BackDoor: Making Microphones Hear Inaudible Sounds

Nirupam Roy, Haitham Hassanieh, Romit Roy Choudhury

University of Illinois at Urbana-Champaign

ABSTRACT

Consider sounds, say at 40kHz, that are completely outside the human's audible range (20kHz), as well as a microphone's recordable range (24kHz). We show that these high frequency sounds can be designed to become recordable by unmodified microphones, while remaining inaudible to humans. The core idea lies in exploiting non-linearities in microphone hardware. Briefly, we design the sound and play it on a speaker such that, after passing through the microphone's non-linear diaphragm and power-amplifier, the signal creates a "shadow" in the audible frequency range. The shadow can be regulated to carry data bits, thereby enabling an acoustic (but inaudible) communication channel to today's microphones. Other applications include jamming spy microphones in the environment, live watermarking of music in a concert, and even acoustic denial-of-service (DoS) attacks. This paper presents *BackDoor*, a system that develops the technical building blocks for harnessing this opportunity. Reported results achieve upwards of 4kbps for proximate data communication, as well as room-level privacy protection against electronic eavesdropping.

1. INTRODUCTION

This paper shows the possibility of creating sounds that humans cannot hear but microphones can record. This is not because the sound is too soft or just at the periphery of human's frequency range. The sounds we create are actually 40kHz and above, completely outside both human's and microphone's range of operation. However, given microphones possess inherent non-linearities in their diaphragms and power amplifiers, it is possible to design sounds that exploit this property. To elaborate, we shape the frequency and phase of sound signals and play them through ultrasound speakers; when these sounds pass through the non-linear amplifier at the receiver, the high frequency sounds are expected to create a low-frequency "shadow". The "shadow" is within the filtering range of the microphone and thereby gets recorded as normal sounds. Figure 1 illustrates the effect. Importantly, the microphone does not require any modification, enabling billions of phones, laptops, and IoT devices to leverage the capability. This paper presents *BackDoor*, a system that develops the technical building blocks for harnessing this opportunity, leading to new applications in security and communications.

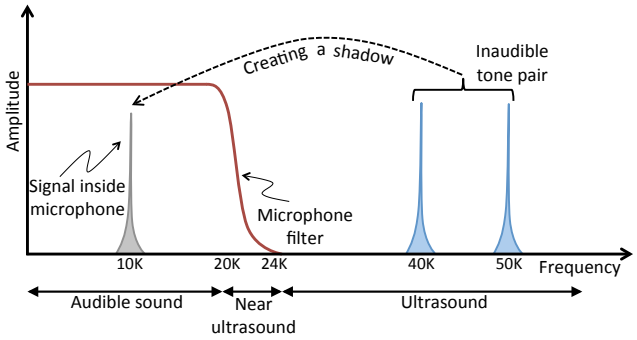

Figure 1: The main idea underlying *BackDoor*.

■ **Security:** Given microphones record these inaudible sounds, it should be possible to silently jam spy microphones from recording. Military and government officials can secure private and confidential meetings from electronic eavesdropping; cinemas and concerts can prevent unauthorized recording of movies and live performances. We also realized the possibility of security threats. Denial-of-service (DoS) attacks on sound devices are typically considered difficult as the jammer can be easily detected. However, *BackDoor* shows that inaudible jammers can disable hearing aids and cellphones without getting detected. For example, during a robbery, the perpetrators can prevent people from making 911 calls by silently jamming all phones' microphones.

■ **Communications:** Ultrasound systems today aim to achieve inaudible data transmissions to the microphone [34]. However, they suffer from limited bandwidth, around 3kHz, since they must remain above human hearing range (20kHz) and below the microphone's cutoff frequency (24kHz). Moreover, FCC imposes strict power restrictions on these bands since they are partly audible to infants and pets [20]. *BackDoor* is free of these limitations. Using an ultrasound-based transmitter, it can utilize the entire microphone spectrum for communication. Thus, IoT devices could find an alternative channel for communication, reducing the growing load on Bluetooth (BLE). Museums and shopping malls could use acoustic beacons to broadcast information about nearby art pieces or products. Various ultrasound ranging schemes, that compute *time of flight* of signals, could benefit from the substantially higher bandwidth in *BackDoor*.

MobiSys '17, June 19–23, 2017, Niagara Falls, NY, USA.

ISBN 978-1-4503-4928-4/17/06…$15.00

DOI: http://dx.doi.org/10.1145/3081333.3081366

This paper focuses on developing the technical primitives that enable these applications. In the simplest case, *BackDoor* plays two tones at say 40kHz and 50kHz. When these tones arrive together at the microphone's power amplifier, they are amplified as expected, but also multiplied due to fundamental non-linearities in the system. Multiplication of frequencies f_1 and f_2 result in frequency components at $(f_1 - f_2)$ and $(f_1 + f_2)$. Given that $(f_1 - f_2)$ is 10kHz in this case, well within the microphone's range, the signal passes unaltered through the low pass filter (LPF). Human ears, on the other hand, do not exhibit such non-linearities and completely filter out the 40kHz and 50kHz sounds.

While the above is a trivial case of sending a tone, *BackDoor* intends to load data on transmitted carrier signals and demodulate the "shadow" after receiving through the microphone. This entails challenges. *First*, The non-linearities we intend to exploit are not unique to the microphone; they are also present in speakers that transmit the sounds. As a result, the speaker also produces a "shadow" within the audible range, making its output audible to humans. We address this by using multiple speakers and isolating the signals in frequency across the speakers. We show, both analytically and empirically, that none of these isolated sounds create a "shadow" as they pass through the speaker's diaphragm and amplifier. However, once these sounds arrive and combine non-linearly inside the microphone, the "shadow" emerges within the audible range.

Second, for communication applications, standard modulation and coding schemes cannot be used directly. Section 4.1 shows how appropriate frequency-modulation, combined with inverse filtering, resonance alignment, and ringing mitigation are needed to boost achievable data rates. *Finally*, for security applications, jamming requires transmitting noisy signals that cover the entire audible frequency range. With audible jammers, this requires speakers to operate at very high volumes. Section 4.2 describes how *BackDoor* is designed to achieve equally effective jamming, but in complete silence. We leverage the *adaptive gain control* (AGC) in microphones, in conjunction with selective frequency distortion, to improve jamming at modest power levels.

The final *BackDoor* prototype is built on customized ultrasound speakers and evaluated for both communication and security applications across different types of mobile devices. Our results reveal the following:

- 100 different sounds played to 7 individuals confirmed that *BackDoor* was completely inaudible.
- *BackDoor* attained data rates of 4 kbps at a distance of 1 meter, and 2 kbps at 1.5 meters – this is 2× higher in throughput and 5× higher in distance than systems that use the near-ultrasound band.
- *BackDoor* is able to jam and prevent the recording of any conversation within a radius of 3.5 meters (and potentially a room-level coverage with higher power [25]). When 2000 English words were played back to 7 humans and a speech recognition software [2], less than 15% of the words were decoded correctly. Audible jammers, aiming at comparable performance, would need to play white noise at a loudness of 97 dBSPL, considered seriously harmful to human ears [19].

In sum, this paper makes the following contributions:

- *Exploits non-linearities in off-the-shelf microphones to enable a "backdoor" from high to low frequencies.* This backdoor permits playback of high frequency sounds that are inaudible to humans and yet recordable through microphones.
- *Builds enabling primitives for applications in acoustic communication and privacy.* The acoustic radio outperforms today's near-ultrasound systems, while jamming raises the bar against eavesdropping.

The subsequent sections expand on these contributions. We begin with an acoustic primer, followed by intuitions, system design, and evaluation.

2. ACOUSTIC SYSTEMS PRIMER

Common Microphone Systems

Any sound recording system requires two main modules – a transducer and an analog-to-digital converter (ADC). The transducer contains a "diaphragm" that vibrates due to sound pressure, producing a proportional change in voltage. The ADC measures this voltage variation (at a fixed sampling frequency) and stores the samples in memory. These samples represent the recorded sound in the digital domain.

A practical microphone needs two more components between the diaphragm and the ADC, namely a *pre-amplifier* and a *low pass filter*. Figure 2 shows the pipeline. The pre-amplifier's task is to amplify the output of the transducer by a gain of around 10× so that the ADC can measure the signal effectively using its predefined quantization levels. Without this amplification, the signal is too weak (around tens of millivolts).

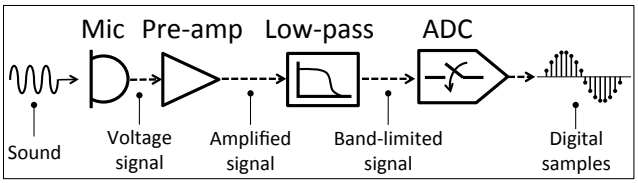

Figure 2: The sound recording signal flow.

As per Nyquist's law, if the ADC's sampling frequency is $f_s Hz$, the sound must be band limited to $\frac{f_s}{2} Hz$ to avoid aliasing and distortions. Since natural sound can spread over a wide band of frequencies, it needs to be low pass filtered (i.e., frequencies greater than $\frac{f_s}{2}$ removed) before the A/D conversion. Since ADCs in today's microphones operate at 48kHz, the low pass filters (LPFs) are designed to cut off signals at 24kHz. Figure 3 shows the effect of the low pass (or anti-aliasing) filter on the recorded sound spectrum.

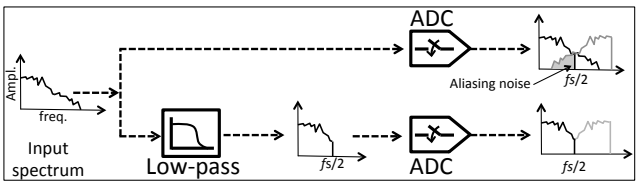

Figure 3: The digital spectrum with and without the (anti-aliasing) low-pass filter.

Sound Playback through Speakers

Sound playback is simply the reverse of recording. Given a digital signal as input, the digital-to-analog converter (DAC) produces the corresponding analog signal and feeds it to the speaker. The speaker's diaphragm oscillates to the applied voltage producing varying sound pressures in the medium, which is then audible to humans.

Linear and Non-linear Behavior

Modules inside a microphone are mostly linear systems, meaning that the output signals are linear combinations of the input. In the case of the pre-amplifier, if the input sound is S, then the output can be represented by

$$S_{out} = A_1 S$$

Here A_1 is a complex gain that can change the phase and/or amplitude of the input frequencies, but does not generate spurious new frequencies. This behavior makes it possible to record an exact (but higher-power) replica of the input sound and playback without distortion.

In practice, however, acoustic amplifiers maintain strong linearity only in the audible frequency range; outside this range, the response exhibits non-linearity. The diaphragm also exhibits similar behavior. Thus, for $f > 25kHz$, the net recorded sound S_{out} may be expressed in terms of the input sound S as follows:

$$S_{out}\Big|_{f>25} = \sum_{i=1}^{\infty} A_i S^i = A_1 S + A_2 S^2 + A_3 S^3 + ...$$

While in theory the non-linear output is an infinite power series, the third and higher order terms are extremely weak and can be ignored. *BackDoor* finds opportunities to exploit the second order term, which can be manipulated by designing the input signal S.

3. CORE INTUITION AND VALIDATION

As mentioned earlier, our core idea is to operate the microphone at high (inaudible) frequencies, thereby invoking the non-linear behavior in the diaphragm and pre-amplifier. This is counter-intuitive because most researchers and engineers strive to avoid non-linearity. In our case, however, we intend to create an inlet into the audible frequency range and non-linearity is essentially the "backdoor". We sketch the basic technique next, followed by some measurements to validate assumptions.

To operate the microphone in its non-linear range, we use an off-the-shelf ultrasound speaker and play a sound S, composed of two inaudible tones $S_1 = 40$ and $S_2 = 50$kHz. Mathematically, $S = Sin(2\pi 40t) + Sin(2\pi 50t)$. After passing through the diaphragm and pre-amplifier of the microphone, the output S_{out} can be modeled as:

$$S_{out} = A_1(S_1 + S_2) + A_2(S_1 + S_2)^2$$

$$= A_1\big\{Sin(\omega_1 t) + Sin(\omega_2 t)\big\} + A_2\big\{Sin^2(\omega_1 t) +$$

$$Sin^2(\omega_2 t) + 2Sin(\omega_1 t)Sin(\omega_2 t)\big\}$$

where $\omega_1 = 2\pi 40$ and $\omega_2 = 2\pi 50$.

Now, the first order terms produce frequencies ω_1 and ω_2, which lie outside the microphone's cutoff. The second order terms, however, is a multiplication of signals, resulting in various frequency components, namely, $2\omega_1$, $2\omega_2$, $(\omega_1 - \omega_2)$, and $(\omega_1 + \omega_2)$. Mathematically,

$$A_2(S_1 + S_2)^2 = 1 - \frac{1}{2}Cos(2\omega_1 t) - \frac{1}{2}Cos(2\omega_2 t) + Cos((\omega_1 - \omega_2)t) - Cos((\omega_1 + \omega_2)t)$$

With the microphone's cut off at 24kHz, all of the above frequencies in S_{out} get filtered out by the LPF, except $Cos((\omega_1 - \omega_2)t)$, which is essentially a 10kHz tone. The ADC is oblivious of how this 10kHz signal was generated and records it like any other sound signal. We call this the "shadow" signal. The net effect is that a completely inaudible frequency has been recorded by unmodified off-the-shelf microphones.

3.1 Measurements and Validation

For the above idea to work with unmodified off-the-shelf microphones, two assumptions need validation. (1) The diaphragm of the microphone should exhibit some sensitivity at the high-end frequencies (> 30kHz). If the diaphragm does not vibrate at such frequencies, there is no opportunity for non-linear mixing of signals. (2) The second order coefficient A_2 needs to be adequately high to achieve a meaningful signal-to-noise ratio (SNR) for the shadow signal, while the third and fourth order coefficients (A_3, A_4) should be negligibly weak. We verify these next.

(1) Sensitivity to High Frequencies: Figure 4 reports the results when a 60kHz sound was played through an ultrasonic speaker and recorded with a programmable microphone circuit. To verify the presence of a response at this high frequency, we "hacked" the circuit using an FPGA kit, and tapped into the signal before it entered the low pass filter (LPF). Figure 4(a) shows the clear detection of the 60kHz tone, confirming that the diaphragm indeed vibrates to ultrasounds. We also measured the channel frequency response at the output of the pre-amplifier (before the LPF): Figure 4(b) illustrates the results. The take away message is that the analog components indeed operate at a much wider bandwidth; it is the digital domain that restricts the operating range.

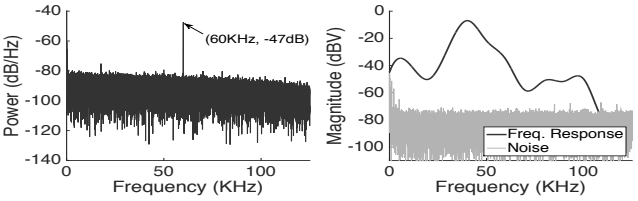

Figure 4: (a) Microphone signals (measured before the LPF) confirm the diaphragm and pre-amplifier's sensitivity to ultrasound frequencies. (b) Full freq. response at the output of the amplifier.

(2) Magnitude of Non-linear Coefficients: Figure 5(a) shows the entire spectrum after the non-linear mixing has occurred, but before the low pass filter (LPF). Except for the shadow at $(\omega_1 - \omega_2)$, we observe that all other frequency spikes are above the LPF's 24kHz cutoff frequency. Similarly, the nonlinear effect on a single frequency – shown in Figure 5(b) – only produces integer multiples of the original frequency, i.e., ω, 2ω, 3ω, and so on. These two types of nonlinear distortions are called *intermodulation* and *harmonic*

distortions, respectively. Importantly, the shadow signal is still conspicuous above the noise floor, while the third order distortion is marginally above noise. This confirms the core opportunity to leverage the shadow.

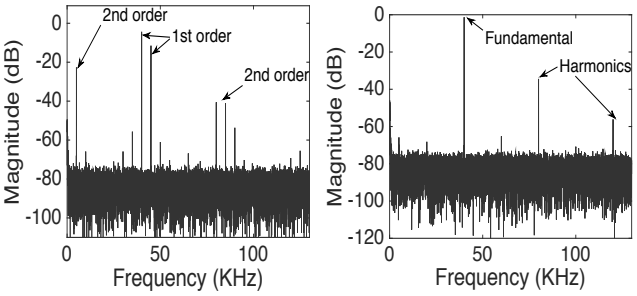

Figure 5: (a)The intermodulation distortion of signal (b) Harmonic distortion.

3.2 Hardware Generalizability

Before concluding this section, we report measurements to confirm that non-linearities are present in different kinds of hardware (not just a specific make or model). To this end, we played high frequency sounds and recorded them across a variety of devices, including smartphones (iPhone 5S, Samsung Galaxy S6), smartwatch (Samsung Gear2), video camera (Canon PowerShot ELPH 300HS), hearing aids (Kirkland Signature 5.0), laptop (MacBook Pro), etc. Figure 6 summarizes the SNR for the shadow signals for each of these devices. The SNR is uniformly conspicuous across all the devices, suggesting potential for widespread applicability.

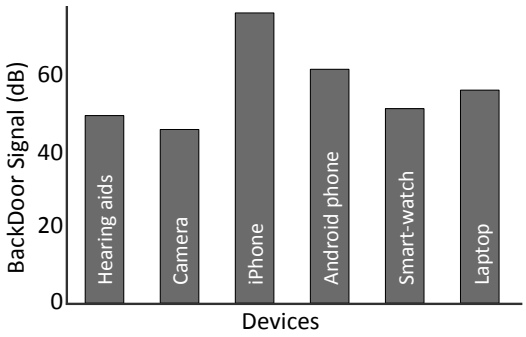

Figure 6: Consistent shadow at 5kHz (in response to 45 and 50kHz ultrasound tones) confirms non-linearity across various microphone platforms.

4. SYSTEM DESIGN

This section details the two technical modules in *BackDoor*: communication and jamming.

4.1 Communication

Thus far, the shadow signal is a trivial tone carrying one-bit of information (presence of absence). While this was useful for explanation, our actual goal is to modulate the high frequency signals at the speaker and demodulate the shadow at the microphone to achieve meaningful data rates. We discuss the challenges and opportunities in developing this communication system.

Failure of Amplitude Modulation (AM)

Our first idea was to modulate a single ultrasound tone, a data carrier, with a message signal $m(t)$. Assuming amplitude modulation [23, 27], this results in $m(t)Sin(\omega_c t)$, where ω_c is a high frequency, ultrasound carrier. Now, if $m(t) = a.Sin(\omega_m t)$, then the speaker should produce this signal:

$$S_{AM} = aSin(\omega_m t)Sin(\omega_c t)$$

Now, when this signal arrives at the microphone and passes through the non-linearities, the squared components of the amplifier's output will be:

$$\begin{aligned} S^2_{out,AM} &= A_2\big\{aSin(\omega_m t).Sin(\omega_c t)\big\}^2 \\ &= -A_2\frac{a^2}{4}\big\{Cos(\omega_c t - \omega_m t) - Cos(\omega_c t + \omega_m t)\big\}^2 \\ &= -A_2\frac{a^2}{4}Cos(2\omega_m t) + (terms\ with\ frequencies \\ &\qquad\qquad above\ \omega_c\ and\ DC) \end{aligned}$$

The result is a signal that contains a $Cos(2\omega_m t)$ component. So long as ω_m, the frequency of the data signal, is less than 10kHz, the corresponding shadow at $2\omega_m = 20$kHz is within the LPF cutoff. Thus, the received sound data can be band pass filtered in software, and the data signal correctly demodulated.

Importantly, the above phenomenon is reminiscent of *coherent demodulation* in conventional radios, where the receiver would have multiplied the modulated signal ($aSin(\omega_m t)Sin(\omega_c t)$) with the frequency and phase-synchronized carrier signal $Sin(\omega_c t)$. The result would be the $m(t)$ signal in baseband, i.e., the carrier frequency ω_c eliminated. Our case is somewhat similar – the carrier also gets eliminated, and the message signal appears at $2\omega_m$ (instead of ω_m). This is hardly a problem since the signal can be extracted via band pass filtering. Thus, the net benefit is that the microphone's non-linearity naturally demodulates the signal and translates to within the LPF cutoff, requiring no changes to the microphone. Put differently, *non-linearity may be a natural form of self-demodulation and frequency translation*, the root of our opportunity.

Unfortunately, the ultrasound transmitter – a speaker with a diaphragm – also exhibits non-linearity. The above property of self-demodulation triggers in the transmitter side as well, resulting in $m(t)$ becoming audible. Figure 7 shows the output of the speaker as visualized by the oscilloscope; a distinct audible component appears due to amplitude modulation. In fact, any modulation that generates *waveforms with non-constant envelopes* [45] is likely to suffer this problem. This is unacceptable and brings forth the first design question: *how to cope with transmitter-side non-linearity?*

Bypassing Transmitter Non-linearity

The design goal at this point is to modulate the carrier signal with data without affecting the envelope of the transmitted signal. This raises the possibility of *angle modulation* (i.e, modulating the phase or frequency but leaving amplitude untouched). However, we recognized that phase modulation (PM) is also unsuitable in this application because of unpredictable noise from phone movements. In particular, the smaller wavelength of ultrasonic signals are easily affected by phase noise and involves complicated receiver-side schemes during demodulation. Therefore, we choose the other alternative of angle modulation: *frequency modulation* (FM). Of

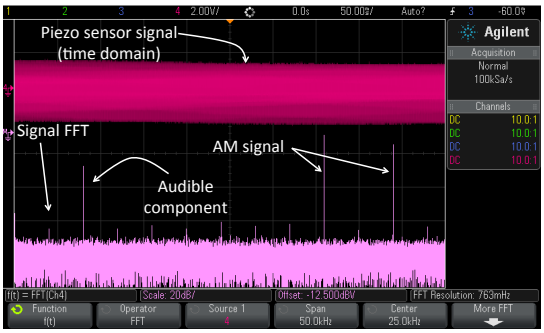

Figure 7: The AM signal produces an audible frequency due to self-demodulation, shown in this oscilloscope screenshot.

course, FM modulation is not without tradeoffs; we discuss them and address the design questions step by step.

FM: No Frequency Translation

FM modulated signals, unlike AM, do not get naturally demodulated or frequency-translated when pass through non-linear transmitter. Assuming $Cos(\omega_m t)$ as the message signal, we have the input to the speaker as:

$$S_{fm} = Sin(\omega_c t + \beta Sin(\omega_m t))$$

Note that the phase of the FM carrier signal should be the integral of the message signal, hence it is $Sin(\omega_m t)$. Now when S_{fm} gets squared due to non-linearity, the result is of the form $(1+Cos(2\omega_c t+otherTerms))$ i.e., a DC component and another component at $2\omega_c$. Hence, along with the original ω_c frequency the transmitter output contains frequency at $2\omega_c$, both above the audible cut-off. Thus nothing gets recorded by the microphone. The advantage, however, is that the output of the speaker is no longer audible. Moreover, as typically the speaker has a low response at high frequencies near $2\omega_c$, the output signal is dominated by the data signal at ω_c as in original S_{fm}.

Second Carrier for Frequency Translation

To get the message signal recorded, we need to frequency-shift the signal at ω_c to the microphone's audible range, without affecting the signal transmitted from the speaker. To achieve this, *BackDoor* introduces a second ultra-sound signal transmitted from a second speaker collocated with the first speaker. Let us assume this second signal is called the *secondary carrier*, ω_s. Since ω_s does not mix with ω_c at the transmitter, the signal that arrives at the microphone diaphragm is simply of the form:

$$S_{fm}^{Rx} = \Big(A_1 Sin(\omega_c t + \beta Sin\omega_m t) + A_1 Sin(\omega_s t)\Big)$$

Note that the first term from the FM modulated ω_c signal, and the second term from the ω_s secondary carrier. Now, upon arriving on the receiver, the microphone's non-linearity essentially squares this whole signal as $(S_{fm}^{Rx})^2$. Expanding this mathematically results in a set of frequencies centered at $(\omega_c - \omega_s)$, and the others at $(\omega_c + \omega_s)$, $2\omega_c$, and $2\omega_s$. If we design ω_c and ω_s to have a difference less than the LPF cutoff, the microphone can record the signal.

Choosing ω_c and ω_s :
As we considered the requirements of the system, the choice of ω_c and ω_s became clear. First, note that the FM-modulated signal has a bandwidth of, say $2W$, ranging from $(\omega_c - W)$ to $(\omega_c + W)$. Thus, assuming that the microphone's LPF cutoff is 20kHz, we should translate the center frequency to 10kHz; this maximizes W that can be recorded by the microphone. Immediately, we know that $(\omega_c - \omega_s) = 10$kHz.

Second, the microphone's diaphragm exhibits resonance at certain frequencies; ω_c and ω_s should leverage this to improve the strength of the recorded signal. Figure 8 plots the normalized power of the *translated signal* for different values of ω_c and ω_s. Given $(\omega_c - \omega_s) = 10$kHz, the resonance effects demonstrate the maximum response when ω_c is 40kHz, and ω_s is 50kHz.

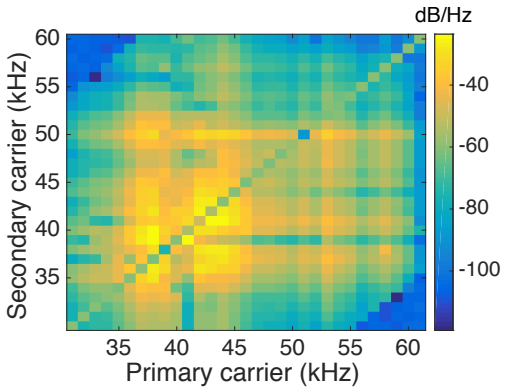

Figure 8: Resonance for various $\omega_c - \omega_s$ values.

Coping with the "Ringing" Effect

The piezo-electric material in the speaker, that actually vibrates to create the sound, behaves as an oscillatory inductive-capacitive circuit. This loosely means that the actual vibration is a weighted sum of input sound samples (from the recent past), and hence, the piezo-electric material has a heavy-tailed impulse response (shown in Figure 9). Mathematically, the output of the speaker can be computed as a convolution between this impulse response and the input signal. Unfortunately, the non-linearity of the speaker impacts this convolution process as well, and generates low frequency components similar to the natural demodulation effect discussed earlier. The result is a "ringing effect", i.e., *the transmitted sound becomes slightly audible even with FM modulation.*

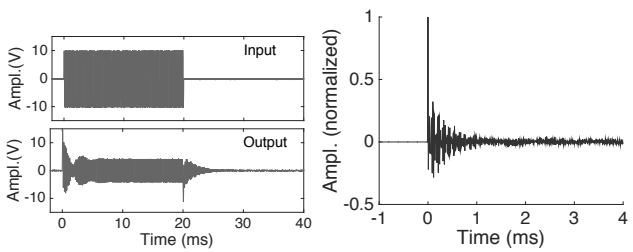

Figure 9: (a) The prolonged oscillation in an ultrasonic transmitter following a 40kHz sine burst input. (b) The impulse response of the ultrasonic transmitter.

To explain the self-demodulation effect, we assume a simplified impulse response 'h':

$$h = \sum_{i=0}^{\infty} k_i \delta(t-i) \approx k_0\delta(t) + k_1\delta(t-1)$$

When an angle modulated (FM/PM) signal 'S' is convolved with 'h', the output 'S_{out}' is:

$$\begin{aligned} S_{out} &= S * h \\ &= sin(\omega_c t + \beta sin(\omega_m t)) * (k_0\delta(t) + k_1\delta(t-1)) \\ &= \quad k_0 sin(\omega_c t + \beta sin(\omega_m t)) \\ &\quad + \quad k_1 sin(\omega_c (t-1) + \beta sin(\omega_m (t-1))) \end{aligned}$$

While S_{out} contains only high frequency components (since convolution is linear), the non-linear counterpart S_{out}^2 mixes the frequencies in a way that has lower frequency components (or shadows):

$$\begin{aligned} S_{out}^2 &= k_0 k_1 cos(\omega_c + 2\beta sin(\frac{\omega_m}{2}) sin(\omega_m t - \frac{\omega_m}{2})) \\ &\quad + (terms\ with\ frequencies\ over\ 2\omega_c\ and\ DC) \end{aligned}$$

Figure 10 shows the spectrum of S_{out} and S_{out}^2, with and without the convolution. Observe the low frequency "shadow" that appear due to the second order term for the convolved signal – this shadow causes the ringing and is noticeable to humans.

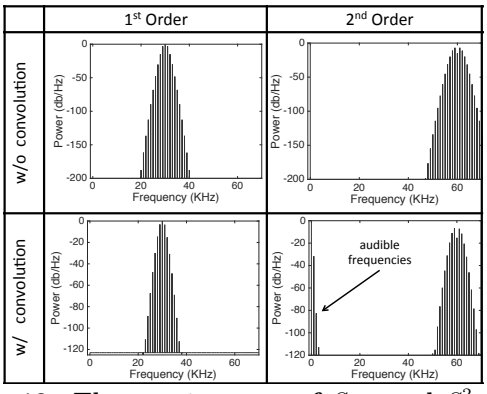

Figure 10: The spectrogram of S_{out} and S_{out}^2, with and without the convolution. The shadow signal appears due to second-order non-linear effects on the convolved signal.

In most speakers, this "shadow" signal is weak; some expensive speakers even design their piezo-electric materials to be linear in a wider operating region precluding this possibility. However, we intend to be functional across all speaker platforms (even the cheapest ones) and aim to be completely free of any ringing whatsoever. Hence, we adopt an inverse filtering approach to remove ringing.

Inverse Filtering to Eliminate Ringing

Our core idea draws inspiration from *pre-coding* in wireless communication, i.e., we modify the input signal S_{fm} so that it remains the same after convolution. In other words, if the modified signal $S_{mod} = h^{-1} * S_{fm}$, then the impact of convolution on S_{mod} results in $h * h^{-1} * S_{fm}$, which is S_{fm} itself. With S_{fm} as the output of the speaker, we do not experience ringing. Of course, we need to compute h^{-1}, i.e., learn the coefficients of the impulse response. For this, we monitor the current passing through the ultrasonic transmitter at different frequencies and calculate the $(k_0, k_1, k_2, ...)$. Fortunately, unlike wireless channels, the response of the transmitter does not vary over time and hence the coefficients of the inverse filter can be pre-calculated. Figure 11(a) shows the frequency response of one of our ultrasound speakers, while Figure 11(b) shows how our inverse filtering scheme curbs the ringing effect.

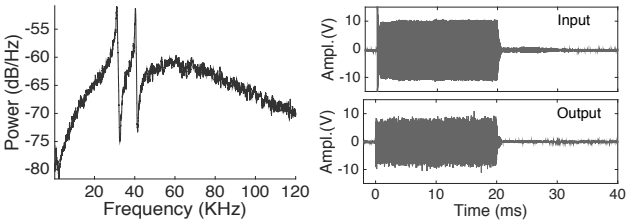

Figure 11: (a) Freq. response of the ultrasonic speaker. (b) Inverse filtering method almost eliminates ringing effect compared to Figure 9

.

Receiver Design

This completes the transmitter design and the receiver is now an unmodified microphone (from off-the-shelf phones, cameras, laptops, etc.). Of course, to extract the data bits, we need to receive the output signal from the microphone and decode them in software. For example, in smartphones, we have used the native recording app, and operated on the stored signal output. The decoding steps are as follows.

We begin by band pass filtering the signal as per the modulating bandwidth. Then, we need to convert this signal to its baseband version and calculate the instantaneous frequency to recover the modulating signal $m(t)$. This signal contains the negative-side frequencies that overlap with the spectrum-of-interest during the baseband conversion. To remove the negative frequencies, we Hilbert Transform the signal, producing a complex signal [29]. Now, for baseband conversion, we multiply this complex signal with another complex signal $e^{-j2\pi(\omega_s - \omega_c)t}$. Here $(\omega_s - \omega_c)$ is 10kHz, i.e., the shifted carrier frequency. This operation brings the modulated spectrum to baseband, centered around DC. The differentiation of its phase gives the instantaneous frequency [40], which is then simply mapped to data bits. Section 5 will present performance evaluation, but before that, we present the techniques for inaudible voice jamming.

4.2 Jamming

Imagine military applications in which a private conversation needs to be held in an untrusted environment, potentially bugged with spy microphones. We envision turning on one/few *BackDoor* devices in that room. The device will broadcast appropriately designed ultrasound signals that will not interfere with human conversation, but will jam microphones in the vicinity. This section targets 2 jamming techniques towards this goal: (1) passive gain suppression, and (2) active frequency distortion. Together, the techniques mitigate electronic eavesdropping.

(1) Passive Gain Suppression

Our core idea is to leverage the *automatic gain control* (AGC) circuit [31, 38, 44] in the microphone to suppress voice conversations. By transmitting a narrowband ultrasound frequency at high amplitude, we expect to force the

microphone to alter its dynamic range, thereby weakening the SNR of the voice signal. We elaborate next, beginning with a brief primer on AGC.

■ **AGC Primer:**
Our acoustic environment has large variations in volume levels ranging from soft whispers to loud bangs. While human ears seamlessly handle this dynamic range, it poses one of the major difficulties in microphones. Specifically, when a microphone is configured at a fixed gain level, it fails to record a soft signal below the minimum quantization limit, while a loud sound above the upper range is clipped, causing severe distortions. To cope, microphones use an Automatic Gain Control (AGC) (as a part of its amplifier circuit) that adjusts the signal amplitude to fit well within the ADC's lower and upper bounds. As a result, the signal covers the entire range of the ADC, offering the best possible signal resolution.

Figure 12 demonstrates the AGC operation in a common MEMS microphone (ADMP401) connected to the line-in port of a Linux laptop running the ALSA sound driver. We simultaneously play 5kHz and 10kHz tones through two different (but collocated) speakers and display the power spectrum of the received sound. Figure 12(a) reports both the signals at around −20dB. However, when we increase the power of the 10kHz signal to reach its AGC threshold (while keeping the 5kHz signal unaltered), Figure 12(b) shows how the microphone reduces the overall gain to accommodate the loud 10kHz signal. This results in a 25dB reduction of the unaltered 5kHz signal.

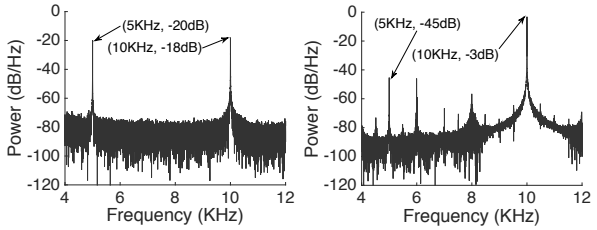

Figure 12: Automatic Gain Control: (a) The 5kHz tone is at −20dB when the amplitude of the 10kHz frequency is at comparable power. (b) The 5kHz tone reduces to −45dB when the amplitude of the 10kHz tone is made to exceed the AGC threshold. Some spurious frequencies also appear due to non-linearities.

■ **Voice Suppression via AGC:**
In line with the above idea, when our ultrasound signal at ω_c passes through the AGC (i.e., before this frequency is removed by the low pass filter), it alters the AGC gain configuration and significantly suppresses the voice signals in the audible frequency. Figure 13 shows the reduction in the received sound power in a *Samsung Galaxy S-6* smartphone when ultrasound tones are played at different frequencies from a piezo-electric speaker. Evident from the plot, the maximum reduction is due to the signal at 40kHz – this is because 40kHz is the resonance frequency of the piezo-electric transducer, and thereby delivers the highest power. In that sense, using the resonance frequency offers double gains, one towards increasing the SNR of our communication signal, and the other for jamming.

This reduction in signal amplitude results in low resolution when sampled with discrete quantization levels at the ADC.

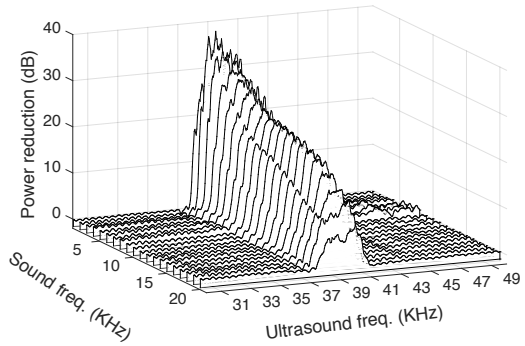

Figure 13: The reduction in sound power due to the AGC: reduction maximum for the 40kHz tone due to the speaker's resonance at this frequency.

In fact, an adequately loud ultrasonic tone can completely prevent the microphone from recording any meaningful voice signal by reducing its amplitude below the minimum quantization level. However, as the electrical noise level is usually higher than the minimum quantization level of the ADC, it is sufficient to reduce the signal power below that noise floor.

Figure 14 shows the reduction in the signal power of a recorded voice segment for 3 different power levels of the 40kHz tone. In practice, an absolute amplitude reduction is difficult unless the speaker uses high power. Importantly, high power speakers are possible with *BackDoor* since the jamming signal is inaudible. *On the other hand, regular white noise audio jammers must operate below strict power levels to not interfere with human conversation/tolerance. This is a key advantage of jamming with BackDoor.* Nonetheless, we still attempt to lower the power requirement by injecting additional frequency distortions at the eavesdropper's microphone.

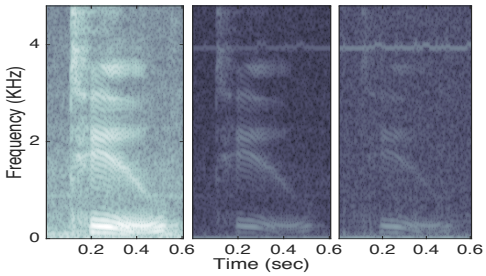

Figure 14: The reduction in signal power of recorded voice segment for 3 power levels (darker is lower power).

(2) Injecting Frequency Distortion

A traditional jamming technique is to add strong white noise to reduce the SNR of the target signal. We first implement a similar technique, but with inaudible band-limited Gaussian noise. Specifically, we modulate the ω_c carrier with white noise, bandpass filtered to allow frequencies between 40kHz to 52kHz only. The 52kHz ω_s carrier shifts this noise to $[0, 12]$kHz, which is sufficient to affect the voice signal.

To improve, we then shape the white noise signal to boost power in frequencies that are known to be important for voice. Note that these distortions are designed in the ultrasound bands (to maintain inaudibility), and hence they are

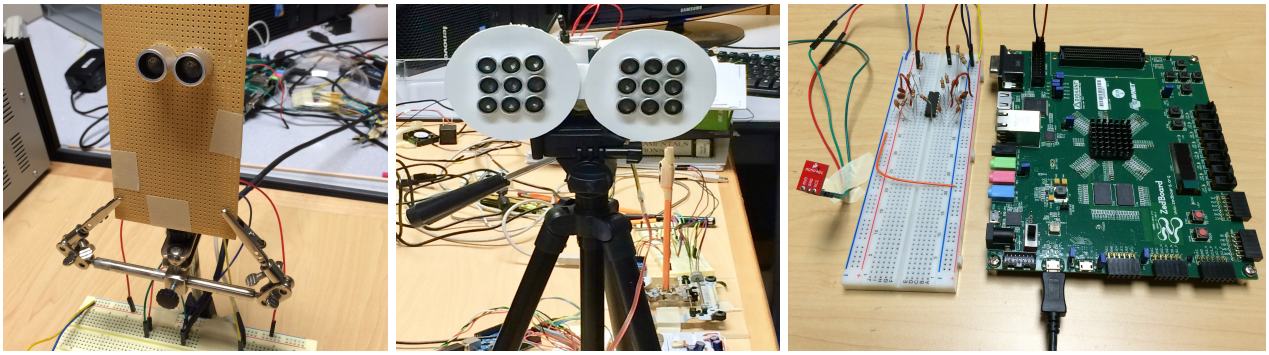

Figure 15: *BackDoor* experimental setup: (a) Two ultrasonic speakers mounted on a circuit board for data communication. (b) A $2Watt$ speaker array system for jamming applications. (c) The FPGA based set up for probing into individual components of the microphone.

played through the ultrasound speakers. Section 5 will report results on word legibility, as a function of the separation between the jammer and the spy microphone.

5. EVALUATION

BackDoor was evaluated on 3 main metrics: (1) human audibility, (2) throughput, packet error rates(PER) and bit error rates (BER) for data communication, and (3) the efficacy of jamming. We summarize the key results here, followed by details.

• Table 1 reports human perception of audibility for *BackDoor* for various frequencies, modulations, and SNR levels. Except for amplitude modulation (AM), all the human volunteers reported complete silence.

• Figure 17 and 18 report the variation of throughput against increasing distance, different phone orientations, and impact of acoustic interference. The results show throughput of 4 kbps at 1 meter away which is 2× to 4× higher than today's mobile ultrasound communication systems.

• Figure 19 compares the jamming radius for *BackDoor* and audible white noise-based jammers. To achieve the same jamming effect (say, $< 15\%$ words legible by humans), we find that the audible jammer requires a loudness of 97 dB-SPL which is similar to a jackhammer and can cause severe damage to humans [19]. *BackDoor*, on the other hand, remains completely silent. Conversely, when the white noise sound level is made tolerable, the legibility of the words was 76%.

We elaborate on these results below, starting with details on our implementation platform.

5.1 Implementation

(1) Transmitter Speakers: Figure 15(a) and (b) show two different transmitter prototypes we have developed, the first one for communication and the other for jamming. The communication transmitter consists of two ultrasonic piezoelectric speakers [33]; each transmits a separate frequency as described in Section 4. A programmable waveform generator (*Keysight 33500b series*) drives the speakers with frequency modulated signals. The signals are amplified using an NE5535AP op-amp based non-inverting amplifier, permitting signals up to 150kHz. The jamming transmitter in Figure 15(b) is composed of two speaker arrays, each array with 9 piezoelectric speakers connected in parallel to generate a 2Watt jamming signal. The signals driving these arrays are first amplified using an *LM380* op-amp based power amplifier separately powered from a constant DC-voltage source. Figure 16 shows the circuit diagram of the speaker array.

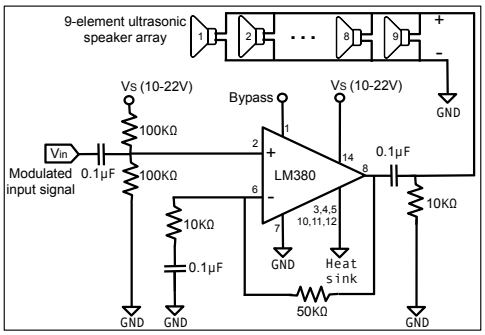

Figure 16: The circuit diagram of the jamming transmitter.

(2) Receiver Microphones: We experiment with two types of receivers. The first is an off-the-shelf *Samsung Galaxy S6* smartphone (released in Aug, 2015) running Android OS 5.1.1. Signals are recorded through a custom Android app using the standard APIs. The second receiver is shown in Figure 15(c) – a more involved setup that was mainly used for micro-benchmarks reported earlier in Sections 3 and 4. This allowed us to tap into different components of the microphone pipeline, and analyze signals in isolation. The system runs on a high bandwidth data acquisition *ZedBoard*, a Xilinx Zynq-7000 SoC based FPGA platform [12], that offers a high-rate internal ADC (up to 1 Msample/sec). A MEMS microphone (*ADMP 401*) is externally connected to this ADC, offering undistorted insights into higher frequency bands of the spectrum.

5.2 Human Audibility Results

We played *BackDoor* signals to a group of 7 users (ages between 27 and 38) seated around a table 1 to 3 meters away from the speakers. Each user reported the perceived loudness of the sound on a scale of 0-10, with 0 being perceived silence. As a baseline, we also played audible sounds and

Reference Mic.	2kHz Tone		5kHz Tone		FM		AM		White Noise	
SNR (dB)	*BackDoor*	Audible	*BackDoor*	Audible	*BackDoor*	Audible	*BackDoor*	Audible	*BackDoor*	Audible
25	0	0.75	0	3.33	0	1.2	0	0.46	0	0.1
30	0	1.5	0	4.08	0	2.3	0.1	1.36	0	0.26
35	0	2	0	4.91	0	3.5	0.1	1.85	0	0.5
40	0	2.67	0	5.42	0	4.2	0.16	2.4	0	0.8
45	0	3.17	0	6.17	0	4.8	0.68	3.06	0	1.24

Table 1: Perceived loudness of *BackDoor* in comparison to audible sounds.

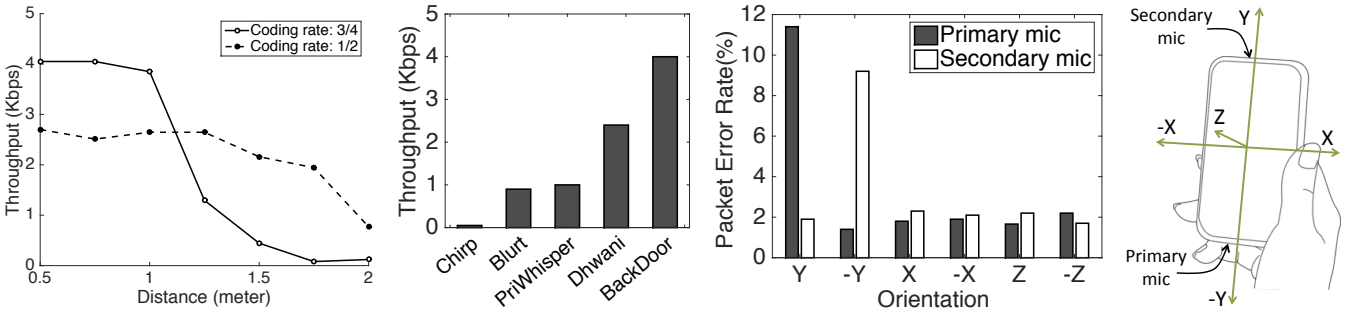

Figure 17: *BackDoor* Communication Results: (a) Throughput vs. Distance, (b) Throughput comparison against related P2P communication schemes. (c) Packet error rate vs. Orientation. (d) Phone orientations.

asked the users to report the loudness levels. A reference microphone is placed at 1m from the speaker to record and compute the SNR (Signal to Noise Ratio) of all the tested sounds. We varied the SNR and equalized them at the microphone for fair comparison between audible and inaudible (*BackDoor*) sounds.

Four types of signals were played: **(1) Single Tone Unmodulated Signals:** In the simplest form, we transmitted multiple pairs of ultrasonic tones ($<40, 42>$ and $<40, 45>$) that generate a single audible frequency tone in the microphone. As baseline, we separately played a 2kHz and 5kHz audible tone. **(2) Frequency Modulated Signals:** We modulated the frequency of a 40kHz *primary carrier* with a 3kHz signal. We also transmitted a 45kHz *secondary carrier* on the second speaker, producing 3kHz FM signal centered at 5kHz in the microphone. As baseline, we played the equivalent audible FM signal on the same speakers. **(3) Amplitude Modulated Signals:** Similar to FM signals, we created these AM signals by modulating the amplitude of 40kHz signal with a 3kHz tone. **(4) White Noise Signals:** Finally, we generated white Gaussian noise with zero mean and variance proportional to the transmitted power, at a bandwidth of 8kHz, band-limited to $[40, 48]$kHz. We also transmit a 40kHz tone on the second speaker to frequency shift the white noise to the audible range of the speaker. As baseline, we create audible white noise with the same properties band-limited to $[0, 8]$kHz and played it on the speakers.

Audibility Vs. SNR

Table 1 summarizes the average of perceived loudness that users reported for both *BackDoor* and audible signals as a function of the SNR measured at the reference microphone. For all types of signals except amplitude modulation (AM), *BackDoor* is completely inaudible to all the users. AM signals are audible due to speaker non-linearity, as described earlier. However, the perceived loudness of *BackDoor* is significantly lower than that of audible signals. Thus, so long we avoid AM, *BackDoor* signals remain inaudible to humans but produce audible signals inside microphones with the same SNR as loud audible signals.

5.3 Communication Results

The *BackDoor* transmitter is the 2-speaker system while the receiver is the Samsung smartphone. The recorded acoustic signal is extracted and processed in MATLAB; we compute bit error rate (BER), packet error rate (PER) and throughput under varying parameters. Overall, 40 hours of acoustic transmission was performed to generate the results.

Throughput

Figure 17(a) reports *BackDoor*'s net end-to-end throughput for increasing separation between the transmitter and the receiver. *BackDoor* can achieve a throughput of 4kbps at 1m, 2kbps at 1.5m and 1kbps at 2m. Figure 17(b) compares *BackDoor*'s performance in terms of throughput and range with state-of-the-art mobile acoustic communication systems (in both commercial products [1, 13] and research [34, 22]). The figure shows that *BackDoor* achieve $2\times$ to $80\times$ higher throughput. This because these systems are constrained to a very narrow communication band whereas *BackDoor* is able to utilize the entire audible bandwidth.

Impact of Phone Orientation

Figure 17(c) shows the *packet error rate* (PER) when data is decoded by the primary and secondary microphones in the phone, placed in 6 different orientations (shown in Figure 17(d)). The aim here is to understand how real-world use of the phone impacts data delivery. To this end, the phone was held at a distance of 1m away from the transmitter, and the orientation changed after each transmission session. The plot shows that except Y and $-Y$, the other orientations are comparable. This is because the $Y/-Y$ orientation align the two receivers and transmitters in almost a straight line, resulting in maximal SNR difference. Hand blockage of the further-away microphone makes the

SNR gap pronounced. It should be possible to compare the SNR at the microphones and select the better microphone for minimized PER (regardless of the orientation).

Impact of Interference

Figure 18(a) reports the bit error rate (BER) variation against 3 different audible interference sources. To elaborate, we played audible interference signals – a presidential speech, an orchestral music, and white noise – from a nearby speaker, while the data transmission was in progress. The intensity of the interference at the microphone was at 70 dBSPL, equaling the level of volume one hears on average in face-to-face conversations. This is certainly much louder than average ambient noise, and hence, this serves as a strict test for *BackDoor*'s resilience to interference. Also, the smartphone receiver was placed 1m away from the speaker, and transmissions were at 2kbps and 4kbps.

Evident from the graph, voice and music has minimal impact on the communication error. On the other hand, white noise can severely degrade performance. Figure 18(b) plots the power spectral density of each interference – the decay beyond 4kHz for voice and music explains the performance plots. Put differently, since *BackDoor* operates around 10kHz frequency, voice and music signals do not affect the band as much as white noise, that remains flat over the entire spectrum.

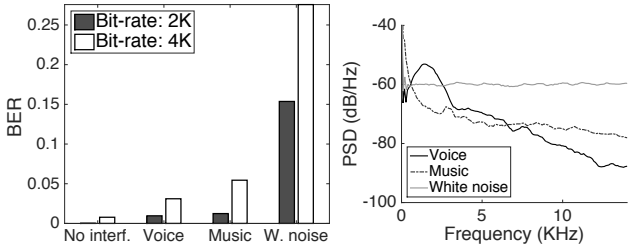

Figure 18: (a) BER vs. Interference. (b) Spectral density of interfering signals.

5.4 Jamming Results

■ **Setup:** Consider the case where Bob is saying a secret to Alice and Eve has planted a microphone in the vicinity, attempting to record Bob's voice. In suspicion, Bob places a *BackDoor* jammer in front of him on the table. We intend to report the efficacy of jamming in such a situation. Specifically, we extract the jammed signal from Eve's microphone and play it to an automatic speech recognizer (ASR), as well as to a group of 7 human users. We define *Legibility* as the percentage of words correctly recognized by each. We plot L_{asr} and L_{human} for increasing jamming radius, i.e., for increasing distance between Alice and Eve's microphone.

We still need to specify another parameter for this experiment – the loudness with which Bob is speaking. Acoustic literature suggests that at social conversations, say between two people standing at arm's length at a corridor, the average loudness is 65 dBSPL (dB of sound pressure level). We design our situation accordingly, i.e., when Bob speaks, his voice at Alice's location 1m away is made to be 70 dBSPL (i.e., Bob is actually speaking louder than general social conversations).

In the actual experiment, we pretend that a smartphone is a spy microphone. Another smartphone's speaker is a proxy for Bob, and the words played are derived from Google's Trillion Word Corpus [10]; we pick the 2000 most frequent words, prescribed as a good benchmark [35]. As mentioned earlier, the volume of this playback is set to 70 dBSPL at 1m away. Now, the *BackDoor* prototype plays an inaudible jamming signal through its ultrasonic speakers to jam these speech signals.

■ **Baseline:** Our baseline comparison is essentially against *audible* white noise-based jammers in today's markets. Assuming *BackDoor* jams up to a radius of R, we compute the loudness needed by white noise to jam the same radius. All in all, 14 hours of sound was recorded and a total of $25,000$ words were tested. The ASR software is the open-source *Sphinx4* library (pre-alpha version) published by CMU [2, 21]. We present the results next.

Audible and Inaudible Jamming Radius

Figure 19(a) plots L_{asr} and L_{human} for increasing jamming radius. Even with a 1W power, a radius of 3.5m (around 11 feet) can be jammed around Bob. We compare against audible noise jammers presented in Figure 19(b). For jamming at the same radius of 3.5m, the loudness necessary for the audible white noise is 97 dBSPL which is the same as a jackhammer and can cause damage to the human ear [19]. Conversely, we find that when the audible white noise is made tolerable (comparable to a white noise smartphone app playing at full volume), the legibility becomes 76%. Thus, *BackDoor* is a clear improvement over audible jammers. More importantly, increasing the power of *BackDoor* jammers can increase the radius proportionally. This can be easily achieved. In fact, current portable Bluetooth speakers already transmit 10× to 20× higher power than *BackDoor* [4, 3]. Audible jammers cannot increase their power to boost the range since they are already intolerable to humans.

Impact of Selective Frequency Distortion

Figure 19(c) shows results when the jamming signal is simply a white noise, without the deliberate distortions of voice-centric frequencies (fricatives, phonemes, and harmonics). Evidently, the performance is substantially weaker, indicating the importance of signal shaping and jamming. Finally, Figure 19(d) shows the confidence scores from ASR for all correctly recognized words. Results show quite low confidence on a large fraction of words, implying that voice fingerprinting and other voice-controlled systems would be easy to DoS-attack with a *BackDoor*-like system.

6. POINTS OF DISCUSSION

Needless to say, there is much room for further work and improvement. We discuss a few points here.

• **Jamming Range:** *BackDoor*'s restriction in the jamming range stems from the attenuation of ultrasound in air and the limited amplitude at which the ultrasound speakers can vibrate, producing low power signals. We have demonstrated a proof-of-concept with 9 speakers that boosts the jamming power level – direct materials cost is around $4. It should be certainly possible to develop a bigger speaker array to significantly increase the power [25]. In some cases (e.g. movie theater) multiple short-range jammers can be used to sufficiently cover the space. The jammers could be wall powered where necessary, and yet, will remain inaudible.

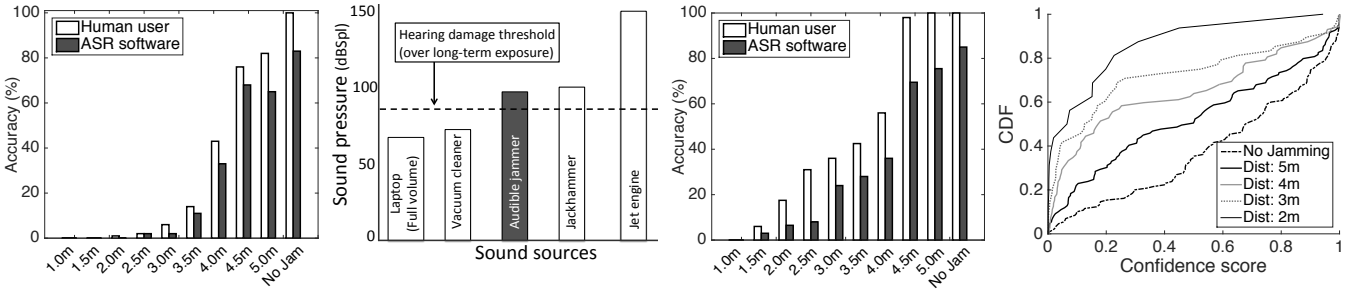

Figure 19: Jamming results: (a) *BackDoor* jams a radius of 3.5m at 2W power. (b) White noise power needed to match *BackDoor* is intolerable. (c) Jamming radius when *BackDoor* uses inaudible white noise, showing importance of selectively jamming voice-centric harmonics. (d) Confidence of speech recognizer.

• **Smarter Spy:** We have assumed a fairly simple attacker planting a single microphone in the vicinity. Multiple microphones, perhaps even with various beamforming capabilities, may be able to extract out the voice from the jamming signal. However, greater sophistication in jamming should be feasible too, such as variation in the jamming signal to prevent channel estimation; even some movements of the speakers. We leave this to future work.

• **Interference with Phone Calls:** Data communication with *BackDoor* can interfere with people talking on the phone nearby. To this end, data communication applications will inherently need to be proximate and at low power. One possibility is an acoustic NFC, but at greater ranges of 1 or 2 feet. Alternatively, the communication could be made spread spectrum so that the interference remains below the noise floor. Our ongoing work is investigating these unresolved issues.

7. RELATED WORK

■ **Literature in Acoustic Non-linearity:** The literature in acoustic signal processing and communication is extremely rich. The notion of exploiting non-linearity was originally studied in the 1957 by Westervelt's seminal theory [43, 42], which later triggered a series of research. The core vision was that non-linearities of the air can naturally self-demodulate signals; when combined with directional propagation of ultrasound signals, it may be possible to deliver audible information over large distances using relatively low power [17, 14, 46]. Recently, there has been a revival of these efforts with *AudioSpotlight* [5], *SoundLazer* [9, 8], and other projects [47, 11, 36]. Our work, however, is opposite of these efforts – we are attempting to retain the inaudible nature of ultrasound while making it recordable inside electronic circuits.

■ **Medical Devices:** Human bones have also been shown to exhibit non-linearities that self-modulate signals, resulting in applications in bone conduction ultrasound hearing aids for severely deaf individuals [28, 15, 16, 37, 32]. Even bone conduction headphones are being considered that exploit similar non-linearities [24].

■ **Assorted Topics Related to *BackDoor*:** A set of recent works bear some degree of relevance to *BackDoor*. *Dhwani* [34] explores in-air sound signals as a short range, ad-hoc data transfer modality. *Chirp* [1] and *Zoosh* [39, 13] have rolled out commercial products using sound for a secure data exchange medium. *GhostTalk* [26] explores various attack scenarios on the consumer electronics using high power electromagnetic interference. Another thread of recent work has looked into watermarking audio-visual media. *Dolphin* [41] enables speaker-microphone communication by embedding data bits on the sound. It adapts the signal parameters in real-time to keep the embedded signal imperceptible to human ears while achieving 500 bps data rate. *Kaleido*[48] proposes a video precoding based solution to prevent videotaping an on-screen show in a theater or on website. It precodes distortions in the video such that it is invisible to humans but severely distorts videotaping (due to specific limitations of the camera). Finally, sound maskers have also been used for protecting private conversation, however, these techniques have been limited to audible frequencies [18, 30, 6, 7]. *BackDoor* differs from the above in the sense that it exploits discrepancies between humans and electronics, ultimately enabling a new capability to the best of our knowledge.

8. CONCLUSION

Device non-linearity has been conventionally viewed as a peril. This paper breaks away from this point of view and discovers various opportunities to harness non-linearity. By carefully designing ultrasound signals, we demonstrate that such signals remain inaudible to humans but are record-able by unmodified off-the-shelf microphones. This translates to new applications including inaudible data communication, privacy, and acoustic watermarking. While our ongoing work is focused on deeper understanding of these capabilities and applications, our longer term goal is focused on generalization to other platforms, such as wireless radios and inertial sensors.

Acknowledgement

We sincerely thank the anonymous reviewers for their valuable feedback. We are grateful to the Joan and Lalit Bahl Fellowship, Qualcomm, IBM, and NSF (award number: 1619313) for partially funding this research.

9. REFERENCES

[1] Chirp technology. http://www.chirp.io. Last accessed 28 November 2016.

[2] Cmu sphinx. http://cmusphinx.sourceforge.net. Last accessed 6 December 2015.

[3] Hight power bluetooth speaker: 12watt. https://www.cnet.com/products/jbl-pulse/specs/. Last accessed 28 November 2016.

[4] Hight power bluetooth speaker: 38watt. http://www.fugoo.com/fugoo-tough-xl/. Last accessed 28 November 2016.

[5] Holosonics webpage. https://holosonics.com. Last accessed 28 November 2016.

[6] Sound masking device. http://www.oeler.com/sound-masking-systems/. Last accessed 28 November 2016.

[7] Sound masking solutions. https://www.speechprivacysystems.com. Last accessed 28 November 2016.

[8] Soundlazer kickstarter. https://www.kickstarter.com/projects/richardhaberkern/soundlazer. Last accessed 28 November 2016.

[9] Soundlazer webpage. http://www.soundlazer.com. Last accessed 28 November 2016.

[10] Top 10000 words from google's trillion word corpus. https://github.com/first20hours/google-10000-english. Last accessed 6 December 2015.

[11] Woody norris ted talk. https://www.ted.com/speakers/woody_norris. Last accessed 28 November 2016.

[12] Zedboard. http://zedboard.org. Last accessed 28 November 2016.

[13] Zoosh technology. http://www.bdti.com/insidedsp/2011/07/28/naratte. Last accessed 28 November 2016.

[14] Bjørnø, L. Parametric acoustic arrays. In *Aspects of Signal Processing*. Springer, 1977, pp. 33–59.

[15] Deatherage, B. H., Jeffress, L. A., and Blodgett, H. C. A note on the audibility of intense ultrasonic sound. *The Journal of the Acoustical Society of America 26*, 4 (1954), 582–582.

[16] Dobie, R. A., and Wiederhold, M. L. Ultrasonic hearing. *Science 255*, 5051 (1992), 1584–1585.

[17] Fox, C., and Akervold, O. Parametric acoustic arrays. *The Journal of the Acoustical Society of America 53*, 1 (1973), 382–382.

[18] Goubran, R., and Botros, R. Adaptive sound masking system and method, June 5 2003. US Patent 20,030,103,632.

[19] Hamby, W. Ultimate sound pressure level decibel table, 2004.

[20] Heffner, H. E., and Heffner, R. S. Hearing ranges of laboratory animals. *Journal of the American Association for Laboratory Animal Science 46*, 1 (2007), 20–22.

[21] Huggins-daines, D., Kumar, M., Chan, A., Black, A. W., Ravishankar, M., and Rudnicky, A. I. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In *in Proceedings of ICASSP* (2006).

[22] Iannucci, P. A., Netravali, R., Goyal, A. K., and Balakrishnan, H. Room-area networks. In *Proceedings of the 14th ACM Workshop on Hot Topics in Networks* (2015), ACM, p. 9.

[23] Jacobs, I. M., and Wozencraft, J. Principles of communication engineering.

[24] Kim, S., Hwang, J., Kang, T., Kang, S., and Sohn, S. Generation of audible sound with ultrasonic signals through the human body. In *Consumer Electronics (ISCE), 2012 IEEE 16th International Symposium on* (2012), IEEE, pp. 1–3.

[25] Kumar, S., and Furuhashi, H. Long-range measurement system using ultrasonic range sensor with high-power transmitter array in air. *Ultrasonics 74* (2017), 186–195.

[26] Kune, D. F., Backes, J., Clark, S. S., Kramer, D., Reynolds, M., Fu, K., Kim, Y., and Xu, W. Ghost talk: Mitigating emi signal injection attacks against analog sensors. In *Security and Privacy (SP), 2013 IEEE Symposium on* (2013), IEEE, pp. 145–159.

[27] Lee, E. A., and Messerschmitt, D. G. *Digital communication*. Springer Science & Business Media, 2012.

[28] Lenhardt, M. L., Skellett, R., Wang, P., and Clarke, A. M. Human ultrasonic speech perception. *Science 253*, 5015 (1991), 82–85.

[29] Lyons, R. G. *Understanding Digital Signal Processing, 3/E*. Pearson Education India, 2004.

[30] McCalmont, A. M. Voice privacy system with amplitude masking, Mar. 25 1980. US Patent 4,195,202.

[31] Mercy, D. A review of automatic gain control theory. *Radio and Electronic Engineer 51*, 11.12 (1981), 579–590.

[32] Nakagawa, S., Okamoto, Y., and Fujisaka, Y.-i. Development of a bone-conducted ultrasonic hearing aid for the profoundly sensorineural deaf. *Transactions of Japanese Society for Medical and Biological Engineering 44*, 1 (2006), 184–189.

[33] Nakamura, T. Piezoelectric speaker, June 3 1986. US Patent 4,593,160.

[34] Nandakumar, R., Chintalapudi, K. K., Padmanabhan, V., and Venkatesan, R. Dhwani: secure peer-to-peer acoustic nfc. In *ACM SIGCOMM Computer Communication Review* (2013), vol. 43, ACM, pp. 63–74.

[35] Nation, P., and Waring, R. Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy 14* (1997), 6–19.

[36] NORRIS, E. Parametric transducer and related methods, May 6 2014. US Patent 8,718,297.

[37] Okamoto, Y., Nakagawa, S., Fujimoto, K., and Tonoike, M. Intelligibility of bone-conducted ultrasonic speech. *Hearing research 208*, 1 (2005), 107–113.

[38] Pérez, J. P. A., Pueyo, S. C., and López, B. C. Agc fundamentals. In *Automatic Gain Control*. Springer, 2011, pp. 13–28.

[39] Sherif, M. H. *Protocols for secure electronic commerce*. CRC press, 2016.

[40] Tretter, S. A. *Communication System Design Using DSP Algorithms: With Laboratory Experiments for the TMS320C6713TM DSK*. Springer Science & Business Media, 2008.

[41] Wang, Q., Ren, K., Zhou, M., Lei, T., Koutsonikolas, D., and Su, L. Messages behind the sound: real-time hidden acoustic signal capture with smartphones. In *Proceedings of the 22nd Annual*

International Conference on Mobile Computing and Networking (2016), ACM, pp. 29–41.

[42] Westervelt, P. J. The theory of steady forces caused by sound waves. *The Journal of the Acoustical Society of America 23*, 3 (1951), 312–315.

[43] Westervelt, P. J. Scattering of sound by sound. *The Journal of the Acoustical Society of America 29*, 2 (1957), 199–203.

[44] Whitlow, D. Design and operation of automatic gain control loops for receivers in modern communications systems. *Microwave Journal 46*, 5 (2003), 254–269.

[45] Xiong, F. *Digital modulation techniques*. Artech House, 2006.

[46] Yang, J., Tan, K.-S., Gan, W.-S., Er, M.-H., and Yan, Y.-H. Beamwidth control in parametric acoustic array. *Japanese Journal of Applied Physics 44*, 9R (2005), 6817.

[47] Yoneyama, M., Fujimoto, J.-i., Kawamo, Y., and Sasabe, S. The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design. *The Journal of the Acoustical Society of America 73*, 5 (1983), 1532–1536.

[48] Zhang, L., Bo, C., Hou, J., Li, X.-Y., Wang, Y., Liu, K., and Liu, Y. Kaleido: You can watch it but cannot record it. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (2015), ACM, pp. 372–385.