Security analysis of rakeness-based compressed sensing

This is the author's version of the article that has been presented at IEEE ISCAS2016
The editorial version of the paper is available at http://dx.doi.org/10.1109/ISCAS.2016.7527215

Security Analysis of Rakeness-Based Compressed Sensing

Mauro Mangia*, Fabio Pareschi†,‡, Riccardo Rovatti‡,*, Gianluca Setti†,*
* ARCES - University of Bologna - via Toffano 2/2 - Bologna - ITALY
† ENDIF - University of Ferrara - via Saragat 1 - Ferrara - ITALY
‡ DEI - University of Bologna - viale Risorgimento 2 - Bologna - ITALY

***Abstract*—Compressed sensing, further to its ability of reducing resources spent in signal acquisition, may be seen as an implicit private-key encryption scheme. The level of achievable secrecy has been analyzed in the most classical settings, when the sensing matrix is made of independent and identically distributed entries. Yet, it is known that substantially improved acquisition can be achieved by tuning the statistics of such a matrix. The effect of such an optimization on the robustness with respect to classical cryptographic attacks is analyzed here.**

I. Introduction

The encryption scheme we consider is based on Compressed Sensing (CS) [1], [2], a method in which a signal represented by a vector $x \in \mathbb{R}^n$ is acquired by applying a linear mapping A (the $m \times n$ *encoding matrix* with $m < n$) to generate the measurements vector $y = \frac{1}{\sqrt{n}} Ax$. To recover x given y, CS leverages its *sparsity*, i.e., the fact that $x = Ds$ for some orthonormal matrix D and vector s with at most $k < m < n$ non-zero entries, as well as the ability of random matrices A to capture such information despite the dimensionality reduction.

Recovery needs the knowledge of A and this naturally leads to see the encoding process as a private-key encryption stage for which x is the plaintext, y is the ciphertext and A is the shared secret. This is the core idea in [3]–[6]. Practical implementations may take advantage of embedding basic secrecy into the same stage that performs parsimonious acquisition by identifying the *key* with the seed of a pseudo-random generator producing A both at the encoder and at the decoder.

When the entries of A are i.i.d. antipodal random variables, robustness to classical attacks is investigated in [7], [8]. Ciphertext-only attacks (COAs) are shown to be ineffective since, when n is large, they may only reveal the average energy of x. Hence, CS-based encryption enjoys *asymptotic circular secrecy*, i.e., it is asymptotically Shannon-secure [7] when the energy of the ciphertext is not an issue. Known-plaintext attacks (KPAs) are also considered, in which the attacker knows both x and y and aims at retrieving A so to identify the key and be able to seed the pseudo-random generator to anticipate future encoding matrices. In this case, robustness comes from the fact that each plaintext-ciphertext is compatible with an enormous number of antipodal matrices among which the true one sits like an indistinguishable *straw* in a haystack.

Yet, it has been recently shown [9], [10] that when the signals to acquire do not distribute their energy uniformly (i.e., when they are not white) sensing performance can be improved by generating each row $a^\top$ of A independently of the others, but with entries whose correlation $\mathcal{A} = \mathbf{E}[aa^\top]$ is adapted to the second-order statistic of x. The design of such optimized rows hinges on the concept of rakeness, i.e., on the ability of measurements in y to capture the energy of x and may greatly enhance the design of CS stages [11].

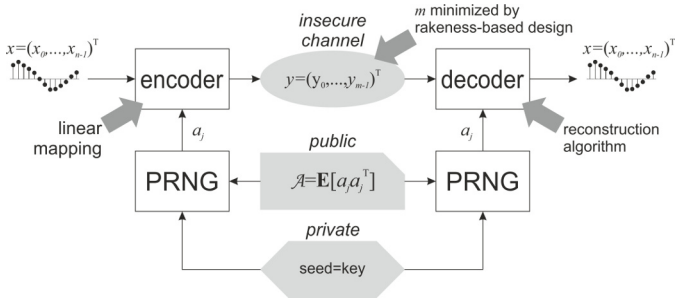

Fig. 1. A cryptographic look at a CS system optimized with a rakeness-based design.

The resulting system becomes as sketched in Figure 1 in which the encoder and the decoder receive the rows a_j of A from two identical Pseudo Random Number Generators (PRNGs) whose seed is the private key and such that $\mathcal{A}$ is publicly known according to Kerckhoff's principle. Hence, differently from classical i.i.d. CS, rakeness-based CS alters the energy of y, whose statistics can be expected to carry information other than the energy of x. Moreover, it uses non-i.i.d. sensing matrices whose statistic is known to attackers and can be exploited, thus suggesting that it may be a weaker encryption scheme.

We here address these issues proving that, even if it is optimized following a rakeness-based design flow, CS-embedded encryption still enjoys asymptotical circular secrecy and thus is robust with respect to COAs. Moreover, an approximate but effective theory is developed allowing to quantify the success chance of KPAs that is still low enough to claim practical security.

II. System model and assumptions

Since the rows of A are independent, we concentrate on one of them $a^\top = (a_0, \ldots, a_{n-1}) \in \{-1, +1\}^n$ and on the corresponding scalar measurement $y = \frac{1}{\sqrt{n}} a^\top x$ where $x = (x_0, \ldots, x_{n-1})^\top$ with $\mathbf{E}[x_j^2] = W_x$. We assume that a and x are slices of zero-mean and independent mixing stochastic processes with power spectra $S_a(f)$ and $S_x(f)$ on which we put the regularity constraint of being square summable. The

This is the author's version of the article that has been presented at IEEE ISCAS2016
The editorial version of the paper is available at http://dx.doi.org/10.1109/ISCAS.2016.7527215

eigenvalues of the correlation matrix $\mathcal{A}$ are $\lambda_j \geq 0$, the eigenvalues of $\mathcal{X} = W_x^{-1}\mathbf{E}[xx^\top]$ are $\mu_j \geq 0$, and the two matrices are assumed to share the same set of orthonormal eigenvectors q_j so that $\lambda_j q_j = \mathcal{A}a_j$ ad $\mu_j q_j = \mathcal{X}q_j$.

By applying the main theorem in [13, chapter 5], we know that $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\lambda_j = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\mu_j = 1$, and that, since $S_a(f)$ is assumed to be square-summable, we may define the finite quantity $\sigma^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\lambda_j^2 = \int_{-1/2}^{1/2} S_a^2(f)\mathrm{d}f$. As an additional assumption on the relationship between a and x, we assume that there is a finite quantity ξ^2 such that $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\lambda_j\mu_j = \xi^2$.

The two parameters σ^2 and ξ^2 are those ultimately affecting robustness against COAs and KPAs.

III. Asymptotic circular secrecy

Robustness against COAs depends on the statistics of $\frac{y}{\sqrt{W_x}} = \frac{1}{\sqrt{n}}\ a^\top\frac{x}{\sqrt{W_x}}$, where the power normalization highlights a scalar product between two unit-power vectors.

Exploiting the assumption that the processes generating the vectors a and x are mixing we may apply the Lindeberg-Feller central limit theorem [14, Theorem 27.4] to recognize that $\frac{y}{\sqrt{W_x}}$ is asymptotically distributed as a zero-mean Gaussian.

To compute the variance, note that since the scalar product is invariant with respect to orthonormal transformation we may compute measurements as $\frac{y}{W_x} = \frac{1}{\sqrt{n}}\ \bar{a}^\top\bar{x}$ where $\bar{a} = Q^\top a$, $\bar{x} = Q^\top\frac{x}{\sqrt{W_x}}$ and Q is the matrix aligning the eigenvectors q_j as columns. The components of $\bar{a}$ and $\bar{x}$ are uncorrelated and $\bar{a}_j$ has variance λ_j while $\bar{x}_j$ has variance μ_j. Overall $\frac{y}{\sqrt{W_x}} = \frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\bar{a}_j\bar{x}_j$ has variance $\frac{1}{n}\sum_{j=0}^{n-1}\lambda_j\mu_j$ whose limit is ξ^2. Overall

$$y \overset{n\to\infty}{\sim} \mathcal{N}\left(0, \xi^2 W_x\right) \tag{1}$$

Hence, the encryption we analyze is asymptotically circularly secret as defined in [7], since $\xi^2 W_x$ is the only information that an attacker may infer from the observation of the ciphertexts y.

The results in [5], [7] are a special case of (1). In fact, when A is made of i.i.d. antipodal entries, $\mathcal{A}$ is the identity matrix, $\lambda_j = 1$ for all j, and, independently of $\mathcal{X}$, the information that leaks is W_x. Yet, asymptotic circular security holds also adopting a rakeness-based design flow [9], [10] that computes an optimized $\mathcal{A}$ exploiting the same eigenvectors of $\mathcal{X}$ and modulating its eigenvalues.

IV. Distance between rows of $\mathcal{A}$ and KPAs

As discussed in [8], once that x and y are known, solving the measurement equation $y = \frac{1}{\sqrt{n}}a^\top x$ for $a \in \{-1,+1\}^n$ is usually not a difficult task. Yet, the solution is not unique and, among the extremely huge amount of solutions, the chance of hitting one that is at least close to the true one is negligible.

Yet, if $\mathcal{A}$ is not the identity, the attacker may exploit such an information. The most elementary way of doing so is to use the same generator employed at the encoder to produce the true a, to generate candidate $\hat{a}$'s and match them with the measurements equation. With this, rows that have been most likely used in the encoding are generated and tried first, hopefully increasing the chance of finding a good approximation of the true a.

To assess the threat of such a KPA, we must know the probability that two sensing vectors a' and a'' (representing the one used by the encoder and the one guessed by the attacker) obeying the same $\mathcal{A}$, have a certain (small) Hamming distance $\Delta(a', a'')$, i.e., differ in $\Delta(a', a'')$ entries. We approximate such a probability for large n and analyze the distribution of $\frac{1}{n}\Delta(a', a'') = \frac{1}{4}\|\frac{a'}{\sqrt{n}} - \frac{a''}{\sqrt{n}}\|^2$ when $(\frac{a'}{\sqrt{n}})^\top x = (\frac{a''}{\sqrt{n}})^\top x = y$.

The idea is to discard the antipodality constraint and focus on the unit-norm vectors $u' = {}^{a'}\!/\!_{\sqrt{n}}$ and $u'' = {}^{a''}\!/\!_{\sqrt{n}}$ modeling them as perturbations of Gaussian vectors with zero mean and correlation $\mathcal{A}$. By now, we neglect the linear constraint due to the measurement equation, to reintroduce it at a later stage.

More formally, we consider zero-mean Gaussian vectors g' and g'' with correlation $\mathcal{A}$ and set $v' = \frac{g'}{\sqrt{n}}$ and $v'' = \frac{g''}{\sqrt{n}}$. Clearly $\mathbf{E}[\|v'\|^2] = \mathbf{E}[\|v''\|^2] = 1$ but we may prove that v' and v'' become very close to unit-norm as $n \to \infty$. To see why, consider the vectors $\gamma' = Qg'$ and $\gamma'' = Qg''$ that are Gaussian but made of uncorrelated and thus independent components with variances λ_j. Since Q does not alter Euclidean, if v is either v' or v'' we have $\|v\|^2 = \frac{1}{n}\|\gamma\|^2 = \frac{1}{n}\sum_{j=0}^{n-1}\left(\gamma_j^2 - \lambda_j\right) + \frac{1}{n}\sum_{j=0}^{n-1}\lambda_j$. The first term is a sum of independent random variables with zero mean and variance

$$\frac{1}{n^2}\sum_{j=0}^{n-1}\mathbf{E}\left[\left(\gamma_j^2 - \lambda_j\right)^2\right] = \frac{1}{n^2}\sum_{j=0}^{n-1}\mathbf{E}[\gamma_j^4] - \lambda_j^2 = \frac{2}{n^2}\sum_{j=0}^{n-1}\lambda_j^2 = \frac{2\sigma^2}{n}$$

Since $\frac{1}{n}\sum_{j=0}^{n-1}\lambda_j = 1$ we have $\|v'\|, \|v''\|^2 \overset{n\to\infty}{\sim} \mathcal{N}\left(1, \frac{2\sigma^2}{n}\right)$. Hence v' and v'' can be seen as unit-norm vectors u' and u'' superimposed with a zero-mean Gaussian-length perturbation with variance $\frac{2\sigma^2}{n}$ (see Figure 2-a).

The squared length of the difference between v' and v'' can be characterized by resorting to the above $\gamma' = Qg'$ and $\gamma'' = Qg''$. In fact Q preserves Euclidean distance and $\|v' - v''\|^2 = \frac{1}{n}\|\gamma' - \gamma''\|^2$ that can be recast into

$$\|v' - v''\|^2 = \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\left[\left(\gamma_j' - \gamma_j''\right)^2 - 2\lambda_j\right] + \frac{2}{n}\sum_{j=0}^{n-1}\lambda_j$$

Since γ_j' and γ_j'' are independent and Gaussian with variance λ_j, the random variables in the above square brackets are independent, have zero mean and variance $8\lambda_j^2$.

Since $\frac{2}{n}\sum_{j=0}^{n-1}\lambda_j = 2$ and $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}8\lambda_j^2 = 8\sigma^2$, we have $\|v' - v''\|^2 \overset{n\to\infty}{\sim} \mathcal{N}\left(2, \frac{8\sigma^2}{n}\right)$. Yet, recall that v' and v'' can be seen as unit vectors u' and u'' plus two perturbations $\overset{n\to\infty}{\sim} \mathcal{N}\left(0, \frac{2\sigma^2}{n}\right)$. These perturbations make the variance of $\|v' - v''\|^2$ larger than that of $\|u' - u''\|^2$ that, is therefore asymptotically distributed as $\mathcal{N}\left(2, \frac{4\sigma^2}{n}\right)$. If the same vectors are rescaled so that their length is r instead of 1, the asymptotic distribution becomes $\mathcal{N}\left(2r, \frac{4\sigma^2 r^2}{n}\right)$.

To exploit this result in our setting, observe that if an antipodal a and its unit-length counterpart $u = \frac{a}{\sqrt{n}}$ are

This is the author's version of the article that has been presented at IEEE ISCAS2016
The editorial version of the paper is available at http://dx.doi.org/10.1109/ISCAS.2016.7527215

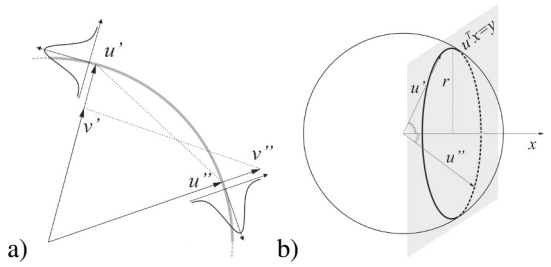

Fig. 2. a) Gaussian normalized vectors v' and v'' as perturbed unit-norm vectors u' and u''. b) The effect of the normalized measurement constraint on the distribution of $\bar{a}'$ and $\bar{a}''$.

constrained to satisfy $\frac{1}{\sqrt{n}}a^\top x = u^\top x = y$ for some x and y, then u belongs to the intersection of the unit-sphere and an hyperplane whose distance from the center of the sphere is $\frac{y}{\|x\|^2}$ (Figure 2-b)). Such an intersection is an $(n-1)$-dimensional sphere with radius $r = \sqrt{1 - \frac{y^2}{\|x\|^2}}$. We approximate such a radius with the value it assumes when the random variables involved in its computation are given their average values.

First, since $y \overset{n\to\infty}{\sim} \mathcal{N}\left(0, \xi^2 W_x\right)$ then y^2 becomes a χ^2 random variable with average $\xi^2 W_x$. Second, since the process generating x is mixing and thus ergodic, $\frac{\|x\|^2}{n} \to W_x$ and thus $\|x\|^2 \simeq nW_x$. All together this gives the approximate value $r \simeq \sqrt{1 - \frac{\xi^2}{n}}$ that can be substituted in the above asymptotic distribution to say that $\|u' - u''\|^2 \overset{n\to\infty}{\sim} \mathcal{N}\left(2\sqrt{1-\frac{\xi^2}{n}}, 4\left(1-\frac{\xi^2}{n}\right)\frac{\sigma^2}{n}\right)$ and thus finally

$$\frac{1}{n}\Delta(a', a'') \overset{n\to\infty}{\sim} \mathcal{N}\left(\frac{\sqrt{1-\frac{\xi^2}{n}}}{2}, \frac{1-\frac{\xi^2}{n}}{4}\frac{\sigma^2}{n}\right) \qquad (2)$$

A. Numerical experiments

To test the above theory and assess how much its asymptotic nature fits finite-n configurations, we consider an input signal coming from a stationary process with an exponential correlation matrix $\mathcal{X}_{j,k} = r^{|j-k|}$ corresponding to the power spectrum $S_x(f) = (1-r^2)/\left(1 + r^2 - 2r\cos(2\pi f)\right)$ that may exhibit high-pass ($-1 < r < 0$), white ($r = 0$), or low-pass ($0 < r < 1$) profiles. Starting from $\mathcal{X}$, the design flow proposed in [9], [10] is used to compute $\mathcal{A}$. From that matrix and from [15] we get the correlation of an easy-to-generate zero-mean jointly Gaussian vector that can be clipped to obtain a.

To explore the design space, the signal dimensionality is taken as $n = 64, 96, 128, 192, 256, 384, 512$ and different spectra are considered for $r = 0, \pm 0.2, \pm 0.5, \pm 0.7, \pm 0.9$. Note that, since the coefficients of the characteristic polynomial of $\mathcal{X}$ depend only on even powers of r, we expect negative and positive values of such a parameter to lead to the same result.

For each configuration we compute 10^5 measurements, each trial characterized by a different random instances of x, a and thus y. The empirical distribution of y is matched against the theoretical prediction (1) in Figure 3 for some n and r. Beyond the good visual agreement, since the secrecy of CS encryption depends on the statistical indistinguishability of

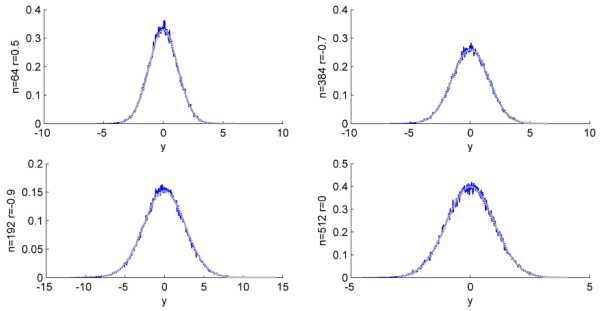

Fig. 3. Match between empirical and theoretical distribution of normalized measurements for different values of n and r.

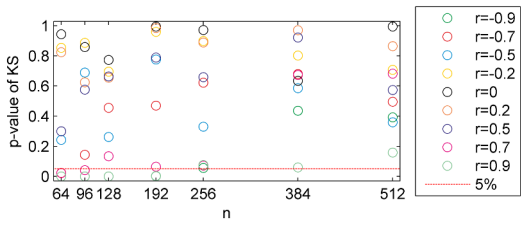

Fig. 4. P-values of the Kolomogorov-Smirnov test for normality on the empirical distribution of measurements for different values of n and r. Dots above the red line correspond to empirical distribution that would commonly be accepted as Gaussian.

those profiles from Gaussian ones, we perform a Kolmogorov-Smirnov test on each set of measurements yielding the corresponding p-value, i.e., the probability that the empirical distribution of an equal number of samples drawn from a true Gaussian distribution deviates more than the observed one from the ideal profile. The results of such tests are reported in Figure 4 against the values of n and along with the 5% significance level that is commonly used to distinguish Gaussian data from non-Gaussian one. Note that, coherently with the asymptotic nature of (1) as n increases, all measurements sets agree with the Gaussian distribution.

To validate (2), for each of the above configurations we simulate many KPAs. In each trial a newly generated signal x is quantized in 14-bits words so that each entry is an integer in $\{-L, \ldots, L\} \setminus \{0\}$ with $L = 8192$, and an antipodal sensing vector a is generated clipping a properly designed Gaussian vector. The corresponding measurement is $y = a^\top x$.

The same generator used for a is exploited to produce a sequence of candidate antipodal vectors $\hat{a}$. When a candidate satisfies $\hat{a}^\top x = y$ the hamming distance $\Delta(a, \hat{a})$ is collected.

The empirical distribution of $\Delta(a, \hat{a})/n$ is matched against the theoretical prediction (2) in Figure 5 for some n and r.

A more quantitative view is given by Figures 6 and 7 in which we report the empirical value against the theoretical value for the average and the standard deviation of $\Delta(a, \hat{a})/n$. In Figure 6 this is done for $r = \pm 0.7$ sweeping $n = 64, \ldots, 512$ while in Figure 7 we set $n = 256$ and sweep $r = 0, \pm 0.2, \ldots, \pm 0.9$. Notwithstanding the simplifying assumptions, theory and simulations agree and, as expected, negative and positive values of r yield substantially the same behavior, confirming the role of the eigenvalues of $\mathcal{X}$.

As an example of how this can be used to assess ro-

This is the author's version of the article that has been presented at IEEE ISCAS2016
The editorial version of the paper is available at http://dx.doi.org/10.1109/ISCAS.2016.7527215

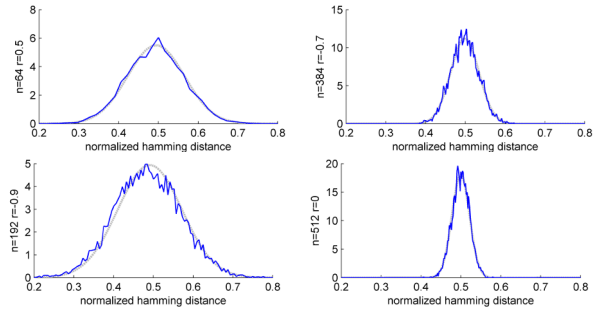

Fig. 5. Match between empirical and theoretical distribution of normalized hamming distances for different values of n and r.

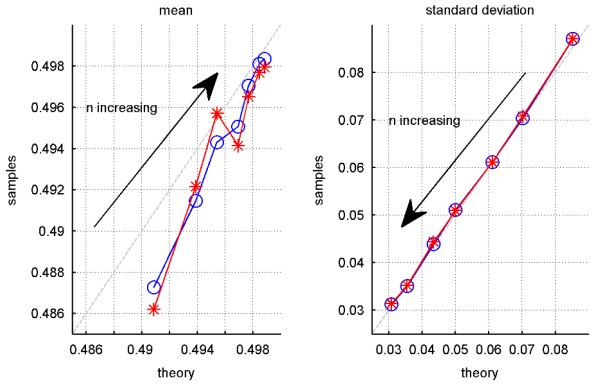

Fig. 6. Match between empirical and theoretical mean and standard-deviations of $\Delta(a,\hat{a})/n$ for $r = 0.7$ (blue circles), $r = -0.7$ (red asterisks) and different values of n.

bustness against KPAs, assume that the rakeness-based CS is optimized to acquire $n = 256$-dimensional chunks of a low-pass process characterized by $r = 0.7$. Since in this case $\Delta(a,\hat{a})/n$ is distributed as $\mathcal{N}(0.497, 0.0019)$, the probability to generate a candidate satisfying the measurement equation and with less than 16 differences from the true sensing vector is $p_{\mathrm{KPA}} = \frac{1}{2}\mathrm{erfc}\left(\frac{0.4970-\frac{16}{256}}{\sqrt{2\times 0.0019}}\right) \simeq 2.13 \times 10^{-23}$.

Under repeated threat of KPA, such a probability can be translated into an estimate of the maximum time that may elapse between two subsequent key changes (a countermeasure that makes all previous KPAs uneffective). In fact, the probability of T repeated failures is $(1-p_{\mathrm{KPA}})^T$ and to ensure that this is not less than a prescribed security level ζ we should have $T \leq \log(\zeta)/\log(1-p_{\mathrm{KPA}})$. For $\zeta = 0.9999$ we obtain a key

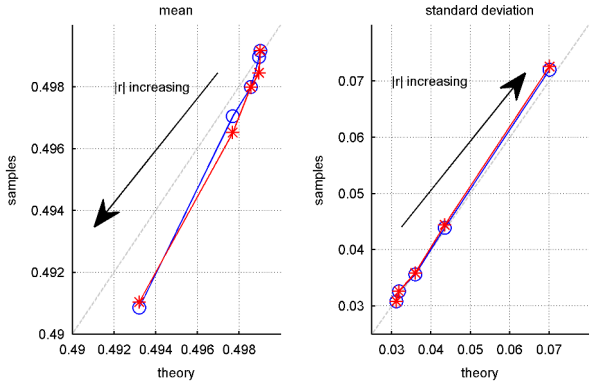

Fig. 7. Match between empirical and theoretical mean and standard-deviations of $\Delta(a,\hat{a})/n$ for $n = 256$ and different positive (blue circles) and negative (red asterisks) values of r.

lifetime not larger than $T = 4.70 \times 10^{18}$ attack opportunities.

Though this is still surely acceptable note that, as it can be intuitively accepted, since rakeness-based CS provides the attacker with additional side information ($\mathcal{A}$) it is somehow less secure. In fact, classical i.i.d. CS can be modeled setting $\mathcal{A}$ to the identity matrix so that $\sigma^2 = \xi^2 = 1$. In this case theory predicts that $\Delta(a,\hat{a})/n$ distributes as $\mathcal{N}(0.499, 0.00973)$, yielding $p_{\mathrm{KPA}} = 8.23 \times 10^{-45}$. To maintain the previous security level the maximum time between key changes is bounded the much larger $T = 1.22 \times 10^{40}$ attack opportunities.

V. Conclusion

The adoption of a rakeness-based design for CS optimizes acquistion performance but affects the secrecy of the implicit encryption. Such an effect can be quantified adopting the theory developed in this paper to find that a non negligible level of security is still achievable.

References

[1] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candes and M. B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[3] Y. Rachlin and D. Baron, "The secrecy of compressed sensing measurements," in *2008 Forty Sixth Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2008, pp. 813–817, 2008.

[4] A. Orsdemir, H. O. Altun, G. Sharma, and M. F. Bocko, "On the security and robustness of encryption via compressed sensing," in *2008 IEEE Military Communications Conference (MILCOM)*. IEEE, 2008, pp. 1–7.

[5] V. Cambareri, J. Haboba, F. Pareschi, R. Rovatti, G. Setti, and K. W. Wong, "A two-class information concealing system based on compressed sensing," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2013, pp. 1356–1359.

[6] J. Barcelo-Llado, A. Morell, and G. Seco-Granados, "Amplify-and-Forward Compressed Sensing as a Physical-Layer Secrecy Solution in Wireless Sensor Networks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 839–850, 2014.

[7] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Low-complexity multiclass encryption by compressed sensing," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2183–2195.

[8] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "On known-plaintext attacks to a compressed sensing-based encryption: A quantitative analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 10, pp. 2182–2195, 2015.

[9] M. Mangia, R. Rovatti, and G. Setti, "Rakeness in the design of analog-to-information conversion of sparse and localized signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 5, pp. 1001–1014, 2012.

[10] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "A rakeness-based design flow for analog-to-information conversion by compressive sensing," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2013, pp. 1360–1363.

[11] F. Pareschi, P. Albertini, G. Frattini, M. Mangia, R. Rovatti, and G. Setti, "Hardware-algorithms co-design and implementation of an analog-to-information converter for biosignals based on compressed sensing," *IEEE Trans. Biomed. Circuits Syst*, 2016.

[12] C. Fantuzzi, S. Simani, S. Beghelli, and R. Rovatti, "Identification of piecewise affine models in noisy environment," *International Journal of Control*, vol. 75, no. 18, pp. 1472–1485, 2002.

[13] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Chelsea Publishing Company, 1984.

[14] P. Billingsley, *Proability and Measure*. Wiley, 2008.

[15] J. H. Van Vleck and D. Middleton, "The spectrum of clipped noise," *IEEE Proceedings*, vol. 54, no. 1, pp. 2–19, 1966.