Channel State Information Based Cryptographic Key Generation for Intelligent Transportation Systems

Soheyb Ribouh, Kelvin Phan, Arnav Vaibhav Malawade, Yassin EL Hillali, Atika Rivenq, Mohammad Abdullah Al Faruque

Abstract—Due to the sensitivity of the information exchanged in Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communication, generating secret keys is critical to secure these communications. As nature is open access, distributed keys are more vulnerable to attacks in the vehicular environment. Physical layer key generation methods using wireless channel characteristics show promise in preventing such attacks, generating keys independently, and removing the need for distribution. In this paper, we present a novel key generation approach in a real vehicular environment based on Channel State Information (CSI), including a new algorithm for key bit extraction. We implemented our algorithm using USRP B210 Software-Defined Radios (SDR) and the industry-standard V2X communication protocol: IEEE 802.11p. The proposed key generation protocol uses the CSI values of each sub-carrier as a source of randomness, from which bits are extracted using a new QAM demodulator quantizer (QAM-Dem-Quan). We compared our technique to state-of-the-art Received Signal Strength (RSS)-based approaches, and show that our method achieves better performance. Moreover, we reached a min-entropy of approximately 70% for the generated keys and a key generation rate of less than 150 µs/key for key lengths ranging from 16 to 128 bits.

Index Terms—Channel State Information (CSI), IEEE 802.11p, Physical layer key generation, QAM-Dem-Quan, V2V security, V2X communications.

I. INTRODUCTION AND RELATED WORK

CONNECTED vehicle technology is expected to see prevalent usage as part of intelligent transportation systems (ITS), which themselves are projected to be widely implemented in urban centers. This expectation is supported by industry trends, such as automakers Volkswagen and Toyota declaring their intent to deploy Vehicle-To-Everything (V2X) communication technology in 2021. Moreover, it is further supported by policy trends like the proposed mandate from the National Highway Traffic and Safety Administration (NHTSA) that would have required all vehicles to have V2X capability by 2020 [1]. These trends themselves are likely motivated by the benefits promised by ITS, which would help resolve current urban mobility issues. ITS have the potential to reduce the severity of up to 80 percent of non-impaired crashes according to the NHTSA and the data collected from these systems can inform better traffic flow management for decreased commute times and, as a result, reduced emissions [2],[3]. Autonomous vehicles have also been the focus of recent transportation advancement, slated to make up 40 percent of vehicle traffic by the 2040s with the promise of better safety, mobility, traffic flow, and energy usage [4]. Through Vehicle-To-Vehicle (V2V) and Vehicle-To-Infrastructure (V2I) communications, a fully realized ITS composed of connected autonomous vehicles promises to be the future of transportation [5],[6],[7].

Given the critical role Vehicle-To-Everything (V2X) wireless communication plays in connected vehicle infrastructure, it thus follows that the security of these communications is equally important. ITS vehicles are highly mobile and so, exchange information in an ad hoc fashion forming Vehicle Ad Hoc Networks (VANETs) or Mobile Ad Hoc Networks (MANETs); however, the growing connectivity between vehicles and the external environments in these VANETs means that there is an increase in attack surfaces and vulnerabilities [8]. As a result, various studies have affirmed the typical security concerns for VANET security standards: authentication, ensuring messages are generated by legitimate users; availability, maintaining access to ITS services; integrity, preventing messages from being tampered with; and privacy/confidentiality, restricting message access to relevant parties [9],[10],[11],[12],[13]. In these studies, current VANET security concerns are presented as unsolved and malicious parties are profiled as those intending to disrupt transportation systems since compromising the security of transportation often involves sabotaging the transmission of critical safety data; therefore, it follows that V2X security is an open research challenge and necessary to the advancement of ITS. Of these security challenges, our concern, and that of the related work described below, is the confidentiality of V2X automotive wireless communications in ITS.

Currently, the most common methods for securing V2X communications utilize a public key infrastructure (PKI) [14]. However, encryption methods based on traditional PKI present some risks [15] and incur significant latency while performing the necessary cryptographic operations of the ITS-S [16]. Furthermore, they require a significant amount of computing resources and power [17]. For V2X applications where safety-critical communications must be completed within 200 ms and embedded hardware with limited computing power is employed, such high latency and power requirements are not viable [18]. These restrictions led to the development of a Vehicular Public Key Infrastructure (VPKI) that mitigates these issues. The Security Credential Management System (SCMS) is a leading candidate for standardizing V2X security in the United States because it can provide data authentication to preserve privacy and provides the additional security measure of being able to revoke of permissions for misbehaving vehicles [19],[20]. Despite these security measures, the risk assessment concluded that the SCMS is still vulnerable to some types of attacks [20]. The current state-of-the-art...
solutions in this research area are thus based on quantum cryptography such as the quantum random number generator [21] and quantum-SCMS [22], which uses quantum theory to perform cryptographic tasks, but this is a cost-prohibitive solution [17]. Physical layer-based approaches that leverage shared wireless channel characteristics in a Pre-Shared Key (PSK) infrastructure have thus emerged as a cost-effective, practical solution to the latency and power issues found in traditional approaches.

In this physical layer approach, measurements of the shared wireless channel are used to seed a key generator individual to each communicating party on opposing sides of a two-way communication link. The shared channel would mean that equivalent symmetric keys are generated without requiring a key exchange step. Existing publications have explored different channel attributes as sources of randomness. A majority of these methods are based on either the Received Signal Strength Indicator (RSSI) or the Channel Impulse Response (CIR), which can be characterized by the CSI values estimated. Existing key generation methods based on entropy sources such as RSSI or CIR are generally assessed on the performance of most, if not all, of the following three components: Quantization, Reconciliation, and Privacy amplification. Compared to RSSI-based approaches, CIR-based quantization methods have demonstrated inherent advantages in the level of secrecy and key generation rate. However, the RSSI-based approach is still a viable method because it is reachable from higher network layers and is typically self-implemented in SDR devices. Key generation rates based on RSSI and CSI have been compared in [23]. One example is the RSSI-based physical layer key generation technique for V2V communication introduced in [24]. In this methodology, RSSI values are measured from a probe signal exchanged between two communicating vehicles through Bluetooth. The RSSI values are then quantized using thresholds to get a sequence of binary values for key generation. Additionally, this paper proposes protocol algorithms to support this key generation technique including scenario mapping, optimization, and cryptographic key derivation. A similar approach was proposed in [25] for securing optical light communication between vehicles. However, RSSI values are not the only sources of entropy in physical-layer applications. The technique proposed in [25] was expanded to the application of optical fiber link security in [26] and [27], where Polarization Mode Dispersion (PMD) was used as a source of randomness for generating keys from optical signals. Similar to the RSSI-based approaches, the collected PMD were divided into many group sizes, then quantized using different upper and lower thresholds for each group size to get binary sequences used for key generation.

In [28], a novel method of extracting bits using CSI-based key generation from the OFDM-FDD system of 3G-LTE is presented; it aims to estimate the uplink CSI of the BS (base station) and UE (user equipment) at the same time by exchanging a known, private probe signal. The amplitude of the estimated CSI is then quantized via a lower and upper threshold to bit values ‘0’ and ‘1’ respectively to generate a key. Simulation results show that this approach supports different scenarios with low complexity. CSI-based methods are further explored in [29], where a new approach called SKECE is proposed. The CSI-based key extraction of this method has three primary steps. First, CSI values are quantized by fixing two adaptive thresholds and assigning “1” to values above the upper thresholds and “0” to values below the lower thresholds; CSI values between the thresholds are dropped. Second, leakage-resilient consistency validation is performed by using an SHA-1 to hash the bitstreams. Finally, weighted key recombination is used in case all bit sequences have mismatches. To resolve this, reconciliation is performed to extract a consistent bitstream using a weighted key recombination method. SKECE was evaluated through various experiments using off-the-shelf 802.11 devices in real-world scenarios.

As an alternative to threshold-based quantization, a new key extraction mechanism called alternating channel quantization was established in [30] and [31]. It is based on the use of nonequivalent quantization sectors on the space of observable complex channels, by adapting the quantization map to the channel observation. This approach has been simulated and tested in an indoor environment using a MIMO system and has been shown to achieve better performance than the grand band methods which use equally probable quantization sectors. Similar to alternating channel quantization, an intelligent key generation mechanism called phase shifting is explored in [32]. It involves converting the shifted phases of the CSI values to constellation points, with a direct quantization after. The simulation results of a typical SISO outdoor channel model show that this method can achieve high efficiency. Nonetheless, all the previously explored methods have some limitations: first, they are not suitable for vehicular fading channels; second, results show low entropy with the decrease of the key generation rate (KGR) at high vehicle velocities; third, the security of these mechanisms has yet to be validated in real-world driving scenarios given that most were tested in simulation or indoor environments.

As described in Section III, the key generation mechanism is non-trivial because of the computational limits of vehicular control systems. Additionally, this paper proposes protocol algorithms to support this key generation technique including scenario mapping, optimization, and cryptographic key derivation. A similar approach was proposed in [25] for securing optical light communication between vehicles. However, RSSI values are not the only sources of entropy in physical-layer applications. The technique proposed in [25] was expanded to the application of optical fiber link security in [26] and [27], where Polarization Mode Dispersion (PMD) was used as a source of randomness for generating keys from optical signals. Similar to the RSSI-based approaches, the collected PMD were divided into many group sizes, then quantized using different upper and lower thresholds for each group size to get binary sequences used for key generation.

In [28], a novel method of extracting bits using CSI-based key generation from the OFDM-FDD system of 3G-LTE is presented; it aims to estimate the uplink CSI of the BS (base station) and UE (user equipment) at the same time by exchanging a known, private probe signal. The amplitude of the estimated CSI is then quantized via a lower and upper threshold to bit values ‘0’ and ‘1’ respectively to generate a key. Simulation results show that this approach supports different scenarios with low complexity. CSI-based methods are further explored in [29], where a new approach called SKECE is proposed. The CSI-based key extraction of this method has three primary steps. First, CSI values are quantized by fixing two adaptive thresholds and assigning “1” to values above the upper thresholds and “0” to values below the lower thresholds; CSI values between the thresholds are dropped. Second, leakage-resilient consistency validation is performed by using an SHA-1 to hash the bitstreams. Finally, weighted key recombination is used in case all bit sequences have mismatches. To resolve this, reconciliation is performed to extract a consistent bitstream using a weighted key recombination method. SKECE was evaluated through various experiments using off-the-shelf 802.11 devices in real-world scenarios.

As an alternative to threshold-based quantization, a new key extraction mechanism called alternating channel quantization was established in [30] and [31]. It is based on the use of nonequivalent quantization sectors on the space of observable complex channels, by adapting the quantization map to the channel observation. This approach has been simulated and tested in an indoor environment using a MIMO system and has been shown to achieve better performance than the grand band methods which use equally probable quantization sectors. Similar to alternating channel quantization, an intelligent key generation mechanism called phase shifting is explored in [32]. It involves converting the shifted phases of the CSI values to constellation points, with a direct quantization after. The simulation results of a typical SISO outdoor channel model show that this method can achieve high efficiency. Nonetheless, all the previously explored methods have some limitations: first, they are not suitable for vehicular fading channels; second, results show low entropy with the decrease of the key generation rate (KGR) at high vehicle velocities; third, the security of these mechanisms has yet to be validated in real-world driving scenarios given that most were tested in simulation or indoor environments.

The remainder of this paper is organized as follows: Section II describes the research challenges and our contributions to solving them, Section III describes our wireless communication and security model, Section IV gives an overview of the proposed key generation process, Section V describes our experimental setups and the experimental results, and Section VI evaluates the success of our proposed method.

II. RESEARCH CHALLENGES AND OUR CONTRIBUTIONS

To overcome the limitations of the aforementioned approaches, it is necessary to solve the following key research challenges:

1) Minimizing performance overhead: For automotive wireless communications or V2X communications, key generation must meet the strict timing constraints of safety-critical messages while operating within the computational limits of vehicular control systems.
2) Selecting a reliable entropy source: The source of entropy for the key generator must preserve the shared physical layer characteristics that produce matching symmetric keys while simultaneously providing sufficient randomness to create secure keys.
To address the previously cited challenges, our main contributions in this paper are as follows:

1) We propose a new method of physical layer key generation for 802.11p-based V2X communication. The bit sequences are extracted from the shared wireless channel, characterized by the estimated CSI values. In this context, we come up with a new quantizer based on a QAM demodulator quantizer (QAM-Dem-Quan) with the addition of a hash function in the output to increase the min-entropy of the key.

2) The proposed approach was implemented on industry-standard software-defined radios (SDR). This solution was validated in several real-world, dynamic vehicular environments (urban and highway). Furthermore, the proposed key generation schema was evaluated for performance and security, demonstrating comparable results in both when compared to the state-of-the-art.

### III. System Model

![System Model Diagram](image)

Given that CSI is the most accurate representation of the wireless channel characteristics and the related research had demonstrated its efficiency in generating secured keys, we establish a wireless communication model centered on the CSI values and their associated characteristics. We use a half-duplex transmitter/receiver pair of vehicular wireless communication nodes (Figure 1), where Vehicle 1 (Alice) represents a connected vehicle that exchanges messages with Vehicle 2 (Bob): another connected vehicle or a Road Side Unit (RSU).

These two vehicles generate a secret symmetric key by exploiting the estimated CSI values to encrypt their communications. The same symmetric key is generated in each vehicle separately, without requiring a Pre-Shared Key step.

The proposed V2X network operates on the IEEE 802.11p standard, known by its industry label of WAVE in the United States and ITS-G5 in Europe. The physical layer of this standard is based on the Orthogonal Frequency Division Multiplexing (OFDM) waveform, which aims to divide the transmitted signal over a large number of sub-carriers. 802.11p uses 64 sub-carriers, where four of them are pilot symbols used for channel estimation.

The transmitted signal based on the OFDM waveform is represented in the time domain as follows:

$$X(t) = \sum_{k=0}^{M-1} X(k)e^{2\pi kt/T}, \ 0 \leq t < T \quad (1)$$

Where $X(k)$ are the transmitted data symbols, $M$ is the number of sub-carriers and $T$ is the OFDM symbol time. However, the received signal in frequency domain over the wireless channel can be written as:

$$Y(k) = X(k)H + W(k) \quad (2)$$

Where $H$ is the wireless channel response and $W$ is the noise in the receiver. After filtering the noise, the received data symbols in each vehicle can be written in the following matrix form:

$$Y_1 = X_2H \quad (3)$$

$$Y_2 = X_1H \quad (4)$$

$Y_1$, $Y_2$ and $X_1$, $X_2$ represent the received and transmitted data symbols at Vehicle 1 and Vehicle 2 respectively and $H$ is the channel matrix, where its coefficients are the CSI values. These CSI values are obtained in the equalization part of the receiver, where they are estimated based on the four transmitted pilot’s sub-carriers. The information transmitted in the pilot’s sub-carriers is known by both the transmitter and the receiver.

In existing research, there are various algorithms for calculating the $H$ matrix’s coefficients. For our model, we have used the least square estimator (LS).(see Equation 5)

$$H_{LS} = Y_{pilot}.X_{pilot}^{-1} \quad (5)$$

By using an interpolation function between the resulting values and the pilot symbols, we retrieve the remaining values of the $H$ matrix which compose the CSI values. The vehicular wireless channel between the transmitter and the receiver is given as a double selective fading propagation channel, which is characterized by the delay spread and the Doppler spread [33]. The base-band time-varying response of the multi-path channel is given by:

$$h(t, \tau) = \sum_{l=0}^{L-1} A_l(t)\delta(\tau - \tau_l(t)) \quad (6)$$

Where $L$, $A_l(t)$, and $\tau_l(t)$ represent the number of non-zero paths, the time-varying complex amplitudes for each path $l$, and the time-varying path delays, respectively. On the other hand, note that in this case the phase of the complex amplitude $A_l(t)$ depends on the variation of the Doppler shift. The propagation of the signal over this channel can induce a Doppler shift to each path in addition to the time delay. As a result, at the receiver side we observe the superposition
of multiple different delayed and frequency shifted versions of the transmitted signal. Due to the reciprocity \cite{34} of the shared wireless channel at the physical layer, we assume that two communicating vehicles estimate the same channel characteristics (CSI) if they are sending messages to each other within the channel’s coherence time \( T_c \) (\( T_c \) is the time interval over which the channel response is considered not varying).

\[
T_c \approx \frac{0.423}{f_d}
\]  

(7)

Where \( f_d \) is the maximum Doppler frequency. In vehicular communication, \( f_d \) can be expressed by the speed difference between the two communicating vehicles \( \Delta V \) as shown below:

\[
f_d = \frac{\Delta V}{c} f_0
\]  

\[
\Delta V = |V_1 - V_2|
\]  

(8)

where \( c \) is the celerity (speed of light) and \( f_0 \) is the communication frequency. As such, the channel characteristics change every coherence time window: \( T_c \). Therefore, the higher the \( \Delta V \) is, the more frequently the channel is changing. However, to extract matching bits to generate the key, CSI values must be extracted within a given coherence time \( T_c \). Otherwise, the channel characteristics will change and it will result in non-matching keys.

A. Security Strength Model

For cryptographic keys, security strength is a measure of the work required to break the encryption through brute force. The formal definition of security strength then, as given by the National Institute of Standards and Technology (NIST), is that an algorithm has “\( X \)-bits security strength” if it takes “\( X \)” number of attempted symmetric keys to guess the correct key \cite{35}. To quantify the randomness and security of a key, we use the concept of min-entropy which is a worst-case entropy estimation that provides a lower bound on the key’s randomness \cite{36}. If \( K \) is the set of all possible randomly generated keys, the equation for min-entropy is as follows:

\[
H_{\infty} = H_{\text{min}} = -\log(\max_{k \in K} \Pr[k = K])
\]  

(9)

\( \Pr[k = K] \) refers to the probability of generating key \( k \in K \). As an extension of this, we consider security strength to be the following:

\[
\text{Security}_{\text{str}} = H_{\text{min}} / \text{Key}_{\text{size}}
\]  

(10)

where \( \text{Key}_{\text{size}} \) is the key length and \( \text{Security}_{\text{str}} \) is a value between 0 and 1. The higher the value the higher the security strength as it provides more bits of entropy.

B. Attack Model

For our methodology, we consider a non-intrusive wireless attack model in which the attacker (Eve) attempts to eavesdrop on the communication between two legitimate parties through a third channel \cite{36}. It is assumed that Eve can capture all the wireless packets and is aware of the characteristics of the V2X network. In this case, if Eve can obtain the same symmetric key, then the system is considered broken.
Algorithm 1: Algorithm for RSSI-Based Physical Layer Key Generation for a Wireless Automotive CPS

**Input:** Measured Signal Strength $RSS$
**Input:** Sample Time Step: $\tau_{step}$
**Input:** Group Size: $G_{size}$
**Input:** Threshold parameter: $\alpha$
**Input:** Required Key Length: $L_{key}$

**Output:** Generated Key: $Key$

1. $L = 0$; $Key = 0$; $RSS_{set} = \emptyset$; $RSS_{filtered} = \emptyset$; $Ke_{yidx} = \emptyset$
2. while $L < L_{key}$ do
   3. for $i = 1$ to $G_{size}$ do
      4. $RSS_{set} = RSS_{set} \cup RSS$
      5. $Wait(\tau_{step})$
      6. $RSS_{filtered} = RSS_{set} * H_{highpass}(t)$
      7. $MeanValue = \text{AverageValue of } RSS_{filtered}$
      8. $Var = \text{VariationValue of } RSS_{filtered}$
      9. $Th_{up} = MeanValue + \alpha * Var$
     10. $Th_{lo} = MeanValue - \alpha * Var$
     11. foreach $RSS_{i} \in RSS_{filtered}$ do
         12. if $RSS_{i} > Th_{up}$ then
             13. $Key = (Key << 1) + 1$
             14. $L = L + 1$
             15. Record $j$ in $Ke_{yidx}$
         else if $RSS_{i} < Th_{lo}$ then
             16. $Key = (Key << 1) + 0$
             17. $L = L + 1$
             18. Record $j$ in $Ke_{yidx}$
     19. Exchange $Ke_{yidx}$; Remove mismatch bits from $Key$
20. Return $Key$

Algorithm 2: Algorithm for CSI-Based Physical Layer Key Generation for a Wireless Automotive CPS

**Input:** Channel State Information $CSI$
**Input:** Required Key Length: $L_{key}$

**Output:** Generated Key: $Key$

1. $L = 0$; $Key = 0$; $CSI_{set} = \emptyset$; $Verified = False$
2. while not $Verified$ do
   3. $GeneratedKeys_{set} = \emptyset$
   4. $CSI_{set} = QAMDemodQuantizer(CSI)$
   5. for $i = 1$ to $Length(CSI_{set})$ in $StepSize = L_{key}$ do
      6. $GeneratedKeys_{set} = GeneratedKeys_{set} \cup CSI_{set}[i : i + L_{key}]$
   7. foreach $TestKey \in GeneratedKeys_{set}$ do
      8. Exchange verification packets;
      9. if Verification Exchange Returns True then
         10. $Verified = True$
         11. $Key = SecureHash(TestKey)$
   12. EndLoop
13. Return $Key$

V. EXPERIMENTAL SETUP

To validate the real-world practicality of our key generation technique and ensure that its performance meets the required metrics, we performed experiments in actual driving scenarios. For these experiments, we used Ettus USRP B210 software-defined radio boards as the communicating vehicles in our wireless network. Two boards represent Alice and Bob, the intended recipients in the message exchange, while the remaining board represents Eve, the malicious eavesdropper. Using GNURadio, the industry-standard IEEE 802.11p wireless protocol is implemented in software and is then run on the Ettus B210 boards to produce IEEE 802.11p transceivers[37]. The transceivers representing Alice and Bob are placed in separate vehicles while the transceiver representing Eve is placed at varying distances relative to Alice. This setup is visualized in Fig. 4. In each driving scenario, RSSI and CSI are collected from each board wherein keys of the following key lengths are generated in post-processing: 16, 32, 64, 128.

Given that real driving scenarios involve factors that vary constantly and cannot be reasonably controlled, we perform tests with regards to velocity, which include an urban/residential scenario(15 mph - 20 mph ) and a highway scenario(40 mph - 50 mph). These experiment parameters common to both RSSI and CSI testing are listed in Table I.

<table>
<thead>
<tr>
<th>Velocity (MPH)</th>
<th>15-20, 40-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Length (Number of Bits)</td>
<td>16, 32, 64, 128</td>
</tr>
</tbody>
</table>

TABLE I: Experimental Setup Parameters common to all scenarios.
A. Real Vehicle Tests for RSSI-Based Key Generation

1) RSSI Data Collection: As presented in Figure 5, RSSI measurements are calculated from raw input data that is collected from a data recorder on the IEEE 802.11p transceiver. Frame decoding refers to the process by which wireless packets are received, synchronized, and decoded. The raw input data is in the form of complex power values that are processed by the following functional blocks: Sync Short, Sync Long, and Frame Equalizer. Sync Short and Sync Long, as noted in the figure, are responsible for detecting potential packet frames and aligning frames respectively. These functional blocks are synchronous GNU Radio blocks, meaning that data is always being periodically received by the transceiver, and so, the data recorder is programmed to only sample complex power values when Sync Short indicates it has detected a frame. Using a custom RSSI measurement function, these complex power values are converted to RSSI values measured in decibel-milliwatts (dBm), producing approximately 14 RSSI values per packet received. Unlike the raw power values, these RSSI values are scalar and can now be used as input data for Key Generation in Algorithm 1. This RSSI calculation, however, introduces latency as it requires costly filters and transformations. Keys are generated using every possible combination of threshold parameter $\alpha$ and Group Size $G_{size}$ listed in Table II.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Size</td>
<td>20, 40, 60, 80, 100</td>
</tr>
</tbody>
</table>

TABLE II: Key Generation parameters for RSSI-based algorithm

2) Correlation of RSSI Values: A subset of 5000 samples from 902,870 RSSI values collected in a real driving scenario test is displayed in Figure 6. Based on the similar shape and amplitude of displayed results, it is apparent that the RSSI values collected at Vehicle A and Vehicle B for the wireless channel communications from Vehicle A to Vehicle B would be highly correlated if they were aligned. This potential high correlation is a result of the shared physical layer or the shared wireless channel between Vehicle A and Vehicle B, which means the channel response should be similar, but the delays caused by the multi-path on the received signal can negatively affect the correlation between the RSSI values collected at Vehicle A and Vehicle B. From [36], it is known that this
correlation allows for the generation of matching symmetric keys as seen when using the threshold-based quantization technique in Algorithm 1. RSSI measurements that are furthest from the mean dBm value (i.e. large spikes in signal strength) are used to generate a key. Cross-correlation techniques were used to determine the delay caused by packet reception timing differences and shift the delayed set of RSSI values accordingly. The result of this is displayed, but despite shifting the correlated values of the RSSI data closer, the values are still clearly not aligned. Potential reasons for this misalignment are elaborated in Section VI.

3) Percentage of Matching Keys and Relevance of Signal Alignment: After collecting RSSI values from tests conducted according to the real driving scenarios described in Section V, the RSSI measurements are used in Algorithm 1 to generate keys in post-processing. Keys are generated separately using Vehicle A values and Vehicle B values independently before performing a reconciliation step, at which point matching keys are those without mismatched bits. In Figure 7, the percentage of matching keys for each key length after generating keys with all possible parameter combinations is listed. The proportion of matching keys for all key lengths are found to be under 10% and demonstrates an inverse relationship with key length, meaning the matching rate decreases as key length increases. However, this relatively low matching rate is only possible in the case that the correlated RSSI measurements are aligned. If these measurements are out of alignment, as seen in the matching rate for unsuccessful alignment, there are 0% matching keys. For these tests, successful alignment was only achieved in the more static urban/parking scenario at low speeds of 15 - 20 mph and only after manually aligning the values in post-processing. Examples of matching keys with varying length achieved in this case can be found in Table III, showing that wireless channel characteristics for V2X communications can be a reliable entropy source for sufficiently strong keys should the keys match. Nonetheless, these matching keys were only produced under more static conditions, which cannot always be expected in normal V2X usage scenarios. At speeds of 40-50 mph in the dynamic highway scenario, signal alignment could not be achieved even in post-processing, and accordingly, the matching rate of 0% follows the previous trend of low matching for unsuccessful alignment. Given the apparent correlation between matching rate and alignment, it thus proves necessary that measures of channel characteristics during packet activity must be aligned.

### TABLE III: Keys Generated from RSSI Data in Urban Scenario Tests

<table>
<thead>
<tr>
<th>Key Length</th>
<th>Generated Keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0010000000000111</td>
</tr>
<tr>
<td>32</td>
<td>00000000001111111111111110100000</td>
</tr>
<tr>
<td>64</td>
<td>000000000011111111111111101000000000001111111111111111</td>
</tr>
<tr>
<td>128</td>
<td>000000000011111111111111101000000000001111111111111111</td>
</tr>
<tr>
<td>256</td>
<td>000000000011111111111111101000000000001111111111111111</td>
</tr>
</tbody>
</table>

B. Real Vehicle Tests for CSI-Based Key Generation

1) CSI Data Collection: As shown in Figure 8, CSI values are collected as output data from the Frame Equalizer functional block, which performs the transceiver’s equalization step by estimating CSI values as described in Section III. Frame decoding is described earlier in the RSSI Data Collection part of Subsection V-A but is essentially the synchronization and decoding of a frame’s complex power values to bits. It is important to note that CSI values are only generated when a packet has been fully received unlike data from Sync Short or Sync Long which sometimes produces false positives for packet reception. Given this dependence on full packet reception, CSI values are not collected from dead air like RSSI values, and, as stated in Section III, a more accurate estimate of channel characteristics while also providing 64 CSI Values per packet received. Using the QAM demodulation-based quantization described in Section IV, the complex CSI values can be converted to bits for generating keys as shown in Algorithm 2.

2) Correlation of CSI Values: A subset of 256 samples from 3264 CSI values collected in a real-world driving scenario test is displayed in Figure 9. From the similar shape and amplitude of the displayed results, it is apparent that the CSI values collected at Vehicle A and Vehicle B for the wireless channel communications between Vehicle A and Vehicle B are highly correlated. This high correlation...
of CSI values is likely a result of the shared wireless channel between Vehicle A and Vehicle B, indicating that the correlated characteristics of shared channels are represented in CSI estimations; therefore, given that CSI values are a measure of channel characteristics like RSSI and RSSI has demonstrated potential as an entropy source [36], we can reasonably conclude that CSI values are a reliable entropy source for generating symmetric keys. Alignment, in this case, is achieved without the use of cross-correlation techniques or the like. Given this successful alignment and the lack of accidental measurements from dead air, all CSI values are produced by real packets and can act as input data for Algorithm 2. At 64 CSI values per packet, CSI estimations can be considered a high yield entropy source.

3) Percentage of Matched Keys in Urban and Highway Scenarios: Once the CSI values have been collected from tests conducted according to the real driving scenarios described in Section V, the CSI values are used in Algorithm 2 to generate keys in post-processing. Keys are generated separately using Vehicle A values and Vehicle B values independently before checking for matching keys or keys that do not have mismatched bits. In Figure 10, the percentage of matching CSI-based keys generated for each key length is listed for both a low-speed Urban Scenario and a high-speed Highway Scenario. For the more static Urban Scenario with low speeds of 15-20 mph, key length and matching rate have a positive correlation wherein the matching rate increases with key length. The reverse is true for the more dynamic Highway Scenario with higher speeds of 40-50 mph, where the matching rate decreases as key length increases. As described in Section IV, our proposed key generation methodology quantizes the received CSI values into a stream of bits that is used to produce the requested key. Each set of keys in both real driving scenarios are generated from a single data set, meaning...
that the keys for each scenario are generated from the finite data gathered in that test. Therefore, the positive trend in the proportion of matched keys for the Urban Scenario is likely not a result of an increase in matching bits, but having the same matching bits for the fewer keys generated at longer key lengths. However, in the Highway Scenario, the more dynamic environment can result in dropped packets or similar interruptions. Given that CSI estimations require fully received packets, the instability of the channel prevents the generation of longer keys and results in a lower key-match rate for greater key lengths as shown in the figure. Examples of the keys generated in the dynamic Highway Scenario are displayed in Table V proving that CSI values capture the unique physical characteristics of shared channels and can thus, be used as a reliable entropy source.

4) Performance Overhead and Security Strength of CSI-Based Key Generation: Regarding performance, the major sources of overhead are data sampling and quantization. Given that there are 64 CSI values collected per packet, which produces a greater number of usable bits after quantization, we ignore data sampling latency since only a few packets are needed to generate each key and the transmission time for few packets is negligible. Thus, we consider quantization to be the main source of performance overhead. As seen in Table V our proposed use of QAM Demodulation for quantization results in execution times on a microsecond scale. Though this latency increases with key length, all of the listed quantization times meet the 200 ms latency requirement for V2X communication set in [36].

Regarding security strength requirements, based on testing results, our proposed CSI-Based Physical Layer Key Generation Methodology meets the uniqueness and min-entropy requirements outlined in Section III-A. The basic premise for our proposed methodology is that separate channels with several wavelengths distance between them are independent of each other and thus, produce different keys even in the same coherence time. In [36], this premise is confirmed when results demonstrated that an attacker positioned several wavelengths distance away experiences different wireless channel characteristics and so, cannot generate the same secret key as parties in the legitimate channel. According to the results outlined in Table V it is clear that keys generated from the channel between Vehicle A and Vehicle C have a mismatch rate greater than 50% with the keys generated from the channel between Vehicle A and Vehicle B. This meets the required uniqueness standard and supports the basic premise for our methodology wherein an attacker that is at least several wavelengths away will not be able to predict/generate the secret key due to experiencing different channel characteristics.

To estimate the min-entropy of our proposed key generation algorithm, we use Equation 9 in Section III-A. We first generate 46,352 8-bit keys from the CSI values collected in the Highway Scenario test to form a set of keys \( K \) that has greater than or equivalent to \( 100 \times 2^8 = 25,600 \) keys, which is the number of keys necessary to reasonably conclude that the calculation produces an average min-entropy. From this set of keys, the key with the most frequent appearance is used to generate \( \Pr_{\text{max}} \) or \( \Pr[k = K] \).

<table>
<thead>
<tr>
<th>Key Length</th>
<th>Average Execution Time (µs)</th>
<th>Average Bit Mismatch Rate for Attacker C (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>18.65</td>
<td>52.38</td>
</tr>
<tr>
<td>32</td>
<td>34.75</td>
<td>52.60</td>
</tr>
<tr>
<td>64</td>
<td>78.45</td>
<td>50.97</td>
</tr>
<tr>
<td>128</td>
<td>149.58</td>
<td>54.29</td>
</tr>
</tbody>
</table>

TABLE V: Average Execution Time for Key Generation and Average Bit Mismatch Rate for Keys Generated by Attacker Vehicle C

As seen in Figure 11 our Key Generation Methodology has an average min-entropy of 70.52% which exceeds that of the RSSI-based method in [36] and industry-standard methods in [38].

![Fig. 11: Comparison of Our Average Min-Entropy to State-of-the-Art](image_url)

Fig. 11: Comparison of Our Average Min-Entropy to State-of-the-Art

### VI. DISCUSSION

A. Minimum Performance Overhead

Given that our proposed key generation algorithm is intended for use in a time-critical setting, it is necessary for the key generation time to consistently meet latency requirements in every scenario. According to [32] qtd. in [36], this hard time limit typically falls within 50 to 200 milliseconds. Assuming no alignment issues, the original RSSI-based Key Generation technique incurs latency from its bit extraction rate metric, which means that it can require more packets depending on the environment and key length. This bottleneck is due in part to variations in the number of usable RSSI values, which result from there being arbitrary RSSI formulas and polling rates across different RSSI hardware sensors. Threshold-based quantization exacerbates this issue because even once all usable RSSI values are collected, not all will be used to generate a key. Multiple packets are commonly required for keys of...
greater length. Data sampling thus becomes a major source of latency alongside the costly RSSI calculation that takes valuable milliseconds to complete. In contrast, our proposed CSI-based Key Generation methodology has demonstrated performance measured consistently in microseconds, well under the 50-200 ms operational range as seen in the experimental results. The source of this successful performance is likely the use of CSI estimations, which provides 64 usable values per packet and the use of the low-latency QAM Demodulator that quantizes these values into 64 or more key bits in microseconds. Due to these mechanisms, our proposed methodology has low overhead in all scenarios as it requires singular packets for most key lengths and efficiently uses all data collected.

B. Reliable Symmetric Keys and Security Strength

As described under research challenges in Section II, matching keys must be reliably generated in various driving scenarios while also having industry-standard security strength. Though it has demonstrated comparable min-entropy to state-of-the-art methods, RSSI-Based Key Generation has displayed success in limited static scenarios. This limited success is further exacerbated by the need for signal alignment, which has been completed in post-processing for this experiment but would require real-time mechanisms in real driving scenarios. Our CSI-Based Key Generation methodology does not suffer from this need for signal alignment because it does not measure dead air or dropped packets, resulting in a significantly higher matching rate and success in both scenarios. Additionally, we have demonstrated that our methodology is safe from eavesdropping, as shown by the high mismatch rate between the eavesdropper, Vehicle C, and Vehicle A. Moreover, given this method’s average-min entropy of 70.52%, it has security strength comparable with state-of-the-art methods and thus, meets the minimum security requirements outlined in [35].

C. Algorithm Complexity

We evaluate the Big-O complexity of our proposed algorithm in terms of the input size (CSIset) and find that our algorithm has a worst-case execution time on the order of $O(n \log n)$. Thus, we can say that our algorithm is a superlinear algorithm where the running time grows approximately in proportion to the CSIset size. Our experimental results shown in Table IV match this assertion.

D. Brute Force Time

The brute force time is the worst-case time needed for an attacker to break a secret key by trying every possible permutation. A key of length $k$ has $2^k$ possible values. The greater the key length, the greater the brute force time and the stronger the key. The brute force time depends on the key length and the attacker’s capabilities (floating-point operations per second). As a result, longer key lengths are generally more preferable as they are more difficult to brute force. As shown in Table IV, the brute force time has been evaluated for various key sizes and levels of attacker capabilities including those of supercomputers (TaihuLight) and quantum computers. Since our work focuses on practical, real-world attacks on individual vehicles, a brute force attack using a TaihuLight supercomputer or a quantum computer are extremely unlikely; however, we present them here to demonstrate that our key lengths and rotation times ensure security with conventional and future computing hardware. From the table, it is clear that keys less than 64 bit are not very secure against brute force attacks; however, we propose rotating 16 and 32-bit keys for every single message to prevent the possibility of an attacker using the broken key to impersonate Alice or Bob. Additionally, confidential information could be reserved for use with 128-bit keys while smaller key sizes are used for immediate safety-critical messages not containing sensitive data.

E. NIST Random Bit Generator Classification

According to the National Institute of Standards and Technology (NIST) recommendations [40], there are two classes of random bit generators (RBGs). The first class uses dedicated hardware or physical experiments to generate random bits, where every bit of output is based on a physical process that is unpredictable; methods in this class are known as Non-Deterministic Random Bit Generators (NRBGs). The second class consists of methods that compute bit sequences deterministically based on pseudo-random number generation methods using specific algorithms; methods in this class are known as Deterministic Random Bit Generators (DRBGs). Thus, our proposed approach can be classified as an NRBG. From NIST recommendations [41], the entropy source model for NRBGs is shown in Figure 12. The entropy source block includes the following components:

1) Noise Source: It is the root of security for the entropy source and the RBG as a whole. It provides the non-deterministic sequences (CSI values in our approach) from the physical process which is the vehicular wireless channel in the setup of our experiments. As shown in the figure, the sampling process includes digitization to convert analog noise to binary data. In our proposed method, since the CSI values are complex numbers, the digitization is performed by the QAM demodulator quantizer (QAM-Dem-Quan).

2) Optional Conditioning: This component aims to increase the entropy of the resulting output bits. In our approach, we use a hash function as our conditioning component as shown in Algorithm 2.
Table VI: Brute Force Time vs Key Size.

<table>
<thead>
<tr>
<th>Key Size (bits)</th>
<th>Time Required at 10^6 FLOPs</th>
<th>Time Required at 10^8 FLOPs</th>
<th>Time Required at 10^17 FLOPs (TaihuLight)</th>
<th>Time Required at 10^18 FLOPs (quantum computer)</th>
<th>Keys Rotation Time (Validity Period)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>65 ms</td>
<td>65 µs</td>
<td>6.5 X 10^-4 ns</td>
<td>6.5 X 10^-5 ns</td>
<td>1 message</td>
</tr>
<tr>
<td>32</td>
<td>1.17 hours</td>
<td>4.2 s</td>
<td>0.42 ns</td>
<td>0.42 X 10^-1 ns</td>
<td>1 message</td>
</tr>
<tr>
<td>64</td>
<td>75.03 X 10^13 years</td>
<td>570.39 s</td>
<td>1.8 X 10^2 s</td>
<td>18 s</td>
<td>1 minute</td>
</tr>
<tr>
<td>128</td>
<td>0.08 X 10^25 years</td>
<td>1.08 X 10^22 years</td>
<td>1.08 X 10^14 years</td>
<td>1.08 X 10^13 years</td>
<td>15 minutes</td>
</tr>
</tbody>
</table>

F. Future Work

As discussed in Section [X], we validated our key generation methodology in multiple scenarios on a real-world testbed of three vehicles equipped with industry-standard V2X hardware. Although we demonstrated good results with our real-world testbed, the practicality and feasibility of CSI-based key generation on larger-scale testbeds with more vehicles remains an open research problem. We leave this for future work.

VII. CONCLUSION

Exploiting the randomness of the wireless channel in the form of estimated CSI, we have presented a physical layer key generation technique that can generate secret symmetric keys to secure automotive wireless communications. Our methodology solves the security challenge of preserving the confidentiality of V2X communications and solves the research challenge of selecting and utilizing a reliable entropy source for generating keys from V2X wireless channels. The results of our real-world tests have demonstrated that our methodology has minimal performance overhead measured in microseconds, well within the expected operational range across various scenarios. These results also showed that the keys generated have an average min-entropy of 70.52% and thus, have comparable security strength to current state-of-the-art methods. In summary, we have validated our CSI-based key generation technique as a practical solution to securing automotive wireless communications.

REFERENCES

[1] S. Abuelsamid, “Toyota Has Big Plans To Get Cars Talking To Each Other And Infrastructure In The U.S.,”

J. W. Wallace, C. Chen, and M. A. Jensen, “Key generation exploiting reciprocal wireless channel modeling, channel estimation, and resource allocation schemes of OFDM.


Soheyb Ribouh received a B.S. degree in Electrical Engineering and Computer Science, from the Engineering School of Technology Algeria (ENT), in 2015 and an M.S. degree in Embedded Systems and Mobile Communications Systems, from the Polytechnic University of the Hauts-de-France (UPHF) in 2017. Currently, he is a Ph.D. Student on Telecommunications at the Polytechnic University of the Hauts-de-France-UPHF. His research interests include Hybrid V2X communications using IEEE 802.11p and cellular technology (4G and 5G), vehicular wireless channel modeling, channel estimation, and resource allocation schemes of OFDM.

Kelvin Phan received a B.S. in Computer Engineering from the University of California, Irvine in 2018 and is currently an M.S. Student at the University of California, Irvine under the supervision of Professor Mohammad Al Faruque. His research interests include the security of wireless V2X communications and the design of intelligent transportation systems for connected/autonomous vehicles.

Arnava Vailbav Malawade received a B.S. in Computer Science and Engineering from the University of California, Irvine in 2018. He is currently an M.S./Ph.D. Student studying Computer Engineering at the University of California, Irvine under the supervision of Professor Mohammad Al Faruque. His research interests include the design and security of cyber-physical systems in connected/autonomous vehicles, manufacturing, IoT, and healthcare.

Yassin El Hillali was born in Chemaia, Morocco, in 1979. He received an M.S degree in 2002, and then a Ph.D. degree in 2005, from the University of Valenciennes, France. Currently, he is an Assistant Professor at the electronics department of the Polytechnic University of the Hauts-de-France (UPHF). He is responsible for ITSCOM platforms (ITS Communications) in the IEMN-DOAE Lab. His research interest includes: signal processing and wireless communication systems (WAVE, ITS5G) applied to intelligent transportation, machine learning dedicated to target detection and recognition using heterogeneous sensors (UWB RADAR, cameras and Lidar) and currently his working on cyber-security for ITS and industrial systems.
Atika Rivenq was born in Marrakech (Morroco) in 1970. She was graduated Engineer from the ENSIMEV engineering school in 1993, received an M.S. degree in electronic engineering in 1993, and then a Ph.D. degree in 1996 from the University of Valenciennes (France). She is a Full Professor at the Department of Electronic Engineering, IEMN Lab, Polytechnic University of the Hauts-de-France (UPHF), France. She is the head of ComNum Group and she is responsible for the SYFRA platform (Systems for smart road applications) with the IEMN-DOAE Lab. Her main activities are in digital communications applied to intelligent transports systems: V2X communications (4G/5G, UWB, ITS-G5, Full Duplex), Cybersecurity and Advanced Perception (Radar, UWB, Detection of vulnerable persons, Deep learning). She Participates in many national and European projects dedicated to C-ITS and inter-vehicle communications especially using ITS-G5 and Cellular systems. Pr. Rivenq has more than 100 publications in international journals, conferences, and workshops in the area of digital communications. She has participated as a General Chair, Member of the Technical Committee, Session Chair, or Program Committee Member of numerous conferences.

Mohammad Abdullah Al Faruque is currently with the University of California Irvine (UCI), where he is an associate professor (with tenure) and directing the Cyber-Physical Systems Lab. Prof. Al Faruque is the recipient of the School of Engineering Mid-Career Faculty Award for Research 2019, the IEEE Technical Committee on Cyber-Physical Systems Early-Career Award 2018, and the IEEE CEDA Ernest S. Kuh Early Career Award 2016. He is also the recipient of the UCI Academic Senate Distinguished Early-Career Faculty Award for Research 2017 and the School of Engineering Early-Career Faculty Award for Research 2017. He served as an Emulex Career Development Chair from October 2012 till July 2015. Before, he was with Siemens Corporate Research and Technology in Princeton, NJ. His current research is focused on the system-level design of embedded systems and Cyber-Physical-Systems (CPS) with special interest in model-based design, multi-core systems, CPS security, etc.

Prof. Al Faruque received his B.Sc. degree in Computer Science and Engineering (CSE) from Bangladesh University of Engineering and Technology (BUET) in 2002, and M.Sc. and Ph.D. degrees in Computer Science from Aachen Technical University and Karlsruhe Institute of Technology, Germany in 2004 and 2009, respectively.

Prof. Al Faruque received the Thomas Alva Edison Patent Award 2016 from the Edison foundation, the 2016 DATE Best Paper Award, the 2015 DAC Best Paper Award, the 2009 IEEE/ACM William J. McCalla ICCAD Best Paper Award, the 2016 NDSS Distinguished Poster Award, the 2008 HiPEAC Paper Award, the 2015 Hellman Fellow Award, the 2015 Kane Kim Fellowship Award, the 2017 DAC Best Paper Award Nomination, the 2012 DATE Best IP Award Nomination, the 2005 DAC Best Paper Award Nomination, the EECS Professor of the year 2015-16 Award, and the 2015 UCI Chancellors Award for Excellence in Fostering Undergraduate Research. Besides 80+ IEEE/ACM publications in the premier journals and conferences, Prof. Al Faruque holds 8 US patents.