SJTU, Shanghai, China
Course Report for Wireless Communication Technology

Defense Against Inaudible Attack

Tiankai Zheng
515021910589

May, 2018

Abstract

Voice controllable systems (VCS) can be used for convenient remote controlling. Although it is now very popular and considered to perform well in identifying, there exists an obvious problem in safety which is called the inaudible attack. Typically, microphones use LPF to ban ultrasound, but the non-linearity will still make it available to attack VCS.

In this course, my teammate and me worked on defense against this kind of inaudible attack based on existing work, simulated both on computer and the real circuit.

Section 1 gives a brief introduction of inaudible attack, section 2 explains the feasibility of the attack and section 3 shows how to defend the attack. Section 4 and section 5 shows my current work and plans in the future.

keywords: Ultrasound; Fourier transform; Attack; Defense

1 Introduction

Nowadays, due to the fast development of technology, hundreds of human-computer interaction mechanism that were considered incredible in the past have emerged, for example, the voice controllable systems(VCS). This technology is now providing us people a wide range of benefits, both in efficiency and comfort. However, every coin has two sides, security problem in VCS should also be taken into account.

The previous work Backdoor [1] showed that using non-linearity effect, it is possible for microphones to hear inaudible sounds, which could lead to serious security problems. It exploited how to generate a word ”shadow” which can be recorded by microphones using simply 2 ultrasound waves, and also worked on inaudible communications. The DolphinAttack [2] succeeded in attacking using inaudible commands by leveraging the self-demodulate feature of the non-linearities. It also gave a primitive solution for this using SVM classifier, but was kind of unreliable.

This report aims at studying the basic theorem of general inaudible attacks, and giving the core idea of the practical solution against inaudible attacks. Details will be discussed in later sections.

2 Inaudible Attack Theorem

2.1 Non-linearity Effect

The base of inaudible attack theorem is called the non-linearity effect.

Amplifier(AMP) is an important component in microphones. After passing this part, weak signals from surroundings can be amplified, thus become available for ADC and later parts to analyze and recognize. However, AMP also take primary responsibility of the non-linearity effects.

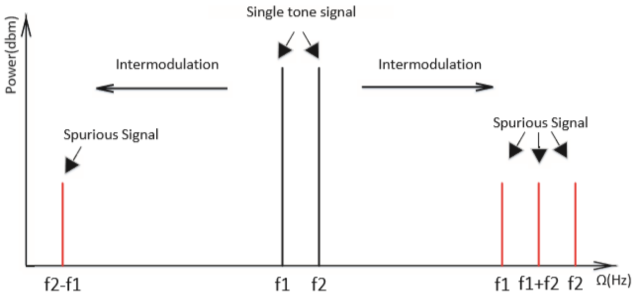


Figure 1: Non-linearity in AMP

Non-linearity effect, in short, is that original tones can be moved and recombined in frequency after passing the microphone, especially the key component AMP, as shown in figure 1. An ideal AMP will directly amplify the input signal without any distortion, which can be written as:

$$V_{out}(t) = GV_{in}(t)$$

In which G denotes the amplification factor.

Yet practically, AMP always suffer from non-linearity effects, thus extra frequency components will be also generated which maybe harmful. Below is the simplest model of a general non-linearity effect:

$$V_{out}(t) = G_0 + G_1 V_{in}(t) + G_2 V_{in}^2(t) + G_3 V_{in}^3(t) + G_4 V_{in}^4(t) + ...$$

Since high-order terms are always relatively weak, only 1-order and 2-order term will be considered. Define $S_{out}(t)$ as the output signal and $S_{in}(t)$ as the input signal, the further simplified model can be written as:

$$S_{out}(t) = A_1 S_{in}(t) + A_2 S_{in}^2(t) \tag{1}$$

Above is the non-linearity model we consider in Mic system. Now we will turn to the feasibility of an inaudible attack.

2.2 Make Ultrasound Audible

In order to make the non-linearity effect into practical uses, now we consider a specified ultrasound wave. Suppose we have a human voice command $m(t)$, an AM modulation will be used to generate the original attacking signal, mathematically:

$$S_{in}(t) = (m(t) + \alpha) * sin(\omega_c t)$$

After passing through the diaphragm and AMP of the microphone, the output signal $S_{out}(t)$ can be calculated as follow mathematically:

$$\begin{aligned} S_{out}(t) =& A_1 S_{in}(t) + A_2 S_{in}^2(t) \\ =& \frac{A_2}{2}\alpha^2 + A_2 \alpha m(t) + \frac{A_2}{2} m^2(t) - \frac{A_2}{2}\alpha^2 cos(2\omega_c t) \\ & - A_2 \alpha m(t) cos(2\omega_c t) - \frac{A_2}{2} m^2(t) cos(2\omega_c t) \\ & + A_1 m(t) sin(\omega_c t) + A_1 \alpha sin(\omega_c t) \end{aligned} \tag{2}$$

Filtering by low pass filter and anti-DC:

$$S_{out}(t) = A_2 \alpha m(t) + \frac{A_2}{2} m^2(t) \tag{3}$$

Obviously, the processed output signal $S_{out}(t)$ contains the complete attack signal $m(t)$, thus it is possible to generate an inaudible attack.

3 Attack Cancellation

3.1 Core Idea

Although inaudible attack can possibly achieve the same effect as real commands, by no means will inaudible attack have complete same characteristics as real ones. Therefore, we can do the defense base on analysis in spectral diagrams.

Consider that the original attack signal has a main frequency component of over 20kHz which is different from the human voice signal, we may choose a single tone which is also over 20kHz to produce some interactions. A good news is, human voice only use the band [0Hz, 4kHz], which indicates that only 8kHz bandwidth have to be occupied to contain an voice modulated signal. We will mainly use the band [10kHz, 20kHz] to do attack detections.

Our main idea is using a single tone to move the attack signal to the [10kHz, 20kHz] band. Using a single tone with frequency ω_1 to couple with the attack signal as the new input signal, we may rewrite the equation 2 as follow:

$$\begin{aligned} S_{out}(t) =& \frac{A_2}{2}(m^2(t)+1)^2 + \frac{A_2}{2}(m^2(t)+1)^2 cos(2\omega_c t) \\ &+ \frac{A_2}{2}(m^2(t)+1)cos[(\omega_1-\omega_c)t] + \frac{A_2}{2} \\ &+ \frac{A_2}{2}(m^2(t)+1)cos[(\omega_1+\omega_c)t] \end{aligned} \tag{4}$$

With some simple adjustments, we can get $\omega_1 - \omega_c \in [10kHz, 20kHz]$. After successfully shifting, the cancellation would be also doable. Now we will give the practical cancellation scheme.

3.2 General Cancellation Scheme

The general cancellation scheme is designed for traditional mobile devices(with low sampling rate). There are mainly 2 steps in our scheme, which is shown in figure 2. Below is the complete scheme:

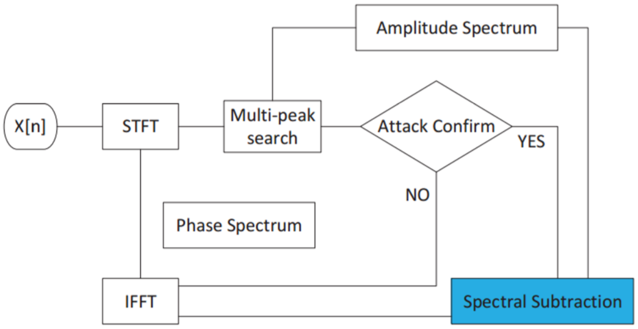


Figure 2: The Process of General Cancellation Scheme

1. **The Multi-tone defense:** Simply move the ultrasound into the band [10kHz, 20kHz] is not enough since you are not able to tell whether the attack signal will use only the frequency next to this band, thus a multi-tone design will be used. Our chosen multi-tones are: [23kHz, 43kHz, 64kHz, 86kHz...]. The reason for choosing will be discussed later.

2. **STFT and IFFT:** Consider that the voice control device is a continuous sensing device, we have to use time-efficient scheme. Therefore, since it is nearly impossible to

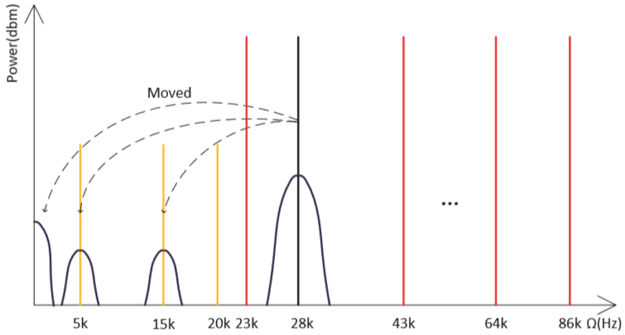


Figure 3: The Process of General Cancellation Scheme

do offset directly in time domain, short time Fourier transform(STFT) with windowing was applied in our design. The advantage of STFT is that we can can achieve a fewer spectrum leakage and high efficiency in time.

3. **Attack Detection:** STFT help obtaining the amplitude and phase spectrum of each frame. Thanks to the multi-tone design, we can obtain 2 shifted frequency component $f_{\omega 1}$ and $f_{\omega 2}$, where $f_{\omega 1} + f_{\omega 2} = 20kHz$, as shown in figure 3. Once we obtain this characteristic, i.e. $f_{\omega 1} + f_{\omega 2} = 20kHz$, we can judge it attack. In detail, we will find the peak values before adding them up since there would be also weak noises, and also set a tolerate bar δf to make it more robust.

4. **Spectral Subtraction:** To achieve the cancellation, we can simply subtract the attack recovered signal by shifted signal in frequency domain. Define $|X_{amp}(k)|$ as the amplitude at kHz, P_h as the peak in [10kHz, 20kHz], the subtracting scheme can be written as:

$$|\bar{X}_{sub}(k)| = \begin{cases} ||X_{amp}(k)| - a|X_{amp}(sum-k)||, (k < sum/2) \\ ||X_{amp}(k)| - b|X_{amp}(P_h+k)||, (k < 4000) \end{cases} \tag{5}$$

Till now, the general attack cancellation scheme is complete. Next, we will discuss about the special design for new devices.

3.3 Cancellation for New Mic System

With the development in technology, nowadays many novel works on acoustic application have sprang up for that they have lower power consumption and are easy to implement, which means that 20kHz cannot meet the need of increasing acoustic applications. Meanwhile, mobile devices are going to equip high-sampling rate chip due to the cost reduction on ADC and DAC. Considering these 2 aspects, we designed another attack cancellation scheme, which is shown in figure 4.

As is shown in the flowchart, we first divide the signal into high frequency part and low frequency part. In high frequency part, We alternate to use Hilbert transform to extract the attack since it would be a waste of time calculating the peak in a wide band. Then we

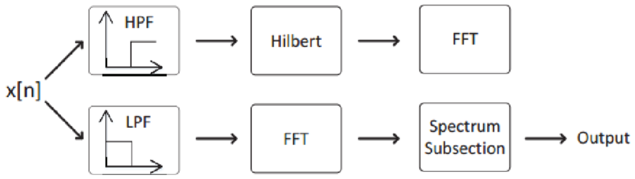


Figure 4: Attack Cancellation Scheme for Devices with High-sampling Chips

will use FT on both the envelope and the low frequency part. Finally, we do subtractions on amplitude spectrum and use IFT to do reconstructions.

4 Evaluation

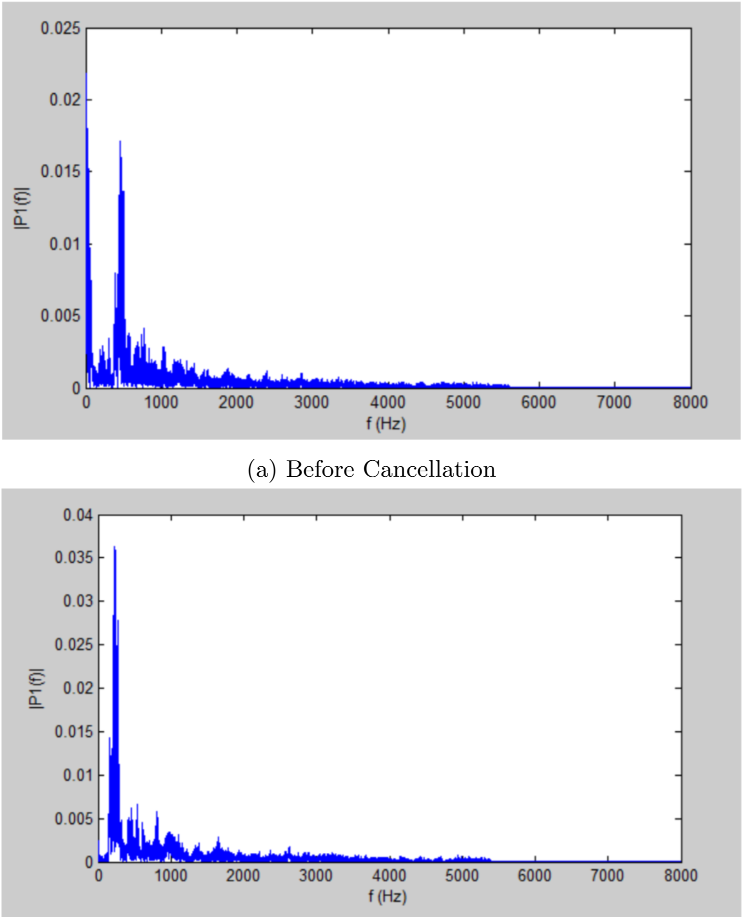
The evaluating works are not finished yet, and the team is currently working on it.

The simulating results of attack cancellation are done on computer, as shown in figure 5. Compare the spectrum diagrams, we can find that the attacking signal is completely canceled.

Currently, we are trying to do simulations on real circuits. One is to generate the chosen single tone groups, i.e. 23kHz, 43kHz...; The other one is to build recording and signal processing circuits for further testing. Complete evaluation will be done in the next few weeks.

5 Conclusion

Our work is based on our junior's current work, to introduce the basic ideas of inaudible attack and the relevant cancellation scheme. Using both mathematical analysis and simulation done on computer, we have confirmed the general attack cancellation scheme with acceptable feasibility, and also brought up the new scheme for novel, high-tech devices. I with my teammate are now working on further real-circuit experiments, and they will be done till the end of the semester to verify our scheme.



(a) Before Cancellation

(b) After Cancellation

Figure 5: Simulation Results of Attack Cancellation

References

[1] N. Roy, H. Hassanieh, and R. Roy Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2017, pp. 2–14.

[2] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 103–117.