Secret Agent Radio: Covert communication through dirty constellations

Aveek Dutta, Dola Saha, Dirk Grunwald and Douglas Sicker

University of Colorado Boulder
Boulder, CO 80309-0430 USA
{Aveek.Dutta, Dola Saha, Dirk.Grunwald, Douglas.Sicker}@colorado.edu

Abstract. In this paper we propose a novel approach to implement high capacity, covert channel by encoding covert information in the physical layer of common wireless communication protocols. We call our technique Dirty Constellation because we hide the covert messages within a "dirty" constellation that mimics noise commonly imposed by hardware imperfections and channel conditions. The cover traffic in this method is the baseband modulation constellation. We leverage the variability in the wireless channel and hardware conditions to encode the covert channel. Packet sharing techniques and pre-distortion of the modulated symbols of a decoy packet allows the transmission of a secondary covert message while making it statistically undetectable to an adversary. We demonstrate the technique by implementing it in hardware, on top of an 802.11a/g PHY layer, using a software defined radio and analyze the undetectability of the scheme through a variety of common radio measurements and statistical tests.

1 Introduction

There are many times when communication needs to be secure. Common and obvious examples include providing security for electronic commerce or privacy for personal matters. At other times, communication must also be *covert*, or undetectable which has a *low probability of intercept* (LPI) or a *low probability of detection* (LPD). LPD communication mechanisms are useful when the very act of communication can raise concerns, such as communication during war-time or during surveillance. Usually it is difficult to detect the receiver of communication mechanisms that exploit the characteristics of radio propagation.

In this paper, we explore methods that provide LPD and LPI for high-bandwidth networks. Our method provides a high-bandwidth covert side-channel between multiple radios using a common wireless network, as indicated in Figure 1. The method is covert because the devices (laptops or smartphones) function as normal devices. Again, the devices "hide in plain sight". Rather than raising suspicions by exchanging encrypted messages with each other or some centralized server, they appear to be conducting normal network communication (browsing web pages, sending mail, streaming multimedia) when in reality, they are able to communicate undetected. The adversary will face great challenge in discovering the side channel because the covert channel is being transmitted by mobile nodes. Monitoring to locate such nodes would require

significant investment or infrastructure, such as monitoring in every coffee shop, bus or public venue where people may be near each other.

The technique uses a common, physical-layer protocol to mask the communication that takes advantage of the hardware imperfections present in commodity hardware, intrinsically noisy channel of wireless communication and receiver diversity. We have implemented this mechanism using software-defined radios, operating in 2.4GHz ISM band, but can also be easily extended to TV whitespaces. Our prototype uses an OFDM waveform. Most consumer electronic devices use OFDM waveforms for high-bandwidth networks (including DVB, DAB, WiFi, WiMAX and LTE), and there are some benefits in "hiding" in such a ubiquitous waveform.

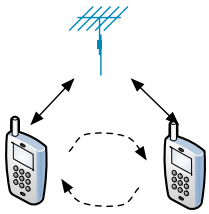
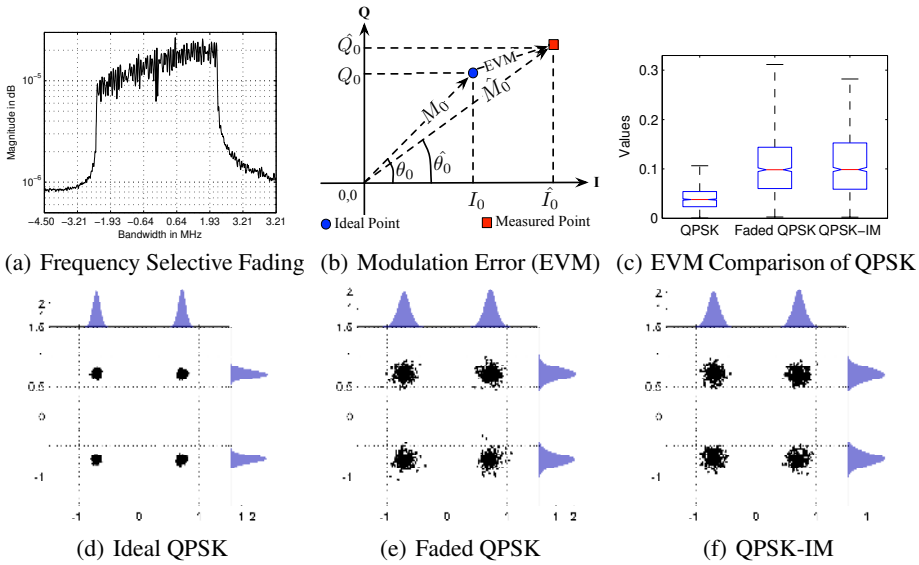


Fig. 1. Undetected Side-Channel Communication

Imperfections in off-the-shelf Network Interface Cards (NICs) [4], coupled with an additive random wireless channel cause the signal to degrade over time and distance. To mask our communication, we "pre-distort" the signal to mimic the normal imperfection of the hardware and Gaussian distortion arising from the channel. This distortion appears as noise to the unobservant receiver, be it the Wi-Fi access point or an adversary. However, a receiver aware of the presence of the signal and its encoding technique can decode the "noise" to reveal the hidden message.

Our motivation for hiding the data in physical layer (analog waveform domain) of common wired and wireless protocols are the following:

- *Hide in plain sight* - Using the physical properties of the transmission medium will allow the covert channel to resemble a common waveform, only distorted by channel noise, or transmitted by a NIC with imperfections.
- *Access to covert channel* - Since the covert channel uses the signal waveform, an adversary is easily abstracted from the covert channel, as opposed to other packet level techniques using higher layers [11]. In our method, the bits of the cover packet are not altered and hence the presence of the covert message is not detected at higher layers, or more specifically in digital domain.
- *Sample collection* - The ubiquitous nature of wireless devices and their localized transmission make it difficult to detect the presence of a covert channel. As opposed to digital contents on the Internet (music, picture, video), which can be accessed from one physical location, acquiring signal waveforms requires hauling expensive, bulky equipment (signal analyzers) to every possible hotspot.
- *Search complexity* - A $500byte$ packet, modulated with QPSK-$1/2$ rate coding, results in $\approx 19KB$ (calculation omitted due to space constraints) of I/Q information. This increases the search space by ≈ 38 times, compared to packet level analysis of a covert channel.
- *Statistically Undetectable* - In higher layer techniques, an adversary can search the header fields (known as unused fields) of a packet stream and find the covert channel [3], whereas in physical layer, the adversary needs to perform several statistical tests on the I/Q samples, which are already tainted by time varying channel noise.



(a) Frequency Selective Fading (b) Modulation Error (EVM) (c) EVM Comparison of QPSK

(d) Ideal QPSK (e) Faded QPSK (f) QPSK-IM

Fig. 2. Characterizing channel and hardware impairments with three waveforms: ideal QPSK, faded QPSK and "impaired," QPSK-IM. The QPSK-IM signal is indistinguishable from QPSK-faded signal using statistical measures.

– *Capacity* - Compared to conventional techniques using higher layers, where only a few unused bits of any header field of a packet is used, our technique can easily utilize 10% of the cover signal to transmit covert messages.

These advantages coupled with relative ease of implementation using now popularized software defined radio, makes this technique extremely useful in providing high capacity covert channels.

2 Characterizing OFDM Signals

Signal quality in wireless channel depends primarily on two factors: channel impairments and hardware impairments. Channel impairments typically range from additive white noise to frequency selective fading and/or hidden terminal and Doppler shifts, which degrade signal properties in time and frequency domain. Figure 2(a) plots the spectrum for an OFDM waveform from a bench measurement that is skewed because of a frequency selective fading in the left-most subcarriers. Similarly, impairments due to various non-linearities in the transceiver pipeline are often reflected in the signal characteristics as well. Since these types of impairments are hardly deterministic, estimating the errors and compensating for them is a non-trivial task.

Signal-to-Noise Ratio (SNR) is a widely used metric, often measured in the time or frequency domain using averaged power measurements. A simple interpretation of the SNR is *"the higher the SNR, the higher the probability that the information can be extracted with acceptable error performance"*. However, high spatial-decorrelation of the wireless channel may render portions of the OFDM signal undecodable even

though a high "average" SNR indicates otherwise. Figure 2(a) is an example of an OFDM spectrum of an ongoing communication that has an average SNR of 21dB but degraded in the frequency domain.

The Error Vector Magnitude (EVM), shown in Figure 2(b) is another metric that measures the deviation of the complex modulation vectors in the I/Q-plane from the ideal position. A bad channel leads to higher dispersion of these vectors and hence higher EVM, which affects the error performance as well. Modulation errors can also be introduced as imperfections in the transceiver hardware itself, which can cause the intended I/Q sample to be transmitted (or received) at a slight offset. In the IEEE 802.11a/g standard [9], this modulation error at the transmitter for a QPSK modulation is mandated to be no more than $10dB$ from an "ideal" I/Q mapping.

Figure 2(c) shows the distribution of EVM (in a boxplot) for three bench measurements of an OFDM waveform using QPSK modulation *where each of the transmissions have the same SNR*. The first measurement is based on an "ideal" transmission with low noise resulting in a low EVM with minimal variance, called ideal QPSK. The second measurement, faded QPSK, from a bench measurement with slightly different antenna orientation, has higher average EVM and wider variance. The difference between ideal QPSK and faded QPSK are due to multipath effects. The last measurement, termed the "impaired QPSK" or QPSK-IM signal, was recorded from a transmitter that predistorted the signal such that the average EVM is 10dB worse than the ideal. On the surface, the QPSK-IM signal appears to have similar properties to faded QPSK – both have higher average EVM and wider variance. Figures 2(d)-2(f) show the three constellations corresponding to the measurements described above. It is indeterminable whether the deterioration in the EVM is due to intentionally introduced noise at the transmitter, or due to imperfections in the hardware that is operating within tolerable limits, or is the result of poor channel quality.

From these examples, it is evident that impairments, whether in the channel or in the hardware, will cause statistical variation in the perceived value of the metrics and that the bounds on these metrics are only loosely defined and can only be formalized by various descriptive statistics and statistical tests.

3 Dirty Constellation

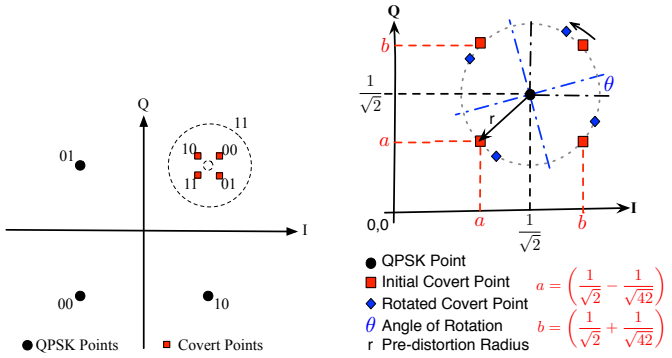
Our method relies on being able to embed one message in another in the wireless channel, but goes well beyond that to then insure that the covert message is undetectable. There are several ways to embed messages by encoding the constellation symbols using bits of two distinct messages [13, 7] but we use a simpler technique that uses existing modulation methods of OFDM.

Using a combination of adaptive modulation and efficient packet sharing using joint constellations we encode the covert channel. If a receiver is aware of our irregular mapping of bits, and it has sufficient SNR for that subcarrier, it is able to decode the covert message while to an uninformed user, the covert constellation points will be treated as random dispersed sample of a low-rate modulation, that reveals an innocuous message.

The key to such covert communication using the physical layer of an OFDM based wireless protocol are four fold: **1)** packets containing covert data must be indistinguish-

able from non-covert packets to all uninformed observers; **2)** the presence of any irregularity in the covert packets has to be kept hidden under rigorous *statistical tests* on the signal; **3)** the covert channel should be non-trivial to replicate, making it secure from spoofing and impersonation; and finally, **4)** it should have high capacity. In this paper we satisfy each of these requirements through a set of techniques.

Requirement 1: Identifying a Covert Channel: Our technique relies on encoding "cover packets" that are transmitted at a low rate (BPSK or QPSK) with supplemental information that can be decoded as an additional QPSK signal by an informed receiver. In the examples below, we use QPSK for both the cover and covert channel.



(a) Cover QPSK constellation (b) Covert constellation

Fig. 3. Encoding Dirty Constellation

In a QPSK encoding, the constellation points encode two bits of information as shown in Figure 3(a). To encode the covert channel, we deflect the placement of the QPSK points. This is similar to having a "covert QPSK" encoding with an origin around the ideal QPSK constellation points of the cover traffic. Figure 3(b) corresponds to the upper right quadrant of the cover QPSK constellation shown in Figure 3(a). To modulate a subcarrier carrying both the cover and covert message, first the cover constellation point (QPSK) is chosen (as per the cover message stream), specifying the quadrant, followed by re-mapping that point to one of the four "covert-QPSK" points around the "cover QPSK" point.

Clearly, the goal is to leave the cover message decodable by standard receivers. Only the covert receiver aware of the joint constellation will decode the subcarriers properly and extract the *two* covert bits to form the hidden packet. An adversary will decode at the base rate or the rate for cover message, as specified in the *signal symbol* of the packet; while the covert points will be treated as noisy points. The cover message could be intended for an access point (as part of a web browsing session) while the covert message can be overheard and decoded by a nearby radio. In this way we implement a covert channel while making it appear as completely innocuous to other users receiving the same transmission.

Requirement 2: Low Probability of Detection: How would an adversary detect such communication? As long as the packet can be decoded, a legacy receiver has no way of knowing how signals are being encoded at the core of the physical layer, because conventional packet decoding is performed by identifying the data rates embedded at

the beginning of the packet which will always contain the base rate (QPSK) information. However, adversaries using measurement equipment like vector-signal analyzers or software defined radios can extract the digital samples from the radio pipeline at different stages of the signal processing. Therefore, our ultimate goal is to provide very low probability of detection not only at the packet level but also at the signal level.

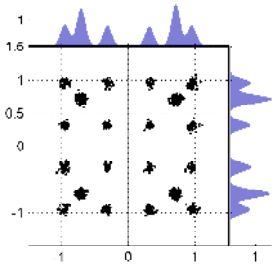


Fig. 4. Constellation without random pre-distortion of the QPSK points and using existing 16QAM points to map the joint covert constellations.

One simple form of analysis is to look at the equalized I/Q vectors of the jointly encoded packet. The presence of the covert constellation at regular interval will appear as distinct point clouds that will set themselves apart from the cover QPSK point cloud and will reveal the presence of the covert channel, as shown in Figure 4.

We solve this problem by changing the I/Q vectors of the covert transmitter in three steps:

Step1: We bring the covert constellation points closer to the ideal QPSK point and re-map the covert constellation points symmetrically around the QPSK points, with a mutual separation of $\frac{2}{\sqrt{42}}$, a distance equal to that of a 64QAM constellation, so that a covert receiver can operate within the operating range of a WiFi receiver.

Step2: We randomize the I/Q vectors of the covert QPSK points with a Gaussian distribution but limit their dispersion to a radius of $\sqrt{\frac{2}{42}}$ as shown in figure 3(b). We call this as the *pre-distortion circle*; pre-distortion of the QPSK signal at the transmitter ensures that the covert constellations are hidden in the cloud of a dispersed (noisy) QPSK point cloud. We introduce imperfections to the transmitted signal in such a way that the average EVM error is equal to or less than 10dB compared to the ideal QPSK constellation points, which is within the limits of hardware anomaly allowed in the IEEE 802.11 standard [9]. Thus, it cannot be ascertained with certainty if the EVM error is due to hardware impairments, channel impairments or intentionally injected distortion.

Step3: To accommodate a higher rate covert channel, *e.g.*, when 50% of the OFDM subcarriers are covert, then at high SNR there is always a finite probability that the covert constellations are visible. To have the covert symbols blend with the pre-distorted QPSK point cloud, the covert symbols are rotated along the circumference of the pre-distortion circle for every subcarrier that is mapped to a covert constellation as shown in Figure 3(b). The rotation is performed using a monotonically increasing angle θ; the transmitter and receiver both start with $\theta = 0^\circ$ at the start of the packet and increment θ for each covert subcarrier. In our implementation we use a 15° counter-clockwise rotation for the covert points.

These 3 steps allow us to hide the covert channel, even when an adversary has access to the I/Q samples of the packet. The adversary will interpret the point cloud as a noisy version of a valid (albeit noisy) QPSK constellation and would not suspect the presence of a covert communication. This compound constellation involving a covert channel hidden within a cover constellation is termed a *"Dirty Constellation"*. However, in order to avoid raising suspicion by any RF fingerprinting algorithms [4], a QPSK-IM

waveform should *always* be used for non-covert transmissions, to avoid sudden changes in the modulation characteristics.

Requirement 3&4: Security and Higher Efficiency: These requirements are considered as an enhancement to the basic scheme of Dirty Constellation. We have implemented 10%, 30% and 50% encoding of subcarriers, as shown in Figure 7, yielding up to $9Mbps$ datarate with QPSK modulation and $3/4$ encoding rate. Using higher modulation constellation, *e.g.*, 256-QAM, we can further increase the capacity of the covert channel by encoding more bits per subcarrier. Due to space constraints we leave this as future work. Finally, we discuss the security aspect in §7.

4 Dirty Constellation on SDR

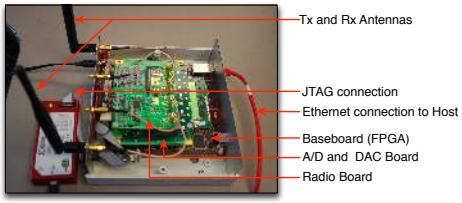
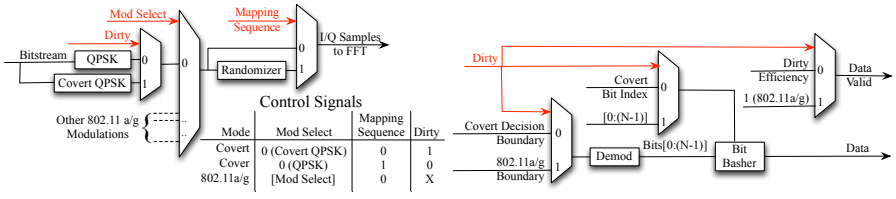


Fig. 5. SDR prototype using Virtex-V FPGA

Hiding a message in a randomized modulation constellation requires a programmable modulator and demodulator. Conventional radio 802.11 PHYs modulate all the subcarriers with one type of pre-defined modulation, For this scheme to work, we used a FPGA-based software defined radio platform based on our previous work [6, 5], as shown in figure 5, and modified the modulator and demodulator to program each subcarrier with different modulations, adding either noise or covert constellations. Figure 6(a) shows the functional diagram of the programmable modulator. The notable parameters in the design are the *dirty* bit and the *mapping sequence* bit which are used to select the appropriate mapping for covert joint constellations and randomize (Gaussian) the cover symbols to engulf the higher order modulation points. The cover and the covert bits are independently packetized as per the 802.11a/g specification and the covert joint symbols are formed by merging the bits of the two packets prior to sending it to the modulator. The merging of packets is performed in software and then fed to the hardware along with the control information to create the Dirty Constellation. The QPSK-IM constellation is generated by using the randomizer unit that emulates an overall modulation error of 10dB, by setting the *dirty* bit to '0' and *mapping sequence* to '1' for all subcarriers.

The decoder employs maximum likelihood decoding and uses pre-defined thresholds to decode the constellation. Figure 6(b) shows the functional diagram of the demodulator. First the covert receiver demodulates the signal using the covert decision boundaries, 64QAM in this case and then extracts the covert bits. Since all subcarriers

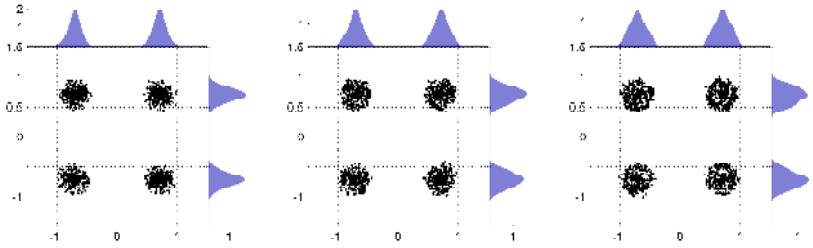


(a) Transmitter Pipeline (b) Receiver Pipeline

Fig. 6. Mod/Demodulator for Dirty Constellation

do not contain the hidden message, the receiver then uses the pre-assigned mapping sequence and its rotation information to filter out the covert subcarriers' information to form the covert packet.

Figure 7 shows an example of Dirty Constellation with varying frequency of the covert channel that has been transmitted by the SDR prototype and captured using a VSA. The I and Q histograms alongside the constellation shows the similarity of the distributions and that they are from the family of normal distributions.



(a) Frequency – 10% (b) Frequency – 30% (c) Frequency – 50%

Fig. 7. Examples of over-the-air transmission of Dirty Constellations with varying embedding frequency using the SDR prototype

5 Experiments and Measurements

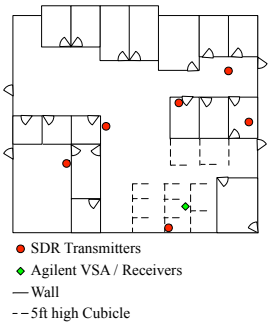


Fig. 8. Node placement

Using the hardware described in §4 as the transmitter, signal samples are collected in a lab/office environment. The transmitter nodes were placed as shown in the Figure 8. The signals were captured using a high-end Agilent vector signal analyzer (VSA) that provides the raw I/Q vectors of the packets transmitted by the SDR nodes. We record data from 6 locations, for ideal QPSK, QPSK-IM and Dirty Constellation with 10% covert channel efficiency. Each dataset contains measurement of 500 data packets of each type per transmit power level. The transmit power is varied in steps of 2.5dB such that the measured SNR at the VSA has a range of 7dB to 20dB. We have chosen this range because 7dB is the minimum SNR required to decode a QPSK packet with 98% packet reception rate. This has been empirically validated using bench measurements using our SDR transceivers. Likewise, 20dB was selected as the upper limit because the EVM doesn't decrease appreciably with higher SNR. After filtering out the required data range we find the average sample size is $10,000$ packets per type. We bin the packets by SNR in bins of size 1dB; each bin contains $500 - 800$ packets per SNR value. We perform all the statistical testing using this dataset which captures a wide range of SNR and channel conditions for all the type of modulations. In these measurements, the VSA is treated as both the covert receiver *and* a very aggressive adversary. As a covert receiver, the messages sent by the different transmitters can be received by the VSA receiver and the covert data can be extracted. As an adversary, the receiver has a high quality measurement device and also acts as the "most aggressive adversary" because it shares the same channel state as the receiver.

6 Analyzing Dirty Constellation

The core idea of testing a sample for adherence to a particular family of signals is performed by comparing test results with a known set of statistics for the same class. Therefore, the first step of the analysis process is to formalize the database of these statistics that characterizes an entire family of signals. In this paper, we intend to compare a Dirty Constellation with a QPSK waveform. We formulate the problem as a hypothesis test, with the null hypothesis:

$\mathcal{H}_0$: Given a random sample from a Dirty Constellation packet, it is statistically same as any other QPSK packet.

Whereas the alternative hypothesis is:

$\mathcal{H}_1$: Given a random sample from a Dirty Constellation packet, it can be statistically identified that it is not a QPSK packet.

In this section, we analyze whether the packets containing covert data can be distinguished from normal packets at the packet level or at the waveform level in the time and frequency domain. The test statistics of standard QPSK signals is lower bounded by the statistics of an "ideal QPSK" packet and upper bounded by a "QPSK-IM" packet. We used "QPSK-IM" packets to mimic a radio with hardware imperfections, but operating within the limits of IEEE 802.11 standard requirements. Each of these bounds have been empirically derived from the measurements collected as described in §5. If the Dirty Constellation sample is within the bounds set for that test then the null hypothesis is "not rejected", meaning that the Dirty Constellation packet is statically indistinguishable from any other QPSK transmission within the expanse of 802.11a/g transmissions using that test.

6.1 Packet Based Analysis

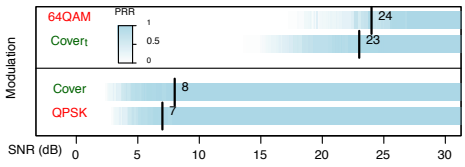


Fig. 9. Packet Reception Rate

Packet based analysis involves looking at parameters that can be extracted at the packet level, or in the digital domain, where there is no trace of the covert packet. To measure if the pre-distortion of the constellation effects the packet reception rate (for both the covert and the cover packet) we performed measurements over a one hop link between two SDR nodes over a wide range of SNR. Figure 9 shows the packet reception rate for the standard modulations used in 802.11a/g and also the SNR required by the intended receiver of the covert packet and the cover packet. The minimum SNR levels required for 98% packet reception rate is marked. For the cover packets, our mechanism is within 1dB of that required by standard 802.11a/g modulation. Given the stochastic nature of the wireless channel and high spatial de-correlation of the nodes, this difference is indistinguishable to an end user (the user would experience greater variance simply by moving their receiver a few inches). The covert receiver requires an SNR of 24dB, similar to the SNR needed to decode a 64QAM packet.

6.2 Signal Domain Analysis

A time varying signal is often characterized either by time-domain measurements (power envelope and peak to average power ratio) or by performing spectral measure-

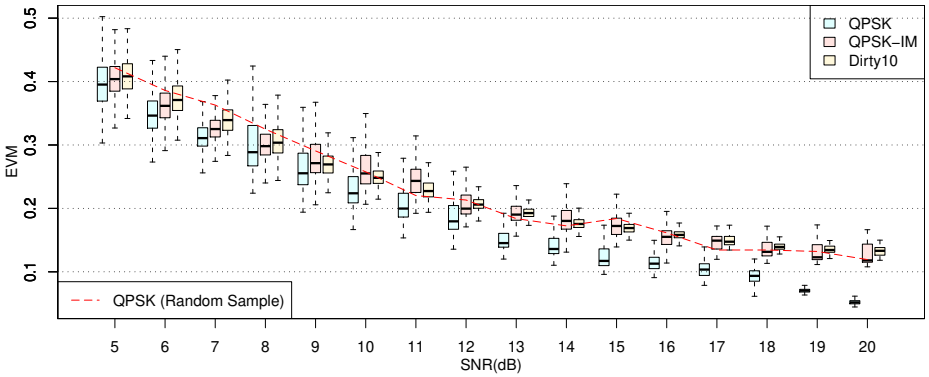


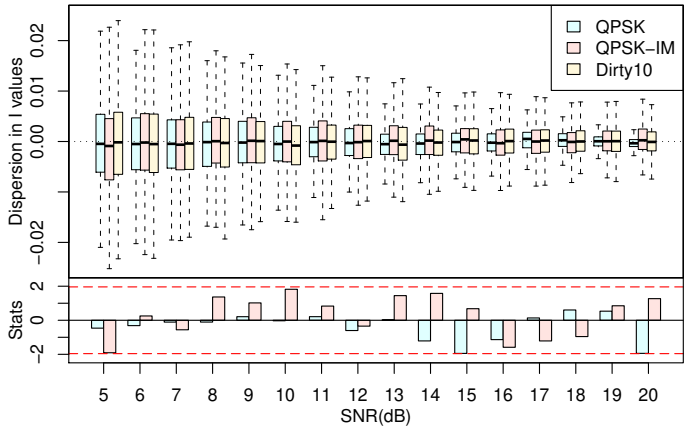
Fig. 10. Distribution of EVM. A faded ideal QPSK sample is also shown.

ments such as power spectral density, phase and magnitude distributions. Since OFDM encodes data in the frequency domain as coefficients of an inverse Fourier Transformation, a frequency domain analysis is of utmost importance and hence we conduct a set of frequency domain analysis, followed by tests in the time domain.

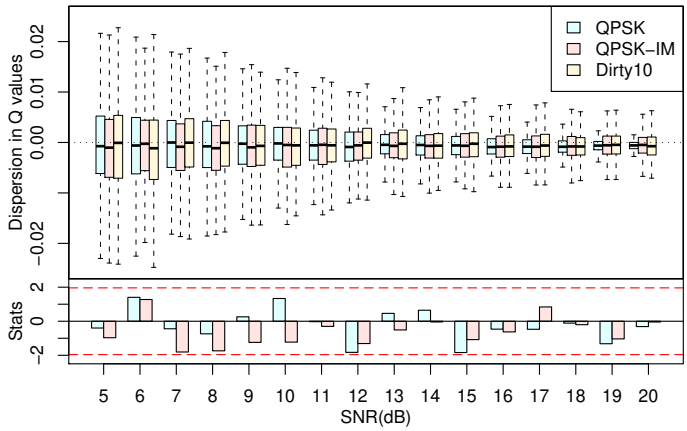
Frequency Domain Tests –

Test 1: EVM of Constellations: The real and imaginary vectors (I & Q) are available at the output of the Fourier transform unit. EVM is the absolute value of the dispersion of the I/Q-vector averaged over all OFDM symbols in a packet. Figure 10 shows EVM with varying SNR for the QPSK and QPSK-IM bounds and for the Dirty Constellation as well. The inter-quartile distances represents the spread of the I/Q vectors as they are degraded by channel noise. The EVM of the Dirty Constellation is distributed within the bounds set for QPSK making it statistically undetectable when compared with the empirical benchmarks. The plot also shows the average of EVM of a frequency faded random QPSK measurement, which emphasizes the non-deterministic effects of the channel that can push the envelope of the set bounds in either direction. That sample has the same parameters and configuration as the "ideal QPSK", but with the antenna moved by 2 inches. We expect the test statistic to be correlated with the variation in the bounds.

Test 2: Measure of I/Q Dispersion: The relative dispersion of the I/Q vectors result in a change in the position of the constellation point. Although all receivers employ channel equalization to compensate for the channel distortion, there are always residual errors that cause the points to violate their respective decision threshold leading to bit errors. Figures 11(a) and 11(b) show how the *deviation* from an ideal QPSK constellation is distributed within the dataset. Deviations in the the Dirty Constellation packets are within the bounds for most of the SNR values. To ascertain that the distributions are indeed similar and highly correlated, and that they are normally distributed about the ideal QPSK constellation, we perform a two sample t-test with the ideal QPSK packet and the QPSK-IM packet. The test statistics for all the SNR are found to be less than the critical value at the 0.05 significance level, as shown in the bottom part of figure 11. This also satisfies the test that the I/Q dispersion for all the three types are distributed in similar fashion and are from the family of normal distribution with statistically similar means.



(a) Dispersion in I vector



(b) Dispersion in Q vector

Fig. 11. Dispersion of I and Q vectors from ideal QPSK mapping. The distribution of the I/Q dispersion is verified with that of ideal QPSK and QPSK-IM using a two sample t-test.

Test 3: Phase and Magnitude Distribution: Often it is important to know how the phase and magnitude vary with the subcarrier index. Figure 13 shows a histogram of the subcarrier phases of all packets in the collected dataset at two SNR levels, low SNR (7dB) and high SNR (18dB). At low SNR the subcarriers undergo distortion over a wider range and so the phases have a wider distribution, while at high SNR the signal is closer to the ideal QPSK signal. However, in both the SNR levels, the phases from the Dirty Constellation packets are distributed similarly to the ideal QPSK and QPSK-IM. The four distinct peaks at multiples of 45° ascertain that Dirty Constellation preserves the phase properties of the QPSK constellation. Similarly, the magnitude distribution across the subcarriers show that the magnitude of the subcarriers in a packet encoded with Dirty Constellation are distributed within the bounds of QPSK waveforms, as shown in figure 12. It is also seen that there is a high degree of correlation among the subcarrier from the three types of packets: the same multipath affects all three transmissions. To show that the distributions are correlated we also show the quantile-quantile (QQ) plot for subcarrier magnitudes of the QPSK-IM and the Dirty

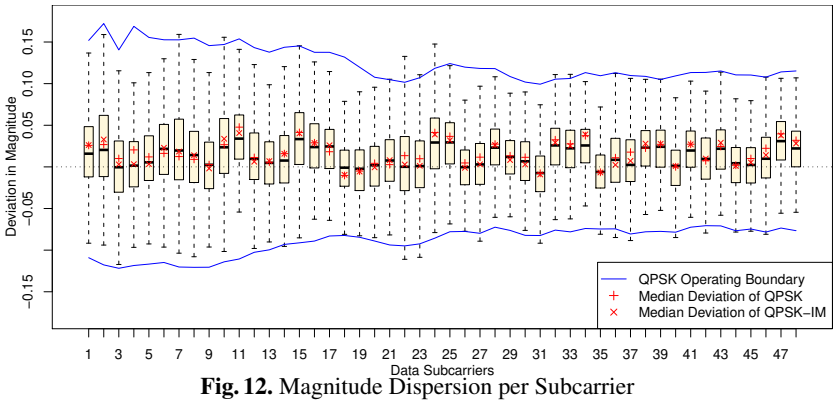


Fig. 12. Magnitude Dispersion per Subcarrier

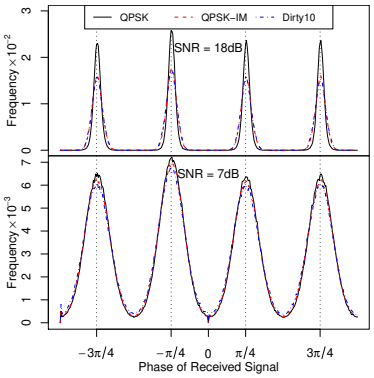
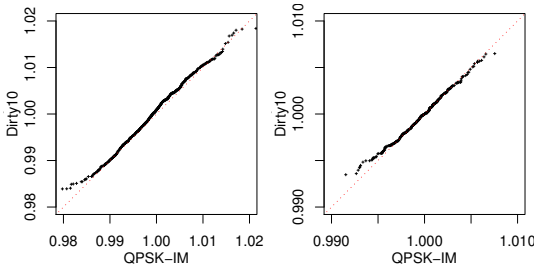


Fig. 13. Phase Distribution



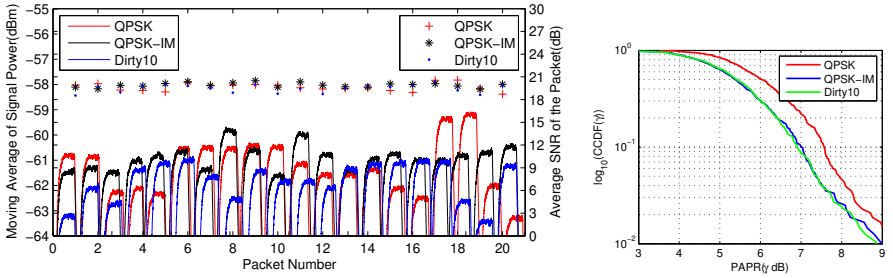
(a) At SNR = 7dB (b) At SNR = 18dB

Fig. 14. QQ-Plot of Magnitude

Constellation packets, as shown in figure 14. The linearity of the QQ plot indicates the signals have similar distributions.

Time Domain Tests –

Test 1: Temporal Variation of Average Signal Power: To test if the Dirty Constellation affects the signal power, we compare the temporal variation with that of a QPSK packet. In an experiment, 20 packets were captured using the VSA for all three types of packet at intervals of $\approx 500ms$. The average power is shown in Figure 15(a). The power envelope for the packets are randomly distributed even though the packets all have similar signal to noise ratios. Therefore, from this test we conclude that our method does not change the average signal power that is different from that of other QPSK packets.



(a) Moving average of signal power (b) CCDF of PAPR

Fig. 15. Time Domain Analysis

Test 2: Peak to Average Power Ratio (PAPR): OFDM can produce spurious increase in the peak power when the packet contains different types of modulations. PAPR is the measure of the spurious increase in power in the time domain. Figure 15(b) shows the complementary CDF (CCDF) of the PAPR for the three packet types. Research [2, 10] shows that the PAPR in 802.11a/g can vary over a wide range with various PAPR optimization techniques. The PAPR for Dirty Constellation falls within that range and follows closely with that of QPSK-IM. Hence it cannot be distinguished as an anomaly compared to the ideal QPSK transmission.

In this section we conducted tests that fail to reject the null hypothesis leading us to conclude that our method is statistically undetectable when compared to known waveforms that spans over a wide range of SNR. The analysis in frequency as well as time domain ensures the completeness of the testing. Thus, we conclude that our method can be successfully used as a covert channel that has very low probability of detection.

6.3 Exceptions

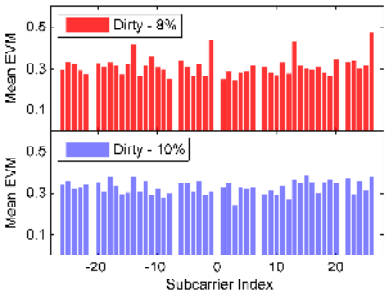


Fig. 16. Average EVM per subcarrier

In this section we provide examples of Dirty Constellation that *are* easily detectable, indicating that the methods and bit-mapping of the covert channel is non-trivial and requires careful analysis before adopting. One would guess that a lower embedding rate is better even though that results in a lower covert data rate. To see that this is not the case, we changed the embedding frequency to $\approx 8\%$ (1 in 12 subcarriers). Figure 16 shows the mean EVM of each data subcarrier of Dirty-8% compared to that of Dirty-10%. Since the Dirty-8% affects 1 in 12 subcarriers, a regular pattern is emphasized in the EVM of certain subcarriers. The mean EVM for Dirty-8% clearly shows that four out of 48 subcarriers has significantly higher EVM. On the contrary, Dirty-10% has a more even distribution of mean EVM in all of its subcarriers because 48 is not evenly divisible by 10.

7 Security

In §6.3 we discussed that mapping of covert channel is a non-trivial problem. This mapping sequence could be generated using a pseudo random number (PRN) sequence generator. Dirty Constellation employs two forms of sequence or pattern: the covert carrier mapping sequence and the angle of rotation for the covert constellation along the pre-distortion circle. While one PR sequence controls the embedding frequency, another specifies the rotation parameters, such as the angle of rotation "θ" for the covert constellation and the direction of rotation. The receiver needs to know which packets contain covert communication as well as the PRN's used to mix the covert message into the cover message. The frequency of covert messages can also be randomly varied without the need for additional coordination. The PRN used to intermix the covert message is synchronized with the receiver at the beginning of a transmission and can vary over time using an agreed-upon PRN based on *e.g.* the time of day. Any existing

encryption method (like AES, DES) can be used in each packet as an added measure to increase the security of the proposed method. However, due to space constraints, we do not analyze the details of the security aspects of this technique in this paper.

8 Related Work

Hiding information has been prevalent since ancient times; however hiding data in digital format is more a recent developments with the popularization of Computer Science. Much of the early work [12] in data hiding with low probability of detection and interception has been done by altering a few bits of the digital representation of an image [15], a sound [8] or video [16] files.

A relatively recent field of study called *network stenography* exploits the redundant fields present in various network protocols headers, like HTTP and TCP. Zander et. al. [17] provides a comprehensive survey of covert channels in computer network protocols. All of the methods detailed in the paper are confined to identifying anomalies or using the protocol properties at the application, transport or the data link layer. Also [11] proposes another scheme to hide data based on utilizing redundant fields in IPv4 header while [3] presents a practical analysis of covert channels in wireless LAN protocols at the transport layer. Information hiding at the application layer of a mobile telephony network has been discussed in [1]. These protocols depend on altering the data itself, which is susceptible to higher probability of interception, when the altered data is tested. Our procedure is significantly different from previous work in the sense that we modify the way of data transmission without altering the bits of any digitally transmitted data. In other words, higher layer stenography operates in the *digital* domain while our method operates in the *analog* domain.

Examples of covert channel implementation utilizing the physical layer are few and far between. A PHY layer based security scheme has been proposed in [14]. However, this method works only when more than one user is available to transmit stenographic packets to a common node. Also it relies on very tight synchronization between multiple transmitter and single receiver entity, which is not a practical assumption in real networks and will lead to erroneous formation of the joint constellations leading to degraded performance. Therefore, comparing to prior work, our method presents a more practical solution to implement covert channels at the PHY layer, while making it secure, high capacity, easily implementable and backward compatible.

9 Conclusion

In this paper, we proposed a technique to implement a covert channel at the physical layer of 802.11a/g wireless protocol. By hiding the covert channel within the perceived noise at the receiver, we can ensure high degree of undetectability. We have implemented the covert communication method using a SDR prototype and present results of a wide variety of statistical tests that confirms the low probability of detection of Dirty Constellation. Higher datarate, very low probability of detection coupled with easy implementation within existing protocol stacks make Dirty Constellation a very successful method to implement covert channels in wireless communication.

References

1. Agaian, S.S., Akopian, D., D'Souza, S.: Wireless steganography. p. 60740G. No. 1, SPIE (2006), http://link.aip.org/link/?PSI/6074/60740G/1
2. Aggarwal, A., Meng, T.: Minimizing the peak-to-average power ratio of ofdm signals using convex optimization. vol. 54, pp. 3099 –3110 (2006)
3. Ahsan, K., Kundur, D.: Practical data hiding in TCP/IP. In: Proc. Workshop on Multimedia Security at ACM Multimedia '02. French Riviera (December 2002)
4. Brik, V., Banerjee, S., Gruteser, M., Oh, S.: Wireless device identification with radiometric signatures. In: Proceedings of the 14th ACM international conference on Mobile computing and networking. pp. 116–127. MobiCom '08, ACM, New York, NY, USA (2008), http://doi.acm.org/10.1145/1409944.1409959
5. Dutta, A., Fifield, J., Schelle, G., Grunwald, D., Sicker, D.: An intelligent physical layer for cognitive radio networks. In: WICON '08: Proceedings of the 4th international conference on Wireless internet (2008)
6. Fifield, J., Kasemir, P., Grunwald, D., Sicker, D.: Experiences with a platform for frequency agile techniques. In: DYSPAN (2007)
7. Ganti, R., Gong, Z., Haenggi, M., Lee, C., Srinivasa, S., Tisza, D., Vanka, S., Vizi, P.: Implementation and experimental results of superposition coding on software radio. In: Communications (ICC), 2010 IEEE International Conference on. pp. 1 –5 (May 2010)
8. Gruhl, D., Bender, W., Lu, A.: Echo hiding. In: Information Hiding: 1st Int. Workshop (Lecture Notes in Computer Science) (1996)
9. IEEE Computer Society : LAN/MAN Standards Committee: Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, http://standards.ieee.org/getieee802/download/802.11-2007.pdf
10. Jayalath, A., Tellambura, C.: Peak-to-average power ratio of ieee 802.11 a phy layer signals. In: Wysocki, T.A., Darnell, M., Honary, B. (eds.) Advanced Signal Processing for Communication Systems. The International Series in Engineering and Computer Science, vol. 703, pp. 83–96. Springer US (2002), http://dx.doi.org/10.1007/0-306-47791-2_7
11. Krätzer, C., Dittmann, J., Lang, A., Kühne, T.: Wlan steganography: a first practical review. In: MM&S;Sec '06: Proceedings of the 8th workshop on Multimedia and security. pp. 17–22. ACM, New York, NY, USA (2006)
12. Petitcolas, F., Anderson, R., Kuhn, M.: Information hiding-a survey. Proceedings of the IEEE 87(7), 1062 –1078 (Jul 1999)
13. Shacham, N.: Multipoint communication by hierarchically encoded data. In: INFOCOM '92. Eleventh Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE. pp. 2107–2114 vol.3 (May 1992)
14. Tsouri, G.R., Wulich, D.: Securing ofdm over wireless time-varying channels using subcarrier overloading with joint signal constellations. EURASIP J. Wirel. Commun. Netw. 2009, 2–2 (2009)
15. Wu, M., Tang, E., Lin, B.: Data hiding in digital binary image. In: Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on. vol. 1, pp. 393 –396 vol.1 (2000)
16. Xu, C., Ping, X., Zhang, T.: Steganography in compressed video stream. In: Innovative Computing, Information and Control, 2006. ICICIC '06. First International Conference on. vol. 1, pp. 269 –272 (2006)
17. Zander, S., Armitage, G., Branch, P.: A survey of covert channels and countermeasures in computer network protocols. Communications Surveys Tutorials, IEEE 9(3), 44 –57 (third 2007)