[Invited] Breaking (and Fixing) Channel-based Cryptographic Key Generation: A Machine Learning Approach

Ihsen Alouani

Breaking (and Fixing) Channel-based Cryptographic Key Generation: A Machine Learning Approach

Ihsen Alouani
IEMN CNRS 8520, Université Polytechnique Hauts-De-France
CSIT, Queen's University Belfast, UK

***Abstract*—Several systems and application domains are undergoing disruptive transformations due to the recent breakthroughs in computing paradigms such us Machine Learning and communication technologies such as 5G and beyond. Intelligent transportation systems is one of the flagship domains that witnessed drastic transformations through the development of ML-based environment perception along with Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication protocols. Such connected, intelligent and collaborative transportation systems represent a promising trend towards smart roads and cities. However, the safety-critical aspect of these cyber-physical systems requires a systematic study of their security and privacy. In fact, security-sensitive information could be transmitted between vehicles, or between vehicles and the infrastructure such as security alerts, payment, etc. Since asymmetric cryptography is heavy to implement on embedded time-critical devices, in addition to the complexity of PKI-based solutions, symmetric cryptography offers confidentiality along with high performance. However, cryptographic key generation and establishment in symmetric cryptosystems is a great challenge. Recent work proposed a key generation and establishment protocol for vehicular communication that is based on the reciprocity and high spatial and temporal variation properties of the vehicular communication channel.**

This paper investigates the limitations of such channel-based key generation protocols. Based on a channel model with a machine learning approach, we show the possibility for a passive eavesdropper to compromise the secret key in a practical manner, thereby undermining the security of such key establishment technique. Moreover, we propose a defense based on adversarial machine learning to overcome this limit.

***Index Terms*—Security, IoT, Machine Learning, Cyber-Physical Systems, Cryptography**

I. Introduction

Due to the breakthroughs in machine learning (ML) and communication technologies, the technological landscape of modern systems is continuously moving towards more ubiquitous, connected and intelligent devices and systems. This development trend offers considerable opportunities towards a fundamental paradigm shift in several application domains such as transportation, energy, health, etc. However, several security and privacy challenges have to be considered for the quest of trustworthiness given the race between attack and defense mechanisms and approaches in this topic [1]–[4].

Intelligent transportation systems are among the most rapidly evolving Cyber-Physical Systems (CPS). With various technologies that are revolutionizing the sector (electric vehicles, autonomous driving, connected cars, artificial intelligence), the automotive ecosystem is shifting toward the Internet of Vehicles (IoV). With communication technologies such as Vehicle to Vehicle (V2V) or Vehicle to Infrastructure (V2I), vehicles would be able to interact actively and in a collaborative manner for a safer and more secure traffic. In the IoV paradigm, vehicles are no longer considered as isolated systems controlled only by human drivers, but rather nodes within an interconnected complex system. The latter should allow secure and high performance communication for a safer driving, optimized traffic and ergonomic services. While the vast majority of vehicular collisions is mainly caused by drivers distraction or insufficient environment perception, wireless communication can considerably reduce the risk of driver-caused collisions and improve traffic safety [5]. In fact, V2V and V2I communication enables vehicles to have a better perception and a more precise understanding of their environment. For this reason, governmental agencies, as well as original equipment manufacturers and research organizations are working towards the definition of wireless vehicular communication protocols (V2X) [6]. The European Commission proposed a legal framework on Cooperative Intelligent Transport Systems (C-ITS). In the EU, the only commercially available short-range V2X technology is called ITS-G5 [7], which is based on the IEEE 802.11 communication standard and standardized in Europe as ETSI EN 302 663 [8]. In the US, the ITS-G5 technology is also referred to as WAVE (wireless access in vehicular environments) technology or DSRC - Dedicated Short-Range Communication [9].

Security Concerns in vehicular Communication Since the vehicular network is based on wireless communication channels, the information is transmitted using broadcasted signals. This makes the vehicular network vulnerable to eavesdropping, message modification, and impersonation attacks. Since sensitive data such as personal information for entertainment applications, financial transactions, security alerts, etc. could be transmitted using V2X communication, the communication confidentiality represents a serious concern. In a project with support from the Defense Advanced Research Projects Agency

(DARPA), researchers have designed and implemented an attack on vehicular infotainment applications and systems like UConnect that results in recalling several vehicles concerned with this vulnerability such as Chrysler [10]. Moreover, the vehicular network is open by construction to ensure access to all users. This sharpens the concerns about future connected and autonomous vehicles that may represent a serious safety threat if the security is not taken seriously from an early stage of the design process. In fact, malicious users could undertake multiple passive or active eavesdropping, or even total hijacking if commands are sent through the wireless channel.

For this reason, a secure channel establishment needs to be set before communication. A secure channel is established through cryptographic systems that provide confidentiality and integrity into communication. Cryptosystems fall under two main categories: Symmetric and Asymmetric. One of the differences between those two categories is the computation load and by consequence energy consumption. In fact, symmetric algorithms, such as the Advanced Encryption Standard (AES), have very high performance and lower energy overhead [11] compared to asymmetric algorithms. On the other hand, asymmetric algorithms such as RSA and Elliptic Curve Cryptography (ECC) are computation intensive and challenging to implement on resource-limited and time-critical devices. Nevertheless, the advantage of asymmetric cryptography is the unnecessary key agreement since the encryption happens using public keys. In symmetric algorithms, the communicating nodes are supposed to share a secret key, which makes secure secret key generation and distribution a serious challenge. Recent work has proposed channel-based key generation mechanisms that leverage two fundamental properties of communication channels to establish random keys for two users: (i) channel reciprocity, and (ii) channel uniqueness [3], [4].

In this paper, we investigate the security of physical layer based key generation and establishment protocols. Specifically, we propose a ML-based approach to estimate the generated key from an eavesdropper. We propose a convolution neural network (CNN) architecture that is trained on an eavesdropper dataset to estimate the generated key at thee victim side with a sliding window manner. Our results and estimation of the key retrieval complexity show a high capacity of reducing the search space to retrieve the key from a simple passive eavesdropper. We also propose adversarial machine learning approach to enhance the security of channel-based key generation.

Contributions. The contributions of this paper can be summarized as follows:

- We investigate the limits of channel-based key generation and establishment in the case of vehicular communication by comprehensively taking into account different propagation scenarios and channel models.
- We propose a ML-based approach that drastically reduces the search space to practically retrieve the key in a short amount of time by a simple passive eavesdropper.
- We propose a countermeasure to the proposed attack by leveraging adversarial attacks to fool the attacker model and deceive the adversary. We believe this is the first defense that uses adversarial attacks as an active countermeasure against another attack.

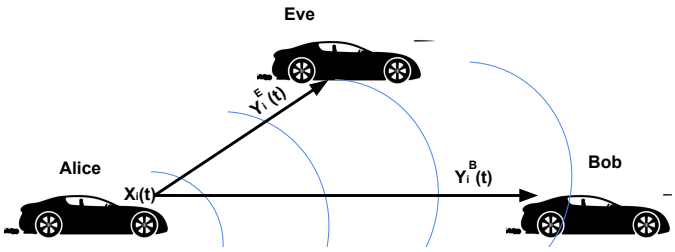

Fig. 1. V2V communication between two legitimate users (Alice and Bob) and a passive eavesdropper (Eve).

II. Background

A. Overview on crypto-systems

A secure channel establishment needs to be set through cryptographic systems before communication. Table I shows a comparative overview on the cryptographic systems and their corresponding characteristics. Cryptosystems fall under two main categories: Symmetric and Asymmetric, and could be hybrid between those two. Symmetric algorithms, such as AES have very high performance and lower energy overhead compared to asymmetric algorithms. However, asymmetric cryptography has an interesting advantage which is the unnecessary key agreement since the encryption and decryption require two different keys, public and private, respectively. On the other hand, symmetric algorithms use the same secret key for encryption and decryption, thereby requiring a secure secret key generation and distribution, which is a serious challenge. Key establishment techniques based on Public Key Infrastructure (PKI) represent the mainstream adopted solution. However, PKI requires a trusted third party which is the certificate authority (CA) that grants digital certificates and revokes compromised ones. In addition to the delay and complexity, especially for the vehicular network application, the centralized aspect of this solution concentrates the risk on the CA. If it is compromised, the whole network becomes vulnerable.

B. Key generation from the communication channel

In a vehicular network, Rayleigh model represents the propagation and the Doppler shift effect in a fast fading channel [12]. In this model, the channel gain H should abide by the following Probability Distribution Function (PDF):

$$PDF_H(H,\sigma) = \frac{H}{\sigma^2} e^{-H^2/(2\sigma^2)} \quad (1)$$

In [3], a channel entropy based key generation technique for V2V communication is proposed. As shown in Figure 1, both legitimate users Alice and Bob are driving, where Alice's

TABLE I
COMPARISON OF EXISTING CRYPTOGRAPHIC ALGORITHMS.

	Symmetric	Asymmetric	Hybrid
Authentication	Message Authenication Code (MAC)	Digital signature	Digital signature on keys, MAC on data
Confidentiality	Data Encryption	Data Encryption	Key encryption with asymmetric, data encryption by symmetric
Key size	32-256 bits	ECC: 256-384 bits RSA: 1024-3072 bits	512-3072 bits for Asym. 32-256 bits for Sym.
Performance	Fast	Slow	Medium

vehicle (A) is communicating with Bob's vehicle (B). Assume the driving velocities for A and B is V_A and V_B , respectively and the velocity difference between these two moving vehicles is $\triangle V$. The coherence time T_c of the communication channel between A and B may be estimated using Equation 2 [12].

$$T_c \approx \frac{0.423}{f_d} \tag{2}$$

If A and B want to generate a key with size of K_{size}, they need to exchange a set of predefined probe signals (can be any kind) to evaluate the randomness of the wireless channel gain H using Equation 1. To have a low mismatch rate, they must exchange each probe signal within the Coherence Time (T_c) interval. Meanwhile, in order to keep bits of the generated key uncorrelated to each other, the time interval defined as τ_{step} between exchanging each probe signal should be no less than T_c. Notice that, as long as the sender A and receiver B share the same τ_{step}, the process of exchanging pre-defined signals is naturally synchronized. In [3], the authors assume there exists a pre-defined τ_{step} for both A and B.

Figure 2 gives a brief overview on the physical layer channel based key generation protocol. After the probe signals are exchanged, a set of measured Received Signal Strength (RSS) values is used to generate secret key bits on each side. The authors then implement a mismatch checking step to remove mismatching bits. During this step, both the sender and receiver will publicly exchange the indexes of the probe signals which are used for generating secret bits, in K_{idx} and remove the mismatched indexes. In this phase,the exchange is publicly broadcasted and a potential attacker might get K_{idx}. Nevertheless, the attacker will not have access to the generated key bits because only the sender and receiver share the bilaterally received RSS values of the probe signals propagated through the channel AB. Based on the reciprocity hypothesis [13] of the wireless channel, if A and B broadcast the probe signals to each other within the coherence time of the wireless channel, the channel gain is symmetric, i.e., $H_{A\rightarrow B}(t) \approx H_{B\rightarrow A}(t)$. However, from an eavesdropper's perspective, the estimated channel gain $H_{A\rightarrow E}(t)$ will be different from $H_{A\rightarrow B}(t)$. Therefore, the key generated between A and B is assumed to be totally independent from the key estimated by the eavesdropper E. Once $K_{AB} = K_{BA}$ is generated securely, the secure channel between A and B is established and the communication can be protected by a symmetric crypto-system using the generated key.

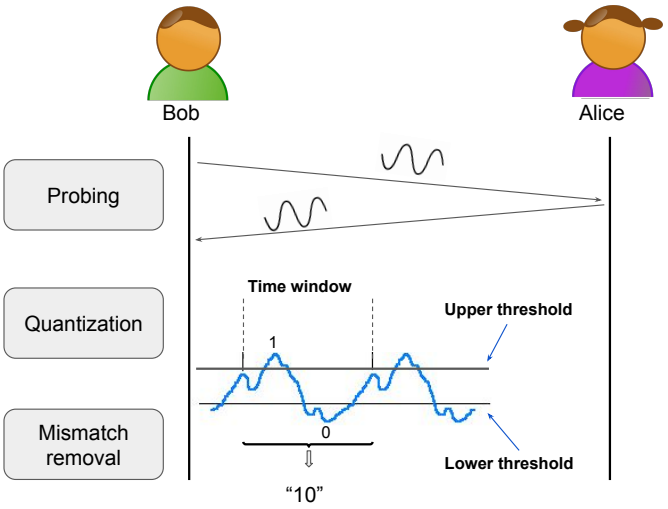

Fig. 2. Channel-based physical layer key generation protocol.

C. Channel models

In this section, we give an overview on the existing propagation models for the received signal strength during a wireless transmission. The wireless communication in different vehicular environments is subject fading and path loss. The path loss is due to the propagation of the electromagnetic wave in the free space. It reflects the attenuation of the signal strength as a function of the distance between the transmitter and the receiver. On the other hand, small-scale power variations are caused by the presence of objects in the environment. In the following, we examine the different fading patterns: path fading, shadowing, multipath fading, and Doppler spreading.

1) Path Loss: Path loss describes the loss of received signal power due to the propagation of the electromagnetic wave. Two path loss models are commonly used in the literature:

- Free-Space Model: it represents the ideal propagation conditions where a single line-of-sight (LOS) path exists between the transmitter and the receiver . The Friis formula [14] describes the power received by the receiver antenna at a distance d from the transmitter antenna:

$$P_r(d) = \frac{P_t \times G_t \times G_r \times \lambda^2}{(4\pi \times d)^2 \times L} \tag{3}$$

where P_t is the power of the transmitted signal, G_t and G_r are, respectively, the gains of the transmitter and receiver antennas, λ is the wavelength and L ($L \geq 1$)

is the loss factor of the system. Free-space path loss is given by:

$$PL_{Free-Space} = 10 \times \log_{10} \frac{P_t}{P_r} \quad (4)$$

- Two-ray-ground model: This model assumes that the received signal is composed of the propagation of the signal emitted in the LOS free space and its reflection on the ground. Unlike the Friis formula, it takes into account the heights of the transmitting antenna h_t and receiver h_r. The received power is predicted by the following equation:

$$P_r(d) = 10 \times \log \frac{P_t \times G_t \times G_r \times h_t^2 \times h_r^2}{d^4 \times L} \quad (5)$$

Two-ray-ground path loss is then expressed in dB by:

$$PL_{2-ray-ground} = -10 \times \log \frac{G_t \times G_r \times h_t^2 \times h_r^2}{d^4 \times L} \quad (6)$$

2) Rician model:

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left(-\frac{x^2+A^2}{2\sigma^2}\right) I_0\left(\frac{Ax}{\sigma^2}\right), & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (7)$$

Where: $I_0(.)$ is defined as the modified zero-order Bessel function of the first kind and A is the dominant signal amplitude.

The Rician distribution is generally described by the factor k which corresponds to the ratio between the power of the LoS path and the average power of multipaths obeying a Rayleigh distribution. k is given by Equation 8.

$$k = \frac{A^2}{2\sigma^2} \quad (8)$$

III. Proposed Attack

In this section, we present our system model, the considered threat model and the proposed attack.

A. System Model

Let Alice and Bob start a key generation protocol from the wireless channel using the Received Signal Strength (RSS) following the methodology explained in Section II-B. As represented in Figure 1, when Alice broadcasts the raw signal $X_i(t)$, Bob receives $Y_i^B(t)$ which is the output of the channel AB.

1) Ideal Environment: Free Space: Bob and Eve receive signals that result respectively from the channels AB and AE. In a free space environment, these signals have one component which ius the Line of Sight (LoS). The LoS component is ruled by the Path Loss model presented in Section II-B. Therefore, Eve can fully estimate the signal strength in Bob (and vice versa in Alice) using Equation 9 below:

$$P_r(B)dBm = 10\log_{10}\left(\frac{P_r(E)}{10^{-3}}\right) + 20\log_{10}\left(\frac{AE}{AB}\right) \quad (9)$$

Where:

- $P_r(B)$ is the RSS in Bob
- $P_r(E)$ is the RSS in Eve
- AE and AB are respectively the distance between Alice and Eve and between Alice and Bob

Notice that there is no unknown element in Equation 9 from Eve's perspective to find the RSS in Bob. This is theoretically a proof that Eve can totally retrieve the generated key between Alice and Bob given that he knows the key generation protocol. However, this model is unrealistic and not useful in vehicular communication environments. In the next subsection, we consider more realistic environments and propose our method to generate the key from a passive eavesdropper.

2) Urban Environment: In urban vehicular communication channels, the transmitter and receiver are moving in a density that can vary from low to high density. This aspect results in different fading statistics depending on the existence of a LOS, shadowing (obstructed LoS) and the proportion of LoS/NoLoS.

An empirical study of the V2V channel in urban, suburban and rural environments [15] where measurements were taken in real driving conditions, shows that the fading envelop is accurately modeled by a Rician distribution [16]. This distribution is expressed in Equation 7 in Section II-B. The K parameter would be changing in inverse proportion with traffic density since the chance of having LOS decreases as traffic density increases on the road. The highest density case tends to result in a Rayleigh fading model. At low traffic conditions, it is very likely to have a LOS since there would be no obstacles in the road between the vehicles.

B. Threat Model

In our scenario two legitimate vehicles, Alice and Bob, are establishing a secure communication channel through a shared secure key agreement based on their wireless fading channel. We consider the presence of another vehicle, Eve, as a *passive* eavesdropper. The eavesdropper is a regular user of the vehicular network that has access to public parameters of the network such as the key generation protocol. We also assume that Eve has the ability to estimate the distance that separates him from Alice, and from Bob. This is possible using narrow-band radars such as Bosch MRR [17] used in Tesla cars. Since the variability of the environment has a direct impact on the entropy of the generated keys, we consider different propagation environments covering several density levels in urban environments as well as rural environment. We also assume that the key generation protocol is not secret.

The assumption made by authors in [3] that there is no correlation between the received signal in Bob and the one received in Eve is questionable. In fact, while the channel Alice-Eve is different from the channel Alice-Bob under realistic distance assumptions, it is not totally de-correlated from one another [18]. Practically, the work in [3] assumes that the eavesdropper has no possible way of modeling the channel between Alice and Bob, based on what he is receiving from the broadcasted $S(t)$ from Alice. We demonstrate in this paper that a careful modeling of the channel can lead a passive eavesdropper to recover a significant part of the

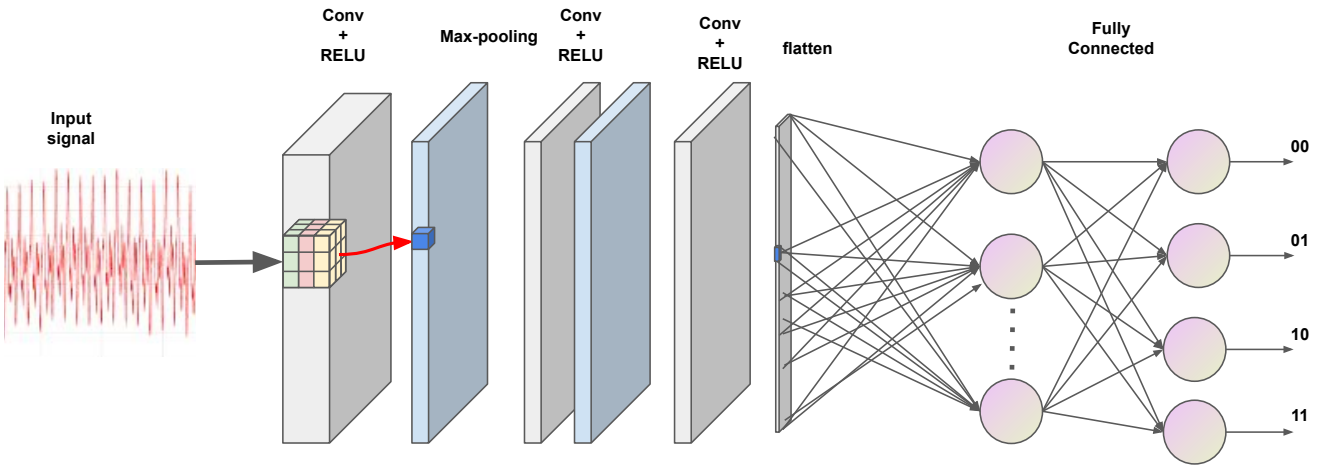

Fig. 3. ML model architecture to estimate the generated key.

signal power variations, and consequently generate a highly correlated key to the one exchanged by Alice and Bob. Eve could then retrieve the same Alice-Bob key through a simple and easy brute forcing.

C. ML-based attack

In this approach, we exploit the correlation between the receiver and the eavesdropper signals to train a ML model to estimate for a given setting the corresponding secret key based on the signal received in the eavesdropper side. We formulate the problem as a classification problem, and we assume the independence between the bits. Therefore, for a given signal, we slice the window of n sub-samples and estimate the class which corresponds to a p-length word. An illustration of the model is show in Figure 3 and the detailed hyperparameters of the architecture are depicted in Table II.

TABLE II
ARCHITECTURE OF THE 1D-CNN.

Layer type	Unit	Output shape	# of Parameters
Conv (ReLU)	(16,3)	(16, 658)	64
Maxpool	(2)	(16, 329)	0
Conv (ReLU)	(32,3)	(32, 327)	1,568
Maxpool	(2)	(32, 163)	0
Conv (ReLU)	(64,3)	(64, 161)	6,208
Maxpool	(2)	(64, 80)	0
Flatten	-	-	5,120
Linear (LReLU)	16	-	81,936
Linear	4	-	84
Total parameters	-	-	89,860

Dataset. We built a dataset of 120000 samples corresponding to the different scenarios as follows:

- Free Space environment: in this case we generate signals from an emitter and collect both signals at the legitimate receiver and the eavesdropper. The propagation channel for this case is a simple path loss for both receivers.
- Urban environment: in this case we model the channel following the Rician model while varying the parameter $k \in [0, 10]$ for different situations.

As for training and validation, the dataset samples are split in a $7:3$ way. We use a learning rate of $1e^{-4}$ and a batch size of 4 for the training. We use Adam optimizer ($\beta 1 = 0.9$ and $\beta 2 = 0.999$) with the negative log likelihood loss function. We train the model for 100 epochs.

The objective of the proposed approach is to generate a bit stream from a received signal by slicing a window that generates 2 bit at a time. While the ideal case is that the generated bit stream is identical to the established key between thee legitimate users, the practical aim is to have as high correlation as possible that would reduce the exploration space in a brute-force follow-up attack.

IV. EXPERIMENTS

In this section, we present the experimental study in different scenarios and propagation environments.

A. Setup

In an urban environment, the propagation is faced by both path loss and multipath fading. The higher traffic density is, the higher the impact of multipath and the lower the impact of line of sight propagation. Hence, to model the wireless channel, we use the Rician model. Varying the K factor in the Rician model translates the variation of the impact of the LoS and by consequence models the traffic density.

B. Results

Figure 4 shows the correlation between the bit stream estimated by the eavesdropper and the actually generated key by the legitimate vehicles. The correlation translates the success probability to unveil the generated key by the eavesdropper. Using our attack model, it is clear from Figure 4 that the lower traffic is around the legitimate vehicles, the higher the probability of key unveiling from a passive eavesdropper.

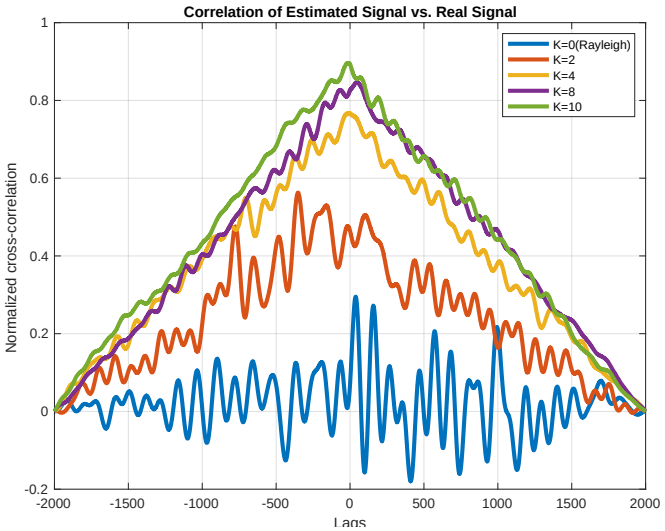

Fig. 4. Correlation between the RSS received by Bob and the RSS estimated at Bob by Eve.

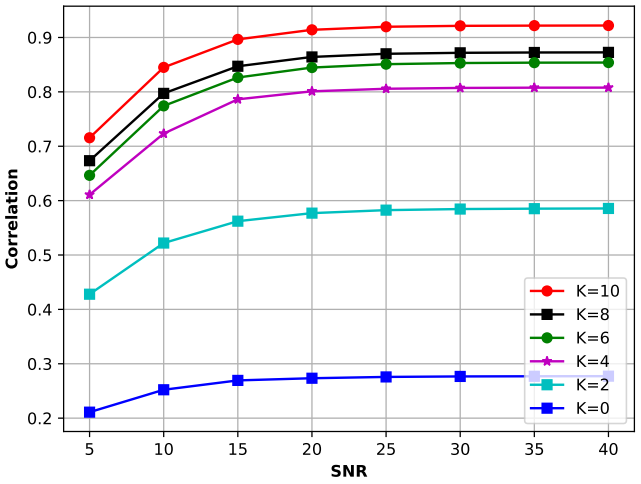

Fig. 5. Correlation between the RSS received by Bob and the RSS estimated at Bob by Eve for different SNR values.

To further evaluate the effectiveness of the attack, we implemented the whole key generation protocol from the eavesdropper perspective.

V. Key Retrieving Difficulty

The environment modeling technique allows the generation of estimated received signal strength whose correlation to the actual received signal strength depends on the environment itself. While this is an interesting finding, a full key generation process is needed as a proof of concept of this attack. In this section, we present an optimized space exploration attack to estimate the difficulty of brute forcing the key and consequently compromising the key agreement. Based on the correlation shown in Figure 4,we build a customized exploration technique that takes into account the specificity of the channel model. In fact, instead of launching a brute force exploration on the whole space, we proceed to a local search. This is explained by the signal correlation we obtain from the estimated RSS. The higher the correlation, the less bits to flip in the exploration. The exploration method is described in Algorithm 1. In this algorithm, we define a local space represented by a window of $N-$bits where N represents the size of the local space to explore. The exploration consists of flipping P bits within the local space, where $P < N$. After scanning a given combination (N, P), we gradually increase the local space as well as the number of bits to flip (P). In the case of low correlation, the size of the local space is high, thereby increasing the possibilities to explore and the exploration time consequently. The initialization is a purely empirical choice and could be adapted to the key size.

Algorithm 1: Difficulty of key brute forcing.

Result: Number of iterations to retrieve the key: $iter$
//Initialize the exploration parameters
if $COR > 0.85$ **then**
    P = 1;
    N = Sizeof(K_e);
else
    if $COR \in [0.75, 0.85]$ **then**
        P = 2;
        N = 4;
    else
        if $COR \in [0.65, 0.75]$ **then**
            P = 4;
            N = 8;
        else
            P = 6;
            N = 12;
        end
    end
end
//Start the exploration
while *True* **do**
    //Explore combinations by flipping P bits out of
    //N-bit windows
    Explore(K_e, N, P);
    if *(Key Retrieved)* **then**
        return($iter$) ;
    else
        if *(N ¡ Sizeof(K_e))* **then**
            N++ ;
        else
            P++ ;
            N = P+1 ;
        end
    end
end

Figure 6 shows the result of key retrieving difficulty in

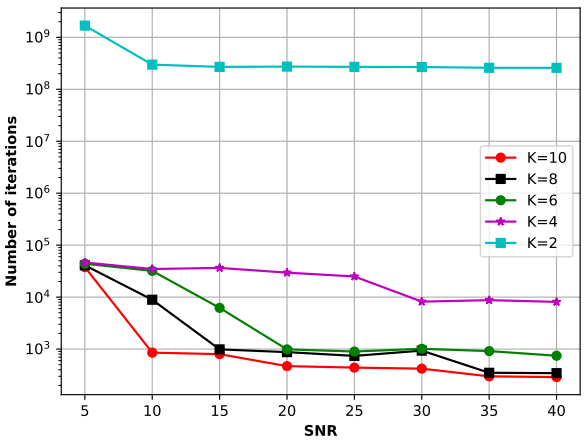

Fig. 6. The number of iterations to retrieve the key vs. Rician K-factor.

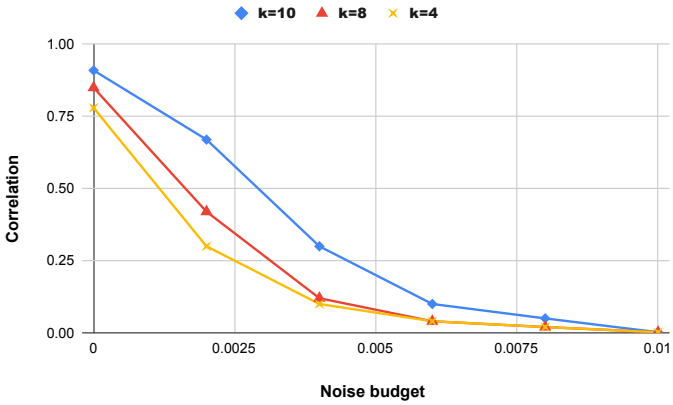

Fig. 7. Attack success vs adversarial noise budget. The attacker success is measured by the correlation between the correct key and the estimated key by the eavesdropper.

terms of number of iterations for a key agreement scenario of 128 bits. In the case of dense traffic, the impact of vehicles shadowing lead to eliminate the LOS component of the received signal. Hence, the only component that shapes the signal power in the receiving part is the multipath NoLOS component. While the key generation process is not compromised in this case, all remaining scenarios are breakable in a very short time.

VI. Defending through Adversarial Attacks

In this section, we provide a defense against our ML-based attack that leverages the inherent vulnerability of ML. More specifically, we exploit adversarial machine learning approaches to generate an adversarial noise that is able to fool the attacker. IN fact, despite their effectiveness and popularity, ML-powered applications suffer from a critical challenge, i.e., adversarial attacks. In fact, by injecting specific perturbation patterns into input data, adversarial attacks can fool the victim model and mislead its cognitive process. With thorough methods proposed, carefully designed adversarial perturbations can be implemented in the real world.

An adversary, using information learned about the structure of the classifier, tries to craft perturbations added to the input to cause incorrect classification. For illustration purposes, consider a CNN used for a classification task. Given an original input x and a target classification model $f()$, the problem of generating an adversarial example $\tilde{x}$ can be formulated as a constrained optimization [19]:

$$\tilde{x} = \arg\min_{\tilde{x}} \mathcal{D}(x, \tilde{x}), s.t.\ f(x) = l,\ f(\tilde{x}) = \tilde{l},\ l \neq \tilde{l} \quad (10)$$

Where $\mathcal{D}$ is a distance metric used to quantify the similarity between two samples and the goal of the optimization is to minimize the added noise, typically to avoid detection of the adversarial perturbations. l and $\tilde{l}$ are the two labels of x and $\tilde{x}$, respectively: $\tilde{x}$ is considered as an adversarial example if and only if the label of the two samples are different ($f(x) \neq f(\tilde{x})$) and the added noise is bounded ($\mathcal{D}(x, \tilde{x}) < \epsilon$ where $\epsilon \geqslant 0$).

Algorithm 2: *attack* function.

Input: a classifier: C with loss J, noise budget: ε, step size: α,input signal: x, label: l, number of iterations: m ;

Output: x^{adv} Initialize $x^{adv} \leftarrow 0$;

for *i= 0 to m-1* **do**

$x_{i+1}{}^{adv} = Clip\{x_i + \alpha sign(\nabla_{x_i}^{adv} J_\theta(C(x_i{}^{adv}), l))\}$;

end

Distance metrics. The adversarial examples and the added perturbations are designed by the attacker to be visually imperceptible by humans. To model this imperceptability, three main metrics to approximate human's perception of visual difference, namely L^0, L^2, and L^∞ [2]. These metrics are special cases of the L^p norm:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \quad (11)$$

These three metrics focus on different aspects of visual significance. L_0 counts the number of pixels with different values at corresponding positions in the two images. L_2 measures the Euclidean distance between the two images x and $\tilde{x}$. L_∞ measures the maximum difference for all pixels at corresponding positions in the two images.

In our case, we try too generate an adversarial patch that is case-agnostic, i.e., it fools the adversary's model regardless of the case or the setting. Practically, the defender will be broadcasting the adversarial patch in parallel with the key generation and establishment protocol. Therefore, we try to generate Input-Agnostic adversarial noise inspired from the universal adversarial perturbations (UAP) [20], which generates an input-agnostic adversarial patch after optimizing over a given dataset. Let $y \in \mathbb{R}^d$ be an input of dimension d that

follows a distribution μ ($y \sim \mu$). The main objective of a UAP is to fool a target model $C(.)$ on almost all inputs sampled from μ. This problem can be formulated as finding a vector δ such that:

$$C(x+\delta) \neq C(x),\ for\ "most"\ x \sim \mu \tag{12}$$

Where δ represents the adversarial patch and must satisfy the following two constrains:

- $\|\delta\|_p \leq \xi$
- $\mathbb{P}_{x\sim\mu}\left(C(x+\delta) \neq C(x)\right) \geq 1-\rho$

The parameter ξ controls the magnitude of the perturbation vector δ, and ρ quantifies the desired fooling rate for all signals sampled from the distribution μ.

The noise generation mechanism is detailed in Algorithm 2, where the noise is updated based on the gradient ascent, i.e., to maximize the error on the ML model (which is in this case the eavesdropper model). The output of this algorithm is an adversarial noise that, when broadcasted reduces the efficiency of the key retrieval attack and hence protects the privacy of the defender. Figure 7 shows the adversarial noise-based defense efficiency in terms of eavesdropper's attack success. Specifically, we explore the correlation between the estimated key by the attacker and the actual key established between the legitimate users vs. the noise magnitude that is broadcasted. The figure depicts that for a noise budget $\varepsilon = 0.004$, the correlation is low enough to make the key retrieval time very costly from the eavesdropper perspective.

VII. Concluding remarks

The interaction with the physical world in applications such as e-Health, intelligent transportation systems, and access control further sharpens the critical aspect of any eventual compromise of such communicating systems. On the other hand, the limited power and resource budget is a constraints that motivates the development of lightweight cryptography for IoT and Edge devices. However, this might come at a high cost in terms of security and privacy.

This paper investigates the limits of channel based key generation in the context of vehicular communication. We show that these systems vulnerable to attacks from different vectors and layers, and can be compromised by. a ML-based approach with a passive eavesdropper threat model. We also exploit the vulnerability of ML models to design a defense based on adversarial noise. The adversarial noise generated by the defender operates as a defensive smart jamming approach that actively deceives that attacker.

References

[1] A. Guesmi, I. Alouani, K. N. Khasawneh, M. Baklouti, T. Frikha, M. Abid, and N. Abu-Ghazaleh, "Defensive approximation: Securing cnns using approximate computing," in *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS 2021. New York, NY, USA: Association for Computing Machinery, 2021, p. 990–1003.

[2] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," 2016.

[3] J. Wan, A. B. Lopez, and M. A. Al Faruque, "Exploiting wireless channel randomness to generate keys for automotive cyber-physical system security," in *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*, April 2016, pp. 1–10.

[4] J. Wan, A. Lopez, and M. A. A. Faruque, "Physical layer key generation: Securing wireless communication in automotive cyber-physical systems," *ACM Trans. Cyber-Phys. Syst.*, vol. 3, no. 2, oct 2018. [Online]. Available: https://doi.org/10.1145/3140257

[5] C. Weiss, "V2x communication in europe from research projects towards standardization and field testing of vehicle communication technology," *Computer Networks*, vol. 55, no. 14, pp. 3103 – 3119, 2011, deploying vehicle-2-x communication. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128611001198

[6] J. H. et al. (2014) Vehicle-to-vehicle communications: Readiness of v2v technology for application. Accessed: 2020-03-5. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/27999

[7] K.-O. Proskawetz. (2016) Deployment of v2x services based on its-g5. Accessed: 2020-03-5. [Online]. Available: https://www.car-2-car.org/press-media/press-releases/press-details/deployment-of-v2x-services-based-on-its-g5-25

[8] ETSI. (2019) Intelligent transport systems (its) its-g5 access layer specification for intelligent transport systems operating in the 5 ghz frequency band. Accessed: 2020-03-5. [Online]. Available: https://www.etsi.org/deliver/etsi_en/302600_302699/302663/01.03.00_20/en_302663v010300a.pdf

[9] H. Rakouth, P. Alexander, A. J. Brown, W. Kosiak, M. Fukushima, L. Ghosh, C. Hedges, H. Kong, S. Kopetzki, R. Siripurapu, and J. Shen, "V2x communication technology: Field experience and comparative analysis," in *Proceedings of the FISITA 2012 World Automotive Congress*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 113–129.

[10] C. Miller and C. Valasek, "Remote exploitation of an unaltered passenger vehicle," *Black Hat USA*, vol. 2015, no. S 91, 2015.

[11] N. Potlapally, S. Ravi, A. Raghunathan, and N. Jha, "A study of the energy consumption characteristics of cryptographic algorithms and security protocols," *IEEE Transactions on Mobile Computing*, vol. 5, no. 2, pp. 128–143, 2006.

[12] *Fading Channel Characterization and Modeling*. John Wiley & Sons, Ltd, 2005, ch. 2, pp. 17–43. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471715220.ch2

[13] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. B. Mandayam, "Information-theoretically secret key generation for fading wireless channels," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 240–254, 2010.

[14] H. T. Friis, "A note on a simple transmission formula," *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, May 1946.

[15] S. Zhu, T. S. Ghazaany, S. M. R. Jones, R. A. Abd-Alhameed, J. M. Noras, T. Van Buren, J. Wilson, T. Suggett, and S. Marker, "Probability distribution of rician k-factor in urban, suburban and rural areas using real-world captured data," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 7, pp. 3835–3839, July 2014.

[16] J. G. Proakis and M. Salehi, *Fundamentals of communication systems*. Pearson Education India, 2007.

[17] Bosch. (2015) Mid-range radar sensor (mrr) for front and rear applications. Accessed: 2019-08-15. [Online]. Available: https://www.bosch-mobility-solutions.com/media/global/products-and-services/passenger-cars-and-light-commercial-vehicles/driver-assistance-systems/predictive-emergency-braking-system/mid-range-radar-sensor-(mrr)/product-data-sheet-mid-range-radar-sensor-(mrr)-2.pdf

[18] M. Edman, A. Kiayias, and B. Yener, "On passive inference attacks against physical-layer key extraction?" in *Proceedings of the Fourth European Workshop on System Security*, ser. EUROSEC '11. New York, NY, USA: Association for Computing Machinery, 2011. [Online]. Available: https://doi.org/10.1145/1972551.1972559

[19] X. Yuan, P. He, Q. Zhu, R. R. Bhat, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *CoRR*, vol. abs/1712.07107, 2017. [Online]. Available: http://arxiv.org/abs/1712.07107

[20] S. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 86–94.