Low-Cost Mitigation of Privacy Loss due to Radiometric Identification[*]

Jason J. Haas
Computer Systems and
Technologies Business Area
Sandia National Laboratories
Albuquerque, New Mexico,
U.S.A.
jjhaas@sandia.gov

Yih-Chun Hu
Dept. of Electrical and
Computer Engineering
University of Illinois at
Urbana-Champaign
1406 West Green Street
Urbana, Illinois, U.S.A.
yihchun@illinois.edu

Nicola Laurenti
Dept. of Information
Engineering
University of Padova
Padova, Italy
nil@dei.unipd.it

ABSTRACT

Recently, there has been much interest in using radiometric identification (also known as wireless fingerprinting) for the purposes of authentication. Previous work has shown that using radiometric identification can discriminate among devices with a high degree of accuracy when simultaneously using multiple radiometric characteristics.

Additionally, researchers have noted the potential for wireless fingerprinting to be used for more devious purposes, specifically that of privacy invasion or compromise. In fact, any such radiometric characteristic that is useful for authentication is useful for privacy compromise. To date, there has not been any proposal of how to mitigate such privacy loss for many of these radiometric characteristics, and specifically no such proposal for how to mitigate such privacy loss in a low-cost manner.

In this paper, we investigate some limits of an attacker's ability to compromise privacy, specifically an attacker that uses a transmitter's carrier frequency. We propose low-cost mechanisms for mitigating privacy loss for various radiometric characteristics. In our development and evaluation, we specifically consider a vehicular network (VANET) environment. We consider this environment in particular because VANETs will have the potential to leak significant, long-term information that could be used to compromise drivers' personal information such as home address, work address, and the locations of any businesses the driver frequents. While tracking a vehicle using visually observable information (e.g., license plates) to obtain personal information is possible, such means require line-of-sight, whereas radiometric identification would not. Finally, we evaluate one of our proposed mechanisms via simulation. Specifically, we evaluate our carrier frequency switching mechanism, comparing it to the theory we develop, and we show the precision with which vehicles will need to switch their physical layer identities given our parameterization for VANETs.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Wireless communication*

General Terms

Theory, Experimentation, Performance, Security

Keywords

vehicular networks, VANET, privacy, radiometric identification, simulation, theory

1. INTRODUCTION

A significant amount of work has been put into preserving privacy in VANETs at the application layer [1, 2, 3, 4]. However, preventing privacy-compromising attacks in VANETs is a network stack-wide problem. If a vehicle can be identified more easily based on its radio's characteristics at one layer than at other layers, then a privacy-compromising attacker will simply use the easier method.

Additionally, radiometric identification has a broader application than just VANETs. Because wireless transmissions have the potential to leak information regarding long-term identity, militaries may find it desirable to obfuscate their transmission signatures so as to avoid being tracked or identified. For example, in order to hide which units are deployed at which places on a battlefield or in a war zone, it may be desirable to obfuscate the radio signature of the communication equipment used by each of the units. Hiding the unit's identity may be useful in preventing adversaries from determining what type of unit is deployed at each location and whether any unit has taken any casualties based on earlier encounters.

Besides militaries, consumer device manufacturers, including automotive manufacturers, may be interested in providing their customers with privacy preserving wireless devices. However, the cost constraints on such enhancements are likely to be tight given the competitive nature of this area of application.

Previous work on radiometric identification (also called wireless fingerprinting) has convincingly demonstrated that commodity hardware can be effectively identified based on the characteristics of its physical-layer (PHY) components [5,

[*]This material is based upon work partially supported by USARO under Contract No. W-911-NF-0710287 and NSF under Contract No. CNS-0953600.

VANET'11, September 23, 2011, Las Vegas, Nevada, USA.
Copyright 2011 ACM 978-1-4503-0869-4/11/09 ...$10.00.

6]. Tolerances in these components give rise to identifiable characteristics in the waveform and modulation errors. In particular, prior work has examined startup transients in the waveform, carrier frequency offset, in-phase and quadrature (I/Q) offsets, preamble modulation quality, and quadrature phase-shift keying (QPSK) modulation errors.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of previous work. We investigate theoretical limits on an attacker's ability to use a specific radiometric identification characteristic to compromise privacy in Section 3. In Section 4, we examine how low-cost techniques can be used to mask the most distinct parts of radiometric identification. In Section 5, we discuss how our suggested mitigations should be incorporated into broader schemes for privacy protection, i.e., wireless nodes should coordinate identity shifts. We present our evaluation via simulation of one of these mitigation strategies in Section 6. Specifically, we parameterize our simulations for the VANET environment. Finally, we present conclusions and other considerations in Section 8.

2. PREVIOUS WORK

Dötzer [2], in his analysis of VANET privacy issues, identified wireless fingerprinting as having the potential to be used to distinguish among vehicles.

Remley et al. [7] studied the feasibility of identifying wireless local area network (WLAN) cards via their "electromagnetic signatures." In the controlled environment of an anechoic chamber, the authors measured two cards from each of three manufacturers using a vector signal analyzer (VSA) and a high-speed oscilloscope. The authors presented figures of the cards' outputs in the frequency domain (from the VSA) and of the start of the signals from each card in the time domain (from the oscilloscope). These figures show that one can visibly distinguish among manufacturers and between each card from a single manufacturer. The authors also presented limited data supporting the hypothesis that the antenna radiation pattern is a distinguishing feature among device manufacturers. The authors concluded that radiometric identification of WLAN cards is feasible, but more testing outside of controlled environments was necessary. The authors also concluded with questioning whether identifying individual cards (even from the same manufacturer) is a reliable mechanism for authentication.

Brik et al. [5] answered this final question in the affirmative and provided a list of radio frequency features of physical hardware, which can be used to identify the specific hardware in use. The authors presented their system, PARADIS, for using radiometric measurements to identify individual radios among a large set of 802.11 commercial devices, all being from the same manufacturer. Using various classification algorithms, the authors were able to identify individual devices under varying channel conditions and across a long span of time with over 99% accuracy. The authors briefly mentioned that their results may have serious privacy implications but left further investigation of privacy issues as future research. It is from this point that we launch our work regarding providing privacy, mitigating the possibility of such privacy-invasive measurements. The authors stated early in their paper that radiometric identification has long been practiced by military and corporate entities, the details of which are unpublished.

Edman and Yener [6] described their experimental setup for determining if radiometric identification can be accomplished reliably with hardware much less expensive than VSAs, as used by Remley et al. and Brik et al. Edman and Yener also used their setup, based on USRP2 software defined radios,[1] to attempt to impersonate laptop-based nodes. That is, by recording many 802.11b frames, they trained their software defined radio attacker to reproduce the modulation attributes (e.g., average phase error and frequency offset) of the legitimate laptop-based nodes. Using the same support vector machine (SVM) classifier as Brik et al., the authors were able to impersonate other nodes with a 55% rate of success. The authors stated motivation for using their setup, that is, using software defined radios rather than VSAs, was to use a more likely real-world setup, one that was much less expensive to deploy than in previous work. However, even software defined radios are far too expensive for automobile manufacturers to justify using them for dedicated short-range communication (DSRC) radios. Consequently, if techniques to modify a transmitter's wireless fingerprint are to be deployed, much cheaper solutions need to be found and demonstrated.

Bauer et al. [8] investigate an eavesdropper using the received signal strength (RSSI) of packets and a k-mean clustering algorithm to link transmissions from different sources in a WLAN. In addition to the authors' experimentation, they provide an overview of privacy techniques and leakage across the network stack. The authors show that an attacker with only a small number of listening posts can distinguish between different transmitters using a vector of RSSIs (each individual measurement coming from a different listening post) with 77-85% accuracy. The authors show that an eavesdropper using this clustering knowledge can additionally use it to increase his accuracy in identifying sources and destinations of encrypted HTTP traffic. Finally, the authors implement two privacy preserving techniques for mitigating privacy loss due to RSSI measurements: directional antennas and transmission power control. Using both of these mechanisms simultaneously, they show clustering accuracy is reduced by almost 50%. In their final discussion, the authors note that their techniques for mitigating privacy loss due to RSSI measurement are imperfect for a number of different reasons. Specifically, transmitters cannot perfectly know or predict the exact fading of wireless signals and so do not know if a certain placement of a directional antenna or a certain lower power setting will result in better or worse interference in the WLAN, or if WLAN connectivity will be sufficiently reliable.

Danev and Capkun [9] experiment with using transmitter start-up transients, custom hardware, and various statistical techniques for classification of transmissions from sensor network nodes. The authors investigate their classifiers' performances in static environments and show that changing antenna polarization (i.e., orientation) alters the shape of the start-up transients.

Any attempt to evaluate the effectiveness of privacy preserving techniques must take into account the efficiency of the estimators used in radiometric identification. General bounds on the estimator's performance that only depend on the channel conditions and the observation interval are the well-known *Cramér-Rao bound* (CRB) [10, Ch. 3], the evaluation of which is intractable in many cases. The *modified Cramér-Rao bound* (MCRB), introduced by D'Andrea et al. [11], provides a looser bound for which more compact expressions can be found, and which we discuss in the next section.

[1] See `http://www.ettus.com`.

3. THEORETIC LIMITS

In order to make standards-compliant devices relatively cheap to manufacture, standards bodies allow for tolerances in the analog waveform generated by a wireless network interface device. Recent work also indicates network interface devices are not always made to strictly follow specifications [12]. Additionally, for a brief period of time before the preamble, a node may emit meaningless but very device-specific transients. Likewise, the modulation parameters allow for some limited amount of error. When some error is systematic based on characteristics specific to a particular hardware device, one can gain information about that device. Many authors have proposed anonymity measures that generalize the popular metric of anonymity set size [13, 14, 15, 16, 17] in terms of information theoretic notions such as entropy [13, 14], mutual information [15] and capacity [16, 17] in the covert channel that is exploited by the attacker. In the case of radiometric identification, the covert channel can be viewed as the cascade of two transformations: one has the transmitter identity T as input and the true value of a transmission parameter X as output, whereas the second takes X as its input and has the corresponding value $\hat{X}$, measured or estimated by the attacker.

The data processing inequality [18, §2.8] for the mutual information between $\hat{X}$ and T, that is the amount of information the attacker can derive from $\hat{X}$ about T, yields

$$I(\hat{X};X) - H(X|T) \leq I(\hat{X};T) \leq \min\{I(\hat{X};X), I(X;T)\}$$

where it is clear that in order to reduce the quantity of information available to the attacker it is necessary to increase the conditional entropy $H(X|T)$ and it is sufficient to reduce $I(X;T)$ at the transmitter side. Both can be achieved by introducing some randomness in the generation of X at each transmitter.

In general, when there are a set of independent errors $X_1, \ldots, X_n$, the information from each error can be aggregated as

$$I(\{X_i\};T) = \sum_{i=1}^{n} I(X_i;T)$$

Thus, even if each form of modulation error contributes only a few bits of information, the aggregate information from all of the errors may be sufficient to identify a single transmitter out of a large group. If we can introduce sufficient noise in the error X that is produced by any particular transmitter, we can drive $H(X|T)$ closer to $H(X)$, reducing the mutual information and reducing the certainty of identification, and thus privacy reduction resulting from wireless fingerprinting.

This task is also related to the problem of *resolvability* of the covert channel from X to $\hat{X}$, that is to measure the minimum amount of randomness that X must possess, so that the distribution of $\hat{X}$ will be indistinguishable from that obtained with a given random input X' to the channel. The solution to this problem was formulated for discrete channels [19], and extended to continuous (and particularly additive white Gaussian noise (AWGN)) channels [20]. The solution is given by X taking with equal probability one of $N = 2^{I(X';\hat{X})}$ properly spaced values, which can then be specified with $I(X';\hat{X})$ bits of precision.

Suppose the system allows the parameter X to take values in the range $(x_0 - \Delta, x_0 + \Delta)$, where x_0 is the standard nominal value, and Δ the tolerance. Let the true parameter X of a randomly chosen transmitter have probability density function

$$p_X(x) = \frac{1}{\Delta} q\left(\frac{x - x_0}{\Delta}\right)$$

where $q(\cdot)$ has support $(-1, 1)$, and thus $p_X(x)$ has support on $(x_0 - \Delta, x_0 + \Delta)$.

The optimal strategy by the attacker would use maximum likelihood (ML) estimation on X to identify the transmitter. In this case, the estimation error $\hat{X} - X$ is asymptotically Gaussian distributed, with zero mean and its variance σ^2 attains by the CRB. Consequently, the probability density function of $\hat{X}$ is

$$p_{\hat{X}}(x) = \frac{1}{\Delta} \tilde{q}\left(\frac{x - x_0}{\Delta}; \frac{\Delta}{\sigma}\right)$$

where

$$\tilde{q}(u;y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} q\left(\frac{u - v}{y}\right) e^{-v^2/2} \mathrm{d}v$$

The mutual information between X and $\hat{X}$ can then be derived as

$$\begin{aligned} I(X;\hat{X}) &= E\left[\log_2 \frac{p_{\hat{X}|X}(\hat{X}|X)}{p_{\hat{X}}(\hat{X})}\right] \\ &= \log_2\left(\sqrt{\frac{2}{\pi e}} \frac{\Delta}{\sigma}\right) - \\ &\quad \int_{-\infty}^{\infty} \log_2\left(2\Delta p_{\hat{X}}(x)\right) p_{\hat{X}}(x) \mathrm{d}x \\ &= i\left(\frac{\Delta}{\sigma}\right) \end{aligned}$$

where

$$i(y) = \log_2\left(\sqrt{\frac{2}{\pi e}} y\right) - \int_{-\infty}^{\infty} \tilde{q}(u;y) \log_2\left(2\tilde{q}(u;y)\right) \mathrm{d}u$$

In the first line of the above equation, we apply the definitions of mutual information and conditional probability. To derive line 2, we apply the expectation and use the fact that $\hat{X} - X$ has a Gaussian distribution.

If we assume X to be uniformly distributed over the interval $(x_0 - \Delta, x_0 + \Delta)$, then $q(u) = \frac{1}{2}$, if $|u| \leq 1$ and $q(u) = 0$ otherwise. This results in

$$\tilde{q}(u;y) = \frac{1}{2}\left[\Phi(y(u+1)) - \Phi(y(u-1))\right]$$

where $\Phi(y)$ is the normal probability cumulative distribution function, and yields approximately

$$i(y) \approx -\frac{1}{2}\log_2(\pi e) + \log_2(y + \alpha) + \frac{1}{2}$$

where

$$\alpha = -\int_{-\infty}^{\infty} \Phi(v) \ln \Phi(v) \, \mathrm{d}v \approx 0.9032$$

A uniform distribution for X yields the maximum value of $I(X;\hat{X})$ and hence

$$N \approx \sqrt{\frac{2}{\pi e}} \frac{\Delta}{\sigma}$$

(we neglect the term $\alpha\sqrt{2/(\pi e)} < 1$) represents the worst-case bound for privacy.

A more common case is that X is Gaussian with mean x_0 and standard deviation $\sigma_X < \Delta/3$, so that the probability of exceeding the allowed range is negligible. In this case, we obtain the well-known formula for a Gaussian channel with Gaussian input $I(X;\hat{X}) = \frac{1}{2}\log_2(1+\sigma_X^2/\sigma^2)$ yielding an approximate value of

$$N = \sqrt{1+\frac{\sigma_X^2}{\sigma^2}} \approx \frac{\sigma_X}{\sigma}$$

As an example, consider the estimation of the carrier frequency offset f_c, which, according to Brik et al. [5], is the most telling statistic for identifying an individual device. D'Andrea et al. [11] derived a MCRB for estimating the frequency offset of a signal, with the following expression

$$\sigma^2 = \frac{3}{2\pi^2 {T_0}^3 B}\frac{1}{\mathrm{SNR}} \tag{1}$$

where T_0 is the total duration of the transmission and B is the bandwidth. The above expression also shows the typical inverse proportionality between CRB and SNR. The MCRB holds in general, as a looser bound than the CRB, under the assumption that the privacy compromiser's estimator is unbiased. If the compromiser receiver can also correctly decode data (which is reasonable in our case, given that f_c is within the accepted range) the MCRB coincides with the actual CRB and it is then asymptotically (for high SNR or high T_0) achieved by ML estimators. Therefore, for a uniform f_c in a $\pm\Delta$ range we get

$$N \approx 2\Delta\sqrt{\frac{\pi B {T_0}^3}{3e}\mathrm{SNR}} \tag{2}$$

and for a Gaussian f_c with standard deviation $\Delta/3$, we get

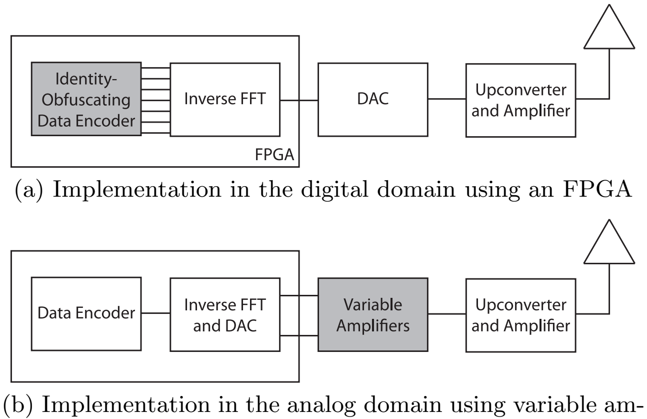
$$N \approx \frac{\Delta\pi}{3}\sqrt{\frac{2BT_0^3}{3}\mathrm{SNR}}.$$

As a result of Equation (2), there are a number of methods for reducing a privacy compromiser's ability to resolve individual vehicles. First, we can force the attacker to only receive with low SNR, but this is impractical because the compromiser is allowed to be anywhere, and consequently, put his receiver anywhere. Reducing the SNR otherwise will result in poorer network performance for the intended uses of the VANET (e.g., lower packet delivery ratio for legitimate vehicles). Second, we can reduce the signal bandwidth, but this has the adverse affect of reducing the rate at which vehicles can communicate, which is also unacceptable. Third, we can reduce Δ, but this requires using crystals of higher quality, which may be expensive. Finally, we can limit the amount of time an attacker can observe by limiting the length of time a transmitter uses an individual frequency, f_c.

In real-world environments, interference must also be considered. However, attackers may be susceptible to varying amounts of intereference. We will discuss intereference further below when we introduce our simulations in Section 6. If the source of the interference comes from many independent sources, the interference can be approximated as Gaussian noise, and SNR can be replaced with SINR. We also leave the further development of theory corresponding to interference to future work.

4. IMPLEMENTATION OF LOW-LEVEL LOW-COST ERROR MODULATION

There may be low-cost ways of introducing new, easily modulated errors that mask the original sources of error. Specifically, for each of the identifying characteristics according to Brik et al. [5], we suggest inexpensive mechanisms that hide the imperfections of the corresponding components.


Identity-Obfuscating Data Encoder
Inverse FFT
FPGA
DAC
Upconverter and Amplifier
Data Encoder
Inverse FFT and DAC
Variable Amplifiers
Upconverter and Amplifier

(a) Implementation in the digital domain using an FPGA

(b) Implementation in the analog domain using variable amplifiers

Figure 1: Implementation design strategies for fingerprint obfuscation. The shaded boxes represent the novel contributions of our proposed mitigations.

4.1 Startup Transients

Obfuscating signals could be modulated even before the preamble is sent in order to minimize the distinctiveness of startup transients. Because previous work in startup transients tends to examine the phase noise during startup, sending an input on the I and Q channels should help obfuscate startup phase noise that is specific to each radio.

4.2 Carrier Frequency Error

Carrier frequency error could be introduced in a variety of ways, including using digital potentiometers to heat a crystal in a crystal oscillator or capacitor in an LC oscillator's tank circuit, thus changing the frequency of the oscillator over time. A technique with lower latency might be changing the voltage on the input to a voltage controlled oscillator (VCO) thus changing the frequency error simply by changing the voltage input to the oscillator. Additionally, frequency synthesizers or direct digital synthesizers could be used, which would allow direct digital control over the carrier frequency. Finally, additional precision may be obtainable using timing rather than a strict hardware-only implementation. For example, to obtain more precision in generating a voltage (i.e., to adjust a VCO), a pulse-width modulated (PWM) signal could be generated and smoothed using a low-pass filter (e.g., a large capacitor and resistor pair). This example solution could allow for a lower-cost digital to analog converter (DAC) than a hardware-only solution. Some precision could be provided by hardware (i.e., a DAC) and the remainder by a timer and a PWM signal.

4.3 Modulation Components

In this category we consider I/Q offsets, preamble modulation quality, and QPSK phase and amplitude modulation errors. In general, any modulation component could be modified in the digital domain, that is, in software, simply by modifying the I and Q values. For example, a transmitter can rotate the entire constellation by ϕ using

$$I' = I\cdot\cos(\phi) - Q\cdot\sin(\phi) \quad \text{and} \quad Q' = Q\cdot\cos(\phi) + I\cdot\sin(\phi)$$

A transmitter can perform this modulation either all the time, during specific parts of the packet (such as the preamble), or only for specific values of I and Q. The main disadvantage to this approach is that it requires a substantial number of extra bits in the DACs required for outputting I and Q, because if the number of extra bits is small, then any movement of I and Q would be large. This disadvantage disappears, however, in multi-carrier systems such as OFDM (which will be used by VANETs) because all of the digital orthogonal frequency-division modulation (OFDM) frequency-domain modulation is done in a high precision Inverse Fast Fourier Transform (IFFT) right before the DAC (as illustrated in Figure 1(a)); inputs to this IFFT can be at an equally high precision.

Another way to introduce changes to modulation is in the analog domain by adding separate variable gain amplifiers for the I and Q signals as soon as they enter the analog domain. A transmitter can thus choose different gains and can shift the points of the constellation as it chooses.

However, if these privacy obfuscating techniques produce transmitters that cannot be identified via wireless fingerprinting but results in a system that cannot receive any data, then the techniques are of little importance. Consequently, we leave to future work the building of such an radiometric identity obfuscating radio. Before any such system can or should be built, an investigation via simulation can be performed to validate the expectations of the privacy gained by any proposed introduction of errors. We will describe our investigation via simulation of techniques for introducing carrier frequency error in Section 6.

5. COORDINATED IDENTITY SHIFT

In certain circumstances, a node may have limited ability to inject noise into a measured phenomena. For example, a node's startup transient may be very short, limiting a node's ability to effectively and substantially modulate that noise. Our goal, however, is not to ensure that the source of *every* packet is indistinguishable from the source of every other packet. In general, when a collection of packets is inherently tied together, such as by protocol or addressing [21], obtaining privacy against wireless fingerprinting does not help overall privacy. As with many other security properties, a privacy-preserving system is only as strong as its weakest link, since an attacker that cannot identify a node with wireless fingerprinting could simply use MAC or network addresses instead. Thus, we propose to only shift identities at the physical layer when there is a shift in higher-layer identities, as previously proposed by Dötzer [2]. In previous work, Jiang et al. [22] developed a protocol for collision-free MAC address changes. In a vehicular system or other system that desires privacy from wireless fingerprinting, a change in higher-layer certificates (as in IPSec [23], TLS [24], or otherwise) should be coordinated with a change in network-layer address (e.g., IPv6 Address Privacy [25]), link-layer address [22], and physical-layer properties.

Another reason for coordinating identity shifts is that the naïve approach of changing PHY-layer identity with each packet gives rise to an attack where the attacker eliminates the introduced noise by averaging over a large number of packets that are all associated with the same higher-layer identity. For example, if a transmitter changes the I/Q offset of each of its packets by offsetting each element of a QPSK constellation equally, an attacker can measure the I/Q offset of all packets corresponding to a single higher-layer identity, thus measuring $\overline{X+Z}$ where X is the I/Q offset introduced by the transmitter's hardware and Z is the average I/Q offset introduced by the transmitter's intentional modifications. However, when averaged over a sufficient number of measurements, Z will exhibit insufficient noise to retain strong privacy, further reinforcing the need for coordinated shifts in identity.

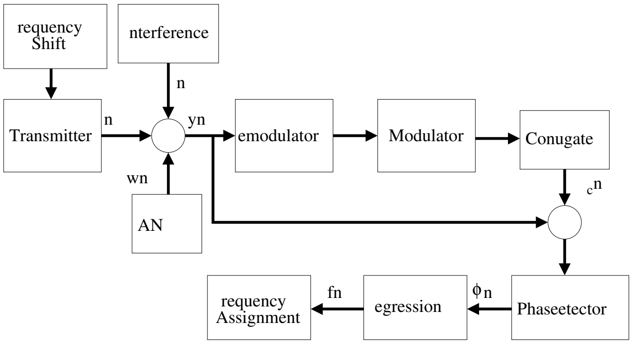


Figure 2: Simulated privacy compromiser model.

6. SIMULATION ANALYSIS

To evaluate the effectiveness of privacy obfuscation techniques, we simulated a privacy compromiser. Our simulations were based on both the VANET environment (802.11p) and a potential hardware radio design in order to evaluate whether or not the attacker can achieve the bound on the number of resolvable identities for carrier frequency offset. By simulating this environment, we are also able to evaluate the effects of interference on a simulated privacy compromiser. Measuring this performance via implementation would be prohibitive in terms of cost, space, and time.

6.1 Attacker Model

Figure 2 shows the model of our simulated privacy compromiser (attacker) (i.e., how our attacker's hardware works to compromise privacy via carrier frequency offset). In our simulations, we made the assumption that the attacker will be able to receive transmitted data and will know the number of transmitters. Thus, the attacker's problem is to decide which transmissions are from the same transmitters. We used QPSK modulation, as QPSK is expected to be used for transmitting VANET safety beacons [26]. We modeled the channel as an AWGN channel, which we also simulated with and without interference from other vehicles. The attacker used maximum likelihood estimation to estimate the carrier frequency via differentiating the phase error and a least-squares regression. By simulating the attacker in this way, he is almost able to achieve the MCRB bound for an AWGN channel.

Because the attacker needs to decide which transmissions belong to the same vehicles, which decision algorithm he uses is an important choice. In order for our proposed wireless identity changing techniques to be useful, as we noted in the previous section, a vehicle's identity at other layers also needs to be changed in a coordinated manner. Thus, we consider that the vehicles an attacker observes enter a mix zone or participate in a silent period, through or during which the observed vehicles change their identities. Thus, the attacker's goal is to match identities between the transmissions he heard before and after the vehicles change their identities. In other words, the attacker observes two sets of packets from each observed vehicle: one set before the mix zone or silent period and one set after the mix zone or silent period.

Because an attacker can easily link transmissions either before or after an identity change, we only consider that the

Table 1: Parameters for simulation analysis of a privacy attacker.

Parameter	Setting
SNR	36 dB
Phase noise	Colored Gaussian (AD 4106)
Frequency switching precision	8, 10, 12, 14, 16, 18, 20, 22, 24 bits
Vehicle group size	20, 40, 60, 80, 100, 120
Packet size	100 bytes
Number of received packets	10

attacker needs to match transmissions and vehicles between sets. The transmissions are likely easily matched either before or after by using any or all of the following information that is likely to be the same across multiple packets: MAC address, network address, signing certificate, etc. Additionally, the attacker should easily be able to link transmissions using the known dynamics of vehicles (e.g., vehicles do not teleport) to link transmissions. This linking must be possible in order for safety applications to be effective. Consequently, we model the number of transmissions that the attacker receives from a single vehicle as a single long packet. For the frequency offset estimator we described above, the attacker need only to match phase between received packets and concatenate the transmissions to emulate this model.

6.2 Methodology

In this section, we describe our methodology for evaluating the effectiveness of the attacker in resolving vehicles' radiometric identities based on carrier frequency offset. we summarize our exploration of parameters in Table 1, and explain how we arrived at those chosen parameters below.

6.2.1 Channel

We modeled the channel as an AWGN channel, per the theory we developed in Section 3. We parameterized the channel using the baseline noise for 802.11p channels at -96 dBm, as set in NS-2 [27]. Considering that an attacker is free to use directional antennas and place his antennas (theoretically) anywhere he chooses, we chose to use a relatively high received signal strength of -60 dBm for all of the packets the attacker receives. Thus, the attacker's SNR was 36 dB.

6.2.2 Phase Noise

We modeled the phase noise of transmitters' carrier frequencies based on the AD4106 frequency synthesizer.[2] We chose to model phase noise using the phase noise data from this part's data sheet because it contained data at 5.8 GHz, which is very close to the nominal 5.18 GHz DSRC carrier frequency. From the data sheet, we used the phase noise measurements of -83.5 dBc at 1 kHz, -85 dBc at 10 kHz, and -115 dBc at 1 MHz. We generated phase noise using this data and a colored Gaussian noise model.[3]

Phase noise serves as an additional noise source that could obscure a vehicle's radiometric identity. Additionally, with the colored Gaussian phase noise model, a vehicle's carrier frequency can drift slightly during and across packets.

6.2.3 Interference

Interference from other vehicles' transmissions may also serve to obscure the radiometric identity of a single vehicle. Specifically, in our simulations we investigated the effects of interference on the ability of an attacker to use carrier frequency offset as a source of radiometric identification. To parameterize our simulations, we used the Illinois VANET simulator [28] to simulate a section of Zürich, Switzerland [29].

The specific trace we chose to simulate and from which we gathered statistics on interference was from the partial trace located in the center of the city (specifically, the Zentrum-Bellevue high density trace). We used the first 1000 seconds of this trace. The trace covered an area of approximately 4.7 kilometers x 4.0 kilometers and contained on average 1280 active vehicles at any time. We chose to simulate such a large trace to reduce edge effects. Figures 3(a) and 3(b) show the probability density function of the number of interfering transmissions and the relative power of interfering packets during this trace, respectively. To generate this data and these figures, we recorded the distributions of the number of interfering transmissions and the power of interfering transmissions. The spike at -99 dB in Figure 3(b) is an accumulation of all interfering transmissions that are -99 dB or lower in power. The powers shown in this figure are relative to the received signal strength of the received packet.

In the simulations below, we will show the attacker's ability to identify vehicles both with and without interference. If an attacker simply deploys (roughly) omni-directional antennas, then the attacker will have to deal with interference. However, it may be possible through a combination of intelligent placement of antennas and the use of directional antennas for the attacker to eliminate (or largely mitigate) the effects of interference. Thus, in our simulations below, we will investigate attackers who are either immune or susceptible to interference.

6.2.4 Signal Duration

An attacker's effectiveness in using carrier frequency offset to resolve a vehicle's identity depends on the duration of the signal, or cumulative signals the attacker receives from the vehicle. We simulated safety messages being 100 bytes long, and the attacker receiving 10 packets from each vehicle. Thus, if vehicles broadcast safety beacons at the rate of 10 Hz, 10 packets corresponds to the attacker overhearing a vehicle across 1 second.

6.2.5 Carrier Frequency Switching

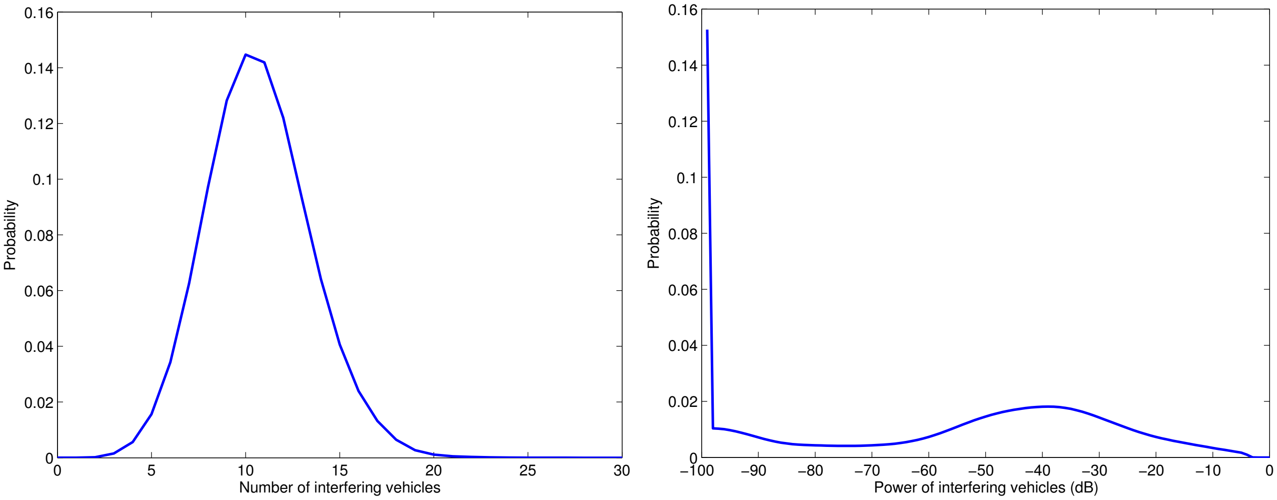
We simulated vehicles switching their carrier frequencies by having them divide the defined range for carrier frequencies, $[f_0 - \Delta, f_0 + \Delta]$ by the number of identities vehicles can assume. We varied the number of identities vehicles could assume by varying the precision of their carrier frequency switching, that is, the *number of bits of precision* with which vehicles could change their carrier frequency. We varied this precision from 8 to 24 bits.

6.2.6 Rank

We chose to use *rank* as the metric to evaluate the effectiveness of our suggested carrier frequency switching mechanism and the attacker's ability to identify vehicles. Rank provides a straightforward metric for understanding an attacker's ability to match vehicles and is both more clear and more broad than entropy. The core idea behind rank is that instead of using a single piece of information, e.g., carrier frequency offset, an attacker may use multiple pieces of in-

[2] See `http://www.analog.com/en/rfif-components/pll-synthesizersvcos/adf4106/products/product.html`.

[3] See `http://www.mathworks.com/matlabcentral/fileexchange/8844-phase-noise`.



(a) PDF of the number of interfering vehicles

(b) PDF of the power of interfering vehicles

Figure 3: Interference data gathered from our simulations of central Zürich, Switzerland.

formation together to finally identify a vehicle. Such other information might take many forms, such as a learned training set of how vehicles behave at an intersection, including which lanes they turn from and the probabilities that they turn or go straight. Thus, an attacker could simply identify a group of most likely candidate vehicles using a single piece of information (e.g., carrier frequency offset) for later refinement with other information.

Rank works as follows. An attacker sorts a list of vehicles in order of most likely match to least likely match, according to the information he has. If a vehicle appears in the x most likely vehicles in this list (i.e., the first x), then the vehicle is identified with rank x. Also, the vehicle is identified with rank $x' \geq x$. Thus, by running multiple trials, we are able to assemble statistics on an attacker's ability to identify vehicles using carrier frequency offset as the probability that an attacker identifies a vehicle within a specific rank, that is $P(\text{Rank} \leq x')$. Additionally, previous work has used rank to evaluate privacy [21].

Using the vehicle trace from Zürich, Switzerland, as we described above, we visually observed the trace and identified a busy intersection. From the results of our earlier simulation of this trace, we determined the effective transmission range in this trace (50% probability of reception at the network layer) to be 152 meters. We recorded the number of active vehicles within 152 meters of the identified intersection, which varied across the duration of the trace between 65 and 121 vehicles. Thus, we modeled the attacker needing to match vehicles within groups of size 20, 40, 60, 80, 100, and 120 vehicles. We consider that the attacker knows the size of the group of vehicles, that the same vehicles are present in both groups, and the number of bits of precision that vehicles are using.

7. SIMULATION RESULTS

First, we measured the performance of our proposed attacker model relative to the MCRB in Equation (1). Simulating our attacker listening to 1000 vehicles (i.e., running 1000 tests) and using only background noise (i.e., no oscillator phase noise, interference, or carrier frequency switching), we found that the variance on the attacker's estimated frequency offsets was $1.44\sigma^2$. Thus, our attacker model is very close to but does not achieve the bound. Again, it may not be possible to achieve the bound since MCRB's are looser lower bounds than actual Cramer-Rao bounds.

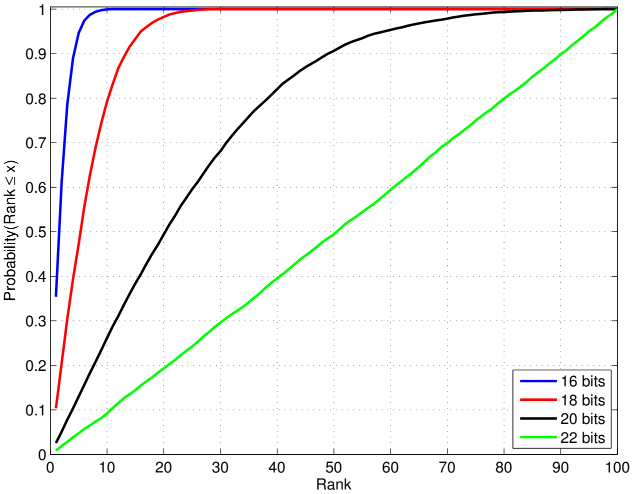
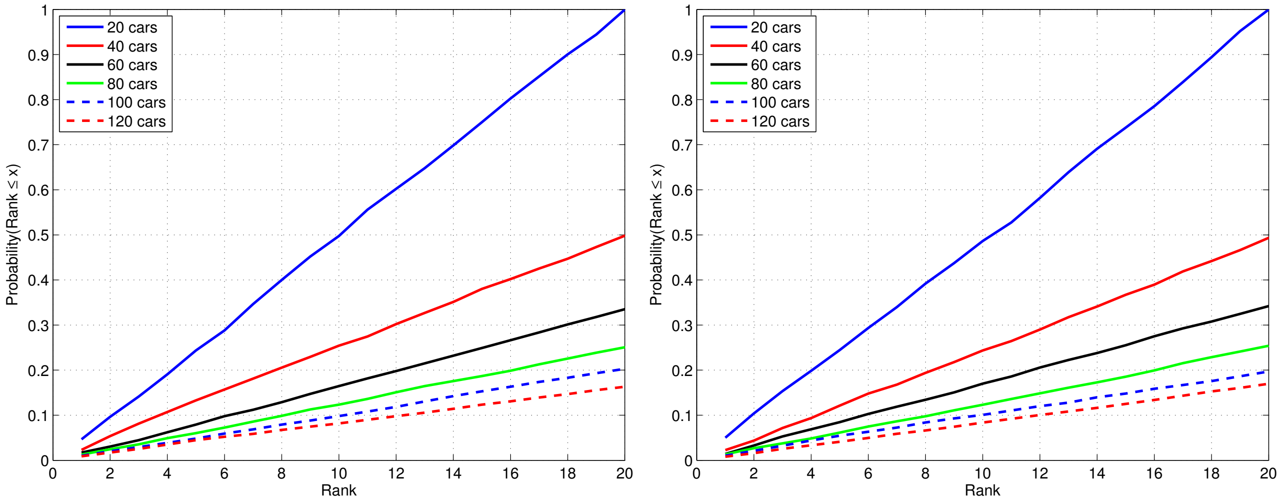


Figure 4: Varying number of bits of precision for frequency switching effects on privacy attacker's ability to rank vehicles for identification based on carrier frequency offset, using a group size of 100 vehicles, without phase noise and without interference.

Additionally, based on the setup we detailed in the previous section, Equation (2) predicts that for an attacker that receives 10 packets, vehicles will require 22.63 bits of precision in varying in their carrier frequencies. We rounded this to the nearest even number of bits, 22, for the following simulations.

7.1 Attacker Estimator Performance

We first began by evaluating the performance of the attacker's estimator by simulating the attacker without the effects of phase noise or interference. Figure 4 shows the attacker's ability to rank vehicles with no phase noise and no interference. This figure shows that vehicles can maintain perfect privacy using 22 bits of precision, as predicted by



(a) With interference

(b) Without interference

Figure 5: Effects of group size on privacy attacker's ability to rank vehicles for identification based on carrier frequency offset.

our theory. Using fewer bits of precision causes the amount of privacy a vehicle can maintain to rapidly decrease.

7.2 Observed Group Size

Next, we investigated the effects of group size on the attacker's ability to identify vehicles based on carrier frequency offset. We enabled phase noise in our simulations. In all results that we present hereafter we enabled phase noise.

Figure 5 shows the effects of group size on the attacker's probability of correctly ranking vehicles for identification, having simulated the attacker with and without interference. Each data series is the result of 100 tests. Theoretically, the worst that the attacker can do in terms of ranking vehicles is $P(\text{Rank} \leq x) = \frac{x}{\text{group size}}$, that is, random guessing. The data series in these figures approximate this limit, inferring that the attacker gains no information from carrier frequency offset, when vehicles use 22 bits of precision for switching their carrier frequency identities.

7.3 Bits of Precision

Next, we varied the number of bits of precision with which vehicles switch their carrier frequency. We experimented with various bits of precision with group sizes of both 20 and 100 vehicles. Again, each data series below represents 100 runs.

7.3.1 100 Vehicles

Figure 6 shows the probability that an attacker assigns a given rank to a vehicle with various numbers of bits of precision for carrier frequency switching and with and without the effects of interference. We performed all runs with a group size of 100 possible vehicles. Thus, in these figures a line with slope $\frac{1}{100}$ represents random guessing.

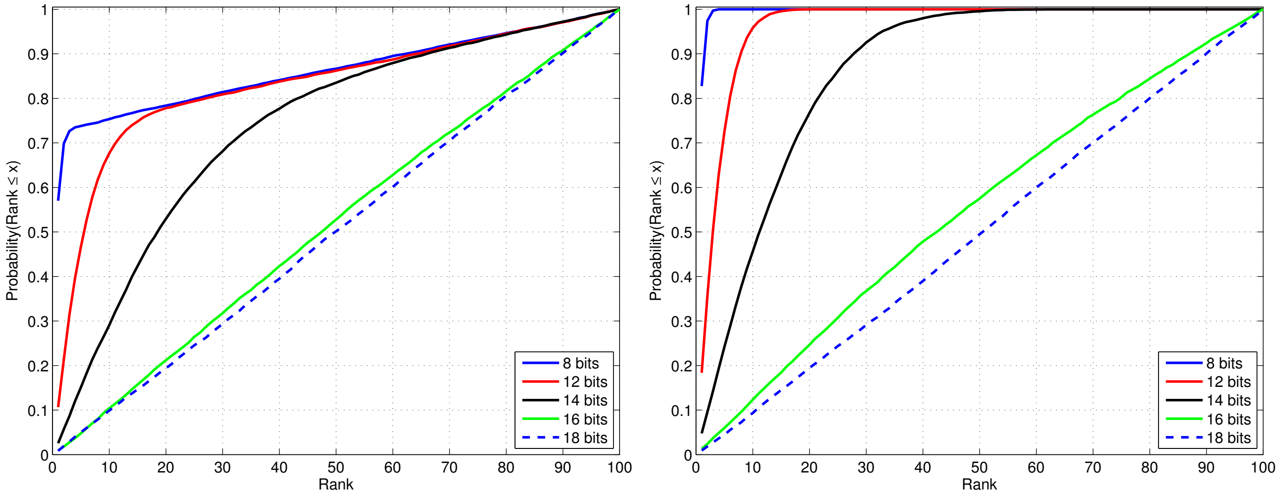
In Figure 6(a), using 18 bits for carrier frequency switching results in the attacker having approximately as good a chance at identifying a vehicle using carrier frequency offset as using random guessing. If vehicles reduce the number of bits of precision they use to 16, the advantage gained by the attacker is minimal but non-zero. However, with only 8 bits, the attacker has a distinct advantage and can even identify a vehicle directly (i.e., rank equal to 1) with greater than a 50% success rate.

Figure 6(b) shows how effective the attacker can be if he is able to ignore or is immune to interference. In this figure, 18 bits also results in attacker performance that approximates random guessing, while 16 bits gives the attacker a small but noticeable advantage. The attacker's advantage with 16 bits is larger without interference than with interference. Reducing the number of bits of precision further still to 8 bits results in the attacker being able to directly identify a vehicle (i.e., rank equal to 1) with greater than 80% accuracy and 97.4% accuracy when considering the two most likely vehicles (i.e., rank less than or equal to 2).

Thus, there is a clear advantage for the attacker that can eliminate or is immune to interference from other vehicles, if vehicles must use less than 18 bits for switching their carrier frequencies. However, vehicles can use 18 bits of precision to maintain perfect privacy, requiring 4.63 fewer bits of precision (compared to the 22.63 bits predicted by Equation 2), independent of the attacker's susceptibility to interference. If vehicles cannot use 18 bits of precision, there is a clear advantage for the attacker that is immune to interference, and his advantage increases rapidly as the number of bits of precision decrease. In other words, to mitigate the attacker's advantage, the carrier frequency switching functionality requires more bits of precision. However, since obtaining more bits of precision results in higher cost (at least for a hardware implementation of carrier frequency switching), maintaining perfect privacy for carrier frequency switching may not be obtainable using a hardware-only solution on a tight budget. On the other hand, some privacy can be preserved with lower cost implementations.

Both of these figures show that phase noise provides 4.63 bits of privacy, independent of the attacker's ability to eliminate interference. If the attacker uses equipment with a higher quality carrier frequency source, vehicles will obtain fewer bits of privacy from phase noise. However, imperfections or non-linearities in the RF front-end may also contribute more noise to the carrier frequency estimation.

Figure 6(a) shows the attacker's ranking ability with interference. The data in this figure appears to approach an



(a) With interference

(b) Without interference

Figure 6: Varying number of bits of precision for frequency switching effects on privacy attacker's ability to rank vehicles for identification based on carrier frequency offset, using a group size of 100 vehicles.

asymptote for series besides the "16 bits" and "18 bits" series. This asymptote does not appear (besides a maximum probability of 1) in Figure 6(b), which shows the same scenarios but without the effects of interference. Consequently, the asymptote shown in Figure 6(a) represents the limit of accuracy an attacker can achieve due to estimation errors induced by interference. Additionally, in each of these figures, as the bits of precision are decreased, the advantage the attacker can gain rapidly increases, quickly approaching each scenario's respective limits.

7.3.2 20 Vehicles

Next, we reduced the group size to 20 vehicles. We omit showing our results for 20 vehicles due to space constraints and because the results were nearly identical to our experiments with 100 vehicles.

8. CONCLUSIONS

Enhancements to PHY-layer privacy could have a very far-reaching effect. Not only would the ability to hide PHY-layer attributes enable VANET privacy, but it could also do the same for standard WiFi or cellular devices. Governments may also find this ability useful for protecting the secrecy of covert operations or the identities of their agents and workers.

Without the ability to provide unlinkability at the PHY layer, upper layer privacy techniques will not be able to provide vehicle privacy. Attackers are likely to take the path of least resistance in terms of using information at various layers to track or identify vehicles.

Governments and manufacturers intend on deploying VANETs in the next 5 years. Manufacturers are unlikely to make major changes to designs of components during that time span. Additionally, vehicle designers need to minimize additional equipment costs to maintain their company's competitiveness. Consequently, low-cost solutions for VANET privacy issues are important for their acceptability to manufacturers and their adoption within such a short time span.

9. REFERENCES

[1] J. Feudiger, M. Raya, M. Félegyházi, P. Papadimitratos, and J.-P. Hubaux, "Mix-zones for location privacy in vehicular networks," in *WiN-ITS 2007*, August 2007.

[2] F. Dötzer, "Privacy issues in vehicular ad hoc networks," in *Proceedings of the Second ACM International Workshop on Vehicular ad hoc Networks*, ACM Press, 2005.

[3] X. Lin, X. Sun, P.-H. Ho, and X. Shen, "GSIS: A secure and privacy-preserving protocol for vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 56, pp. 3442–3456, November 2007.

[4] J. J. Haas, Y.-C. Hu, and K. P. Laberteaux, "The Impact of Key Assignment on VANET Privacy," *Security and Communication Networks*, vol. 3, pp. 233–249, September 2009.

[5] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *MobiCom '08: Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, (New York, NY, USA), pp. 116–127, ACM, 2008.

[6] M. Edman and B. Yener, "Active attacks against modulation-based radiometric identification," Tech. Rep. 09-02, Rensselaer Polytechnic Institute Department of Computer Science, 2009.

[7] K. Remley, C. Grosvenor, R. Johnk, D. Novotny, P. Hale, M. McKinley, A. Karygiannis, and E. Antonakakis, "Electromagnetic signatures of WLAN cards and network security," in *Signal Processing and Information Technology, 2005. Proceedings of the Fifth IEEE International Symposium on*, pp. 484–488, Dec. 2005.

[8] K. Bauer, D. Mccoy, B. Greenstein, D. Grunwald, and D. Sicker, "Physical Layer Attacks on Unlinkability in Wireless LANs," in *PETS '09: Proceedings of the 9th International Symposium on Privacy Enhancing Technologies*, (Berlin, Heidelberg), pp. 108–127, Springer-Verlag, 2009.

[9] B. Danev and S. Capkun, "Transient-based identification of wireless sensor nodes," in *IPSN '09: Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, (Washington, DC, USA), pp. 25–36, IEEE Computer Society, 2009.

[10] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory.* Prentice Hall, 1993.

[11] A. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchronization problems," *Communications, IEEE Transactions on*, vol. 42, pp. 1391–1399, Feb–Apr 1994.

[12] B. Han, L. Ji, S. Lee, B. Bhattacharjee, and R. Miller, "All bits are not equal - a study of IEEE 802.11 communication bit errors," in *IEEE INFOCOM 2009*, pp. 1602–1610, April 2009.

[13] C. Diaz, S. Seys, J. Claessens, B. Preneel, and K. Esat-cosic, "Towards measuring anonymity," in *Workshop on Privacy Enhancing Technologies, PET 2002*, pp. 54–68, 2002.

[14] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *Workshop on Privacy Enhancing Technologies, PET 2002*, pp. 41–53, 2002.

[15] Y. Zhu and R. Bettati, "Anonymity vs. information leakage in anonymity systems," in *IEEE International Conference on Distributed Computing Systems ICDCS'05*, pp. 514–524, 2005.

[16] I. Moskowitz, R. Newman, D. Crepeau, and A. Miller, "Covert channels and anonymizing networks," in *ACM Workshop on Privacy in the electronic society, WPES*, pp. 79–88, 2003.

[17] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden, "Anonymity protocols as noisy channels," *Information and Computation*, vol. 206, pp. 378–401, 2008.

[18] T. Cover and J. Thomas, *Elements of Information Theory.* John Wiley & Sons, 2006.

[19] T. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, pp. 752–772, 1993.

[20] T. Han and S. Verdu, "The resolvability and the capacity of AWGN channels are equal," *IEEE International Symposium on Information Theory, ISIT*, p. 463, 1994.

[21] Y.-C. Hu and H. J. Wang, "A framework for location privacy in wireless networks," in *Proceedings of the ACM SIGCOMM Asia Workshop 2005*, (Bejing, China), ACM, April 2005.

[22] T. Jiang, H. J. Wang, and Y.-C. Hu, "Preserving location privacy in wireless LANs," in *MobiSys '07: Proceedings of the 5th International Conference on Mobile Systems, Applications and Services*, (New York, NY, USA), pp. 246–257, ACM, 2007.

[23] S. Kent and R. Atkinson, "Security architecture for the internet protocol." RFC 2401, Nov. 1998.

[24] T. Dierks and E. Rescorla, "The transport layer security (TLS) protocol version 1.2." RFC 5246, Aug 2008.

[25] T. Narten, R. Draves, and S. Krishnan, "Privacy extensions for stateless address autoconfiguration in IPv6." RFC 4941, Sep 2007.

[26] D. Jiang, Q. Chen, and L. Delgrossi, "Optimal data rate selection for vehicle safety communications," in *VANET '08: Proceedings of the Fifth ACM International Workshop on VehiculAr Inter-NETworking*, (New York, NY, USA), pp. 30–38, ACM, 2008.

[27] Q. Chen, F. Schmidt-Eisenlohr, D. Jiang, M. Torrent-Moreno, L. Delgrossi, and H. Hartenstein, "Overhaul of IEEE 802.11 modeling and simulation in ns-2," in *MSWiM '07: Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, (New York, NY, USA), pp. 159–168, ACM, 2007.

[28] J. J. Haas, Y.-C. Hu, and K. P. Laberteaux, "Real-world VANET security protocol performance," in *Proceedings of the IEEE Globecom 2009, Symposium on Selected Areas in Communications*, IEEE, December 2009.

[29] V. Naumov, R. Baumann, and T. Gross, "An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces," in *MobiHoc '06: Proceedings of the 7th ACM International Symposium on Mobile ad hoc Networking and Computing*, (New York, NY, USA), pp. 108–119, ACM, 2006.