Dhwani: Secure Peer-to-Peer Acoustic NFC

Rajalakshmi Nandakumar, Krishna Kant Chintalapudi, Venkata N. Padmanabhan,
Ramarathnam Venkatesan
Microsoft Research India

ABSTRACT

Near Field Communication (NFC) enables physically proximate devices to communicate over very short ranges in a peer-to-peer manner without incurring complex network configuration overheads. However, adoption of NFC-enabled applications has been stymied by the low levels of penetration of NFC hardware.

In this paper, we address the challenge of enabling NFC-like capability on the existing base of mobile phones. To this end, we develop *Dhwani*, a novel, acoustics-based NFC system that uses the microphone and speakers on mobile phones, thus eliminating the need for any specialized NFC hardware. A key feature of Dhwani is the *JamSecure* technique, which uses self-jamming coupled with self-interference cancellation at the receiver, to provide an information-theoretically secure communication channel between the devices. Our current implementation of Dhwani achieves data rates of up to 2.4 Kbps, which is sufficient for most existing NFC applications.

Categories and Subject Descriptors

C.2.m [**Computer Systems Organization**]: COMPUTER - COMMUNICATION NETWORKS—*Miscellaneous*

Keywords

NFC, Wireless, Security

1. INTRODUCTION

Near-Field Communication (NFC) enables low data rate, bidirectional communication between devices within close proximity, usually within a few centimeters, in a peer-to-peer manner. The key advantage of NFC is that it eliminates the need for cumbersome network configuration efforts required to set up a communication channel using alternatives such as Bluetooth or WiFi. This is due to its inherent property of *association by physical proximity* — if two devices can communicate using NFC, then it implies that they must be co-located. As an example, using an NFC enabled mobile phone, a user can make payments by simply bringing the phone close to a reader at the checkout counter, without having to first identify the reader or connect to it.

Several NFC-based applications have been proposed or demonstrated, e.g., contact-less payment, access control, social networking, ticketing, museum services, etc. In many cases, NFC is used to automatically initiate and set up a high data rate communication channel such as WiFi or Bluetooth. However, the adoption of these applications has been stymied by the low levels of penetration of NFC hardware, estimated to be just 3-5% [11] among mobile phones worldwide and only about 12% [10] even in an advanced market such as the U.S., as of 2012. Even as far out as 2016, the penetration is expected to be under 50%. Correspondingly, the prevalence of NFC-enabled point-of-sale (POS) terminals is also low — under 5% today and expected to rise to only about 49% globally by 2017 [9]. Even disregarding the optimism that usually colours such forecasts, it seems likely that the majority of phones and POS terminals globally will *not* be NFC-enabled even 3-4 years from now. Thus, the opportunities for using NFC applications such as peer-to-peer transfers or contact-less payments will remain rather limited.

Can we enable NFC-like functionality on today's devices? We answer this question in the affirmative by presenting *Dhwani*, a novel, acoustics-based system that uses the existing microphones and speakers on phones to enable NFC, thus, eliminating the need for specialized NFC hardware. As in conventional NFC, where communication through magnetic coupling is confined to a short range, acoustic communication in Dhwani is confined to a short range (few cm). Thus, similar to conventional NFC, Dhwani enables the "association by proximity" functionality needed for applications such as P2P transfers and contact-less payments.

A key advantage of Dhwani over conventional NFC is that it is a purely software-based solution, that can run on legacy phones, including feature phones, so long as they have a speaker and a microphone. Consequently, much of the installed base of phones today could use Dhwani to perform P2P NFC communication. That said, the use of acoustic communication means that, unlike conventional NFC, Dhwani is *not* amenable to implementation in passive tags.

A second significant advantage of Dhwani over conventional NFC is in terms of information-theoretic, physical-layer security. As discussed in Section 3.1, the security model in Dhwani is that the devices seeking to communicate are trusted and immune to tampering. However, in their midst might be one or more eavesdroppers. Conventional NFC does not incorporate any security at the physical or MAC layers since the short range of communication (about 10 cm) is in itself presumed to offer a degree of protection. However, in [16], the authors demonstrate that it is possible to snoop on NFC communications from a distance of 20-30 cm using an oscilloscope and a standard tag antenna. The authors also conjecture that with a more sophisticated sniffer antenna, such snooping should be possible from a distance of a meter or more.

SIGCOMM'13, August 12–16, 2013, Hong Kong, China.
Copyright 2013 ACM 978-1-4503-2056-6/13/08 ...$15.00.

Dhwani provides security at the physical layer using a novel self-jamming technique, *JamSecure*, wherein the receiver intentionally jams the signal it is trying to receive, thereby stymying eavesdroppers, but then uses self-interference cancellation to successfully decode the incoming message. The security thus obtained is information-theoretic, i.e., Dhwani inherently prevents the leakage of information to an eavesdropper. This is in contrast to cryptographic security, which is based on assumptions about computational hardness. Even if cryptographic security protocols are employed at the higher layers, Dhwani enables key exchange without the need for any shared secret or certificates to be set up a priori. This is a significant advantage, since creating a public key infrastructure (PKI) spanning billions of devices would be challenging.

In order to enable Dhwani we implemented an Acoustic Software Defined Radio (ASDR) on the mobile devices that uses speakers and microphones to receive and transmit data. Our ASDR design had to address several challenges unique to the nature of the acoustic signal propagation and speaker-microphone characteristics. For example, we found the gain of the speaker-microphone combination in phones to be extremely non-uniform across the range of frequencies (frequency selectivity), presumably due to the mechanical properties of their electro-mechanical parts (*e.g.,*vibrating membranes). Further, the high degree of ringing in the acoustic channel (reverberations), compared to Radio Frequency (RF), rendered the existing RF self-interference cancellation techniques inadequate. Consequently, for Dhwani, we present a novel and efficient technique for self-interference cancellation, which takes advantage of the fact that the jamming sequence can be predetermined by the receiver.

We present the design and implementation of Dhwani, an analysis of its security properties, and an experimental evaluation on mobile devices such as phones and laptops. To sum up, the main contributions of our work are

- A characterization of the acoustic hardware and environment in the context of mobile phones.
- An Acoustic Software Defined Radio suitable for operation on mobile phones.
- The JamSecure self-jamming technique for providing information-theoretic, physical-layer security.

2. AN NFC PRIMER

As described in Section 1, NFC enables configuration-free low data rate communication between two devices in close physical proximity. NFC standards (ISO/IEC 18092/ECMA-340, NFC IP-1, ISO/IEC 14443) have evolved from RFID technology. However, while RFID readers can read tags up to distances of a few meters, NFC readers are designed to read at distances of a few centimeters.

NFC devices can operate either in an *active* mode, in which the device (e.g., a reader) generates its own electromagnetic field, or in a *passive* mode, in which the device (e.g., a tag) is powered by the electromagnetic field generated by another device in its proximity. There are three modes of NFC interaction available for a mobile device such as a phone:

- **Read/Write:** An NFC-enabled phone, operating in active mode, can Read/Write data from/to a passive tag.
- **Peer-to-Peer (P2P):** Two NFC-enabled phones, each operating in active mode, can exchange data.
- **Card Emulation:** An NFC-enabled phone can emulate a smart card, allowing an active reader to read from it.

In this paper, we limit ourselves to the P2P mode of NFC.

How NFC works. Current day NFC technology works on the principle of magnetic induction. Each NFC device is equipped with an antenna coil. Typically, one of the devices initiates communication by passing a current through its antenna coil. This current generates a magnetic field, which then induces current in the receiving device's antenna coil. Thus, the two devices essentially form an air-core transformer. Data is transmitted by modulating the current passed through the transmitter coil.

Existing NFC standards employ Amplitude Shift Keying (ASK) in the 13.56 MHz spectrum, with a bandwidth of about 1.8 MHz. Three different data rates are supported: 106, 212 and 424 Kbps. Typically, Manchester coding with 10% modulation is used, implying that the low and high amplitudes are 10% off on either side of the carrier amplitude.

NFC is intended only for small data transfers; e.g., NFC tags are typically equipped with a memory size of 96 to 512 bytes. Often, when a large amount of data needs to be transferred, NFC is only used to set up the initial connection for a higher data rate standard such as Bluetooth or WiFi. Instant user gratification is an important requirement of NFC, so the communication delay should not exceed a few seconds.

Security in NFC. The air interface and data link layer for NFC does not include any provision for security (NFCIP-1 [4]), with information being transmitted in the clear. For the P2P mode of NFC, newer security standards, layered on top of the data link layer, have been defined. NFC-SEC [7] defines the framework for security services, including a shared secret service and a secure channel service. The actual security protocols are specified in NFC-SEC-01 [6], including Elliptic Curves Diffie-Hellman (ECDH) for key agreement and the AES algorithm for data encryption and integrity.

However, as noted in [5], NFC-SEC-01 does not protect against man-in-the-middle attacks because no entity authentication can be provided when the peer NFC devices do not share any secret a priori. It is further noted that the practical risk of a man-in-the-middle attack is low due to the short operating distance, but that users should be aware of and carefully evaluate the potential vulnerability in their setting.

The authors in [13] discuss various attacks, including eavesdropping and data modification, that could be mounted on NFC at the physical layer. They report eavesdropping ranges of 1m and 10m, respectively, for the passive and active modes. Furthermore, an attacker can perform data modification (particularly with the 10% modulation that is commonly employed) by injecting signal energy during a "low" period to make the corresponding amplitude higher than in the following "high" period, thereby flipping the corresponding bit.

Compared to the NFC security enhancements, such as NFC-SEC-01, the physical-layer security provided by Dhwani avoids the possibility of man-in-the-middle attacks by allowing the peers to securely establish a secret without requiring any a priori shared secret or third-party communication.

3. DHWANI - THE KEY IDEAS

The goal of Dhwani is to enable NFC-like functionality, i.e., configuration-free short-range communication, in a wide array of existing mobile devices, while also ensuring physical-layer security. In this section, we present an overview of Dhwani highlighting the key novel aspects.

3.1 Security Model in Dhwani

The security goal of Dhwani is to ensure the secrecy and integrity of messages exchanged between a transmitter and receiver pair located within close proximity (a few centimeters), in the presence of attackers. In this section we make the following assumptions about Dhwani's operation and security model:

- Both transacting devices (transmitter and receiver) are trusted entities. These devices are assumed to function correctly and execute the Dhwani protocol faithfully. Any failure is presumed to be only accidental (e.g., due to a power outage).
- The attacker is presumed to be capable of mounting both passive (e.g., eavesdropping) and active attacks (e.g., message insertion). However, we assume that the attacker is unable to directly tamper with the trusted entities or alter their functioning.
- The communication range of the transacting devices is limited to a few centimeters.

The above assumptions are consistent with the NFC model, wherein association, and the consequent transaction, happen implicitly through physical proximity. So, for instance, users who swipe their NFC-capable cards at a point-of-sale (POS) terminal are presumed to have satisfied themselves about the authenticity of the POS terminal, say based on its location in the check-out area of a reputable store. The only concern would be the possibility of attackers lurking in the vicinity. Note that as discussed in Section 8, Dhwani's security can be potentially subverted, albeit with great difficulty, using sophisticated directional antenna or antenna arrays. However, we believe that Dhwani raises the bar for active and passive attackers significantly compared to the state of the art.

3.2 Acoustic Characterization

While the use of the acoustic channel for NFC offers the promise of a broad footprint, we have to contend with the peculiarities of both the acoustic hardware (speakers and microphones) in mobile devices and the acoustic environment. While acoustic communication has been studied with specialized hardware and in specific domains such as underwater communication, we are not aware of prior work on characterizing off-the-shelf mobile devices in the context of over-the-air communication, as we present in Section 4. We find a high degree of ambient noise, significant ringing (reverberations), and highly frequency selective fading due to the electromechanical nature of the speakers and microphones. These findings inform the design of the Acoustic Software Defined Radio and also the JamSecure technique in Dhwani.

3.3 Acoustic Software Defined Radio

Dhwani provides an Acoustic Software Defined Radio (ASDR) service, which applications can use to transmit or receive packets. As described in Section 5, Dhwani's ASDR implements almost all of the functionality of a standard modern day radio, including OFDM modulation and demodulation, error correction coding, etc. However, a key difference in Dhwani's ASDR compared to traditional RF radios is that it has no notion of a carrier frequency and a separate baseband. The reason is that the ADC is able to sample at a rate (44 KHz) that is sufficient for the entire acoustic bandwidth supported by the speaker and microphone. A sampling rate of 44 KHz allows operating (at best) in the 0-22 KHz band. Consequently, Dhwani implements a carrier-less OFDM over the entire 0-22KHz band, simply suppressing (i.e., nulling) sub-carriers that are not suitable for use, either because of the ambient noise (Section 4.1) or because of the speaker and microphone characteristics.

3.4 JamSecure

JamSecure is a novel self-jamming technique used by the receiver in Dhwani to cloak the message being transmitted by the transmitter, thereby preventing an attacker from receiving the message. Figure 1 depicts the key idea behind JamSecure. Transmitter A wishes to transmit a message M to receiver B while an eavesdropper E attempts to listen to message M. As A transmits its message with a power P_A dBm, simultaneously, B jams A's transmission by transmitting a Pseudorandom Noise (PN) sequence with power P_B dBm. The PN sequence is generated afresh for each secure reception and is known only to B. The eavesdropper E can only overhear the combination of the message M from A and the jamming noise from B. The received Signal to Noise Ratio (SNR) at E will thus be $P_A - P_B$ dB. If P_B is high enough, then information-theoretically, E will not be able to extract any useful information about M. While B also receives a combination of its own jamming and the message M, it performs Self Interference Cancellation (SIC), *i.e.,* subtracts the (known) jamming signal from the received signal, in an attempt to retrieve M. Since SIC is not perfect in practice, suppose that B can cancel IC dB of its own signal. Then, the SNR seen by B is $P_A - P_B + IC$ dB. If IC is "high enough", B will be able to retrieve the message M from A.

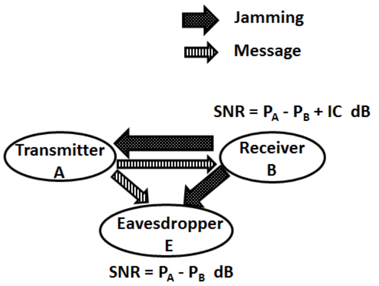

Figure 1: JamSecure

While SIC is conceptually simple, the characteristics of the acoustic hardware and channel make it challenging to directly perform channel estimation for SIC. Instead, as discussed in Section 6.2, we employ a hybrid offline-cum-online approach, which works with a predetermined library of PN sequences, and random combinations thereof.

As discussed in Section 8, Dhwani's approach to physical-layer security can be viewed as a fusion of Wyner's wiretap model [21] (by ensuring differential SNR for the intended receiver versus an attacker) and Shannon's one-time pad [20] (through the use of a pseudo-random jamming noise). As such, this approach is not confined to acoustic communication and could, in principle, be employed in other contexts too, e.g., to enable an RFID reader to securely read a tag.

3.5 Other Aspects of Dhwani

We briefly touch on a couple of other elements of Dhwani's design, including pointers to later sections for elaboration.

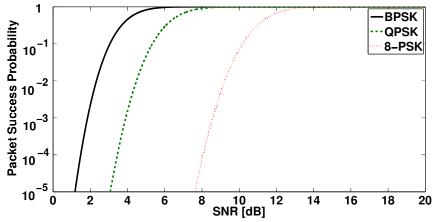

Figure 2: PSR-SNR Curves for various physical layer modulations

How much jamming is needed? For each physical-layer modulation technique, the SNR at the receiver imposes a theoretical lower bound on the Bit-Error-Rate (BER), and hence an upper bound on the Packet Success Rate (PSR) for error-free reception. Figure 2 depicts the best possible PSR that can be achieved for a 256-bit packet, as a function of SNR, for BPSK, QPSK, and 8-PSK. The key observation is that in each case, PSR falls very sharply around a certain SNR threshold; e.g., with QPSK, just a 4dB drop in SNR (from 6dB to 2dB) causes PSR to fall by 5 orders of magnitude.

In Dhwani we need to ensure that the receiver injects enough noise that the SNR at the eavesdropper is to the left of the chosen curve in Figure 2 while, at the same time, the SNR at the receiver itself, with the benefit of SIC, is to the right. We discuss how Dhwani achieves this balance in Section 7.2.

Scrambling the message. Receiving a message with errors might still leak information by allowing the attacker to retrieve parts of it. To address this issue, Dhwani uses a scrambler prior to transmitting the message, which ensures that even a single bit of error in the scrambled message would result in a large number of bit errors in the unscrambled message. We repurpose AES, which is designed for ensuring message secrecy, for scrambling instead (Section 7.1).

4. THE ACOUSTIC CHANNEL

The design of any communication system depends fundamentally on the characteristics of the communication medium or the communication channel. Specifically, for Dhwani there are three key properties that influence its design: *ambient noise*, *acoustic channel*, and *acoustic propagation*. In this section, we characterize these three properties, both qualitatively and quantitatively, through measurements using various mobile devices in different settings.

4.1 Ambient Noise

A key requirement of Dhwani is that it must operate in public spaces such as malls and cafes where the ambient (acoustic) noise can cause significant interference. To characterize this interference, we measured the received acoustic power in a range of environments such as malls, cafes, and office conference rooms at various times. First we measured the noise floor of the mobile device in an isolated, silent room. Next we collected ambient sound samples on the same device in various venues. Figure 3 depicts the ratio of ambient sound energy to the noise floor as a function of frequency measured on a Samsung Galaxy S2 phone in two public venues – payment counter at a popular mall, and a cafe during busy hours.

As seen from Figure 3, the ambient noise in both the mall and the cafe can be significantly high – up to 25-30dB (1000 times) above the noise floor of the phone at frequencies below 1.5KHz. Even at frequencies up to 5KHz, the ambient interference can be as high as 10dB (10 times). This is because while human voices rarely exceed 1KHz, several public venues (including the ones in Figure 3) have background music or televisions which contribute to the noise at higher frequencies. The cafe had a higher ambient noise than the mall, not only due to human chatter but also because the background music and television sounds were louder. Beyond 6KHz however, the ambient interference is almost close to noise levels and becomes negligible after 8KHz. *These observations imply that 6KHz forms a lower limit for the operation of Dhwani.*

4.2 The Channel Transformation

When a digital acoustic signal $s(k)$ is transmitted, a distorted version $r(k)$ is received at the receiver. Specifically, if the signal is represented as a sum of M sinusoids, $f_1, f_2, \cdots f_M$ (Fourier Transform representation), then in the received signal, each of these sinusoids experiences frequency-dependent attenuation $a(f_i)$, and phase distortion $\Delta\phi(f_i)$ as follows:

$$\begin{aligned} s(k) &= \sum_{i=1}^{i=M} \cos\left(2\pi f_i \frac{k}{F_s} + \phi_i\right) \\ r(k) &= \sum_{i=1}^{i=M} a(f_i) \cos\left(2\pi f_i \frac{k}{F_s} + \phi_i + \Delta\phi(f_i)\right) \end{aligned} \tag{1}$$

Eqn 1 can be represented in complex form as:

$$\begin{aligned} s_{cplx}(k) &= \sum_{i=1}^{i=M} e^{j2\pi f_i \frac{k}{F_s} + \phi_i} \\ r_{cplx}(k) &= \sum_{i=1}^{i=M} \left[a(f_i) e^{j\Delta\phi(f_i)}\right] e^{j2\pi f_i \frac{k}{F_s} + \phi_i} \end{aligned} \tag{2}$$

In Eqn 2, the complex number $a(f_i)e^{j\Delta\phi(f_i)}$ is referred to as the channel gain at frequency f_i.

4.2.1 Frequency Selectivity

Frequency selectivity refers to selective attenuation of certain frequencies in the transmitted signal. There are two key reasons for frequency selectivity in Dhwani – *microphone/speaker selectivity* and *multipath*.

Microphone-Speaker Frequency Selectivity. Sound is a mechanical wave. Consequently, speakers and microphones have mechanical components (*e.g.,*vibrating membranes) required for electro-mechanical conversion. Frequency selectivity arises because of the inability of these components to faithfully reproduce tones of certain frequencies. Even though most mobile phones today allow for an acoustic sampling rate of up to 44KHz, implying a maximum operating frequency of 22KHz, their speaker/microphones components are typically designed for human speech, and their performance degenerates significantly at higher frequencies.

Multipath. Multipath (echo) is common in sound propagation and leads to the superposition of several time delayed (and attenuated) copies of the transmitted signal at the receiver. The net effect of multi-path is spreading of the received signal in time, and constructive/destructive interference at various frequencies, leading to frequency selectivity.

Examples of Acoustic Channels. Figure 4 depicts the *frequency response* (the function $a(f)$ in Eqn 1) of a Speaker-Microphone channel for three different acoustic communication links — a Samsung Galaxy S2 phone to a HP Mini laptop, a Samsung Galaxy S2 to HTC Sapphire and finally a HP Mini to a Samsung Galaxy S2. The frequency responses were measured by transmitting tones of frequencies between 100Hz and 20KHz, from one device to another, while placing the devices within 10cm of each other. The frequency responses in Figure 4 are normalized to the maximum power received for any single tone during the course of the measurement.

An ideal frequency response should be a line at 0dB horizontal to the x-axis, indicating that all frequencies experience the same overall attenuation from transmission to reception. However, Figure 4 shows that this is far from being the case. We make two key observations from the figure:

- *Attenuation at high frequencies:* In all cases, we see a significant degradation at higher frequencies, especially after 12KHz. This implies that if we use a frequency band for communication that spans beyond 12KHz, there will be a significant information loss corresponding to the part of the band beyond 12KHz.
- *Notches:* As seen from Figure 4, the frequency responses for all pairs of devices is extremely uneven and has deep notches (valleys) in various parts of the spectrum, even at frequencies much lower than 12KHz. This unevenness causes the shape of the received waveform to be distorted relative to the transmitted waveform, resulting in decoding errors.

Based on these observations, we conclude that Dhwani should avoid frequencies beyond 12KHz while also working around the notches at the lower frequencies.

Multipath vs. Speaker-Microphone Frequency Selectivity. The frequency selectivity evident in Figure 4 arises from a combination of both the characteristics of electro-mechanical components of the

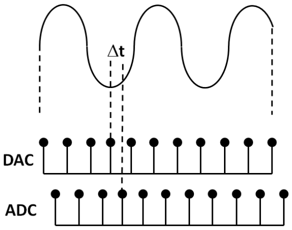

Figure 7: Phase Distortion due to sampling offset.

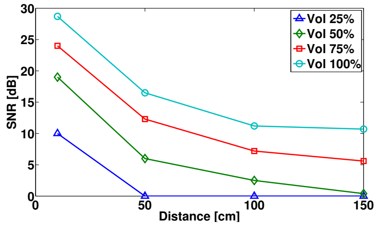

Figure 8: Decay of received SNR with distance for various volume settings

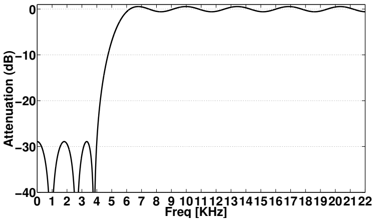

Figure 9: Response of the Filter used in Dhwani

was motivated by the fact that it is particularly well-suited to frequency-selective communication channels. Our OFDM radio allows the choice of various sub-carrier modulation schemes such as BPSK, QPSK, 16QAM, *etc.*, and includes basic error correction coding mechanisms. Rather than describing the well-known aspects of OFDM, we focus in this section only on the aspects that are unique to our implementation.

5.1 Ingress Filter

In order to be immune to typical ambient noise, the Dhwani receiver first applies a digital filter. As seen in Figure 3, and described in Section 4, ambient noise can be as high as 25-30dB above the noise floor at frequencies below 1.5KHz and up to 10dB at 5KHz. The ambient noise above 6KHz is typically negligible. Consequently, at the receiver, we use a High-Pass Finite Impulse Response filter, which allows only frequencies greater than 6kHz. Figure 9 shows the frequency response of the filter we used. As seen in the figure, the filter attenuates all frequencies below 4KHz by 30dB, thus practically annulling the effects of all ambient noise. The filter provides close to 0dB gain for frequencies greater than 6KHz, and consequently allows higher frequencies to pass.

5.2 OFDM Design

A key difference between the radio-frequency OFDM radios and Dhwani is the absence of a carrier wave in the latter (unlike WiFi, for instance, which uses a 2.4GHz carrier). This is because, unlike WiFi, the ADC (at the microphone) and DAC (at the speaker) can sample at a rate commensurate with the entire available bandwidth of 22KHz. The OFDM subcarriers are thus spread over the entire 22KHz in our implementation. While increasing the number of subcarriers helps combat high frequency selectivity, it also adversely effects the Peak to Average Power Ratio (PAPR). In our implementation we choose a sub-carrier width of 171 Hz (128 subcarriers in the range 0-22KHz). At a sampling rate of 44KHz, this leads to an OFDM symbol length of about 5.8ms.

Choosing an operating bandwidth. The ingress filter filters frequencies below 6KHz, while the speaker/microphone do not transmit/receive well at high frequencies. Consequently, a suitable operating bandwidth needs to be chosen. In our implementation we chose 1KHz of bandwidth in the range 6-7KHz as our operating bandwidth, which is a conservative choice that works well across all the platforms we tested. [2] Choosing an operating bandwidth of 6-7KHz in our system corresponds to transmitting zero energy in the remaining sub-carriers.

Real to Complex Signal Representation. The key advantage of the complex representation of a signal (Eqn 2) is that the phase of each sample can be readily extracted from the ratio of its real (in-phase) and imaginary (quadrature-phase) parts. In contrast, the real representation (Eqn 1) is not amenable to such a computation. Ready access to phase is crucial for many operations such as preamble correlation, demodulation *etc.*In a typical radio, the mixer (responsible for mixing *i.e.,* up/down conversion from the carrier frequency) generates both real and complex samples from the received real signal at the carrier frequency, as a part of mixing process. In Dhwani, however, since the sound-card provides only 16-bit real samples, the receiver does not have the luxury of being provided complex samples.

The Dhwani receiver uses *negative sideband suppression* to convert real signal to its complex representation. This scheme relies on the property that the complex representation can be obtained from its real counterpart by setting all its negative frequency components to zero. Suppose that the received digital signal is $s_{real}(k)$ and has a total of N samples. The first step in this scheme is to compute the N-point Fourier transform S_k. Then, all the negative sideband Fourier coefficients $S(k), k > N/2$ are set to zero to obtain $S_+(k)$. Thereafter, the corresponding complex representation s_{cplx} is obtained by taking an N point inverse Fourier transform of $S_+(k)$.

Cyclic Prefix. As depicted in Figure 6, reverberations in the acoustic channel last over 25ms (time taken to decay to noise floor or 0dB SNR). To combat Inter Symbol Interference (ISI) due to ringing in the channel (Section 4), we experimentally found that a 4ms long cyclic prefix worked well in practice in all environments we tested for modulations such as BPSK and QPSK. This is because the energy of the reverberations decay by more than 10dB after about 4ms, which is sufficient to allow reliable detection of QPSK symbols despite inter-symbol interference.

Preamble. For synchronizing the OFDM transmitter and receiver, a preamble precedes each transmitted packet. For our preamble, we used a chirp whose frequency increased from f_{min} to f_{max} in the first half and then decreased back to f_{min} as follows,

$$P(t) = \begin{array}{ll} e^{i\pi a t^2} & for \quad t < T \\ e^{i\left[\phi_0 + f_{max}(t-T) - \pi a (t-T)^2\right]} & for \quad T < t < 2T \\ a = \frac{f_{max} - f_{min}}{T} & \\ \phi_0 = \pi a T^2 & \end{array} \qquad (3)$$

The reason for the choice of this chirp was twofold. First, the chirp has a very low Peak-to-Average Power Ratio (PAPR), which makes it easier to detect compared to a standard OFDM-based preambles that have a higher PAPR. Second, experiments suggested that having the chirp frequency to first increase and then decrease led to a more accurate synchronization than using a chirp where frequency simply increases (or decreases). In our implementation, we chose f_{max} = 16KHz and f_{min} = 6KHz and T = 5.81ms (i.e., 256 samples long, given the sampling rate of 44KHz). One problem we found was that since the amplitude of the preamble is much larger than that of the OFDM transmission (due to low PAPR), a 3ms cyclic prefix proved insufficient for shielding from ISI the training symbols that immediately followed the preamble. Consequently, we padded each preamble with a silence period of

[2] Note that use of more advanced error correcting codes than used in our current implementation of Dhwani may allow the use of wider bandwidths.

roughly 4ms that allowed ringing of the preamble to subside significantly and reduced channel estimation errors.

Achieved Data Rates. The data rate achieved by Dhwani depends on the operating acoustic bandwidth (1 KHz in our current implementation), the modulation and error correction codes being used. In our current implementation, Dhwani achieves 2.4 Kbps corresponding to 8-PSK with about 80% PSR, around 95% PSR for QPSK (1.6Kbps) and 100% for BPSK (800bps) without any error correction. So for a short transfer of say 100 bytes, as would be typical of NFC transactions, Dhwani would take under a second. We believe that these rates could be further improved through the use of better error correcting codes with higher modulation schemes such as 16-QAM or 64QAM and wider bandwidths than 1KHz.

6. JAMSECURE

As described in Section 3, in Dhwani, a receiver defeats an eavesdropper by jamming the transmissions from the sender. It then uses *Self-Interference Cancellation (SIC)* to decode the transmission despite jamming. Consequently, there are two key goals in the design of jamming and SIC in Dhwani:

- *Security:* The jamming should be random and powerful enough that an eavesdropper is unable to cancel out the jamming and retrieve the message.
- *Effective SIC Cancellation:* At the same time, SIC must be good enough for the receiver to decode the message.

In this section we start by explaining the basic techniques used for SIC and what makes SIC in Dhwani especially difficult. We then describe *JamSecure*, a novel jamming technique that allows efficient SIC at the receiver while making cancellation practically impossible for an adversarial eavesdropper.

6.1 SIC Primer

The fundamental difficulty in performing SIC in Dhwani is that the transmitted signal $s(k)$, is affected by the speaker, microphone, and multipath, altering the received signal $r(k)$ (as discussed in Section 4). There has been a significant amount of research in terms of Self Interference Cancellation (SIC) in the context of full-duplex communication for radio frequency wireless communication. We present a quick overview of these methods and explain why these are not suitable for Dhwani.

Analogue SIC : In [14], the transmitted signal is fed back over a delayed path, attenuated, and then subtracted from the received signal. A key advantage of using this approach is that it allows for the detection of weak signals from a distant transmitter, by avoiding ADC saturation [14]. However, ADC saturation is not an issue in Dhwani since both the transmitter and the receiver are located within close proximity. Further, one of the design goals of Dhwani is that it should work on off-the-shelf components without any hardware additions.

Channel Estimation based SIC : If the communication channel is linear, then it can be modeled by a digital filter $H(t)$ whose Fourier transform corresponds to the complex *channel gains* $a(f)e^{\Delta\phi(f)}$ of the acoustic channel (Section 4). The key challenge then, is to accurately estimate the channel gains of the acoustic channel. Typically, the channel estimation is performed by the transmitter sending a well-known training signal, $p_{xmit}(t)$, prior to $s(t)$. The receiver then computes the channel gains, and hence $H(t)$, using the received version, $P_{rec}(t)$.

In multipath environments, the length of this filter (or, equivalently, the frequency resolution at which channel gains must be estimated) corresponds to the duration for which the channel reverberations (ringing) lasts. As described in Section 4, reverberations last for several tens of milliseconds (Figure 6). At a sampling rate of 44KHz, this corresponds to a filter with a response that lasts over a few thousand samples. The need for estimating such a large filter accurately limits the performance of this method in Dhwani.

6.2 Design of JamSecure

As discussed in Section 4, the frequency selectivity of the acoustic channel in Dhwani arises from two sources — the electro-mechanical components in the speaker/microphone, and multipath in the ambient environment. Also, as noted there, the effect of electro-mechanical components is a significant cause for frequency selectivity. Note that this is unlike RF, where antennas are specifically chosen not to be frequency selective in the operating bandwidth and most of the frequency selectivity is due to multipath. For SIC in Dhwani, the self interference channel primarily comprises that between the device's own speaker and microphone, which is constant for any given device. Consequently, if the (static) effect of the electro-mechanical components were estimated ahead of time, say during the initial configuration, then the task of channel estimation at runtime becomes much easier.

Training Phase. During initial configuration as part of the training phase in JamSecure, the device transmits a library of PN sequences PN^i_{xmit}, $i = 1, \ldots, M$ of length N samples each, with each sequence being preceded by a preamble (a chirp, as discussed in Section 5). The device also simultaneously records the received versions of the corresponding PN sequences PN^i_{recv} at the microphone, using the preamble to determine the start of each received PN sequence in the library. In the rest of the paper we shall refer to M as the *library size*.

Generating the Jamming Signal. In order to generate a jamming signal, $JamPN^J_{xmit}$ for the J^{th} transaction, the receiver first chooses a random subset of K PN sequences from the library, $PN^{n^J_i}_{xmit}$,$i = 1,\cdots,K$, n^J_i being the index of the i^{th} randomly chosen PN sequence among K. The m^{th} sample of the jamming sequence is generated as,

$$JamPN^J_{xmit}(m) = \sum_{i=1}^{i=K} \frac{1}{K} PN^{n^J_i}_{xmit}(m) \tag{4}$$

In the rest of the paper we shall refer to K as the *mixing factor*.

If the acoustic channel remained exactly the same for every transaction and there were no sampling offset errors (discussed in Section 4), then based on the linearity of the channel, the received version of this PN sequence can be written as,

$$JamPN^J_{recv}(m) = \sum_{i=1}^{i=K} \frac{1}{K} PN^{n^J_i}_{recv}(m) \tag{5}$$

In practice, however, the received signal will be different from Eqn 5, as the sampling offsets and preamble synchronization errors are non-zero, and the multipath environment changes. Consequently, as we discuss later in this section, JamSecure estimates and compensates for these effects at runtime.

Choosing Library Size (M) and Mixing Factor (K). The key objective of M and K is to thwart the eavesdropper from learning/predicting the PN sequence generated by the receiver. For example, upon hearing the jamming sequence several times, an eavesdropper may learn the sequence accurately, and perform jamming cancellation at its end. Given a large number of possible combinations — M choose K — the eavesdropper receives a new sequence each time and it becomes hard for it to learn this library of PN sequences. For example, if $M = 1000$ and $K = 5$, the number of possible sequences increases to 10^{15}. Thus, even for small values of $K = 5$, it becomes computationally intractable for the eavesdropper to learn the library of sequences and perform interference cancellation. Further, the library could be refreshed periodically to ensure that it cannot be learned even over a long period.

Dealing with sampling offset. As described in Section 4, a sampling offset of Δt introduces a phase error of $2\pi f \Delta t$ at frequency f. Consequently, compensating for sampling offset is equivalent to shifting the phase of the frequency component corresponding to f in the signal's Fourier representation by $2\pi f \Delta t$. To achieve this we first compute the Fourier transform of the received sequence and then shift each frequency component's phase by multiplying with the complex cosine $e^{j2\pi f \Delta t}$, f being the frequency. Finally, we obtain the delayed version $PN^i_{recv,\Delta t}$ by taking an inverse Fourier transform. This entire procedure can be written as,

1. Compute the N point Fourier Transform of PN^i_{recv}, denoted as Φ^i_{recv}.

2. Compute sampling offset version $\Phi^i_{recv,\Delta t}(f) = \Phi^i_{recv}(f)$ $e^{j2\pi f \Delta t}$, here f is the frequency and spans from $-\frac{F_s}{2}$ to $\frac{F_s}{2}$, F_s being the sampling rate.

3. Compute the N point Inverse Fourier Transform of $\Phi^i_{recv,\Delta t}$ to obtain $PN^i_{recv,\Delta t}$.

One challenge that remains is that the receiver does not know Δt. To address this problem, during pre-configuration time, we also pre-compute and store several phase-shifted versions $PN^i_{recv,\Delta t}$ corresponding to $0 < \Delta t < \frac{1}{F_s}$. The best delayed version is then recovered by comparing the first few samples (512 in our implementation) of the received jamming signal and matching it to the closest sampling offset version corresponding to Δt_{opt} based on minimizing squared error distance. The received jamming version is then computed as,

$$JamPN^J_{recv,\Delta t_{opt}}(m) = \sum_{i=1}^{i=K} \frac{1}{K} PN^{n_i}_{recv,\Delta t_{opt}}(m) \qquad (6)$$

Dealing with Synchronization Errors. Sampling offset correction only corrects sub-sample synchronization errors. However, in order to cancel perfectly, the receiver must know the exact sample at which it started receiving the jamming sequence. Each jamming sequence is preceded by a chirp preamble (as described in Section 5) that helps the receiver synchronize itself to the jamming sequence. This method, however, by itself sometimes results in errors of up to a few samples. To exactly determine the offset, we compare the received PN signal at a few different sampling offsets with the precomputed $JamPN^J_{recv,\Delta t}$ for various values of Δt, to correctly identify both the start of PN sequence and the sampling offset.

Dealing with Multi-path. For dealing with multi-path, Dhwani takes the channel estimation based approach, but with input signal as the optimal delay-shifted version $JamPN^J_{recv,\Delta t_{opt}}$ of the *received* signal rather than the transmitted sequence. Working with the received signal greatly simplifies the channel estimation since most of the channel effects due to the electro-mechanical components of the speaker and microphone are already compensated for. Thus, Dhwani uses the first few samples of the (known) PN sequence received (512 in our implementation) and computes a FIR filter $H(t)$ that can transform $JamPN^J_{recv,\Delta t_{opt}}(m)$ (which only factors in the transformation due to static factors such as the speaker and microphone) into the actual received samples (which is also impacted by the dynamic multi-path environment). It then applies this filter to the rest of the $JamPN^J_{recv,\Delta t_{opt}}$, to transform it suitably, and then subtract it from rest of the received jamming signal.

7. PUTTING IT ALL TOGETHER

Figure 10 shows the overall architecture of Dhwani.

Transmitter Overview. As seen in Figure 10, at the transmitter, the message is first scrambled to ensure that bit errors result in the entire message being corrupted at the receiver. As discussed in Section 3, this is important in order to ensure that the eavesdropper cannot benefit from extracting parts of the message that are error free. The OFDM radio (described in Section 5) then transmits the scrambled message over the air using the speaker. The jamming detector continuously monitors the ambient jamming level to ensure that there is "enough" jamming to prevent any eavesdropper from receiving the message. Upon detecting a drop in jamming levels below a "safe" threshold, it simply directs the OFDM transmitter to abort the transmission mid-way.

Receiver Overview. As depicted in Figure 10, the received signal first passes through the JamSecure module. The JamSecure module transmits the jamming signal over the speakers, while simultaneously performing SIC on the received signal, as described in Section 6. An additional function that the JamSecure module performs is to estimate the appropriate jamming power needed to simultaneously ensure that (a) an eavesdropper cannot decode the message, (b) the receiver, with the benefit of SIC, can decode the message, and (c) a concurrent transaction at a distance of 1.5m or greater is not interfered with. Requirement (c) just imposes an upper bound on the jamming power, as discussed in Section 4.3. This then leaves the task of balancing requirements (a) and (b). The estimation of jamming power for this purpose is performed with the help of the transmitter, as described in Section 7.2. The OFDM receiver then decodes the message, after which the descrambler retrieves the original message. Successful reception of a packet is indicated by a 24-bit CRC check.

In the rest of this section we shall describe two key components of the system that have not been described so far: the *scrambler/descrambler* and the *jamming power estimator*.

7.1 Scrambler-Descrambler

As noted in Section 3.5, Dhwani uses scrambling prior to encoding and modulation, to amplify the impact of bit errors, thereby rendering the received (scrambled) message unreadable and preventing any information leakage. While a special-purpose scrambler could be designed, we simply re-purpose the widely-available and highly-efficient Advanced Encryption Standard (AES) [3] scheme. Whereas AES is typically used for encryption, with a private key that is kept secret, we use it with a well-known key, since our objective is to achieve the desired error propagation, not secrecy. The block-size in AES equals the key length — 128, 192, or 256 bits — which allows the possibility of sending a short NFC message (with padding) as a separately encrypted block or longer messages as multiple blocks. When the receiver, with the benefit of SIC, retrieves an error-free copy of the scrambled message, it is able to unscramble it with the well-known key. However, an eavesdropper, who typically suffers several bits of error will be unable to decode the message, even knowing the key.

7.2 Jamming Power Estimator

As described in Section 3, the message success rate at the receiver experiences a precipitous drop when the SNR falls below a certain threshold, SNR_{min}. For most modulation schemes this threshold can be determined experimentally. The jammer in JamSecure should ensure the following,

- before SIC, the SNR is low enough to guarantee several bit errors, and

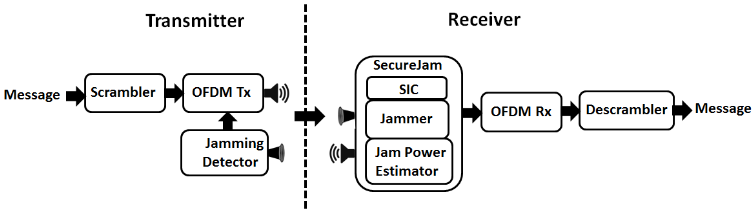

Figure 10: Dhwani System Overall

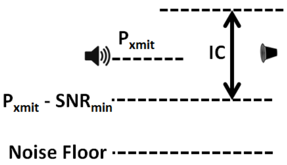

Figure 11: Jamming Power Estimation

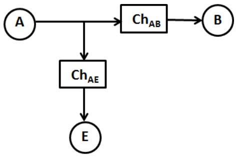

Figure 12: The Wyner wire-tap security model

- after SIC, the SNR is sufficient to decode the packet.

In other words, the jammer should guarantee that the SNR of the eavesdropper is below SNR_{min}, while the same for the receiver is greater than the SNR_{min}.

While, in our experiments we do not see a significant variation in the amount of cancellation achieved by SIC from location to location (as reported in Section 9), in general it may be influenced by the environment. To safeguard against a scenario where SIC may not perform as well as intended, for every transmission the amount of available SIC must be estimated. Further, since the received power levels of the transmission can vary across transmissions, this should also be estimated for each transmission. In order to accomplish this, a Dhwani transaction starts by the transmitter transmitting some known bits to the receiver (also the jammer). The jamming power estimator (in the receiver) uses this transmission to determine the transmit power P_{xmit} in dB of the sender[3]. Soon after this, the jammer begins transmitting its own PN sequence, performs SIC and determines the amount of cancellation ICdB that it can obtain. Based on its estimates of IC, P_{xmit} and SNR_{min}, the receiver then determines the P_{Jam} using the relation:

$$P_{jam} = P_{xmit} - SNR_{min} + IC. \qquad (7)$$

As seen from Figure 11, $P_{xmit} - SNR_{min}$dB is the maximum noise that the receiver can tolerate after SIC. Hence, JamSecure can afford to jam at a power $P_{xmit} - SNR_{min} + IC$. The eavesdropper will experience a SNR of $SNR_{min} - IC$. Consequently, as see from Figure 2 in Section 3, $IC > 5dB$ will ensure that the eavesdropper cannot receive the packet since message success rate will drop to almost 0%.

One interesting issue arises when P_{xmit} is so low that P_{xmit} $- SNR_{min}$dB is below the noise floor of the receiver. In this case, Dhwani's receiver itself is incapable of receiving this packet. Further, jamming below noise floor of the receiver is not meaningful. The eavesdropper however, can potentially have an ultra-low noise receiver that might have an SNR advantage and could still decode the packet. In this scenario, Dhwani simply jams with a high power, making sure that even the eavesdropper is unable to decode the packet successfully.

8. SECURITY ANALYSIS

In general, there are two approaches to achieving secure communication: *information-theoretic* and *cryptographic*. The information-theoretic approach is based on Shannon's information theory (as we elaborate on in this section), while the cryptographic approach (e.g., RSA) relies on the computational hardness of problems such as prime factorization. Our approach to security in Dhwani is information theoretic but in no way precludes the use of cryptographic techniques, which can always be implemented over and above Dhwani.

8.1 Information-Theoretic Security in Dhwani

The classical information-theoretic approach to security is *Shannon's one-time pad (OTP)* encryption [20]. Suppose device A needs to communicate a message M to device B securely in the presence of an eavesdropper E who can see all messages. Then, A and B first share a secret random string ω, of length equal to that of M, over an independent channel not accessible to E. This, ω, is one-time pad, i.e. it can be only used once and cannot be reused for any subsequent message. A then transmits the message $M' = M \oplus \omega$ to B. Knowing ω, B can extract M from M'. This approach is provably guaranteed to be secure.

Wyner's wiretap model. Given the obvious difficulties in Shannon's approach of setting up a shared secret between A and B, over an alternate channel, for every message, Wyner's wiretap model [21] takes a different approach, depicted in Figure 12. In this model, A and B communicate over a channel Ch_{AB} that is less noisy than the channel Ch_{AE} via which E eavesdrops. *The key proven result in the Wyner's wire-tap model is that, if Ch_{AE} is even slightly more noisy than Ch_{AB}, there exists an error correcting mechanism (e.g., error correcting codes) that A can use over Ch_{AB} that will appear identical to noise for E but can be perfectly decoded at B.* The channel that E listens on is called a *Wyner's wiretap channel*.

Dhwani's approach. Dhwani's approach falls primarily under the purview of Wyner's wiretap model, since the channel to the eavesdropper is noisier than that of the intended receiver due to jamming coupled with self-interference-cancellation. Consequently, the transmitter can use an error correcting mechanism (e.g., error correcting codes) that is sufficient for correcting errors in the less noisy channel, Ch_{AB}, but not so for the more noisy channel, Ch_{AE}, for the eavesdropper.

Further, since the jamming sequence is generated pseudo-randomly for each transaction and never reused, it can be viewed as a random string for one-time-pad encryption [4]. However, unlike Shannon's OTP, Dhwani does not apply the OTP encryption at the transmitter itself and may be vulnerable to certain attacks such as those based on shielding and directional antennas, which undermine the Wyner wiretap assumption.

8.2 Security Attacks on Dhwani

As noted in Section 3.1, Dhwani seeks to defend against both passive and active attacks on a pair of proximate, communicating nodes, which are assumed to be trustworthy. In this section, we discuss various security attacks on Dhwani, assuming that node A intends to transmit a message M securely to B, while C is a malicious node.

Man-in-the-middle and replay attacks. When A transmits M to B, a co-located eavesdropper E can receive it and try to transmit a

[3] Note that P_{xmit} is really the sender's power received at the jammer.

[4] While Shannon's OTP encryption and Wyner's wire-tap model analyse the security of binary channels, these results hold even for analogue signals since they can be translated into corresponding binary message through demodulation.

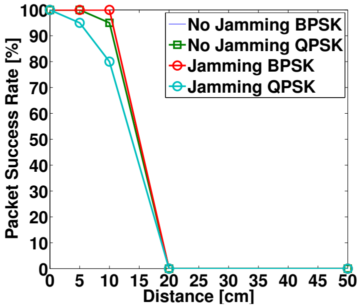

Figure 13: Communication Range of Dhwani

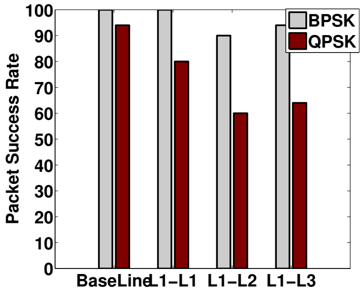

Figure 14: Packet Success Rates of Dhwani at various locations

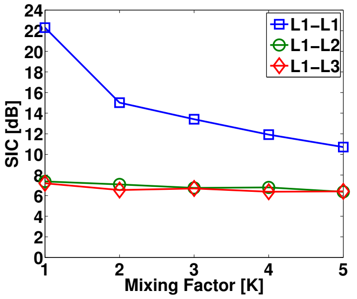

Figure 15: SIC achieved in JamSecure for increasing values of K

modified (or unmodified) version to someone else, pretending to be A. However, since E has no way of receiving any meaningful data, given the jamming from B, these attacks will remain ineffective.

DOS attacks A co-located device E could transmit its own jamming signal, to disallow meaningful communication between A and B. While E may succeed in disrupting communication between A and B, there will be no loss of security since E cannot recover the data transmitted by A.

Placement attacks. This attack is based on the presumption that there might be vantage points where B's jamming is not as effective and so E could recover A's transmission. Consider three devices a sender A, a receiver B and an eavesdropper E located as depicted in Figure 18. Suppose that the distance between A and B is d and that between A and E is D. The received acoustic power typically decays with distance x as $x^{-\gamma}$, where $\gamma > 2$. The SNR at the eavesdropper is thus given by,

$$SNR = \frac{S}{J}\left[\frac{D^2 + d^2 - 2Dd\cos\theta}{D^2}\right]^{\frac{\gamma}{2}} \tag{8}$$

where J is the jammer's power and S the sender's power at a unit distance. From Eqn 8 is is clear that SNR decreases monotonically with increasing D for any given θ and in fact the maximum SNR occurs at $D = 0$. Thus, the most advantageous position for the eavesdropper is to be co-located with the sender. As described in Section 7, the sender transmits only upon ensuring there is enough jamming to ensure that it cannot decode its own transmission. Since no eavesdropper can enjoy at better SNR than the sender, it follows that an eavesdropper cannot decode the message either.

In the above argument, we do not consider the effects of multipath and near-field acoustic power decay. For instance, in theory, it is possible that at certain locations, the jamming signals arriving along multiple paths may all interfere destructively. At such locations, the eavesdropper might enjoy an SNR high enough to enable decoding. While it is hard to claim that such scenarios will never occur, in our tests we could not find any such vantage points, as discussed in Section 9.5.

Stopping Attacks. This attack arises specifically because of Dhwani's reliance on the receiver, B, to jam A's transmission. If B were somehow disabled, then E could receive M in the clear. In fact, as it disables B, E can start emitting its own jamming signal, to keep A in the dark about B's disablement. However, as discussed in Section 3.1, in Dhwani's security model, a deliberate attack by E to disable B is not within scope. However, if B were to fail accidentally (e.g., lose power), A would detect that the jamming has ceased and stop transmitting immediately, as noted in Section 7.

Directional reception and shielding attacks. In these attacks, E uses a highly directional microphone (say using an microphone array formed by a coordinated set of attacker nodes) that is aimed at the speaker of A to boost the signal from A relative to the jamming noise from B, or alternatively, uses physical shielding aimed at B to reduce its jamming noise relative to the signal from A. The net effect in either case is an improvement in the signal-to-noise ratio, increasing E's chances of decoding A's transmission despite the jamming by B. While such attacks are possible in theory, these would be extremely difficult to mount in practice because of the close proximity of A and B, with the typical separation between them being a few cm. For example, the attacker has to be able to focus the directional microphone (or beamform) into a narrow region of only a few cm in order to selectively avoid the jammer. Also, since sound travels freely around obstacles, it is not feasible to selectively shield the jamming noise emanating from B, short of placing the shield right next to B (cloaking B's speaker), which again is hard to do undetected.

9. RESULTS

In this section, we present an evaluation of Dhwani and quantify several performance aspects such as its range, efficacy of self interference cancellation, and achieved packet success rates for different modulation schemes.

9.1 Operating Range of Dhwani

For Dhwani to be suitable for NFC, the ability to communicate must be limited to a very small range. As discussed in Section 3, each transmitter was pre-configured based on measurements to be limited to a range of 10cm. In these experiments, we evaluate how sharply the drop off in the communication range of Dhwani is. To answer this, we measured the Packet Success Rate (PSR) for Dhwani by transmitting 100 packets between a Samsung Galaxy S2 and a HP mini, at each of several different distances separating the devices. Figure 13 shows the PSR as a function of distance, with and without jamming, for two different modulations BPSK and QPSK. The best case SIC was used in the case with jamming (see Section 9.2 below). As seen from Figure 13, the PSR sharply falls to 0 at a distance of about 20cm, even in the absence of any jamming. So the communication range is indeed small.

Location	L1	L2	L3	L4
L1 (small room)	22.3	6.9	7.2	6.7
L2 (large room)	7.3	18.3	7.4	7.9
L3 (open pantry)	7.2	6.5	24.1	9.5
L4 (cafeteria)	7.4	8.7	8.3	12.9

Table 1: SIC (in dB) obtained at various locations.

9.2 Effectiveness of SIC

In this experiment we evaluate the effectiveness of Dhwani's SIC. As described in Section 6, Dhwani generates PN sequences during training and uses the recorded versions of these sequences to perform SIC. If the devices are trained at one "training" location and then taken to another "test" location, while the channel characteristics of the speaker and microphone do not change, the multi-path environment may change. To evaluate the impact of this change, we tested Dhwani in four different locations – a small room, a large room, an open pantry area and a cafeteria, which we shall refer to as L1, L2, L3, and L4, respectively. At each of the locations, Dhwani was trained by generating a library of PN sequences, and then tested by performing SIC at all other locations.

In Table 1 the value of i^{th} row and j^{th} column is the amount of SIC obtained when the PN sequence was generated at L_i was used to cancel at location L_j, the results being averaged over 10 trials. As seen from Table 1, the cancellation is close to 20dB (i.e., only $\frac{1}{100}$ of the jamming signal remains) whenever the testing and training locations are the same. This is what we expect to be the case when the receiver is a fixed installation, such as a POS terminal. Further, unsurprisingly, when the training and test locations are non-matching, the effectiveness of SIC degenerates to 6-7dB. This quick degeneration of SIC performance with even a limited amount of multipath shows drop from 20 to 6-7 indicates that even though multi-path is not the significant reason for frequency selectivity of the acoustic channel it does have significant impact on SIC. However, per the discussion in Section 3.5, even this reduced amount of SIC is sufficient for secure communication.

9.3 Impact of Jamming on Packet Success Rate

To evaluate the impact of jamming, we measured packet success rates for 256-bit packets at various locations with SIC based on the recorded PN sequences from various locations. Figure 14 depicts the packet success rates when the initial configuration was performed at L1 and Dhwani was tested at locations L1-L3. The packet success rate is measured by transmitting 100 packets and counting the number of packets whose CRC did not fail. In order to establish a baseline, we first measured the PSR without any jamming, which is referred to as the *Baseline* in Figure 14. As seen from the figure, in the absence of jamming, for BPSK modulation, the packet success rate is 100% while it is about 95% for QPSK modulation. In the presence of jamming and SIC with the configuration and testing locations both being L1, BPSK still gives a 100% packet success rate while it reduces to about 80% for QPSK. When Dhwani is tested at other locations, the PSR is around 90% for BPSK and 60-70% for QPSK. Note that in the event of a packet loss, retransmissions can be used to recover the packet. However, our current implementation does not have any retransmissions. A key result *in all these experiments was that in the presence of jamming, zero out of 100 packets were received successfully at each of the locations when we did not apply any SIC (as would be the situation of the eavesdropper).* This corresponds to the scenario where the eavesdropper is exactly co-located at the receiver.

9.4 Performance of JamSecure with increasing Mixing Factor

In the previous results we demonstrated the effectiveness of SIC when jamming was done using a single PN sequence Mixing Factor K = 1 in Eqn 6. As discussed in Section 6, when receivers are located at fixed installations, an eavesdropper may have the advantage of time to learn the library of PN sequences. The mixing factor $K > 1$ can then be used to thwart the eavesdropper from learning the library since the eavesdropper's search space increases exponentially as K. Clearly as K is increased, the performance of SIC is expected to degrade. This is because, the estimation errors of each of the constituent PN sequences add up and result in larger errors.

Figure 15 depicts the SIC achieved by JamSecure for various values of K — when the training and testing locations were the same (L1) and when the training and testing locations were different (trained in L1 and tested in L2,L3). As seen from Figure 15, when the training and testing locations are the same, as expected SIC does degenerate as K as it increases from 1 to 5, however, even at $K = 5$ the SIC is as high as 10dB. When training and testing locations are different (L1-L2 and L1-L3) the achieved SIC only decreases slightly as K increased from 1 to 5 and is typically between (6-7dB). Figure 16 depicts the Packet Success Rate (PSR) corresponding to the scenarios in Figure 15 for BPSK modulation. As depicted in Figure 16 the PSR is almost 100% when the receiver is trained and tested in the same location and over 90% for values of K between 1 and 5.

9.5 Multipath Effects on Jamming

To answer the question of whether there are special vantage points where jamming is not very effective and so the eavesdropper could enjoy a high SNR, we conducted experiments using the topology depicted in Figure 18. We placed an eavesdropper E at various locations around the sender A, and measured the received SNR in the presence of jamming from the receiver B. Figure 17 shows the received SNR at placement locations centred around the sender at three different distances D = 2,5 and 10cm and θ at 45^o intervals. As seen in Figure 17, at all these locations, while there are variations in SNR, the observed SNR is typically less than 0dB, indicating that there is no vantage point where an eavesdropper may be placed to achieve successful reception.

10. RELATED WORK

We have presented background and related work in the context of NFC (Section 2) and SIC (Section 6.1). We now discuss prior work on phone based acoustic communication and physical-layer security.

10.1 Acoustic Communication

[17] provides an extensive review of acoustic communication as a wireless technology for short and long distance communication. Existing techniques use On Off Keying (OOK) modulation and achieve up to 270bps at short distance of under 30cm. Our current implementation of Dhwani uses OFDM and achieves 5x these data rates.

Wimbeep [1] and Zoosh [2] offer acoustic communication systems targeting location-based advertising and mobile payments. While the technical details of these start-up offerings are not available, the description of Zoosh at [8] indicates that it operates in the ultrasonic band (beyond 20 KHz), offers a low bit rate of 300 bps (presumably because of the poor speaker and microphone characteristics beyond 20 KHz), and limits the communication range to 15cm to provide security. We believe that security based on range-limitation alone is inherently risky since the eavesdropper could use an ultra-sensitive microphone to boost the effective reception range. Hence Dhwani's emphasis on information-theoretic security at the physical layer.

10.2 Physical-Layer Security

The related work closest to Dhwani is IMDShield [12], which aims to secure communication to and from implantable medical devices (IMDs). IMDs are not amenable to using cryptographic

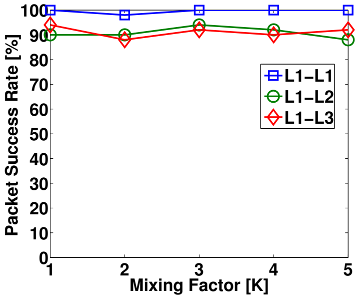

Figure 16: Packet Success Rate With Increasing Mixing Factor for BPSK modulation

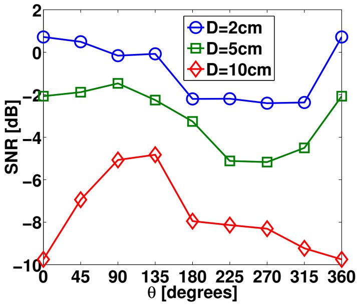

Figure 17: SNR for the eavesdropper at various locations around the sender

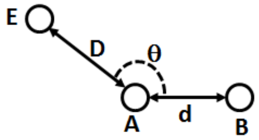

Figure 18: Placement attack analysis

techniques due to limited device memory and the need for immediate access in critical scenarios. IMDShield is a base-station that attempts to provide secure communication to and from the IMDs without requiring any alteration to the devices themselves. IMDShield continuously listens to IMD transmissions and transmits a jamming signal to secure it from eavesdroppers. Similar to Dhwani, the IMD base-station can perform self-interference cancellation and extract the message transmitted from the IMD. It then relays this message securely to the intended receiver using suitable encryption mechanisms. Similarly, in order to disallow malicious devices from reprogramming the IMD, the IMDSheild upon detecting a spurious transmission actively jams it and prevents the IMD from being programmed. Dhwani's novelty compared this work lies in (a) it being a software-only solution with no additional hardware, and (b) the JamSecure technique, which uses a pre-computed library based approach to jamming and SIC (Sec 6.2).

In Radio Telepathy [19], the authors propose a scheme where every pair of nodes can agree on a cryptographic key without actually performing a key exchange. The key idea is that since the wireless channel between a pair of nodes is symmetric, a common key can be extracted independently at each node from the channel response characteristics from a single transmission. In [15], the authors explore the practical limitations of extracting viable cryptographic keys using channel response information. In [18] the authors propose attack cancellation – a technique where sensor nodes in a sensor network collaboratively jam fake transmissions to defend against battery depletion attacks.

11. CONCLUSION

In this paper, we have presented Dhwani, a software-only acoustic NFC system that is accessible to the large base of existing mobile devices. The design of Dhwani is informed by our characterization of acoustic hardware and environment, and includes several novel elements. Chief among these is the receiver-based, self-jamming technique called JamSecure, which provides information-theoretic, physical-layer security. Our experimental evaluation point to the suitability of Dhwani for secure NFC communication.

Acknowledgements

We thank the anonymous reviewers and our shepherd, Shyamnath Gollakota, for their feedback.

12. REFERENCES

[1] Wimbeep. https://sites.google.com/site/wimbeep/.

[2] Zoosh. http://www.naratte.com/.

[3] Advanced Encryption Standard (AES), Nov 2001. U.S. Federal Information Processing Standard Publication 197, http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf.

[4] Near Field Communication Interface and Protocol (NFCIP-1), Dec 2004. ECMA-340 Standard (2nd Edition), http://www.ecma-international.org/publications/standards/Ecma-340.htm.

[5] NFC-SEC Whitepaper, Dec 2008. http://www.ecma-international.org/activities/Communications/tc47-2008-089.pdf.

[6] NFC-SEC-01: NFC-SEC Cryptography Standard using ECDH and AES, Jun 2010. ECMA-386 Standard (2nd Edition), http://www.ecma-international.org/publications/standards/Ecma-386.htm.

[7] NFC-SEC: NFCIP-1 Security Services and Protocol, Jun 2010. ECMA-385 Standard (2nd Edition), http://www.ecma-international.org/publications/standards/Ecma-385.htm.

[8] Start-Up Naratte Launches Novel Ultrasonic Near-Field Communications Solution, Jul 2011. http://www.bdti.com/InsideDSP/2011/07/28/Naratte.

[9] 86% of POS terminals in North America will accept NFC payments by 2017, Jun 2012. http://www.nfcworld.com/2012/06/07/316112/berg-86-percent-of-pos-terminals-in-north-america-will-accept-nfc-payments-by-2017/.

[10] At Villanova University, NFC Technology Being Tested, Mar 2012. http://www.todaysfacilitymanager.com/2012/03/at-villanova-university-nfc-technology-being-tested.

[11] NFC specialist Tapit to raise US$8m for international expansion, Jul 2012. http://www.nfcworld.com/2012/07/26/317057/nfc-specialist-tapit-to-raise-us8m-for-international-expansion/.

[12] S. Gollakota, H. Hassanieh, B. Ransford, D. Katabi, and K. Fu. They Can Hear Your Heartbeats: Non-Invasive Security for Implanted Medical Devices. In *SIGCOMM*, 2011.

[13] E. Haselsteiner and K. Breitfuss. Security in Near Field Communication (NFC). In *Workshop on RFID Security*, Jul 2006.

[14] M. Jain, J. I. Choi, T. Kim, D. Bharadia, K. Srinivasan, S. Seth, P. Levis, S. Katti, and P. Sinha. Practical, Real-time, Full Duplex Wireless. In *Mobicom*, 2011.

[15] S. Jana, S. Premnath, M. Clark, S. Kasera, N. Patwari, and S. Krishnamurthy. On the Effectiveness of Secret Key Extraction from Wireless Signal Strength in Real Environments. In *Mobicom*, 2009.

[16] H. Kortvedt and S. Mjolsnes. Eavesdropping Near Field Communication. In *The Norwegian Information Security Conference (NISK)*, Nov 2009.

[17] A. Madhavapeddy, D. Scott, A. Tse, and R. Sharp. Audio Networking: The Forgotten Wireless Technology. *IEEE Pervasive Computing*, 2005.

[18] I. Martinovic, P. Pichota, and J. B. Schmitt. Jamming for Good: A Fresh Approach to Authentic Communication in WSNs. In *WiSec*, 2009.

[19] S. Mathur, N. M, C. Ye, and A. Reznik. Radio-telepathy: extracting a secret key from an unauthenticated wireless channel. In *Mobicom*, 2008.

[20] C. E. Shannon. Communication Theory of Secrecy Systems. *Bell Systems Technical Journal*, 28, Oct 1949.

[21] A. Wyner. The Wire-Tap Channel. *Bell Systems Technical Journal*, 54, 1974.