most attackers. Hackers target login credentials to gain unauthorized access to various accounts, such as email, social media, banking, or corporate systems. Hence it has been the subject of many research [8, 10, 38, 66, 75, 81, 82]. Some researchers have focused on special types of passwords. For example, Panda et al. targeted only numerical passwords [56]; Halevi et al. worked on passwords consisting of lower-case English alphabets [32]; and Ponnampalath et al. investigated extracting passwords composed of English words [57].
4. **Text:** Texts may also contain sensitive or private information. Consequently, some researchers have been focused on extracting typed text from ambient noises [8, 10, 31, 38, 38, 66, 75, 81, 82, 82]. Typically excluding typos, a text consists of punctuation marks and words that can be found in a dictionary. Moreover, texts adhere to a structure and often follow some recognizable patterns. These features allow researchers to improve their method by filtering out less likely options.
5. **PIN Code:** PIN stands for Personal Identification Number. PIN codes are a particular type of password consisting of some digits; usually, the length of PIN codes in a system is fixed and known, and they are used as a convenient way for authenticating people. PIN codes are widely used in ATM and POS devices to authorize transactions and access funds, mobile devices as unlocking codes, and security systems and digital locks to grant access to restricted areas or buildings. Hence, we categorized them as an independent category. Researches like [2, 56, 63–65] focused on PIN codes.
6. **Miscellaneous:** In 2015, Anand et al. [3] designed an acoustic side-channel attack against the tapping-based rhythmic password mechanism [76], which Wobbrock had introduced in 2009. Wang et al. proposed a method to counter fast typists in 2016. When a typist presses two consecutive keys in a short interval, the signals of the two keystrokes become mixed. As a result, methods that rely on individual keystroke signals are ineffective. Their approach works by considering two combined keystrokes and utilizes the blind signal separation technique to separate the mixed signals [74].

## 4 Attack Strategies

This section explores the predominant approaches for identifying keystrokes using acoustic signals. These procedures may involve several stages, including the keystroke signal detection stage. Note that the accuracy of the keystroke signal detection stage significantly influences the overall outcome [82]. We classify the attack strategies into the following categories: timing-based, frequency-based, and Geometry-based approaches.

## 4.1 Timing-based Analysis

Timing information is a type of side-channel information related to typing patterns. Timing information includes the time between the start or end of two keystrokes, the duration of each key press, the time from keystroke push to keystroke release, and so on [27], [47]. Liu et al. [48] suggest user-independent inter-keystroke timing attacks on PINs using a human cognitive model. The findings indicate that their attack methods yielded favorable performance results.

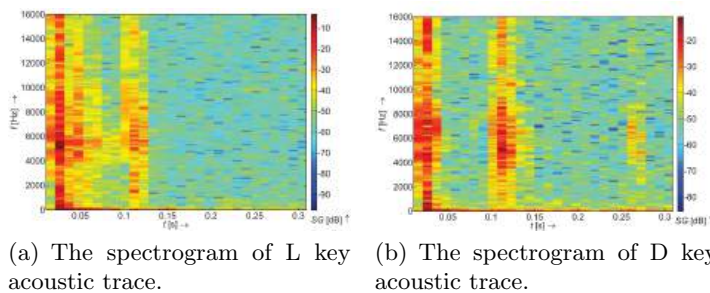
## 4.2 Frequency-based Analysis

This technique focuses on analyzing acoustic signals in the frequency domain. Typically, researchers employ the Fast Fourier Transform (FFT) to convert signals from the time domain into the frequency domain, which serves as the base of their recognition system. Additionally, researchers leverage cepstrum features, a common choice in speech analysis and recognition systems. Empirical evidence demonstrates that cepstrum features, especially mel-frequency cepstral coefficients (MFCCs) [82], tend to outperform plain FFT coefficients when processing sound signals. MFCC applies the Fourier transform to yield the frequency content of the input signal. The attackers require these features from the signal data to detect and classify each keystroke.

**Signal Processing and Statistics Techniques** In [56], Panda et al. propose a side-channel attack on a 4–6 digit random PIN key and achieve a 60% accuracy in recovering them. Their sample victims were asked to memorize a six-digit number and type it repeatedly to adapt themselves to this new PIN sequence. Then, they captured the victims’ keystroke sounds to build their dataset. They generated a feature vector of recorded signals using a Fast Fourier Transform (FFT) and used it as input for various techniques, such as Cross-Correlation and Euclidean distance. The study by Halevi et al. [33] introduces a time–frequency decoding technique for identifying passwords. It also meticulously investigates the influence of typing style on detection accuracy. The findings reveal that when employing the same typing style (hunt and peck) for both data training and decoding, the optimal success rate for accurately detecting a typed key stands at 64% per character. Furthermore, the study demonstrates that altering the typing style to touch typing during the decoding phase diminishes the success rate to approximately 40% per character. This approach can potentially decrease the entropy of the search space for random passwords by up to 57% per character.

**Classical Machine Learning-based Techniques** Machine learning is crucial in extracting pertinent features from acoustic data, facilitating more accurate key recognition. These models are trained to discern patterns within acoustic signals linked to different keystrokes, enabling them to predict the pressed keys effectively. Machine learning algorithms, including Support Vector Machines (SVM), Random Forests, and neural networks, can be trained using labeled acoustic data to classify recorded sounds into specific key presses. Furthermore, machine learning models can be deployed for real-time keystroke detection and inference, making them valuable tools that raise concerns about potential security breaches. [82] uses three different ML methods, including Neural Network, Linear classification(Discriminant), and Gaussian Mixture, to classify acoustic signals. In the study by Anand et al. [2], they utilize FFT coefficients and MFCC as distinguishing attributes for individual keystrokes. These features are employed in supervised machine learning methods, including Logistic Regression and Random Forest, to analyze random passwords and numeric PINs. In the paper by Halevi et al. [32], they employ a method based on frequency-domain features. They

utilize Mel-Frequency Cepstral Coefficients (MFCC) as inputs for neural networks and apply a 10 ms window with a 2.5 ms window step size, computing 13 MFCCs per window. They analyze a total duration of 40 ms for each keypress event. They use Matlab’s *newpnn()* function to create the neural network. In [6], a proposed attack utilizes a neural network for recognizing the pressed keys. The authors employed the JavaNNS neural network simulator to construct a backpropagation neural network. They utilized the Fast Fourier Transform (FFT) extracted from an 8–10 ms window around the key press peak as the feature. Cecconello et al. [14] employed a Logistic Regression classifier to categorize a dataset containing ten samples for each of the 26 English alphabet keys. This classification was carried out using a 10-fold cross-validation [28] approach. The authors assessed the classifier’s accuracy using various spectral features, including FFT coefficients, cepstral coefficients, and MFCC.



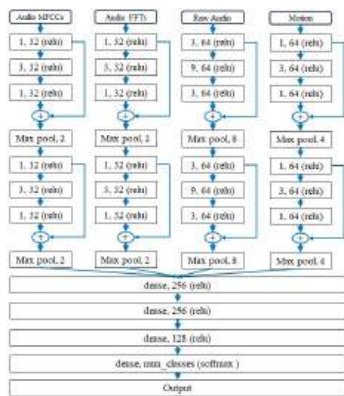
**Fig. 4.** Spectrograms depicting two distinct keystrokes [52].

**Deep Learning-based Techniques** Applying FFT to the signal gives us only frequency values, and we lose track of time information. The system won’t be able to tell what was recorded first if we use these frequencies as features. Here, Spectrograms come into the picture. The visual representation of frequencies of a given signal with time is called a Spectrogram. The idea is to break the audio signal into smaller frames(windows) and calculate the Discrete Fourier Transform (DFT) for each window. This way, we will get frequencies for each window, and the window number will represent the time. In a spectrogram representation plot, one axis represents the time, the second axis represents frequencies, and the colors represent the magnitude (amplitude) of the observed frequency at a particular time. (Bright colors represent strong frequencies). This is a 2D matrix representing the frequency magnitudes and time for a given signal. Now, think of this spectrogram as an image (e.g. Figure 4). This reduces the problem to an image classification problem. Harrison et al. [34] employed a deep learning model to classify laptop keystrokes. They captured keystroke sounds using the integrated microphone of a smartphone and utilized mel-spectrograms as input for an image classification model named *CoAtNet*. Their proposed classifier achieved an accuracy of 95% under certain conditions, all without utilizing a language model. Furthermore, they conducted training on keystroke sounds recorded through the video-conferencing software Zoom, attaining an accuracy of 93%. As shown in Figure 5 in the study by Giallanza et al. [29], researchers designed a system that combined convolutional and recurrent neural networks for keystroke detection and identification, respectively. This system was tested in an experiment involving 20 participants who



typed naturally while conversing. The findings indicated that mobile phone arrays could identify approximately 41.8% of keystrokes and 27% of typed words accurately in a noisy setting, even without personalized user training. To assess the potential security implications of this attack, the authors transformed the machine learning models into a real-time system capable of distinguishing keystrokes using an array of mobile phones. The system’s performance was evaluated through trials with a single user typing under varying conditions.

In the paper by Akinbi et al. [1], audio signals are initially transformed into audio images using the spectrogram representation of the Mel-Frequency Cepstral Coefficients (MFCC) method. Subsequently, a vision transformer-based approach called *ConvMixer* is employed on these spectrogram images to identify passwords based on their audio characteristics. The authors achieved an average accuracy score of 92.44% in their experimental work. Additionally, a comparison is made with two pre-trained convolutional neural network (CNN) models, namely *ResNet18* and *VGG16*. This comparison underscores the potential of the proposed approach as a compact acoustic keylogger system. Toreini and colleagues [72] harnessed advanced techniques in signal processing and machine learning to discern individual Enigma keys. Their experimentation involved the detection of four distinct peaks within each sound sample, employing an LPC-based method in the preprocessing stage. In the feature extraction process, they derived MFCC (Mel-frequency cepstrum coefficients) from the audio samples. The highest recognition rate of 92.18% was achieved by utilizing an Artificial Neural Network classifier. In the paper by Martinasek et al. [52], a spectrogram (as shown in Figure 4) serves as the input for a standard multi-layer neural network trained using the backpropagation learning algorithm, with no additional tools applied to enhance classification results. The recording setup involved a laptop with an integrated microphone in an office setting. The resulting average success rate for first-order classification in this experiment was 72.3%.

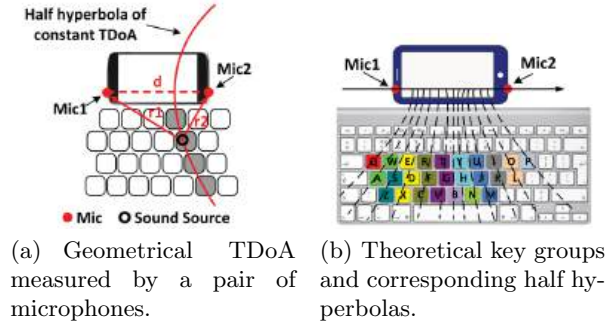


**Fig. 5.** Convolutional network architectures used for keystroke classification [29]. (\*ReLU: rectified linear unit, \*Max Pool: an operation to create a downsampled feature map)

### 4.3 Geometry-based Attacks

These techniques are based on the geometric positions of keys on keyboards and are employed to estimate the positions of pressed keystrokes [81]. A key can be identified by computing the differences

in arrival times of a sound wave at two or more microphones [26]. Geometry-based approaches can detect and identify pressed Keys Regardless of the type and meaning of the entered text, so they can work on random texts such as passwords. Using various approaches such as Time Difference of Arrival (TDoA), Triangulation, and differential audio analysis (DAA), attackers estimate the candidate area of a typed key.

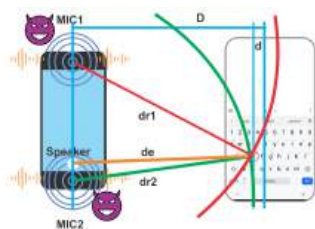


**Fig. 6.** This is an illustration of the geometrical TDoA on a single phone and the theoretical key groups [46].

**Time Difference of Arrival and Triangulation Attacks** Zhu et al. [81] introduces an innovative technique that offers context-independent keystroke recovery, even when inputted randomly. Their methodology involves employing two or more smartphones to capture acoustic emanations resulting from keystrokes [61]. This technique then processes the collected acoustic signals, employing a geometrical framework that leverages the TDoA methodology to pinpoint the physical location of the pressed key. The TDoA technique calculates the variations in the distance between the key and each microphone, ultimately enabling the estimation of a set of potential key positions on a plane. This candidate key set is progressively narrowed down through the collaborative effect of multiple microphone pairs. Their experimental results demonstrate an impressive success rate of over 72.2% in accurately recovering keystrokes. Cheng and colleagues [16] present an approach for determining keystroke positions through the TDoA technique and subsequently reconstructing words via a dictionary. This attack employs just a single phone and doesn't necessitate any training. The outcomes reveal a 45.2% success rate in correctly identifying words within the top 50 candidates across the tested vocabulary. Moreover, for words exceeding 10 characters, the success rate improves to 46.3% within the top 25 candidates. As shown in Figure 6, by employing the TDoA technique to locate keystrokes, researchers can eliminate the need for labeled training data or linguistic context. Experiments conducted with three different types of keyboards and off-the-shelf smartphones demonstrate scenarios, where this system can recover 94% of keystrokes, as reported in [46].

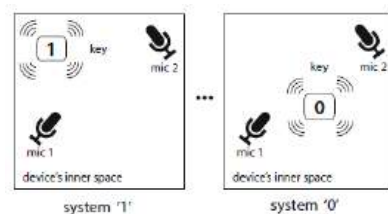
Another related approach to this kind of attack is an acoustic triangulation attack. This attack involves determining an object's position by measuring acoustic waves produced by a keystroke [12]. Triangulation is a method for determining the distance of a point through the application of triangle-based principles. The distance is considered as one of the sides of a triangle, computed by measuring its angles and other sides. Triangulation is a widely employed approach for object localization,

finding applications in fields such as surveying, navigation, and astrometry [26]. Fiona et al. [26] calculated the differences in arrival times of sound waves at two microphones installed at specific locations beside the keyboard. They employed both the maximum peak position approach and the correlation approach, achieving a recognition rate of up to 80% with a 5-minute computation. In Ranade’s study [60], the triangulation method achieved an accuracy of 87.5% in distinguishing between four keys on an ATM keypad. In the paper by Bai et al. [8], they introduce a scheme that utilizes a single smartphone with dual microphones for conducting a side-channel attack. Their approach includes an efficient environment estimation scheme to overcome challenges related to microphone positioning variability and the scarcity of training data. The proposed scheme predicts keystrokes based on their locations and achieves a correct identification rate of 91.2%.



**Fig. 7.** Foundational concepts of touchscreen keystroke localization using acoustic signal attenuation [49].

**Differential Audio Analysis** In [69], the authors introduce a method called *DAA* (differential audio analysis) for analyzing the differential attributes of sounds recorded by two microphones situated within the device’s empty space. This analysis is conducted through the transfer function between the two captured signals. The method is employed on four-digit PIN pads. The authors successfully identify all 1200 keystrokes from two independently tested devices of the same model, yielding classification rates of 100% and 99.8% for the first two models and 63% for the third model (Figure 8).

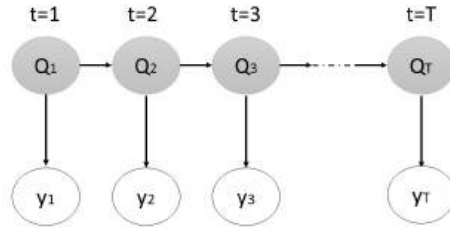


**Fig. 8.** An ensemble involving the device’s internal structure, key position, and two microphones creates a system necessitating transfer function estimation. The key classification relies on these derived transfer functions. [69].

**Phase Shifts and the Doppler Effect** As shown in the 7, the study by Lu et al. [49] explores the reduction of acoustic signal strength, demonstrating that a user’s typing fingers can be pinpointed by assessing the attenuation of acoustic signals captured by smartphone microphones. This attenuation is then harnessed to localize individual keystrokes and subsequently assess discrepancies arising from background noise. To enhance the precision of keystroke localization, *KeyListener* additionally monitors finger movements during input through phase shifts and the Doppler effect, which aids in mitigating errors linked to the acoustic signal attenuation-based localization technique. Moreover, the application incorporates a binary tree-based search strategy to deduce keystrokes contextually. This approach achieves a near 90% accuracy in inferring the correct keystroke among the top 5 candidates, accompanied by a top-5 error rate of approximately 6%.

## 5 Post Processing Technologies

This section outlines methods commonly reliant on language models to improve keystroke identification quality. These methods contribute to the enhancement of word or sentence recognition rates. In fact, this method is used to filter incorrect classified characters to improve the recognition rate [72].



**Fig. 9.** Illustration of a Hidden Markov Model (HMM) trace, with each vertical segment corresponding to a time step. Within each time step, the upper node  $q_t$  represents a character pair, while the lower node  $y_t$  represents the observable variable related to consecutive keystrokes [67].

### 5.1 Hidden Markov Models (*HMMs*)

As illustrated in Figure 9, Hidden Markov Models (*HMMs*) are a type of machine learning model. *HMMs* is a statistical model used for modeling systems with hidden states that generate observable outcomes [9, 50]. In fact, Hidden Markov models are engineered to handle data that can be represented as a ‘sequence’ of observations over time. They have applications in various fields, including speech recognition, natural language processing, bioinformatics, and more. *HMMs* are used to model data sequences, where the underlying process is assumed to be a Markov process with hidden states that influence the observed data [59]. In the scenario of an acoustic side-channel attack on a keyboard, an *HMM* is a model that undergoes training using a collection of text to anticipate the probable word or character at specific positions within a sequence. For example, if the result of an identification method is ‘clasc’, a *HMM* could be used to infer that ‘c’ was, in fact, a falsely classified ‘s’ [34]. In the work by Zhuang et al. [82], the authors employ a variant of the Viterbi

algorithm tailored for second-order *HMMs*, resembling the approach outlined in Thede et al. [70]. They have demonstrated that the language model correction greatly enhances the correct recovery rate for words. Additionally, they utilize the *HMM* in an unsupervised manner for keyboard key recognition. This approach achieves an average recovery rate of 87.6% for words and 95.7% for characters.

## 5.2 Spelling Checker

Utilizing a spell checker represents a straightforward method to capitalize on language-related insights. In the study by Zhuang et al. [82], they employed the Aspell [7] spell check tool on identified text and observed specific enhancements. Nevertheless, standard spell checkers have restrictions regarding the types of spelling errors they can address (for instance, a maximum of two incorrect letters in a word). Their functionality is optimized for typical errors made by human typists rather than the errors made by acoustic emanation classifiers. Hence, their effectiveness in this context is understandably restricted.

## 5.3 N-GRAM Language Model

The spelling correction approach discussed earlier doesn't consider word frequency relationships or grammatical concerns. For instance, certain words are more prevalent than others, and there are established rules for constructing sentences. The spelling correction process might approve "Ir fact" as accurate because "Ir" is a recognized word, even if the intended phrase is likely "In fact" [82]. A potential solution involves employing an n-gram language model that probabilistically represents word frequency and the associations between neighboring words [37]. To be precise, in [82], trigrams are integrated with the aforementioned spelling correction technique, and sentences are structured using a graphical model.

# 6 Methods and Results Comparison

In this survey, we have delved into the feasibility and mechanics of acoustic side-channel attacks on keyboards, which stem from the sounds and vibrations generated during the act of typing. We explored the distinctive sounds produced by various keys, underpinned by the physics of keyboards and the diverse typing styles of users.

Throughout our analysis, we have highlighted the practicality of these attack scenarios, demonstrating that attackers can exploit not only physical proximity but also remote and indirect methods to record keystroke sounds. This poses a significant threat to the security and privacy of users, as attackers can capture sensitive information without direct physical access to the victim's device.

We conducted a thorough analysis of different methods used to launch acoustic side-channel attacks on keyboards. The presented comparison tables offer insights into the landscape of both timing-based and geometry-based attack strategies (Table 1), as well as signal processing, classic machine learning-based, and deep learning-based approaches (Table 2). We evaluated key features and performance metrics across various categories to gain insights into the diverse strategies used by researchers in this field. It is evident that the choice of methodology and key features greatly influences the accuracy of these attacks, with some achieving remarkably high success rates. However, each approach also comes with its own set of limitations, such as dependencies on typing style, keyboard models, environmental noise, and the need for sizable training datasets. These are the typical constraints that generally affect most methods at their core:

- **Typing Style Dependency:** Regardless of the approach, the accuracy of acoustic side-channel attacks is inherently tied to the typing style of the user, making it challenging to generalize across different users.
- **Keyboard Model Dependency:** All approaches exhibit a degree of dependency on the specific keyboard model being used, limiting their applicability to various keyboard types.
- **Environmental Noise Sensitivity:** The presence of environmental noise can significantly impact the accuracy of these attacks, posing a common challenge across methodologies.
- **Training Data Requirements:** Many of these methods require substantial training datasets, which may not always be readily available or feasible to collect.

The timing-based analysis involves using a range of techniques, such as hidden Markov models, cross-correlation, and distance metrics, to identify patterns. Usually, there are common limitations to these techniques, such as typing errors, ecological validity <sup>4</sup>, and handling multiple identical time intervals. Geometry-based approaches introduce innovative techniques like Time Difference of Arrival (TDOA) and acoustic triangulation, demonstrating commendable keystroke recovery rates, albeit with some restrictions regarding precise key positioning and acoustic signal source angles.

Signal processing methods, classic machine learning, and deep learning-based approaches emphasize the significance of feature extraction, including FFT coefficients, MFCC features, and spectrograms. While classic machine learning models like hidden Markov models and support vector machines exhibit robust performance, deep learning models, such as convolutional and recurrent neural networks, achieve promising accuracy rates, often surpassing 90% in certain situations. Furthermore, the limitations associated with each technique, such as ecological validity and recognition accuracy, should be carefully considered in real-world scenarios.

This survey equips researchers, practitioners, and security experts with a comprehensive understanding of the current state of acoustic side-channel attacks on keyboards. By illuminating the strengths, weaknesses, and potential challenges posed by these approaches, we aim to encourage further research and innovation in the realm of keyboard security. As the digital landscape continues to evolve, safeguarding against acoustic side-channel attacks remains a critical endeavor, and this survey serves as a resource in that pursuit.

---

<sup>4</sup> Ecological validity measures of how well a test predicts behavior in real-world scenarios.

**Table 1.** Comparison of Timing-Based and Geometry-Based Acoustic Side Channel Attack Approaches on Keyboards

Approach	Cite	Methodology	Key Features	Accuracy	Limitations	Attack Target Data
Timing-based Analysis	[27]	Hidden Markov Model	The time interval between keystrokes	search space is reduced by at least two orders of magnitude	Multiple of the same time intervals	ATM and door keypads
	[47]	Cross-correlation	Inter-keystroke timing dictionary built from a human cognitive model	top-3, top-10, top-25, top-50, top-100 = 75%, 80%, 81%, 83%, 88%	Typing styles, Typing errors, Ecological validity	PIN
	[56]	Cross-Correlation	Hold time, release time, and time interval between keystrokes	60% chance to re-cover the PIN key	The size of the training dataset	4-6 digit random PIN
	[48]	Cosine similarity	Inter-keystroke timing dictionary	guessing the correct PIN is above 10%.	Typing styles, Typing errors, Ecological validity	PIN
Geometry-based approaches	[81]	Time Difference of Arrival	Acoustic Signal	recover 72.2% of Keystrokes	Angle of Smartphones(sound recorder)	Keyboard
	[61]			86% accuracy	number of sound recorder devices	Keyboard
	[46]			recover 94% of keystrokes	the sampling rate, Microphone Placement	Keyboard
	[16]			a 45.2%success rate	Microphone's Placement, Word length	Keyboard
	[26]	Acoustic Triangulation		80%accuracy	cannot make precise judgments on the positions of close keys	Keyboard
	[60]			87.5% accuracy	Microphone resolution	ATM keypad
	[8]			correctly identify 91.2% of keystrokes with 10-fold cross-validation	keyboard model	keyboard text
	[69]	Differential Audio Analysis		A classification rate of 99.8%.	Non-audible Pin pads	PIN pads
	[49]	Phase Shifts and the Doppler Effect		90% on a top-5	Microphone's placement	keyboards of touch screen

**Table 2.** Comparison of Frequency-Based Acoustic Side Channel Attack Approaches on Keyboards

Appr.	Cite	Methodology	Key Features	Accuracy	Limitations	Attack Target Data
Signal Processing	[57]	Distance Metric Approach	FFT and raw audio signals	60% recognition rate	Typing styles, Keyboard model changing	Password
	[82]	Hidden Markov Model	cepstrum features	96%, 90%, 80%	Constraints of English language	English text and random passwords on keyboard
Classic Machine Learning-based	[75]	K-means	coefficients of the Mel-frequency Cepstrum	90%	Typing speed, Ambient noise	keyboard text
	[74]	Support Vector Machines	Blind Source Separation (BSS) and Independent Component Analysis (ICA), and MFCC features	78.4% recognition accuracy	Microphone's placement	Two combined keystrokes
	[2]	Logistic Regression, Random Forest, Linear Nearest Neighbor Search	FFT coefficients and MFCC	accuracy of 74.33%	Noisy Audio Signals	Random passwords, PINs
	[58]	Support Vector Machines anomaly detection and Back Propagation Neural Network	spectrum features	99.47% keystroke detection rate, a 97.27% recognition accuracy and 84.55% content recovery accuracy	Combined keys	keyboard text
	[32]	Time-frequency decoding	MFCC features	detecting correctly, the typed key is 64% per character	Typing style	Random passwords
	[6]	JavaNNS neural network	time-FFT	recognized correctly 79% of clicks	Silent keyboards, touchscreen or touch stream	ATM keypads
	[10]	Cross-correlation	FFT Coefficients	90% of finding the correct word	Sound duration, word length, repeated characters, Shift keys, punctuation marks and digits	7-13 characters from typing on a keyboard
	[19]	Logistic Regression	Mel-frequency cepstral coefficients (MFCC)	top-5 accuracy of 91.7%	Typing style and keyboard model	random key pressed on keyboard
	[14]	Logistic Regression	MFCC	top-5 accuracy of 91.7%	keyboard model	random key pressed on keyboard
	Deep Learning-based	[34]	Self-attention transformer	Mel-Spectrograms	95% accuracy	typing style and randomized passwords
[29]		Convolutional and recurrent neural networks	Cepstrum	ability to detect 41.8% of keystrokes and 27% of words	Microphone's placement, and keyboard models	Keyboard
[1]		Vision transformer based (ConvMix)	Spectrogram	92.44% accuracy	The number of samples	Passwords
[72]		Neural Networks	MFCC coefficients	success rate of 84%	-	Enigma keyboard
[38]		Hidden Markov Model and neural network newpnn() and Spellchecker	FFT and Cepstrum	81.47%	lack of public dataset	Keystroke sequence
[66]		Recurrent neural network	FFT	92.60%	keyboard Model, Environmental noise and microphone placement	Keyboard text
[52]		Multi-layer perceptron	Spectrogram	72.3% success rate	Size of the spectrogram matrix	Keyboard



## References

1. Akinbi, A., Deniz, E., Ismael, A.M., Rashid, Z.N., Sengur, A.: Password-sniffing acoustic keylogger using machine learning. Available at SSRN 4431909 (2023)
2. Anand, S.A., Saxena, N.: Keyboard emanations in remote voice calls: Password leakage and noise (less) masking defenses. In: Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy. pp. 103–110 (2018)
3. Anand, S.A., Shrestha, P., Saxena, N.: Bad sounds good sounds: Attacking and defending tap-based rhythmic passwords using acoustic signals. In: International Conference on Cryptology and Network Security. pp. 95–110. Springer (2015)
4. Angel, S., Kannan, S., Ratliff, Z.: Private resource allocators and their applications. In: 2020 IEEE Symposium on Security and Privacy (SP). pp. 372–391. IEEE (2020)
5. Ashokkumar, C., Giri, R.P., Menezes, B.: Highly efficient algorithms for aes key retrieval in cache access attacks. In: 2016 IEEE European symposium on security and privacy (EuroS&P). pp. 261–275. IEEE (2016)
6. Asonov, D., Agrawal, R.: Keyboard acoustic emanations. In: IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004. pp. 3–11. IEEE (2004)
7. Atkinson, K.: “aspell (2005)
8. Bai, J.X., Liu, B., Song, L.: I know your keyboard input: A robust keystroke eavesdropper based on acoustic signals. In: Proceedings of the 29th ACM International Conference on Multimedia. pp. 1239–1247 (2021)
9. Baum, L.E., Petrie, T.: Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics* **37**(6), 1554–1563 (1966)
10. 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 (2006)
11. Brumley, D., Boneh, D.: Remote timing attacks are practical. *Computer Networks* **48**(5), 701–716 (2005)
12. Canistraro, H.A., Jordan, E.H.: Projectile-impact-location determination: an acoustic triangulation method. *Measurement Science and Technology* **7**(12), 1755 (1996)
13. Cardaioli, M., Conti, M., Balagani, K., Gasti, P.: Your pin sounds good! on the feasibility of pin inference through audio leakage. arXiv preprint arXiv:1905.08742 (2019)
14. Ceconello, S., Compagno, A., Conti, M., Lain, D., Tsudik, G.: Skype & type: Keyboard eavesdropping in voice-over-ip. *ACM Transactions on Privacy and Security (TOPS)* **22**(4), 1–34 (2019)
15. Chen, S., Wang, R., Wang, X., Zhang, K.: Side-channel leaks in web applications: A reality today, a challenge tomorrow. In: 2010 IEEE Symposium on Security and Privacy. pp. 191–206. IEEE (2010)
16. Cheng, K., Li, W., Zhang, L., Ma, X., Chen, J.: Dictionary attacks based on tdoa using a smartphone. In: 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). pp. 154–159. IEEE (2022)
17. Cheng, P.: Acoustic-channel attack and defence methods for personal voice assistants. Lancaster University (United Kingdom) (2020)
18. Cheng, P., Bagci, I.E., Roedig, U., Yan, J.: Sonarsnoop: Active acoustic side-channel attacks. *International Journal of Information Security* **19**(2), 213–228 (2020)
19. Compagno, A., Conti, M., Lain, D., Tsudik, G.: Don’t skype & type! acoustic eavesdropping in voice-over-ip. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. pp. 703–715 (2017)
20. Das, A., Borisov, N., Caesar, M.: Do you hear what i hear? fingerprinting smart devices through embedded acoustic components. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. pp. 441–452 (2014)
21. Das, D., Golder, A., Danial, J., Ghosh, S., Raychowdhury, A., Sen, S.: X-deepsca: Cross-device deep learning side channel attack. In: Proceedings of the 56th Annual Design Automation Conference 2019. pp. 1–6 (2019)

22. Deepa, G., SriTeja, G., Venkateswarlu, S.: An overview of acoustic side-channel attack. *International Journal of Computer Science & Communication Networks* **3**(1), 15 (2013)
23. Ellis, S.: Predicting keystrokes using an audio side-channel attack and machine learning. <https://github.com/sam4llis/Predicting-Keystrokes>, accessed: 2023-09-26
24. Fang, S., Markwood, I., Liu, Y., Zhao, S., Lu, Z., Zhu, H.: No training hurdles: Fast training-agnostic attacks to infer your typing. In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. pp. 1747–1760 (2018)
25. Ferrigno, J., Hlaváč, M.: When aes blinks: introducing optical side channel. *IET Information Security* **2**(3), 94–98 (2008)
26. Fiona, A.H.Y.: Keyboard acoustic triangulation attack. The Chinese University Of Hong Kong: bachelors thesis (2006)
27. Foo Kune, D., Kim, Y.: Timing attacks on pin input devices. In: *Proceedings of the 17th ACM conference on Computer and communications security*. pp. 678–680 (2010)
28. Fushiki, T.: Estimation of prediction error by using k-fold cross-validation. *Statistics and Computing* **21**, 137–146 (2011)
29. Giallanza, T., Siems, T., Smith, E., Gabrielsen, E., Johnson, I., Thornton, M.A., Larson, E.C.: Keyboard snooping from mobile phone arrays with mixed convolutional and recurrent neural networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **3**(2), 1–22 (2019)
30. Golder, A., Das, D., Danial, J., Ghosh, S., Sen, S., Raychowdhury, A.: Practical approaches toward deep-learning-based cross-device power side-channel attack. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* **27**(12), 2720–2733 (2019)
31. Gupta, H., Sural, S., Atluri, V., Vaidya, J.: Deciphering text from touchscreen key taps. In: *IFIP Annual Conference on Data and Applications Security and Privacy*. pp. 3–18. Springer (2016)
32. 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 (2012)
33. Halevi, T., Saxena, N.: Keyboard acoustic side channel attacks: exploring realistic and security-sensitive scenarios. *International Journal of Information Security* **14**(5), 443–456 (2015)
34. Harrison, J., Toreini, E., Mehrnezhad, M.: A practical deep learning-based acoustic side channel attack on keyboards. In: *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. pp. 270–280. IEEE (2023)
35. Inci, M.S., Gulmezoglu, B., Irazoqui, G., Eisenbarth, T., Sunar, B.: Cache attacks enable bulk key recovery on the cloud. In: *International Conference on Cryptographic Hardware and Embedded Systems*. pp. 368–388. Springer (2016)
36. Irazoqui, G., Inci, M.S., Eisenbarth, T., Sunar, B.: Wait a minute! a fast, cross-vm attack on aes. In: *International Workshop on Recent Advances in Intrusion Detection*. pp. 299–319. Springer (2014)
37. Jurafsky, D.: Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. *Probabilistic linguistics* **21**, 1–30 (2003)
38. Kelly, A.: Cracking passwords using keyboard acoustics and language modeling. University of Edinburgh, <http://citeseerx.ist.psu.edu/viewdoc/download> p. 54 (2010)
39. Kim, H., Joe, B., Liu, Y.: Tapsnoop: Leveraging tap sounds to infer tapstrokes on touchscreen devices. *IEEE Access* **8**, 14737–14748 (2020)
40. Kocher, P., Horn, J., Fogh, A., Genkin, D., Gruss, D., Haas, W., Hamburg, M., Lipp, M., Mangard, S., Prescher, T., et al.: Spectre attacks: Exploiting speculative execution. *Communications of the ACM* **63**(7), 93–101 (2020)
41. Kocher, P.C.: Timing attacks on implementations of diffie-hellman, rsa, dss, and other systems. In: *Annual International Cryptology Conference*. pp. 104–113. Springer (1996)
42. Kovacs, E.: Security week. “Hard drive LED allows data theft from air-gapped PCs (2015)
43. Lampton, B.W.: A note on the confinement problem. *Communications of the ACM* **16**(10), 613–615 (1973)

44. Lerman, L., Bontempi, G., Markowitch, O., et al.: Power analysis attack: an approach based on machine learning. *Int. J. Appl. Cryptogr.* **3**(2), 97–115 (2014)
45. Lipp, M., Schwarz, M., Gruss, D., Prescher, T., Haas, W., Fogh, A., Horn, J., Mangard, S., Kocher, P., Genkin, D., Yarom, Y., Hamburg, M.: Meltdown: Reading kernel memory from user space. In: 27th USENIX Security Symposium (USENIX Security 18) (2018)
46. 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 (2015)
47. LIU, X.: When keystroke meets password: Attacks and defenses. Singapore Management University (2019)
48. Liu, X., Li, Y., Deng, R.H., Chang, B., Li, S.: When human cognitive modeling meets pins: User-independent inter-keystroke timing attacks. *Computers & Security* **80**, 90–107 (2019)
49. Lu, L., Yu, J., Chen, Y., Zhu, Y., Xu, X., Xue, G., Li, M.: Keylistener: Inferring keystrokes on qwerty keyboard of touch screen through acoustic signals. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications. pp. 775–783. IEEE (2019)
50. Maleki, H., Valizadeh, S., Koch, W., Bestavros, A., van Dijk, M.: Markov modeling of moving target defense games. In: Proceedings of the 2016 ACM Workshop on Moving Target Defense. p. 81–92. MTD '16, Association for Computing Machinery, New York, NY, USA (2016). <https://doi.org/10.1145/2995272.2995273>, <https://doi.org/10.1145/2995272.2995273>
51. Marques, D., Muslukhov, I., Guerreiro, T., Carrico, L., Beznosov, K.: Snooping on mobile phones: Prevalence and trends. In: Twelfth Symposium on Usable Privacy and Security (SOUPS 2016). vol. 2, p. 77. USENIX Association (2016)
52. Martinasek, Z., Clupek, V., Trasy, K.: Acoustic attack on keyboard using spectrogram and neural network. In: 2015 38th International Conference on Telecommunications and Signal Processing (TSP). pp. 637–641. IEEE (2015)
53. Monaco, J.V.: Sok: Keylogging side channels. In: 2018 IEEE Symposium on Security and Privacy (SP). pp. 211–228. IEEE (2018)
54. Narain, S., Sanatinia, A., Noubir, G.: Single-stroke language-agnostic keylogging using stereo-microphones and domain specific machine learning. In: Proceedings of the 2014 ACM conference on Security and privacy in wireless & mobile networks. pp. 201–212 (2014)
55. Neale, G., et al.: Investigating the feasibility of keyboard acoustic attacks. Department of Computer Science—University of Auckland, <https://www.cs.auckland.ac.nz/courses/compsci725s2c/archive/termpapers/gneale.pdf> p. 12 (2006)
56. Panda, S., Liu, Y., Hancke, G.P., Qureshi, U.M.: Behavioral acoustic emanations: Attack and verification of pin entry using keypress sounds. *Sensors* **20**(11), 3015 (2020)
57. Ponnamp, S.: Keyboard acoustic emanations attack: An empirical study (2013)
58. Qin, Z., Du, J., Han, G., Yong, G., Guo, L., Wang, L.: Lol: localization-free online keystroke tracking using acoustic signals. *Soft Computing* **23**(21), 11063–11075 (2019)
59. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**(2), 257–286 (1989)
60. Ranade, V., Smith, J., Switala, B.: Acoustic side channel attack on atm keypads (2009)
61. Rosmansyah, Y., et al.: The microphone array sensor attack on keyboard acoustic emanations: Side-channel attack. In: 2017 International Conference on Information Technology Systems and Innovation (ICITSI). pp. 261–266. IEEE (2017)
62. Sabra, M., Maiti, A., Jadliwala, M.: Zoom on the keystrokes: exploiting video calls for keystroke inference attacks. arXiv preprint arXiv:2010.12078 (2020)
63. Schlegel, R., Zhang, K., Zhou, X.y., Intwala, M., Kapadia, A., Wang, X.: Soundcomber: A stealthy and context-aware sound trojan for smartphones. In: NDSS. vol. 11, pp. 17–33 (2011)
64. Shumailov, I., Simon, L., Yan, J., Anderson, R.: Hearing your touch: A new acoustic side channel on smartphones. arXiv preprint arXiv:1903.11137 (2019)

65. Simon, L., Anderson, R.: Pin skimmer: inferring pins through the camera and microphone. In: Proceedings of the Third ACM workshop on Security and privacy in smartphones & mobile devices. pp. 67–78 (2013)
66. Slater, D., Novotney, S., Moore, J., Morgan, S., Tenaglia, S.: Robust keystroke transcription from the acoustic side-channel. In: Proceedings of the 35th Annual Computer Security Applications Conference. pp. 776–787 (2019)
67. Song, D.X., Wagner, D., Tian, X.: Timing analysis of keystrokes and timing attacks on {SSH}. In: 10th USENIX Security Symposium (USENIX Security 01) (2001)
68. de Souza Faria, G., Kim, H.Y.: Identification of pressed keys by acoustic transfer function. In: SMC. pp. 240–245 (2015)
69. de Souza Faria, G., Kim, H.Y.: Differential audio analysis: a new side-channel attack on pin pads. *International Journal of Information Security* **18**(1), 73–84 (2019)
70. Thede, S.M., Harper, M.: A second-order hidden markov model for part-of-speech tagging. In: Proceedings of the 37th annual meeting of the Association for Computational Linguistics. pp. 175–182 (1999)
71. Timon, B.: Non-profiled deep learning-based side-channel attacks with sensitivity analysis. *IACR Transactions on Cryptographic Hardware and Embedded Systems* pp. 107–131 (2019)
72. Toreini, E., Randell, B., Hao, F.: An acoustic side channel attack on enigma. *School of Computing Science Technical Report Series* (2015)
73. Van Eck, W.: Electromagnetic radiation from video display units: An eavesdropping risk? *Computers & Security* **4**(4), 269–286 (1985)
74. Wang, J., Ruby, R., Wang, L., Wu, K.: Accurate combined keystrokes detection using acoustic signals. In: 2016 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN). pp. 9–14. *IEEE* (2016)
75. Wit, E., Houtenbos, T.: All your keystrokes are belong to us. *academia.edu* (2014)
76. Wobbrock, J.O.: Tapsongs: tapping rhythm-based passwords on a single binary sensor. In: Proceedings of the 22nd annual ACM symposium on User interface software and technology. pp. 93–96 (2009)
77. Yarom, Y., Falkner, K.: {FLUSH+ RELOAD}: A high resolution, low noise, l3 cache {Side-Channel} attack. In: 23rd USENIX security symposium (USENIX security 14). pp. 719–732 (2014)
78. Yu, J., Lu, L., Chen, Y., Zhu, Y., Kong, L.: An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing. *IEEE Transactions on Mobile Computing* **20**(2), 337–351 (2019)
79. Zarandy, A., Shumailov, I., Anderson, R.: Hey alexa what did i just type? decoding smartphone sounds with a voice assistant. *arXiv preprint arXiv:2012.00687* (2020)
80. Zhou, M., Wang, Q., Yang, J., Li, Q., Xiao, F., Wang, Z., Chen, X.: Patternlistener: Cracking android pattern lock using acoustic signals. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. pp. 1775–1787 (2018)
81. Zhu, T., Ma, Q., Zhang, S., Liu, Y.: Context-free attacks using keyboard acoustic emanations. In: Proceedings of the 2014 ACM SIGSAC conference on computer and communications security. pp. 453–464 (2014)
82. Zhuang, L., Zhou, F., Tygar, J.D.: Keyboard acoustic emanations revisited. *ACM Transactions on Information and System Security (TISSEC)* **13**(1), 1–26 (2009)