Bringing PHY-based Key Generation into the Field: An Evaluation for Practical Scenarios

René Guillaume
Corporate Research
Robert Bosch GmbH
Stuttgart, Germany
rene.guillaume@de.bosch.com

Fredrik Winzer,
Christian T. Zenger, Christof Paar
Horst Görtz Institute for IT-Security
Bochum, Germany
fredrik.winzer@rub.de

Andreas Czylwik
Chair of Communication Systems
University Duisburg-Essen
Duisburg, Germany
czylwik@nts.uni-duisburg-essen.de

***Abstract*—The need for secured communication between computationally weak wireless devices has driven the development of novel key generation protocols. The generation of symmetric cryptographic keys out of wireless channel properties turned out to be a promising approach comprising easy distribution of secret keys. Various schemes for extracting secret keys have been proposed during recent years, making the generation protocol more and more efficient for individual applications. However, often these schemes were evaluated based on theoretical models and without considering practical effects. We present a system for PHY-based key generation with two legitimate users as well as a passive attacker of equivalent power and analyze results from practical measurements in real world scenarios. Furthermore we extend practical constraints by considering heterogeneous setups and show the impact onto representative performance indicators.**

I. Introduction

In times of mobile and wireless networks, protecting confidential data has become crucial for broad acceptance of modern communication systems. As the number of communicating nodes increases and devices do not necessarily come with a proper user interface, the distribution of secret keys can turn out to be tedious. Physical layer security brought up a suitable solution, that is, extracting symmetric keys from reciprocal properties of wireless communication channels. For legitimate nodes (e.g., Alice A and Bob B) the observed reciprocal property can be understood as a common source of randomness. Properly detecting the variation of such random process is essential for generating high quality keys. While fluctuations in time or frequency domain facilitate the randomness of the secret key pair, spatial diversity of the observed channel pretends a third party (e.g. Eve E) to observe the same random sequence as A and B as long as E is not within some minimum distance to A or B. In this paper, we evaluate the performance of PHY-based key generation on basis of a fully implemented key extraction protocol. We present results for key generation rates and achieved key quality in terms of randomness, but also show how well a passive attacker performs in gathering observations correlated to those of legitimate nodes. We enhance our results by accounting for different scenarios comprising static/dynamic, indoor/outdoor, and homogeneous/heterogeneous setups.

II. Related Work

Over the last years, different aspects of PHY-based key generation have been of increasing interest and came into the focus of some considerable research works. Various authors dealt with the design and improvement of quantization schemes, i.e. mapping a sequence of a measured channel property onto a binary sequence. While some of them focused on improving the actual quantization method regarding specific performance aspects, e.g. [1]–[3], others considered peripheral signal processing [4]–[6] and combined it with their individual quantization technique in order to improve the performance of the overall key generation scheme [7]–[9]. In [7], Patwari et al. proposed the so-called High-Rate Uncorrelated Bit Extraction (HRUBE) framework. The authors basically address non-simultaneous measurements due to half duplex communication inherent in Time-Division Duplexing (TDD) as well as temporal correlation of subsequent channel measurements due to oversampling of the channel. The combined application of different reciprocity enhancement and decorrelation schemes has been recently analyzed in [10]. As it is shown, reciprocity enhancement can increase the probability of agreeing secret keys, while decorrelation can cope with additional redundancy introduced by smoothing the measured sequence. The energy aggregation achieved by decorrelation transformation is used to compress the data and to decrease the overhead by processing redundant data.

However, the above-mentioned works mainly dealt with isolated aspects of PHY-based key generation or assumed theoretical models as the underlying system. In this work, we present a complete key generation scheme according to [11] and apply it to measurements taken with two Raspberry Pi platforms representing A and B. We also consider a third party E as an eavesdropper. Inspired by the work of Jana et al. [2], performance evaluation is done for typical application scenarios, implying static and mobile, as well as indoor and outdoor use cases. In addition to our experiments with homogeneous setups, we present results from heteregeneous systems where the devices are using different frontends. Moreover, for analyzing the quality of the generated cryptographic keys, we apply a subset of tests taken from the test suite recommended by the National Institute of Standards and Technology (NIST). In difference to previous works, we additionally utilize Volf's Decomposed Context Tree Weighting (DCTW) [12] in order to compress the generated sequence and, thus, get a measure of redundancy included in the used key material.

III. Principles and Preliminaries

PHY-based key generation is based on measuring reciprocal properties of a common wireless channel. The underlying system model is shown in Fig. 1. We assume a wireless communication system with two legitimate nodes A and B and a passive attacker E.

Legitimate nodes A and B measure reciprocal properties of the physical channel, denoted by h_{BA} and h_{AB}. An appropriate property fulfilling the reciprocity requirement is, e.g., the Channel Impulse Response (CIR). However, it can be any function of the wireless channel, e.g., the Received Signal Strength Indicator (RSSI), possibly taken from different subcarriers of a multicarrier transmission system. Due to imperfect reciprocity, in practice it holds that $h_{BA} \approx h_{AB}$. Depending on eavesdropper E's distance to both legitimate nodes, E's observations h_{AE} and h_{BE} are barely correlated to h_{BA} and h_{AB}.

As it has been proposed in literature, the process for extracting keys from the communication channel can be defined by a modular structure of specific buidling blocks. The structure we are considering in this work is based on the model proposed in [11] and is shown in Fig. 2. The figure shows that our model includes the fundamental blocks *Channel Measurement*, *Quantization*, and *Information Reconciliation* which are necessary to derive a binary, aligned random sequence from an observed channel property. Additionally, the model considers *Pre-Processing* to decrease the disagreement probability and to reduce redundancy, *Randomness Testing* to assure the quality of the key, as well as *Privacy Amplification* for final entropy compression. In order to guarantee the participating nodes have generated identical keys, *Key Verification* completes the key generation process.

IV. Raspberry Pi Demonstrator Platform

We implemented a data exchange protocol on three Raspberry Pi Model B platforms (A, B, and E). All devices are equipped with a TP-Link TL-WN722N WiFi USB adapter using an Atheros AR9002U chipset and operating in 802.11n monitor mode. For heterogeneous setups they can be equipped either with TP-Link TL-WDN3200 WiFi USB adapters using Ralink RT5572 chips or with WiPi OYR-COMFAST88 adapters using Ralink RT2870 chipsets instead. In order to establish common channel probing, A periodically sends management frames to B and waits for its reply. If no response was detected after some specified time, A retransmits the according frame. The passive attacker E also receives these request and response pairs. When receiving such a probe, all three devices extract RSSI values and, thus, can measure a channel-dependent sequence over time.

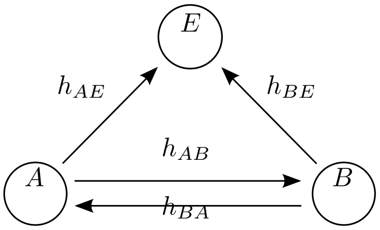

Fig. 1. Legitimate nodes A and B measure reciprocal properties of the physical channel, denoted by h_{BA} and h_{AB}. An attacker E's observations h_{AE} and h_{BE} are dependent on its relative position and are usually less correlated to h_{BA} and h_{AB} than h_{BA} is to h_{AB}.

For evaluation the measured sequences are stored and processed locally on a monitoring laptop. There we have implemented the key generation protocol illustrated in Fig. 2. The pre-processing step is optional and comprises reciprocity enhancement and decorrelation schemes (cf. [10]). In this work, pre-processing will be utilized for dynamic scenarios as will be shown later. Quantization is obligatory. A 2-bit multilevel quantization scheme is used as it is proposed in [8] with a blocklength of 250 samples. In order to cope with disagreements due to imperfect channel reciprocity, in Information Reconciliation we utilize syndrome-based alignment with BCH codes as it is proposed in [13] and [14]. Here, we consider a $(255, 47, 85)$ BCH code. The sequences' randomness are tested by applying a subset of tests from the statistical test suite proposed by the NIST [15] and for Privacy Amplification the SHA-3 hash function according to [16] is used. Finally, we assume a key verification protocol by which two nodes A and B can detect disagreeing keys.

V. Measurements & Results

We run measurements under various scenarios. These include indoor and outdoor environments with static as well as with mobile communication nodes. Intermediate scatterers etc. can potentially be mobile and, hence, block the Line Of Sight (LOS). However, we distinguish

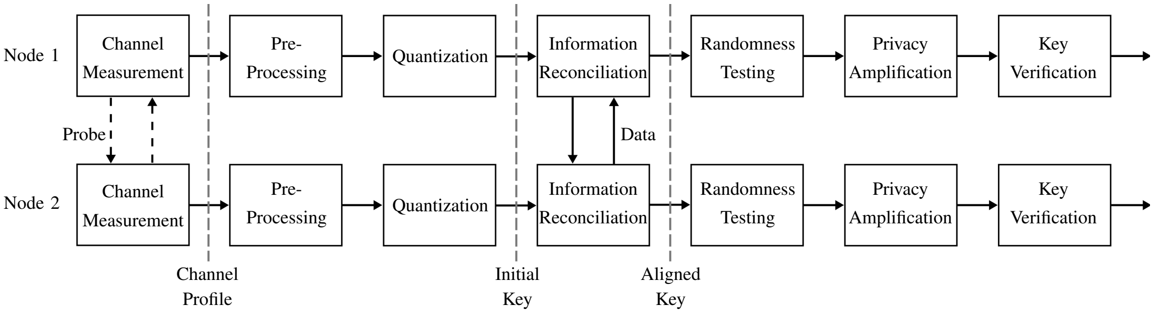

Fig. 2. Modular structure of the considered key generation process.

between LOS and Non Line Of Sight (NLOS) if a massive object like a concrete wall is blocking the LOS permanently. Generally, we probe the channel once every 1 ms and choose a timeout of 10 ms, not including delays due to packet losses etc. for a period of 10 min. For all measurements attacker E is close to node A with a fixed distance of 15 cm. As a reference, we consider homogeneous setups, that is, the use of the same type of WiFi adapter for all three communication nodes. Then, the TL-WN722N adapter of node B is replaced by the TL-WDN3200 adapter and the WiPi adapter such that E (attacking A) still has the same constraints as A.

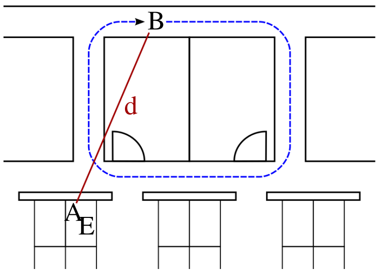

Fig. 3. Static and dynamic indoor scenario.

Fig. 3 illustrates the setup for an indoor scenario in a big room office environment (I-S1). In the assumed static scenario (red), A and E are placed on a desk, while node B is located behind a wall (NLOS) with a distance d of 7.5 m. The office is nearly empty, so there are only few moving obstacles. Additionally, we consider a dynamic scenario (I-D1), where nodes A and E are again at a fixed location, while B moves along the given trajectory (blue) with regular walking speed leading around two meeting rooms. Its distance to A varies in a range from 5 to 12 m. Due to packet losses, an effective probing rate of 170 samples/s is achieved. In Fig. 4 and Fig. 5, some collected samples of A, B and E are exemplary plotted versus the sampling indices, thus, versus time for I-S1 and I-D1, respectively. As can be noticed in Fig. 4, the mean signal strength is very static over time for all three sequences. There is only small variation between three adjacent RSSI levels which are basically random and not introduced by fading effects. Consequently, the correlation between A's (BA) and B's (AB) sequences is very low and will result in high disagreement rates. In contrast, the observations by A and B shown in Fig. 5 are highly correlated with common variation over a range of nearly 20 dB. The observation of E, however, is barely correlated and will lead to only random agreements with the secret key of A and B.

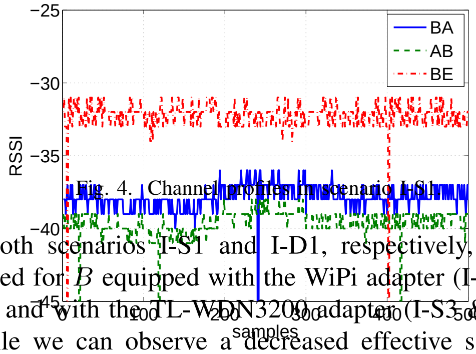

Fig. 4. Channel profiles in scenario I-S1.

Both scenarios I-S1 and I-D1, respectively, are repeated for B equipped with the WiPi adapter (I-S2 & I-D2) and with the TL-WDN3200 adapter (I-S3 & I-D3). While we can observe a decreased effective sampling rate of 122 samples/s with the WiPi adapter, with the TL-WDN3200 adapter it increases to 230 samples/s. However, in case of the TL-WDN3200 stick the vendor specific computation of RSSI values comes into play: Fig. 7 shows 500 samples of the newly equipped node B. Obviously, the TL-WDN3200, equipped with another WiFi chip than the TL-WN722N stick, has another maximum RSSI of -22 dB and, hence, B cannot resolve measurements reciprocal to those of A. As a consequence, the TL-WDN3200 adapter is not suitabel for secret key generation in our setup as it requires advanced calibration methods and will be not further considered in the following analysis.

Next, we consider an outdoor scenario located at a

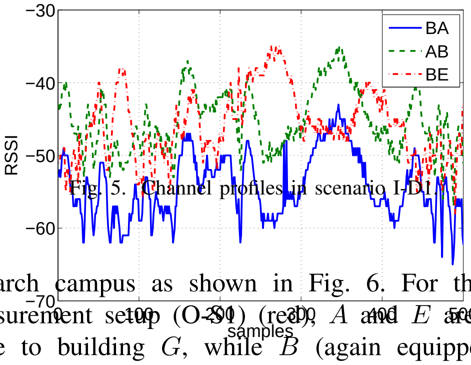

Fig. 5. Channel profiles in scenario I-D1.

research campus as shown in Fig. 6. For the static measurement setup (O-S1) (red), A and E are placed close to building G, while B (again equipped with TL-WN722N) is located around the corner such that the LOS is blocked. The distance between A and B is 20 m. Equivalently to the indoor scenario, we also perform measurements for a dynamic use case (O-D1), where B moves along the given path (blue) in walking speed between two buildings and next to two tiny lakes, while A and E stay at the same location as in O-S1. The distance between A and B varies in a range from 10 to 30 m. Again, the effective probing rate is 170 samples/s is achieved. Both scenarios O-S1 and O-D1 are repeated for B equipped with the WiPi adapter (O-S2 & O-D2). The samples collected in O-D1 are plotted versus the sampling indices in Fig. 8. It can be noticed that the measurements vary in a range of nearly 40 dB. The variation is dominated by slow fading effects introduced by the circular movement of node B. Although such variation introduces high correlation between the sequences of A and B, node E observes equivalent fading behavior, since the distance between B and E changes in the same manner as between B and A. To prevent E from successfully generating A's and B's secret key, during *Pre-Processing* we substract a moving average function from the nodes' individual measurements and, hence, get rid of slow fading in dynamic scenarios.

In order to quantitatively express the channel reciprocity between A and B respectively between B and E, we compute Pearson's correlation coefficient ρ as proposed in [17]. We show the effect of slow fading by calculating ρ_{raw} not considering pre-processing and ρ_{pp} after substracting a moving average. Another way to measure the imperfect reciprocity is to calculate the Bit Disagreement Rate (BDR) after quantization of the measured sequence. Since the multibit scheme according to [8] is applied, we can extract two bits per sample, but have to accept a tradeoff of increased BDR compared to more robust schemes like, e.g., the scheme proposed in [1]. The results for all scenarios are summarized in Tab. I. It can be noticed that, in dynamic scenarios for A and B the coefficient ρ_{raw} is higher than 90 %. Although this seems to be a good result for achieving a low amount of disagreements, also E's measurements correlates with $\rho_{\text{raw}} > 70\,\%$. This results from slow fading effects. Once we substract a moving average, A and B can maintain their advantage over E as the results for ρ_{pp} show. The correlation coefficients as well as the BDR slighty reflect the disadvantage due to different hardware. Additionally, it is easy to see that, when relying on RSSI values, movement drastically decreases the BDR and, therefore, allows faster key generation. For all scenarios, the eavesdropper E's bit sequence differs by more than 40 %.

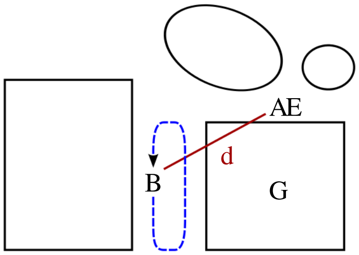

Fig. 6. Static and dynamic outdoor scenario.

TABLE I
CORRELATION COEFFICIENTS AND BIT DISAGREEMENT RATES.

	ρ_{pp}		ρ_{raw}		BDR	
Index	AB-BA	AB-BE	AB-BA	AB-BE	AB-BA	AB-BE
I-S1	0.203	0.072	0.706	0.199	0.469	0.499
I-S2	0.077	0.055	0.053	0.211	0.498	0.498
I-D1	0.835	0.373	0.956	0.795	0.180	0.414
I-D2	0.840	0.315	0.928	0.718	0.187	0.430
O-S1	0.093	0.077	0.482	0.588	0.470	0.487
O-S2	0.075	0.060	0.105	0.022	0.495	0.488
O-D1	0.815	0.449	0.990	0.954	0.221	0.381
O-D2	0.671	0.347	0.952	0.849	0.301	0.422

Once binary random sequences with $\text{BDR} > 0$ have been extracted from the channel, a key alignment protocol can cope with disagreements. Since we apply a $(255, 47, 85)$ BCH code, up to 42 bit can be corrected. Obviously, some blocks of the measured sequences can have more than 42 disagreeing bits, so these blocks will be discarded once this has been detected. However, because error correcting codes can only detect a limited amount of erroneous bits, disagreeing blocks are potentially not detected, such that they are not discarded. Consequently, even after Information Reconciliation the sequences of A and B can disagree. Therefore, the assumed key verification protocol detects if a secret

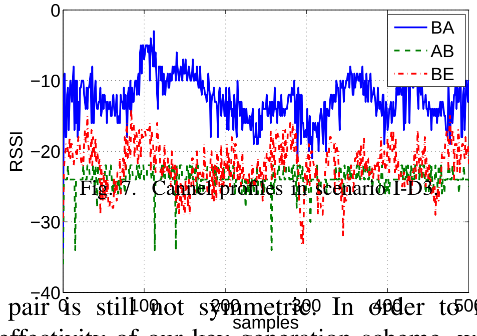

Fig. 7. Cannel profiles in scenario I-D3.

key pair is still not symmetric. In order to measure the effectivity of our key generation scheme, we define the Key Bit Generation Rate (KBGR) as the amount of agreeing key bits generated over the time it took to extract the amount of samples needed to generate this key. Hence, as we have collected a sufficient amount of data to generate multiple keys, we can divide the number of bits of all agreeing keys by the duration of the respective measurement. We assume the generated keys should have a length of 128 bit. This is in line with today's standards and is the recommended key length for long-term protection [18]. Because the utilized SHA-3 hash function allows to define different input lengths, but cannot increase the entropy per bit, the input sequence should have a length of at least 128 bit. However, the generated bit sequence may contain some redundancy, decreasing the secret key's level of security since some bit would be easier to predict. In order to compress the generated sequence and, thus, get a measure of redundancy included in the used key material we utilize Volf's Decomposed Context Tree Weighting (DCTW) [12] which is found to be a good compressor in [19]. Additionally, by exchanging a syndrome during Information Reconciliation, A and B disclose some information to the attacker E. In case of a (n, k, d) BCH code, $(n - k)$ bit per code word would be revealed. If the secret key shall have an uncertainty of 128 bit, such disclosures have to be considered when calculating the KBGR.

Tab. II summarizes the KBGR and the compression rate ν_c computed by DCTW for all scenarios, where we define ν_c as the proportion of compressed and uncompressed data. Depending on the scenario, the amount of uncompressed, aligned data varies in a range from 152000 to 330000 bit. Under the assumption that every syndrome contains the same level of redundancy as the rest of the data, the effective Key Bit Generation Rate $\mathrm{KBGR_{eff}}$ is given by

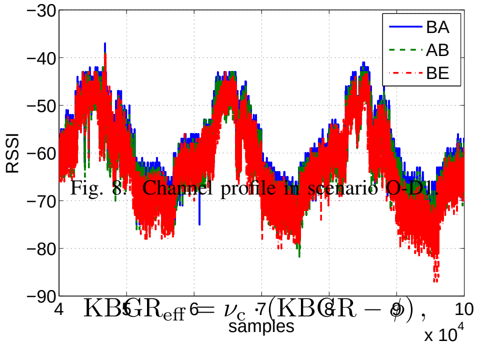

Fig. 8. Channel profile in scenario O-D1.

$$\mathrm{KBGR_{eff}} = \nu_c \cdot (\mathrm{KBGR} - \phi)\,, \qquad (1)$$

where ϕ denotes the amount of disclosed bits per time. Thus, we get an estimate $\mathrm{KBGR_{eff}}$ of how fast non-redundant, secret information can be generated in specific setups. As can be noticed from Tab. II, in static scenarios the KBGR is significantly lower than in a dynamic case. This results from very high BDR (cf. Tab I) and only few data packets which can be corrected by the considered reconciliation scheme. With approximately $45 - 58\,\%$, the compression rate ν_c indicates lower redundancy in dynamic than in static scenarios resulting in a compression rate of $20 - 36\,\%$. Hence, in dynamic scenarios a sampling rate of 130 to 170 samples/s implies an oversampling of $\sim$ 2. As a consequence of equation (1), we achieve an effective generation rate $\mathrm{KBGR_{eff}}$ of 2 to 13 bit/s in heterogenous and 8 to 13 bit/s in homogeneous, dynamic setups. Consequently, if a 128 bit key is need, basically it can be extracted in only 2-5 seconds. But if we follow strict requirements for high quality keys, the generation time can extend to up to 64 seconds. This prolongs in case of static setups. Nevertheless, previous assumptions still can be naive as disclosed syndromes may contain no redundancy and reveal maximum information of $(n - k)$ bit. In such a case, $\mathrm{KBGR_{eff}}$ will decrease even further.

TABLE II
KEY BIT GENERATION RATES.

	KBGR [bit/s]	ν_c	$\mathrm{KBGR_{eff}}$ [bit/s]
I-S1	0.84	0.308	0.05
I-S2	0.41	0.196	0.01
I-D1	158.59	0.529	15.45
I-D2	124.63	0.582	13.37
O-S1	2.53	0.360	0.17
O-S2	0.64	0.295	0.04
O-D1	85.47	0.502	7.90
O-D2	25.06	0.449	2.08

Finally, the NIST test suite is used to test the collected,

TABLE III
RELATIVE AMOUNT OF PASSED NIST TESTS.

	I-D1	I-D2	O-D1	O-D2
Serial	54.7 %	31.1 %	30.8 %	13.3 %
Runs	58.9 %	45.9 %	42.3 %	6.7 %
Frequency	74.7 %	54.1 %	53.8 %	40.0 %
BlockFrequency	94.7 %	93.2 %	84.6 %	73.3 %
FFT	47.7 %	37.8 %	73.1 %	86.7 %
Cumulative Sum	32.6 %	27.0 %	15.4 %	13.3 %

uncompressed data sets for statistical defects. Similarly to [10], we apply a subset of six tests, as some other tests require a huge amount of data to work properly. Due to the same reason, we cannot apply these tests to measurements from static scenarios, since a sufficient amount of data could not be generated to get reliable results. We divide the reconciled sequences into blocks of 1024 bit such that every test can be applied at least 50 times for each scenario. The output of each test is an indicator called *p-value*. A tested sequence passes a test if the computed p-value is greater than 0.01. Tab. III lists the relative amount of passes for each test and scenario. It can be noticed that most tests are passed more often by sequences taken from indoor scenarios. The discrete Fourier transform test is the only exception and detects signficantly more defects for indoor than for outdoor scenarios. Also the pass rates slightly decrease in case of heterogeneous setups. In general, the pass rates are quite low, what could be expected as the underlying data is uncompressed and still contains redundancy. Also other statistical defects cannot be excluded. E.g. the results from the FFT test could imply remaining defects from circular movements, which were not properly compensated by substracting the moving average function.

VI. CONCLUSION

In this work, we anaylse the performance of a practical implementation for extracting secret keys from a wireless channel. Based on channel characteristics measured for individual use cases, we provide practically motivated evaluation metrics and show that key generation is possible for homogeneous and heterogeneous systems as well. However, this is hardware dependent, as heterogeneous setups may require extensive calibration in order to work properly. We also consider a passive attacker and analyze its performance in extracting the same key as two legitimate nodes. The different scenarios show that constraints from setup and environment can affect the key generation performance, although statistics do not reflect individual environments or the type of setup. Optimization according to specific channel conditions should be part of future research work.

VII. ACKNOWLEDGEMENT

This work is supported by the Federal Ministry for Education and Research (BMBF) within the project *Providing Physical Layer Security for the Internet of Things (Prophylaxe)* (16KIS0006 and 16KIS0010).

REFERENCES

[1] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, “Radio-telepathy: Extracting a secret key from an unauthenticated wireless channel,” in *MobiCom ’08*, ACM, 2008.

[2] S. Jana, S. Premnath, M. Clark, S. Kasera, N. Patwari, and S. Krishnamurthy, “On the effectiveness of secret key extraction from wireless signal strength in real environments,” in *MobiCom ’09*, pp. 321–332, ACM, 2009.

[3] J. Wallace, C. Chen, and M. Jensen, “Key generation exploiting mimo channel evolution: Algorithms and theoretical limits,” in *EuCAP 2009*, pp. 1499–1503, Mar 2009.

[4] S. Yasukawa, H. Iwai, and H. Sasaoka, “Adaptive key generation in secret key agreement scheme based on the channel characteristics in ofdm,” in *ISITA 2008.*, pp. 1–6, Dec 2008.

[5] C. Chen and M. Jensen, “Secret key establishment using temporally and spatially correlated wireless channel coefficients,” *Mob. Comp., IEEE Trans. on*, vol. 10, no. 2, pp. 205–215, 2011.

[6] S. T. Ali, V. Sivaraman, and D. Ostry, “Zero reconciliation secret key generation for body-worn health monitoring devices,” in *Proc. of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*, WISEC ’12, pp. 39–50, 2012.

[7] N. Patwari, J. Croft, S. Jana, and S. Kasera, “High-rate uncorrelated bit extraction for shared secret key generation from channel measurements,” *Mobile Computing, IEEE Trans. on*, vol. 9, pp. 17–30, Jan 2010.

[8] A. Ambekar, M. Hassan, and H. Schotten, “Improving channel reciprocity for effective key management systems,” in *ISSSE 2012*, pp. 1–4, Oct 2012.

[9] J. Croft, N. Patwari, and S. K. Kasera, “Robust uncorrelated bit extraction methodologies for wireless sensors,” IPSN ’10, pp. 70–81, ACM, 2010.

[10] S. Gopinath, R. Guillaume, P. Duplys, and A. Czylwik, “Reciprocity enhancement and decorrelation schemes for PHY-based key generation,” in *Globecom 2014 Workshop - Trusted Communications with Physical Layer Security*, Dec. 2014.

[11] C. T. Zenger, A. Ambekar, F. Winzer, T. Pöppelmann, H. D. Schotten, and C. Paar, “Preventing scaling of successful attacks: A cross-layer security architecture for resource-constrained platforms,”

[12] P. Volf and E. U. of Technology, *Weighting Techniques in Data Compression: Theory and Algorithms*. 2002.

[13] Y. Dodis, B. Kanukurthi, J. Katz, L. Reyzin, and A. Smith, “Robust fuzzy extractors and authenticated key agreement from close secrets,” *Information Theory, IEEE Transactions on*, vol. 58, pp. 6207–6222, Sept 2012.

[14] M. Edman, A. Kiayias, Q. Tang, and B. Yener, “On the security of key extraction from measuring physical quantities,” *CoRR*, vol. abs/1311.4591, 2013.

[15] NIST, “800-22,” *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, 2000.

[16] G. Bertoni, J. Daemen, M. Peeters, and G. V. Assche, “Keccak sponge function family main document.” Submission to NIST (Round 2), 2009.

[17] R. Guillaume, C. Zenger, A. Mueller, C. Paar, and A. Czylwik, “Fair comparison and evaluation of quantization schemes for phy-based key generation,” in *19th International OFDM Workshop 2014; Proceedings of*, Aug 2014.

[18] ECRYPT-II, “Yearly report on algorithms and keysizes (2012),” *D.SPA.20 Rev. 1.0, ICT-2007-216676*, 2012.

[19] R. Begleiter, R. El-Yaniv, and G. Yona, “On prediction using variable order markov models,” *Journal of Artificial Intelligence Research (JAIR)*, vol. 22, pp. 385–421, 2004.