Physical-Layer Cyberattack and Interference Resilient Automotive Radars

ONUR TOKER, (Member, IEEE)
Department of Electrical and Computer Engineering, Florida Polytechnic University, Lakeland, FL 33805, USA
e-mail: otker@floridapoly.edu

This work was supported in part by the NSF under Grant 1919855, in part by the Advanced Mobility Institute under Grant GR-2000028 and Grant GR2000029, and in part by the Florida Polytechnic University Startup under Grant GR-1900022.

ABSTRACT In this paper, we present a physical-layer attack and interference resilient automotive radar system, and derive analytical upper bounds for the probability of not detecting an attack, and the probability of false attack alarm. We consider a quite general attack model and prove that if the attack signal level is above a defined relative threshold, both the probability of false attack alarm and the probability of not detecting an attack converge to zero exponentially with the number of samples acquired during a single chirp, and the number of chirps used in a frame. We also derive an analytical formula for this relative threshold, and prove that by selecting shorter frame durations, and using lower noise RF equipment, the threshold can be made as small as possible. Basically, by proper selection of radar parameters arbitrarily small attack signals can be detected almost always with almost no false alarms. We also present a numerical example using real measured data obtained from two 77 GHz automotive radars operated at the same time. Also using real data, we show that the proposed system reduces the negative effects of undetected weak attacks which are below the above mentioned threshold.

INDEX TERMS Automotive radars, attack resilience, cybersecurity, interference resilience, physical-layer attacks.

I. INTRODUCTION

Autonomous vehicles (AV), and advanced driver assistance systems (ADAS) rely on highly advanced sensors including millimeter wave radars, lidars, and multiple camera based vision systems. Each of these sensor technologies have their own pros and cons. Millimeter wave radars are quite effective in target distance and velocity measurements, direction estimation, and object identification [1]. A standard frequency modulated continuous wave (FMCW) AV radar offers no protection against physical-layer attacks, but using advanced signal processing techniques, a high degree of attack and interference resilience can be achieved. The report [2], [3] has a discussion of different attack scenarios for connected vehicles and AV radars. In [4], an attack resilient automotive radar design was introduced and its effectiveness was demonstrated using simulations. An similar attack resilient design was discussed in [5] and its effectiveness was demonstrated using real measured data, but without any theoretical performance analysis. In this paper, we present a new attack resilient automotive radar system, derive analytical bounds on its attack detection performance, prove that the total risk can be made as small as possible, and demonstrate all of these using real measured data obtained from Texas Instruments 77 GHz automotive radars. All experimental data are obtained using the real-time raw data capture software developed in [6], [7].

A common technique for cyberattack detection is to use an estimator, and generate an alarm if a significant difference is observed between estimated and measured values. Regardless of the problem setup, a significant difference could be an indication of something unusual. Depending on how the difference is scored and thresholded, it is possible to define a variety of different attack detectors, for example see [8]–[11], and references therein. Also in [12]–[14], similar techniques are discussed. These attack detectors have an estimator which can be an analytical expression derived from a probabilistic model, or can be completely artificial intelligence (AI) based. However, for all such designs false attack alarms are always possible because observed differences may be due to simply natural measurement noise and/or estimation error. In general, thresholds should be selected to minimize false alarms without significantly degrading the attack detection capability, see [15] for a related edge case where gradually increasing...
attacks are used to bypass detection and to destabilize the overall system. These estimator based techniques are known to be quite useful for detecting attacks at the communication layer, but the accuracy of the estimation model is of critical importance for reliable operation. Furthermore, the estimation models may need to be updated due to changing environmental conditions and hence may be difficult to maintain for proper system operation.

In [16], [17], physical-layer attacks on automotive radar sensors are discussed. The first one is based on estimators and the second one is based on a technique called spatio-temporal challenge-response (STCR) which does not depend on an estimation model. A related technique called, physical challenge-response authentication (PyCRA), is introduced in [18]. Both STCR and PyCRA techniques can be quite robust because they do not need an accurate estimator for reliable operation. For physical-layer attacks, there is a sensor which may or may not be connected to a communication network, and attacks occur at the physical level, i.e. adversarial agents generate physical signals to directly interfere with the sensor’s measurement process, confuse the sensor and hence the embedded system processing the sensor output [2], [18]. If the processing nodes, i.e. the embedded systems with sensors, are also communicating with each other, then attacks at the communication layer is also possible. But, that is a different attack scenario which has been heavily investigated in the literature. In a good engineering design, both physical-layer and communication layer (if exists) attack detectors should be used to improve safety and reliability. In this work, we focus on physical-layer attacks described in [2], and our approach is inspired by [16], [18]. Finally, the author would like to cite [19]–[21] as related published results.

We will simply use the term “attack” for an adversarial agent generating either non-random or random-like electromagnetic waves to confuse the radar sensor. Undetected attacks may cause false object detection, false object classification, not detecting an existing object, and/or increased background noise. We propose a new radar sensor system architecture, derive analytical upper bounds of false attack alarms, and prove that these can be made as small as possible by proper selection of radar parameters and RF hardware. Experimental setup, testing and verification is presented as an additional supporting evidence, and not as the main evidence. Although we have recorded radar data for the following cases:

- A pedestrian walking in front of a radar
- A vehicle is driven towards the radar
- Attack signal generator stationary
- Attack signal generator on a secondary moving vehicle

there are millions of different real-life scenarios that we cannot generate because of lack of time and resources. Therefore, experimental results should be interpreted as just as experimental data to provide additional evidence.

B. MIMO RADARS

Most millimeter wave radars, including the ones shown in Fig. 2, have multiple transmit (TX) and receive (RX) antennas. This allows radars to transmit and receive in different directions without mechanically rotating radar sensor, and also helps to achieve narrower beam-width. Although in our theoretical analysis there is only a single transmit and single receive antenna, our results apply to all multiple TX and multiple RX antenna radars (MIMO radars) as well. Because, mathematically a MIMO system can be viewed as a single TX single RX system with narrower beam-widths, and dynamically reconfigurable directivity. For the experimental measurements, we have considered only the forward looking direction, but we would like to reiterate that experimental data is used only as a secondary supporting evidence.

C. ORGANIZATION OF THE PAPER

This paper is organized as follows: In Section II, we present the proposed system architecture, and in Section III, we outline our mathematical model. In Section IV, we summarize our main theoretical results which are basically analytical bounds on $P_F$ and $P_M$, and in Section V we use real

---

**FIGURE 1.** Attack resilient AV radar block diagram: Attack Detector DSP detects attacks, and AV Sensor DSP detects/classifies objects.

**FIGURE 2.** Experimental setup: The mm-wave radar on the left is the attack generator, and the one on right is the radar sensor. The green FPGA board is used for raw data capture.
II. SYSTEM ARCHITECTURE

The block diagram shown in Fig. 1 provides a high level overview of the proposed system architecture. For the sake of simplicity, the 90-degree phase shifter before the mixer, the low pass filter after the mixer, and separate analog to digital converters (ADC) for real and imaginary parts are not shown. This FMCW radar system has a voltage controlled oscillator (VCO), and two digital signal processing (DSP) subsystems, one for attack detection, which is addressed in this paper, and another one for the 2D FFT based range-velocity heatmap generation, and the subsequent object detection and classification. Attack detection DSP can generate both false attack alarms, and may miss (not detect) certain attacks. In the proposed system, effects of all noise sources and equipment imperfections are represented by an equivalent noise term, \( a \), and the effect of the attack signal is represented by an equivalent disturbance term, \( a \), see Fig. 1.

A. ATTACK MODEL

The attack signal is modeled as an equivalent additive disturbance, \( a \), just after the ADC stage as shown in Fig. 1. The equivalent attack signal \( a \) is not assumed to be a random process, instead we adopt a worst case approach and derive upper bounds on \( P_F \) and \( P_M \) which are valid for any attack signal. Basically, we mathematically prove that all attacks with \( yy \) above a certain threshold are almost always detected, false attack alarm rate is almost zero, and experimentally demonstrate that weak power attacks below the detection threshold will look like additional background noise.

The radar sensor is assumed to have a physically uncloneable function (PUF) based binary random number generator [22], [23], which is used to determine the sign of the slope of the chirp signal. We assume that, no matter which algorithm is used by the adversarial agent, these PUF generated random numbers will look really random for the adversarial agent. This is the most fundamental assumption made in this paper. In principle, the strength of the proposed attack resilient radar system is limited by the predictability of PUF based random number generators. However, to the best of author’s knowledge this is a known difficult prediction problem, and hence justifies the resilience of the proposed system.

Overall system performance is measured by two critical probabilities, \( P_F \), false attack alarm rate, and \( P_M \), probability of not detecting an attack. System performance can also be characterized by the total risk, which is \( P_F + P_M \). Note that, our false attack alarm rate, \( P_F \), is not the probability of the AV Sensor DSP detecting a target when there is no real target, which is also called as a false alarm in radar literature. Similarly, our \( P_M \) is not the probability of the AV Sensor DSP not detecting a target when there is a real target, which is also called as miss in radar literature. The results of this paper are valid for both deterministic and random attack scenarios, including spoofing and interference. To the best of author’s knowledge, all previously published papers use either simulated attacks and/or specific experiments for testing and verification. One of our differentiating contributions is the theoretical analysis of \( P_F \) and \( P_M \), derivation of analytical upper bounds, and proving that the risk \( P_F + P_M \) can be made as small as possible.

B. RADAR DETAILS

Theory of automotive radars are discussed in [24], [25]. Basically, an AV radar outputs a new measurement result (range-velocity map) in every \( T_m \) seconds, which is also called as the measurement cycle or frame. Within a single measurement cycle, we have \( N_r \) non-overlapping and uniformly separated chirp windows \( I_0, \ldots, I_{N_r-1} \), each of which having \( t_c \) seconds duration. The time between the beginning of two consecutive chirp windows is \( t_d \), and usually \( t_d > t_c \).

In Fig. 3, we see a complete frame for \( N_r = 6 \), and some of the chirps have positive frequency slope and shown in blue color, and others have negative frequency slope and shown in red color. In a standard FMCW AV radar, chirps do not have this random frequency slope pattern. Instead, they are either all positive, or all negative, or in alternating positive-negative format. However, the proposed random slope pattern is enough to achieve a high degree of attack resilience.

![FIGURE 3. A complete frame: A sequence of chirps with PUF generated random slope. Chirps with blue color have positive frequency slope, whereas the ones with red color have negative frequency slope.](image)

The ADC frequency is \( f_s \) and a total of \( N_s = t_c f_s \) samples are acquired during each chirp window. Basically, for every measurement cycle or frame, we have a \( N_r \times N_s \) complex matrix as the measured data. For the Texas Instruments AWR1642 series 77 GHz radars used in this work, default parameters values are \( T_m = 40 \text{ ms} \), corresponding to 25 frames per sec, \( N_r = 128 \) corresponding to 128 chirps per frame, and \( t_d = 160 \mu \text{s}, t_c = 25.6 \mu \text{s} \) corresponding to \( N_r = 256 \) samples at \( f_s = 10 \text{ MHz} \) ADC. The received I/Q data from a single chirp can be viewed as a complex \( 1 \times N_s \)
row vector, and all of the row vectors corresponding to the chirps in a single frame can be viewed as a $N_r \times N_c$ complex matrix. The AV radar DSP basically computes the 2D FFT of this complex matrix to generate the range-velocity heatmap. Usually, windowing, zero padding, crop/resizing, normalization, and optional background subtraction are applied as well, see [7] and references therein. The AWR1642 radar has indeed 4 receive antennas, and for the numerical examples based on experimental data we use the forward looking direction by simply averaging all receive antenna data to form the $N_r \times N_c$ complex matrix.

For all experiments, 2 GHz radar bandwidth is used resulting 7.5 cm range resolution. Although the AWR1642 has a total of six antennas (2 TX and 4 RX), and capable of beamforming, the author used a single TX single RX configuration just to prove that the proposed attack detector works even for minimalistic radar designs. The author used Python for coding the detector logic, but AWR1642 has a DSP processor, and an ARM processor which can be used for coding the detector logic. Both proposed detectors require only polynomial computation time, and both have mostly standard vector operations, norms, and loops which are not that complicated to code using a C compiler.

III. MATHEMATICAL MODEL

In this section, we summarize our notation and the mathematical model. A Gaussian distributed random variable with mean $\mu$ and standard deviation $\sigma$ is denoted by $N(\mu, \sigma)$, i.i.d. stands for independent identically distributed, and $\| \cdot \|_2$ is the Euclidean norm. There is one important notational convention adopted throughout the rest of this paper: A discrete-time signal of length $N_s$, which is basically a function defined on $\{0, \cdots, N_s - 1\}$, and its column vector representation, which is a $N_s \times 1$ vector/matrix, are considered as the same. Basically, we use both the function notation, and the column vector notation at the same time; if a matrix and a signal appear one after the other in an algebraic expression, column vector notation at the same time; if a matrix and a signal appear one after the other in an algebraic expression, column vector notation is implied, but if square brackets are used, function representation is implied. Furthermore, we use the term almost always if a probability is close to 1, and almost never if a probability is close to 0.

We define $S$ as the set of all functions, $\rho$, from the domain set $\{0, \cdots, N_r - 1\}$ to the range set $\{-1, +1\}$. Elements of $S$ are called signature functions, and the random signature generator block shown in Fig. 1 selects a new random signature, $\rho \in S$, for each measurement cycle or frame. Furthermore, $\rho(k) = +1$ means that during the interval $I_k$ the VCO is driven by a positive slope ramp signal resulting an up-chirp (Chirp with positive frequency slope), and $\rho(k) = -1$ means that during the interval $I_k$ the VCO is driven by a negative slope ramp signal resulting a down-chirp (Chirp with negative frequency slope). Basically, the sign of the chirp used during the interval $I_k$ is equal to $\rho(k)$, and $\rho(k)$ is a modeled as a ±1 valued random variable with equal probabilities for each. In an FMCW radar system, velocity resolution is equal to $\delta v = \frac{\lambda}{2N_r\lambda_d}$, where $\lambda$ is the wavelength, and $N_r$ is usually chosen as a large value to achieve good velocity resolution. The set $S$ will be a quite large with $2^{N_r}$ elements. For the Texas Instruments AWR1642 series 77 GHz radars used in this work, the default $N_r$ is 128, which will result a set $S$ with more than $10^{38}$ elements.

The physical signal received by the RX antenna goes through all of the RF and analog blocks, and at the output of the ADC stage it is called as the received data. This complex signal is denoted by $R_k[m]$, where $k = 0, \cdots, N_r - 1$ is the index of the chirp interval $I_k$, and $m = 0, \cdots, N_r - 1$ is the sample index within the chirp window. Furthermore,

$$R_k[m] = n_k[m] + x_k[m] + a_k[m],$$

where $n_k[\cdot]$ is the equivalent noise term, $x_k[\cdot]$ is the term representing reflections from objects, and $a_k[\cdot]$ is the equivalent disturbance term representing the attack signal, see Fig. 1. We assume that $n_k[m]$’s are all i.i.d. $N(0, \sigma)$ with circular symmetry. The AV radar system generates its own random signature function $\rho \in S$ for each frame, but does not know $a_k[m]$’s. The adversarial agent is assumed to have full information about all of the details of the radar DSP blocks, except for the randomly generated signature $\rho$. Because of the rapid switching pattern of the AV radar, and microsecond level chirp durations, we assume that the adversarial agent cannot determine the value of $\rho(k)$ and use it for an intelligent attack while the $k^{th}$ chirp is still active. This is a fundamental assumption made in this paper.

In the following, we list our assumptions in a more rigorous format, and provide an engineering interpretation.

Independence Assumptions: The radar sensor generates $\rho(k)$ without knowing the value of $a_k$. The adversarial agent generates $a_k$ without knowing the value of $\rho(k)$. The noise $n_k$ is independent of both $\rho(k)$ and $a_k$. Noise terms, $n_k[m]$, are i.i.d. $N(0, \sigma)$.

For any signal, $s[m]$, defined for $m = 0, \cdots, N_s - 1$, we define the RMS level as

$$\|s\|_r := \sqrt{\frac{1}{N_s} \sum_{m=0}^{N_s-1} s^2[m]} = \|s\|_2 \sqrt{N_s},$$

where, $\| \cdot \|_2$, is the Euclidean norm. Note that, this is a measure of signal strength, and is also a norm. Furthermore, the signal power is equal to $\|s\|_r^2$.

Environmental Assumptions: The frame duration is short, and changes in objects’ locations are relatively small within a single frame. This is a fundamental assumption for 2D FFT based range-velocity heat-maps, and moving target detection. For attack resilience, we impose the following weaker environmental condition

$$\|x_k\|_r - \|x_k\|_r < \epsilon_0(\|x_k\|_r + \|x_\ell\|_r),$$

(1)

holds for all $k, \ell = 0, \cdots, N_r - 1$. In a typical case, $\epsilon_0$ will be small. Given an upper bound for the velocity of moving objects in an environment, reducing the frame duration will force $\epsilon_0$ to decrease. Basically, this assumptions can be interpreted as “the frame duration should be selected short
for the given relative speeds of objects in that particular environment”. Later in the paper, both this assumption and the value of $\epsilon_0$ is discussed using real experimental data.

Attack Signal Level (Relative Threshold): The attack signal average RMS level is above the threshold level defined by the following inequality

$$
\left( \frac{1}{N^r} \sum_{k=0}^{N^r-1} \| a_k \|_r \right) \geq C_1 \sigma + \epsilon_0 C_2 \left( \frac{1}{N^r} \sum_{k=0}^{N^r-1} \| x_k \|_r \right), \quad (2)
$$

where

$$
C_1 = \frac{8 + 2 \epsilon_0}{1 - \epsilon_0}, \quad \text{and} \quad C_2 = 4 \left( 1 - \epsilon_0 \right)^2.
$$

Basically, it will be proved that if the attack signal is above a minimal threshold relative to the noise level and radar signal strength, then the proposed system will detect the attack with a very high probability. These constants $C_1$ and $C_2$ are used to define this minimal threshold. Because of the existence of both $\sigma$ and $\| x_k \|_r$, $k = 0, \ldots, N^r - 1$, terms, (2) is not an absolute threshold, but it is a relative threshold. If the reflected signal $x_k$’s RMS levels are well above the noise level (i.e. SNR is high), then the second term in (2) is expected to be dominant. In a typical case, $\epsilon_0$ will be small, and we can write $C_1 \approx 8$, and $C_2 \approx 2$. The rationale behind this assumption is the following: Attack signals which are small compared to noise, $n_k$, and reflected signal, $x_k$, will have a relatively small effect on range-velocity heat-map generation, and object detection and identification. Basically, what is proved in this paper can be summarized as follows: Attack signals above the minimal level defined in (2) are detected almost always with almost no false alarms. However, attack signals below this minimal level may or may not be detected that easily. For a good system design, this minimal level defined in (2) should be reasonably small, because in the worst case, attack signals below this minimal level may be completely missed, and missed attack signals should have very small signal levels to make sure that their effect on Radar DSP is small and comparable to system noise. This observation implies that, we need to have a low noise analog hardware so that the $\sigma$ is small, and we have a short frame duration so that $\epsilon_0$ is small as well. Note that, short frame durations require both better analog and faster digital subsystems.

In the rest of this section, we will prove two useful lemmas. First, we define time-domain projection matrices $P_{+1}$ and $P_{-1}$ as

$$
P_{+1} = F_{N^r}^{-1} \cdot \Pi_{+1} \cdot F_{N^r}, \quad P_{-1} = F_{N^r}^{-1} \cdot \Pi_{-1} \cdot F_{N^r},
$$

where $F_{N^r}$ is the $N^r \times N^r$ matrix with $(F_{N^r})_{k, \ell} = e^{-2\pi ik\ell/N^r}$, i.e. the matrix representation of the $N^r$-point DFT operation, $F_{N^r}^{-1}$ is the matrix representation of the $N^r$-point inverse DFT operation, and $\Pi_{+}$ and $\Pi_{-}$ are frequency-domain projection matrices defined as

$$
\Pi_{+} = \begin{bmatrix} 0_{N^r/2} & 0_{N^r/2} \\ 0_{N^r/2} & 0_{N^r/2} \end{bmatrix}, \quad \Pi_{-} = \begin{bmatrix} 0_{N^r/2} & 0_{N^r/2} \\ 0_{N^r/2} & I_{N^r/2} \end{bmatrix},
$$

where $0_{N^r/2}$ and $I_{N^r/2}$ are $(N^r/2)$-dimensional zero and identity matrices respectively. For a given discrete time signal of length $N^r$, we define $s^+ = P_{+1} s$ and $s^- = P_{-1} s$. Note that $s = s^+ + s^-$, and one can consider $s^+$ as the component of $s$ with only non-negative frequencies, and $s^-$ as the one with only negative frequencies. Furthermore, by simple matrix algebra it follows that $< s^+, s^- > = 0$, i.e. as column vectors $s^+$ and $s^-$ are orthogonal. This decomposition will play a critical role in detector design, and theoretical proofs. We now present two lemmas which will be used repeatedly throughout the rest of this paper.

**Lemma 1:** Projected noise components, $(F_{N^r} n_k^{\pm}[m])$, are i.i.d. $N(0, \sigma \sqrt{N^r})$ for $m \in [0, (N^r - 1)/2]$, and 0 otherwise. Similarly, $(F_{N^r} n_k^{\pm}[m])$ are i.i.d. $N(0, \sigma \sqrt{N^r})$ for $m \in [(N^r + 1)/2, N^r - 1]$, and 0 otherwise.

**Proof:** By definition,

$$
E\{F_{N^r} n_k^{\pm}[m] \cdot (F_{N^r} n_{\ell}^{\pm}[m])^*\} = E\{\Pi_{+1} \cdot F_{N^r} \cdot n_k \cdot n_{\ell}^* \cdot (F_{N^r})^* \cdot \Pi_{+1}^*\},
$$

which reduces to

$$
\Pi_{+1} \cdot F_{N^r} \cdot \sigma^2 \delta_{k,\ell} n_{\ell} \cdot (F_{N^r})^* \cdot \Pi_{+1}^* = \delta_{k,\ell} n_{\ell} \sigma^2 \Pi_{+1}.
$$

This proves the claim for $(F_{N^r} n_k^{\pm}[m])$, $m \in [0, (N^r - 1)]$. Proof of the other part is similar. □

**Lemma 2:** $P_{\rho(k)} x_k = x_k$ and $P_{-\rho(k)} x_k = 0$.

**Proof:** Consider an FMCW radar system with maximum range selected properly to avoid aliasing. In that case, if the VCO is driven with a positive slope ramp signal, the complex representation of the mixer output will have an increasing phase and positive frequency components. On the other hand, if the VCO is driven with a negative slope ramp signal, the complex representation of the mixer output will have a decreasing phase, and negative frequency components. Since all of the hardware imperfections are assumed to be captured by the equivalent noise term, we get $P_{\rho(k)} x_k = x_k$ and $P_{-\rho(k)} x_k = 0$. □

**IV. MAIN RESULTS**

In this section, we define the attack detectors used in the Attack Detector DSP block shown in Fig. 1, and summarize the bounds on false attack alarm rate, $P_F$, missed attack alarm rate, $P_M$, and total risk $P_F + P_M$.

**A. DEFINITIONS OF ATTACK DETECTORS**

We define detectors $d_1(k, \eta)$, and then $D_1(p, \eta)$ by their alarm trigger conditions. For a given signature function $\rho \in \mathcal{S}$, and interval index $k \in [0, \ldots, N^r - 1]$, the detector $d_1(k, \eta)$ generates an alarm iff

$$
\| P_{-\rho(k)} R_k \|_r \geq \eta \sigma, \quad (3)
$$

and the ratio of the left hand side to the right hand side is called the normalized detector output for $d_1(k, \eta)$. For a given measurement cycle or frame, if at least $p$ of these detectors generate an alarm, then the detector $D_1(p, \eta)$ generates an alarm. We will use the shorthand notation, $D_1$, without any parenthesis and arguments for the specific detector $D_1(1, 1)$. 


The normalized detector output for \( D_1 \) will be maximum of normalized detector outputs of \( d(k, 1) \)'s for \( k = 0, \cdots, N_r - 1 \).

For a given measurement cycle or frame, the detector \( D_2(\epsilon, \eta) \) generates an alarm iff
\[
\| P_{p(\ell)} R_k \| r - \| P_{p(\ell)} R_{\ell} \| r \geq \eta \sigma + \epsilon (\| P_{p(\ell)} R_k \| r + \| P_{p(\ell)} R_{\ell} \| r),
\]

is true for some \( k, \ell \in \{ 0, \cdots, N_r - 1 \} \). For this equation, the maximum of the ratio of the left hand side to the right hand side is called normalized detector output for \( D_2(\epsilon, \eta) \), and the maximum is over \( k, \ell \in \{ 0, \cdots, N_r - 1 \} \). In the following, we will use the shorthand notation, \( D_2 \), without any parenthesis and arguments for the specific detector \( D_2(\epsilon_0, 2 + 2 \epsilon_0) \).

In summary, for a given measurement cycle or frame, the detector \( D_1 \) generates an alarm iff
\[
\exists k \in \{ 0, \cdots, N_r - 1 \}, \| P_{-p(\ell)} R_k \| r \geq \sigma, \tag{4}
\]
and the detector \( D_2 \) generates an alarm iff
\[
\exists k, \ell \in \{ 0, \cdots, N_r - 1 \}, \| P_{p(\ell)} R_k \| r - \| P_{p(\ell)} R_{\ell} \| r \geq (2 + 2 \epsilon_0) \sigma + \epsilon_0 (\| P_{p(\ell)} R_k \| r + \| P_{p(\ell)} R_{\ell} \| r). \tag{5}
\]

The inequalities (4) and (5) are basically the formal definitions of the detectors \( D_1 \) and \( D_2 \) proposed in this work, and they both measure the level of an anomaly which is expected to be very low when there is no attack. Note that, both detectors \( D_1 \) and \( D_2 \) have polynomial time computational complexity, and both are easy to implement as software inside the Attack Detector DSP block, see Fig. 1.

**B. ANALYTICAL BOUNDS ON \( P_F, P_M, \) AND THE RISK**

Consider an AV radar system using detectors \( D_1 \), and \( D_2 \) as defined above. The false attack alarm rate, \( P_F \), satisfies
\[
(1 - \epsilon_1) N_r (0.906)^{N_r} \leq P_F \leq 2N_r (0.927)^{N_r}, \tag{6}
\]
where the upper bound is always valid, but for the lower bound part \( \epsilon_1 = 0.5(N_r - 1)(0.906)^{N_r} \) and \( N_r \) is assumed to be a multiple of 128. These bounds imply that, for a low \( P_F \), a high \( N_r \) value is necessary, and sufficient provided that the \( N_r/N_s \) ratio is bounded from below and above by positive numbers.

The probability of not detecting an attack, \( P_M \), is bounded by
\[
P_M \leq (0.500)^{N_1} + (2N_r + 1)(0.927)^{N_r}. \tag{7}
\]

Note this upper bound (7) is valid for any attack signal above the threshold defined in (2). In other words, (7) is not simply a bound for the expected value of the probability of not detecting an attack, it is indeed a worst case deterministic upper bound. It is also clear that, for a low \( P_M \), a high \( N_r \) and a high \( N_s \) will be sufficient, provided that the \( N_r/N_s \) ratio is bounded from below and above by positive numbers.

As \( N_r, N_s \to \infty \) with the ratio of \( N_r/N_s \) kept fixed, the risk function, \( P_F + P_M \), converges to 0 exponentially. This can be interpreted as the system being asymptotically perfect, i.e. the risk function, \( P_F + P_M \), can be made smaller than any given threshold by choosing large enough \( N_r \) and \( N_s \).

**C. CFAR APPROACH FOR THRESHOLD SELECTION**

In this paper, we do not address perception or target detection algorithms, and the thresholds used in such algorithms. We only consider attack detectors and their thresholds. However, we would like present a short discussion on constant false alarm rate (CFAR) based threshold selection for target detection and also a CFAR based threshold selection for cyberattack detection.

CFAR for target detection: All attacks which are below the detection threshold will result a slight increase in background noise as shown in all of the examples presented in Section V. Therefore, if a CFAR type algorithm is used for target detection, weak attacks below the detection threshold will result a slight increase in CFAR target detection threshold (Compare Fig. 10 and Fig. 11 of Section V). For such weak attacks below the detection threshold, targets will be slightly more difficult to identify, and this may cause certain very weak targets to be missed/undetected. Although it is always possible to find edge cases, the proposed attack detectors are sensitive, Fig. 10 and Fig. 11 are very similar, and the increase in CFAR target detection threshold due to weak undetected attacks will be small. In other words, undetected weak attacks will definitely have some negative effect on target detection, but these will be really small.

CFAR for cyberattack detection: As in all hypothesis testing problems, there is always a trade-off between \( P_F \), probability of generating a false cyberattack alarm, and \( P_M \), probability of not detecting (missing) a cyberattack. Changing an attack detector threshold either increases \( P_F \) and decreases \( P_M \), or decreases \( P_F \) and increases \( P_M \). In most applications, both \( P_F \) and \( P_M \) needs to be reasonably small, and hence the selection of a good attack detector threshold is important. Note that, we propose analytic detector thresholds which result very small \( P_F \) and \( P_M \) values. Therefore, an adaptive threshold which may result an even smaller \( P_F + P_M \) value seems to be not so critical from an engineering perspective. In other words, the proposed analytic thresholds already perform really quite well, and result in very small \( P_F \) and \( P_M \) values.

**V. NUMERICAL EXAMPLE USING REAL RADAR DATA**

In this section, we will present a numerical example to illustrate the main concepts. Although theoretical proofs will be presented later in the paper, we use real experimental data to illustrate how \( \epsilon_0, C_1, \) and \( C_2 \) can be computed, experimentally demonstrate that the detectors indeed work, and also experimentally demonstrate that weak attack signals below the detection threshold will appear as additional background noise.

Our test setup shown in Fig. 2 has two 77 GHz Texas Instruments AWR1642 radars connected to a laptop computer, one of them is acting as an attack/interference generator, and the
other is used as a regular AV radar sensor. Although we have done multiple different types of tests, and we will use the following two cases for illustration purposes: (E1) A pedestrian approaching the AV radar sensor, Fig. 4, and (E2) A vehicle approaching the AV radar sensor, Fig. 5. For real-time data acquisition, the software developed in [7] is used. For the default parameters $N_r = 128$, and $N_s = 256$, the bound (6) implies that the probability of false attack alarm, $P_F$, is bounded by $9.57 \times 10^{-7}$. Similarly, the bound (7) implies that the probability of not detecting an attack, $P_M$, is bounded by $9.61 \times 10^{-7}$. Note that, this upper bound on $P_M$ is valid under the assumption that the attack signal level is above the minimal threshold defined by (2), which is discussed in the next paragraph.

Note that $x_k$, $a_k$, and $n_k$’s are indeed unit-less 16-bit values after the ADC, see Fig. 1. By turning off transmitters, and reading only system noise, we observed that $\sigma \approx 10$. To determine $\epsilon_0$, we used the measurement results of the experiments E1 and E2, and observed that $\epsilon_0 = 0.06$ and $\epsilon_0 = 0.09$ will work. As expected, for an environment with larger size objects moving at higher speeds, $\epsilon_0$ turns out to be larger. Given a bound on the size of moving objects, their maximum speed, and the frame duration, a comprehensive set of experiments and measurements are required for the selection of $\epsilon_0$, however we have used only 10 different recorded experimental data to select a large enough $\epsilon_0$. We would like to reiterate that reducing the frame duration will reduce $\epsilon_0$. We simply choose a slightly more conservative (larger) value of $\epsilon_0 = 0.1$, and the $C_1$ and $C_2$ values in (2) will be 9.11 and 4.93 respectively. The relative threshold for the attack signal level will be

$$\left( \frac{1}{N_r} \sum_{k=0}^{N_s-1} \|a_k\|_\rho \right) \geq 9.11\sigma + 4.93\epsilon_0 \left( \frac{1}{N_r} \sum_{k=0}^{N_s-1} \|x_k\|_\rho \right),$$

which directly follows from (2). Basically, any attack signal above this threshold is almost always detected with almost no false alarms. Negative effects of attacks below this threshold are discussed later in this section.

In Fig. 4, and Fig. 5, radar range-velocity heatmaps are presented in normalized format, and displayed in parula colormap instead of gray-scale. The bright points on the central horizontal line correspond to reflections from the stationary background. An approaching pedestrian appears as a vertical strip, whereas an approaching vehicle appears as bright dot, see [26] for a detailed discussion. For testing and verification of the proposed attack detectors, we have first recorded the AV radar data for experiments E1 and E2. Then, we have turned off the AV radar’s transmitter, activated the second radar unit to emulate an attack, and recorded the signal received by the AV radar’s receiver. This recorded attack signal is later scaled with different values, and added to the initial AV radar data using the equation

$$R_k = A_k + sB_k, \quad k = 0, \cdots, N_r - 1,$$

where $s$ is the scaling factor, $A_k$ is the recorded radar signal when there is no attack (i.e. the second radar, the attack radar, is off), and $B_k$ is the recorded attack signal when the second radar, the attack radar, is active, but the AV radar’s transmitter is off. For illustration purposes, we are using a single deterministic attack scenario, but a larger set of random and deterministic attack signals can be used for a more comprehensive experimental testing and verification of the theoretical results proved in this paper.

In Fig. 6, normalized detector outputs of the detectors $D_1$ and $D_2$ are plotted with respect to the attack signal scaling factor $\sigma$. This plot is generated using a two randomly generated signature functions $\rho \in S$, but as observed in [5], generating different random signature functions results similar curves. Note that, normalized detector outputs were defined in the previous section.

As the attack signal level is increased, attack detectors $D_1$ and $D_2$ output larger and larger values, and when the thresholds defined in (4) and (5) are reached, an alarm is generated. But, weak attack signals do not trigger an alarm. A natural question to ask is the following: What are the negative effects of an attack signal which is just below this detection threshold, i.e. worst case effects of undetected weak attacks. In Fig. 7, we see the effects of an attack signal just below this detection threshold. Compared to Fig. 4, there is some
noise but the vertical strip corresponding to the approaching pedestrian is still clearly visible. For a standard FMCW radar with no attack detectors and attack resilience, the same attack signal will result a more noisy radar range-velocity heatmap shown in Fig. 8. Because, the randomized signature function basically swaps the energy of $x_k$’s randomly between left and right portions of the frequency spectrum, which reduces the negative effects of undetected attack signals. Attack signals which are above the threshold will have much worse effects on a standard FMCW radar. The proposed detector system detects such attacks almost always with almost no false alarms.

### A. MULTI-TARGET EXPERIMENT

In this subsection, we use the same radar system for another experiment with multiple targets. In this new experiment, we have two pedestrians and two vehicles as shown in Fig. 9 and Fig. 10. Both Fig. 9 and Fig. 10 represent the same scene, same data with no attack signal. In Fig. 9, all targets are marked and to achieve maximum possible zoom level in double-column format, the colorbar and the axis labels are deleted. In Fig. 10, we have the same scene, same data, but targets are not marked on the image, i.e. the image is untouched, but the colorbar and axis values are shown so that comparison can be made with Fig. 11, and Fig. 12.
the proposed attack detector will detect the attack and reject the radar image, i.e. will not forward it to the perception algorithm. However, if no attack resilience is implemented, stronger attack signals may result major errors in the perception algorithm output, for example pedestrians detection may fail.

VI. FALSE ATTACK ALARMS

In this section, we will first analyze the false attack alarm rate of the detector, \( D_1 = D_1(1, 1) \), and derive analytical upper and lower bounds for \( P_F(D_1) \). Basically, we will prove that

\[
(1 - \epsilon_1)N_r(0.906)N_s \leq P_F(D_1) \leq N_r(0.927)^{N_s} \tag{8}
\]

where the upper bound is always valid, but for the lower bound part \( \epsilon_1 = 0.5(N_r - 1)(0.927)^{N_s} \) and \( N_s \) is assumed to be a multiple of 128. In a typical case, \( N_s \) and \( N_r \) will be large, and \( \epsilon_1 \) will be small. For the Texas Instruments AWR1642 radar used in this work, the default parameters are \( N_s = 256 \), \( N_r = 128 \), which results \( \epsilon_1 = 2.37 \times 10^{-7} \), false attack alarm probability upper bound \( 4.79 \times 10^{-7} \), and lower bound \( 1.35 \times 10^{-9} \).

In the last subsection, we will analyze the false attack alarm rate of the detector \( D_2 = D_2(\epsilon_0, 2 + 2\epsilon_0) \), and derive the upper bound

\[
P_F(D_2) \leq N_r(0.927)^{N_s}. \tag{9}
\]

It is easy to see that, the inequalities (8) and (9) imply (6).

A. ANALYSIS OF THE DETECTOR \( d_1(k, 1) \)

In this subsection, we focus on the false attack alarm rate of \( d_1(k, 1) \). Consider a \( k \in \{0, \ldots, N_s - 1\} \), and assume that \( \rho(k) = +1 \) (The proof the \( \rho(k) = -1 \) case will be similar). If there is no attack signal present, \( P_{-1}R_k = P_{-1}n_k + P_{-1}x_k + P_{-1}a_k \), will reduce to \( P_{-1}R_k = n_k^r \) by Lemma 2. Therefore, \( \|P_{-1}R_k\|_2^2 \) will be equal to \( \|F_Nn_k^r\|^2/N_s^2 \). By using Lemma 1, we see that the random variable \( \sum_{m=0}^{N_s-1} |(F_Nn_k^+)[m]|^2 \) is a sum of squares of \( N_s/2 \) i.i.d. \( N(0, \sigma \sqrt{N_s}) \) random variables. Namely, it is a \( \chi^2 \) type random variable with \( N_s \) degrees of freedom. If \( \rho(k) = -1 \), then we have \( \|P_{+1}R_k\|_2^2 = \|F_Nn_k^t\|^2/N_s^2 \), and \( \sum_{m=0}^{N_s-1} |(F_Nn_k^-)[m]|^2 \) will be a sum of squares of \( N_s/2 \) i.i.d. \( N(0, \sigma \sqrt{N_s}) \) random variables, i.e. will be a \( \chi^2 \) type random variable with \( N_s \) degrees of freedom.

The following lemma is basically an upper bound for the false attack alarm rate of the detector \( d_1(k, 1) \).

**Lemma 3:** For any \( k \in \{0, \ldots, N_s - 1\} \),

\[
Pr \left( \sum_{m=0}^{N_s-1} |(F_Nn_k^+)[m]|^2 \geq N_s^2\sigma^2 \right) \leq (2/e)^{N_s/4}, \tag{10}
\]

\[
Pr \left( \sum_{m=0}^{N_s-1} |(F_Nn_k^-)[m]|^2 \geq N_s^2\sigma^2 \right) \leq (2/e)^{N_s/4}, \tag{11}
\]

and \( P_F(d_1(k, 1)) \leq (2/e)^{N_s/4} \).
Proof: We will only prove (10) as the proof of (11) is similar. The upper bound on $P_F(d_l(k, 1))$ will be proved at the end.

Consider a random variable $Y = (1/h^2) \sum_{m=0}^{q-1} y_m^2$ with $y_m$’s i.i.d. $N(0, h)$ random variables. The Chernoff bound [27] implies that

$$Pr(Y \geq a) \leq e^{-sa/M_Y(s)} \text{ for } s \in (0, 1/2)$$

where $M_Y(s) = (1 - 2s)^{-q/2}$ is the moment generating function of $Y$. For a given $s \in (0, 1/2)$, let $z = 1/(1 - 2s)$ and $a = zq$. The Chernoff bound reduces to

$$Pr\left(\left(1/h^2\right)\sum_{m=0}^{q-1} y_m^2 \geq zq\right) \leq z^{q/2} e^{(1-z)h/2}.$$  

Finally, if we substitute $z = 2$, we get

$$Pr\left(\sum_{m=0}^{q-1} y_m^2 \geq 2qh^2\right) \leq (2/e)^{q/2}.$$  

Therefore,

$$Pr\left(\sum_{m=0}^{N/2-1} |(F_{N_s} n_k^s)[m]|^2 \geq 2(N_s/2)\sigma^2 N_s\right) \leq (2/e)^{N_s/4},$$

which implies the inequality (10) using Lemma 1. The proof of (11) is similar.

Note that $\|P_{-1} R_k\|_r = \|F_{N_s} n_k^r\|_r/N_s$ when $\rho(k) = +1$, and $\|P_{+1} R_k\|_r = \|F_{N_s} n_k^r\|_r/N_s$ when $\rho(k) = -1$. Therefore, (10) is an upper bound for the false attack alarm rate, $P_F(d_l(k, 1))$, when $\rho(k) = +1$, and (11) is an upper bound for $\rho(k) = -1$. 

B. UPPER BOUND FOR $P_F(D_1)$

The detector $D_1$ generates an alarm iff any of the detectors $d_l(k, 1), k = 0, \cdots, N_r - 1$, generates an alarm. Hence

$$P_F(D_1) \leq \sum_{k=0}^{N_r-1} P_F(d_l(k, 1)) \leq N_r(2/e)^{N_s/4} < N_r(0.927)^{N_s},$$

which implies the upper bound in (8).

C. LOWER BOUND FOR $P_F(D_1)$

If $Y_q = (1/h^2) \sum_{m=0}^{q-1} y_m^2$ with $y_m$’s i.i.d. $N(0, h)$ random variables, then $Y_q$ will be a $\chi^2$ random variable. It is trivial to see that, if $q = Lq_0$ with $L, q_0$ positive integers, then

$$Pr(Y_q \geq 2q) \geq Pr(Y_{q_0} \geq 2q_0)^L.$$  

Therefore, if $q$ is a multiple of 64, i.e. $q = 64 L$ with $L$ a positive integer, then

$$Pr(Y_q \geq 2q) \geq Pr(Y_{64} \geq 128)^L = (Pr(Y_{64} \geq 128)^{1/64})^L.$$  

By using a $\chi^2$ calculator, we get

$$Pr(Y_{64} \geq 128)^{1/64} \geq 0.822.$$  

Therefore, if $q$ is a multiple of 64, $Pr(Y_q \geq 2q) \succeq (0.822)^q$. This observation will be used in the following proof.

Lemma 4: If $N_s$ is a multiple of 128, then $(0.906)^{N_r} \leq P_F(d_l(k, 1))$ and

$$(1 - \epsilon_1)N_r(0.906)^{N_s} \leq P_F(D_1)$$

where $\epsilon_1 = 0.5(N_r - 1)(0.927)^{N_s}$.

Proof: Consider $k \in \{0, \cdots, N_r - 1\}$, and assume that $\rho(k) = +1$ (The proof the $\rho(k) = -1$ case is similar). The detector $d_l(k, 1)$ generates a false alarm if

$$\|P_{-1} R_k\|_r^2 = \|F_{N_s} n_k^r\|_r^2/N_s^2 \geq \sigma^2.$$  

We know that $\|F_{N_s} n_k^r\|_r^2/N_s^2$ is a sum of $N_s/2$ i.i.d. $N(0, \sigma)$ random variables. Therefore

$$P_F(d_l(k, 1)) = Pr(\|F_{N_s} n_k^r\|_r^2/N_s^2 \geq \sigma^2) \geq (0.822)^{N_s/2},$$

where for the rightmost inequality, we used the observation made just before the lemma. Since, $(0.822)^{1/2} > 0.906$, we get $(0.906)^{N_r} \leq P_F(D_1)$.

To prove the lower bound on $P_F(D_1)$, we need a separate argument. Let $f$ be the probability that the detector $d_l(k, 1)$ generates a false attack alarm. Note that $f$ is independent of $k$. The probability that none of the detectors, $d_l(k, 1), k = 0, \cdots, N_r - 1$ generates an alarm will be $(1 - f)^{N_r}$, and the probability that at least one generates an alarm will be $1 - (1 - f)^{N_r}$, which is the same as $P_F(D_1)$. Therefore,

$$P_F(D_1) = 1 - (1 - f)^{N_r} \geq N_s f - N_r(1 - f)^{N_r}/2$$

Using $(0.906)^{N_r} \leq f \leq (0.927)^{N_s}$, we get the lower bound of $P_F(D_1)$.

Therefore, we have completed the derivation of both upper and lower bounds in (8).

D. UPPER BOUND FOR $P_F(D_2)$

In this section, we derive an upper bound for the false attack alarm rate of $D_2$. Note that we set the attack signal, $a_k = 0$ for the false attack alarm rate analysis. By definition, the detector $D_2(\epsilon_0, \eta)$ generates an alarm iff

$$\|P_{\rho(k)} R_k\|_r - \|P_{\rho(\ell)} R_\ell\|_r \geq \eta \sigma + \epsilon_0(\|P_{\rho(k)} R_k\|_r + \|P_{\rho(\ell)} R_\ell\|_r),$$

is true for some $k, \ell \in \{0, \cdots, N_r - 1\}$. Note that, we use the shorthand notation, $D_2$, without any parenthesis and arguments for the specific detector $D_2(\epsilon_0, 2 + 2\epsilon_0)$.

An upper bound for $P_F(D_2)$ can be obtained as follows: The environmental assumptions imply

$$\|P_{\rho(k)} R_k\|_r - \|P_{\rho(\ell)} R_\ell\|_r < \epsilon_0(\|P_{\rho(k)} R_k\|_r + \|P_{\rho(\ell)} R_\ell\|_r),$$

for all $k, \ell = 0, \cdots, N_r - 1$. The inequalities (10) and (11) in Lemma 3 imply that, with a probability of at least $1 - N_r(2/e)^{N_s/4}$, we have

$$\|P_{\rho(k)} n_k\|_r < \sigma, \text{ for all } k \in \{0, \cdots, N_r - 1\}.$$  

Therefore, with a probability of at least $1 - N_r(2/e)^{N_s/4}$, we have

$$\|P_{\rho(k)} R_k\|_r - \|P_{\rho(\ell)} R_\ell\|_r < (2 + 2\epsilon_0)\sigma + \epsilon_0(\|P_{\rho(k)} R_k\|_r + \|P_{\rho(\ell)} R_\ell\|_r),$$

Figure 4: If $N_s$ is a multiple of 128, then $(0.906)^{N_r} \leq P_F(d_l(k, 1))$ and

$$(1 - \epsilon_1)N_r(0.906)^{N_s} \leq P_F(D_1)$$

where $\epsilon_1 = 0.5(N_r - 1)(0.927)^{N_s}$.
Note that, this inequality is equivalent to the detector $D_2 = D_2(\epsilon_0, 2 + 2\epsilon_0)$ is not generating a false attack alarm. Therefore, $P_F(D_2) \leq N_r(2/\epsilon)^{N_r/4} \leq N_r(0.927)^{N_r}$, hence the derivation of (9) is complete.

**VII. MISSED ATTACK ALARMS**

In this section, we will consider the probability of an attack not being detected by $D_1$ and $D_2$, and derive analytical bounds on the probability of miss $P_M$.

**Theorem 1:** The probability that both of the attack detectors $D_1 = D_1(1, 1)$, and $D_2 = D_2(\epsilon_0, 2 + 2\epsilon_0)$ miss (do not detect) an attack is bounded by

$$P_M \leq \left(\frac{2}{e}\right)^{N_r/2} + \max\left\{\left(\frac{1}{2}\right)^{N_r}, 2N_r\left(\frac{2}{e}\right)^{N_r/2}\right\}.$$  \hspace{1cm} (12)

In general, $P_M \leq (0.500)^{N_r} + (2N_r + 1)(0.927)^{N_r}$ and in most cases the second term will be dominant.

**Proof:** The prove this result, we will consider three separate cases, but before that, we need a preliminary discussion.

We would like to find an upper bound for the probability of the detector $d_{\rho_1}(k, 1)$ missing an attack. Basically, we will show that, this probability $P_M(d(k, 1)) = Pr(\|P_{-\rho(k)}R_k\|_r < \sigma)$ is small if the attack signal power is large compared to the noise level.

Consider a $\rho \in S$, and $k \in \{0, \cdots, N_r - 1\}$. Without loss of generality, assume that $\rho(k) = 1$ (The proof of the $\rho(k) = -1$ case is similar). Because of Lemma 2, we have $P_{-\rho}R_k = a_{\rho} + n_{\rho}$, and by using the triangle inequality for the norm $\| \cdot \|_r$,

$$Pr(\|P_{-\rho}R_k\|_r < \sigma) \leq Pr(\|n_{\rho}\|_r > \|a_{\rho}\|_r - \sigma).$$

If $\|a_{\rho}\|_r \geq 2\sigma\sqrt{N_r}$, equivalently $\|a_{\rho}\|_r > 2\sigma$, then

$$Pr(\|n_{\rho}\|_r > \|a_{\rho}\|_r - \sigma) \leq Pr(\|n_{\rho}\|_r > \sigma) \leq (2/e)^{N_r/2},$$

where the rightmost inequality follows from Lemma 3. In summary, our preliminary discussion implies that if $P_{-\rho}R_k \|_r \geq 2\sigma$, then the attack will be detected “almost certainly” by $d(k, 1)$.

$$P_M(d(k, 1)) \leq (2/e)^{N_r/4} \leq (0.927)^{N_r}.$$  

Now, we start the proof of the theorem and complete it by analyzing three separate cases.

**Case 1:** This case is characterized by the following condition:

$$\forall \rho \in S, \exists k \in \{0, \cdots, N_r - 1\}, \text{ such that } \|P_{-\rho}a_k\|_r \geq 2\sigma.$$  

Here the $k$ is allowed to depend on the specific $\rho$. Based on the preliminary discussion presented above, we have

$$P_M \leq P_M(d(k, 1)) \leq (2/e)^{N_r/4}. $$

**Case 2:** This case is characterized by the following conditions:

$$\exists \rho_0 \in S, \forall k \in \{0, \cdots, N_r - 1\}, \|P_{-\rho_0}a_k\|_r \leq 2\sigma,$$

and

$$\forall \rho_1 \in S \text{ with } \rho_1 \neq \rho_0, \exists k \in \{0, \cdots, N_r - 1\}, \|P_{-\rho_1}a_k\|_r \geq 2\sigma.$$  

The first one is simply the negation of the condition defining Case 1, and but the second one is different. In this case, for any $\rho_1 \in S$ with $\rho_1 \neq \rho_0$, at least one of the detectors $d_{\rho_1}(k, 1)$ generates an alarm almost always, and hence $D_1 = D_1(1, 1)$ detects the attack almost always. More precisely, the probability of miss is bounded by

$$P_M \leq (1/2)^{N_r} + (1 - (1/2)^{N_r})(2/e)^{N_r/4}$$

where the first term is the probability of the random signature generator generating the signature $\rho_0$, and the second term is an upper bound for the combined probability of miss for all of the other signature functions in $S$.

**Case 3:** This case is characterized by the following conditions:

$$\exists \rho_0 \in S, \forall k \in \{0, \cdots, N_r - 1\}, \|P_{-\rho_0}a_k\|_r < 2\sigma$$

and

$$\exists \rho_1 \in S \text{ with } \rho_1 \neq \rho_0, \forall k \in \{0, \cdots, N_r - 1\}, \|P_{-\rho_1}a_k\|_r < 2\sigma.$$  

The first one is simply the negation of the condition defining Case 1, and the second one is the negation of the extra condition used for the definition of Case 2. Overall, Cases 1, 2, and 3 cover all possible cases.

Without loss of generality, we can assume that $\rho_0$ and $\rho_1$ differ at 0, i.e. $\rho_0(0) \neq \rho_1(0)$, equivalently $-\rho_0(0) = \rho_1(0)$. In this case, we will have

$$P_{\rho_0}(a_0) \leq 2\sigma, \text{ and } P_{-\rho_1}(a_0) \leq 2\sigma.$$  

If the detector $D_2 = D_2(\epsilon_0, 2 + 2\epsilon_0)$, do not generate an alarm, we have

$$\|P_{\rho_0}(R_k)\|_r - \|P_{\rho_0}(R_0)\|_r < (2 + 2\epsilon_0)\sigma + \epsilon_0(||P_{\rho_0}(R_k)\|_r + ||P_{\rho_0}(R_0)\|_r),$$

for all $k \in \{0, \cdots, N_r - 1\}$. Our assumptions about the environment implies

$$\|P_{\rho_0}(x_k)\|_r - \|P_{\rho_0}(x_0)\|_r < \epsilon_0(||P_{\rho_0}(x_0)\|_r + ||P_{\rho_0}(x_k)\|_r).$$

By using the last two inequalities, we get

$$\|P_{\rho_0}(a_k + n_k)\|_r - \|P_{\rho_0}(a_0 + n_0)\|_r \leq (2 + 2\epsilon_0)\sigma + \epsilon_0(||P_{\rho_0}(R_k)\|_r + ||P_{\rho_0}(R_0)\|_r)$$

$$+ \epsilon_0(||P_{\rho_0}(x_k)\|_r + ||P_{\rho_0}(x_0)\|_r),$$

and hence

$$\|P_{\rho_0}(a_k)\|_r - \|P_{\rho_0}(a_0)\|_r \leq \|P_{\rho_0}(n_k)\|_r + \|P_{\rho_0}(n_0)\|_r$$

$$+ (2 + 2\epsilon_0)\sigma + \epsilon_0(||P_{\rho_0}(R_k)\|_r + ||P_{\rho_0}(R_0)\|_r)$$

$$+ \epsilon_0(||P_{\rho_0}(x_k)\|_r + ||P_{\rho_0}(x_0)\|_r).$$
Since, $\|P_{\rho(0)}a_0\|_r < 2\sigma$, we have

$$\|P_{\rho(k)}a_k\|_r < 2\sigma + \|P_{\rho(0)}R_k\|_r + \|P_{\rho(0)}a_0\|_r + (2 + 2\varepsilon_0)\sigma + \varepsilon_0(\|P_{\rho(0)}R_k\|_r + \|P_{\rho(0)}a_0\|_r),$$

and hence

$$\|P_{\rho(k)}a_k\|_r < 2\sigma + (1 + \varepsilon_0)(\|P_{\rho(0)}R_k\|_r + \|P_{\rho(0)}a_0\|_r) + 2\varepsilon_0(\|P_{\rho(0)}R_k\|_r + \|P_{\rho(0)}a_0\|_r).$$

After rearranging the terms, we get

$$(1 - \varepsilon_0)\|P_{\rho(0)}a_k\|_r < (4 + 2\varepsilon_0)\sigma + (1 + \varepsilon_0)(\|P_{\rho(0)}R_k\|_r + \|P_{\rho(0)}a_0\|_r) + 2\varepsilon_0(\|P_{\rho(0)}R_k\|_r + \|P_{\rho(0)}a_0\|_r),$$

and hence

$$\|P_{\rho(0)}a_k\|_r < \frac{4 + 2\varepsilon_0}{1 - \varepsilon_0}\sigma + \frac{1 + \varepsilon_0}{1 - \varepsilon_0}(\|P_{\rho(0)}R_k\|_r + \|P_{\rho(0)}a_0\|_r) + \frac{2\varepsilon_0}{1 - \varepsilon_0} \|P_{\rho(0)}R_k\|_r,$$

where we used Lemma 2, and $\|x_k\|_r < (1 + \varepsilon_0)/(1 - \varepsilon_0)\|x_0\|_r$, which is a direct consequence of the environmental assumptions.

If the detector $D_1$ misses the attack as well, then for all $k \in \{0, \ldots, N_r - 1\}$ we will have $\|P_{\rho(0)}R_k\|_r < \sigma$, and by using the triangle inequality

$$\|P_{\rho(0)}a_k\|_r < \|P_{\rho(0)}a_k\|_r + \sigma.$$ 

Now, we have bounds for both $\|P_{\rho(0)}a_k\|_r$ and $\|P_{\rho(0)}a_k\|_r$.

At this point, we will invoke the results proved in the false alarm rate of $D_1$, indeed specifically $d(k, 1)$: With a probability of at least $1 - (2N_r + 1)(2/e)^{N_r/2}$, the following will be true

$$\|P_{\rho(0)}a_0\|_r < \sigma, \quad \|P_{\rho(0)}R_k\|_r < \sigma, \quad \|P_{\rho(0)}R_k\|_r < \sigma,$$

for any $k = 0, \ldots, N_r - 1$. With the help of these inequalities, we get

$$\left(\frac{1}{N_r} \sum_k \|a_k\|_r\right) < \frac{8 + 2\varepsilon_0}{1 - \varepsilon_0}\sigma + \frac{4\varepsilon_0}{1 - \varepsilon_0} \left(\frac{1}{N_r} \sum_k \|x_k\|_r\right).$$

However, our assumption about the attack signal RMS level implies

$$\left(\frac{1}{N_r} \sum_k \|a_k\|_r\right) \geq \frac{8 + 2\varepsilon_0}{1 - \varepsilon_0}\sigma + \frac{4\varepsilon_0}{1 - \varepsilon_0} \left(\frac{1}{N_r} \sum_k \|x_k\|_r\right),$$

contradicts with this. Therefore, with a probability of at least $1 - (2N_r + 1)(2/e)^{N_r/4}$, we get a contradiction. In summary, if we are in Case 3, with a probability of at least $1 - (2N_r + 1)(2/e)^{N_r/4}$, detectors $D_1$ and/or $D_2$ will generate an attack alarm.

Now, we can write the following bound

$$P_M \leq \max\{(1/2)^{N_r} + (2/e)^{N_r/4}, (2N_r + 1)(2/e)^{N_r/4}\},$$

which completes the proof.

VIII. CONCLUSION

In this paper, we have proposed a cyberattack and interference resilient automotive radar system, and derived analytical upper bounds on false attack alarm probability, $P_F$, and the probability of not detecting an alarm, $P_M$. It is proved that, attacks which are above a defined relative threshold are almost always detected, with almost no false alarms, more precisely $P_F$, and $P_M$ converge to zero exponentially as $N_r, N_s \to \infty$ with $N_r/N_s$ kept constant. This relative threshold can be made as small as possible by using lower noise analog hardware, and shorter frame durations which require higher processing power. All of these theoretical results are illustrated on a numerical example using real data from 77 GHz automotive radars. Worst case effects of undetected weak attacks are also analyzed, and it has been shown that the random signature function technique reduces the impact of such weak attacks. The proposed basic design strategy is choosing radar parameters such that $P_F, P_M$, and the relative threshold is small so that undetected weak attacks will have negligible effect on the radar range-velocity heatmap, and attacks above this minimal threshold are almost always detected with almost no false alarms.

REFERENCES

[1] S. M. Putole, M. Torfak, D. Wang, and M. Ali, “Automotive radars: A review of signal processing techniques,” IEEE Signal Process. Mag., vol. 34, no. 2, pp. 22–35, Mar. 2017, doi: 10.1109/MSP.2016.2628914.
[2] C. Bhat, “Cybersecurity challenges and pathways in the context of connected vehicle systems,” Data-Supported Transp. Oper. Planning Center, Austin, TX, USA, Tech. Rep. 134, Feb. 2018. Online. Available: https://ctr.utexas.edu/wp-content/uploads/134.pdf
[3] S. Alland, W. Stark, M. Ali, and M. Hegde, “Interference in automotive radar systems: Characteristics, mitigation techniques, and current and future research,” IEEE Signal Process. Mag., vol. 36, no. 5, pp. 45–59, Sep. 2019, doi: 10.1109/MSP.2019.2908214.
[4] O. Toker, S. Alsweiss, J. Vargas, and R. Razdan, “Design of an automotive radar sensor firmware resilient to cyberattacks,” in Proc. IEEE SoutheastCon, Raleigh, NC, USA, Mar. 2020, pp. 1–7.
[5] O. Toker and S. Alsweiss, “Design of a cyberattack resilient 77 GHz automotive radar sensor,” Electronics, vol. 9, no. 4, p. 573, 2020, doi: 10.3390/electronics9040573.
[6] O. Toker and B. Kuhn, “A Python based testbed for real-time testing and visualization using TI’s 77 GHz automotive radars,” in Proc. IEEE Veh. Netw. Conf., Los Angeles, CA, USA, Dec. 2019, pp. 1–8.
[7] O. Toker, S. Alsweiss, and M. Abid, “A computer vision based testbed for 77 GHz mmWave radar sensors,” in Proc. IEEE SoutheastCon, Raleigh, NC, USA, Mar. 2020, pp. 1–7.
[8] A. Sargolzaei, K. K. Yen, M. N. Abdelghani, S. Sargolzaei, and B. Carbunar, “Resilient design of networked control systems under time delay switch attacks: application in smart grid,” IEEE Access, vol. 5, pp. 15901–15912, 2017.
[9] A. Sargolzaei, K. Yazdani, A. Abbaspour, C. D. Crane, and W. E. Dixon, “Detection and mitigation of false data injection attacks in networked control systems,” IEEE Trans. Ind. Informat., vol. 16, no. 6, pp. 4281–4292, Jun. 2020.
[10] A. Sargolzaei, K. K. Yen, and M. N. Abdelghani, “Preventing time-delay switch attack on low frequency control in distributed power systems,” IEEE Trans. Smart Grid, vol. 7, no. 2, pp. 1176–1185, Mar. 2016.
ONUR TOKER (Member, IEEE) received the B.S. degrees in electrical and electronics engineering, mathematics and physics from Bogazici University, Istanbul, and the M.S. degrees in electrical engineering and mathematics and the Ph.D. degree in electrical engineering from The Ohio State University, Columbus, OH, USA. He worked as a Graduate Researcher with the Technical University of Eindhoven, The Netherlands, and then with the University of California at Riverside, Riverside, CA, USA. He worked for LiDAR company, Quanergy, (Sunnyvale, CA, USA), and then for Augmented Reality Company, DAQRI, (Sunnyvale, CA, USA, and Pasadena, CA, USA). He is currently an Associate Professor of Electrical and Computer Engineering with Florida Polytechnic University. His research interests include FPGA-based hardware acceleration, mmwave and THz radar and communication systems, mmwave software defined radios, cybersecurity, and attack resilience. He has published more than 100 technical articles, and was the PI and co-PI for several funded projects.