Man-in-the-Middle Attack Resistant Secret Key Generation via Channel Randomization

Yanjun Pan
University of Arizona
Tucson, AZ, USA
yanjunpan@email.arizona.edu

Ziqi Xu
University of Arizona
Tucson, AZ, USA
zxu1969@email.arizona.edu

Ming Li
University of Arizona
Tucson, AZ, USA
lim@arizona.edu

Loukas Lazos
University of Arizona
Tucson, AZ, USA
llazos@arizona.edu

ABSTRACT

Physical-layer based key generation schemes exploit the channel reciprocity for secret key extraction, which can achieve information-theoretic secrecy against eavesdroppers. Such methods, although practical, have been shown to be vulnerable against man-in-the-middle (MitM) attacks, where an active adversary, Mallory, can influence and infer part of the secret key generated between Alice and Bob by injecting her own packet upon observing highly correlated channel/RSS measurements from Alice and Bob. As all the channels remain stable within the channel coherence time, Mallory’s injected packets cause Alice and Bob to measure similar RSS, which allows Mallory to successfully predict the derived key bits. To defend against such a MitM attack, we propose to utilize a reconfigurable antenna at one of the legitimate transceivers to proactively randomize the channel state across different channel probing rounds. The randomization of the antenna mode at every probing round breaks the temporal correlation of the channels from the adversary to the legitimate devices, while preserving the reciprocity of the channel between the latter. This prevents key injection from the adversary without affecting Alice and Bob’s ability to measure common randomness. We theoretically analyze the security of the protocol and conduct extensive simulations and real-world experiments to evaluate its performance. Our results show that our approach eliminates the advantage of an active MitM attack by driving down the probability of successfully guessing bits of the secret key to a random guess.

KEYWORDS

Security, physical-layer key generation, channel randomization, reconfigurable antenna

1 INTRODUCTION

In the past two decades, the idea of physical-layer key generation has drawn significant attention from the security community [12, 16, 26, 30]. To generate a pairwise key, Alice and Bob measure the inherent common randomness in the physical wireless channel. This randomness is unique to the location and time that measurements take place. Physical-layer key generation has been shown to be information-theoretically secure against passive eavesdroppers and does not assume any prior shared secrets between Alice and Bob, nor does it require a public key infrastructure. It complements upper-layer security primitives and has been touted as a less expensive and more flexible solution to replace public-key cryptography than competing alternatives (e.g., quantum cryptography [16]).

The key assumptions exploited by physical-layer key generation schemes are channel reciprocity and spatial signal decorrelation. Specifically, the channel reciprocity property implies that the signal distortion (attenuation, delay, phase shift, and fading) is identical in both directions of a link, which enables the measurement of common randomness by legitimate devices. The spatial decorrelation property indicates that in rich scattering environments, two receivers located a few wavelengths away will experience uncorrelated channels. These properties have been demonstrated to hold in practice and the latter is essential for securing the physical-layer key generation process against eavesdroppers [15, 31]. However, an active attacker can bypass this assumption by injecting its own signals to influence and infer the derived key bits between Alice and Bob. For example, Eberz et al. [9] proposed a practical man-in-the-middle (MitM) attack against RSS-based key generation protocols [12, 16], where Mallory exploits the same channel characteristics as Alice and Bob do. Mallory awaits for attack opportunities when the channel from Mallory-to-Alice and Mallory-to-Bob are similar, to inject packets that cause a similar channel measurement at both Alice and Bob. This simple and practical attack strategy enables the adversary to recover up to 47% of the secret key bits generated.

The root cause of this vulnerability is that even in relatively rich scattering environments, the coherence time of the channel is likely to span multiple channel probing rounds. This allows the adversary to measure for opportunities over one probing round before attacking the subsequent probing round with high probability of predicting the extracted bits due to the channel similarity. In this paper, we aim at defending against such MitM attacks, by proposing a channel randomization-based key generation protocol. With the key observation that the MitM attack has to be launched in a wait-then-attack manner, our idea is to break the channel coherence and reciprocity across different probing rounds to prevent successful bit inference/injection from a MitM attacker, while preserving these properties within each probing round for key generation. To realize this goal, we utilize a reconfigurable antenna at one of the legitimate transceivers, where in each probing round the antenna mode is randomly chosen from a large set of diverse antenna modes, which effectively randomizes the channel to/from the adversary.

The main contributions of this work are three-fold:

- We are the first to propose a channel randomization-based approach to prevent active MitM attacks against physical layer key generation protocols. The main innovation is in the quantization phase of the protocol, where one of the legitimate devices is equipped with a pattern reconfigurable antenna. Randomly switching the antenna mode randomizes the channel state in each probing round, thus preventing the adversary from predicting the outcome of packet injection attacks and consequently preventing the adversary from predicting bits of the secret key.
• We theoretically analyze the security of our scheme against the MitM attack by deriving key security metrics such as the key efficiency rate, the key recovery rate, and the probability of correctly guessing the entire key generated by Alice and Bob, under realistic multipath channel models.

• We validate our theoretical analysis using extensive simulation and real-world experiments with commodity devices. Empirical results show that our channel randomization-based key generation approach greatly reduces the successful key bit guessing probability of the MitM attack, making it less effective than random guess.

This paper is organized as follows: In Sec. 2, we give an overview of existing physical-layer key generation protocols and the practical MitM attack proposed in [9]. We present our channel randomization-based physical-layer key generation protocol in Sec. 3. We theoretically analyze its security in Sec. 4. In Sec. 5, we show evaluate the protocol security experimentally. In Sec. 6, we summarize the related works. We present a discussion on a more advanced adversary model in Sec. 6 and conclude in Sec. 7.

2 PRELIMINARIES

Signals traveling over a wireless channel are modified by the channel in a way that is unique to the transmit-receive pair. Physical layer-based key generation protocols exploit this radio propagation characteristic to privately extract secret bits between two parties, Alice and Bob, who do not share any prior secrets. The majority of physical-layer key generation protocols consist of the three basic phases shown in Fig. 1, namely the quantization phase, the information reconciliation phase and the key verification phase. We describe each phase in detail.

2.1 Physical-layer Key Generation

2.1.1 Quantization Phase. The quantization phase aims at converting channel measurements between Alice and Bob to an initial bit stream that will be used for key generation. This is a two-step process involving probing and quantization. During probing, Alice and Bob exchange probing signals to measure channel properties such as RSS or CSI.

In practice, Alice transmits a probe first, allowing Bob to obtain a channel measurement. The probing signal contains a preamble that enables synchronization, channel estimation, and frequency offset correction at Bob, as well as the ID of Alice for sender identification (although no sender authentication is provided). Other payload fields may be included. Bob samples the received signal, which can be represented as

\[ r_b^n(t_i) = h_b(x_a(t_i))s(t_i) + n_b(t_i) \]  \hspace{1cm} (1)

where \( r_b^n(t_i) \) denotes the signal received at Bob during time \( t_i \) when Alice transmits, \( h_b(t_i) \) is the impulse response of the A – B channel at time \( t_i \), \( s(t_i) \) is the known probe signal transmitted by Alice, and \( n_b(t_i) \) is the noise at Bob. From the received signal samples, Bob can further extract the relevant channel measurements that will be used for quantization such as the average RSS or the CSI amplitude. We denote a channel measurement obtained by Bob over a probe packet during the \( i^{th} \) round as \( x_b(i) \). Similarly, Bob transmits its one probe to Alice, allowing her to obtain channel measurement \( x_a(i) \).

With the completion of the \( N \)th probing round, Alice and Bob proceed to quantization by independently selecting quantization thresholds. A commonly-used method is the multi-threshold approach [9, 16], where a device \( u \) independently selects two thresholds \( q_u^L \) and \( q_u^U \) as:

\[ q_u^L = \mu(x_u) + \beta \sigma(x_u) \quad q_u^U = \mu(x_u) - \beta \sigma(x_u). \]  \hspace{1cm} (2)

Here, \( \mu(x_u) \) and \( \sigma(x_u) \) are the mean and standard deviation of \( x_u \), respectively and \( \beta \) is a weight parameter with \( 0 < \beta < 1 \). The quantization function \( Q(x) \) for a measurement \( x \) is defined as

\[ Q(x) = \begin{cases} 0, & x < q_- \\ 1, & x > q_+ \end{cases} \]  \hspace{1cm} (3)

To improve robustness, the quantization can be implemented with excursions [12, 16], which are defined as \( e \) consecutive channel measurements in \( x_u \) (or \( x_b \)) that exceed \( q_+ \) or are below \( q_- \). To identify excursions, Alice first goes through \( x_u \), determines the locations of excursions, and publicly sends the index list of excursion locations \( L_u \) to Bob. Bob checks his own measurements \( x_b \) at the indices specified in \( L_u \) to determine whether an excursion also occurs. Bob then identifies all indexes in \( L_b \) that also produce an excursion in \( x_b \), and sends these indexes to Alice in list \( L_b \). Finally, Alice and Bob quantize each excursion in \( L_b \) and construct the initial bitstreams \( S_u \) and \( S_b \), respectively.

2.1.2 Information Reconciliation Phase. Ideally, if the channel between Alice and Bob is perfectly reciprocal, the quantized sequences \( S_u \) and \( S_b \) at Alice and Bob would match perfectly. In practice, there is a mismatch between \( S_u \) and \( S_b \) because the channel is measured
In this paper, we focus on an adversary whose goal is to infer the bits in the pairwise key agreed between Alice and Bob. By learning the symmetric key, the adversary can later compromise the secrecy of the communications between the two parties. The bit inference can be achieved either through passive eavesdropping [10, 24] or active attacks [9, 32]. Note, that we do not address other type of adversary goals such as authentication. This security property can be achieved in conjunction with key generation using a number of additional out-of-band or in-band methods [6, 18, 21].

Passive attacks: In a passive attack, one or more eavesdroppers overhear the probes of Alice and Bob to infer the generated secret key, without actively interrupting/modifying their communication. This attack model has been widely studied in past work (e.g., [10, 24]). The main defense against eavesdropping exploits the fast channel decorrelation with distance in rich scattering environments. When Mallory is situated several wavelengths away from Alice and Bob, the channel becomes independent from that between the legitimate parties, thus preventing the correct channel measurement and quantization.

Active attacks: In this paper, we focus on active MitM attacks that have been shown to be successful even if the adversary is several wavelengths away from Alice and Bob. Unlike traditional MitM attacks, the attacker does not aim at establishing a separate key with Alice and Bob, since that would require blocking any direct communication between Alice and Bob. Moreover, such a MitM attack are easily detected when authentication is incorporated intokey generation. For instance, if authentication is performed by a visual out-of-band channel, the establishment of different key between Alice and Bob is detected.

A practical MitM attack for key inference is proposed by Eberz et al. [9]. In this attack, Mallory first identifies an attack opportunity on probing round $i$. She then reactive jams the transmitted probes on round $i + 1$ and injects her own probes. This is possible because probes cannot be authenticated in the absence of a common secret. Specifically, let $RSS^a_y$ denote the RSS measured at $x$ when $y$ transmits. Mallory identifies an attack opportunity during the $i^{th}$ probing round when the following condition is met [9]:

$$\big|RSS^a_y - RSS^b_y\big| < d \ \text{AND} \ (RSS^a_y > q_+ \ \text{OR} \ RSS^b_y < q_-) \quad (4)$$

where $d$ is a threshold that defines the maximum RSS difference between probes. The opportunity incorporates two criteria. The first criterion requires the difference between RSS values received from Alice and Bob to be smaller than $d$, which indicates a similar channel from Mallory to both Alice and Bob. The second criterion detects if the RSS measurements exceed any of the quantization thresholds, so that Alice and Bob are likely to use that round to extract a bit during quantization. This criterion also decides the bit that is guessed by Mallory. If the injection attack is implemented after observing $RSS^a_y > q_+$, then Mallory guesses the injected bit to be 0. The prerequisite for successfully guessing the bit extracted by Alice and Bob during round $i + 1$ is that the channel remains the same as in round $i$. If both rounds are within the channel coherence time, Mallory can predict the RSS at both Alice and Bob and infer the extracted bit with high probability.

Fig. 2 shows an example of this attack strategy, where an attack opportunity is found at round 2, with $RSS^a_2 \approx RSS^b_2$ and $RSS^a_2 > q_+$. Mallory follows-up with a jamming and injection attack in the next round. Due to the channel coherence, it follows that $RSS^a_3 \approx RSS^b_3$ and $RSS^a_3 \approx RSS^b_3$. Besides, $RSS^a_3 \approx RSS^b_3$ and $RSS^a_3 \approx RSS^b_3$ due to channel reciprocity. Hence, at round 3, the RSS values at Alice and Bob are $RSS^a_3 \approx RSS^b_3 > q_+$ and $RSS^a_3 \approx RSS^b_3 > q_+$ since $RSS^a_3 \approx RSS^b_3$. Correspondingly, the measurement at round 3 will be converted into bit 1, and Mallory can correctly guess it. Note that although the MitM proposed in [9] is implemented based on RSS, it can be easily extend to CSI by letting Mallory find opportunities upon the amplitude of CSI.
3 RAKG: RA-BASED KEY GENERATION

The main vulnerability exploited by the MitM attack described in the previous section stems from the similarity of the M-A and M-B channels over several probing rounds, once an opportunity is identified. This allows Mallory to predict the impact of the injected probes with high probability. To defend against this attack, our basic idea is to reduce the channel coherence time to the duration of a single probing round by using a reconfigurable antenna (RA) at either Alice or Bob. Without loss of generality, let Alice be equipped with an RA with a total of $U$ antenna modes. By randomly selecting the antenna mode $u$ on every probing round, Alice can randomize the M-A (and A-B) channel on every probing round. Therefore, even if Mallory identifies an opportunity on the $i^{th}$ round, the channel has changed when she performs the injection attack on round $i + 1$. On the other hand, the A-B channel remains reciprocal within the same probing round, so that Alice and Bob can still extract a common bit from the randomized channel.

3.1 Channel Randomization with an RA

An RA is one antenna that can swiftly reconfigure its radiation pattern, polarization, and frequency, or combinations of them by rearranging its antenna currents [4]. We choose an RA for channel randomization as it provides a diversity of antenna states [8]. For example, a type of parasitic-layer based pattern-reconfigurable antenna uses p-i-n diodes as switches, where 12 switches can give up to $2^{12} = 4096$ different configurations [8].

Let Alice be equipped with an RA, whereas Bob is equipped with a conventional omnidirectional antenna (OA). Mallory can have any type of antenna, including OA, directional antenna, or an RA. The time-varying multipath channel that incorporates the impact of an RA is described as [14]:

$$h(u_j, t_i) = \sum_{l=0}^{P} g(u_j, \theta_l) a_l(t_i),$$

where $u_j$ is the antenna mode selected, $P$ is the number of signal paths, $g(u_j, \theta_l)$ is the antenna gain under mode $u_j$ at direction $\theta_l$, and $a_l(t_i)$ is the fading parameter of the $l$-th path at time $t_i$. More specifically, $a_l(t_i) = a_l(t_i)e^{-j\phi_l(t_i)}$, where $a_l(t_i)$ and $\phi_l(t_i)$ are the amplitude and phase of the fading, respectively. From Eq. 5, we can see that although Mallory can know $g(u_j, \theta_l)$ for any given antenna mode, she does not know which antenna mode is selected by Alice. Therefore, the A-B and M-A channels will be randomized between probing rounds. However, the M-B channel will remain the same between several probing rounds because Bob is equipped with an OA.

3.2 Protocol Details

The key focus of the RAKG protocol is to ensure that the initial bit-streams independently extracted by Alice and Bob during the quantization phase are not predictable by Mallory. Once this security property is established, the security of the entire protocol follows as the information reconciliation and key verification phases are not influenced by the MitM attack. Hence, we focus our attention on securing the quantization phase. The novel quantization phase of RAKG is described in the following steps, which are also shown in Fig. 3. The information reconciliation and key verification phases are the same as the ones shown in Fig. 1.

3.2.1 The Quantization Phase in RAKG. There are two sub-phases:

Probing with channel randomization:

1. At the beginning of each probing round, Alice randomly selects an antenna mode and sends a probe to Bob.
2. Bob measures the channel and responds to Alice with his own probe.
3. Alice measures the channel.
4. Alice and Bob repeat steps 1-3 for $N$ probing rounds. At the end of the $N$ rounds, each has a vector of channel measurements $x_a = \{x_a(1), x_a(2), \ldots, x_a(N)\}$ and $x_b = \{x_b(1), x_b(2), \ldots, x_b(N)\}$, respectively.

Initial bitrate construction:

5. Alice and Bob independently select thresholds $q^a_+$, $q^a_-$, $q^b_+$, and $q^b_-$ according to Eq. (2).
6. Alice compares her channel quantizes her channel measurements $x_a$ and sends Bob the list of indexes $L_a$ for which the channel measurements are above $q^a_+$ or below $q^a_-$. Bob checks his corresponding measurements in $x_b$, and makes a list $L_b$ of all indexes with measurements that agree with those of Alice. Bob sends $L_b$ to Alice.
7. Alice and Bob quantize $x_a$ and $x_b$ at each index in $L_a$ and $L_b$, respectively using the quantization function in Eq. (3) and generate initial bitstreams $S_a$ and $S_b$, respectively.

Note that in RAKG, we set the excursion length $e$ equal to one. This is because channel randomization on a per round basis drastically reduces the excursion length. A short excursion length has been shown to increase the number of extracted bits, but also increasing the probability of bit mismatches, thus establishing two competing factors in the rate of secret key extraction [16]. However, it is beneficial to our protocol in terms of security.

3.2.2 An Illustrative Example. Fig. 4 shows an example timeline of the channel randomization in RAKG and how it can defend against a MitM attack. For clarity, we denote the time that Bob and Alice receive a probe during the $i^{th}$ probing round as $t_i$ and $t_i'$, respectively. In the first probing round, Alice selects antenna mode $u_1$, and no attack opportunity is found by Mallory. In round two, Alice randomly and uniformly selects another antenna mode $u_2$. Mallory identifies an attack opportunity because $RSS_m^a(t_2) \approx RSS_m^b(t_2)$ and $RSS_m^a(t_2) > q_+$. In the third probing round, Alice...
We use the following two metrics proposed in [9] to quantify the channel measured by Bob at \( t \) if the channels stayed coherent across rounds two and three:

\begin{align*}
\text{RSS}_a(t) - \text{RSS}_b(t) &< d \\
\text{RSS}_a(t) > q_+ &\quad \text{AND} \quad \text{RSS}_b(t) < q_-.
\end{align*}

4.2 Theoretical Analysis

To analyze the three metrics defined in Sec. 4.1, we first state Mallory’s bit guessing behavior.

**Bit Guessing Behavior:** As mentioned in Sec. 2.2, Mallory’s bit guessing behavior is decided by \( \text{RSS}_m^a \) in the observed opportunity. We define opportunity 0 (\( O_0 \)) and opportunity 1 (\( O_1 \)) as:

- **opportunity 0 (\( O_0 \)):**
  \( |\text{RSS}_m^a - \text{RSS}_m^b| < d \quad \text{AND} \quad \text{RSS}_m^a < q_- \quad \text{RSS}_m^b < q_- \)

- **opportunity 1 (\( O_1 \)):**
  \( |\text{RSS}_m^a - \text{RSS}_m^b| < d \quad \text{AND} \quad \text{RSS}_m^a > q_+ \quad \text{RSS}_m^b > q_+ \).

If Mallory finds an \( O_0 \), she guesses the bit injected into the next probing round as 0. Similarly, Mallory guesses the injected bit as 1 when \( O_1 \) is observed. Here we define the opportunity with both \( \text{RSS}_m^a \) and \( \text{RSS}_m^b \) because they must have the same relationship with \( q_+ \) or \( q_- \) to generate a bit. Hence, we evaluate the attack of Mallory more precisely by considering them jointly.

Based on \( O_0 \) and \( O_1 \), here we differentiate the key recovery efficiency for bit 0 and bit 1 at a round i as \( p_0 \) and \( p_1 \), which is captured by the following conditional probabilities, respectively:

\begin{align*}
p_0 &= P \left( \text{RSS}_m^a(i) < q_- \quad | \quad O_0 \right. \left. \quad \text{at} \quad i = 1 \right), \quad (6) \\
p_1 &= P \left( \text{RSS}_m^a(i) > q_+ \quad | \quad O_1 \right. \left. \quad \text{at} \quad i = 1 \right). \quad (7)
\end{align*}

**Proposition 4.1.** When antenna modes are independently and uniformly selected for each probing round, \( p_0 \) and \( p_1 \) only depend on the state of the M-A channel during round i. That is, \( p_0 \) and \( p_1 \) can be simplified as:

\begin{align*}
p_0 &= P \left( \text{RSS}_m^a(i) < q_- \right), \quad (8) \\
p_1 &= P \left( \text{RSS}_m^a(i) > q_+ \right).
\end{align*}

The proof can be found in Appendix, Sec. 8.2.

Intuitively, Proposition 4.1 states that when the channel is randomized with independently and uniformly selected antenna modes, the channel states are independent from round to round. Using the RA multipath channel model in Eq. (5), we derive the closed-form for \( p_0 \) and \( p_1 \) as follows:

**Proposition 4.2.** Let the M-A channel follow the Rician model with fading parameters \( \gamma_{ma}(u) \) and \( \gamma_{ma}(u) \) under antenna mode. If Alice selects antenna modes independently and uniformly at each probing round, and opportunities are found upon CSI amplitude, \( p_0 \) and \( p_1 \) can be simplified to:

\begin{align*}
p_0 &= \frac{1}{|U|} \sum_{u \in U} \left( 1 - Q_1 \left( \frac{\gamma_{ma}(u)}{\gamma_{ma}(u)} \cdot \frac{q_-}{\gamma_{ma}(u)} \right) \right), \quad (9) \\
p_1 &= \frac{1}{|U|} \sum_{u \in U} Q_1 \left( \frac{\gamma_{ma}(u)}{\gamma_{ma}(u)} \cdot \frac{q_+}{\gamma_{ma}(u)} \right), \quad (10)
\end{align*}

where \( Q_1(\cdot) \) is the Marcum Q-function, and \( |U| \) is the cardinality of the set of available antenna modes.
The proof can be found in Appendix, Sec. 8.3.

Here, we consider Rician fading since we focus on indoor environments, but the results can be extended to other channel models such as Rayleigh fading.

Using the individual probabilities $p_0$ and $p_1$, we can now evaluate the PMF of the number of bits guessed by Mallory using combinatorial arguments. This is derived in the following proposition.

**Proposition 4.3.** For a key of length $t$ extracted by Alice and Bob with RAKG, let $n$ be the number of opportunities identified by Mallory, and $n_0$ be the number of opportunities for injecting bit 0. The PMF of correctly guessing $m$ injected bits and the average key recovery efficiency and key recovery rates are given by:

$$
\mathbb{P}(M = m) = \sum_{i=0}^{m} \binom{n}{i} p_0^i (1 - p_0)^{n_i} p_1^{m-i} (1 - p_1)^{n-n_0-(m-i)}
$$

$$\mathbb{E}[\text{KRE}] = \frac{n_0 p_0 + (n - n_0) p_1}{n}, \quad \mathbb{E}[\text{KRR}] = \frac{n_0 p_0 + (n - n_0) p_1}{t}.
$$

The proof is a direct application of the binomial distribution assuming independence for each opportunity due to channel randomization. For recovering the entire key, Mallory guesses the bits being attacked with the bit guessing strategy having a success probability of $p_0$ and $p_1$, respectively. For the remaining bits that do not present an attack opportunity Mallory can take a random guess. Then the key guessing probability is

$$p_{\text{key}} = 0.5^{t-n} p_0^{n_0} p_1^{n-n_0}.
$$

In practice, Mallory may not need to correctly guess the entire key. By exploiting the error correction capacity of the information reconciliation phase, Mallory can afford up to $w$ errors, where $w$ is the error correcting capability of the code used for reconciliation.

### 4.3 Evaluation of Security Metrics

In this section, we evaluate the RAKG security via numerical examples and simulation. Specifically, we placed Alice, Bob, and Mallory at coordinates $A(5, 0)$, $B(15, 0)$ and $M(10, 5\sqrt{3})$, respectively, within an area of $20 \times 10$. Each multipath channel consisted of one LoS path and two NLoS paths produced by two randomly placed scatters within the setup area. We used the profile data of a customized RA that was designed for channel randomization purpose in [22]. The RA has 360 antenna modes, achieved by rotating a log periodic antenna by one degree.

We first computed $p_0$ and $p_1$ from the RA profile according to Eqs. (9) and (10), by sampling one million channel states under each antenna mode. This resulted in $p_0 = 0.7399$ and $p_1 = 0.9114$. Then, we evaluated the performance of our RAKG protocol by implementing the key generation process and the MitM attack. For comparison, we also implemented our protocol when Alice was equipped with an OA. We call this protocol as OAKG. We set the number of probing rounds to 50 million and the $\beta$ parameter for quantization to 0.4. We assumed that the channel coherence time lasts for 10 probing rounds.

First, we compared the randomness of the generated key with OAKG and RAKG, to show the impact of our RAKG protocol on increasing channel randomness. In Fig. 5, we show the first 500 bits of the initial bitstream generated by Alice. The generated key indicates significant randomness with our RAKG protocol. Further, we tested the key randomness using NIST test suite [3] with the $10^3$-bit length bitstream. The test results are typically in the form of a $p$-value, which must exceed 0.01 for a pass. We summarized the test results in Table 1, where the $p$-values are all significantly greater than 0.01 for RAKG, while only one of them is beyond 0.01 for OAKG.

We further present the performance of RAKG on defending against the MitM attack and compare it with the OAKG scheme, in Table 2. We can see that the performance of the adversary with RAKG was much worse than that with OAKG. Noticeably, the $\text{KRE}$ and $\text{KRR}$ of RAKG is about 1/2 and less than 1/3 of OAKG’s, respectively. Besides, we observed that the number of recovered 0-bits is always less than that of 1-bits in both cases. The reason is that in this simulation, when the adversary’s measurements from both transceivers are below $\varphi_0$, these two measurements are not significantly correlated and close to each other. In contrast, for

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>OAKG</th>
<th>RAKG</th>
<th>OAKG</th>
<th>RAKG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunities (%)</td>
<td>20.14</td>
<td>20.46</td>
<td>12.71</td>
<td>12.71</td>
</tr>
<tr>
<td>Resulting key length</td>
<td>30473949</td>
<td>54358764</td>
<td>32272550</td>
<td>32272550</td>
</tr>
<tr>
<td>$\text{KRE}$ (%)</td>
<td>47.51</td>
<td>47.66</td>
<td>97.46</td>
<td>97.46</td>
</tr>
<tr>
<td>$\text{KRR}$ (%)</td>
<td>47.39</td>
<td>5.17</td>
<td>19.90</td>
<td>19.45</td>
</tr>
</tbody>
</table>

Table 1: $p$-value of NIST test results

Table 2: Simulation results of the MitM attack, assuming that the channel coherence time lasts for 10 probing rounds
We implemented RAKG and OAKG using five commodity Wi-Fi routers (TP-link 4300) and the Atheros CSI tool [29]. Alice and Bob operated in broadcast mode to exchange probes, whereas Mallory monitored both the M-A and M-B channels. However, due to the limitation of this CSI tool’s working mode, it can only operate under the broadcast mode. Thus, except for the three routers (shown in Fig. 6 in blue) to measure $h_{ab}$, $h_{ba}$, and $h_{am}$, we added two more routers (shown in Fig. 6 in purple) at Bob and Mallory’s side to measure $h_{bm}$. We did not implement jamming and injection attacks in our experiments but simulated the attack with collected data, since [27] already demonstrated a reactive jamming attack with a success rate above 99.9%. In this paper, we focus on evaluating the performance of our RAKG protocol defending against the MitM attack instead of showing the feasibility of the attack.

We set the probe sampling rate to 20Hz, which translates to a gap of 50ms between probes. For RAKG, Alice was equipped with an RA of 3×3 square shaped metallic pixels that are connected by 12 p-i-n diode switches with fast reconfigurability. The radiation patterns of some typical modes can be found in [21], and we switched antenna modes among 4096 available antenna modes. However, we only have the data of the antenna patterns for 253 modes, which prevent us from obtaining $p_0$ and $p_1$ among all antenna modes. Hence, for the experiment results, we did not compute $p_{key}$.

### 5.2 Metrics for Protocol Performance

Since our main contribution lies in the quantization phase, we use the following three performance metrics to quantify the performance of the generated random bits (the initial bitstream after reconciliation), mainly from the efficiency and randomness point of view. These are complementary to the metrics defined in Sec. 4.1 for the MitM attack.

**Bit mismatch rate:** the ratio of bits that do not match between Alice’s and Bob’s initial bitstreams.

**Entropy:** captures the uncertainty of a random variable. We estimate the entropy of a bitstream with the approximate entropy test of NIST test suite [3], since it is a measure to quantify the amount of regularity and unpredictability of fluctuations over a time series.

**Secret bit rate:** The average number of secret bits extracted per collected sample. It is measured in terms of final output bits produced after fixing the bit mismatches with information reconciliation, and subtracting the bits revealed to the adversary.

### 5.3 Real-world Experiments

In this section, we report results derived from data collected from real-world experiments under different scenarios, and compare the performance of RAKG with OAKG.

#### 5.3.1 Experiment A: static indoor environment, same room.

We performed our first experiment inside an apartment unit’s living room, where the distance of A-B, M-A, and M-B is 3m, 1.4m, and 2.5m, respectively. The living room was almost empty, and there was no human activity during the experiment. Thus, there was minimal interference to the wireless channel. Fig. 7 shows the raw amplitude of CSI collected by Alice, Bob, and Mallory plotted against the index of probing round. As expected, there are not many variations for all the channels with OAKG, while the A-B and M-A channels are much randomized with RAKG. More importantly, as also demonstrated by [12], when the physical channel is stable, the channel reciprocity is low when OAKG was applied (the curves for Alice and Bob do not follow each other). The variations of the static channel are caused by hardware imperfections and thermal effects that are
not reciprocal, as a result, the bit mismatch is relatively high (about 0.5, shown in Table 3) for OAKG. Thus, it is not possible to extract the secret bit at a fast rate. In contrast, for RAKG, the A-B channel can extract secret bits at a rate of about 0.4, as shown in Table 3. This is because the channel variations caused by antenna mode change dominated the channel noise.

Table 3: Protocol performance with MitM attack when \( d = 3 \)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>RAKG</th>
<th>OAKG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mismatch fraction</td>
<td>0.2117</td>
<td>0.1315</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.4914</td>
<td>0.5336</td>
</tr>
<tr>
<td>Secret bit rate</td>
<td>0.3909</td>
<td>0.4551</td>
</tr>
</tbody>
</table>

5.3.2 Experiment B: indoor environment with human activity, same room. Next, we did the second experiment in our lab in which involved some human activities. Specifically, we put all the routers on the floor as shown in Fig. 8, and let a person move around the area with a vacuum. Fig. 9 shows the raw amplitudes of CSI collected by Alice, Bob, and Mallory. Compared with the results from Experiment A (Fig. 7), we can see that all the channels are randomized, and there is a high degree of reciprocity for the A-B channel. The results of MitM attack are shown in Table 4. We can see that under RAKG, both the number of opportunities and key recovery rate were reduced. However, we notice that: 1) The revealed bits are mostly 0-bits under both RAKG and OAKG; this can be explained by the fact that most of the opportunities the adversary found were type \( O_0 \) in this experiment. However, this is not a real issue if the overall number of bits revealed to Mallory is sufficient, since he is not interested in a specific key. 2) In this scenario, the result of key recovery efficiency for RAKG is higher than that of the simulation results. This is because the RA we used in the experiment is designed to steer several directions from communication, while we selected the antenna mode in each probing round from all switch combinations at random. The antenna modes with a low gain in the direction of the A-B channel caused packet losses. The channel measurements were likely generated from antenna radiation patterns with large enough gains. Hence, several adjacent CSI measurements can be similar and bring opportunities to Mallory. Nevertheless, the overall key recovery rate is still low. In conclusion, since an RA with more distinct antenna patterns was used in the simulation, the key recovery efficiency of it was much lower than the experiment results.

Table 4: MitM attack results for Experiment B

<table>
<thead>
<tr>
<th>( d )</th>
<th>RAKG</th>
<th>OAKG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunities [%]</td>
<td>6.39</td>
<td>10.19</td>
</tr>
<tr>
<td>Bits recovered (0/1)</td>
<td>11/4</td>
<td>15/7</td>
</tr>
<tr>
<td>Resulting key length</td>
<td>310</td>
<td>310</td>
</tr>
<tr>
<td>KRR [%]</td>
<td>78.95</td>
<td>73.33</td>
</tr>
<tr>
<td>KRE [%]</td>
<td>47.5</td>
<td>6.94</td>
</tr>
</tbody>
</table>

5.3.3 Experiment C: indoor environment with human activities, different rooms. We performed this experiment in different rooms of an apartment unit and put routers on the desk. The layout is shown in Fig. 10. We involved human activities by letting a person walk in the living room back and forth. The MitM attack results are shown in Table 5, which were similar to the results of Experiment B. Both the key recovery rate and key recovery efficiency of OAKG are reduced significantly compared with Experiment B. However, the MitM attack performs better for OAKG if Mallory is closer to Alice or Bob. In contrast, the distance has only a small impact on the MitM attack results under RAKG. This is because when the channel is proactively randomized with an RA, the antenna mode has a more significant impact on the multipath gains than human activities. Thus, it makes the channel correlation less sensitive to the distance.
Table 5: MitM attack results for Experiment C

<table>
<thead>
<tr>
<th></th>
<th>RAKG</th>
<th>RAOG</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Opportunities [%]</td>
<td>7.09</td>
<td>8.43</td>
</tr>
<tr>
<td>Bits recovered (0/1)</td>
<td>9.2</td>
<td>12.3</td>
</tr>
<tr>
<td>Resulting key length</td>
<td>270</td>
<td>291</td>
</tr>
<tr>
<td>KRE [%]</td>
<td>58.82</td>
<td>65.22</td>
</tr>
<tr>
<td>KRR [%]</td>
<td>3.70</td>
<td>3.54</td>
</tr>
</tbody>
</table>

6 RELATED WORK

6.1 Physical-Layer Secret Key Generation

Since Wyner proposed the wiretap channel in 1975 [28], many works advanced the theoretical aspects of physical-layer secret communications. Later, Maurer proposed the idea of common randomness, in which two parties can both tune to a common radio signal source and extract a secret key from it [17]. Mathur et al. [16] first proposed a practical RSS and CSI based key generation algorithm for indoor scenarios. Later on, Jane et al. [12] investigated the key generation rate for RSS-based key generation protocols in various channel conditions. It was shown that only highly dynamic mobile scenarios support key generation at a high rate. For static environments, the key generation rate is too low as the channel lacks randomness. To alleviate this problem, various approaches have been proposed. For example, MIMO channel coefficients were used to increase the key generation rate by Jon et al. [25]. Also, context-based pairing protocols utilize the randomness inherent in the observed environment to enable devices equipped with the corresponding sensors to extract the key from observed surrounding contents [20, 23].

Some previous works also used channel randomization to increase the secret key rate in key generation protocols. Specifically, Aono et al. [1] also adopted reconfigurable antennas to randomize the channel and relied on channel reciprocity to generate keys. Mehmood et al. [19] characterized the impact of RA complexity on key generation performance in different static propagation environments.

However, the goal of the above works is to only defend against passive eavesdroppers and increase the secrecy/key rate. Under an active attacker, the adversary can bypass the assumption of channel decorrelation even if it is far away from the legitimate devices. In particular, Eberz et al. [9] showed the possibility of a MitM attack against physical-layer key generation protocols [16] by injecting packets to legitimate transceivers upon observing opportunities, where the attacker can infer up to 47% of the generated key. Several countermeasures were suggested by their work, such as timing-based attack detection, but none of them can effectively prevent the MitM attack.

6.2 Trust Establishment without Prior Secrets

Trust establishment methods aim to establish authenticated secret keys between two or more parties over the wireless channel. Their goal is to achieve key confidentiality against eavesdroppers, as well as message integrity and authentication against active attacks. They usually rely on a cryptographic key exchange protocol such as Diffie-Hellman, and require out-of-band (OOB) secure channels for authentication [2, 5]. However, OOB channels often require additional hardware interfaces and human interaction, which affects cost, compatibility, and usability. Other works such as I-code [7] proposed in-band approaches to protect message integrity and indirectly authenticate devices through their presence, by exploiting ON-OFF keying and error detection codes. These works still rely on traditional cryptographic primitives, and thus do not enjoy information-theoretic secrecy for key establishment.

6.3 Enhancing PHY-layer Security with Channel Randomization

Channel randomization is viewed as one of the proactive/dynamic defense (or moving target defense) mechanisms. It has been applied for enhancing the security of wireless communications, in terms of countering both passive and active attacks.

Apart from key generation protocols, for PHY-layer secret communication in general, Haitham et al. [11] defend against multi-antenna eavesdroppers in RFID communication, by rotating eight directional antennas equipped at the transmitter’s side using a fan motor and randomly selecting the antennas. Pan et al. proposed ROBin [22] to enhance the security of orthogonal-blinding based secret communication against known-plaintext attacks by multi-antenna eavesdroppers, also by varying the channel state with rotating directional antennas.

To defend against signal cancellation attacks (or correlated jamming) and protect message integrity, Pan et al. [21] combine ON-OFF keying modulation with channel randomization via a reconfigurable antenna. The changing channel prevents the attacker from generating the desired signal to inject, since it would require exact channel states. The signal cancellation attack is quite different from the MitM attack considered in this work, while ON-OFF keying requires PHY-layer modifications.

For PHY-layer key generation, it is challenging to defend against active attacks at the same time. Our work gives a low-cost approach to prevent the MitM attack without using OOB channels, nor requiring changes to the PHY-layer protocol.

7 DISCUSSION

Our adversary model directly follows that proposed in [9], which does assume the adversary exploits the same properties as legitimate transceivers: the broadcast nature of wireless channels and channel reciprocity. But in practice, the adversary can have better attack strategies combining channel estimation with power control. The probe signals sent by Alice and Bob consist of preamble, header, payload, and CRC. Hence, Mallory can estimate the M-A and M-B channels with the preamble. A better attack strategy for Mallory is to jam the legitimate packets once the channel is estimated, and immediately inject packets with different power, to cause similar channel estimates at Alice and Bob. In this way, Mallory does not need to wait for opportunities, he can keep jamming and injection until a significant portion of the generated key is revealed.

All the physical-layer key generation approaches without additional device authentication mechanisms can be vulnerable to this type of MitM attack. But fortunately, we can detect and prevent this attack in the first place. From the perspective of Alice (or Bob), she will receive two consecutive incoming packets in a short time period. The first one cannot be decoded since it is a legitimate
packet jammed by Mallory, while the second injected packet can be decoded. In this case, both Alice and Bob can know that the MitM attack is launched, so that they discard the corresponding estimate. In this way, the MitM attack becomes trivial. The attack implemented by Mallory can only deny the key generation process between Alice and Bob, but never get any knowledge about the generated key.

8 CONCLUSIONS

In this paper, we proposed a physical-layer key generation protocol which is resistant to MitM attacks via channel randomization. We leveraged an RA at one of the legitimate transceivers to proactively randomize the channel state across different channel probing rounds. We theoretically analyzed the key guessing capability of the MitM adversary and showed that our approach can be significantly reduce using channel randomization. We conducted extensive simulations and real-world experiments to evaluate the performance of channel randomization. Results show that our RAKG protocol can successfully prevent the MitM attack by reducing its advantage to nearly random guess.

REFERENCES

APPENDIX

8.1 Key Agreement Scheme (Schürmann and Sigg [23])

Denote the set of possible words and codewords of RS with a given antenna mode \( n, k \). Alice and Bob can derive a key from the key derivation function \( \delta \) and Eq. (5), we described the value of the time-varing channel at time \( t_i \) during the \( i \)-th probing round as \( t_i \) and \( t_i' \), respectively. Denote the antenna mode used at round \( i \) as \( u_i \). Assume that \( O_1 \) is found at probe round \( i - 1 \), then Mallory implements the attack at \( t_i \) and gets the key bit as 1, then we can express \( p_1 \) as:

\[
p_1 = \mathbb{P} \left( \tilde{h}_{ma}(U_i, Ama(t_i)) > q_+ \bigg| \tilde{h}_{am}(U_i-1, Aam(t_i-1)) < d, \tilde{h}_{am}(U_i-1, Aam(t_i-1)) > q_+, \tilde{h}_{bm}(Aam(t_i)) + q_+ \right)
\]

where \( \tilde{h}_{xy} \) denotes the amplitude of the \( x-y \) channel. To consider the worst case for the protocol, we assume the physical channel remains static from finding an opportunity until finishing the attack, which means:

\[
A_{xy}(t_i-1) = A_{xy}(t_i') = A_{xy}(t_i) \quad \forall x-y \text{ channel}
\]

Besides, with channel reciprocity, we have

\[
A_{yx}(t_i) = A_{xy}(t_i) \quad \forall x-y \text{ channel}
\]

Then we ignore the index of time in CSI and use \( A_{xy} \) to denote the time-varying multipath components, Eq. (15) can be rewritten as:

\[
p_1 = \mathbb{P} \left( \tilde{h}_{ma}(U_i, Ama) > q_+, \tilde{h}_{mb}(Amb) < q_+ \bigg| \tilde{h}_{am}(U_i-1, Aam) < d, \tilde{h}_{am}(U_i-1, Aam) > q_+, \tilde{h}_{bm}(Amb) > q_+ \right)
\]

where we simplify Eq. (18) with channel coherence and reciprocity. Obviously, for the channel \( h(U, A) \), the antenna mode part \( U \) and the multipath part \( A \) are independent. Hence, Eq. (19) can be expressed as:

\[
p_1 = \int_{-\infty}^{+\infty} \mathbb{P} \left( \tilde{h}_{am}(U_i, aam) > q_+ \bigg| \tilde{h}_{am}(U_i-1, aam) < d_{max}, \tilde{h}_{am}(U_i-1, aam) > q_+, \tilde{h}_{bm}(Amb) > q_+ \right) f_{Aam}(aam) daam
\]

by using the property that the joint probability of two independent random variables is the product of their marginal probabilities, where \( f_{Aam}(aam) \) denotes the distribution of \( A_{am} \) that includes all the time-varying multipath components of the M-A channel.

As the antenna modes for different probing rounds are independently and uniformly selected, then \( h_{am}(U_i-1, aam) \) and \( h_{bm}(U_i, aam) \) are independent. Besides, the M-A, M-B, and A-B channels are
where each time-varying multipath component can be expressed as $A_l(t) = X_l(t) + jY_l(t)$, where $X_l(t)$ and $Y_l(t)$ are independent and identically distributed (iid) stationary time-varying Gaussian random variables with the same variance $\sigma^2_l$, and mean $\mu_{x,l}$, $\mu_{y,l}$, respectively.

For indoor scenario, denote $l = 0$ as the LoS component, and $l = 1, \cdots, P$ as the NLoS components, we have:

$$\mu_{x,l} = \begin{cases} \mu_{x,0}, & l = 0 \\ 0, & l = 1, \cdots, P \end{cases}$$

$$\mu_{y,l} = \begin{cases} \mu_{y,0}, & l = 0 \\ 0, & l = 1, \cdots, P \end{cases}$$

Then Eq. (26) can be rewritten as:

$$h(u, A(t)) = \sum_{l=0}^{P} g(u, \theta_l) A_l(t)$$

where $X(u, t)$ and $Y(u, t)$ are iid stationary time-varying Gaussian random variables with the same variance $\sigma^2(u)$. The mean of them is $\mu_x(u)$ and $\mu_y(u)$, respectively.

$$\sigma^2(u) = \sum_{l=0}^{P} (g(u, \theta_l) \sigma_0)^2 = \sigma_0^2 \sum_{l=0}^{P} g^2(u, \theta_l)$$

$$\mu_x(u) = \sum_{l=0}^{P} g(u, \theta_l) \mu_{x,l} = g(u, \theta_l) \mu_{x,0}$$

$$\mu_y(u) = \sum_{l=0}^{P} g(u, \theta_l) \mu_{y,l} = g(u, \theta_l) \mu_{y,0}$$

which means for any antenna mode $u$, the amplitude of the multipath channel $(h(u, A(t)))$ follows Rician distribution with PDF:

$$f(u) = \frac{x}{\xi^2(u)} e^{-\frac{x^2 + \xi^2(u)}{2 \xi^2(u)}} I_0 \left( \frac{\xi u}{\xi^2(u)} \right)$$

and the mean and the variance of $\hat{h}(u, A(t))$ are

$$\mu(u) = \xi(u) \sqrt{\pi/2} L_{1/2} \left( -\frac{\mu(u)}{2 \xi^2(u)} \right)$$

$$\sigma^2(u) = 2 \xi^2(u) + \mu^2(u) - \frac{\pi \xi^2(u)}{2} L_{1/2} \left( \frac{\mu(u)}{2 \xi^2(u)} \right)$$

where $L_{1/2}()$ is the the Laguerre polynomial, and $L_{1/2}^2()$ is the square of it.

With the distribution of $\hat{h}(u, A(t))$ above, Eq. (23) can be further expressed as:

$$p_1 = \mathbb{P} \left( \hat{h}_{ma}(U_i, A_{ma}) > q_+ \right)$$

$$= \sum_{u \in U} \mathbb{P} \left( \hat{h}_{ma}(u, A_{ma}) > q_+ \right) p_U(u)$$

$$= \sum_{u \in U} Q_1 \left( \frac{\hat{v}_{ma}(u)}{\sigma_{ma}(u)} \right) \frac{q_+}{\sigma_{ma}(u)}$$

$$= \frac{1}{|U|} \sum_{u \in U} Q_1 \left( \frac{\hat{v}_{ma}(u)}{\sigma_{ma}(u)} \right) \frac{q_+}{\sigma_{ma}(u)}$$

where $p_U(u)$ is the PMF of antenna mode selection, which is $\frac{1}{|U|}$ by assuming the antenna mode is uniformly selected.

Similarly,

$$p_0 = \frac{1}{|U|} \sum_{u \in U} \left( 1 - Q_1 \left( \frac{\hat{v}_{ma}(u)}{\sigma_{ma}(u)} \right) \right)$$

where $\hat{v}_{ma}(u)$ is the mean of the multi-path channel $(\hat{h}(u, A(t)))$.