Hard Drive of Hearing: Disks that Eavesdrop with a Synthesized Microphone

Andrew Kwong[1], Wenyuan Xu[2], and Kevin Fu[1]

[1]University of Michigan spqr.eecs.umich.edu
[2]Zhejiang University usslab.org

***Abstract*—Security conscious individuals may take considerable measures to disable sensors in order to protect their privacy. However, they often overlook the cyberphysical attack surface exposed by devices that were never designed to be sensors in the first place. Our research demonstrates that the mechanical components in magnetic hard disk drives behave as microphones with sufficient precision to extract and parse human speech. These unintentional microphones sense speech with high enough fidelity for the Shazam service to recognize a song recorded through the hard drive. This proof of concept attack sheds light on the possibility of invasion of privacy even in absence of traditional sensors. We also present defense mechanisms, such as the use of ultrasonic aliasing, that can mitigate acoustic eavesdropping by synthesized microphones in hard disk drives.**

1. Introduction

Magnetic hard disk drives (HDDs) continue to persist in everything from legacy laptops to server racks. Because of their critical role in a wide variety of applications, hard drives make an appealing target for both cyber criminals and nation states alike. Kaspersky describes how an advanced hacking organization, dubbed the "Equation Group," developed malware that reflashes its host machine's hard drive's firmware to gain advanced persistence [1]. Other researchers have shown how even modestly funded adversaries can compromise HDD firmware to create highly stealthy backdoors for subverting machines [2].

While these incidents have motivated researchers to investigate the threat of hard drive malware from the digital side, little attention has been given to the cyberphysical side-channel attack surface exposed by a hard drive's mechanical components. Due to the complexity of their read/write mechanisms and the granularity at which a hard drive must track its head, hard drives possess certain characteristics that respond to the oscillations in air pressure caused by acoustic waves. This raises the possibility of using a hard drive as an unintentional microphone, thereby allowing attackers to eavesdrop on speech in the vicinity of the drive.

In this paper, we demonstrate how an adversary could leverage HDD firmware resident malware to extract human speech by measuring the offset of the read/write head from the center of the track that it is seeking. Modern hard drives use this offset, known as the Position Error Signal (PES), in a feedback control loop; the microprocessor takes the PES (Figure 1) as input for actuating the read/write head by use of a voice-coil motor (VCM) [3].

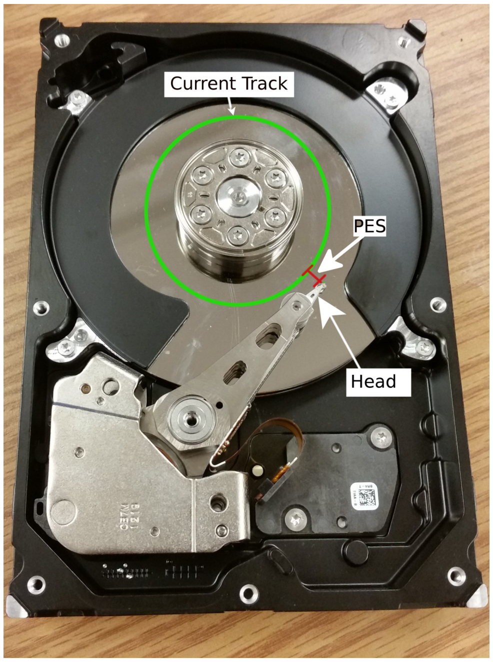


Figure 1: **The Position Error Signal (PES) measures the offset of the read/write head from the center of the track.**

For both read and write operations, the head can tolerate deviation from the center only on the order of nanometers. Accordingly, PES measurements are taken at a very fine granularity. These extremely precise measurements are sensitive to vibrations caused by the slightest fluctuations in air pressure, such as those induced by human vocalizations.

Extracting speech from the PES, however, is complicated due to a weak signal-to-noise-ratio (SNR). Imperfections in the eccentricity of the platters, thermal drift, and turbulence from the rapid rotation of the disks all contribute to a large quantity of noise in the signal [4]. Through a mixture of digital filtering techniques in both the time domain and the frequency domain, however, we have managed to sufficiently clean the signal such that human speech can be completely reconstructed under certain conditions.

To prove the existence of this acoustic side-channel, we physically probed the PES directly from the hard drive under operation. This is sufficient to explore the possible information leakage available to an attacker with firmware access.

We validate our side-channel attack by performing signal analysis and using Shazam to recognize songs recorded through the hard drive's PES. Furthermore, we evaluate the intelligibility of the reconstructed speech signal through objective measures of speech quality [5] [6].

Our attack sheds light on the potential for invasion of privacy even in the absence of traditional sensors. Security conscious individuals may make efforts to remove or disable any and all sensors (i.e. placing tape over a laptop's camera, removing the built-in microphone, etc.), but they will often neglect the possibility of how non-sensors can be synthesized into intrusive sensors: in our case, hard disk drives into microphones. Our contributions towards addressing the cyberphysical side channels in HDDs are the following:

- We model how the mechanical components in hard disk drives lend themselves into becoming unintentional sensors.
- A proof of concept attack demonstrates how an attacker can use a hard disk drive as a microphone to extract human speech. We evaluate our side-channel by (1) performing qualitative signal analysis, (2) using the Shazam service to identify songs recorded through the hard disk drive, and (3) quantitatively analyzing recovered audio through objective measures of speech intelligibility.
- We discuss defenses in both software and hardware, and make suggestions on how manufacturers and end users can mitigate risks with ultrasonic masking and sound dampening.

2. Background

Synthesizing a microphone from a hard drive relies on the similarities between its mechanical components and those of a microphone.

2.1. Acoustic Waves and Microphones

Human speech is entirely encoded in acoustic waves that propagate as oscillations in air pressure. As such, microphones record audio signals by measuring these small changes in air pressure. They can accomplish this through use of a diaphragm that oscillates back and forth with the fluctuations in air pressure induced by acoustic waves; the microphone then produces as output a voltage in proportion to the distortion of the diaphragm. This analog value, taken as a function of time, then represents the oscillations in air pressure that compose the acoustic wave.

2.2. Hard Drive Mechanics

Hard drives read and write from the magnetic platters by making use of a small magnetic slider, called the read/write head, that floats just 5 nm above the surface [7]. This head must follow the center of the track with extreme precision, and in the case of high performance drives, can deviate from the center of the track by no more than 7 nm [8] [9]. As such, HDD's make use of a high precision feedback control loop wherein the offset from the center of the track, called the Position Error Signal (PES), is fed back to the microprocessor so that it can actuate the read/write arm with a voice coil motor (VCM). As shown in Figure 1, the hard drive controller computes the PES by reading out magnetic signals known as "servo bursts" from special sectors on the disk, called servo sectors. Each track contains the same number of servo sectors, and within a given track, the servo sectors are laid out in even intervals [10]. This gives the PES its periodicity, with the frequency proportional to the angular velocity of the disk.

Keeping the read/write head within the allowable margins is a challenging task due to a plethora of noise sources, all of which contribute to disk run-out, which is a measure of how much the slider's rotational path differs from a perfect circle. Disk run-out due to imperfections in the eccentricities of the platters, along with turbulence from the spinning disks, creates white noise that falls out across the spectrum. Furthermore, expansion in aluminum components due to thermal drift can result in alterations in PES 300 to 400 times greater than the width of a track [8]. Finally, acoustically induced vibrations, which are the signals we wish to measure for our attack, also work to push the head off track. Thus, the read/write head assembly approximately functions as a crude diaphragm.

3. Eavesdropping

This section describes our assumptions on the attacker, and details how such an adversary might carry out our acoustic eavesdropping attack.

3.1. Threat Model

We assume the attacker can reflash the HDD's firmware; this is necessary because the ATA protocol does not expose the PES. An attacker can gain this privilege in one of two primary ways: reflashing it entirely through software, or by intercepting HDDs before they reach the end user.

With reflashing, the attacker can use traditional methods such as binary exploitation, drive-by downloads, or phishing attacks, to compromise the operating system upon which the HDD is attached. Then, the attacker abuses his root privileges to update the firmware over the SATA connection. This is what Kaspersky says the Equation Group accomplished with its Grayfish trojans [1]. By reverse engineering an off-the-shelf HDD, Zaddach et al. [2] demonstrate how even modestly funded attackers can practically carry out such an operation. We emphasize that even on a device that typically has a microphone installed in it, having root access on the device does not necessarily give the attacker access to a microphone; a privacy-minded user could have disabled the microphone in the BIOS, or even have physically disconnected it. Security conscious individuals have even taken to modifying their devices by adding a switch that only closes the microphone's circuit when switched on [11].

The adversary can also gain firmware access by conducting man-in-the-middle (MITM) attacks against users attempting to download legitimate firmware updates for their hard drives. Even when manufacturers deploy security sensitive downloads with SSL/TLS, they are still potentially unsafe, as the POODLE [12], LOGJAM [13], and DROWN [14] attacks against the SSL and TLS protocols themselves have demonstrated. In a related attack vector, our adversary can employ social engineering and spear phishing techniques in order to direct naive users to attacker controlled websites, where they then proceed to download malicious firmware updates.

The attacker can gain access to the HDD's firmware by intercepting the HDD and planting malware on the device before it ever reaches its destination. For instance, nation states have reportedly intercepted routers [15] and CD-ROMs [16] to plant malware. Furthermore, physical access to HDDs at the factory itself places HDDs at risk of tampering; in 2007, Seagate Maxtor drives manufactured in China shipped with preinstalled malware [17]. In this scenario, even full disk encryption does not mitigate our attack. Since full disk encryption aims to prevent data theft by storing the encryption keys separately from the data on the disk, the hard drive malware would be unable to tamper with user files and subsequently gain access to the microphone. They are, however, still able to record audio and write it to disk, to be recovered at a later time.

In both cases, we make the assumption that no digital signatures are in use. We believe that this is a mild assumption, as only a few of the newer models of HDDs implement this security feature; none of the HDDs that Zaddach et al. [2] reverse engineered signed their firmware updates.

Given that our attack is largely OS independent, we make no assumptions on which operating system is in use. The attacker's goal is simply to reconstruct human speech spoken in proximity to the hard drive.

3.2. Speech Exfiltration

There are two primary ways by which the attacker can exfiltrate the data recovered through our attack: (1) through a reverse shell over the internet, and (2) by storing the audio on disk and physically recovering it later.

Reverse Shell Exfiltration. Once the attacker has firmware access, he can leverage the compromised drive to exfiltrate the recorded audio over the internet, thereby leaving no physical trace of the attack. When the hard drive, which acts as the basis of trust, is under the attacker's control, there are a number of ways by which an attacker can accomplish this. One such example: by modifying the `.bashrc` and `/etc/shadow` files on a Linux machine, the malicious drive can trivially establish a reverse shell with root privileges for the remote attacker. Zaddach et al. [2] demonstrated the practicality of such a backdoor by implementing a similarly stealthy covert channel, in which a remote attacker was able to read and write arbitrary blocks to a back-doored hard drive over the internet.

Now that the attacker has privileged arbitrary code execution, it is then a simple matter for the malicious firmware on the hard drive to stream the captured audio over its SATA or SCSI interface to the attached machine, which the remote attacker then extracts over the internet via the reverse shell.

Recording to Disk. If the reverse shell is not an option, perhaps because the hard drive is installed in an air-gapped system, the malicious firmware can instead store the audio on the disk itself. The firmware can accomplish this covertly by writing the captured audio to the System Area, which can contain over 400 MB of unused space [18]. The advantage of using this reserved space is that its blocks are not exposed via any external interface, thus rendering it hidden to antivirus and forensics tools alike. Kaspersky explains how the Equation Group used the System Area for this purpose, so as to covertly store data to be recovered at a later time [1].

4. How a Hard Drive Hears

Given the physical structure of a hard drive's read/write components, we hypothesize that the PES measurements from the head can be interpreted in the same manner that the values out of a microphone's analog to digital converter (ADC) are. That is, an acoustic wave's oscillations in air pressure will displace the head in the same manner that acoustic waves oscillate a microphone's diaphragm. Thus, the PES readings will directly approximate an acoustic wave's instantaneous amplitude.

Since it is unclear what the exact relation is between this acoustic interference and the PES, we tested this hypothesis using an HDD and measured its PES under various external acoustic inputs. We chose the Seagate Barracuda 7200.12 1TB hard disk as our target because, after examining all of the Seagate F3 drives, it was the only one that exposed the pin by which we extract the PES.

4.1. Measuring the Position Error Signal

In our threat model where the adversary has firmware access on the hard drive, the microcontroller reads out and extracts the full 16-bit PES. Patents from Seagate [19] and Western Digital [20] [21] describe how the HDD's main microcontroller is responsible for reading the PES offsets and then computing the required adjustments, thereby demonstrating that the PES is indeed available to the HDD's main microcontroller. We verified this to be the case with the Barracuda 7200.12 by commanding the hard drive controller to output various aggregate statistics on the PES over a specified number of revolutions, and even received ASCII art representations of the PES over the serial diagnostic port, which connects directly to the hard drive controller.

In our case, however, we instead measured the PES by partially reverse engineering the hard drive's debugging interface, which exposes sufficient information for exploring the side-channel leakage available to an attacker with complete firmware access. So as to illustrate some of the limitations imposed by this approach, we briefly describe the process for pedagogical reasons.

We first attached to the hard drive's serial diagnostic port. The interface allowed us to command the drive to output 8 bits at a time of the 16-bit PES on the "AMUX" pin. While it is likely that we would obtain a higher signal to noise ratio if we had access to all 16 bits, we found that 8 bits was sufficient as a proof of concept. To find this pin, we first probed all exposed pads on the HDD's printed circuit board and observed the output under an oscilloscope. Then, to narrow down our list of candidates for the desired pin, we computed the expected frequency by the following reasoning: if we know that for a particular HDD, n is the number of servo sectors per track, then the read/write head will pass over n servo sectors per revolution, and thus report n evenly spaced samples per period. Given that the platters rotate at f_r revolutions per minute (RPM), we can then compute the frequency f_s as $f_s = f_r \cdot n$. In the case of the Barracuda 7200.12, with parameters $f_r = 7200$ RPM and $n = 288$, we can compute the rate at which our signal is sampled by the following:

$$\begin{aligned} f_s &= f_r \cdot n \\ &= 120 \text{ Hz} \cdot 288 \\ &= 34,560 \text{ Hz} \end{aligned}$$

As such, we expect to see a square wave of frequency 34,560 Hz on the output of the "AMUX test pin." After probing all exposed pins, we found only one such candidate; we confirmed our guess by toggling the PES output and verified that the output of the pin also toggled on and off. With the AMUX pin identified, we were able to read out 8 bits of the hard drive's real-time PES.

4.2. Sampling Rate

Using the frequency f_s from the previous subsection, we find that the sampling rate comes out to 34.56 kHz.

The Nyquist-Shannon sampling theorem states that, given a signal where the highest frequency component is f, sampling the signal at a rate of $2f$ is sufficient to reconstruct the signal. Thus, the hard drive's sampling rate can perceive signals of up to $\frac{34,560}{2}$ Hz= 17.28 kHz.

This covers almost all of the audible range in humans, which spans from 20 Hz–20 kHz; furthermore, given that the Plain Old Telephone System (POTS) uses a sampling rate of just 8 kHz, our sampling rate is more than sufficient to eavesdrop on human speech. Having an even higher sampling rate than POTS offers the additional advantage of reducing the chances of noise above the Nyquist threshold aliasing in the band of our expected signal. While telephony system designers do not typically expect a significant amount of very high frequency noise to be present, there is no reason to assume the same for a hard drive's PES. In fact, as the spectrogram of the PES in Figure 2a shows, a sampling rate of just 8 kHz would have resulted in the dark red line just above 8 kHz aliasing over our signal in the 80–260 kHz band. This is known as "out-of-band" noise.

4.3. Sampling Granularity and Resolution

Because of the hard drive's need for extreme precision in actuating the read/write arm, the PES is sampled at both a very fine granularity and a high resolution. Technical specifications from Seagate show that PES is a 16-bit value, with a granularity equal to just $\frac{1}{2^{12}}$ of the width of one track.

If we desire even more resolution, we can leverage the high sampling rate to oversample and obtain a higher effective resolution [22]. In practice, however, we find no need for the additional resolution.

Due to the manner in which we extract the PES from the hard drive, we can only read out 8 bits per sample. This is problematic; if we extract the low 8 bits of the PES we see substantial clipping under normal operation, which leads to degradation of the signal and the introduction of undesirable harmonics. If we examine only the high 8 bits, minor disturbances in the PES are masked, as is most of our signal. Thus, we must make a trade off between granularity and distortion of the signal. Experimentally, we have found that using the 3rd least significant bit through the 10th least significant bit yields the best compromise, as it is the lowest set of bits such that no clipping occurs. While this effectively reduces our granularity by a factor of four, signal clipping is minimal. Under our threat model, however, the attacker would have access to the full 16-bit signal, and would likely be able to recover a cleaner signal; we leave this possibility to future research.

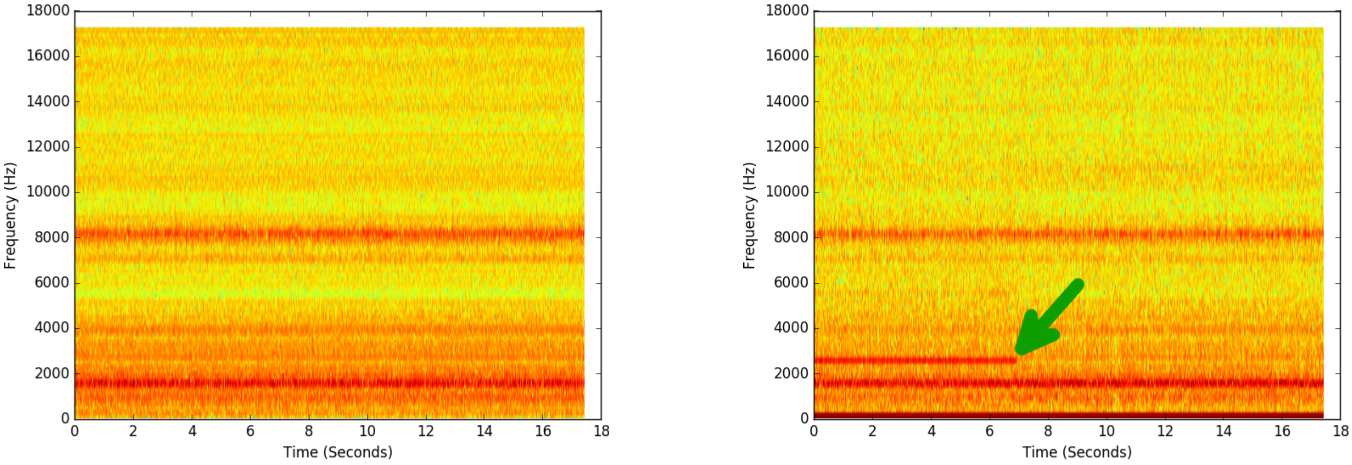
Given the sampling rate of 34.56 kHz and a 16-bit resolution, SATA 3.0's native transfer rate of 6.0 Gbit/s is more than sufficient for streaming captured audio.

4.4. Linearity

In order to function like a microphone, our hard drive blackbox must approximate a linear time invariant (LTI) system. The actuator's attempts to minimize the PES by use of the voice coil motor have the potential to compromise the "sinusoidal fidelity" of the system; that is, an input to the system of a given frequency may not yield an output of equal frequency [23]. To test this property of our hard drive, which is essentially a blackbox, we subjected it to varying frequencies and observed an increase in the corresponding bands in the PES's frequency spectrum. The PES responded strongly to 2.5 kHz, which guided our decision to conduct further tests with this frequency. The spectrogram in Figure 2b confirms our assumption of linearity.

4.5. Noise

As can be seen in Figure 2, the heaviest bands of noise are concentrated just above 8 kHz and just below



(a) **Baseline spectrogram of the position error signal. The baseline exhibits heavy, persistent noise just above 8 kHz and below 2 kHz.**

(b) **Spectrogram of the position error signal while the drive is subjected to a 2.5 kHz tone for the first 7 seconds. As indicated by the green arrow, it is easy to see when the tone stops.**

Figure 2: **Comparison of PES spectrograms with and without being subject to an external tone**

2 kHz. Since these bands don't overlap with the fundamental frequencies of adult speech, this noise can be removed by using linear filtering techniques.

Even after using such filters, however, a substantial amount of noise remains in the signal's frequency band. We speculated that noise due to disk runout is periodic with the rotation of the disk and thus can be reduced with signal subtraction in the time domain. After looking at PES averages for a given servo sector over many revolutions, however, we found that the baseline noise is uniformly distributed and as such requires a different approach. Our non-linear filtering techniques are described in Section 5.2.

4.6. Directionality and Orientation

To investigate how the hard drive's orientation and the direction of the oncoming waves affected our measurements, we played a 2.5 kHz tone at equal intensities from all five exposed sides while the bare hard drive assumed three different orientations: face up, face down, and on its side. We then repeated the measurements, only with the hard drive housed in the CSE-M35T-1B external HDD enclosure. The high frequency of the 2.5 kHz tone aids in this measurement, due to how acoustic waves diffract. Lower frequencies "bend" around objects more easily; by using a high frequency tone, we minimize this diffraction, and thus isolate the directionality of the tone.

While observing the response in the frequency domain, we found that the hard drive's PES responded very poorly to tones coming from the sides of the hard drive. Given that the PES is a measurement of the read/write head's horizontal, and not vertical, displacement, this is surprising. It is likely that the weak response is then due to the hard drive's thin profile from the side capturing less energy from the oncoming wave, as the wave vibrates less of the drive's surface area. The PES's response to tones coming from the bottom and top of the hard drive was substantially stronger, as the results in Table 1 show.

We did not observe any noticeable differences in the three different orientations of the hard drive. This is unsurprising, given that user manuals for hard drives commonly state that as long as the drive is placed on a flat surface, orientation does not impact the drive's operation.

5. Speech Recovery

This section details how an adversary with access to the PES can begin to extract human speech.

5.1. Audio Extraction

If HDD components do indeed function sufficiently as a microphone, the PES values will roughly approximate instantaneous air pressure readings; furthermore, the PES sampling rate of 34.56 kHz is more than sufficient to extract speech, given that the Plain Old Telephone System uses a sampling rate of 8 kHz. This allows us to treat the string of PES readings as linear pulse-code modulation values, corresponding to samples of an audio signal. We can simply write the PES values into a WAV file with the appropriate sampling rate, and then use digital signal processing algorithms designed for speech recognition.

5.2. Digital Signal Processing

The unprocessed signal taken from the HDD is incredibly noisy, and without further processing, the raw audio is completely unintelligible. However, by exploiting certain spectral properties of human voice and making use of both linear and non-linear filtering algorithms, we demonstrate

	Bare Drive			Enclosed Drive		
	bottom	side	top	side	front	top
Response	13 dB	9 dB	16 dB	10 dB	21 dB	18 dB

TABLE 1: **dB increase in corresponding band while playing 2.5khz tone at 90 dBA from different directions.**

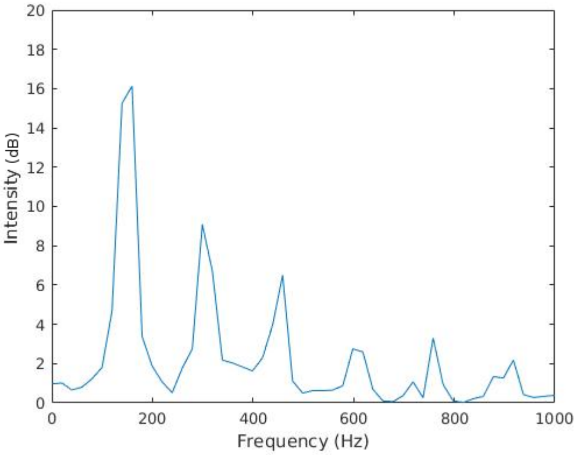


Figure 3: **Frequency spectrum of the male speaker's voice, taken over a 50 ms window. When observing sufficiently short segments, the spectrum of human voice exhibits peaks in smaller sub-bands. A time-variant filter can pass these peaks while rejecting the troughs.**

that the signal can be cleaned up to a sufficient extent to parse human speech.

Filtering. The naive, initial approach to any noise reduction problem is to examine the frequency domain and cut out the unwanted frequencies. The spectrogram in Figure 2 reveals heavy, persistent noise in the PES under standard operation, particularly around 8.1 kHz and 1.8 kHz. Since an average male's fundamental frequency lies between the range of 85 Hz to 180 Hz, while a female's lies between 156 Hz and 255 Hz [24] [25], we can easily remove these with a linear bandpass finite-impulse-response filter that passes only frequencies between 80 Hz and 260 Hz. Even after linear filtering, however, there remains a substantial amount of noise that overlaps with our passband, and as such can not be dealt with by a linear filter.

Given that the white noise and our signal overlap in both the time and frequency domains, we make use of a non-linear time-variant filter. To attenuate the in-band noise, we use a technique known as spectral noise gating. As illustrated in Figure 3, while the spectra of human speech and white noise may overlap over the length of a recording, the human speech's energy is largely concentrated in separable bands when observed over a very short frame, i.e. 16 ms. This is in contrast to white noise, which remains spectrally flat when looking at both short and long segments.

Thus, with spectral noise gating we take advantage of the briefly separate spectrums of noise and signal and use a time variant filter that operates over short windows, passing frequencies above the noise floor and rejecting others [23].

6. Evaluation

Despite the fact that hard drives were not designed to function as microphones, the mechanics of their internal components allow them to sense acoustic waves to some degree. To understand the limits of what a hard drive can hear, we explore three major questions:

- What are the physical limits of the PES in detecting acoustic waves? (clear information leakage at 75 dBA)
- How difficult is it to automatically recognize structured sound patterns (e.g., music)? (Shazam recognizes at 90 dBA)
- How difficult is it to recover unstructured conversations? (yields speech recordings recognizable to human ear at 85 dBA)

Our evaluation answers these questions by performing signal analysis on the input and output to the hard drive; by using objective measures of speech quality for spoken phrases; by testing how well the Shazam service can recognize music recovered from the PES; and by analyzing the feasibility of advanced techniques such as acoustic arrays of hard drives.

6.1. Experimental Setup

In all of our experiments, the hard drive lies enclosed in the CSE-M35T-1B SuperMicro HDD enclosure, which comes attached with a San Ace 92 9GV0912P1H03 8500 RPM 42 Watt fan [26], mimicking the typical usage of an external hard drive or server rack. In this section, in addition to the baseline testing, we present results in which we drive the fan separately at max power so as to simulate an exaggeratedly loud internal fan that may be present in certain desktops or datacenters; furthermore, we conduct experiments while continuously writing a large file to the HDD.

When playing our audio samples at 75 dBA, which is comparable to a loud conversation, we are able to recover muffled recordings; however, in order to yield a large signal to noise ratio (SNR) for the purpose of demonstrating our proof of concept attack, our audio samples are played at a volume of 85 dBA. While this is louder than what can be expected in most practical scenarios, we aim only to demonstrate the presence of such a side-channel, and expect that an attacker using state of the art filtering and voice recognition algorithms can substantially amplify the channel's strength.

In our setup shown in Figure 4, we are careful to physically separate the hard drive from the speaker so as

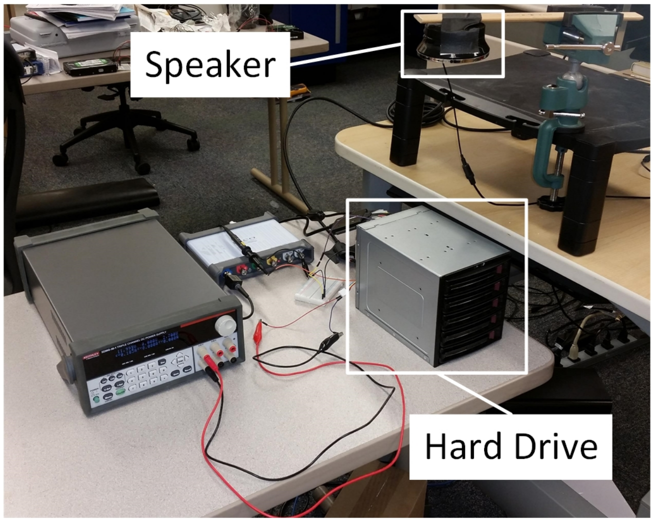


Figure 4: **The speaker is positioned 10 inches directly above the HDD enclosure. It is suspended from a ruler that is attached to a neighboring, but not connected desk, so as to eliminate any mechanical coupling between the hard drive and the speaker.**

to eliminate any possible mechanical coupling. When a speaker projects a tone, the rapid oscillation of its diaphragm vibrates the speaker itself, and transfers this energy into the platform upon which it rests. This energy can then potentially transfer into the hard drive, and thus vibrate the read/write head. We accomplish this separation by suspending the speaker from a ruler, which is mounted on a clamp attached to a separate table. This assures that the transmission of our audio is solely the result of acoustic waves, and not from the vibration of a shared surface.

The samples are recordings of Harvard sentences [27], which are specifically designed to feature phonemes at the same frequency that they naturally appear in spoken English. The female sample is from list 1, while the male sample is from list 57. The specific audio samples are taken from the Open Speech Repository [6].

All tests were conducted using the Seagate Barracuda 7200.12 1 terabyte hard disk, due to the fact that it was the only drive that has the "AMUX pin" exposed on the top side of the printed circuit board. We leave the question of whether or not certain drives yield better results to future work.

6.2. Signal Analysis

A simple side-by-side comparison of the time domains of the original signal and the one extracted by the hard drive presents clear evidence of information leakage. In Figure 5, both signals are from the male sample; the signal on the top is the filtered and processed audio extracted from the hard drive, while the bottom is of the original audio sample. The speech, as annotated, was taken from list 57 of the Harvard sentences [27]:

Paint the sockets in the wall dull green. The child crawled into the dense grass. Bribes fail where honest men work. Trample the spark, else the flames will spread.

The spikes in amplitude seen in the recovered signal clearly align with corresponding spikes in the original. To quantify the similarity between the input and the output to the hard drive, we also computed the discrete cross correlation between the two time series, as shown in Figure 5. The cross correlation, as defined by the following formula:

$$(f \star g)[n] \overset{\mathrm{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m]\ g[m+n].$$

is a measure of the dot-product of the two signals as a function of time displacement, also known as lag. As such, the sliding dot product will hit its maximal value when the peaks align with peaks, and troughs with troughs. The large spike when the lag is equal to zero indicates a strong correlation between the unshifted original and recovered audio. This is to be expected, as the two audio samples are already time aligned. Furthermore, this result agrees with our assessment of the hard drive microphone's linear properties. Billing's book on nonlinear system identification [28] states that cross correlations between the input and output of nonlinear systems can yield false positives, even in the presence of strong correlations.

We can observe the response of the hard drive in Figure 6, which displays the power density spectrums of both the source signal and of the recovered audio through the hard drive. The lack of any energy beyond 4 kHz is due to the fact that the source audio is a WAV file that was recorded at a sampling rate of 8 kHz. Thus, the presence of such noise in the recovered audio is a combination of noise from the hard drive itself and artifacts of the spectral noise gating process. Such artifacts can arise from discrimination errors, wherein bands within the signal fall too close to the noise floor and are subsequently filtered out.

Also of note is the severe attenuation of the lower frequencies compared to the higher bands. This treble heavy response is likely a consequence of vibrational resonance. A system's natural frequency is the frequency at which the system vibrates when the vibrating force is removed, and resonance occurs when the input force oscillates at a frequency that divides the natural frequency. At these resonant frequencies, input to the system results in particularly high amplitude responses.

In Dutta's dissertation [29] on hard disk performance in the presence of noise, he utilized finite element analysis to show that the hard drive components that affect the read/write head respond most to frequencies in the 2–8 kHz range, corresponding to their natural frequencies. We then conclude that resonance is largely responsible for the hard drive's treble heavy response.

Previous research has demonstrated that a substantial amount of information can be recovered simply by ob-

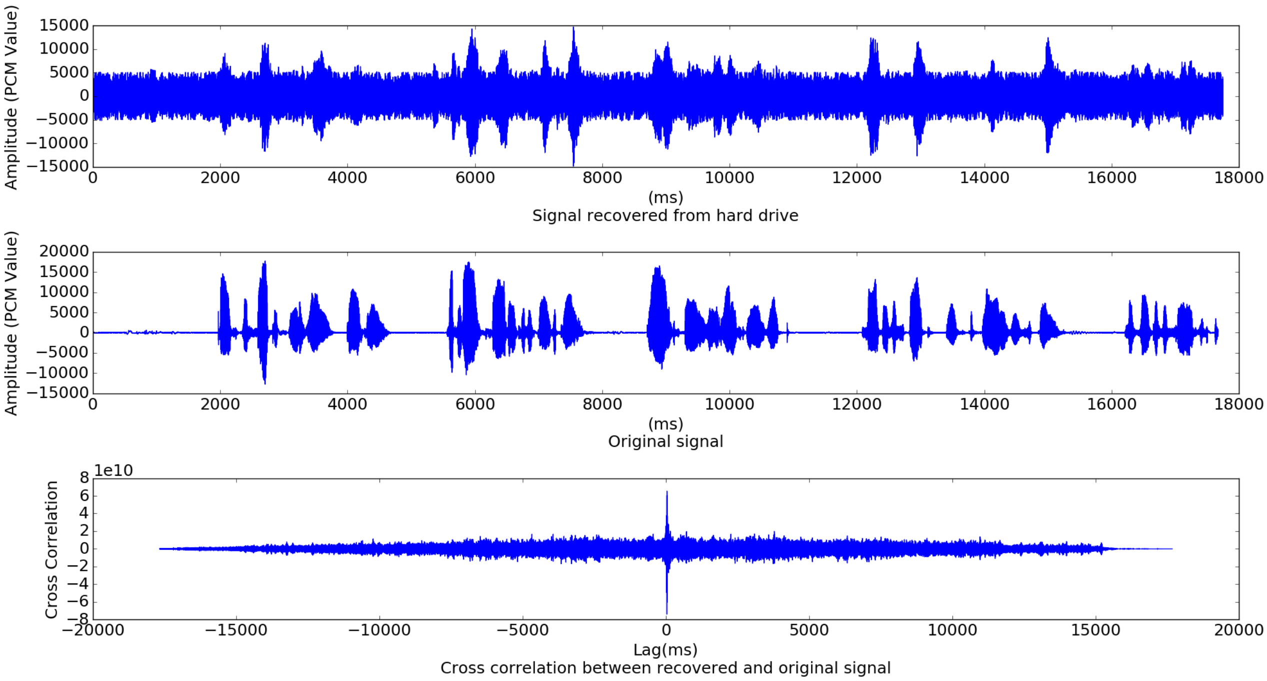


Figure 5: **Time domain comparison of the original audio and the recovered audio after it has been cleaned through digital signal processing techniques. The recording is of the first two Harvard sentences from list 57, spoken by a male. The bottom graph plots the cross correlation between the two signals, with a sharp spike at lag=0 seconds. This indicates a strong correlation between the two signals.**

serving very coarse measurements of human speech [30] and using pattern recognition and knowledge of the spoken language's structure. In our case, however, we are actually able to listen to the processed audio and successfully parse the recorded speech.

We qualitatively found that placing the drive in the enclosure actually amplified the audio signal. This somewhat counterintuitive result can be potentially attributed to the observation that the larger size of the chassis presents a greater surface area to oncoming waves, thereby enabling it to absorb more energy and then transmit it to the drive within. It is also possible that the tighter, confined space within the enclosure causes the waves to reflect and superpose, thereby increasing the amplitude.

Additionally, while driving the fan at maximum power to simulate a noisy environment, we found that the intelligibility of the signal degraded only slightly. By observing the frequency spectrum of the PES while the fan was active, we found that the noise due to the fan results in very narrow peaks at 200 Hz and its harmonics; as such, despite the fan being quite loud, the periodic nature of the noise is easily filtered. Likewise, recording the PES while writing a large file to the HDD resulted in very little loss of intelligibility. The only noticeable degradation was a moderately periodic clipping noise, which we attribute to the read/write head seeking between tracks.

We subjectively found that it is actually the male voice that is more intelligible. Given the treble heavy response of the hard drive, this is surprising; we attribute this to the heavy band of noise just below 2 kHz present in the baseline. Even though virtually no adult female's fundamental frequencies will fall in that range, a female voice's harmonics overlap that range to a much stronger extent. As those harmonics are filtered out along with the noise, the recorded audio loses enough intelligibility such that it is actually rendered less intelligible than that of the male. This hypothesis is in agreement with the results seen in Figure 6, wherein a large dip in the 2 kHz band can clearly be seen.

Sample audio recordings can be found at, https://www.dropbox.com/sh/q5du7yzcoenq5u6/AACsr-cDTRy7xxKBIWAfv_UUa?dl=0. We back up our qualitative claims with objective measurements.

6.3. Signal to Noise Ratio

The first manner in which we quantitatively characterized the effectiveness of our side-channel was by making use of the Laboratory for the Recognition and Organization of Speech and Audio's SNReval Objective measures of speech quality/SNR [5]. These measures are MATLAB script implementations of commonly used measurements compiled from academic research, and are designed for characterizing distorted speech. Following is a brief description of the measures:

1) **NIST STNR:** NIST Signal to Noise Ratio. Estimates the signal to noise ratio as

$$\frac{peak\_speech\_power}{mean\_noise\_power}$$

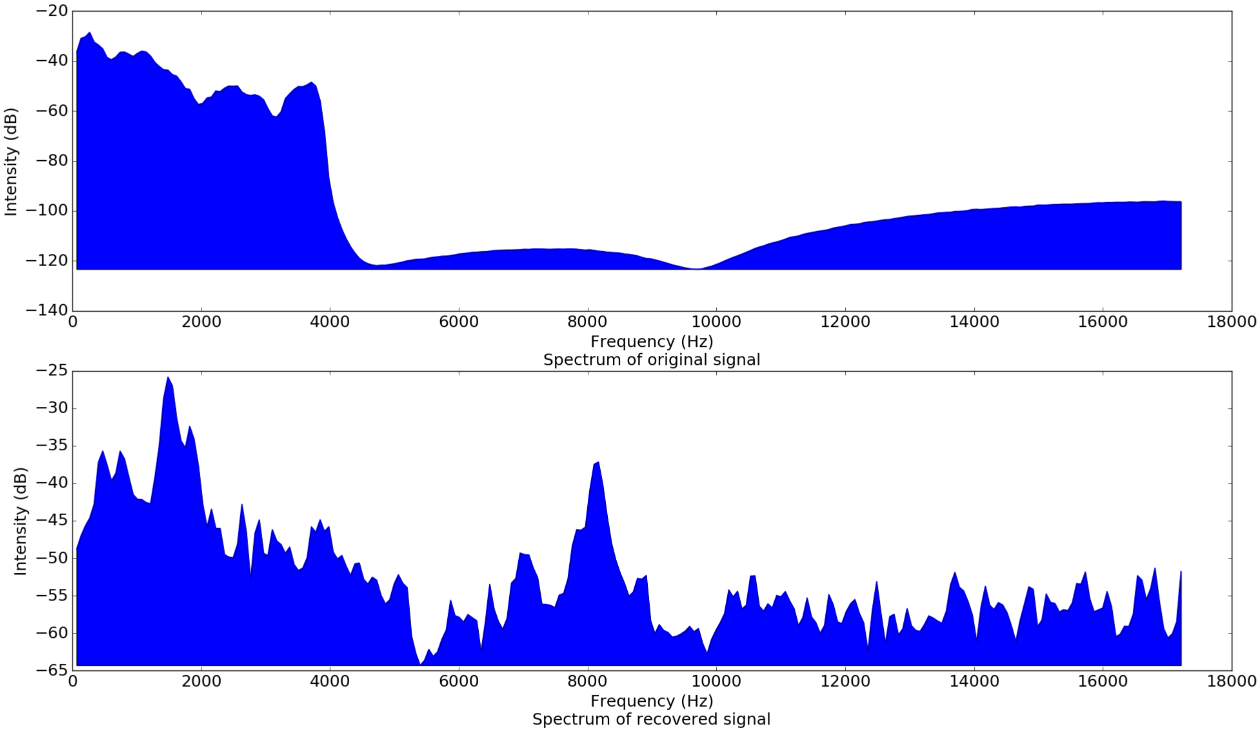


Figure 6: **Frequency spectra of the original signal, on top, and of the recovered signal, on the bottom. Note the treble heavy response.**

where power is computed as the variance of a given signal

2) **PESQ MOS:** Perceptual Evaluation of Speech Quality Mean Opinion Scores. Estimates intelligibility of speech, and is recommended by the International Telecommunication Union's Telecommunication Standardization Sector [31].

Between these two measures, PESQ is most closely correlated with the intelligibility of human speech [32]. The values obtained from computing these measures over the recovered audio obtained through the hard drive are displayed in Table 2; for reference, we also computed the same values over recordings obtained through an actual microphone, as displayed in Table 3, and through the bare drive, while it was not housed in an enclosure in Table 4.

As expected, both the female and male recordings degrade to some degree when recorded through the hard drive as opposed to a real microphone. While the SNR measurement exhibits a large drop, PESQ drops by a lesser extent. This is significant because the potential of the side channel is most closely aligned with the intelligibility of the recovered audio, and not the presence of objectionable noise. This also supports our previous qualitative claims, as the PESQ MOS values don't vary much in both the setup where the fan is driven at max power and the setup where a large file is continuously written. This minimal loss in fidelity enables the attacker to write the recorded audio to disk, to be retrieved at a later time if extraction over the internet is not possible.

6.4. Shazam Recognition

We further validate our side-channel by using the mobile application Shazam to identify songs played at the hard drive. By doing so, we demonstrate the possibility of using the hard drive as a microphone to match audio patterns.

Shazam operates by extracting and storing the most robust features of over 8 million songs and storing them in a database [33]. They accomplish this by computing spectrograms of the songs, and then reducing the songs to a series of the spectrograms' peaks. These "spectral fingerprints" serve as references against which the audio sample in question is compared. Thus, Shazam's recognition algorithm amounts to identifying which frequencies are present in the audio.

To test the hard drive's limitations at recording complex audio, we played Iron Maiden's song, "The Trooper," and used the hard drive to record it. We found that when played at 90 dBA, Shazam was able to correctly identify the song from the recording, despite the audio sounding like completely unintelligible noise to a listener's ear. While both powering the fan at maximum power and writing continuously to the HDD, the threshold increased to 94 dBA.

Notably, the requisite amplitude is substantially higher than what we played the recordings of human speech at. There are a few explanations for why such a high sound pressure level is required. The primary reason is that we are unable to use the DSP techniques described previously

	Baseline		Under Write Workload		Fan at Maximum Power	
	NIST STNR	PESQ MOS	NIST STNR	PESQ MOS	NIST STNR	PESQ MOS
Male Speaker	8.0 dB	1.7 dB	11.2 dB	1.7	7.8 dB	1.6
Female Speaker	6.2 dB	1.7 dB	7.8 dB	1.5	3.5 dB	1.7

TABLE 2: **SNReval measurements while using the hard drive in the enclosure to record audio.**

	NIST STNR	PESQ MOS
Male Speaker	11.8 dB	1.8
Female Speaker	12.8 dB	2.1

TABLE 3: **SNReval measurements when using a microphone to record audio.**

	NIST STNR	PESQ MOS
Male Speaker	5.5 dB	1.4
Female Speaker	1.5 dB	1.9

TABLE 4: **SNReval measurements while using the bare hard drive to record audio.**

to filter the dirty, recovered audio in any effective manner. In fact, at the threshold volumes, Shazam is unable to recognize songs from the filtered audio. This is likely a result of the manner in which Shazam matches spectral fingerprints against its database. Since a song's spectrum is much wider than that of conversational speech, linear filtering is again ineffective for removing wide-band white noise. While spectral noise gating proved effective in the case of human speakers, it is not helpful in this case, as most of a song's energy is not concentrated in human voice. Thus, the momentary spectral separation of noise from the signal, as shown in Figure 3 does not arise, and the signal lies too close to the noise floor. This results in discrimination problems, wherein bands containing the signal are misclassified as noise and are subsequently removed; this destroys the spectral fingerprint, and renders Shazam incapable of recognizing the songs.

Thus, we attribute the necessity of high volumes to Shazam's classification algorithms rendering our signal processing useless. However, our results demonstrate that a hard drive can approximate a microphone closely enough to capture very complex waveforms.

6.5. Potential Improvements

In this section we discuss potential situations and algorithms that can aid in the recovery process.

6.5.1. Multiple Hard Drives. We consider the situation wherein the attacker has access to more than a single hard drive in close proximity. This can arise when a single machine, either a desktop or laptop, is using a combination of internal and external hard drives. Other possibilities include a conference room with multiple hard drives, or even a server room.

The presence of N hard drives opens up the possibility of using signal averaging to strengthen the signal in comparison to the noise. With signal averaging, we simply construct the cleaned signal by computing the average over the N corresponding measurements from the hard drives. Intuitively, this works because the average of the common signal is simply the signal itself, while the average of white noise will tend to its mean, by the Law of Large Numbers. This technique yields an improvement in the SNR by a factor of N [34].

One complication to this technique is the possible difference in phase between the wave as it hits the different hard drives. Signal averaging relies on the signal samples being time aligned, which may not be the case. Given the speed of sound at 343 m/s, a 500 Hz wave will have a wavelength of

$$\frac{343 \text{ m/s}}{500 \text{ Hz}} = 0.68 \text{ m}$$

meaning that in the conference room setting, the signals from the hard drives can easily be more than an entire wavelength out of phase. To remedy this, we can again use cross correlation to time-align the samples.

One situation that is likely to benefit from signal averaging is that where the attacker controls a multitude of hard drives within a data center. Given the large number of receivers, a linear improvement in the SNR will yield a very substantial increase in the intelligibility of the extracted speech.

A reasonable concern would be one with regards to the loud volume of background acoustical noise in data centers. Counter intuitively, however, we claim that this noise actually acts only to improve our attack. Acoustic noise within a data center reaches volumes of up to 80 dBa [35], and results in people raising their voices to communicate. This loud noise, however, largely originates from very cyclic processes: namely, fans and electronics hum. As such, we can use linear band-stop filtering techniques to remove this noise. Verbally communicating humans, however, don't typically make use of such noise reduction techniques, and as a result resort to shouting and otherwise raising their voices. This strengthens the signal our side-channel attempts to extract, and thus makes our attack even more practical in the data center setting.

In most multiple hard drive settings, it is unlikely that any form of adaptive noise cancellation can be employed, as the closed nature of a hard drive's spinning platters sufficiently isolates them so as to diminish any correlations in noise. When the noise present in multiple receivers bears no correlation, adaptive noise canceling is rendered useless. It may be possible, however, that hard drives stacked upon one another exhibit correlations in their noise due to mechanical coupling, thus leaving open the possibility of using adaptive noise cancellation. We leave this question to future work.

Moreover, in the presence of multiple hard drives, we can effectively increase the sampling rate. This is because N hard drives will take N times as many samples of the same signal as a single drive. The algorithms discussed in [36] demonstrate how to account for time and gain mismatches to allow us to leverage multiple hard drives that may not be equally spaced or oriented with respect to the signal source. By oversampling at a rate of

$$N \cdot f_n,$$

where f_n is the nyquist frequency, we gain $\log_2(N)$ additional bits of resolution [37]. With a higher sampling rate, we can increase our sampling resolution, and thus further improve the hard drive's sensitivity as a microphone. This could potentially work to allow the hard drive to pick up softer sounds at a greater range.

6.5.2. Repetition. By building off the same principles that take advantage of multiple sensors, we can also leverage multiple repetitions of a given utterance to improve upon our attack. In typical conversations, certain words may be repeated more than once, resulting in a very similar acoustic wave being picked up by a single hard drive. This essentially provides additional samples of the same signal, which can then be averaged to improve the SNR. In order to find repetitions of certain utterances, we can make use of what is known as auto-correlation. This is simply the cross correlation of a signal with itself, and similar utterances will result in the largest peaks within the auto-correlation. The lag associated with these peaks is then the offset at which we are likely to find repeated utterances.

7. Defenses

Eavesdropping through the use of our acoustic side-channel leverages the very same mechanisms that allow high density, high performance hard drives to operate. Thus, defenses must take into consideration the trade off between security and performance, and cannot ignore the strict constraints that a hard drive operates under. In this section we discuss the range of defensive measures one can take to mitigate this attack against privacy; we both examine how retroactive defenses can secure already deployed hard drives and how manufacturers can take steps towards securing future hard drives.

7.1. Ultrasonic Masking

Since our techniques for improving our side-channel rely on increasing the SNR, an obvious approach for mitigating the attack is to work to reduce the SNR. We can accomplish this by either increasing the noise or decreasing the strength of the signal.

To accomplish the former, we propose the idea of ultrasonic acoustic masking. Simply using a white noise generator to mask over the frequencies wherein human speech is contained has the undesirable side effect of being noticeable and annoying to nearby humans. By leveraging the phenomenon known as aliasing, however, we can make use of an acoustic mask that is imperceptible to humans.

Aliasing is a phenomenon where a signal matches the amplitude of a different signal at each point that it is sampled, despite being of a different frequency. This occurs when a sinusoid of frequency f is sampled at a sub-Nyquist frequency f_s, and results in the signal becoming indistinguishable from

$$|f - \ell \cdot f_s|, \ell \in \mathbb{Z}.$$

Then, we can leverage this effect to create a sound mask that lies above the hard drive's sampling rate, such that it aliases and masks over the same frequency bands as human voice. Figure 7 illustrates how the acoustic mask sits just above the sampling rate of the hard drive, yet is aliased so that it ends up in the baseband spectrum, where it masks human voice. In the case of our Seagate Barracuda 7200.12 1 terabyte hard disk, the sampling rate of 34.56 kHz is well above the 20 kHz ultrasonic threshold.

While this mitigation does have the benefit of being undetectable by humans, designers must be careful to find an appropriate threshold for the intensity of the sound mask. The National Counterintelligence and Security Center gives guidelines for acoustic masking and states that the intensity of the mask must exceed the level of the conversations [38]. However, given the treble heavy response of hard drives, the high frequency tones generated in defense have the potential to disturb the normal operations of the hard drive; if the added noise results in the head being completely unable to stay on track, the hard drive is then rendered useless. If the ultrasonic mask lacks sufficient power, however, it will not be effective in mitigating the side-channel. We leave the exploration of this trade off to future research.

7.2. Securing Future Disks

We have thus far discussed only solutions that can be deployed retroactively to protect systems already in use. We now discuss potential mitigations and steps manufacturers can take towards protecting future hard drives during the development cycle.

The primary challenge with this is that any attempts to mitigate the side-channel at the hard drive level must also be mindful of impairing the performance of the hard drive. As such, preventative measures such as reducing the sampling rate or resolution are not practical due to the performance of the hard drive's reliance on exactly these qualities. Thus, mitigations are limited to decreasing the read/write head's susceptibility to acoustic interference.

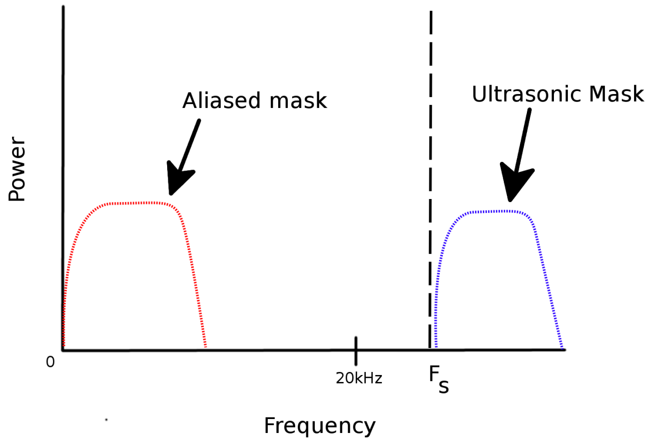


Figure 7: **The ultrasonic mask generates white noise in the region just above the hard drive's sampling rate f_s. Due to the insufficient sampling rate, the mask is aliased over the region just above 0 Hz.**

One approach to doing this is to simply build the hard drive with more effective acoustic dampening built into the sides. As we can see from the hard drive's dampened response to acoustic waves approaching from the side, this approach can substantially attenuate external noise. A potential drawback of this approach is that the acoustic insulation may also act to insulate heat, and thereby impair the hard drive's performance.

A similar defense involves increasing the resonant frequencies of the read/write head assembly such that it will have a weak response to human vocalizations. This can be accomplished by increasing the rigidity of the arm.

8. Related Work

The research that most closely resembles our own is Michalevsky et al.'s work on extracting speech from a smartphone's gyroscope [36]. While both their work and our own leverage acoustically induced perturbations in sensitive mechanical components to eavesdrop on proximal human speech, our work differs in the primary challenges and limitations. Whereas Michalevsky et al.'s primary obstacle was the low (100 Hz) sampling rate of the gyroscope, the sampling rate of our hard drive was sufficient for perceiving most of the audible range. On the other hand, we had to overcome the presence of a large amount of noise present in spinning disks. Another considerable difference exists in the experimental setup; they conducted their experiments by playing audio through speakers sharing a common surface with the phone. In contrast, we mechanically decoupled the hard drive from the speaker so as to isolate the acoustic signal.

One recent work that has explored the topic of turning hard drives into microphones is Ortega's presentation at EkoParty 2017[1]. His work proposes extracting speech by measuring hard drive write latencies from user space. This side-channel differs from our own in matters of timing precision and the requisite audio volume. In order for acoustic waves to have any impact on read/write latencies, the head must be pushed far off track; Sandahl et al. [39] demonstrate in their study that amplitudes upwards of 110 dB are required to begin affecting latencies. Additionally, timing delays introduced by the operating system's multiplexing of the hard drive will add unpredictable skew to latency measurements, whereas the even spacing of servo sectors yields a PES with equally spaced delays between samples.

Roy et al. [40] demonstrated how to jam microphones with inaudible sounds. Their work differs from our own defense in that they leveraged non-linearities specific to microphones, while our ultrasonic mask relies on aliasing due to insufficient sampling rates.

Previous work has demonstrated that subjecting hard drives to extremely loud tones can result in degradation of performance. Nickerson et al. [41] swept hard drives from 1 to 16 kHz and identified which tones resulted in severe reductions in throughput. Dutta's thesis [29] confirmed these results, demonstrated that the most sensitive frequencies corresponded to the resonant frequencies of the HDD's internal components, and used finite element analysis to gain deeper insight into how acoustic waves interact with the head stack assembly. Bolton et al. [42] further examined acoustically induced throughput loss, and designed an attenuator controller for reducing the impact of acoustic interference on HDDs.

Sensor Side Channels. Information leakage through unintended means occurs through mediums known as side channels. Various researchers have demonstrated examples of how sensors can leak information that they were never designed to measure. Marquardt et al. showed how to leverage an accelerometer's readings to recover keystrokes from a nearby keyboard [43]. Biedermann et al. [44] used the magnetometer on a smartphone to deduce the activities of a hard drive, due to the magnetic fields produced by the read/write head. Michalevsky et al. demonstrated how to geolocate a smartphone by measuring its power consumption [45]. Owusu et al. showed how attackers can recover passwords input to a smartphone's touch screen by observing accelerometer readings [46]. Guri et al. utilized speakers' near identical circuitry to approximate microphones for the purpose of eavesdropping [32]. In contrast to these works, our own acoustic side channel extracts information through a device that was never intended to function as an acoustic sensor in the first place.

Hard Drive Security. There exists a substantial body of work investigating the implications of HDD malware. Zaddach et al. demonstrated the ease with which even modestly funded attackers can reverse engineer a HDD's malware and reflash it to implement an incredibly stealthy back door into a system [2]. In a line of research that accomplished the

[1] https://www.youtube.com/watch?time_continue=1&v=ntw32kYDryM

opposite goal of our own, Guri et al. demonstrated how to use the acoustic emanations given off by the movement of a HDD's head to establish a covert channel [47]. Guri et al. again used a hard drive for a covert channel by modulating arbitrary bits over the hard drive's flashing LED [48].

9. Discussion

Although solid state drives are increasingly encroaching on hard drives' portion of the market share, many legacy systems, desktops and laptops alike, still rely upon spinning disks. In fact, even in 2017 worldwide sales for hard drives in PCs nearly doubled that of solid state drives [49] [50]. As such, a large number of computer users are presently at risk from this side-channel attack. Furthermore, as hard drive technology continues to advance, and bit density and rotational frequencies increase, the need for extremely high precision feedback control loops for positioning the read/write head will become even more necessary. The implication is that hard drives will become even more well suited to functioning as microphones, making the need for a solution all the more urgent.

Firmware Security. Our research demonstrates yet another risk that hard drive malware presents. As such, we view our proof of concept acoustic side channel as a call to action for hard drive manufacturers to adopt simple defensive measures that have already been proven effective in other domains (i.e. web and mobile security).

Simply cryptographically signing firmware updates is the single most effective way to prevent the spread of hard drive firmware malware. Though previous research has demonstrated that determined attackers can bypass digital signatures in some cases via side channels, timing attacks, or mathematical weaknesses [51] [52], making use of signatures significantly increases the effort required on the behalf of attackers. Moreover, when distributing updates for hard drives that have no support for verifying digitally signed firmware, manufacturers should adopt TLS to prevent MITM attacks.

While they were unable to accurately measure just how prevalent the use of digital signatures is, Zaddach et al. [2] subjectively found that few do in practice. They also found that hard drives only verify signatures at load time, and are thus still vulnerable to run time injections. Furthermore, should any such vulnerability exist, it would be easily exploitable as none of the drives they examined took measures to mitigate classical binary exploits, such as address space layout randomization (ASLR) or data execution prevention (DEP).

Future Directions. A hard drive is able to function as a microphone due to its ability to detect vibrations at an extremely high resolution. This opens up the possibility for other vibrationally induced side channels. For example, it may be possible to extract keystrokes from nearby keyboards or mice; by leveraging the sensitive data input to the keyboard, an attacker could then create a virus that spreads between laptops that share common surfaces, much the same way that a biological virus spreads.

In a *write* side channel, an attacker may be able to use the hard drive's voice coil motor to drive the read/write head the same way it would drive a speaker's diaphragm. Thus, the hard drive operates as a speaker, and can potentially inject arbitrary voice commands into nearby voice control systems such as Siri, Google Now, Alexa, and others.

Additionally, our work sheds light on a broader area of research that has been historically neglected. While substantial effort has been invested into exploring the limitations of what side channels are available to sensors, few have considered the threat surface that is exposed by devices that were never designed to be sensors in the first place. Beyond hard drives, a printer must also stabilize its head, and thus may offer a similar side channel.

Furthermore, this cyberphysical attack surface is only expected to grow as the Internet of Things intrudes further into our personal lives. As such, we believe this line of research will open up a rich set of research directions to pursue that will only become more relevant with time.

10. Conclusion

Our work demonstrates the threat posed by an overlooked attack vector; that is, the potential for non-sensing devices to infringe upon privacy through a cyberphysical side channel. In particular, we have leveraged a hard drive's capability to act as an unintentional microphone to extract and parse human speech. Despite only having access to a subset of the full position error signal, we were still able to validate the side-channel through use of the Shazam service and by exploring the hard drive's properties, both qualitatively and quantitatively.

We then examined the challenges of defending both systems that are already deployed and those yet to be, and explored the fundamental trade offs between security and performance. As a consequence of our work, we recommend that security and privacy sensitive systems should adopt solid state drives.

Acknowledgments

This research is supported in part by NSF CNS-1330142, NSFC 61472358, and a gift from Analog Devices, Inc. The views and conclusions contained in this paper are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of NSF or ADI. We thank Srdjan Capkun, Simson Garfinkel, the anonymous reviewers, and SPQR lab colleagues for feedback on early drafts; Lauren Biernacki and Mark Gallagher for reproducing the experiments; and CERT/CC for vendor facilitation.

References

[1] "Equation group: Questions and answers," Kaspersky, Tech. Rep., 2015, https://securelist.com/files/2015/02/Equation_group_questions_and_answers.pdf.

[2] J. Zaddach, A. Kurmus, D. Balzarotti, E.-O. Blass, A. Francillon, T. Goodspeed, M. Gupta, and I. Koltsidas, "Implementation and implications of a stealth hard-drive backdoor," in *Proceedings of the 29th Annual Computer Security Applications Conference (ACSAC)*. ACM, 2013.

[3] T. Yamaguchi, Y. Soyama, H. Hosokawa, K. Tsuneta, and H. Hirai, "Improvement of settling response of disk drive head positiong servo using mode switching control with initial value compensation," in *IEEE Transactions on Magnetics*, 1996.

[4] S. Xiong and D. Bogy, "Position error signal generation in hard disk drives based on a field programmable gate array (FPGA)," in *Microsystem Technologies*, vol. 19, 2013.

[5] Objective measures of speech quality/SNR. Last accessed: 2018-05-01. [Online]. Available: http://labrosa.ee.columbia.edu/projects/snreval/

[6] "Open speech repository," last accessed: 2018-05-01. [Online]. Available: http://www.voiptroubleshooter.com/open_speech/american.html

[7] J.-G. Zhu, "New heights for hard disk drives," *Materials Today*, vol. 6, pp. 22–31, July 2003.

[8] "Exceeding capacity, speed and performane expectations," Seagate, Tech. Rep., 2011, https://www.seagate.com/files/staticfiles/docs/pdf/whitepaper/seagate-acutrac-TP624.1-1110US.pdf.

[9] A. Dayes and J. Treder, "Drive Performance-TMR," last accessed: 2018-05-01. [Online]. Available: http://www.logicsmith.com/performance.html

[10] S. Hu and B. Vikramaditya, "Servo control techniques for track following on hard disk drive spin stand testers," in *2014 American Control Conference*, 2014.

[11] "Cell phone privacy modification," last accessed: 2018-05-01. [Online]. Available: http://stahlke.org/dan/phonemute/

[12] B. Möller, T. Duong, and K. Kotowicz, "This POODLE bites: exploiting the SSL 3.0 fallback," 2014.

[13] D. Adrian, K. Bhargavan, Z. Durumeric, P. Gaudry, M. Green, J. A. Halderman, N. Heninger, D. Springall, E. Thomé, L. Valenta *et al.*, "Imperfect forward secrecy: How Diffie-Hellman fails in practice," in *Conference on Computer and Communications Security (CCS)*. ACM, 2015, pp. 5–17.

[14] N. Aviram, S. Schinzel, J. Somorovsky, N. Heninger, M. Dankel, J. Steube, L. Valenta, D. Adrian, J. A. Halderman, V. Dukhovni *et al.*, "DROWN: Breaking TLS Using SSLv2." in *USENIX Security Symposium*, 2016, pp. 689–706.

[15] G. Greenwald, *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State*. Picador, 2014.

[16] D. Goodin, "How "omnipotent" hackers tied to NSA hid for 14 years-and were found at last," 2015, last accessed: 2018-05-17. [Online]. Available: https://arstechnica.com/information-technology/2015/02/how-omnipotent-hackers-tied-to-the-nsa-hid-for-14-years-and-were-found-at-last/

[17] "Maxtor basics personal storage 3200," last accessed: 2018-05-01. [Online]. Available: http://web.archive.org/web/20080411051058/http://www.seagate.com/www/en-us/support/downloads/personal_storage/ps3200-sw

[18] A. Berkman, "Hiding data in hard-drive's service areas," Recover Information Technoloies LTD, Tech. Rep., 2013, https://dl.packetstormsecurity.net/papers/general/SA-cover.pdf.

[19] D.-D. Chang and R. Wong, "Hardware PES calculator," Patent 6 130 798, 2000.

[20] T. Ueda, K. Satoh, H. Ono, and T. Wada, "Disk drive and write control method for a disk drive," Patent 6 111 714, 1998.

[21] A. Dhanda, T. Hirano, T. Semba, and S. Yamamoto, "Magnetic Recording Disk Drive With Position Error Signal (PES) Blocks in The Data Tracks for Compensation of Track Misregistration," Patent 9 412 403, 2016.

[22] J. M. Madapura, "Achieving higher ADC resolution using oversampling," Microchip, Tech. Rep., 2008, http://ww1.microchip.com/downloads/en/AppNotes/Achieving%20Higher%20ADC%20Resolution%20Using%20Oversampling%2001152A.pdf.

[23] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub, 1997.

[24] I. R. Titze and D. W. Martin, *Principles of Voice Production*. Prentice Hall, 1998.

[25] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice*. Cengage Learning, 2000.

[26] "San Ace 92 Datasheet," last accessed: 2018-05-01. [Online]. Available: http://www.farnell.com/datasheets/1878751.pdf?_ga=2.178000229.1377951826.1522458792-2128987597.1522171362

[27] "Harvard sentences," last accessed: 2018-05-01. [Online]. Available: http://www.cs.columbia.edu/~hgs/audio/harvard.html

[28] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[29] T. Dutta, "Performance of hard disk drives in high noise environments," Master's thesis, Michigan Technological University, 2017.

[30] C. V. Wright, L. Ballard, F. Monrose, and G. M. Masson, "Language identification of encrypted VOIP traffic: Alejandra y Roberto or Alice and Bob?" in *USENIX Security Symposium*, 2007, pp. 43–54.

[31] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 229–238, 2008.

[32] M. Guri, Y. Solewicz, A. Daidakulov, and Y. Elovici, "SPEAKE(a)R: Turn speakers to microphones for fun and profit," in *11th USENIX Workshop on Offensive Technologies (WOOT)*, 2017.

[33] A. Wang, "An industrial-strength audio search algorithm," in *Proceedings of the 4th International Conference on Music Information Retrieval*, 2003.

[34] U. Hassan and M. S. Anwar, "Reducing noise by repetition: introduction to signal averaging," *European Journal of Physics*, vol. 31, no. 3, p. 453, 2010, last accessed: 2018-05-01. [Online]. Available: http://stacks.iop.org/0143-0807/31/i=3/a=003

[35] D. Miljković, "Noise within a data center," in *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016 39th International Convention on*. IEEE, 2016, pp. 1145–1150.

[36] Y. Michalevsky, D. Boneh, and G. Nakibly, "Gyrophone: Recognizing speech from gyroscope signals," in *USENIX Security Symposium*, 2014.

[37] *Silicon Labs*, last accessed: 2018-05-01. [Online]. Available: https://www.silabs.com/documents/public/application-notes/an118.pdf

[38] "Technical specifications for construction and management of sensitive compartmented information facilities." [Online]. Available: https://www.dni.gov/files/NCSC/documents/Regulations/Technical-Specifications-SCIF-Construction.pdf

[39] D. Sandahl, A. Elder, and A. Barnard, "The impact of sound on computer hard disk drives and risk mitigation measures," Tyco, Michigan Technical University, Tech. Rep., 2015, https://www.ansul.com/en/us/DocMedia/T-2016367.PDF.

[40] 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.

[41] M. L. Nickerson, K. Green, and N. Pai, "Tonal noise sensitivity in hard drives," in *Proceedings of Meetings on Acoustics 166ASA*, vol. 20, no. 1. ASA, 2013, p. 040006.

[42] C. Bolton, S. Rampazzi, C. Li, A. Kwong, W. Xu, and K. Fu, "Blue Note: How intentional acoustic interference damages availability and integrity in hard disk drives and operating systems," in *Proceedings of the 39th Annual IEEE Symposium on Security and Privacy*, May 2018.

[43] P. Marquardt, A. Verma, H. Carter, and P. Traynor, "(sp)iPhone: decoding vibrations from nearby keyboards using mobile phone accelerometers," in *Conference on Computer and Communications Security (CCS)*. ACM, 2011.

[44] S. Biedermann, S. Katzenbeisser, and J. Szefer, "Hard drive side-channel attacks using smartphone magnetic field sensors," in *Financial Cryptography and Data Security: 19th International Conference*, 2015.

[45] Y. Michalevsky, A. Schulman, G. A. Veerapandian, D. Boneh, and G. Nakibly, "Powerspy: Location tracking using mobile device power analysis." in *USENIX Security Symposium*, 2015.

[46] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang, "Accessory: password inference using accelerometers on smartphones," in *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, 2012.

[47] M. Guri, Y. Solewicz, A. Daidakulov, and Y. Elovici, "Acoustic data exfiltration from speakerless air-gapped computers via covert hard-drive noise ('diskfiltration')," in *22nd European Symposium on Research in Computer Security (ESORICS)*, 2017.

[48] M. Guri, B. Zadov, E. Atias, and Y. Elovici, "LED-it-GO: Leaking (a lot of) Data from Air-Gapped Computers via the (small) Hard Drive LED," in *DIMVA 2017. Detection of Intrusions and Malware, and Vulnerability Assessment - 14th International Conference*, 2017.

[49] "HDD still dominate the storage wars," last accessed: 2018-05-01. [Online]. Available: http://datastorageasean.com/daily-news/hdd-still-dominate-storage-wars

[50] "Shipments of hard and solid state disk (HDD/SSD) drives worldwide from 2015 to 2021 (in millions)," last accessed: 2018-05-01. [Online]. Available: https://www.statista.com/statistics/285474/hdds-and-ssds-in-pcs-global-shipments-2012-2017/

[51] D. Boneh, R. A. DeMillo, and R. J. Lipton, "On the importance of checking cryptographic protocols for faults," in *EUROCRYPT*, 1997.

[52] E. Sidorov, "Breaking the Rabin-Williams digital signature system implementation in the crypto++ library," *IACR Cryptology ePrint Archive*, vol. 2015, p. 368, 2015.