CovertBand: Activity Information Leakage using Music

RAJALAKSHMI NANDAKUMAR[†], ALEX TAKAKUWA[†], TADAYOSHI KOHNO, SHYAMNATH GOLLAKOTA, Paul G. Allen School of Computer Science & Engineering, University of Washington
[†]Co-primary Student Authors

This paper contributes a novel method for low-cost, covert physical sensing and, by doing so, surfaces new privacy threats. We demonstrate how a smartphone and portable speaker playing music with embedded, inaudible signals can track multiple individuals' locations and activities both within a room and through barriers in 2D space. We achieve this by transforming a smartphone into an active sonar system that emits a combination of a sonar pulse and music and listens to the reflections off of humans in the environment. Our implementation, CovertBand, monitors minute changes to these reflections to track multiple people concurrently and to recognize different types of motion, leaking information about where people are in addition to what they may be doing. We evaluated CovertBand by running experiments in five homes in the Seattle area, showing that we can localize both single and multiple individuals through barriers. These tests show CovertBand can track walking subjects with a mean tracking error of 18 cm and subjects moving at a fixed position with an accuracy of 8 cm at up to 6 m in line-of-sight and 3 m through barriers. We test a variety of rhythmic motions such as pumping arms, jumping, and supine pelvic tilts in through-wall scenarios and show that they produce discernibly different spectrograms from walking in the acoustic reflections. In tests with 33 subjects, we also show that even in ideal scenarios, listeners were unlikely to detect a CovertBand attack.

CCS Concepts: • **Security and privacy** → *Privacy compromises*; • **Human-centered computing** → *Smartphones*; *Mobile devices*;

Additional Key Words and Phrases: Active Sonar; Music processing; Information leakage; Motion tracking

ACM Reference format:
Rajalakshmi Nandakumar[†], Alex Takakuwa[†], Tadayoshi Kohno, Shyamnath Gollakota. 2017. CovertBand: Activity Information Leakage using Music. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 1, Article 87 (September 2017), 24 pages. https://doi.org/10.1145/3131897

1 INTRODUCTION

Smart devices and appliances are becoming increasingly prevalent, but as a consequence of adding these connected devices such as smart TVs, phones, and hubs like the Amazon Echo [17] to our homes, there are an increased number of connected speakers and microphones with access to our private environment. This provides a lot of value for consumers, but there are also privacy threats involved with increased connected sensing capabilities. In this paper, we show that in the case of microphones and speakers there are privacy leaks possible with today's devices that go beyond the ability to simply record conversations in the home. For example, what if an attacker could remotely co-opt your television to track you as you move around, without you knowing? Further, what if that attacker could figure out *what* you were doing in addition to where you were? Could she

2474-9567/2017/9-ART87
https://doi.org/10.1145/3131897

even figure out if you were doing something *with* another person? A positive answer could leak information about user activities that are inaudible to a microphone and so far have been considered to be private.

While there has been significant research interest in the use of RF for localization and activity recognition (see §6), no existing RF mechanisms using Wi-Fi hardware on commodity devices — routers, laptops, and smartphones — permit device-free localization of unsuspecting victims in *either* through-barrier or remote-attack scenarios. We create CovertBand, which, for the first time, transforms commodity devices with microphones and speakers into active sonar systems to track users and differentiate between different classes of motion. At a high level, we transmit acoustic pulses in the 18-20 kHz range from the speaker and track reflections from the human body on the microphones. To accomplish our goals, we had to overcome two key challenges:

(1) *How to perform passive localization using acoustic signals.* Due to the nature of indoor environments, there are significant multipath effects from static reflectors. To address this, we borrow Orthogonal Frequency-Division Multiplexing (OFDM), a modulation technique commonly used in wireless communication systems such as Wi-Fi and LTE [36]. OFDM's strong autocorrelative properties allow CovertBand to function in the presence of multi-path, where a signal bounces off multiple objects in the environment before arriving at the receiver. This lets the receiver perform channel correlation to estimate the multi-path effects in the transmitted signal.
(2) *How to perform acoustic localization through barriers.* A naïve solution to this challenge would simply increase the volume of pulses in the 18-20 kHz range until enough sufficient energy pentrates the barrier, reflects off a subject, passes through the wall, reflects off of a subject, and returns to the receiver. However, CovertBand uses speakers on existing devices which are not specifically built to transmit in the 18-20 kHz range at high volume. As a result, they create harmonics in audible frequency ranges. To mitigate this effect, we show how to mix the harmonics with cover music. We also show how to choose OFDM symbols that music can best conceal.

We implemented CovertBand on a Samsung Galaxy S4 with common audio devices, including 4 portable speakers [7–9, 31] and a home theater system [22]. To demonstrate the potential for privacy attacks on varied devices, we implemented CovertBand on a 42 inch SHARP TV [48]. We ran experiments in five homes in the Seattle area to demonstrate CovertBand's ability to help an attacker both localize victims and leak information about activities even in scenarios where those activities are not audible.

We summarize our experimental results below:

- CovertBand can track multiple subjects independently through barriers in a 2D plane. We ran experiments in five homes to track both a single subject and multiple subjects and found that we could localize with tracking error comparable to the state-of-the-art in RF localization [13, 34]. Specifically, CovertBand localized walking subjects with a mean tracking error of 18 cm and subjects moving in a fixed position with a mean tracking error of 8 cm. For comparison, WiTrack2 [13], which uses custom FMCW radar hardware, has an accuracy of 10.9–19.2 cm when tracking moving subjects through walls.
- We evaluated CovertBand's range through a variety of materials using a portable speaker [31], showing that it can track at up to 6 m without barriers and 3 m in through-wall scenarios.
- CovertBand can differentiate between rhythmic and linear motions. We tested a variety of rhythmic motions — pumping arms, jumping, and supine pelvic tilts — in through-wall scenarios and show that they produced discernibly different spectrograms from walking.
- We compared performance in the 18-20 kHz range across multiple speakers to prove CovertBand could work on a wide variety of hardware. We also demonstrated CovertBand on an LG G4 connected to a Sharp TV [48] without fine tuning our algorithms to demonstrate the possibility of sensing with diverse sets of hardware without device-specific training.

- We evaluated our ability to conceal CovertBand with music by playing unmodified songs and songs with an additional sonar signal back-to-back in random order for 33 subjects in an isolated environment. We found that subjects could correctly differentiate only 58% of the pairs, which is close to random guessing, showing that even in ideal scenarios victims are unlikely to to detect the attack.

Contributions. To the best of our knowledge, we are the first to demonstrate active sonar for through-barrier sensing on a wide range of commodity devices available to standard consumers. Specifically:

- We demonstrate the first device-free localization capability on commodity devices in both through-barrier and remote-attack scenarios. We show how to perform localization, tracking, and motion classification for multiple subjects in a 2D plane using changes in the audio channel.
- We ran experiments in five real homes to show that attacks are possible with our prototype. In particular, we show through multiple scenarios that an attacker can use active sonar to glean information about victims through walls, even when the attacker cannot see the victim nor hear any movements, and that such an attack is feasible using many common, off-the-shelf devices.
- We show how to conceal the attack from a victim by mixing active sonar pulses with music. We ran user studies to evaluate our methods, showing that such an attack could be done covertly to avoid detection even in ideal scenarios.
- We reflect on the broader implications of this work, including privacy implications and future research in this space. Specifically, we demonstrate the feasibility of potential threats through three case-studies, including spying on: 1) multiple people in a dormitory room when they are engaging in private activities, 2) a person's private activities in a bathroom (even when these activities are inaudible from outside), 3) remote victims by adding CovertBand-based malware to gadgets, like phones and TVs, that are commonly present in homes.

We note that our work intends to show the possibility of information leakage with commodity speakers and microphones. Maximizing the range and resolution for different materials and configurations, or building applications to utilize this capability is beyond the scope of this paper.

2 MOTIVATION AND GOALS

We begin by considering several motivating scenarios. From these scenarios, we derive our key goals.

2.1 Scenarios

These scenarios survey the utility of understanding the feasibility of covert, through-barrier sensing. Such attacks provide a new avenue for leaking information about obscured activities even in the presence of background or cover noise.

National intelligence. Imagine a spy (Alice) entering a foreign country. She rents a hotel room adjacent to an individual (Bob), whom she intends to discretely and covertly monitor. To be a good spy, Alice cannot enter the country with dedicated surveillance hardware, and she cannot acquire any suspicious new hardware while in-country. But she still wants to monitor her neighbor to know when he is in the bedroom or bathroom, an ideal opportunity to enter the apartment and gather additional information). Bob does not know that he is being monitored, or even that he is the potential target of monitoring. Alice would benefit from using a covert monitoring mechanism, something she could run on her phone and that would avoid arousing Bob's suspicion.

Vigilante Justice. In some cases, revealing certain private activities can be dangerous to victims. For example, many countries or non-government entities persecute pre-mairtal or other sexual partnerships [25, 28, 51]. We note that in many of these cases, vigilantes do not seek conclusive evidence before condemning victims; as such, the possibility of even circumstantial evidence could pose security threats for these individuals.

Remote Hacking of Phones and Smart TVs. We also consider attacks that leverage devices already inside a victim's home. Because our attack requires access only to a speaker and microphone, an attacker can leverage many devices that already exist in the home environment. Smart TV apps and voice assistants, like Amazon Echo [17] and Google Home [21], already have access to speakers and microphones and let users install applications. A remote adversary who compromises one of these devices, perhaps via a Trojan application in an app store or via a remote exploit, could use our methods to remotely glean information about an individual's home activities. An attacker could also find more surreptitious ways to execute such an attack. For example, a streaming music app with voice control has all the permissions (speaker and microphone) needed to execute our attack. As a simple example, an attacker could utilize the advertising library embedded inside a music application to determine whether the user is near the phone when an ad is played.

2.2 Goals and Non-goals

Inspired by the preceding scenarios, we enumerate five goals for our system:

(1) *Major motion detection and tracking.* The system should be able to detect motion and perform 2D tracking for each of several individuals in an environment.

(2) *Distinguish between movements.* Our technique should be able to convey information about the type of movement occurring.

(3) *Re-purpose commonly available devices.* Our technique should be implementable on devices that people might already own for several reasons. First is cost, ensuring that our approach is affordable enough to be commonly available. Second, using devices commonly found in household environments increases the potential attack surface. Further, using such devices provides "plausible deniability": a camera installed in an environment leaves physical evidence, and the purchase of dedicated radar equipment leaves little ambiguity as to someone's intended monitoring activities.

(4) *Through-barrier sensing.* Each of the previous capabilities should be possible despite the presence of common barriers, e.g., walls, doors, windows, etc.

(5) *Not detectable to unknowing target.* An unknowing target, unaware of our this type of attack, should have a low probability of detecting our attack.

We do not attempt to avoid detection by *knowing* targets. If targets know that they might be the subjects of monitoring, then they might be able to detect it by setting up sensors. Because this is true of any existing, active through-wall imaging techniques, including radar, we argue that this is a reasonable non-goal.

To our knowledge, we are the first to study, demonstrate, and evaluate an attack using the preceding goals. More specifically, we are unaware of any solution that performs through-barrier detection and tracking on common devices. Note that RF-based systems that perform through-barrier device-free detection and tracking [12, 14, 15, 41–43, 62, 63, 69] require custom hardware, such as USRPs, or are limited to Wi-Fi chipsets with access to channel state information (CSI) in the vicinity of the device [52, 60, 64]. Commodity smartphones do not typically provide software-level access to CSI information. Thus, we know of no existing solution that tracks through barriers with commodity smartphone devices.

3 COVERTBAND DESIGN

Section Outline. This section is organized as follows: §3.1 describes the attack surfaces and attacker requirements. §3.2 describes our choice of signal, and §3.3 describes how we play this signal and our method for obscuring it. Sections §3.4 and §3.5 explain how to use the signal to calculate the distance to moving objects and how to use that distance to perform 2D localization.

3.1 Adversary Model

CovertBand enables two unique attack surfaces.

(1) The first is a remote attack where an attacker compromises speakers and microphones already in a victim's home. This may be as innocuous as a music or video application with access to the microphone and speakers on a smartphone, Amazon Echo, or a Smart TV. Because CovertBand uses common devices, it can use an over-permissioned or malicious application, a common and well-known problem with mobile applications [44], to monitor an individual's location and activity. In principle, the hardware need only have multiple microphones and a single speaker, a configuration common in smarthphones and home assistants (e.g., Amazon Echo has a 7-microphone array). The attacker can likely reference hardware specifications to get information about speaker locations and microphones but must make some assumptions about device location. An Amazon Echo and television, for example, are unlikely to move. Thus, if attackers learn their location (for example, in a bedroom), they could use that information to breach privacy. State-of-the-art RF approaches generally do not permit such remote attacks because through-wall approaches require specialized hardware (USRPs and FMCW radar arrays) not present in a victim's home [13, 43]. Other approaches use common Wi-Fi access points but require multiple access points and multiple devices in the environment at known locations and under attacker control. Further, they require a training phase that may demand victim cooperation [35, 45, 47, 65].

(2) The second attack is a through-wall scenario where the attacker places a speaker and microphones near a barrier to sense obscured activities. Though this attack works best when victims are in the forward direction relative to the speaker, it: (1) does not require any particular speaker and microphone placement, (2) can be executed with a diverse combination of speakers and microphones, and (3) can be placed anywhere along (or up) the barrier as long as the victim is within range. Existing sonar approaches for through-barrier sensing require more specialized hardware and setups. For example, the DoD funded a through-wall sonar detector [12] with specialized hardware, which was meant for presence detection, not localization. Finally, while the principles in our paper can generalize to two (stereo) or more speakers and a single microphone or a synthetic aperture by moving the speakers in a line, demonstrating these generalizations are beyond the scope of this paper.

3.2 Strong Autocorrelative Signal

CovertBand leverages autocorrelation to identify the beginning of an echo from a human. For this reason, we selected OFDM signals for our sonar pulses. OFDM, a modulation technique commonly used in wireless communication systems including Wi-Fi [26] and LTE [36], has strong autocorrelative properties. These properties let it work in the presence of multi-path reflections, where a signal bounces multiple objects in the environment before arriving at the receiver. The receiver can thereby perform channel correlation to estimate the multi-path effects in the transmitted signal.

3.3 Signal Generation at the Speaker

CovertBand generates OFDM symbols in the upper audible range (18-20 kHz), which we play through a connected portable speaker. Since our phones' microphones accept samples at 48 kHz, by the Nyquist condition, the effective bandwidth is 24 kHz. We divide this bandwidth into 64 subcarriers, each with a width of 375 Hz. We assign random data (either 1 or -1) to the seven subcarriers between 48 and 54 that correspond to the frequencies of 18 to 20 kHz and perform an IFFT on these 64 values to generate a 200-sample OFDM symbol in the time domain. When sampling at 48 kHz, a 200 sample OFDM symbol forms a pulse that spans 4.2 ms. Though radios have oscillators that let them transmit quadrature and in-phase components, and hence transmit the complex

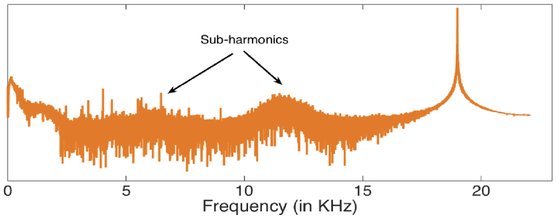


Fig. 1. The figure plots the frequency spectrum of the signal recorded in a smartphone when the speaker plays a 19 kHz tone. While playing it creates sub-harmonics in lower frequencies. Recorded in a quiet lab environment by placing a speaker directly in front of a smartphone microphone.

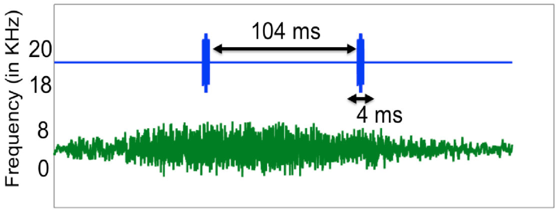


Fig. 2. The figure shows the sonar OFDM signals (18-20 kHz) mixed with the cover music (sub-10 kHz). Generated in software to show the separation in frequency space between the OFDM symbols and the cover music.

numbers output by the IFFT, speakers only accept 16-bit real numbers, so we transmit the amplitude of the IFFT samples and discard the phase.

Although not all adults can hear frequencies at 18-20 kHz, when played through an off-the-shelf portable speaker at high volume there are audible sub-harmonics at lower frequencies. Fig. 1 shows the spectrogram of a 19 kHz tone played from a JBL portable Bluetooth speaker and recorded using a smartphone. The plot shows sub-harmonics in the 11 kHz range. To hide these sub-harmonics, we combine the OFDM symbols with an audible song (see Fig. 2). Specifically, we play music continuously in the 0.1-8 kHz range and transmit the OFDM symbol every 105 ms, i.e., about 9.6 OFDM symbols every second. Since the song and OFDM symbols use different frequency ranges, we can isolate the OFDM symbol at the receiver with a high pass filter, removing both our song and any environmental or cover noise below the 18-20 kHz band.

We note two key points about our design. First, we generated a number of random OFDM symbols and found that each symbol has different audibility level when played on a speaker. When we compared the structure of the OFDM symbol with its audibility level, we found that OFDM symbols that do not ramp up to full volume near the beginning or end of the signal are much less audible than symbols with a sudden change in amplitude (i.e., a crescendo and decrescendo, rather than immediate spikes at the beginning or end of the symbol). We use one such symbol in our design, though we did not excessively optimize OFDM symbols for this purpose.

Second, the songs selected effected the detectability of our sonar signal. We found that songs with more percussive events easily obscured the sonar signal but songs and speeches with many silent pauses were unable to mask certain elements of the signal. In our design, attackers can modify the ratio of song volume to sonar signal volume to better hide the OFDM signal. We found that song volumes higher than a quarter of the sonar signal volume were sufficient to our ears. That is, songs played at much lower volumes than the signal proved sufficient to mask it in our tests. See§4.5 for the evaluation of covertness.

3.4 Computing Distance from Microphone

When the speaker plays our sonar signal, sound waves reflect off both static objects in the environment and moving persons before being reaching the microphone. To find the distance of the people from the microphone, CovertBand performs to steps. It: (1) estimates the channel correlation for each transmitted OFDM symbol to find all reflectors, and (2) compares consecutive correlation profiles in time to seek moving reflectors, which we assume to be humans (but could be other moving objects).

Step 1: Generating the channel correlation profile. To generate the correlation profile, we pass the recorded signal through a high-pass filter to remove the song and isolate sonar signal reflections. Fig. 3 shows a sample correlation profile at a specific time instance. Each peak corresponds to an object's echo. We can find the distance to the

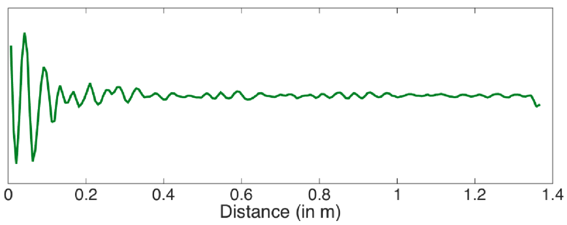


Fig. 3. The figure plots an example correlation profile for a time instance recorded through a door in one of our five home experiments. The peaks represent all the major reflectors including the static objects and the human subject present at the corresponding distances.

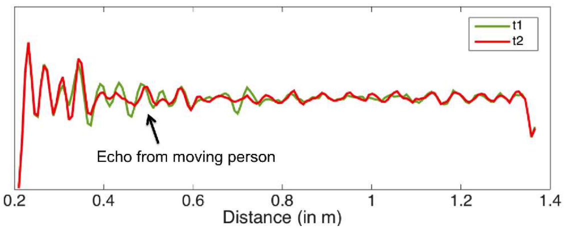


Fig. 4. The figure plots an example correlation profile at two time instances t1 and t2 also recorded in a real home environment. The peaks shift farther to a distance of 0.45 m when the person moves away from the smartphone.

object using the speed of sound and the observed time delay. Since OFDM has strong autocorrelation properties, the correlation profile is accurate within a range of 2 to 3 samples (1-2 cm sampling at 48 kHz).

Step 2: Identifying the echo from the person. When a human moves, the resulting echo occurs at different distances over time. Fig. 4 shows the correlation profile at two time instances separated by 0.02 s when the subject moves from 0.45 m to 0.5 m. We clearly see the change in the echo's position. We extract this change by performing a consecutive subtraction of the channel correlation profiles every 110 ms. We should note that the consecutive subtraction operation removes the constant echo from all the static objects in the environment. During human motion, we see a significant change in echoes mainly at the distance corresponding to the individual's location. This change occurs in a range of distances corresponding to the human profile (height and weight). However, we also see some minor changes at slightly large distances due to the dynamic multi-path i.e. a multi-path that changes when the human moves. For example an echo from the human might reach a nearby strong static reflector first before reaching the receiver. In most cases, the dynamic multi-path causes small changes at larger distance than the direct path and hence we identify the smallest distance at which the difference is above a relative threshold to identify the new location of the human. CovertBand uses 60% of the maximum change as the minimum threshold. The dynamic multipath causes larger changes in the echo only when there is a strong reflector very close to the human subject. Since the reflector is basically in contact with the human subject, this leads only to a small error of less than 10 cm.

To perform the correlation, enough energy must pass through the barrier, regardless of path, and reflect back to the microphone. We find this to be the case in our real-world experiments. Our intuition is that, in normal environments, sufficient gaps and holes (light switches, power outlets, door frames, windows, etc.) let sound propagate and return to the microphone for sensing.

Multipath. In cases where the wall is not permeable to our sonar signal, the large correlation will correspond to an indirect path that over-estimates distance. This would cause error in our experiments. However, our experiments in realistic environments with various configurations show that some energy does pass in a sufficiently direct path to moving reflectors. Although the direct path may not be the highest energy reflection, we can see that it moves in the same way as do the higher energy reflections, indicating that it is a shorter path to the same moving object. We use the lowest-distance reflection to get an upper bound on the individual's distance.

In principle, we can infer the actual position of an individual even without *any* energy traveling along the direct path. One potential technique would build on [67], which uses angle of arrival to locate the source of reflections even if the direct path is fully blocked. We did not need to do this for our proof of concept information leakage demonstration because sufficient energy passed directly perform sensing.

Multiple people. When multiple people move in an environment, consecutive subtraction of correlation profiles will show changes at distances corresponding to each human. We record on two microphones so multiple people at the same distance from one microphone are likely to show up as distinct reflectors on the other microphone, unless they are, in fact, in the same position. To identify each individual, we first scan through the entire correlation profile and attribute each distance where the difference in the correlation profile exceeds the minimum threshold to a single person. In our implementation, we attribute motions that occur within a distance of 20 cm to the same person. We use the distance that corresponds to the maximum difference in the correlation profile in each group to compute the individual's location. Note that multi-path reflections corresponding to a single subject move at a fixed rate and distance. Using this, we can disambiguate between reflections from different subjects. In some theoretical situations where two subjects are in precise locations and move in concert at the same time, we will see only one change in the correlation profile and thus recognize one person even when there are two separate individuals. This highly unlikely situation is beyondthe scope of a commodity system such as CovertBand.

3.5 Tracking with Multiple Microphones

We can track an individual's location using the distance from multiple microphones. Note that the distance measured in the previous step sums of the distance from the speaker to the individual and the distance from the individual back to the microphone. In a two dimensional space, given this distance, the human can be at any point in the 2D plane along an ellipse with the locations of the speaker and microphone as the foci and the measured distance being twice the length of the major axis. Thus, each microphone creates an ellipse; the intersection of the ellipses from multiple microphones provides the individual's location in the plane. In the case where one cannot assume the subject is moving in a 2D plane, we would need a third microphone (for example, one plugged into the audio jack) or a different device to do trilateration in 3D. For example, the Amazon Echo has a 7 speaker array [27] that would help immensely for 3D localization.

Since the phone has two microphones, the individual's position can be at any location that occurs at the intersection of the two corresponding ellipses. While two ellipses can intersect at four different points, in our case both ellipses share a common focus (i.e., location of the speaker). Hence, they intersect at only two points that are symmetrical along the line joining the two microphones on the smartphone. These two points lie on either side of the barrier and can therefore be used to disambiguate the motion on either side.

4 EXPERIMENTAL EVALUATION

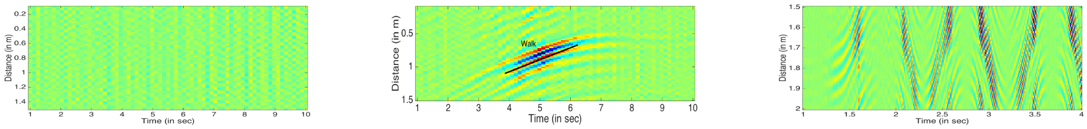
Here we evaluate CovertBand's ability to achieve our goals from §2. Because sonar has not been used with these goals in mind, — namely for covert, through-barrier sensing on commodity devices — we aim to show the constraints of our design choices from §3 and demonstrate feasibility in scenarios that represent realistic threats.

We had five main goals from §2:

(1) Identify different classes of motion and activities.
(2) Track multiple people in a 2-D environment.
(3) Evaluate range through different materials.
(4) Ensure that our methods work with existing, cheap, off the shelf hardware.
(5) Conceal the attack from the victims.

In this section, we outline our experiments, implementation details and present results. We performed experiments for each of our goals to show feasibility in real world environments.

Implementation Details. We implement our design of CovertBand as a third party Android app that does not require rooting the phone. The app uses the *AudioTrack* API to play the acoustic signals and the *AudioRecord* API to record simultaneously on both microphones in stereo. Our design requires the following from the phone:



(a) No action

(b) Walking

(c) Repeated pelvic tilts

Fig. 5. Spectrograms of the change in the correlation profile at all distances (increasing from the top in the y-axis of the figure) over time (x-axis) for three different cases a) When the person is stationary, there is no change. b) For a linear motion like walking, the maximum change occurs at the new location of the user. c) For a rhythmic activity like pelvic-tilt exercise, the changes repeat over time.

(1) transmit acoustic signals at 18-20 kHz, (2) sample the received signals on the microphones at 48 KHz, and (3) have two microphones to achieve 2D tracking (recall that a smartphone with a single microphone can be used to estimate the distance but not 2D position). Many Android phones including Samsung Galaxy S4, Samsung Galaxy S5 and HTC One satisfy the above requirements.

For each of our experiments, we used a Samsung Galaxy S4 connected to a portable speaker [31] through the audio jack. Like most common smartphones, the S4 has two microphones, one at the top and one at the bottom, separated by approximately 15 cm. Our Android app transmits a song along with our sonar signal through the speaker and records the backscattered signals using the two microphones as described in §3. We record and process the raw sample data and send the results to a laptop over Bluetooth for offline processing. We note that this attack can potentially be done on the phone, rather than offline, if needed.

Microphone Orientation. For each of our experiments, we placed the phone on its side, so that the two microphones (one at the top of the phone and one at the bottom) were in the same horizontal plane and perpendicular to the direction of the subject. For distinguishing between activities, any orientation should work. However, for some of the experiments discussed later, such as 2D tracking, the only relevant property of this orientation is that the two microphones are in the same plane as the target, as we can only make inferences in this plane. With more microphones we could potentially sense in 3D.

4.1 Distinguishing Between Activities

First we demonstrate CovertBand's ability to help an attacker infer information about *what* a person is doing using two basic methods: (1) inference based on characteristics of motion and (2) inference based on timing.

Inference based on characteristics of motion. We show how CovertBand can potentially enable an attacker to differentiate between different classes of movements even when subjects are in different body positions and orientations. Specifically, we focus on two classes of motion: (1) linear motion (the subject walks in a straight line) and (2) periodic motion (pelvic tilt where the subject remains in approximately the same position (lying on his or her back on the floor) but performs a periodic exercise). These motions are sufficiently different that we should be able to differentiate them by looking at the spectrograms, but are also realistic enough to potentially enable privacy leakage. For example, (1) models information that might be of interest to intelligence community members, e.g., to track the location of a target within a room and (2) could be used to infer sexual activity, for which the importance of protecting might vary depending on the target's culture and cultural norms or might vary depending on the target's public visibility, e.g., celebrity status or political status.

To run these experiments, we placed our phone and speaker 20 cm from a standard interior wall [30]. A subject 1 m from the inside of the wall was asked to perform each of the above activities. We then transmit our covert sonar signal and track the changes in the echoes as a function of time as described in §3. Fig. 5 shows the spectrogram plots for the different activities. The x-axis denotes the time and the y-axis is the distance of the subject from a single microphone at the smartphone. The spectrogram plots the difference in the echoes received where the differences are computed over successive 10 ms durations. The plots show that:

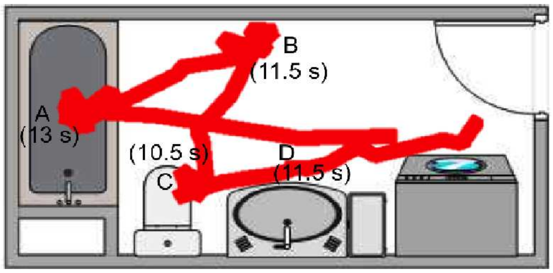


Fig. 6. We were able to localize the person in different areas of the bathroom. According to our sonar readings, Bob spent 13s at station A, 11.5 s at station B, 10.5 s at station C and 11.5 s at station D. (Ground truth: 19.5 s at A, 13.4 s at B, 12 s at C and 14 s at D). Trajectory line thickened for visibility.

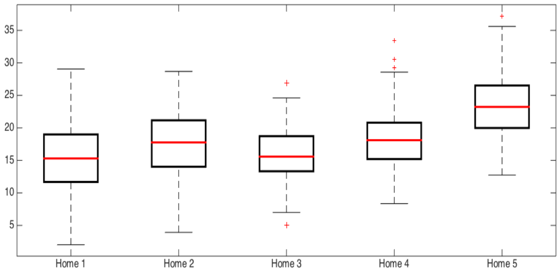


Fig. 7. Experimental Setup 1: Box and whiskers plot of tracking error for one moving subject in the bathroom. We calculate error by comparing observations with trajectories based on starting and ending points marked on the floor.

- First, when there is no activity on the other side of the wall, as expected, we do not see any significant changes in the echoes as received by the microphone.
- As the subject walks towards the phone on the other side of the wall, we see a strong change in consecutive echoes occurring at decreasing distances from the phone. The black line in Fig. 5(b) shows the actual distance of the subject as a function of time. By looking at the areas where the changes in the echoes are the highest, we can see that CovertBand accurately tracks the distance.
- When the subject performs a repetitive motion from a stationary position, we see a repetitive signal at a fixed distance (1.5 m).

We also tested other rhythmic motions, such as jumping and pumping arms with the subject in a standing position. The plots look similar to the pelvic tilt in that they are clearly repetitive, but have different energies and distances associated with them. Though this is clearly an example of a rudimentary classification, it requires no training phase to generate the data above that enables an additional privacy breach. More sophisticated attackers could potentially train models to do more accurate classification or detect additional types of movements. And though we do not aim to do gesture recognition in this work, even recognizing a motion as repetitive may be sufficient as a privacy threat in situations where circumstantial evidence can be damning. We also note that CovertBand should still be able to differentiate between linear and rhythmic motion in cases where subjects come into contact with stationary objects in the environment. We verified in §4.2 that CovertBand could detect the motion and localize a person sitting in an office chair, performing rhythmic motions.

Inference based on timing. As we will see in §4.2, we can use CovertBand to do 2D tracking of subjects even through walls, which can further leak information about potentially private activities. To demonstrate this, we show a scenario where one subject (Bob) pretended to go through a routine in the bathroom while the other (Alice) used CovertBand to track his movements. We placed the speaker setup 15 cm outside the bathroom door and performed four trials during which Bob spent less than 20 seconds doing each of the following: showering, drying off on the scale, sitting on the toilet, and brushing his teeth. *During the experiment, the bathroom fan was ON and we could not hear Bob performing any of the activities inside the bathroom.*

Using CovertBand, we were able to localize Bob at different areas of the bathroom as he performed different tasks. Fig. 6 shows a 2D mapping of Bob's movements with observed timing as he stopped in each location. We were also able to download the publicly available floorplan for this particular apartment, which allowed us to map each of the stops to different stations within the bathroom. For example, we can guess that Bob was probably using the sink during D, or showering during A. Floorplans for many apartments and hotels are available online.

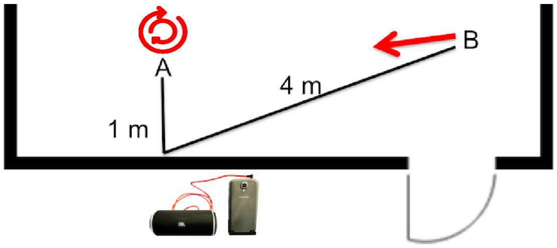


Fig. 8. Experimental scenario 2: Multiple people, multiple motions. An example of one of the bedroom layouts. Subject A twisted at the hips, while subject B walked toward him. For all experiments, doors/windows were closed.

We note that the notably higher error for timing while Bob was at station A is due to Alice's inability to differentiate between Bob standing in the shower, and Bob opening/closing the curtain or simulating drying off directly outside the shower. If we had simulated Bob showering for a more realistic amount of time, the relative error would be much more reasonable.

During this experiment, the participant did not spend a realistic amount of time at each station, rather simply paused and used a timer to record how long he stayed at each stop. We can see that being able to localize Bob within 20cm can tell us a lot about which part of the room he is in and with what he could be interacting.

4.2 2D Tracking

As we mentioned above, CovertBand can track 2D movements using echoes from multiple microphones on a phone (recall §3.4). Here we strive to demonstrate and quantify our 2D location tracking ability by performing experiments in real-world setups both from within the home and from outside of it.

Home Environment: In our first setup, we ran two sets of experiments in each of five real-world homes in a metropolitan area. For each, we asked volunteers of different height and weight to perform various actions. We confirmed in separate tests that larger subjects (largest was 6'3", 180 lbs) reflected more energy, than smaller ones (smallest was 5'3", 130 lbs). However, though they reflected more energy, this did not affect our ability to localize them or change the observed error. In some cases, larger people were easier to detect, but caused more observed error as they reflected energy from a wider space.

Experimental scenario 1 - Single subject. For the first experiment, we ran three trials at each location where we placed the phone and speaker on a chair **outside a closed wooden bathroom door** and asked a volunteer to walk along a straight 1 m line marked on the bathroom floor. The thicknesses of the wooden doors were standard [29], but some were hollow and some were solid wood. Also, in a couple of homes the fan inside the bathroom was on, and in all the homes we could not visually see or hear the subject performing the activity. For each trial, we compared the 2D trajectory computed by our system with the marked trajectory on the inside of the bathroom. Fig. 7 shows the mean 2-D tracking error across the bathrooms in the five homes. While we see a variation in the errors across the homes, owing to differences in the bathroom door material and the natural variation in movements across different subjects and trials, across the five home environments the mean tracking error for the bathroom experiments was just 18 cm.

In a similar scenario, we were also able to localize a subject in close contact with a strong multi-path reflector. To demonstrate this, we placed a large metal sheet next to a subject, who performed rhythmic motion. As mentioned in §3.4, the close proximity of a strong multi-path reflector causes dynamic multi-path. However, we were still able to correctly localize the moving subject within an average error of 13 cm.

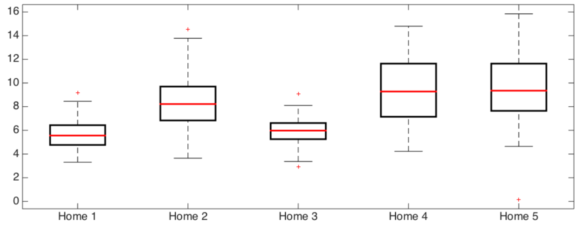


Fig. 9. Experimental Scenario 2: Box and whiskers plot of tracking error for the twisting subject from Fig. 8 (Subject A). The error is smaller because he is stationary (only moves his torso).

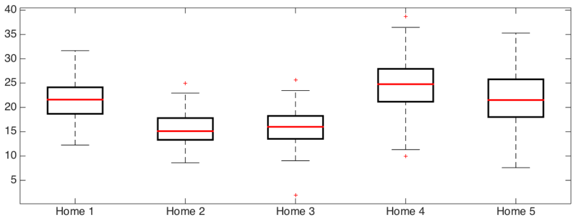


Fig. 10. Experimental scenario 2: Box and whiskers plot of tracking error for the walking subject from in Fig. 8 (Subject B). We compute error based on the trajectory marked on the floor.

Experimental scenario 2 - Multiple subjects, multiple motions. We placed the speaker and phone **outside a bedroom** to detect and localize subjects inside. In three of five homes, we placed the setup outside the bedroom wall; in the the other two, we placed it outside a closed bedroom door. All walls and doors were standard and all doors were closed for the experiments. To show that we can simultaneously track two people, we asked one person to stand 2 m from the wall closest to the speaker and continuously move his torso in a rhythmic fashion. At the same time, a second person walked 2 m towards the first person. For these experiments, distances and orientations were a bit different in each setup depending on the layout of the room. The experimental setup for one of these layouts is depicted in Fig. 8. For subject A, we calculated the tracking error as the distance between the computed 2D location and the true location. For subject B, we computed the tracking error as the difference between the computed trajectory and the direct line between the starting and ending points.

In the bedroom experiments, the mean tracking error was 8 cm for the subject twisting in a fixed location, but 20 cm for the walking subject. The discrepancy in tracking error between the subjects is in part because we measured the accuracy against a reference line which was directly between the starting point and the ending point. In reality, it is likely that the walking subject deviated slightly from this direct line while moving 2 m toward the twisting subject. For example, even for a subject walking directly on the trajectory line, normal walking motion includes arms, separated by more than 20 cm, swinging with pendulum-like motion. As such, we do not try to optimize our results for errors less than 50 cm. Despite this, all calculated errors were still less than 25cm (less than 1 foot) and in each trial we were able to both localize each movement and identify it as either the linear movement or the rhythmic one. Though the experiments for Fig. 7 were done in the same respective homes as those for Fig. 9 and 10, they were done in different rooms with different layouts, so we expect some deviation in the results. However, the mean tracking error for the bathroom experiments are similar to errors for the walking subject in the bedroom setup, owing to the similarity in movement, the similarity of materials within homes and similar variability in subject trajectory.

Experimental Scenario 3: Multiple people, convergence. To demonstrate this behavior, we performed two more experiments to confirm that CovertBand could deal with two subjects converging to a single location. In these experiments, both subjects performed a whole body motion by walking towards and away from each other inside a room with the speaker setup placed outside a standard interior wall. Specifically, for the first experiment, the two subjects were present at 1 m and 1.7 m from the wall and walked towards each other; for the second experiment, both the subjects started from the same point 1.5 m from the wall and walked in diagonally opposite directions. Fig. 12 shows the CDF of the 2-D localization errors of the two subjects. From the results, we found that we were able to localize both the subjects with an average error of less than 20 cm. When the subjects get sufficiently close, i.e., less than 20 cm, our thresholds begin to treat them as a single person. As they walk away, the threshold separates them into two distinct people again.

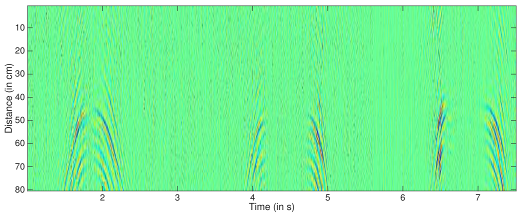


Fig. 11. Spectrogram shows the seated person performing the rythmic motion when the other subject stands in contact with the chair and the subject.

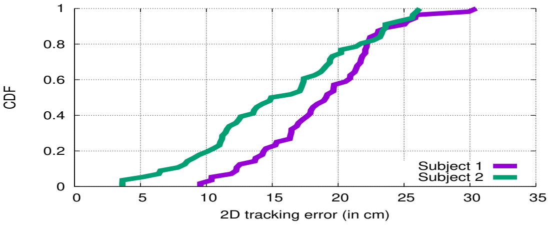


Fig. 12. CDF of localization errors for two subjects when they are walking towards each other as in experimental scenario 3.

A Note About Multiple People. As described in *Experimental Scenario 3*, when multiple people are far apart, CovertBand can distinguish between them, but when they come within close contact, it treats them as a single person. One person moving rhythmically and one static will be treated as a single person moving rhythmically. Furthermore, if both are moving linearly together, they will simply be treated as a single walking person, but they will likely reflect more energy. Because CovertBand is only accurate to 18 cm, we leave it out of the scope of this work to distinguish situations where there is only one person in a location vs. when there are two. An attacker using CovertBand would have to deduce this using prior knowledge. For example, she could use CovertBand to observe that there were previously two people moving toward each other. We note that an attacker can make inferences about activities in a bedroom — for example, where two people converge to a location, spend time performing rhythmic activity in the same location, and then separate some time later — without obtaining concrete evidence. As mentioned in §2.1, this could be a significant potential security threat for certain individuals. However, as we discuss in §7, CovertBand will not allow an attacker to identify either subject once they have separated.

Two people moving rhythmically in the same position will simply look like rhythmic motion (albeit with potentially different frequencies) on a spectrogram, similar to the pumping arms scenario, where the subject's arms were not moving in unison. We also note that a static person between the attacker and a moving subject will act like a stationary object in between a victim and attacker, serving to attenuate the sonar signal to some degree. To show that CovertBand can perform this type of tracking when subjects are in contact with stationary objects in the environment, we ran a similar experiment where the stationary person sat in an office chair and performed rhythmic motion while the walking subject walked 1 m to that position and stood next to the seated subject (who was performing rhythmic motion) both in contact with the chair and the person. CovertBand correctly recognized that rhythmic motion and was able to localize both the seated and walking subjects. When they came within 20 cm of each other, it treated them as a single person continuing rhythmic motion. Our system was able to localize this subject with an average error of 9.57 cm. Fig. 11 shows the spectrogram for the seated person performing rhythmic motion. We can see that despite contact with a static object and another human, we are still able to see the rhythmic motion of the subject.

Overall, we find that resolution and accuracy were not drastically affected in any of our real-world experiments despite very different circumstances, building structures and layouts. This implies that in common scenarios, CovertBand can be resilient to changes in location and, unlike some alternative approaches, does not require a training phase before performing through wall detection and tracking.

Spying through external walls and doors: In addition to detecting and localizing subjects across rooms inside a home environment, we also tested CovertBand through external doors and walls. These barriers are thicker and are often designed so that any outsider cannot visually see or hear anything that happens within the home.

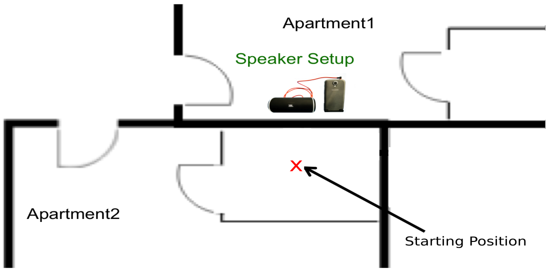


Fig. 13. The figure shows the speaker layout placed in apartment1 close to the wall shared with apartment2. The subject walked in a line toward the speaker on the other side of the drywall.

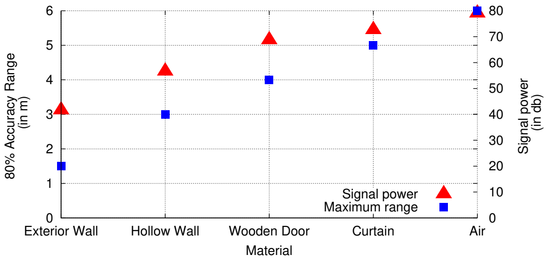


Fig. 14. **Effect of barrier material.** The figure plots the measured signal power and the maximum range at which we can accurately locate an individual through five different barrier materials.

We conducted experiments in three different locations — one apartment and two standalone homes — to test our ability to do 2D localization through external materials. Specifically, we tested a 4.5 cm thick solid wood door, a 30 cm thick exterior foundational basement wall composed of a combination of different wood paneling and sheet-rock, and an external double-paned window with curtains drawn. In each experiment, we placed the speaker close to the barrier and ensured that all windows and doors were shut. In the apartment, this meant placing the speaker in a neighboring apartment along a shared wall.

In all experiments, the speaker was placed 5 cm from the barrier on the house's exterior and a test subject walked a distance of 1 m inside the home. We then localize this subject over time using their walking motion and compute the average of shortest perpendicular distance to the original trajectory as the localization error.

We found that we were able to localize the subject within an error of about 30 cm in the case of the external door an window. As an example, Fig. 13 shows the speaker setup and the measured trajectory for the experiment that we conducted in the apartment. For that experiment, we placed the speaker setup along the shared wall in the neighboring apartment an asked a subject to walk toward the wall on the other side. We were able to localize this subject with an average error of 30 cm. However, we were not able to track the subject through the foundational basement wall due to much thicker external materials. However, an attacker could potentially increase the range and penetrate thicker materials by using a better speaker with higher output power.

4.3 Evaluating Range through Materials

To estimate upper bounds on CovertBand's ability to sense in different environments, we ran experiments through various barriers in a two-bedroom apartment and measured the maximum ranges at which CovertBand could detect movement. In particular, we placed the speaker with no barrier as a baseline and 30 cm from the following visual obstructions: a hollow wooden door, a hollow interior wall, and a hollow exterior wall. In each case, we performed the experiments in the context of the home so as to study feasibility in realistic situations. The interior walls were all standard hollow walls [30] and the doors were made of hollow wood were standard interior doors [29]. The external walls were also drywall and had an estimated thickness of 10 cm.

For each situation, we placed the Samsung Galaxy S4 connected to a JBL portable Bluetooth speaker around 30 cm from the barrier. We also placed an Amprobe sound pressure meter [5] 20 cm away on the other side of the barrier to measure the attenuation. We asked a volunteer to take two steps (around 60 cm) on the other side of the barrier at various distances so that we can calculate the maximum distance at which CovertBand can successfully detect the motion. The volunteer repeats the motion five times at each distance until we cannot successfully detect it with an error of less than 30 cm, 80% of the time, i.e., four of the five times. This is sufficient in our opinion to constitute a privacy breach.

Fig. 14 shows the pressure values reported by the sound pressure meter and the maximum distance at which we can detect the movement correctly in 80% of our trials. In order to count as correctly detected, we require that CovertBand track and localize the movement with an error of less than 30 cm from the ground truth trajectory, marked on the floor with tape. Logically, as the attenuation caused by the barrier increases, the distance at which we can detect the motion decreases. The maximum distances are around 6, 5, 4, 3 and 1.5 meters for air, curtain, wooden door, hollow wall and exterior wall, respectively. In addition to using speakers capable of higher volumes, we mention a few ways to increase this range across all materials in §7. Though these results are only for a single apartment, our experiments in five other homes in later sections showed similar results. We note that our system can detect movement with a lower probability (e.g., one out of five) at a further distance.

4.4 Using Common Existing Hardware

For our detection to work, the speaker needs to transmit signals at the higher end of the audible frequency range. As such, rather than re-running all of our experiments in homes with many different speakers, we find it equally effective to compare the relative volumes at which a variety of speakers can play frequencies in this range.

We compared 4 cheap portable speakers, including the JBL [31] we used in the above experiments, and a home theater system. All were able to play tones at comparable volumes at frequencies from 200 Hz-20 kHz with the exception of a Bose Bluetooth speaker, which had a noticeable dip at 17.5 kHz. We also noticed more speaker "clipping", across all portable speakers at full volume when we played some of the higher frequencies. Despite this, as we demonstrate in other sections, the power is still sufficient for the purposes of our attack and the "clipping" noises don't harm our ability to perform tracking. The clipping is a result of the speaker not being able to drive the speaker cone with the expected amplitude sufficiently fast, so even at 90% volume the clipping disappears. Better speakers capable of higher volumes would not necessarily have this problem. The home theater system we tested, for example, did not exhibit clipping at any of the volumes we measured.

We believe that all five of the speakers we tested are fully capable of executing our attack with ranges from 2-6 m, with the caveat that the signal may need to be altered to use the Bose speaker. The only other components we need are a pair (or more) of microphones that can sample at 48 kHz or higher. Many common smartphones, such as the Galaxy S4/5, LG G4, and HTC One, have microphones that satisfy this requirement.

Demonstrating information leakage with smart TVs. As we mentioned, the ability to use common hardware for this type of attack opens up a variety of potential devices for an attacker to use to get information about a target; an attacker with access to a speaker and microphone that already exist in the environment can potentially leverage them to glean information about remote targets. We show that such an attack is possible if an attacker can get an over-permissioned application on a smart TV. To demonstrate, we use CovertBand to play some of our altered songs through the television speakers of a 42 inch SHARP TV [48]. Though all of our previous tests were done on a Samsung Galaxy S4 with software tuned for the orientation of the speakers on that device, an attacker may not know the exact orientation and power of speakers on a victim's device, or the relative position with respect to the speaker they now control. To account for this, we used the dual microphones from a larger LG G4 phone (with microphones in different positions) and placed it on a table next to the TV. As we mean to demonstrate feasibility, we conservatively did not change any of the software to tune it for the new hardware. We had a volunteer stand two meters in front of the television and perform repetitive motions for a short interval in a stationary position. We found that using the standard TV front-facing speakers we could correctly estimate distance of a subject within a maximum error of up to 30 cm.

4.5 Evaluating Covertness of our Design

Our final goal was to conceal the signal, and its subharmonics, against unknowing attackers. We designed an experiment to get an estimate of an upper bound on detectability by testing whether subjects who were both

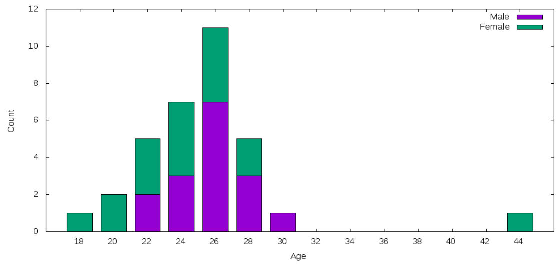


Fig. 15. Histogram of ages and genders for test subjects from experiments in §4.5.

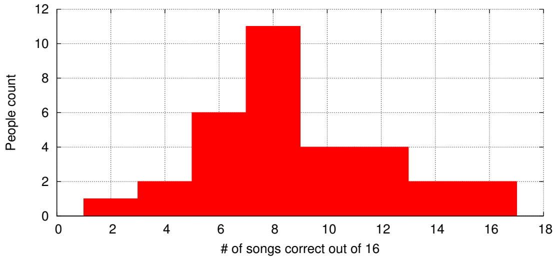


Fig. 16. Distribution of detection scores. Random guessing would on average result in 8 correct songs.

familiar with the experiment and knowledgeable of the signal could differentiate between unaltered songs and songs containing our sonar signal. Our logic is that subjects who have been exposed to the signal, who are not obscured by barriers, and who have been told that the signal will be present should detect our attack at higher rates than unknowing victims.

Experimental Setup. Participants were asked to identify which of two clips (played back-to-back in a random order) was the original unmodified song and which was the clip with our added sonar signal. The volume of the added sonar signal was half the volume of the song (6 dB lower). Our subjects (17 female, 16 male) spanned ages between 18 and 45 and were composed of students and staff from different entities in our organizations as well as other local individuals. None of the subjects were monetarily compensated. All the experiments reported in this were given an exemption by our organization's IRB. Fig. 15 shows a histogram of ages and genders.

We used sixteen different songs chosen from a list of popular songs and selected ones with sufficient amounts of percussion. We did not use any sophisticated methods for choosing songs or try to maximize our ability to hide the signal within. The songs are listed in the appendix. Each participant was seated in front of a Beats Bluetooth speaker [8] with no barrier blocking the speaker. They sat within a 30 degree angle of the speaker's face and were allowed to move around as close or far as needed. To give our victims as much power as possible, we spent the first few minutes of each experiment training their ears by playing our signal un-obscured. We also allowed them to ask for a replay of any clip at any time. Every subject we tested was able to identify the sonar signal when played without music cover. They reported hearing something similar to static.

Results. We expect our subjects to be able to guess the correct clip with slightly better than 50% accuracy. Intuitively, this is because with no information, we would expect them to succeed at 50%. However, we have given our subjects a number of advantages and have not tailored our signal to hide specifically within songs. Further, some subjects are likely to be able to hear frequencies at which we transmit [4], driving up the expected percentage of correct answers.

Our results fall mostly in line with our hypothesis. Overall, our 33 subjects guessed the correct clip 58% of the time, for a mean of 9.3 out of 16, a median of 9, and a standard deviation of 3.077.

We also asked each of our subjects to record any pairs for which they were very confident they had correctly identified the signal. Of the 528 total trials, this resulted in 71 claims of high confidence, of which 60 (85%) were chosen correctly. These high confidence claims tended to be of two varieties: 1) Three subjects claimed to be able to hear the sonar signal clearly in almost every song. Of those, two chose the correct clip in all sixteen trials. One (Female, 21) was able to hear the high frequency tone in one of two ears. The other (Female, 22) heard both the clipping and the high frequency sound with ease. The third (Male, 27) guessed incorrectly twice before figuring out that he could feel the clipping sound if he leaned in directly in front of the speaker. 2) There was one song for which almost everyone was able to identify the sonar signal due to many short silent periods in the song. A third of subjects marked this song with high confidence and all of those subjects correctly identified the modified version. Many explicitly mentioned this song after the conclusion of the experiment. If we remove this song from the statistics, even with the three subjects above included, the detection rate drops to 56%. If we additionally remove the three subjects who were able to consistently detect readings, we show a detection rate of under 53%, less than 3% above random guessing ($p = 0.12$).

Most of our subjects admitted that they guessed randomly on every trial in the experiment despite the lack of barriers, close proximity, and prior training.

Key Observations. When played at full volume (as we did in our experiments), whether or not we obscured with music, there was an audible clipping sound due to the sonar signal. However, reducing the volume by just ten percent removes the clipping, so we would expect even lower detection rates at more modest volumes.

Through-Barrier Follow-Up. To get an upper bound on detectability in through-barrier scenarios, we re-tested two of the three subjects who claimed to hear the sonar signal in every song in a through-the-wall scenario. We placed them inside one of the rooms we used for experiments in §4.2 and placed the speaker [31] along the same wall we used before. Using the exact same songs — which the subjects were able to detect from the unobstructed trials — we asked them if they could identify the sonar signal. We only collected subjective feedback on their experience due to a very small sample size and limited subject availability. Though both were able to hear the unobscured signal through the wall, when mixed with music they admitted to having low confidence on over two thirds of pairs. Subjects had the greatest success when lying on the floor to listen through the space at the bottom of the door, indicating they were able to identify the signal much better when unobstructed.

Improvements. We believe that a sophisticated attacker could hide the signal during percussive events in the music on a per-song basis, making it incredibly difficult for most people to detect. We also randomly generated our OFDM symbols. An attacker who really wanted to make it undetectable could generate a tone specifically meant to avoid clipping noises (some symbols were better than others) and avoid playing the signal at 100% volume. Future work could inclue specifically analyzing the ratios of music to OFDM symbol and tailor the volumes to match. However, even simple improvements, such as removing the symbols during silent portions of songs would reduce detectability based on our responses. We also note that very young children and animals often hear higher frequencies more clearly, and thus may be able to detect our signal. We are currently placing these detection mechanisms out of scope.

In summary, for most people, detection was very difficult despite all the advantages we gave our subjects. We hypothesize that for adult targets who are not aware of the presence of the signal, detection would be exceedingly unlikely even without a wall or other barrier. Even subjects who can hear the added signal may just think there is something strange with the speaker or music, like static from the radio.

5 DEFENSES

In this section we discuss some defenses against CovertBand attacks.

(1) Victims could prevent some versions of this attack by soundproofing their homes. This may be infeasible for many people, especially those who want windows that can open to the outside. However, one could soundproof more private areas of the home and remove speakers and microphones to prevent CovertBand attacks.

(2) Another defense that does not require structural changes involves jamming signals in the victim's inaudible range. However, jamming in the 18-20 kHz range may be audible to pets or children and requires playing inaudible sounds whenever a victim notices music near private areas. Furthermore, an attacker could allay suspicion by simulating natural sounds, like birds chirping or leaves ruffling, to do sparse sensing without explicit cover traffic (music). It does not always seem feasible to play sounds like this, though playing random noise across all potential frequencies in the inaudible range in private areas would thwart the attack. We note that although

background noise such as loud music can thwart eavesdroppers, it will not thwart CovertBand, which filters out signal in the audible range.

(3) One could set up a dedicated sensor, like a Raspberry pi with an attached microphone, that listens for transmissions at frequencies that exceed a victim's hearing threshold. A potential defense could combine this with jamming, sensing when there is a potential external sonar signal and responding by jamming with a known pseudorandom sequence in a comparable frequency band.

(4) Finally, a simple smartphone application could be built to fool or jam a CovertBand attack. Upon detecting high frequencies, the app could match the frequency range and signal power and jam with random noise in that band. We verified the effectiveness of such a defense against CovertBand attacks by conducting an experiment in the bedroom of one of our test homes. With the speaker setup 20 cm outside the wooden door of the bedroom, a subject walked 1 m towards the door from the inside. At the same time, we set up another smartphone as a jammer inside the bedroom 2 m from the door and played a random OFDM signal in the same frequency range of 18-20 khz from the smartphone speaker. We repeated this experiment with the jammer set to play at five different volume levels 6, 8, 10, 12 and 15 (in android phones, the volume level ranges from 0 to 15). For this particular layout, with the jamming volume set to 8, i.e. 50 % of output power, the tracking error increased to 64 cm. At volumes higher than that, our system could not detect the motion of the subject. We note a possible extension to this defense: given that CovertBand repeats OFDM symbols, the defending phone app could even spoof locations and activities by transmitting altered signals at comparable amplitudes during expected intervals. In this way, a phone left in a particular location or carried by a potential victim could be set up to do detection and jam/spoof locations.

We note that although these defenses are effective to varying degrees and have different drawbacks, they rely on victims to understand that they may be under attack and take actions to mitigate the harm.

6 RELATED WORK

Acoustic Systems. Acoustic transmissions have been used extensively to localize devices in systems such as Cricket [49]. [6, 23, 53, 70] demonstrated the feasibility of localizing and determining the direction of device movement using acoustic transmissions. Unlike prior work, we demonstrate the feasibility of tracking users without instrumenting them with any devices.

Prior work leverages Doppler shifts from inaudible acoustic transmissions to perform gesture recognition [11, 19]. These designs enable recognition of a pre-defined set of gestures in close vicinity to the mobile devices using the resulting Doppler shifts. [57, 66] achieve finger-level localization close to a phone; these techniques do not use active sonar but instead localize the sounds of fingers tapping on a surface. FingerIO [39] uses active sonar to track finger motion around a wearable device, but it is not designed to go through barriers. ApneaApp [38] detects breathing movements using acoustic transmissions on the phone with a range of one meter but is not designed to operate through barriers. Medical imaging uses high-frequency (> 1 MHz) ultrasound signals to perform imaging inside the body [37]. These systems require the acoustic transducer to be placed directly on the skin's surface. Sonar imaging has also been used extensively in underwater settings [1, 24, 50].

Many sonar-based approaches address through-wall detection to aid Law Enforcement. However, these approaches currently require non-standard equipment placed directly against a wall [2, 12, 62]. As such, they would be expensive, forego plausible deniability, and could not be leveraged by remote attackers.

A rich body of work focuses on mapping environments using acoustics, for example [16]. However, this work does not pertain to ours in execution or motivation. In general, it uses first order echoes to reconstruct environments and is not designed to work through barriers or track movement.

RF-based Designs. Prior work has proposed radio-based solutions for human motion detection [14, 15, 43], localization [13, 14, 43], and gesture recognition [33, 40, 41]. While promising, none of the RF-based designs have been demonstrated to work on off-the-shelf smartphones. Specifically, these designs use expensive, ultra-wideband transceiver and/or specialized hardware that are not available on mobile devices. Further, they typically require multiple antennas separated by half a wavelength, which is difficult on smartphones due to their size constraints. In particular, [13, 14, 43] use radar hardware that transmits, receives and processes 500 MHz to 3 GHz wideband signals and requires multiple antennas. Researchers have recently proposed ultra-wideband radar designs that operate in the millimeter wave [72] and terahertz bands of the electromagnetic spectrum [20, 54] where the wavelengths are significantly smaller, permitting multiple antennas to be packed together. Further, it remains to be seen whether the power and processing required for such wideband signals could be achieved on smartphone-grade consumer devices.

Recent work also leverages Wi-Fi for human motion detection and gesture recognition. [15] does human motion detection (walking forward and backward) in through-wall scenarios using 20 MHz wide transmissions but requires specialized interference cancellation hardware that is not available on commodity devices. [41] extracts Doppler shifts on RF transmissions to perform gesture recognition in through-wall scenarios. [10] uses specialized, full-duplex hardware to track finger strokes using RF signals. [61] extracts the minute changes that occur on a loudspeaker when playing a song using Doppler effects from wireless signals. However, these require custom hardware processing (e.g., software radios) and do not work with commodity devices.

Wi-Fi gestures [40] can enable recognition of a pre-defined set of gestures in the vicinity of an Intel Wi-Fi chipset. WiDraw [52] tracks arm motion using transmissions from 20-30 other Wi-Fi devices. These systems work only when the user is close to the Wi-Fi chipset and have not been demonstrated in through-barrier scenarios. Work on tomography imaging [63, 71] tracks motion by deploying 10-30 sensors spread throughout the environment that measure the attenuation between every pair of sensors. WiDir [64] estimates the direction of motion using Wi-Fi CSI values; however, CSI information is not available on commodity smartphones at the software level. In contrast, to the best of our knowledge, ours is the first work to demonstrate user motion detection or 2D tracking through walls and barriers using just a smartphone and a Bluetooth speaker, opening up a new attack vector.

Finally, thermal imaging cameras have been designed to interface with smartphones and detect the heat radiated by humans using an infrared sensor array [18, 46]. These cameras can detect changes in heat radiation patterns and can hence see in the dark or detect pipes within walls. However, they cannot be used to see through walls or even glass surfaces [55].

Summary of limitations and comparisons to prior work. While there has been extensive work on device tracking [6, 23, 34, 49, 53, 70], we focus our comparisons on device-free tracking, which does not require the victim to carry a device. Here, we compare our work to previous work in the device-free localization and tracking space in both sonar and RF domains. Table 1 lists those works — grouped by approach — and their capabilities and quantitative results. To be representative of the general approach, we have used the strongest listed results from each group of citations that are comparable to our work.

Of the compared systems, only CovertBand and the sonar gesture papers [38, 39, 59] can be implemented on commodity hardware. The gesture-based papers, however, do not focus on full-body detection. The Wi-Fi solutions exploit existing Wi-Fi infrastructure in a space to do 2D localization. However, they require control over multiple access points and network devices connected to the localization software in order to function. For example, to achieve 70 cm accuracy, [56] uses four access points and seven laptops to do localization in the home, in effect creating an 11 antenna array. Further, because they work by looking at disturbances in multiple wireless streams, they need those devices to get good coverage of the space, and need to know the locations of

	CovertBand	FMCW Hardware [3, 13, 14, 43]	Software Radio [15, 32]	Wi-Fi [35, 45] [47, 56, 65]	Gesture Sonar [38, 39, 59]	Coupled Acoustic [12]
Can be done on COS hardware	●			🟣	●	
Detects moving subjects	●	●	●	●	🟣	●
Detects stationary subjects	🟣	●	🟣	●	●	🟣
Works in non-LOS	●	●	●	●		●
Differentiates different types of motion	●	●	●		●	
Enables remote attacks	●			🟣	🟣	
2D localization	●	●	●	●	🟣	
Range in ideal circumstances	6 m	20 m	4.9 m	20 m	0.3 m	13.7 m
2D Accuracy	18 cm	10.9 cm	80 cm	70 cm	0.8 cm	-

Table 1. Comparing CovertBand constraints and capabilities against similar approaches in RF and Sonar. Cells with ● are considered to meet the criteria. Cells with 🟣 satisfy the criteria with constraints. Blank cells do not satisfy the criteria.

each device. They also require training or a background collection phase, which would have to be redone should any device move.

All compared approaches permit detection of moving subjects (though we have marked the gesture sonar papers as “satisfy with constraints" because they were not designed for human level tracking). However, CovertBand, software radio-based approaches [15, 32], and the DoD solution [12] can detect stationary people only when there is sufficient movement, i.e., arm movement or twisting motions. The Wi-Fi ecosystem approach [35, 45, 47, 65] and the FMCW approaches [13, 43] can localize static people using either breathing motion or by monitoring changes in an already mapped environment.

Of the attacks possible on commodity hardware, we mark the Wi-Fi approaches as “satisfy with constraints" because they require a detailed understanding of the WiFi AP and device placements in the environment and would require compromising many devices. The very small range of the sonar-based gesture papers have earned it the same score, though microphones could be utilized solely for eavesdropping purposes.

Most of these localization approaches are fairly particular about the placement of sensing equipment. As mentioned, the Wi-Fi papers require a detailed understanding of device placements and retraining if devices move. The DoD approach [12] requires placing the acoustic coupler directly against the wall. CovertBand, the FMCW papers, and the software radio papers benefit from directionality, though they do not necessarily need subjects to be in any particular direction: for RF papers, range will be best when the subject is in the main lobe of the antenna. Similarly, CovertBand gets best results when subjects are in the forward direction of the speaker owing to the directional nature of most commodity speakers.

The main constraint of our work is range. Because RF propagates well through materials, it permits a longer range in through-barrier scenarios. We discuss some methods to increase range in §7. The custom FMCW builds [43] can increase transmit power to penetrate thick materials and measure up to 20 m. Similarly, because Wi-Fi approaches use existing Wi-Fi infrastructure, they can in principle work as long as multiple Wi-Fi streams intersect a location. This depends greatly on the wireless network’s layout. Alternatively, because the coupled acoustics [12] are placed directly against the wall, they eliminate one of the biggest reflectors (known as “the flash problem"), turning the barrier into a speaker of sorts. This enables for much farther propagation than is achievable with audio on commodity hardware.

As noted, RF approaches without significant antenna arrays are far less accurate due to the vastly larger RF wavelengths. CovertBand benefits from shorter acoustic wavelengths and speed of sound, permitting accuracies similar to those of large antenna arrays in the FMCW approaches [13, 14, 43]. The DoD solution does only presence detection, not localization, so its accuracy is not listed. In effect, the Wi-Fi approach also has a large array. As mentioned above, To achieve 70 cm accuracy [56] effectively uses 11 antennas. The figures listed for the gesture sonar papers come from FingerIO [39], which was designed for tracking finger movement very close to the phone. As such, range and accuracy are very small. The best comparison to CovertBand is probably WiTrack2 [13], which operates at slightly better ranges around 10 m and accuracies around 10.9 cm. We calculated the accuracy using median x and y errors for the first detected subject (likely an underestimate). Again, however, this approach will not enable remote attacks as it cannot be done on commodity equipment.

7 DISCUSSION AND CONCLUSION

With a proof-of-concept prototype that uses active sonar pulses in the 18-10 kHz range played on commonly available devices that already exist in many homes, we show that an attacker can glean information about *what* a person is doing even when that attacker can neither hear the person nor see his movements. This section outlines our system's limitations as well as future research opportunities.

Achieving a Higher Accuracy and Resolution. One could incorporate phase-based algorithms [39, 70] to achieve a higher resolution than that demonstrated in this paper. One could also use multiple phones or move the phone along a straight line (potentially in conjunction with the accelerometer) to improve localization and gesture detection. Moving the phone creates a virtual microphone array by taking measurements at different points in space. One could then use tradition angle of arrival algorithms to gain both resolution and accuracy even in the absence of a direct path, enabling the sensing of more subtle motions such as the movement of hands, arms, or even fingers.

Achieving a Higher Range. We evaluate CovertBand's range in a variety of materials, showing that it can track at up to 6 m without barriers and 3 m in through-wall scenarios. While this in itself is a privacy leakage, further research is required to achieve better range. For example, we currently place the microphone next to the speaker to make it easier to administer experiments. This limits our range because the close proximity of the speaker means we can play only a limited volume before we exceed the microphone's capabilities. However, an attacker could use any method to supply audio. Therefore, it is possible to simply position the microphone far from the speaker, for example, by connecting to it over Bluetooth, allowing the use of significantly higher volumes to increase range. Additionally, an attacker could use longer OFDM symbols and perform correlation-based techniques to decode the minute changes due to human motion at larger distances. As discussed in §4.5, a sophisticated attacker could improve CovertBand by using various clever methods to hide the sonar signal. Some of the more subtle methods including tailoring each signal to match the song in cadence and distortive sound. More sophisticated methods may even emulate natural sounds or natural frequencies — simulating a car driving, truck moving, or a jackhammer — allowing for the use of lower frequency sonar pulses which would permit much higher volumes and better sound propagation on existing commodity speakers.

Tracking More than Two Subjects. Our current implementation is tested to track up to two concurrent subjects. In principle, as long as the subjects are at different distances from the microphones, CovertBand can distinguish more than two subjects. One can further generalize the techniques described in this paper to leverage more than two microphones, potentially on multiple devices, to achieve higher angular resolution. If the microphones are placed on different sides of the subject, this could also help solve the near-far problem, where one of the subjects is much closer to the microphones and can have much stronger reflections.

Expanding the Set of Activities Classified. Our paper explores a limited set of activities like pumping arms, jumping, and supine pelvic tilts, which in certain contexts can expose private activities. However, one could explore the use of active sonar to achieve imaging of the environment and a much richer set of activity recognition. At a high level, if we can combine the acoustic reflections from multiple devices spread across a home, one can start creating images of the environment with a higher resolution.

Identifying Individual Subjects. Our implementation can distinguish between users in different locations, but it cannot identify them. Therefore, it cannot continue to track a particular person if two people move to

the same location and then separate. Similarly, CovertBand can not currently differentiate between movement caused by different objects, such as a dog, a fan, or a human. However, there may be differences in the types of motion caused by each. For example, prior work has demonstrated that gait information can be used to identify human subjects. In principle, one could extract gait information from the acoustic reflections and achieve subject identification. Similar work has been done in the RF space [58, 68]. One could imagine that these techniques could generalize to classify certain movements as non-human or as generated by a particular object, like a fan.
Generalizing to More Devices CovertBand could also adapt to instances where an attacker has access to only a single microphone by using multiple speakers. A system with stereo speakers and a single microphone would still create two ellipses which share a focus at the microphone. For example, a smart television with a single microphone will almost assuredly have stereo sound; this could be used to track motion in a 2-D plane.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their insights. This work was funded in part by a Google Faculty Award, Sloan fellowship, the Short-Dooley Professorship, and the National Science Foundation under awards CNS-1420654 and CNS-1565252. We would also like to thank Anna Kornfeld-Simpson, Elliot Saba, Lucy Simko, Paul Vines and Sandy Kaplan for their comments and insight.

APPENDIX

Songs Used (See §4.5 for details): American Woman - Lenny Kravitz, Bad - Michael Jackson, Barbara Streisand - Duck Sauce, Baby Don’t Cry - 2pac, Five Hours - Deorro, What Goes Around Comes Around - Justin Timberlake, Guerilla Radio - Rage Against The Machine, Ice Ice Baby - Vanilla Ice, O.P.P. - Naughty By Nature, Revolutionary Warfare - Nas, Rockefeller Skank - Fatboy Slim, Save Me San Francisco - Train, Hey, Soul Sister - Train, Spoils - Protest the Hero, Uptown Funk - Bruno Mars, What Would You Do? - City High

Speakers Used (See §4.4): JBL [31], Beats [8], Auvio [7], Bose [9], Bose Acoustimass Home Theater [22]

REFERENCES

[1] 2016. VideoRay ROV Sonar, VideoRay Imaging Sonar. (2016). http://www.videoray.com/homepage/options/sonar/blueview-imaging-sonars.html.

[2] Acoustic Localization 2017. Acoustic Localization through wall/ceiling/floor. (2017). https://redecomposition.wordpress.com/acousticvideo/.

[3] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. Capturing the human figure through a wall. *ACM Trans. Graph.* 34 (2015), 219:1–219:13.

[4] HO Ahmed, JH Dennis, O Badran, M Ismail, SG Ballal, A Ashoor, and D Jerwood. 2001. High-frequency (10–18 kHz) hearing thresholds: reliability, and effects of age and occupational noise exposure. *Occupational Medicine* 51, 4 (2001), 245–258.

[5] Amprobe 2016. Amprobe sound power meter. (2016). http://www.amprobe.com/amprobe/usen/environmental-test/sound/amp-sm-10.htm?PID=73334.

[6] Md Tanvir Islam Aumi, Sidhant Gupta, Mayank Goel, Eric Larson, and Shwetak Patel. 2013. DopLink: Using the Doppler Effect for Multi-device Interaction. In *UbiComp*.

[7] Auvio 2016. Auvio Portable Bluetooth Speaker. (2016). http://www.amazon.com/Portable-Bluetooth-Speaker-PBT1000-Gadgets/dp/B00D2D46S4.

[8] Beats 2016. Beats Pill. (2016). http://www.beatsbydre.com/beatspill.html.

[9] Bose 2016. Bose Sound Link 2. (2016). https://www.bose.com/products/speakers/wireless_speakers/soundlink_mini_ii.html.

[10] Bo Chen, Vivek Yenamandra, and Kannan Srinivasan. 2015. Tracking Keystrokes Using Wireless Signals. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*.

[11] Ke-Yu Chen, Daniel Ashbrook, Mayank Goel, Sung-Hyuck Lee, and Shwetak Patel. 2014. AirLink: Sharing Files Between Multiple Devices Using In-air Gestures. In *UbiComp*.

[12] Ivan Cowie. 2008. *Through-Wall Surveillance for Locating Individuals Within Buildings.* Technical Report. Time Domain Corporation and United States of America. Final Scientific and Technical Report.

[13] Fadel Ddib, Zachary Kabelac, and Dina Katabi. 2015. Multi-Person Localization via RF Body Reflections. In *NSDI*.

[14] Fadel Ddib, Zach Kabelac, Dina Katabi, and Robert C Miller. 2014. 3D Tracking via Body Radio Reflections. In *NSDI*.

[15] Fadel Ddib and Dina Katabi. 2013. Seeing Through Walls Using WiFi!. In *SIGCOMM*.

[16] Ivan DokmaniĂĞa, Reza Parhizkara, Andreas Walthera, Yue M. Lub, and Martin Vetterlia. 2013. Acoustic echoes reveal room shape. In *Proceedings of the National Academy of Sciences*, Vol. 110. 12186–12191.

[17] Echo 2016. Amazon Echo. (2016). http://www.amazon.com/Amazon-SK705DI-Echo/dp/B00X4WHP5E.

[18] FLIR 2016. FLIR One Thermal Imaging. (2016). http://www.flir.com/flirone/explore.cfm.

[19] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. SoundWave: Using the Doppler Effect to Sense Gestures. In *CHI*.

[20] Ruonan Han, Yaming Zhang, Youngwan Kim, Dae Yeon Kim, H. Shichijo, E. Afshari, and O. Kenneth. 2012. 280GHz and 860GHz image sensors using Schottky-barrier diodes in 0.13 um digital CMOS. In *ISSCC*.

[21] Home 2017. Google Home. (2017). https://madeby.google.com/home/.

[22] Home Theater 2016. Costco: Bose Acoustimass 10 Series V Home Theater Onkyo Bundle. (2016). http://www.costco.com/Bose%C2%AE-Acoustimass-10-Series-V-Home-Theater-Onkyo-Bundle.product.100147047.html.

[23] Wenchao Huang, Yan Xiong, Xiang-Yang Li, Hao Lin, XuFei Mao, Panlong Yang, and Yunhao Liu. 2014. Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones. In *INFOCOM*.

[24] Humming 2016. Humming Bird Side Imaging. (2016). http://www.humminbird.com/Category/Technology/Side-Imaging/.

[25] Hunted 2016. Hunted: The War Against Gays in Russia. (2016). http://www.hbo.com/documentaries/hunted-the-war-against-gays-in-russia.

[26] IEEE. [n. d.]. IEEE 802.11g-2003: Further Higher Data Rate Extension in the 2.4 GHz Band. In *Standard, 2003*.

[27] ifixit 2017. Amazon Echo Teardown. (2017). https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953.

[28] Incest 2016. Man killed on suspicion of incest. http://timesofindia.indiatimes.com/city/ahmedabad/Man-killed-on-suspicion-of-incest/articleshow/35271666.cms. (2016).

[29] Interior Door Dimensions 2017. Standard Inside Door Sizes. (2017). http://homeguides.sfgate.com/standard-inside-door-sizes-84805.html.

[30] Interior Wall Dimensions 2017. Building Requirements for Partition Walls. (2017). http://homeguides.sfgate.com/building-requirements-partition-walls-62677.html

[31] JBL 2016. JBL Flip 2. (2016). http://www.jbl.com/bluetooth-speakers/JBL+FLIP+II.html.

[32] Kiran Raj Joshi, Dinesh Bharadia, Manikanta Kotaru, and Sachin Katti. 2015. WiDeo: Fine-grained Device-free Motion Tracing using RF Backscatter. In *NSDI*.

[33] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. 2014. Bringing Gesture Recognition To All Devices. In *NSDI*.

[34] Manikanta Kotaru, Kiran Raj Joshi, Dinesh Bharadia, and Sachin Katti. 2015. SpotFi: Decimeter Level Localization Using WiFi. *Computer Communication Review* 45 (2015), 269–282.

[35] Qiang Lin and Yuan Yue. 2015. Device-Free Passive Human Detection Using Wi-Fi Technology: Current State and Future Trend. *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)* (2015), 1717–1723.

[36] LTE 2016. An Introduction to LTE, 3GPP LTE Encyclopedia. (2016). https://sites.google.com/site/lteencyclopedia/home.

[37] mobisante 2016. Smartphone Ultrasound: The MobiUS SP1 System. (2016). http://www.mobisante.com/products/product-overview/.

[38] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaneil Watson. 2015. Contactless Sleep Apnea Detection on Smartphones. In *MobiSys*.

[39] Rajalakshmi Nandakumar, Vikram Iyer, Shyamnath Gollakota, and Desney Tan. 2016. FingerIO: Fine-Grained Finger Tracking Using Active Sonar. In *CHI*.

[40] Rajalakshmi Nandakumar, Bryce Kellogg, and Shyamnath Gollakota. 2014. Wi-Fi Gesture Recognition on Existing Devices. *CoRR* abs/1411.5394 (2014).

[41] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-Home Gesture Recognition Using Wireless Signals. In *MOBICOM*.

[42] Radar Through Wall 2012. Through-the-Wall Sensors for Law Enforcement. (October 2012). https://www.justnet.org/pdf/00-WallSensorReport-508.pdf.

[43] T.S. Ralston, G.L. Charvat, and J.E. Peabody. 2010. Real-time through-wall imaging using an ultrawideband multiple-input multiple-output (MIMO) phased array radar system. In *ARRAY*.

[44] Franziska Roesner, Tadayoshi Kohno, Alexander Moshchuk, Bryan Parno, Helen J. Wang, and Crispin Cowan. 2012. User-Driven Access Control: Rethinking Permission Granting in Modern Operating Systems. *2012 IEEE Symposium on Security and Privacy* (2012), 224–238.

[45] Ahmed Saeed, Ahmed E. Kosba, and Moustafa Youssef. 2014. Ichnaea: A Low-Overhead Robust WLAN Device-Free Passive Localization System. *IEEE Journal of Selected Topics in Signal Processing* 8 (2014), 5–15.

[46] Seek 2016. Seek Thermal Camera Review: Smartphone thermal vision in a tiny package. (2016). http://www.cnet.com/products/seek-thermal-camera/.

[47] Moustafa Seifeldin, Ahmed Saeed, Ahmed E. Kosba, Amr El-Keyi, and Moustafa Youssef. 2013. Nuzzer: A Large-Scale Device-Free Passive Localization System for Wireless Environments. *IEEE Transactions on Mobile Computing* 12 (2013), 1321–1334.

[48] Sharp TV 2017. Sharp LC-42SB45UT 42" 1080p LCD TV . (2017). https://www.amazon.com/Sharp-LC-42SB45UT-42-1080p-LCD/dp/B001F0QS9G.

[49] Adam Smith, Hari Balakrishnan, Michel Goraczko, and Nissanka Priyantha. 2014. Tracking Moving Devices with the Cricket Location System. In *Mobisys*.

[50] sonardyne 2016. Sonardyne: Sound in Depth. (2016). http://www.sonardyne.com/.

[51] Stoned 2015. Afghan woman stoned to death for 'adultery'. (2015). http://www.cnn.com/2015/11/04/asia/afghanistan-taliban-woman-stoning/.

[52] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. [n. d.]. WiDraw: Enabling Hands-free Drawing in the Air on Commodity WiFi Devices. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom '15)*.

[53] Zheng Sun, Aveek Purohit, Raja Bose, and Pei Zhang. 2013. Spartacus: Spatially-aware Interaction for Mobile Devices Through Energy-efficient Audio Sensing. In *Mobisys*.

[54] H. Tanoto, J. H. Teng, Q. Y. Wu, Z. N. Chen, S. A. Maier, B. Wang, C. C. Chum, G. Y. Si, and A. J. Danner. 2013. Retraction: Greatly enhanced continuous-wave terahertz emission by nano-electrodes in a photoconductive photomixer. In *Nat Photon*.

[55] Thermal Fact 2016. Thermal Imaging: Facts versus Fiction. (2016). https://pr-infrared.com/about-thermal-imaging/thermal-imaging-facts-vs-fiction/.

[56] Ju Wang, Hongbo Jiang, Jie Xiong, Kyle Jamieson, Xiaojiang Chen, Dingyi Fang, and Binbin Xie. 2016. LiFS: low human-effort, device-free localization with fine-grained subcarrier information. In *MobiCom*.

[57] Junjue Wang, Kaichen Zhao, Xinyu Zhang, and Chunyi Peng. 2014. Ubiquitous Keyboard for Small Mobile Devices: Harnessing Multipath Fading for Fine-grained Keystroke Localization. In *MobiSys*.

[58] Wei Wang, Alex X. Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *UbiComp*.

[59] Wei Wang, Alex X. Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *MobiCom*.

[60] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: Device-free Location-oriented Activity Identification Using Fine-grained WiFi Signatures. In *MOBICOM*.

[61] Teng Wei, Shu Wang, Anfu Zhou, and Xinyu Zhang. 2015. Acoustic Eavesdropping Through Wireless Vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*.

[62] Norbert Wild. 2001. *Ultrasonic through-the-wall surveillance system.* Technical Report. International Society for Optics and Photonics. 167–176 pages.

[63] J. Wilson and N. Patwari. 2011. See-Through Walls: Motion Tracking Using Variance-Based Radio Tomography Networks. *IEEE Transactions on Mobile Computing* (2011).

[64] Dan Wu, Zhang Daqing, Chenren Xu, Yasha Wang, and Hao Wang. 2016. WiDir: walking direction estimation using wireless signals. In *UBICOMP*.

[65] Jiang Xiao, Kaishun Wu, Youwen Yi, Lu Wang, and Lionel M. Ni. 2013. Pilot: Passive Device-Free Indoor Localization Using Channel State Information. *2013 IEEE 33rd International Conference on Distributed Computing Systems* (2013), 236–245.

[66] Robert Xiao, Greg Lew, James Marsanico, Divya Hariharan, Scott Hudson, and Chris Harrison. 2014. Toffee: Enabling Ad Hoc, Around-device Interaction with Acoustic Time-of-arrival Correlation. In *MobileHCI*.

[67] Jie Xiong and Kyle Jamieson. 2012. Towards fine-grained radio-based indoor location. In *HotMobile*.

[68] Qinyi Xu, Yan Chen, BeiBei Wang, and K. J. Ray Liu. 2017. Radio Biometrics: Human Recognition Through a Wall. *IEEE Transactions on Information Forensics and Security* 12 (2017), 1141–1155.

[69] Lei Yang, Qiongzheng Lin, Xiangyang Li, Tianci Liu, and Yunhao Liu. 2015. See Through Walls with COTS RFID System!. In *MOBICOM*.

[70] Sangki Yun, Yi-Chao Chen, and Lili Qiu. 2015. Turning a Mobile Device into a Mouse in the Air. In *MobiSys*.

[71] Y. Zhao and N. Patwari. 2014. Robust Estimators for Variance-Based Device-Free Localization and Tracking. *IEEE Transactions on Mobile Computing* (2014).

[72] Yanzi Zhu, Yibo Zhu, Ben Y. Zhao, and Haitao Zheng. 2015. Reusing 60GHz Radios for Mobile Radar Imaging. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom '15)*.

Received February 2017; revised May 2017; accepted July 2017