SECRET KEY GENERATION SCHEME FROM WIFI AND LTE REFERENCE SIGNALS

Christiane L. Kameni Ngassa (Thales Communications and Security (TCS), Gennevilliers, France; Christiane.Kameni@thalesgroup.com); Renaud Molière (TCS; Renaud.Moliere@thalesgroup.com); François Delaveau (TCS; Francois.Delaveau@thalesgroup.com); Taghrid Malzoum (Telecom ParisTech (TPT), Paris, France; Taghrid.Mazloum@telecom-paristech.fr); Alain Sibille (TPT; Alain.Sibille@telecom-paristech.fr)

ABSTRACT

Physical layer security has emerged as a promising approach to strengthen security of wireless communications. Particularly, extracting secret keys from channel randomness has attracted an increasing interest from both academic and industrial research groups. In this paper, we present a complete implantation of a Secret Key Generation (SKG) protocol which is compliant with existing widespread Radio Access Technologies. This protocol performs first the Quantization of the Channel State Information (CSI), then Information Reconciliation and Privacy Amplification. We also propose an innovative algorithm to reduce the correlation between quantized channel coefficients. Finally we assess the performance of our protocol by evaluating the quality of secret keys generated from real field WiFi and LTE probe signals in various propagation environments.

1. INTRODUCTION

Recent news highlighting security failures of public wireless communication systems have recalled the limits of the cryptographic key distribution approach and the urge to improve security of the information exchanged over the air interface [1, 2, 3, 4]. The emergence of Physical layer Security (Physec) has provided an alternative approach for designing robust secret keys by leveraging the intrinsic randomness of wireless channels. This technique is referred to as Secret Key Generation (SKG) [5].

In § 2, we detail the typical scenario where two legitimate users (Alice and Bob) can communicate securely in presence of an eavesdropper (Eve), and how this principle works. The vast majority of existing works on SKG use the Received Signal Strength Indication (RSSI) since it is easily accessible. However, RSSI does not capture the entire richness of the channel as it ignores the phase of channel coefficients, which usually provide more randomness than the power of the signal. In this paper we present a full SKG scheme based on full Channel State Information (CSI) or its

Fourier transform (Channel Frequency Response – CFR, §3). Our SKG protocol is composed of the following steps: Channel estimation (§3), Channel Coefficient de-correlation (§4), Quantization of the CSI (§5), Information Reconciliation (§6) and Privacy Amplification (§7).

In order to evaluate the performance of our scheme (quality of the generated keys, complexity of the processing), we apply our secret key generation protocol to real field WiFi and LTE networks (§8). Signals being captured in several indoor and outdoor locations, keys are computed from Channel Frequency Responses extracted from these real field records. Then, randomness analyses and NIST tests are used to access their quality.

2. SECRET KEY GRENERATION PRINCIPLE

2.1. Communication scenario

The legitimate users Alice and Bob attempt to communicate securely in presence of an eavesdropper Eve. For this, Alice and Bob extract a common secret key from their channel estimates. When Eve is located at a distance of a few wavelengths from Bob, her channel measurements will be de-correlated of the legitimate channel and therefore any measure of Eve will be de-correlated to the secret key.
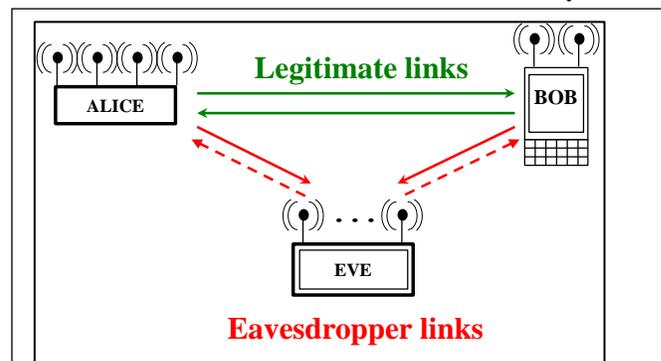

Figure 1: Communication scenario

2.2. On channel randomness

The main reasons why a secret key can be extracted from the random radio propagation are the following.

In indoor and outdoor environments, waveforms transmitted from Alice to Bob and Eve follow multiple paths and come across various obstacles with distinct angles of incidence. As a result, they are altered very differently when they are received by Bob and Eve. A few wavelengths is enough to ensure a complete de-correlation between Bob and Eve's channels, especially when the scatterers' Angular Spread (AS) is large [6].

Besides, due to complex wave propagation and unpredictable scatterers in the communication channel, Eve cannot predict or recover the legitimate channel.

Finally, in TDD mode and for each carrier, waveforms from Alice to Bob hit obstacles in the forward and return direction with the same angle of incidence. Therefore the legitimate users see the same randomness and thus have similar channel measurements. This phenomenon is referred to as "channel reciprocity".

Consequently, the channel coefficients measured by Alice and Bob characterize the legitimate link and cannot be reconstructed by Eve. Thus, Alice and Bob can use this shared pool of randomness to generate secret keys.

2.3. Secret key generation steps

The proposed SKG protocol is composed of the following steps:

Channel Estimation: the first step of the SKG scheme estimates the radio channel and computes CSI or CFR

Channel Coefficient de-correlation: in this second step, we apply a new algorithm to select channel coefficients with low cross correlation. This optimizes the randomness selection in stationary environments.

Quantization: this step uses the Channel Quantization Alternate (CQA) algorithm introduced by Wallace to quantize selected channel coefficients [5], that minimizes key mismatch between the legitimate users Alice and Bob.

Information Reconciliation: this step corrects the remaining mismatch between Alice and Bob keys. We employ secure sketch and error correcting codes to correct Bob's errors on Alice's key. To do so, Alice has to send the secure sketch over the public channel, possibly leaking a controlled amount of information to the eavesdropper Eve.

Privacy Amplification: this step improves the randomness of the secret key by decreasing its length and removes the information leaked to Eve. To do so, we use hash functions. This final step guarantees that the generated secret key is fully de-correlated from the key computed by the eavesdropper.

Note: searching for practical implementation inside communication devices, we focused in each step on most robust and simple algorithms. For example, we choose a simple algebraic forward error correcting (FEC) code to reconcile Alice and Bob keys and a classical family of 2-universal hash function in the privacy amplification step [7].

3. CHANNEL ESTIMATION

When considering an Orthogonal Frequency Division Multiplexing signal (OFDM, such as encountered in WiFi and LTE networks) in the frequency domain, the component of the Channel Frequency Response (CFR) H_f quantifies the fading applying on each subcarrier. In a sampled system, considering a finite response and band, the kth frequency component f_k of the CFR can be calculated as follows.

$$\hat{H}_f(k) = Y(f_k)/X(f_k)$$

where Y is the received signal, and X is the emitted signal (or reference signal).

In the time domain, an equivalent Channel Input Response (CIR) estimation can be deduced from the CFR by IFFT, as follows:

$$\hat{H}_{IFFT} = IFFT(\hat{H}_f)$$

When considering now TDMA of CDMA wave forms encountered in 2G and 3G radio Access technologies (RAT), CIR can be computed directly in the time domain by applying filter estimations techniques to reference signal X.

4. CHANNEL DECORRELATION

Secret key bits should be completely random to keep them unpredictable by Eve, therefore any deterministic component in the radio propagation channel should be removed. Same apply to any time or frequency correlation between quantized bits: the quantization algorithm should not only generate bits with equal probability but also the channel coefficients that are quantized to generate these bits should be as random and de-correlated as possible.

The goal of this step is to decrease the negative effect of channel correlation by a careful selection of the channel coefficient to be quantized.

First, time correlation is decreased between channel coefficients. To do so:

- Channel coefficients computed at a given time acquisition, constitute a frame.
- Cross-correlation coefficients are computed between two consecutive frames
- Only frames with low cross-correlation coefficient (above a given threshold T_t) are selected.

Then, same procedure applies to frequency correlation:

- Cross-correlation coefficients are computed between two consecutive frequency carriers
- Only frequency carriers for which the cross-correlation coefficient is above a given threshold T_f are selected. In addition, lowest and highest frequency carriers are dropped.

Finally, Alice sends to Bob the position of the channel coefficients over the public channel. Hence, Eve also knows which coefficients were dropped and which ones were selected but she does not have any additional information on their value. Therefore there is no information leakage during the channel de-correlation step.

5. QUANTIZATION

After measuring the radio channel, Alice and Bob jointly employ an algorithm to quantize the channel taps that they have estimated in order to generate a common sequence of key bits from their instantiation of the shared channel, under reciprocity assumption.

However, due to noise and channel estimation errors, Alice and Bob may disagree on some key bits. Several quantization algorithms employing censoring schemes have been developed to limit this mismatch between Alice and Bob keys.

A typical censoring algorithm defines guard band intervals and discards any channel measurement falling into it [5]; leading to an inefficient exploitation of channel measurements and to a lower number of generated key bits.

Thus, other schemes employ different quantization maps where each one is adapted to the channel observations, e.g. channel quantization alternating (CQA) algorithm [5]. The principle consists in choosing the adaptive quantization map where the current observation is less sensitive to mismatch. Consequently, we apply the CQA algorithm to complex channel coefficients to generate secret key bits.

6. INFORMATION RECONCILIATION

This step suppress remaining mismatches between Alice and Bob keys by using secure sketch based on error-correcting codes [8]. The key computed by Alice is considered as the secret key and Bob wants to retrieve Alice's key using the key K_b he extracts from his channel measurements. The processing can be described as follows:

Alice:
- selects a random codeword c from an error-correcting code C
- computes the secure sketch $s = K_a \oplus c$
- sends s to Bob over the public channel

Bob:
- subtracts s from its computed key K_b :
$$c_b = K_b \oplus s (= K_b \oplus K_a \oplus c)$$
- decodes c_b to recover c and gets $\hat{c}$
  - computes K_a by shifting back and gets:
$$\widehat{K_a} = \hat{c} \oplus s$$

Perfect reconciliation is achieved when Bob perfectly retrieves the random codeword chosen by Alice, meaning that $\hat{c} = c$. As a result, no mismatch occurs between Alice and Bob keys ($K_a = K_b$).

Therefore the secure sketch s, sent over the public channel, allows the exact recovery of the secret key without revealing the exact value of the key.

However, s might leak some information on the secret key over the public channel as Eve can also use the secure sketch to retrieve the secret key K_a.

Thus, a final step is then necessary to suppress the leaked information and to improve the quality of the secret key.

7. PRIVACY AMPLIFICATION

The objective of the privacy amplification step is to erase the information leaked to Eve on the secret key during the information reconciliation step and to improve the randomness of the key.

For our SKG scheme we interpret the secret key K as an element of the Galois Field $GF(2^n)$ and we choose the following two-universal family of hash functions [9] where n is the number of bits of the key K.

For $1 \leq r \leq n$ and for $a \in GF(2^n)$, the functions $\{0,1\}^n \to \{0,1\}^r$ assigning to the key K the first r bits of key $a. K \in GF(2^n)$ define a two-universal family of hash functions. r is the final length of the secret key.

In practice, at each new key computation, the parameter a is randomly chosen by Alice who sends it to Bob over the public channel. Alice and Bob then compute the product $a. K \in GF(2^n)$.

The hash mechanism spreads any bit error all over the final key $(a. K)_{r\ bits}$ (first r bits of $a. K$), thus:
- When Eve tries to recover the initial key K (at the reconciliation step), any error on K will make the final key $(a. K)_{r\ bits}$ unusable for her.
- Bob has to perfectly recover the initial key K (i.e. reconciliation should be perfectly achieved) in order to get the usage of the final key $(a. K)_{r\ bits}$.
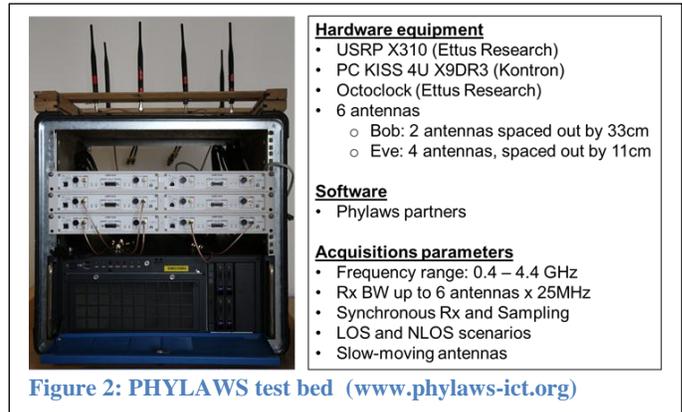

Hardware equipment
- USRP X310 (Ettus Research)
- PC KISS 4U X9DR3 (Kontron)
- Octoclock (Ettus Research)
- 6 antennas
  - Bob: 2 antennas spaced out by 33cm
  - Eve: 4 antennas, spaced out by 11cm

Software
- Phylaws partners

Acquisitions parameters
- Frequency range: 0.4 – 4.4 GHz
- Rx BW up to 6 antennas x 25MHz
- Synchronous Rx and Sampling
- LOS and NLOS scenarios
- Slow-moving antennas
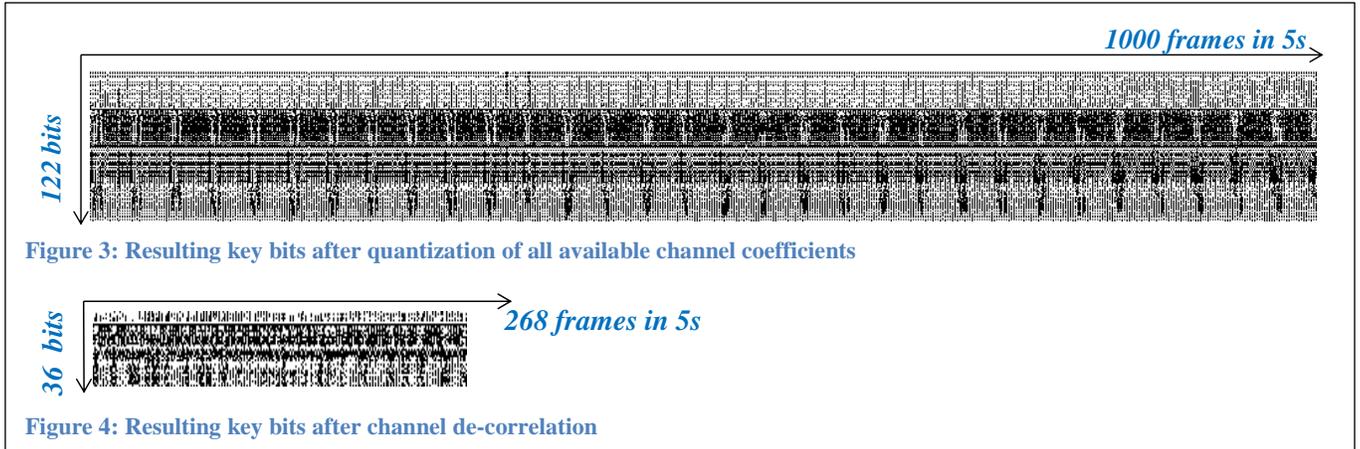
Figure 2: PHYLAWS test bed (www.phylaws-ict.org)

Figure 3: Resulting key bits after quantization of all available channel coefficients

Figure 4: Resulting key bits after channel de-correlation
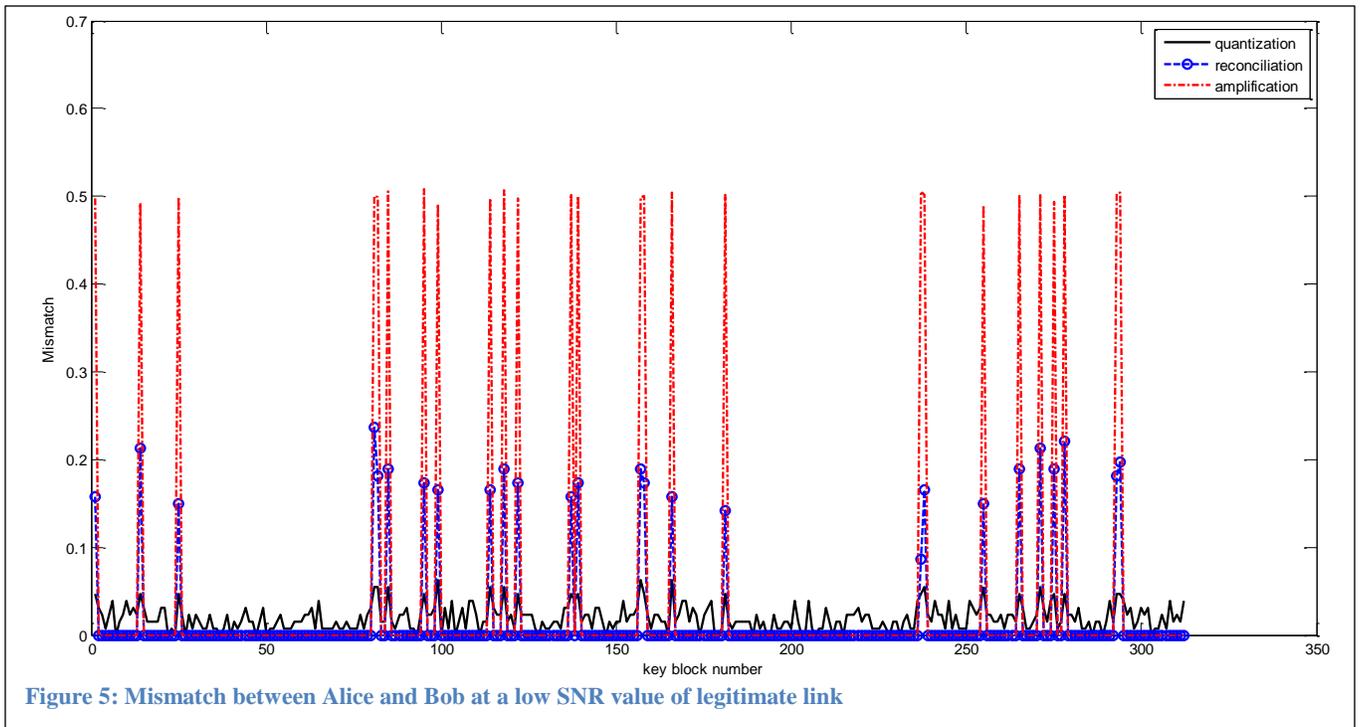

Figure 5: Mismatch between Alice and Bob at a low SNR value of legitimate link

8. EXPERIMENTAL RESULTS

In this section we generate keys from real LTE and WiFi signals acquired using the PHYLAWS test bed [10] and we analyze their quality and the processing complexity.

8.1. Impact of channel de-correlation

Figure 3 is relevant to a very stationary the propagation environment (empty indoor tennis court, static Alice and Bob, static scatterers) and shows the direct output of the CQA algorithm (§5) with 4 Quantization Regions (QR). CFR computed from LTE signals over 5 seconds (frequency: 2627.5MHz, bandwidth: 1.4 MHz) produced 1000 frames detections and 122 secret bits per frames. However, we can notice a repetitive pattern on the generated keys meaning that CFR coefficients are highly correlated in time and in frequency. This high correlation represents a major vulnerability as the generated secret key bits will not be random enough.

Figure 4 shows key bits obtained on the same record with the same processing after applying our channel coefficient selection (§4) on the original CFRs: the correlation between bits has significantly decreased both in time and frequency (our algorithm managed to extract the repeating pattern of the key bit). However the price to pay is fewer secret key bits.

Figure 6: Mismatch between Alice and Bob
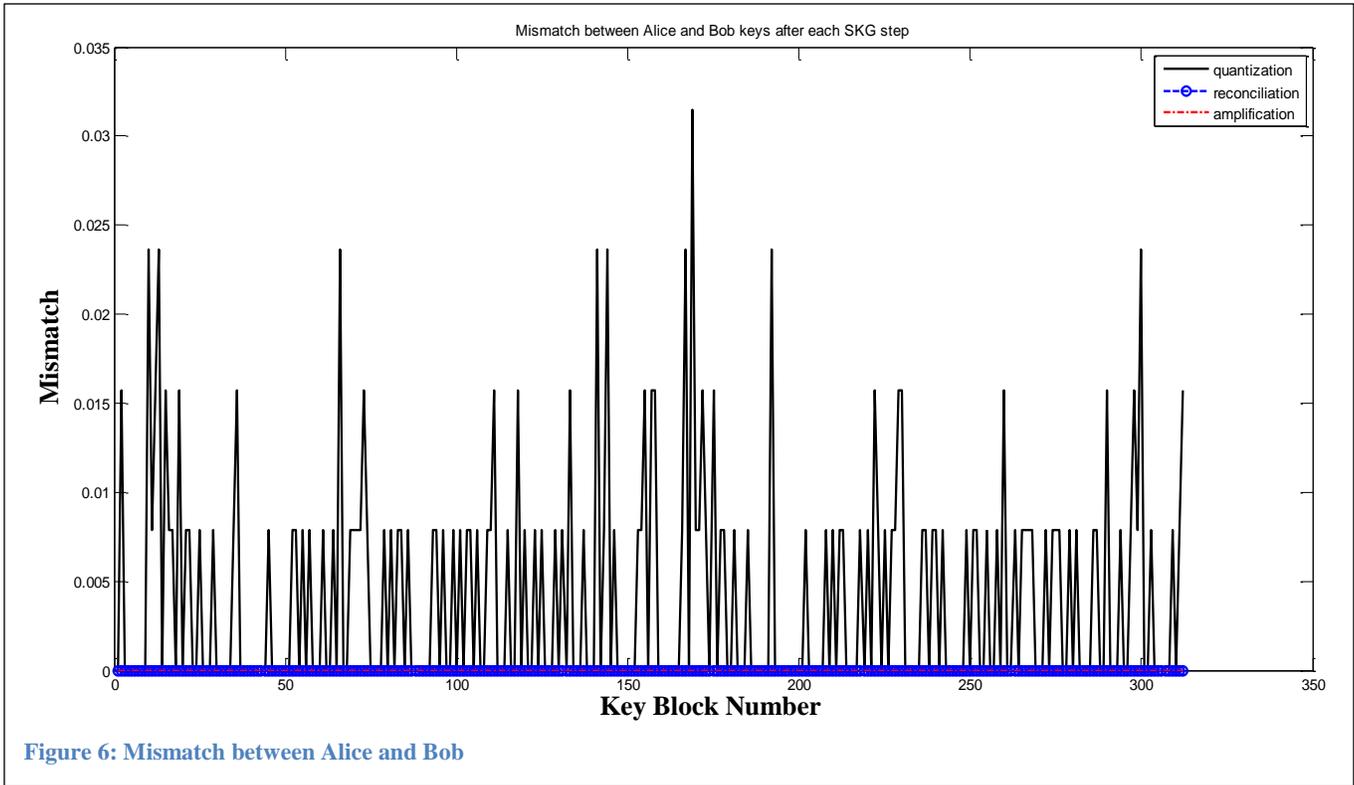
8.2. Analysis of the BER between Alice's and Bob's keys

In this section we analyze the key bit error rate (or "key mismatch") between Alice and Bob's generated keys. 320 keys of 127 bits were generated under a Wifi carrier (IEEE 802.11a, frequency 2462 MHz, Bandwidth: 20 MHz) in an open space environment (office) and three different SNR (20 dB, 25 dB and 28 dB) were considered.

For each SNR we plot the mismatch of Alice and Bob keys after each step of our SKG scheme.
- Black curves represent the mismatch after quantization
- Blue curves represent the mismatch after information reconciliation
- Red curves represent the mismatch after privacy amplification

Our quantization step (§5) uses the CQA Algorithm with 4 regions. Our reconciliation step (§6) uses secure sketch based on a (127, 92, 11) BCH code. Our privacy amplification step uses the 2-universal family of hash functions of §7.

Figure 5 plots the key mismatch between Alice and Bob for a low value of the SNR: here the number of errors is much higher than the error-correction capability of the BCH code, and key mismatched remain. A more powerful FEC code would optimize the information reconciliation.

We note that when the information reconciliation step fails, it increases the key mismatch compare to its value after the quantization step. Moreover the privacy amplification induces two extreme behaviors.
- When there is no error between the Alice and Bob's key, the mismatch remains null
- However, for any non-zero value, the mismatch is driven to 0.5. Thus privacy amplification increases the confusion on key mismatches when Bob's does not success to extract the same key than Alice. Same applies to Eve.

Figure 6 shows the same results when considering SNR=28 dB. Here all errors were corrected by the information reconciliation step thus Bob and Alice generate the same keys after privacy amplification.

8.3. Analysis of the BER between Eve and Bob

In this section, under the same WiFi carrier as above, we evaluate the number of errors that Eve makes on Bob's key: we analyze the Bit Error Rate between Eve and Bob when Eves applies the same process than Bob to Alice's signals with some antenna advantage (Eve has four antennas for her CFR estimations while Bob has only two antennas).

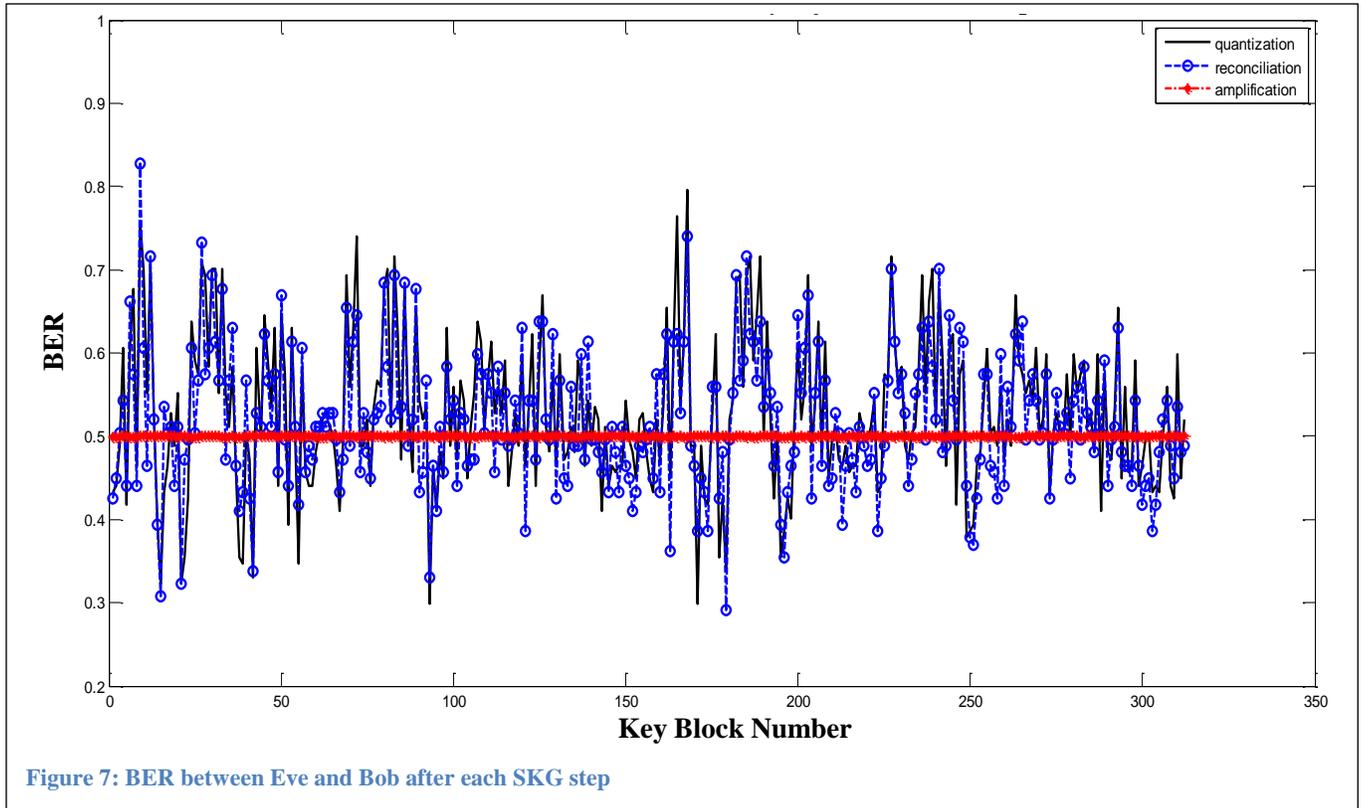Figure 7: BER between Eve and Bob after each SKG step

Figure 7 plots the BER between Eve and Bob for each of the 320 generated keys.

We note that BER does not change much after information reconciliation. However, after privacy amplification, the BER is driven to 0.5 after. In other words, reconciliation has low impact on Eve but privacy amplification highly increases the confusion of Eve on Bob's key and ensures that Eve's key is de-correlated from Bob's Key.

Figure 7 shows that Eve's BER after privacy amplification is 0.5. Thus, Eve has no information on the value for each bit of Bob's key. Hence, further investigation showed that no vulnerability occurred to particular bits.

Nevertheless, theoretically, information was leaked during the information reconciliation step (exchange of the secure sketch s). Therefore a corresponding number of bits should be removed from the key.

Denote N the length of the FEC code used for information reconciliation and R the rate of the code. The secure sketch s sent over the public channel leaks information on $N(1-R)$ bits of Bob's key. Therefore the secret key length should be decreased to $N*R$.

8.4. Analysis of the randomness of the keys

In this section we study secret keys computed from LTE signals (Frequency: 2627.5 MHz, Bandwidth: 1.4 MHz) and WiFi Carrier (2462 MHz, Bandwidth: 20 MHz).

- LTE Carrier - Indoor environment (classroom) with static antennas and limited mobility of scatterers).
  - Figure 8 shows quantization outputs: the key are not totally random..
  - Figure 9 shows privacy amplification outputs: it confirms that this step provides an extra level of security, improves the randomness of the keys and ensures the independency of Eve's computed key from the secret key shared by Alice and Bob.
- LTE Carrier - Urban outdoor environment with static antennas and mobile people and cars:
  - Figure 10 shows quantization outputs: the key are numerous but not totally random.
  - Figure 11 shows privacy amplification outputs: it confirms that the key are numerous and that key randomness is significantly improved.

In addition, figures 10 and 11, when compared to figures 8 and 9, show that more keys are generated in outdoor environments compared to indoor environments.
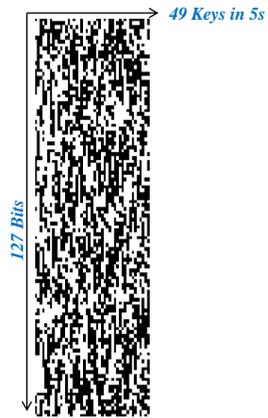
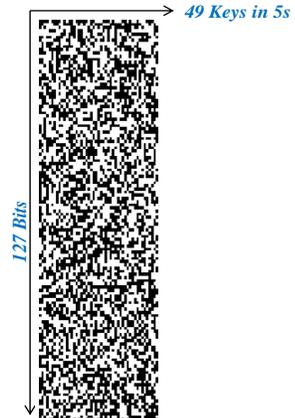**Figure 8: Key bits after quantization (LTE, indoor classroom, 2627.5 MHz)**

Figure 9: Key bits after privacy amplification (LTE, indoor classroom, 2627.5 MHz)

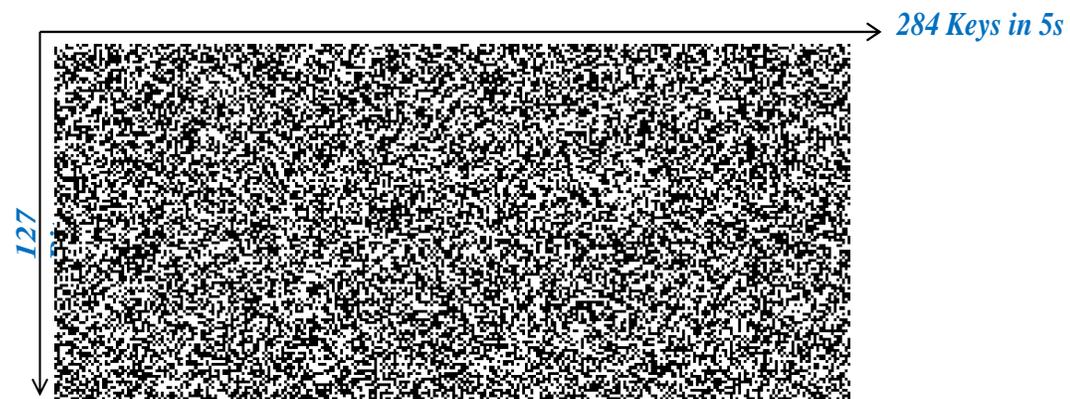**Figure 10: Key bits after quantization (LTE, 2627.5 MHz, outdoor urban street)**

Figure 11: Key bits after privacy amplification (LTE, 2627.5 MHz, outdoor urban street)
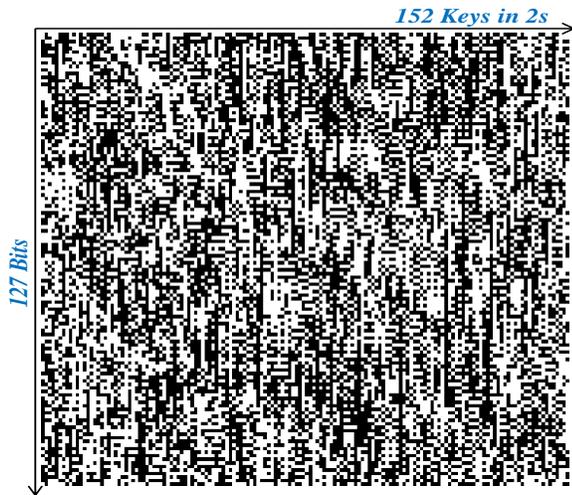
152 Keys in 2s

127 Bits

Figure 12: Key bits after quantization (WiFi, 2462 MHz , indoor LOS)

152 Keys in 2s

127 Bits
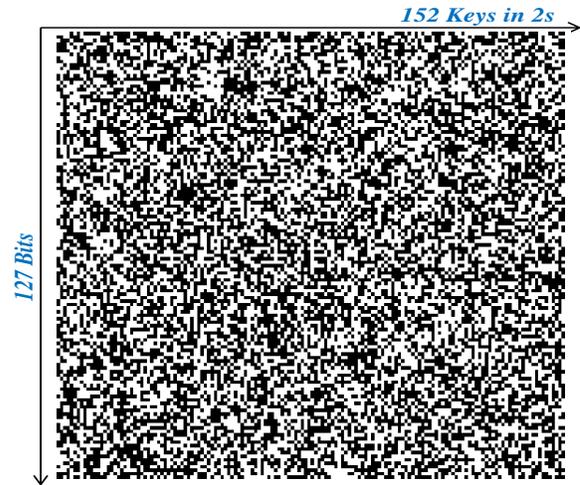
Figure 13: Key bits after privacy amplification (WiFi, 2462 MHz, indoor LOS)

171 Keys in 2s

127 Bits

Figure 14: Key bits after quantization (WiFi, 2462 MHz , indoor NLOS)

171 Keys in 2s

127 Bits

Figure 15: Key bits after privacy amplification (WiFi, 2462 MHz, indoor NLOS)

- <u>Wifi Carrier - Indoor environment (open space office) with slow mobile antennas and LOS configuration).</u>
  Figure 12 and figure 13 show that a significant number of keys are generated thanks to the mobility of the antennas and that randomness is quite good (even not perfect) after quantization
- <u>Wifi Carrier - Indoor environment (open space office) with Slow mobile antennas and NLOS configuration).</u>
  Figure 14 and figure 15 show that a larger number of keys are generated with convenient random properties just after quantization.

Finally, all these figures not only show the following trend
- more mobility and richness in the channel provide more keys of better random quality
- secret keys can be rapidly generated: 49 keys in 5 seconds in a static environment to 152 keys in 2 seconds when antennas are mobile.

Table 1: Frequency monobit test results

LTE	Indoor (2.6GHz)	Outdoor (2.6GHz)		WIFI	LOS (2.4 GHz)	NLOS (2.4 GHz)
Quantization	98% (48/49)	99% (281/284)		Quantization	87% (132/152)	100% (171/171)
Amplification	100% (49/49)	100% (284/284)		Amplification	99% (151/152)	100% (171/171)

Table 2: Run test results

LTE	Indoor (2.6GHz)	Outdoor (2.6GHz)		WIFI	LOS (2.4 GHz)	NLOS (2.4 GHz)
Quantization	27% (13/49)	80% (228/284)		Quantization	84% (128/152)	99% (169/171)
Amplification	100% (49/49)	100% (284/284)		Amplification	98% (149/152)	99% (170/171)

8.5. NIST Statistical tests

In this section we evaluate the quality of the keys by performing two randomness tests defined in the NIST Statistical Test Suite [11].

The tables above provide the results for keys generated by using the previous records of LTE and WiFi signals.

- **NIST frequency mono-bit test**

The goal of this test is to determine whether the numbers of 0s and 1s in the key are approximately the same as would be expected for a truly random sequence.

Table 1 provides the percentage of keys that successfully passed the frequency mono-bit test for the previous LTE and WiFi signals.

According to the results, almost all the keys pass the test after quantization and the privacy amplification increase the percentage of successful keys to 99% and 100%.

- **NIST runs tests**

The goal of this test is to determine whether the oscillation between 0s and 1s is too fast or too slow compared to what it is expected for a truly random sequence.

Table 2 provides the percentage of keys that successfully passed the runs test for the previous LTE and WiFi signals.

When considering quantization only, and according to the previous results,
- only a small percentage of keys generated in the indoor environment with limited mobility passed the tests
- a high percentage of keys generated with dispersive channels passed the test.

Note about the LTE Indoor case after quantization:
- Most of the keys that did not pass the runs test passed the frequency mono-bit test which is less stringent (since the CQA algorithm divides the CFR in equi-probable regions, it is expected that the number of 0s and 1s in each key should be approximately equal, which matches the frequency mono-bit test).
- The runs test better captures the randomness of a sequence. (Since CFRs captured on 1.4 MHz bandwidth only in indoor environment were a little correlated, the keys steam after quantization provides time and frequency correlation which are rejected).

Note about the benefit of privacy amplification:
After privacy amplification step, the success to NIST test is always improved, even in the static indoor environment. This final step of our SKG scheme appears really necessary for processing low dispersive radio environments and narrow band signals.

9. CONCLUSION

After recalling the basic schemes and principle of Secret Key Generation, and after describing particular implementation case to WiFi and LTE carrier, this paper outlined practical results performed in various radio-environments.

In dispersive radio-environments (with some scatterers and some mobility), a significant number of keys (of hundreds of bits each) can be extracted in a very short time under Wifi carriers and under LTE carriers. These keys have basically low cross correlation at the output and are quite robust to correlation attacks since the quantification step.

In stationary environments (with very few scatterers and no mobility, such as encountered in some indoor cases, in IoT applications, etc.) and when no channel coefficient de-correlation algorithm is applied, the extracted keys may be highly correlated and this vulnerability can be exploited by Eve to recover Bob's key.

Still in stationary environments, the quantization processing takes a large benefit of our channel coefficient de-correlation algorithm: the key rate is quite decreased but the extracted keys present low cross correlation and are robust to a correlation attack.

In any case, the proposed simplified reconciliation step with classical FEC codes provides a significant resilience of the key agreement between Alice and Bob. Only the FEC capability has to be adapted to the Signal to Noise ratio at receiving.

In any case, the proposed simplified amplification step with classical 2-Universal hash functions provides significant resilience of the key randomness against Eve's attacks, with a limited reduction of the Key lengths: NIST statistical tests were used to show that the keys shared by Alice and Bob are no correlated to the keys extracted by Eve.

To the best of our knowledge, this is the first work on a full secret key generation scheme with experimental CSI results using real field WiFi and LTE signals. Our promising results are evidence that the studied Secret Key Generation scheme can provide significant secrecy capabilities to users of public Radio Access technologies and that it can be practically implemented in existing wireless communication systems with minor modifications of the software architecture of nodes and terminals.

10. REFERENCES

[1] ZEIT, "Wie Merkels Handy abgehört werden konnte," 18 12 2014. [Online]. Available: http://www.zeit.de/digital/datenschutz/2014-12/umts-verschluesselung-umgehen-angela-merkel-handy.

[2] Ccc-Tv, «SS7map : mapping vulnerability of the international mobile roaming infrastructure,» [En ligne]. Available: https://media.ccc.de/v/31c3_-_6531_-_en_-_saal_6_-_201412272300_-_ss7map_mapping_vulnerability_of_the_international_mobile_roaming_infrastructure_-_laurent_ghigonis_-_alexandre_de_oliveira#video.

[3] I. Surveillance, «Rayzone-piranha-lte-imsi-catcher,» [En ligne]. Available: https://insidersurveillance.com/rayzone-piranha-lte-imsi-catcher/.

[4] T. intercept, «The Great SIM Heist, Hows Spies kept the key of the Encrypton Castle,» [En ligne]. Available: https://firstlook.org/theintercept/2015/02/19/great-sim-heist/.

[5] J. Wallace and R. Sharma, "Automatic secret keys from reciprocal MIMO Wireless channels: measurement and analysis," *IEEE Trans. on Info. Foren. and Sec.,* vol. 5, no. 3, pp. 381-392, Sep 2010.

[6] H. D. X. He, «Is link signature dependable for Wireless Security?,» *proceeding IEEE INFOCOM,* pp. 200-204, 2013.

[7] U. Maurer et S. Wolf, «Secret-key agreement over unauthenticated public channels. II. Privacy amplification,» *Information Theory, IEEE Transactions on,* vol. 49, n° %14, pp. 839-851, 2003.

[8] Y. Dodis, L. Reyzin et A. Smith, «Fuzzy extractors: How to generate strong keys from biometrics and other noisy data,» *Advances in cryptology-Eurocrypt,* pp. 523-540, 2004.

[9] C. Bennett, G. Brassard, C. Crepeau and U. Maurer, "Generalized privacy amplification," *Information Theory, IEEE Transactions on,* vol. 41, no. 6, pp. 1915-1923, 1995.

[10] PHYLAWS, «www.Phylaws-ict.org,» [En ligne].

[11] NIST (National Institute of standards and technology), «A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications,» 2010.