Physical Layer Authentication Using Information Reconciliation

Atsu Kokuvi Angélo Passah, Rodrigo de Lamare, Arsenia Chorti

Physical Layer Authentication Using Information Reconciliation

Atsu Kokuvi Angélo Passah*†, Rodrigo C. de Lamare*‡, and Arsenia Chorti†§
*Department of Electrical Engineering (DEE), CETUC, PUC-Rio, Brazil
†ETIS Laboratory UMR 8051, ENSEA, CY Cergy Paris University, CNRS, France
‡School of Physics, Engineering and Technology, York University, United Kingdom
§Barkhausen Institut gGmbH, Germany

***Abstract*—User authentication in future wireless communication networks is expected to become more complicated due to their large scale and heterogeneity. Furthermore, the computational complexity of classical cryptographic approaches based on public key distribution can be a limiting factor for using in simple, low-end Internet of things (IoT) devices. This paper proposes physical layer authentication (PLA) expected to complement existing traditional approaches, e.g., in multi-factor authentication protocols. The precision and consistency of PLA is impacted because of random variations of wireless channel realizations between different time slots, which can impair authentication performance. In order to address this, a method based on error-correcting codes in the form of reconciliation is considered in this work. In particular, we adopt distributed source coding (Slepian-Wolf) reconciliation using polar codes to reconcile channel measurements spread in time. Hypothesis testing is then applied to the reconciled vectors to accept or reject the device as authenticated. Simulation results show that the proposed PLA using reconciliation outperforms prior schemes even in low signal-to-noise ratio scenarios.**

***Index Terms*—Physical layer authentication, physical layer security, information reconciliation.**

I. Introduction

The emergence of new generation wireless systems such as large scale, heterogeneous, Internet of things(IoT) networks, brings about a lot of security issues. In this context, the computational complexity of classical cryptographic schemes that use public key encryption node authentication can be a limiting factor in terms of performance, inducing considerable delays [1], [2]. Furthermore, standard wireless communication systems do not employ security protocols at the physical layer. Physical layer security [3] is currently being considered as the means to address such problems in next generation networks [4] by incorporating them in new security protocols. Authentication is a key element in security protocols, in particular, physical layer authentication (PLA) takes advantage of channel characteristics, similarly to hardware layer security, e.g., physical unclonable functions, that proposing authentication exploiting variations during hardware fabrication processes.

Recently, various channel-based PLA schemes have been studied. However, the wireless medium is subject to random variations over time. To account for this fact, the works in [5] and [6] studied an authentication scheme based on the channel impulse response (CIR) while also integrating additional multipath delay characteristics of the wireless channel into the authentication framework. Additionally, the studies in [5] and [6] used a two dimensional quantization method to mitigate random variations of amplitudes and delays. Moreover, [6] exploited the time-varying propagation delay to enhance authentication accuracy. In [7], the authors proposed two PLA schemes based on CIR without using a quantization algorithm in contrast with [5] and [6]. The idea was to avoid quantization errors introduced by the algorithm that could degrade the authentication performance. The first method used directly the estimated channel to make decisions about authentication, while the second approach exploited the channel correlation coefficient to enhance the performance.

To address the problem of channel state information (CSI) variations over time, we consider in this paper an approach based on error-correcting codes in the form of reconciliation. The proposed method employs a Slepian-Wolf coding scheme with polar codes that allows the reconciliation of discrepancies between different channel measurements in order to authenticate legitimate users. The authentication decision is therefore based on the comparison between reconciled vectors followed by hypothesis testing. In addition, we derive closed-form expressions of the probability distribution of the hypothesis test statistical variable and of the probability of false alarm and detection.

The remainder of this paper is organized as follows. Section II presents the authentication system model and explains the authentication phases. In Section III, the proposed approach is described. We provide performance analyses in Section IV by deriving closed-form expressions of the probability of false alarm and detection. Then, simulation results are presented in Section V and the paper is concluded in Section VI.

II. System Model

A standard wireless communication system of three nodes denoted by Alice, Bob and Eve is considered, where Alice and Bob are the legitimates nodes and Eve is an adversary. In this scheme, Bob wants to authenticate Alice while Eve is an active attacker that attempts to impersonate her. The objective is to design a scheme based on the CSI to distinguish Alice and Eve, both of which are equipped with a single antenna

($N_u = 1$, where $u \in \{a, e\}$. The indices a and e denote respectively Alice and Eve. Bob is equipped with N_b antennas.

We assume that the communication takes place in a rich scattering environment so that channel characteristics of different users are spatially uncorrelated when the distance between them is greater than half a wavelength [8]. Therefore, Alice's and Eve's estimated channels at Bob are uncorrelated. The communication is divided into two steps: the training phase and the authentication phase, as shown in Fig. 1.

1) Training phase: This phase occurs offline. Bob estimates the Alice's CSI. More precisely, in time slot t, Bob estimates the CSI $\mathbf{h}_a(t) \in \mathbb{C}^{1\times N_b}$ of Alice, where $h_{ai} \sim \mathcal{CN}\left(0, \sigma_h^2\right)$, $i = 1, \ldots, N_b$, which will be used as a reference in the authentication phase to decide whether the user is Alice or not.

2) Authentication phase: During the second online phase, in a subsequent time slot $t+m$, Bob takes new CSI measurements of $\mathbf{h}_u(t+m) \in \mathbb{C}^{1\times N_b}$, $u \in \{a, e\}$, that may either come from Alice or Eve and needs to make an authentication decision based on the CSI obtained in the training phase. Without loss of generality, we assume $m = 1$. When the transmitted signal at $t+1$ comes from Eve, $h_{ei}(t+1) \sim \mathcal{CN}\left(0, \sigma_h^2\right)$. We consider a scenario where the channel behaviour changes slowly over time. The channel between the same transmitter-receiver pair can then be well described by a first order Gauss-Markov process [7]. Thus, the channel between Alice and Bob in the time slot $t+1$ is given by

$$\mathbf{h}_a(t+1) = \beta \mathbf{h}_a(t) + \sqrt{1-\beta^2}\mathbf{n}_a \tag{1}$$

where β is the channel correlation coefficient and $\mathbf{n}_a$ is a measurement noise vector, $n_{ai} \sim \mathcal{CN}\left(0, \sigma_h^2\right)$. The noise vector $\mathbf{n}_a$ is statistically independent of $\mathbf{h}_a$.

We are going to present in the following section, our proposed approach based on Slepian-Wolf decoding to reconcile discrepancies over the CSI measurements in time, depicted through the system model in Fig. 1.

III. Proposed approach

Our goal is to use reconciliation in order to reduce the impact of inconsistencies over the CSI observed in different time slots. Reconciliation is a standard technique used in physical unclonable function based authentication (typically referred to as fuzzy extractors) and secret key generation from channel measurements. It is proposed in this work, for the fist time, to be used in channel-based PLA. In each phase, the channel estimation is quantized and the output vectors at time t and $t+1$ are treated as the codewords at the input of the reconciliation, as shown in Fig. 1. Note that during the offline first phase, not only the original CSI but also helper data (i.e., syndrome side information) need to be stored. Then, to make a decision during the online phase, a hypothesis test is performed by Bob in order to distinguish Alice from Eve.

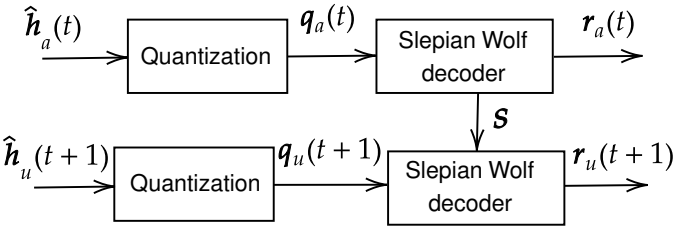

Figure 1: Proposed authentication scheme: the training phase at time t and the authentication phase at time $t+1$, $u \in \{a, e\}$

1) Training phase: The measurement by Bob at time t is given by

$$\hat{\mathbf{h}}_a(t) = \mathbf{h}_a(t) + \mathbf{z}(t) \tag{2}$$

where $\mathbf{z}(t) \in \mathbb{C}^{1\times N_b}$ is a zero mean complex Gaussian noise so that $z_i(t) \sim \mathcal{CN}\left(0, \sigma_z^2\right)$, $i = 1, \ldots, N_b$. Bob collects M samples of these measurements in a row vector $\left[\hat{\mathbf{h}}_{a1} \| \hat{\mathbf{h}}_{a2} \| \ldots \| \hat{\mathbf{h}}_{aM}\right] \in \mathbb{C}^{1\times MN_b}$. By extracting the real and imaginary parts of this row vector and by concatenating them, we get the vector $\mathbf{x}_a(t) \in \mathbb{R}^{1\times N}$ where $N = 2MN_b$. For a purpose of simplicity, $\mathbf{x}_a(t)$ is then quantized using the 1-bit threshold quantizer (3) and output $\mathbf{q}_a(t) \in \{0,1\}^{1\times N}$.

$$Q(x_j) = \begin{cases} 1, & \text{if } x_j \geq \gamma \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where the threshold γ is the mean of $\mathbf{x}$, i.e., $\gamma = \mathrm{mean}(\mathbf{x})$

2) Authentication phase: Similarly to the previous step, the channel measurements at Bob's side is given by

$$\hat{\mathbf{h}}_u(t+1) = \mathbf{h}_u(t+1) + \mathbf{z}(t+1) \tag{4}$$

where $u \in \{a, e\}$, $\mathbf{z}(t+1) \in \mathbb{C}^{1\times N_b}$, $z_i(t+1) \sim \mathcal{CN}\left(0, \sigma_z^2\right)$, $i = 1, \ldots, N_b$, and the quantized vector is $\mathbf{q}_u(t+1) \in \{0,1\}^{1\times N}$. Based on the principle of Slepian Wolf decoding with polar codes [9], $\mathbf{q}_a(t)$ and $\mathbf{q}_u(t+1)$ are decoded using the cyclic redundancy check (CRC) successive cancellation list decoding. It enhances the performance of the polar code in finite blocklengths [10], [11]. CRC assists the decoder in choosing the right decoding path from a list of possibilities. The reconciliation outputs reconciled vectors $\mathbf{r}_a(t)$ and $\mathbf{r}_u(t+1) \in \{0,1\}^{1\times K}$. Note that the decoding of $\mathbf{q}_u(t+1)$ uses the syndrome $\mathbf{s}$ from the decoding of $\mathbf{q}_a(t)$. For a well-designed reconciliation scheme with the right choice of code rate, the reconciled vectors should be the same in the normal case and should be different in the spoofing case. Thus, in an ideal situation, Bob expects $\mathbf{r}_a(t) = \mathbf{r}_a(t+1)$ in the normal case and $\mathbf{r}_a(t) \neq \mathbf{r}_e(t+1)$ in the spoofing case. Bob uses a hypothesis test to distinguish the normal case and the spoofing case (5). H_0 denotes the normal case, i.e., the user is Alice at $t+1$, H_1 denotes the spoofing case, i.e., the user is not

Alice, η_{th} is the decision threshold and η is the hypothesis test statistical variable.

$$\begin{cases} H_0 : \ \eta = d\left(\mathbf{r}_a(t), \mathbf{r}_a(t+1)\right) \leq \eta_{th} \\ H_1 : \ \eta = d\left(\mathbf{r}_a(t), \mathbf{r}_e(t+1)\right) > \eta_{th} \end{cases} \tag{5}$$

A suitable choice to distinguish Alice from Eve is to calculate the bit error, that is to determine the number of bit positions where $\mathbf{r}_a(t)$ and $\mathbf{r}_u(t+1)$ are different. That is well represented by the Hamming distance between $\mathbf{r}_a(t)$ and $\mathbf{r}_u(t+1)$. η is therefore as follows,

$$\eta = d\left(\mathbf{r}_a(t), \mathbf{r}_u(t+1)\right) = \sum_{n=1}^{K} \left| r_{a,n}(t) - r_{u,n}(t+1) \right| \tag{6}$$

where $d(.,.)$ is the Hamming distance.

To better position this work and highlight the merits of the proposed approach, let us present briefly other possible approaches in [7] and [12]. In [7], channel measurements $\hat{\mathbf{h}}_a(t)$ (2) and $\hat{\mathbf{h}}_u(t+1)$ (4) was considered for the hypothesis test as follows:

$$\begin{cases} H_0 : \ \eta_p \leq \eta_{p_{th}} \\ H_1 : \ \eta_p > \eta_{p_{th}} \end{cases} \tag{7}$$

where the statistical variable η_p was given by the square of the ℓ_2-norm between $\hat{\mathbf{h}}_a(t)$ and $\hat{\mathbf{h}}_u(t+1)$ as given by

$$\eta_p = \sum_{i=1}^{N_b} \left(\hat{h}_{a,i}(t) - \hat{h}_{u,i}(t+1) \right)^2. \tag{8}$$

A key-based PLA method was presented in [12]. Note that key-based methods assumes that legitimate nodes Alice and Bob agreed on the same secret keys before the transmission. The steps are described as follows:

- The user requests the authentication by sending data packets to Bob.
- Authentication inquiry:
  Bob generates then $\mathbf{s}_b = [s_{b,1}, s_{b,2}, \ldots, s_{b,N_b}]^{\mathrm{T}}$, $s_{b,i} = \exp(j\theta_{b,i})$ and $\theta_{b,i}$ is uniformly distributed over $[0, 2\pi)$. The signal received by the user at time t is $x_{u,i}(t) = h_{u,i}(t)s_{b,i} + z_{b,i}(t)$, $i = 1, \ldots, N_b$ where $h_{u,i}(t) \sim \mathcal{CN}\left(0, \sigma_h^2\right)$ and $z_{b,i}(t) \sim \mathcal{CN}\left(0, \sigma_z^2\right)$.
- Authentication response:
  The user gets the phase $\theta_{u,i}$ of $x_{u,i}(t)$ and sends the response signal $\mathbf{s}_u = [s_{u,1}, s_{u,2}, \ldots, s_{u,N}]^{\mathrm{T}}$ to Bob, where $s_{u,i} = \exp\{j\left(\mathcal{M}\left(k_{u,i}\right) - \theta_{a,i}\right)\}$. $\mathcal{M}(\cdot)$ is the secret key mapping function. It is given by: $\mathcal{M}\left(k_i\right) = 0$ if $k_i = [00]$, $\mathcal{M}\left(k_i\right) = \pi/2$ if $k_i = [01]$, $\mathcal{M}\left(k_i\right) = \pi$ if $k_i = [11]$ and $\mathcal{M}\left(k_i\right) = 3\pi/2$ if $k_i = [10]$, where k_i is the key bits. The shared key bits between the legitimate nodes are $k_{a,i} = k_{b,i}$.
- Authentication completion:
  Bob received a signal $x_{b,i}(t+1) = h_{u,i}(t+1)s_{u,i} + z_{b,i}(t+1)$ and gets his phase as previously. He then calculates $y_i = x_{b,i}(t+1)s_{b,i}$ to remove $\theta_{b,i}$.

The hypothesis test is formulated as

$$\begin{cases} H_0 : \mathbf{K}_{t+1} = \mathbf{K}_B \\ H_1 : \mathbf{K}_{t+1} \neq \mathbf{K}_B \end{cases} \tag{9}$$

The hypothesis test statistic variable $\eta_p \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{p_{th}}$ is given by $\eta_p = \mathrm{Re}\left\{\sum\left(\exp^{-j\mathcal{M}(\mathbf{K}_B)} \otimes \mathbf{y}\right)\right\}$ where $\mathrm{Re}(\cdot)$ calculates the real part.

We focus in the next section on the analysis where we derive the closed-form expressions of the performance metrics.

IV. Performance Analysis

Based on hypothesis testing, the performance metrics are the probability of false alarm (type I error) and the probability of detection. They are given respectively by $\mathrm{P_{FA}} = \Pr\left(\eta > \eta_{th} \mid H_0\right)$ and $\mathrm{P_D} = \Pr\left(\eta > \eta_{th} \mid H_1\right)$. Thus, we need to determine first the probability distribution of η under H_0 and under H_1.

The behaviour of a Hamming distance (6) allows to determine the closed-form expressions of the probability distribution of η under H_0 and H_1 in the following propositions.

Proposition 1: Under H_0, η follows a binomial distribution of parameters K and p_0, i.e. $\eta \sim \mathbb{B}(K, p_0)$.

$$P(\eta = n|H_0) = \binom{K}{n} p_0^n (1-p_0)^{K-n} \tag{10}$$

where p_0 is the bit error probability during the decoding.

Proof: The Hamming distance η can be modeled as the sum of Bernoulli random variables, where each bit has a certain probability of being in error. Let's consider X_i, a Bernoulli random variable representing the error in the i^{th} bit position.

$$\begin{cases} X_i = 1 \ \text{,if error} \\ X_i = 0 \ \text{,otherwise} \end{cases} \tag{11}$$

$\implies \eta = \sum_{i=1}^{K} X_i$. X_i, $i = 1, \ldots, K$ are independent and identically distributed $\implies$ η follows a binomial distribution of parameters K and p_0, $\eta \sim \mathbb{B}(K, p_0)$, where p_0 is the bit error probability under the normal case H_0. The pdf of η is then given by $P(\eta = n|H_0) = \binom{K}{n} p_0^n (1-p_0)^{K-n}$.

Proposition 2: Under H_1, η follows a binomial distribution of parameters K and p_1, i.e. $\eta \sim \mathbb{B}(K, p_1)$.

$$P(\eta = n|H_1) = \binom{K}{n} p_1^n (1-p_1)^{K-n} \tag{12}$$

where p_1 is the bit error probability during the decoding.

Proof: It is very similar to that of Proposition 1 by replacing p_0 by p_1 $\implies$ $P(\eta = n|H_1) = \binom{K}{n} p_1^n (1-p_1)^{K-n}$.

In propositions 1 and 2, bit error probabilities p_0 and p_1 can be estimated with the bit error rate by simulation. Fig. 2 shows the comparison between simulation results and closed-from expressions of the probability distribution of η under H_0 and H_1. The closed-from expression perfectly matches the simulation result under both hypotheses for a $SNR = 10dB$.

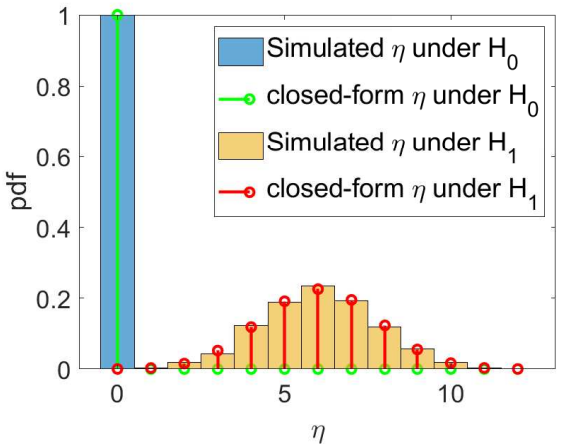

Figure 2: Simulated vs closed-from expression: code rate $= 0.01$, $SNR = 10dB$, $p_0 \approx 0$ and $p_1 \approx 0.5025$

Based on the proposition 1, the closed-form expression of the false alarm probability is given by

$$P_{FA} = \sum_{n=\eta_{th}+1}^{K} \binom{K}{n} p_0^n (1-p_0)^{K-n}. \tag{13}$$

Based on the proposition 2, the closed-form expression of the probability of detection is given by

$$P_D = \sum_{n=\eta_{th}+1}^{K} \binom{K}{n} p_1^n (1-p_1)^{K-n}. \tag{14}$$

V. Numerical Results

The simulation parameters are defined as follows. The number of antennas at the receiver, i.e., Bob, is $N_b = 32$, the channel correlation coefficient $\beta = 0.9$, the channel variance $\sigma_h^2 = 1$, $M = 16$, the codelength $N = 2MN_b = 1024$. The signal-to-noise ratio is defined as $SNR = \frac{\sigma_h^2}{\sigma_z^2}$.

First, we investigate the impact of the code rate on the reconciliation scheme. Fig. 3 shows the detection probability as a function of the code rate for a $SNR = 15dB$ and a false alarm rate of 10^{-3}. We observe that the detection probability is almost equal to 1 for code rates less than 0.2. For code rates greater than 0.2, it becomes very low. In this case, as the code rate increases, it decreases towards 0. The result can be improved in higher SNR scenarios or by increasing the code length.

We compare our results with the prior results in [7] and [12]. In Fig. 4, we present the receiver operating characteristic (ROC) curve that represents the probability of detection as a function of the probability of false alarm for a $SNR = 5dB$ and a code rate of 0.01. As the P_{FA} increases, the P_D increases for all the schemes. We observe that not only the proposed reconciliation method performs better than the prior ones but also P_D is always very close to 1 even for very low probabilities of detection. We actually have approximately 99.97% increase in the detection probability for

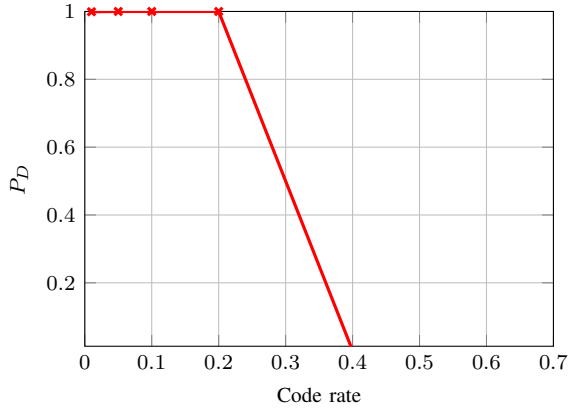

Figure 3: P_D vs code rate: $P_{FA} = 0.001$, $SNR = 15dB$

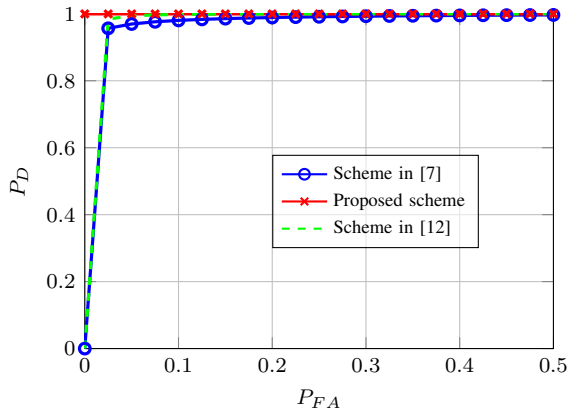

Figure 4: ROC curve: code rate $= 0.01$, $SNR = 5dB$

a $P_{FA} \approx 0\%$. That can be explained by the fact that using the error-correcting code allows to reconcile subsequent channel measurements. We also notice that the prior scheme in [12] performs better than the one in [7] with a small gap.

Fig. 5 presents the detection probability as a function of the SNR for a code rate of 0.01 and a false alarm probability $P_{FA} = 10^{-3}$. First, as the SNR increases, the P_D increases for all the schemes. This is due to the fact that, as the SNR increases, the estimation errors decline. Second, our reconciliation scheme performs better than prior results even for low SNRs but the work in [7]'s performance is very close for SNRs greater than $10dB$. We notice that the method in [12]'s performance is better than the one in [7] for SNRs less that $5dB$ but is worse than [7] for SNRs greater than $5dB$.

VI. Conclusion

This paper investigates the problem of PLA using error-correcting polar code to reconcile discrepancies between channel measurements and to distinguish between the normal case and the spoofing case. We provide closed-form expressions for the false alarm probability and the detection probability. In a scenario of $SNR = 15dB$ and a false alarm rate of 10^{-3}, results show that that the probability of detection is close to

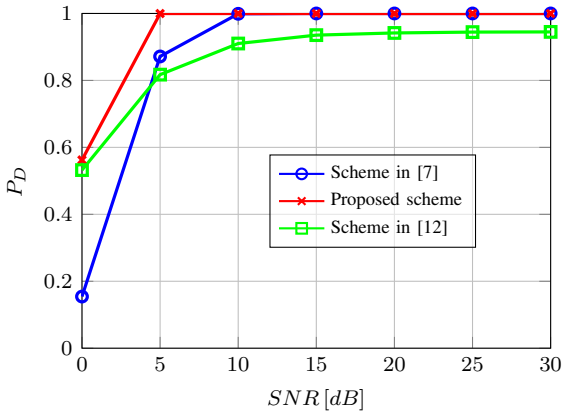

Figure 5: P_D vs SNR: code rate $= 0.01$ with a $P_{FA} = 0.001$

one for code rates less than or equal to 0.2 and very small for code rates greater that 0.2. Simulation results confirm also that our reconciliation-based method has better performance than prior schemes. The fact that we use a small value of 10^{-3} for the false alarm probability allows to confirm the performance of our work in practical systems that need very low false alarm probabilities.

ACKNOWLEDGMENT

Atsu Kokuvi Angélo Passah has been supported by the ELIOT ANR-18-CE40-0030 and FAPESP 2018/12579-7, FAPERJ project. Arsenia Chorti has been partially supported by the EC through the Horizon Europe/JU SNS project ROBUST-6G (Grant Agreement no. 101139068), the ANR-PEPR 5G Future Networks project, the ELIOT ANR-18-CE40-0030 and FAPESP 2018/12579-7, FAPERJ project and the CYU INEX-PHEBE projects.

REFERENCES

[1] Shakiba-Herfeh, M., Chorti, A., Vincent Poor, H. (2021). *Physical Layer Security: Authentication, Integrity, and Confidentiality.* In: Le, K.N. (eds) Physical Layer Security. Springer, Cham.

[2] M. Mitev, A. Chorti, H. V. Poor and G. P. Fettweis, "What Physical Layer Security Can Do for 6G Security," in *IEEE Open Journal of Vehicular Technology,* vol. 4, pp. 375-388, 2023.

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

[4] B. Aazhang et al., "Key drivers and research challenges for 6G ubiquitous wireless intelligence",(white paper), Sep. 2019

[5] F. J. Liu, X. Wang and S. L. Primak, "A two dimensional quantization algorithm for CIR-based physical layer authentication," 2013 *IEEE International Conference on Communications (ICC)*, Budapest, Hungary, 2013, pp. 4724-4728.

[6] J. Liu and X. Wang, "Physical Layer Authentication Enhancement Using Two-Dimensional Channel Quantization," in *IEEE Transactions on Wireless Communications,* vol. 15, no. 6, pp. 4171-4182, June 2016.

[7] N. Xie, J. Chen and L. Huang, "Physical-Layer Authentication Using Multiple Channel-Based Features," in *IEEE Transactions on Information Forensics and Security,* vol. 16, pp. 2356-2366, 2021.

[8] Jakes William C. *Microwave Mobile Communications.* IEEE, 1994.

[9] E. Arikan, "Source polarization," *2010 IEEE International Symposium on Information Theory,* Austin, TX, USA, 2010, pp. 899-903.

[10] I. Tal and A. Vardy, "List decoding of polar codes," *2011 IEEE International Symposium on Information Theory Proceedings,* St. Petersburg, Russia, 2011, pp. 1-5.

[11] M. Shakiba-Herfeh and A. Chorti, "Comparison of Short Blocklength Slepian-Wolf Coding for Key Reconciliation," *2021 IEEE Statistical Signal Processing Workshop (SSP),* Rio de Janeiro, Brazil, 2021, pp. 111-115.

[12] X. Lu, J. Lei, Y. Shi and W. Li, "Physical-Layer Authentication Based on Channel Phase Responses for Multi-Carriers Transmission," in *IEEE Transactions on Information Forensics and Security,* vol. 18, pp. 1734-1748, 2023.