Improving Quantization for Channel Reciprocity based Key Generation

Paul Walther*, Carsten Janda*, Elke Franz*, Mathias Pelka†,
Horst Hellbrück†, Thorsten Strufe*, and Eduard Jorswieck*

*TU Dresden, Dresden, Germany
{<firstname>.<lastname>}@tu-dresden.de

†CoSA
Fachhochschule Lübeck, Lübeck, Germany
{<firstname>.<lastname>}@fh-luebeck.de

***Abstract*—Quantization, and the fact that channel characteristics are independent and identically distributed so far have received only little attention in reports about actual implementations of physical layer key generation schemes. They are merely assumed for channel reciprocity based key generation, although the secret key generation significantly relies on them.**
We set out to design a quantization preprocessing as well as an online quantization scheme which favours i.i.d. and uniform distribution of the generated values to achieve high entropy and key rates, and calculate the resulting mutual information between communication partners in a large, realistic measurement study. Our experiments indicate a remarkable increase in mutual information, and underline the applicability to various quantization and key generation schemes.

***Index Terms*—Physical Layer Security, PhySec, Channel Reciprocity based Key Generation, Secret Key Generation, Quantization**

I. Introduction

Physical Layer Security (PhySec) is an actively discussed and promising contender for energy-efficient secure wireless communication [20], [26]. One direction of PhySec aims at secure key generation (SKG) between two legitimate communicating parties Alice and Bob, relying on the exclusivity of the characteristics of their shared channel. Both parties perform noisy and somewhat diverging measurements of realisations of the common randomness, and implementing Channel Reciprocity-based Key Generation (CRKG), they agree on a shared secret key [28].

PhySec key establishment protocols have been formally analysed, and their information theoretical security is proven [4]. The proofs rely on three major assumptions regarding the input values, i.e., the bit sequences obtained from the channel estimates: reciprocity between the legitimate partners, an independent identical distribution (i.i.d), and high entropy [5]. Reciprocity is needed in order to generate matching keys between Alice and Bob. The *channel reciprocity theorem* covers this assumption and states that reciprocity is given for their common channel [25]. The obtained measurements should to be i.i.d. in order to prevent predictability of future values from existing ones, which would result in predictable and, thereby, insecure key material. The entropy of the input values should be as close as possible to the maximum to deliver sufficient innovative bits as key material. Low entropy would either result in predictable key bits or in a low secret key rate. In order to maximise the entropy, a uniform distribution of the input values should be targeted [6]. In conclusion, these three requirements are crucial for CRKG.

The quantization step transforms the obtained measurements to bit vectors. It hence is the natural step in the PhySec SKG to optimize towards these requirements. In conventional transmission processes, a quantization scheme intends to reconstruct the original samples as good as possible from the potentially noisy transmission. As CRKG aims to extract random channel characteristics by treating the obtained values as a source of randomness, this approach does not fit.

Since the subsequent steps for SKG can only reduce the available entropy and, thereby, the effective key rate, the quantization should leverage as much randomness with high entropy as possible. Furthermore, the quantization can facilitate an independent distribution in time by decoupling the channel characteristics from the obtained samples. In combination, the quantization step is a suitable point to yield independent and uniformly distributed input values.

Existing approaches tend to rely on global knowledge, i.e., they use the average over *all* measurements to define a threshold for quantization. Since SKG works with subsequent channel measurements, this global knowledge is an unrealistic assumption. To avoid its use, several existing schemes use buffering of data points [2], [17], [24], [12], [16]. Thereby, local knowledge is created, on which similar calculations can be performed. Buffering comes with the disadvantage of performance hits – since the protocol flow has to wait until the next buffer frame is filled, the effective performance drops. We hence suggest to avoid buffering and use the subsequent measurements directly.

Current implementations do not fully consider the whole

This work is partly supported by the German Research Foundation (DFG) through CRC 912 “HAEC”, the Cluster of Excellence cfaed and by the Sächsische AufbauBank (SAB) through cluster 3125 “SATURN”.

978-1-5386-4413-3/18/$31.00 ©2018 IEEE

potential of the quantization step within the CRKG process. Therefore, we propose a new preprocessing step for quantization, which extracts the random channel characteristics. We additionally propose a new quantization scheme that provides uniformly distributed resulting values. A further advantage of the proposed processing steps is its *online* nature: Each new measurement is treated independently or only with knowledge of *previous* values. This allows to instantaneously start the SKG process as soon as the measurements are taken. Hence, no global knowledge or buffering is needed as in existing approaches.

We have evaluated our schemes against related methods (e.g.,[2], [17], [24]) with focus on the three major requirements for CRKG: i.i.d input values, their high entropy and reciprocity. The independence of the bit strings is assessed via the autocorrelation of the sequences. Entropy and reciprocity are evaluated via the mutual information between the legitimate partners. The results of our evaluation confirmed the suitability of the suggested approaches.

In this paper, we make the following *contributions*:

- introduction of a new online quantization scheme targeting high entropy through a uniform distribution,
- introduction of a new online preprocessing approach to effectively extract the channel characteristics,
- in depth analysis of main parameters of these schemes,
- comparison against existing approaches via real world measurements, especially regarding the entropy and distribution of the resulting bit vectors.

The remaining paper is structured as follows: Section II describes the CRKG system model and its general assumptions. Section III analyses state of the art implementations and their quantization. In Section IV, the new preprocessing and quantization approaches are introduced. Section V discusses the evaluation setup and the resulting findings. Section VI summarizes and gives an outlook.

II. Background

CRKG is an implementation of the source model of PhySec SKG described by Maurer and Ahlswede [18], [1], which intends to derive symmetric keys in the presence of a passive attacker. Two legitimate parties and an eavesdropper observe three different observations X, Y, Z from a common distribution P_{XYZ} of a random variable. By exchanging messages over an authenticated noiseless public channel, the legitimate partners can derive a common secret key from these observations unbeknownst to the eavesdropper. This exchange is divided into 4 steps: randomness sharing, advantage distillation, information reconciliation, and privacy amplification [4].

It has been proven that this procedure is information theoretically secure and that protocols fulfilling the respective requirements do exist [4]. Additionally, these proofs also hold, if the eavesdropper has additional knowledge about the channel by visual inspection of the environment or even by measuring the room itself. Furthermore, it was shown that this process requires significantly less energy than traditional key exchange primitives like ECDH [26].

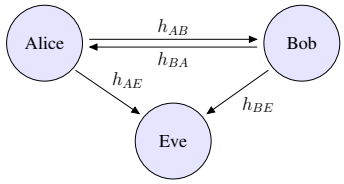


Fig. 1. Generic system model for CRKG: Alice and Bob measure the reciprocal channel and thereby obtain their estimates $\hat{h}_{AB}$ and $\hat{h}_{BA}$. Eve overhears this communication and estimates her own channels $\hat{h}_{AE}$ and $\hat{h}_{BE}$.

In general, the legitimate nodes Alice and Bob exchange messages over their channel (Fig. 1). Each partner measures the relevant channel characteristics of the received signals to determine the channel estimates $\hat{h}_{AB}$ and $\hat{h}_{BA}$[1]. In parallel, Eve listens to all transmissions and measures the same metrics as Alice and Bob. Thereby, she can estimate the channels $\hat{h}_{AE}$ and $\hat{h}_{BE}$. Following the reciprocity assumption, the channels $\hat{h}_{AB}$ and $\hat{h}_{BA}$ are highly correlated, whereas $\hat{h}_{AB}$ and $\hat{h}_{AE}$ (resp. $\hat{h}_{BA}$, $\hat{h}_{BE}$) are less correlated. Thus, Alice and Bob see roughly the same, whereas Eve sees a statistically totally different channel. Mathur et al. [17] claim that a distance of half a wavelength λ is sufficient to alter Eves channel estimates in such a way, that no inference to $\hat{h}_{AB}$ and $\hat{h}_{BA}$ is possible. This phenomenon originates from Jakes uniform scattering model and is called *spatial decorrelation* [8][2].

Since Eve cannot extrapolate the channel characteristics of $\hat{h}_{AB}$ from her observations, these properties can subsequently be used as input for key generation.

Quantization of the observed measurements belongs to the *randomness sharing* step. The quantization can further be divided into the quantization decision and a possible preprocessing. In general, the preprocessing is a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, which operates directly on the measurement values. The successive quantization transforms the analogue, time-discrete input measures into bits as a function $f : \mathbb{R}^n \rightarrow \{0,1\}^m$.

Each measurement acquired from the channel is denoted as x_t. Here, t denotes the start time of the measuring. Subsequently, n data points are recorded and composed into the vector $x_t = (x_t^1, x_t^2, \ldots, x_t^n), x_t^i \in \mathbb{R}$. The preprocessing transforms x_t into the vector x_t^*, which is in turn quantized into a bit string $b_t = (b_t^1, b_t^2, \ldots, b_t^m), b_t^i \in \{0,1\}$.

This bitstring is the final result of the quantization and will then be passed into the upcoming processing step. It is important to notice, that the conversion from measurement values into bit vectors also implicitly transforms their distribution. Hence, the quantization dictates the distribution of the input values for the whole CRKG process. Therefore, it is crucial for the whole process that this input values are in conformity with the assumed properties.

Furthermore, it is important to notice that this quantization does not target the *transmitted signals* but the *random channel characteristics*. Hence, it is inherently different to traditional

[1]Actually, the channels are time-dependent ($\hat{h}_{XY}(t)$). This time variation can be well modelled as an additional additive channel estimation error and is therefore included in our model. The notation is only omitted due to brevity.

[2]This assumption is quite accepted, but it may not be fulfilled in reality [23].

signal estimation and quantization, which has been studied thoroughly and is well understood.

III. State of the art

Since this work targets quantization, the related work is selected and examined with a special focus towards this step.

Existing CRKG schemes mainly focus on the protocol steps after the quantization (Sec. II). None of the presented works explicitly targets a uniform distribution or high entropy of the extracted randomness via optimization of the quantization step.

Further, current schemes are slow ($< 2\,bits/s$, [10]) compared to the proposed use as pseudo one time pad [15]. Hence, the performance of all protocol steps should be improved.

In general, existing CRKG implementations can be divided into two groups according to the channel characteristics used: either the received signal strength indicator (RSSI) or the complex channel state information (CSI). Both values represent, with different resolution, the transmitted energy, which was distorted during the transmission.

The RSSI value is a single value representing the qualitative signal strength of the whole received transmission. Thus, only few data points can be obtained from a single measurement. But this value is very common within hardware implementations, hence, even common off-the-shelf (COTS) hardware delivers it to higher layers. Scheme using this value (e.g., [2], [17], [12]) apply a common approach for quantization: Several measurements (e.g., 384) are buffered and the mean or median is defined as threshold for 1-bit quantization. Wallace and Sharma [24] expanded this scheme to MIMO systems.

The CSI includes all information about the transmission, hence, a single measurement yields much more data points for further usage. Obtaining the CSI is considered a rather specialized use case, for which in general customized hardware is needed. Nevertheless, there are efforts to make this information available on commodity hardware (e.g., Intel 5300 NIC [11]). CSI-based schemes like [16] and [27] also use buffering to calculate a threshold based on the Cumulative Distribution Function.

These quantization approaches are all based on the common idea to buffer several measurements to calculate statistics, hence, they are not considered *online*. Tope and McEachen [22] proposed an online scheme which uses the difference of consecutive RSSI values as reciprocal characteristic. However, this procedure produces low entropy key material and is thereby susceptible to attacks [12].

There are two major aspects of CRKG schemes: First, buffering is widely used, albeit it inherently introduces delays. Second, none of them evaluates the quantization step with respect to the requirements mentioned in Sec. II. Since this influence is neglected, the quantization schemes could no be optimized towards the fulfillment of those requirements.

IV. Approach design

To address the mentioned open points, we introduce a novel preprocessing step as well as two new quantization approaches. All of them work *online* and are designed to fully comply with the CRKG requirements.

Since efforts like [11] make CSI available on COTS hardware, we will focus on the usage of CSI.

Tab. I shows the notation used in the following.

TABLE I
Variables used

t	point in time	θ	quantization threshold
x_t	vector of measurement at t	n	length of vector x_t
	channel estimation at t	s_t	sliding window at t
x_t^*	preprocessed vector	w	window size
b_t	resulting bit vector	m	quantization width

A. Preprocessing

Channel Characteristics Estimation – CCE

The CRKG process relies on reciprocal channel characteristics, which vary when the channel changes. In general, these alternating characteristics are small compared to the overall channel estimate. Hence, the straight forward approach of applying the quantization directly on the whole channel estimate is biased by the magnitude of the estimate. In order to diminish this bias, the smaller channel characteristics need to be extracted from the total channel estimate.

To achieve this separation, we propose the following new preprocessing step: The overall channel estimate is interpreted as a composite of a static part (for the current situation) and the "interesting" characteristics. The static component will be approximated and subsequently deduct from the channel estimation. Thereby, the reciprocal channel characteristics will be extracted from the remaining part as well as normalized. This result is then passed on to the remaining CRKG steps.

For the approximation of the static part, we propose to apply a moving average over the last w measurements. The moving average is chosen because it is easy to implement in hardware while still delivering high estimation performance [21].

$$s_t^i = avg(x_t^i, x_{t-1}^i, \ldots, x_{t-w}^i), i \in \{1, 2, \ldots, n\} \quad (1)$$

$$s_t = (s_t^1, s_t^2, \ldots, s_t^n), s_t^i \in \mathbb{R} \quad (2)$$

$$x_t^* = x_t - s_t \quad (3)$$

To summarize, we first approximate the static part s_t at time t by applying the moving average to each data point within a measurement over the last w measurements (Eq. (1)). The averaging function avg is either the arithmetic mean or the median. In Eq. (3), this approximation s_t is deduct from the obtained measurements x_t. The resulting vector x_t^* represents the channel characteristics at time t.

B. Quantization

In general, quantization defines one or more thresholds θ, which divide the domain of the quantization function in different ranges. These ranges are subsequently labelled with the targeted bit strings (e.g. through Gray codes [9]). In our scenario, only 1 bit quantization was considered, hence, only a

single threshold θ is needed. We quantize by setting values $\geq \theta$ to 1 and values $< \theta$ to 0. Quantizing schemes aiming for more than 1 bit can easily be deduced from this base case. More advanced quantizer (e.g., vector quantizer using the Max Lloyd algorithm) are unsuitable for the current use case because of their increased complexity [7]. In the subsequent description, the averaging function avg may be again the arithmetic mean or the median.

Local Moving Average – LMA

The reasoning for the CCE preprocessing can also be applied to the quantization decision per se. Hence, the static part s_t is again estimated by applying a moving average over the last w measurements. The resulting values are then used as thresholds for each sample.

$$\theta_t = (\theta_t^1, \theta_t^2, \ldots, \theta_t^n) \tag{4}$$

$$\theta_t^i = avg(x_t^i, x_{t-1}^i, \ldots, x_{t-w}^i), i \in \{1..n\} \tag{5}$$

This quantization approach is effectively the proposed CCE scheme used as threshold determination. Hence, it is expected that their combination will not yield large improvements. Nevertheless, it is still considered as quantization scheme in order to compare its performance against the other approaches.

Results Based Adaption – RBA

The idea of RBA is to adapt the quantization decision so that its outcome is as close as possible to the uniform distribution.

This is realised by adapting the threshold θ based on the resulting number of zeros and ones after quantization. The underlying assumption is that the threshold is biased towards one side, if either outnumbers the other. Hence, the threshold needs to be adapted towards the opposite direction (8).

The width of the adaption is reduced in every step in order to converge towards an "optimal" value. After reducing the step size for 8 steps, it is kept stable for further adaptions - this aims for a quick convergence towards the optimal value without neglecting later values through too small step sizes.

$$y_t = \alpha \cdot avg(x_t^1, x_t^2, \ldots, x_t^n) \tag{6}$$

$$\alpha = \begin{cases} \frac{1}{2^t} & \text{if } t \in \{0, 1, 2, \ldots, 8\} \\ \frac{1}{2^8} & \text{if } t > 8 \end{cases} \tag{7}$$

$$\theta_t = \begin{cases} y_t & \text{if } t = 0 \\ \theta_{t-1} - y_t & \text{if } |\{b_{t-1}^i = 0, i \in \{1, 2, \ldots, n\}\}| > \lfloor \frac{n}{2} \rfloor \\ \theta_{t-1} + y_t & \text{if } |\{b_{t-1}^i = 0, i \in \{1, 2, \ldots, n\}\}| < \lfloor \frac{n}{2} \rfloor \\ \theta_{t-1} & \text{if } |\{b_{t-1}^i = 0, i \in \{1, 2, \ldots, n\}\}| = \lfloor \frac{n}{2} \rfloor \end{cases} \tag{8}$$

C. Sliding window size

The proposed mechanism has two major parameters: the averaging function used and the size of the "window" w. The median was chosen as averaging function, because of its robustness against outliers. The window size was determined by the following analysis of the measurements.

Since the obtained measurements describe the channels between Alice and Bob, they inherently vary strongly in between them. According to the channel model, these measurements are interpreted as events of a Gaussian distribution. To evaluate the stability of a given window size w, we took a sample of size w from the realisations of this distribution and calculated the selected average function. Then we computed the difference between this average and all observations outside of the current sample. The resulting estimation of the random, reciprocal part of the transmission is also expected to also be a Normal distribution, since this differencing does not alter the underlying distribution. Finally, we computed the standard deviation for each sequence of differences w.r.t. to the window size w.

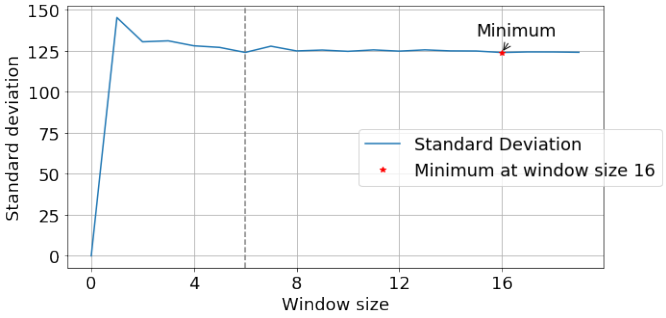


Fig. 2. Resulting standard deviation with respect to the window size used $w = \{1, 2, \ldots, 20\}$

Fig. 2 shows these standard deviations for $w = \{1, 2, \ldots, 20\}$. It is visible that the deviation decreases after the first peak, which denotes the convergence towards the "true" average. Further, the decrease is only minimal for $w > 6$, depicting that a stable solution has been found.

It has to be noted that the values depend on the current system setup. Altering the setup would probably result in different progressions. However, determine w again for each new situation would not scale. Thus, we suggest to settle on a value, which balances resource usage and generality – being big enough to retain generality claims, but also being small enough to only use a reasonable amount of resources.

V. Evaluation

The evaluation targets the three main requirements for the input sequences originating from the assumptions described in Section I. Hence, as evaluation metrics the mutual information between Alice and Bob and the autocorrelation of the respective single sequences were chosen.

The mutual information is defined as:

$$I(A;B) = H(A) + H(B) - H(A,B) \tag{9}$$

$$= \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} p(a,b) \cdot log \frac{p(a,b)}{p(a)p(b)} \tag{10}$$

Here, A and B denote the quantized random bit sequences derived by Alice and Bob from the channel estimates. $I(A;B)$ incorporates the two major components of the key establishment. On the one hand, the innovativeness of the single source is included through the entropy of each sequence. On the other hand, the reciprocity is included by means of the common distribution (in turn the joint entropy) of both sequences.

Further, the mutual information also represents the maximal achievable secure key generation rate. This rate is defined in

relation to Eves sequence (E):

$$C_s \leq min(I(A;B), I(A,B|E)) \tag{11}$$

Following the decorrelation arguments from Sec. II, we assume $I(A,B|E)) = 0$. Hence, $I(A;B)$ between Alice and Bob represents the maximal secure key rate [4]. Since the secret key rate is only governed by $I(A;B)$, this metric will also be used to evaluate the security of the new schemes.

For the autocorrelation, the sequence of quantized bit strings is interpreted as stochastic process, on which the Pearson correlation with itself is calculated for different lags. If the sequence is i.i.d., the autocorrelation peaks at the offset 0. All other offsets will yield significantly lower correlation values.

The obtained measurements are subsequent data points in the time domain. Evaluation will be done in this domain (*TIME*) and in the frequency domain (*FREQ*). Processing the data within the frequency domain is reasonable, since current transmission protocols us this domain for multiplexing or modulation. Processing within the time domain is reasonable, since it might result in higher information rates as stated by the data processing inequality [3], [13].

For the evaluation of our newly proposed schemes, we also implemented the following baseline approach. **In Measurement Average – IMA**: This method applies an average function over the current measurement and sets the threshold according to the output: $\theta_t = avg(x_t^1, x_t^2, \ldots, x_t^n)$. Using this approach with the median as averaging function is the most common amongst the related work.

Applying this method over all available measurements would be what was called quantization with global knowledge. Using this scheme with subsets of the available measurements is equal to buffering approaches like [17], [2], [24]. The authors of [10] define it as a baseline for quantization processes. All approaches described in Section III implement IMA.

A. Measurement Setup

To prove the feasibility of the new approaches, it was intended to maximise the resulting key rate – therefore, the channel was sampled with very high bandwidth to obtain as much characteristics as possible. Nevertheless, the proposed schemes will also work with lower sampling rates.

The evaluation uses data delivered by two wireless sensor nodes measuring the reciprocal channel impulse response (CIR) using ultra-wideband signals with 500 MHz bandwidth in the 4 GHz band. Messages are transmitted with impulse radio ultra-wideband which supports short pulses in the time domain, with a typical duration of less than 3 ns. The channel impulse estimation evaluates the perfect autocorrelation sequence properties of the preamble sequence of IEEE 802.15.4a compliant messages. The preamble sequence is a set of ternary symbols and is carefully designed to exploit the perfect autocorrelation properties to remove the side-lobes in the periodic correlation sequence. Peaks in the correlation correspond to the CIR of the channel [19]. We correlate the received preamble sequence against the expected preamble sequence and estimate the CIR using a leading-edge detection algorithm [14]. In our hardware, the sample interval is 1 ns, which allows detection of multipath components with a spatial resolution of 30 cm. The resolution of the ADC is 12 bit and is implemented with COTS components.

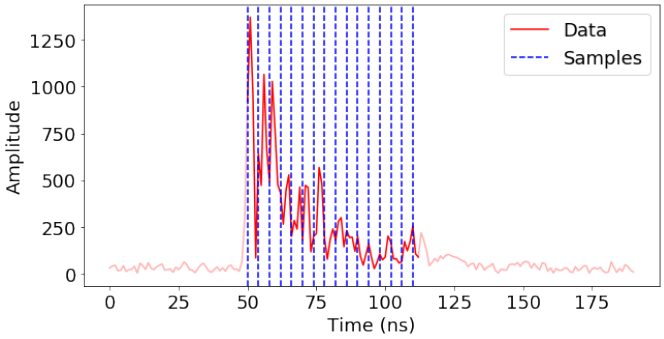


Fig. 3. Realizing a reduced width $m < n$ for quantized bit vector by implementing subsampling

Alice and Bob transmit messages under line-of-sight conditions in our laboratory and in the adjacent hallway. After reception of a packet, they estimate their CIR. Alice and Bob are stationary and five meters apart. The receivers are placed 1 m above ground to ensure optimal transmission quality. The line-of-sight path is interrupted when people move along causing random changes of the CIR. Alice and Bob transmit messages every 370 ms, generating 6 CIRs per second. Single measurements consist of 200 data points. We performed measurements for 6 days, generating more than 20 GB of data.

B. Parameters for Processing

To increase the maximal bit rate, we need to create as many bits as possible from each measurement. But to create meaningful statistics we should use fewer bits. If each x_t is quantized into a bit vector b_t of width m, than each sequence will have 2^m possible values. $P(A,B)$ will than in turn generate a matrix with 2^{2m} values. Thus, a quantization width m requires at least 2^{2m} measurements so that each possible event may occur once.

One reasonable value for performance is $m = 16$, although this would lead to the need of $2^{16} * 2^{16} = 4,294,967,296$ measurements. Our measurement lasting for 6 days yielded $1,336,180$ data points. So by $log_2(1336180) = 20.3497$ we have approximately 2^{20} values. Thus, with a quantization width of 10, we have at least one possible data point for each event of the joint probability distribution.

Fig. 3 depicts how the quantization width m was realised. The figure shows an instance of the obtained measurements. After the removal of the surplus values in the front and the back (noise and line of sight component), only the part bearing relevant information is left. Subsequently, m data points were taken in equidistant intervals as subsamples. The 1-bit quantization then yields a m bit wide bit string as result.

According to Jana et al. [12], this step makes this processing approach lossy. The lossiness has the disadvantage of reducing the available information. On the other side, the increased gap between the sub samples decreases the correlation between

TABLE II
RESULTING METRICS FOR DIFFERENT 16-BIT QUANTIZATION STRATEGIES WITH AND WITHOUT THE PROPOSED PREPROCESSING

	Without CCE preprocessing				With CCE preprocessing			
Method	$I(A;B)$	% of $I(A;B)_{max}$	$P(a)=0$	$P(b)=0$	$I(A;B)$	% of $I(A;B)_{max}$	$P(a)=0$	$P(b)=0$
Global	2.0722 bpcu	12.95%	9.25%	9.73%				
Block10	2.1992 bpcu	13.74%	9.45%	9.91%				
Block5	2.2664 bpcu	14.16%	9.55%	10.00%				
TIME IMA	3.0602 bpcu	19.13%	77.80%	76.34%	6.4074 bpcu	40.05% (**+20.92**)	80.33%	80.32%
TIME RBA	6.5645 bpcu	41.03%	31.97%	22.49%	10.0126 bpcu	62.58% (**+21.55**)	0.54%	0.20%
TIME LMA	10.8637 bpcu	67.90%	0.02%	0.04%	11.0478 bpcu	69.05% (**+1.15**)	0.01%	0.02%
FREQ IMA	1.6955 bpcu	10.60%	91.68%	92.53%	6.8652 bpcu	42.91% (**+32.31**)	80.36%	80.36%
FREQ RBA	2.5142 bpcu	15.71%	73.93%	77.05%	5.6722 bpcu	35.45% (**+19.74**)	0.23%	1.38%
FREQ LMA	11.2162 bpcu	70.10%	0.00%	0.00%	11.2671 bpcu	**70.42%** (**+0.32**)	0.00%	0.00%

TABLE III
RESULTING METRICS FOR DIFFERENT 10-BIT QUANTIZATION STRATEGIES WITH AND WITHOUT THE PROPOSED PREPROCESSING

	Without CCE preprocessing				With CCE preprocessing			
Method	$I(A;B)$	% of $I(A;B)_{max}$	$P(a)=0$	$P(b)=0$	$I(A;B)$	% of $I(A;B)_{max}$	$P(a)=0$	$P(b)=0$
Global	0.3950 bpcu	3.95%	11.13%	7.13%				
Block10	0.3907 bpcu	3.91%	5.57%	3.42%				
Block5	0.3812 bpcu	3.81%	2.93%	1.37%				
TIME IMA	0.0770 bpcu	0.77%	53.61%	54.59%	0.1845 bpcu	1.68% (**+0.91%**)	76.95%	76.76%
TIME RBA	0.3531 bpcu	3.53%	0.00%	0.00%	0.6395 bpcu	6.39% (**+2.86%**)	0.00%	0.00%
TIME LMA	1.7916 bpcu	16.29%	0.00%	0.00%	1.8079 bpcu	16.44% (**+0.15%**)	0.00%	0.00%
FREQ IMA	1.2378 bpcu	11.25%	77.49%	77.83%	0.1362 bpcu	1.24% (**-11.1%**)	77.39%	77.39%
FREQ RBA	1.2623 bpcu	12.62%	1.76%	5.08%	0.5567 bpcu	5.57% (**-7.05%**)	0.00%	0.00%
FREQ LMA	1.8983 bpcu	17.26%	0.00%	0.00%	1.9047 bpcu	**17.32%** (**+0.06%**)	0.00%	0.00%

data points within a measurement, since adjacent samples are stronger correlated than samples further apart.

Finally, an important parameter for the quantization process is the window size of the proposed preprocessing. According to the preliminary test (Fig. 2), the value $w = 6$ was used for the evaluation.

C. Results and Discussion

The results of the evaluation are presented in Tab. II and III. These tables have the following columns:

- *Method*: Describes the combination of data representation and quantization schemes used.
- $I(A;B)$: The mutual information between the legitimate communications partners achieved by the current scheme.
- *% of* $I(A;B)_{max}$: Relates the resulting mutual information to the maximal achievable mutual information $I(A;B)_{max} = H_{max} = \mathrm{ld}(2^m) = m$, which also represents the maximal achievable secure key rate. The main purpose here is to create a comparable metric over different values of m.
- $P(a) = 0$/$P(b) = 0$: The percentage of events in the event space of sequence A (resp. B), which have the empirical probability 0.
  Some quantization schemes do not yield uniformly distributed results. In this case, certain quantization results will occur with high probability, whereas others never occur, i.e., their relative frequency will be zero (Fig. 4(a)). A high percentage shows that a lot of possible quantization results do have a relative frequency of zero, which indicates a biased quantization scheme. Hence, this metric is an indicator regarding the uniform distribution of the resulting bit strings.

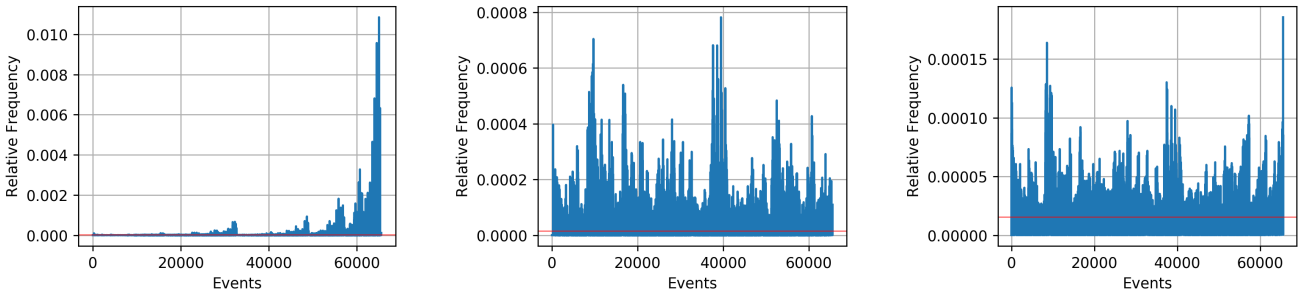
The tables combine the results for the proposed preprocessing as well as for the proposed quantization schemes. The left part of the table describes the different quantization approaches without the proposed preprocessing. The right side shows the same procedures only with the preprocessing added.

For all schemes, the median was used as averaging function. IMA was additionally implemented for buffering approaches with global knowledge (*Global*) and local knowledge using buffer sizes of 5 (*Block5*) and 10 (*Block10*).

a) Mutual Information / Achievable secure key rate: Tab. II shows results for the quantization width $m = 16$. In general the proposed preprocessing step increments the achieved mutual information in all cases. The minor increment for the LMA quantization scheme is salient. This stems from the fact that both schemes internally use the same mechanism of sliding average functions. But even in this case, adding the preprocessing increased the achievable mutual information.

Further, the preprocessing makes the online schemes more effective than those with global or buffered knowledge, which yield mutual information in the range from $12 - 14\%$.

The proposed quantization schemes perform significantly better than IMA as well as the global/local knowledge approaches. The mutual information increased by up to 59.5 percent points. Additionally, the percentage of results with probability zero was reduced, e.g., from 77.8% to 0.02%.



(a) No preprocessing and IMA (b) Preprocessing and IMA (c) Preprocessing and RBA

Fig. 4. Relative frequency of the single events with different processing approaches at 16 bit. A event is the occurrence of one of the 2^m possible bit strings. The red line shows the "optimal" value $p = 1/2^m$, which would be achieved by a perfectly uniform distribution.

Thus, the resulting distribution is significantly closer to the uniform distribution.

Tab. III reports results for quantization width $m = 10$. Overall, the results are significantly worse. The maximal achieved percentage for mutual information drops from 70.42% to 17.32%; the improvements through CCE are only minor and in the frequency domain even negative.

Nevertheless, the proposed quantization schemes still outperform the baseline approach and global/local knowledge schemes. There are increases of up to 15.52 points.

The significant drop of performance between 16 and 10 bit as well as the negative impact of CCE within 10-bit results $(-11, 1\%)$ is assumed to originate from the increased subsampling intervals. By increasing the distance of single data points, the characteristic reciprocal attributes are not captured any more. Since the process builds on these attributes, the resulting mutual information drops. Subsampling sizes of $m = \{4, 6, 8, 10, 12, 14, 16\}$ strengthen this assumption: The mutual information drops exponentially with decreasing m.

However, this only affects the evaluation and not the proposed solutions per se. The reduction of the quantization width was only introduced to create sound statistics. In real world scenarios, the quantization width would not be reduced, since the focus would shift towards high bit rates. Hence, the channel characteristics would reliably be captured, which enable the described positive results of the proposed solutions.

b) Uniform distribution of resulting bit strings: According to Sec. I and Sec. II the occurrences of the single bit vectors should be i.i.d. as well as uniformly distributed. With uniform distribution, each event (occurrence of bit vector b_j) would have the same probability $p(b_j) = \frac{1}{2^m}$. Hence, no possible bit vector should have a probability of 0. Both Tab. II and III show instances where the percentage of events with probability 0 is significantly larger than zero.

On the one hand, there are percentages of 92.5% for schemes without preprocessing. This shows that these schemes do not result in uniformly distributed bit vectors. It is worth to bring to mind, how the input data without any preprocessing is structured (Fig. 3). After a peak at the start the signal slowly decreases towards the end of the transmission. Throughout the decrease, there might occur peaks, e.g., when energy via reflected paths arrive at the receiver. If this time domain data is used, this structure is also visible in the distribution of the bit vectors. In practice, this is visible, e.g., by applying the IMA approach: During the quantization decision, the first values (the "peak") will almost surely be quantized into ones, because they clearly are in the upper range of the value space. As shown in Fig. 4(a) the greater values appear clearly more often than the smaller ones (since their first bits are all "1").

On the other hand, the majority of the proposed online schemes with CCE show a percentage of 0% for events with probability 0. This can be interpreted as indicator that the resulting bit vector distribution is closer to uniform. An intermediate step is the result of IMA with CCE as depicted in Fig. 4(b). Although the preprocessing managed to distribute the events more evenly, the value range of the frequency shows that there are gaps within the graph. For $m = 16$, the probability of uniformly distributed events would result in $\frac{1}{2^{16}} = 1,53\text{e-}5$, but the actual single values lie in the range of 1e-4. Hence, there are still events with probability 0 (Tab. II).

Finally, the distribution in Fig. 4(c) (CCE with RBA) is close to this "optimal" value. Tab. II confirms, that there are no events with probability 0.

Altogether, LMA combined with CCE preprocessing yields the best results with a mutual information of 69.05% of the quantization width $m = 16$ for time domain data and 70.42% for frequency domain data.

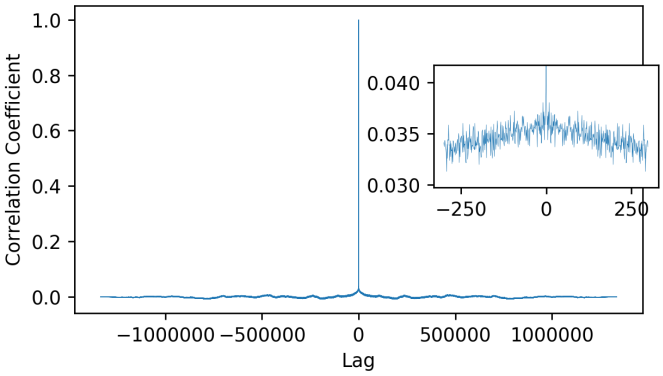


Fig. 5. Example autocorrelation function for IMA with $m = 16$.

c) Independence of quantized bit strings: The autocorrelation shows a clear peak for a lag of 0 (Fig. 5).

The next correlation maxima have values of 0.0397, 0.038 and 0.0371 at the lags 1, 2, and 93. These low correlation coefficients for lags $\neq 0$ suggest that no linear relation between the subsequent bit vectors exist. The very low correlation combined with the single peak at lag 0 strongly indicates an independent distribution of the underlying sequence. All schemes yield similar curves (in similar value ranges) for their autocorrelation function. Hence, it can be assumed that all schemes fulfil the described i.i.d. assumption.

D. Security considerations

The CRKG process is proven to be secure (Sec. II). The security proofs are based on the privacy amplification, which is not influenced by our approaches. Since our new schemes solely target the randomness sharing and, thereby, the assumed preconditions, these security proofs still hold. Additionally, the real world measurements showed that the new approaches result in bit strings, which are significantly closer to the assumptions of those proofs than the bit strings delivered by existing ones (especially regarding the uniform distribution).

VI. Conclusion

Within this paper, we thoroughly investigated the quantization step of CRKG. We proposed a new approach for preprocessing (CCE) and two new approaches for quantization (LMA, RBA). These novel schemes target the essential assumptions of i.i.d bit sequences with high entropy and reciprocity between the communication partners. In contrast to existing solutions, they all work in a purely online fashion. To evaluate the approaches, real world measurements of communication channels were obtained and processed with the proposed schemes and for comparison with existing quantization schemes.

Further, it was investigated how the main parameter for CCE and LMA, the window size w, can be determined. Based on the real world measurements, it was concluded that $w = 6$ is a sufficient window size in this scenario.

The mutual information was chosen for comparison, since it depends on the entropy of the single measurements as well as on the reciprocity between the legitimate communication partners. In addition, it represents the maximal achievable secure key rate. The results clearly showed that the CCE leveraged greater mutual information in almost all scenarios, with differences up to $+32.3$ percent points. The same holds for proposed quantization schemes with increases up to $+21.9$ percent points for RBA and $+59.5$ for LMA. Further, the distribution of the resulting bit vectors produced by the suggested approaches are closer to uniformity. Finally, our results revealed, that directly using the data residing in the time domain does not provide an additional advantage.

Although the results paint a clear picture, several ensuing questions arise. The influence of the quantization width m is not finally settled, since it constitutes another trade-off between efficiency and correlation. The reduction of m also yielded an unexpectedly strong reduction of the resulting mutual informations.

In upcoming studies, the influence of the proposed schemes will be further investigated in two ways: On one side, the inherent entropy of the mutual information will be examined further – which in turn influences the quality of the key material and the resulting key rate. On the other side, the information leaked towards Eve needs to be examined, since this a crucial information for the privacy amplification step.

References

[1] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[2] T. Aono *et al.*, "Wireless secret key generation exploiting the reactance-domain scalar response of multipath fading channels: RSSI interleaving scheme," in *Wireless Technology*, 2005.

[3] N. J. Beaudry and R. Renner, "An intuitive proof of the data processing inequality," arXiv [quant-ph], Tech. Rep., 2011.

[4] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[5] C. Chan *et al.*, "Multivariate mutual information inspired by secret-key agreement," *Proc. of the IEEE*, vol. 103, no. 10, pp. 1883–1913, 2015.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[7] M. Garey *et al.*, "The complexity of the generalized lloyd - max problem," *IEEE Trans. on Information Theory*, vol. 28, no. 2, 1982.

[8] A. Goldsmith, *Wireless Communications*. Cambridge univ. press, 2005.

[9] F. Gray, "Pulse code communication," US Patent US2 632 058, 1953.

[10] R. Guillaume *et al.*, "Fair comparison and evaluation of quantization schemes for phy-based key generation," in *Int. OFDM Workshop*, 2014.

[11] D. Halperin *et al.*, "Tool Release: Gathering 802.11N Traces with Channel State Information," *SIGCOMM CCR.*, vol. 41, pp. 53–53, 2011.

[12] S. Jana *et al.*, "On the effectiveness of secret key extraction from wireless signal strength in real environments," in *Mobile Computing and Networking*. ACM, 2009.

[13] J. B. Kinney and G. S. Atwal, "Equitability, mutual information, and the maximal information coefficient," in *Proceedings of the National Academy of Sciences of the USA*, vol. 111, no. 9, 2014, pp. 3354–3359.

[14] M. J. Kuhn, J. Turnmire, M. R. Mahfouz, and A. E. Fathy, "Adaptive leading-edge detection in UWB indoor localization," in *Radio and Wireless Symposium (RWS), 2010 IEEE*. IEEE, 2010, pp. 268–271.

[15] L. Lai, Y. Liang, H. V. Poor, and W. Du, "Key generation from wireless channels," *Physical Layer Security in Wireless Comm.*, pp. 47–92, 2013.

[16] H. Liu *et al.*, "Fast and practical secret key extraction by exploiting channel response," in *INFOCOM*, 2013.

[17] S. Mathur *et al.*, "Radio-telepathy: Extracting a secret key from an unauthenticated wireless channel," in *Mobile Computing and Networking*. ACM, 2008.

[18] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. on Information Theory*, vol. 39, no. 3, 1993.

[19] Z. Sahinoglu and S. Gezici, "Ranging in the IEEE 802.15. 4a standard," in *Wireless and Microwave Technology Conference*, 2006, pp. 1–5.

[20] Y.-S. Shiu *et al.*, "Physical layer security in wireless networks: A tutorial," *IEEE Wireless Communications*, vol. 18, no. 2, 2011.

[21] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub., 1999.

[22] M. Tope and J. McEachen, "Unconditionally secure communications over fading channels," in *Milcom*, 2001.

[23] W. Trappe, "The challenges facing physical layer security," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 16–20, 2015.

[24] J. W. Wallace and R. K. Sharma, "Automatic secret keys from reciprocal MIMO wireless channels: Measurement and analysis," *Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 381–392, 2010.

[25] R. Wilson, D. Tse, and R. A. Scholtz, "Channel identification: Secret sharing using reciprocity in ultrawideband channels," *Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 364–375, 2007.

[26] C. Zenger *et al.*, "Exploiting the Physical Environment for Securing the Internet of Things," in *New Security Paradigms Workshop*, 2015.

[27] J. Zhang *et al.*, "Secure key generation from OFDM subcarriers' channel responses," in *Globecom Workshops*, 2014.

[28] X. Zhou, L. Song, and Y. Zhang, *Physical Layer Security in Wireless Communications*. CRC Press, Nov. 2013.