PRNU-leaks: facts and remedies

Fernando Pérez-González
Signal Processing in Communications Group
Atlanttic Research Center
Vigo, Spain
fperez@gts.uvigo.es

Samuel Fernández-Menduiña
EE Department
Imperial College
London, UK
sf219@ic.ac.uk

***Abstract*—We address the problem of information leakage from estimates of the PhotoResponse Non-Uniformity (PRNU) fingerprints of a sensor. This leakage may compromise privacy in forensic scenarios, as it may reveal information from the images used in the PRNU estimation. We propose a new way to compute the information-theoretic leakage that is based on embedding synthetic PRNUs, and presesent affordable approximations and bounds. We also propose a new compact measure for the performance in membership inference tests. Finally, we analyze two potential countermeasures against leakage: binarization, which was already used in PRNU-storage contexts, and equalization, which is novel and offers better performance. Theoretical results are validated with experiments carried out on a real-world image dataset.**

***Index Terms*—Fingerprint, PRNU, Leakage, Information theory, Membership inference.**

I. Introduction

The PhotoResponse Non-Uniformity (PRNU) is a multiplicative pattern due to unique imperfections in the manufacturing of imaging sensors that serves as a fingerprint that can be used in forensic applications to solve camera attribution/identification problems and to detect intentional manipulations [1], [2]. Being the PRNU a very weak signal, its extraction for a given sensor requires a number of images known to be captured by it and preprocess those images with a denoising filter to remove the interference they cause. Unfortunately, this denoising is not perfect and leaves in the estimated PRNU traces from the original images used during the extraction. Such leakage represents a threat to privacy, even more so considering that the images involved in criminal investigations are often of a very sensitive nature, as in cases involving sexually-oriented crimes. Since law-enforcement agencies customarily share other fingerprints (like those provided by Microsoft's PhotoDNA tool) in order to uncover child-pornography filesharing networks, there is a danger that this practice is extended to camera fingerprints without properly assessing the privacy risks of PRNU-leakage. We must remark that with ever increasing image sizes a good performance may be achieved by fingerprint detectors even with PRNUs extracted with few images, but this may give a false impression of security, because, as we will see, the leakage increases considerably when the number of images decreases.

GPSC is funded by the Agencia Estatal de Investigación (Spain) and the European Regional Development Fund (ERDF) under project WINTER (TEC2016-76409-C2-2-R). Also funded by Xunta de Galicia and ERDF under projects Agrupación Estratéxica Consolidada de Galicia accreditation 2016-2019 and Grupo de Referencia ED431C2017/53.

To the best of our knowledge, this, together with the companion paper [3], is the first time this leakage is explicitly identified. In [3] we discuss two measures to quantify it: one based on the mutual information, and another based on the success rate of a membership inference test. We acknowledge, however, that such leakage was indirectly exploited in the so-called *triangle test* to counteract the *PRNU copy attack* [4].

In this paper we take a further step with respect to [3] in three directions: 1) quantifying the leakage by proposing a synthetic-embedding procedure to exactly evaluate the mutual information and compare the results with the lower bound proposed in [3]; 2) a definition of a Signal to Noise Ratio (SNR) for the membership inference test that compresses its performance down to a single indicator; 3) the proposal and analysis of two measures for mitigating the leakage, based on postprocessing the estimated PRNU. One of them, namely, equalization of the PRNU, is shown to reduce the leakage while preserving the PRNU detection performance, so it is strongly recommended for storing and sharing PRNUs.

The rest of the paper is organized as follows: in Sect. II we discuss the existence of PRNU leakage and its mitigation; in Sect. III we quantify, and provide bounds and approximations to the leakage; Sect. IV contains the results of experiments carried on an image dataset, and Sect. V presents our conclusions.

A. Notation

Matrices, written in boldface, represent luminance images. All are assumed to be of size $M \times N$. The (k,l)th pixel of image $\mathbf{X}$ is refered to as $X[k,l]$. Given two matrices, $\mathbf{X}$ and $\mathbf{Y}$, its Hadamard product $\mathbf{Z} = \mathbf{X} \circ \mathbf{Y}$ is such that $Z[k,l] = X[k,l] \cdot Y[k,l]$, and the Hadamard division $\mathbf{Z} = \mathbf{X} \oslash \mathbf{Y}$ such that $Z[k,l] = X[k,l]/Y[k,l]$ provided that $Y[k,l] \neq 0$ for all k,l. The Frobenius cross-product of $\mathbf{X}$ and $\mathbf{Y}$ is defined as $\langle \mathbf{X}, \mathbf{Y} \rangle_F \doteq \mathrm{tr}\left(\mathbf{X}^T\mathbf{Y}\right)$, where $\mathrm{tr}(\cdot)$ denotes trace and T transpose.

II. Leakage existence and mitigation

The starting point for PRNU estimation is the simplified sensor output model presented in [1] and summarized as

$$\mathbf{Y} \doteq (\mathbf{1} + \mathbf{K}) \circ \mathbf{X} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y}$ is the output of the sensor, $\mathbf{K}$ is the multiplicative PRNU term, $\mathbf{X}$ is the noise-free image and $\mathbf{N}$ collects all the non-multiplicative noise sources.

The PRNU $\mathbf{K}$ can be estimated from a set of L images $\{\mathbf{Y}^{(i)}\}_{i=1}^{L}$ coming from the same sensor by constructing the set of residuals $\mathbf{W}^{(i)} \doteq \mathbf{Y}^{(i)} - \hat{\mathbf{X}}^{(i)}$, $i = 1, \cdots, L$, where $\hat{\mathbf{X}}^{(i)}$ is the result of passing $\mathbf{Y}^{(i)}$ through a denoising filter (in this paper, we have used the most popular one from [7]). Then, the estimate becomes $\hat{\mathbf{K}} = \left(\sum_{i=1}^{L} \mathbf{W}^{(i)} \circ \hat{\mathbf{X}}^{(i)}\right) \oslash \mathbf{R}$, where $\mathbf{R} \doteq \sum_{i=1}^{L} \hat{\mathbf{X}}^{(i)} \circ \hat{\mathbf{X}}^{(i)}$. In [3] we show that $\hat{\mathbf{K}}$ can be written as

$$\hat{\mathbf{K}} = \mathbf{\Omega} \circ \mathbf{K} + \mathbf{N}_k, \tag{2}$$

where $\mathbf{\Omega} \doteq \left(\sum_{i=1}^{L} \mathbf{\Omega}^{(i)} \circ \hat{\mathbf{X}}^{(i)} \circ \mathbf{X}^{(i)}\right) \oslash \mathbf{R}$ is a function of the used images. Experiments reported in [5] show that $\mathbf{N}_k$ can be well-modeled by an independent Gaussian process with variance at the (k,l)th position denoted by $\gamma^2[k,l]$.

As we discuss and prove through information-theoretic lowerbounding in [3], there is a significant amount of leakage from the images $\{\mathbf{Y}^{(i)}\}_{i=1}^{L}$ into $\hat{\mathbf{K}}$. This is quite apparent in Fig. 1d in which the local variance $\gamma^2[l,k]$ of $\hat{\mathbf{K}}$ estimated with a 9×9-pixel window is represented as $\log(0.5 + 1/\gamma^2[k,l])$, together with 3 of the 25 images (Figs. 1a through 1c) used for the estimation.

A. Leakage mitigation

In this paper we analyze two possible countermeasures against information leakage. One is to simply binarize the PRNU estimate $\hat{\mathbf{K}}$ by sample-wise taking its sign. We will denote the output as $b(\hat{\mathbf{K}})$. The rationale for the binarization is that it has proven to be effective as a way of reducing the storage size for the PRNU while somehow preserving the performance in detection [8]. The other remedy, which we will show in this paper to be more effective, is to simply divide $\hat{\mathbf{K}}$ by an estimate of its local standard deviation $\hat{\mathbf{\Gamma}}$ in an attempt to remove the contextual traces of images $\{\mathbf{Y}^{(i)}\}_{i=1}^{L}$ that are visible in $\mathbf{\Gamma}$ (see Fig. 1). We will denote this *equalized* PRNU by $n(\hat{\mathbf{K}})$ which can be formally written as $n(\hat{\mathbf{K}}) \doteq \hat{\mathbf{K}} \oslash \hat{\mathbf{\Gamma}}$.

III. Leakage quantification

A. Novel information-theoretic bounds and approximations

In this section we present a novel approach to compute $I(\mathbf{N}_k, \hat{\mathbf{K}})$ which, as justified in [3], serves as a lower bound to the information leakage $I(\{\mathbf{Y}^{(i)}\}_{i=1}^{L}, \hat{\mathbf{K}})$ of the the set of images used for the estimation $\{\mathbf{Y}^{(i)}\}_{i=1}^{L}$into the estimated PRNU. In [3], the lack of knowledge of $\mathbf{\Omega}$ motivates the search for a further lower bound on $I(\mathbf{N}_k, \hat{\mathbf{K}})$ which through a water-filling argument was shown to be:

$$I(\mathbf{N}_k, \hat{\mathbf{K}}) \geq \frac{1}{2} \sum_{l,j} \log \left(1 + \frac{2}{\sqrt{1 + 4/(\mu \cdot \gamma^2[l,k])} - 1}\right) \tag{3}$$

where μ is the solution to the equation

$$\frac{1}{2} \sum_{k,l} \gamma^2[l,k] (\sqrt{1 + 4/(\mu \cdot \gamma^2[l,k])} - 1) = P \tag{4}$$

and P is the total available PRNU energy, which can be obtained by computing two indepedent estimates of $\hat{\mathbf{K}}$ by splitting the set $\{\mathbf{Y}^{(i)}\}_{i=1}^{L}$ into two disjoint subsets. If $\hat{\mathbf{K}}_1$, $\hat{\mathbf{K}}_2$ denote those estimates, then P can be estimated as $\hat{P} = \langle \hat{\mathbf{K}}_1, \hat{\mathbf{K}}_2 \rangle_F$.

We remark that the main obstacle towards knowing $\mathbf{\Omega}$ is the fact that even for large L, the estimations $\hat{\mathbf{K}}$ are dominated by the estimation noise $\mathbf{N}_k$, so the effect of $\mathbf{\Omega}$ is not directly observable. In order to overcome this problem, in this paper we take a novel approach: we embed a synthetic PRNU $\mathbf{K}_s^{(j)}$ with variance $\sigma_{k,s}^2$ following the basic multiplicative model in (1) (where noise is set to zero) in the set $\{\mathbf{Y}^{(i)}\}_{i=1}^{L}$ and estimate the corresponding $\hat{\mathbf{K}}_s^{(j)}$ which, according to the model (2) will be such that $\hat{\mathbf{K}}_s^{(j)} = \mathbf{\Omega} \circ \mathbf{K}_s^{(j)} + \mathbf{N}_k^{(j)}$. This procedure is repeated for R different synthetic PRNUs, $\hat{\mathbf{K}}_s^{(j)}$, where R is large enough to drive the estimation error well below the strength of $\mathbf{\Omega} \circ \mathbf{K}_s^{(j)}$; in practice, this means using R larger than 10^6. Then, $\mathbf{\Omega}$ can be estimated as

$$\hat{\mathbf{\Omega}} = \frac{\sum_{i=1}^{R} \hat{\mathbf{K}}_s^{(j)} \mathbf{K}_s^{(j)}}{R \cdot \sigma_{k,s}^2} \tag{5}$$

The estimation of the distribution of $\mathbf{N}_k$ is simpler: thanks to the dominance of $\mathbf{N}_k$ in $\hat{\mathbf{K}}$, $\gamma^2[k,l]$ is estimated as the local variance of $\hat{\mathbf{K}}$ within a square window centered at (k,l). With this, the mutual information becomes

$$I(\mathbf{N}_k, \hat{\mathbf{K}}) = \frac{1}{2} \sum_{l,k} \log_2 \left(1 + \frac{\gamma^2[k,l]}{\sigma_k^2 \cdot \omega^2[k,l]}\right) \tag{6}$$

which requires knowing σ_k^2, that is, the variance of the true PRNU. Note that this is a parameter intrinsic of the camera. We propose next a way to estimate it. The expected value of P defined above is $\mathrm{E}\{P\} = \sigma_k^2 \langle \mathbf{\Omega}_1, \mathbf{\Omega}_2 \rangle_F$, where $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are the respective gains for $\hat{\mathbf{K}}_1$ and $\hat{\mathbf{K}}_2$ in model (2). Both $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ can be estimated following the procedure of embedding synthetic PRNUs described above. Then, we construct the estimate $\hat{\sigma}_k^2 = \langle \hat{\mathbf{K}}_1, \hat{\mathbf{K}}_2 \rangle_F / \langle \hat{\mathbf{\Omega}}_1, \hat{\mathbf{\Omega}}_2 \rangle_F$.

The expression in (6) can be approximated by noticing that the standard deviation in $\mathbf{\Omega}$ is small relative to its mean, so one can substitute $\omega[k,l]$ by $\bar{\omega} \doteq \sum_{k,l} \omega[k,l]/(M \cdot N)$, that is,

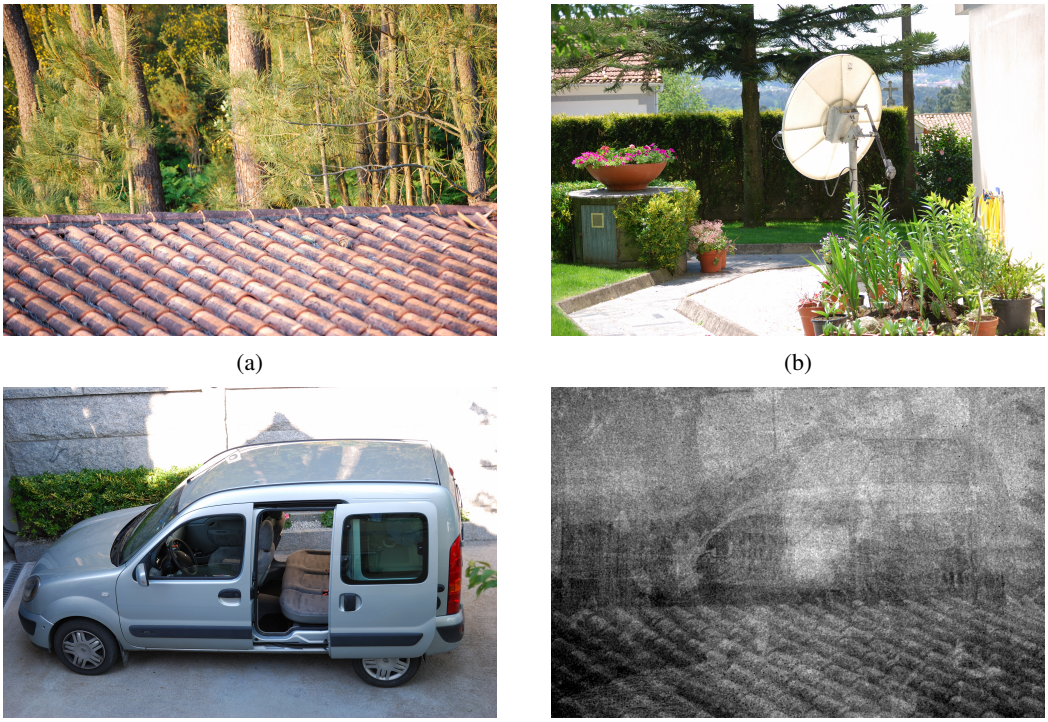
$$I(\mathbf{N}_k, \hat{\mathbf{K}}) \approx \frac{1}{2} \sum_{l,k} \log_2 \left(1 + \frac{\gamma^2[k,l]}{\sigma_k^2 \cdot \bar{\omega}^2}\right) \tag{7}$$

Similarly, we can produce a further approximation by replacing $\gamma[k,l]$ above by its mean value $\bar{\gamma} \doteq \sum_{k,l} \gamma[k,l]/(M \cdot N)$. This yields

$$I(\mathbf{N}_k, \hat{\mathbf{K}}) \approx \frac{M \cdot N}{2} \log_2 \left(1 + \frac{\bar{\gamma}^2}{\sigma_k^2 \cdot \bar{\omega}^2}\right) \tag{8}$$

B. Mutual information for the binarized PRNU

As it is obvious, binarization will render a mutual information $I(\mathbf{N}_k, b(\hat{\mathbf{K}}))$ of at most one bit per pixel, and due to the data processing inequality will be upperbounded by $I(\mathbf{N}_k, \hat{\mathbf{K}})$. To derive the mutual information for the binarized PRNU, notice first that, conditioned on $\omega^2[k,l]$ and $\gamma^2[k,l]$



(a) (b) (c) (d)

Fig. 1: Several images taken with the NikonD60 camera, and local variance of the corresponding estimated PRNU.

at the (k,l) position, $\hat{K}[k,l]$ will be zero-mean Gaussian distributed; therefore $\text{sgn}(\hat{K}[k,l])$ will take probability $1/2$ at both $-1,+1$, and then the entropy $h(b(K[\hat{k},l])) = 1$ bit for all k,l. On the other hand, conditioned on $N_k[k,l] = n_k$, $\text{sgn}(\hat{K}[k,l])$ will take the value -1 (resp. $+1$) with probability $1-q$ (resp. q) where $q = Q(-n_k/(\sigma_k\omega[k,l]))$ and $Q(x) \doteq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$.

Therefore, if $H_b(q)$ denotes the binary entropy of a Bernoulli process with probability q, the conditional entropy can be computed through the following integral (the indices $[k,l]$ are omitted for compactness):

$$h(b(\hat{K})|N_k) = \frac{1}{\sqrt{2\pi\gamma^2}}\int_{-\infty}^{\infty} H_b\left(Q\left(\frac{t}{\sigma_k\omega}\right)\right)e^{-\frac{t^2}{2\gamma^2}}dt \quad (9)$$

and then, the sought mutual information is

$$I(\mathbf{N}_k, b(\hat{\mathbf{K}})) = M\cdot N - \sum_{l,k} h(b(\hat{K}[k,l])|N_k[k,l]) \quad (10)$$

C. Mutual information for the normalized PRNU

So far, we have assumed that the matrix $\mathbf{\Gamma}$ of standard deviations governing $\mathbf{N}_k$ is known and available to an adversary (through estimation). In other words, although it was not made explicit, we have been computing $I(\hat{\mathbf{K}}, \mathbf{N}_k|\mathbf{\Gamma})$ for a fixed realization of $\mathbf{\Gamma}$. We have done so by modeling $\mathbf{N}_k$ as $\mathbf{\Gamma}\circ\mathbf{S}$, where the elements of $\mathbf{S}$ are i.i.d. zero-mean, unit-variance Gaussian, and we have determined $I(\hat{\mathbf{K}}, \mathbf{S}|\mathbf{\Gamma})$. Since from practical considerations all the elements of $\mathbf{\Gamma}$ are strictly positive, it follows that Hadamard division of $\hat{\mathbf{K}}$ by $\mathbf{\Gamma}$ is invertible and thus it does not change the mutual information, i.e., $I(n(\hat{\mathbf{K}}), \mathbf{S}|\mathbf{\Gamma}) = I(\hat{\mathbf{K}}, \mathbf{S}|\mathbf{\Gamma})$. This can be readily checked in the formulas in Sect. III-A, where it is clear that dividing $\gamma^2[l,k]$ and $\omega^2[l,k]$ by the same quantity does not alter the mutual information. However, if we consider $\mathbf{\Gamma}$ as stochastic, then it is clear that there will also be leakage in $\hat{\mathbf{K}}$ from $\mathbf{\Gamma}$ and in turn from $\{\mathbf{Y}^{(i)}\}_{i=1}^L$ (see Fig. 1). This leakage, measured by $I(\hat{\mathbf{K}}, \mathbf{\Gamma})$, can be incorporated into $I(\hat{\mathbf{K}}, \mathbf{N}_k)$ thanks to the chain rule of mutual information as $I(\hat{\mathbf{K}}, \mathbf{N}_k) = I(\hat{\mathbf{K}}, \mathbf{S}|\mathbf{\Gamma}) + I(\hat{\mathbf{K}}, \mathbf{\Gamma})$. Unfortunately, the actual computation of $I(\hat{\mathbf{K}}, \mathbf{\Gamma})$ is very difficult as the distribution of $\mathbf{\Gamma}$ does not admit a simple modeling. Finding a lower bound for this term has undeniable interest and is left for future research.

For our purposes, it suffices to justify that $I(n(\hat{\mathbf{K}}), \mathbf{\Gamma}) \approx 0$, that is, equalizing the PRNU practically prevents estimating $\mathbf{\Gamma}$: although Hadamard division of $\hat{\mathbf{K}}$ by $\mathbf{\Gamma}$ would in theory leave information about the latter in $(\mathbf{\Omega}\circ\mathbf{K})\oslash\mathbf{\Gamma}$, such term would be negligible w. r. t. the equalized $\mathbf{N}_k$, thus making it impractical for an adversary to learn $\mathbf{\Gamma}$.

D. Membership inference

In [3] we introduce a membership inference test (MIT) [6] to quantify the extent at which it is possible to identify the images used in the PRNU estimation. This is a binary hypothesis test that, given a PRNU estimate of a certain camera and an image taken by that camera, determines whether the image was used in the estimation. While more indirect than the measure provided by the mutual information, the inference is able to

capture some of the leakage that is harder to evaluate, such as $I(\hat{\mathbf{K}}, \mathbf{\Gamma})$, see previous section. The detection statistic is the normalized cross-correlation (NCC):

$$J_{\mathsf{NCC}} \doteq \frac{1}{MN-1} \sum_{l,j} \frac{(\hat{K}[l,j] - \hat{\mu}_k)}{\hat{\sigma}_k} \cdot \frac{(W^{(t)}[l,j] - \hat{\mu}_t)}{\hat{\sigma}_t} \quad (11)$$

where $\mathbf{W}^{(t)}$ denotes the residual corresponding to the image under test, and $\hat{\mu}_k$ and $\hat{\mu}_t$ and $\hat{\sigma}_k^2$ and $\hat{\sigma}_e^2$ are the sample means and variances of $\hat{\mathbf{K}}$ and $\mathbf{W}^{(t)}$, respectively. Here, the null hypothesis (i.e., H_0) corresponds to $\mathbf{W}^{(t)}$ not having been used in the estimation of $\mathbf{K}$, and the alternative (i.e., H_1) otherwise.

In [3] the performance of the MIT for different values of L and cameras is measured by the corresponding ROC curves. Whereas ROC curves provide more complete information, they make hard a direct assessment of the impact of an increasing number of images L in the estimation. Albeit the distribution of the test statistic in (11) is not Gaussian for either of the hypotheses, we propose to utilize the following definition of Signal to Noise Ratio (SNR) in which we make explicit the dependence with L:

$$\mathrm{SNR}(L) \doteq \frac{\hat{\mu}_{H1}(L) - \hat{\mu}_{H0}(L) - Q^{-1}(P_{FP})\hat{\sigma}_{H0}(L)}{\hat{\sigma}_{H1}(L)} \quad (12)$$

where P_{FP} is the target false positive probability, $Q^{-1}(\cdot)$ is the inverse of the Q-function introduced above, and $\hat{\mu}_{H0}(L)$ and $\hat{\sigma}_{H0}(L)$ (resp. $\hat{\mu}_{H1}(L)$ and $\hat{\sigma}_{H1}(L)$) are the sample mean and standard deviation of the NCC for the null (resp. the alternative) hypothesis.

IV. Experiments

We have conducted experiments to quantify the information-theoretic leakage and the membership inference performance on a database of uncompressed images taken with 9 commercially available cameras listed in Table I. The number of images per camera ranges from 122 (Canon1100D#2) to 316 (Canon1100D#1).

In our first set of experiments we aim at comparing the different bounds and approximations to $I(\hat{\mathbf{K}}, \mathbf{N}_k|\mathbf{\Gamma})$ presented in Sect. III-A. In particular, we remind that in [3] we provide only the lower bound in (3), but here we have presented a synthetic approach that allows us to estimate the gains $\mathbf{\Omega}$ needed for the exact mutual information. The drawback of this synthetic approach is that it is very time consuming, as in practice $R = 2 \cdot 10^6$ in (5). Therefore, to speed up the experiments we have used for each camera a set of 25 randomly chosen images from which the 512×512 central patches were cropped out. To generate representative values of $\mathbf{\Gamma}$ we have produced 5 estimates obtained from the 512×512 central patches of L randomly selected images, where $L \in \{25, 50\}$. The window used to estimate the local variance has size 9×9 pixels. On the other hand, to estimate the total PRNU energy P, we have followed the procedure discussed in [3] on 50 randomly selected images for each camera.

Camera	(6)	(3)	(7)	(8)	(10)
NikonD60	1.661	1.657	1.651	1.666	0.755
Canon1100D#1	1.209	1.204	1.199	1.205	0.662
Canon1100D#2	1.798	1.797	1.794	1.804	0.781
Canon1100D#3	1.432	1.428	1.425	1.433	0.712
NikonD3000	1.469	1.452	1.448	1.458	0.719
NikonD3200	1.333	1.332	1.332	1.340	0.692
NikonD5100	1.961	1.96	1.959	1.969	0.802
Canon600D	0.824	0.822	0.822	0.827	0.535
NikonD7000	1.473	1.472	1.470	1.478	0.722

TABLE I: Mutual information bounds and approximations (in bits per pixel) for different cameras in the dataset. Last column corresponds to binarized PRNUs. $L = 25$.

ΔI	D60	1100#1	D300	D3200	D5100	600D	D7000
$\hat{\mathbf{K}}$	0.431	0.293	0.360	0.302	0.572	0.181	0.319
$b(\hat{\mathbf{K}})$	0.086	0.068	0.084	0.077	0.095	0.064	0.072

TABLE II: Decrease in mutual information for the standard (6) and binarized (10) PRNUs when L goes from 25 to 50.

From each of the distributions of $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ we draw 10^6 samples that are used to evaluate the sums in Eqs. (6-8) for a total of 10^{12} combinations. This is repeated for the 5 instances of $\mathbf{\Gamma}$ and the results are averaged. For the case of the binarized PRNU, since the integral must be evaluated numerically, we have reduced the number of total samples of $(\mathbf{\Omega}, \mathbf{\Gamma})$ to 10^6 in order to make the estimate computationally tractable.

We present the results for $L = 25$ in Table I. The small differences that can be observed with respect to the bound reported in [3] are mainly due to dependences between P (or $\mathbf{\Omega}$) and $\mathbf{\Gamma}$ that are indirectly discarded here by using a fixed set of 25 images to estimate $\mathbf{\Omega}$. The first remarkable conclusion from the reported results, besides validating that (3) is a lower bound to the leakage, is that both the bound and the approximations are very tight. Therefore, for future evaluations of the leakage it will not be necessary to resort to the very expensive estimation of $\mathbf{\Omega}$, and even a very simple expression as (8) will be sufficient in most cases. The second conclusion is that PRNU binarization lowers the leakage by more than 50% in most cases; in addition to the reduction that is expected in $I(\hat{\mathbf{K}}, \mathbf{\Gamma})$. Table II, second row, reflects for 7 of the cameras in the dataset the reduction in the leakage when L is increased to 50. Notice that in the case where $\bar{\gamma}^2/(\sigma_k^2 \cdot \bar{\omega}^2) \gg 1$ one should expect a reduction of 0.5 bpp when doubling L, but this rarely happens in practice. In contrast, the leakage reduction from the binarized PRNU (third row of the table) is much smaller, thus reflecting that binarization is worthwhile only when the PRNU is estimated from few images.

In our second set of experiments, concerned with the membership inference test (MIT), we have plotted (Fig. 2) the SNR as defined in (12) for two of the cameras in the dataset (Nikon D7000 and Nikon D60) and different numbers L of images used in the PRNU estimation, to show how the SNR approximately decreases linearly with L for the standard, binarized and equalized PRNUs. It is interesting to observe that while the binarized PRNU gives only marginal improvements, equalizing the PRNU is

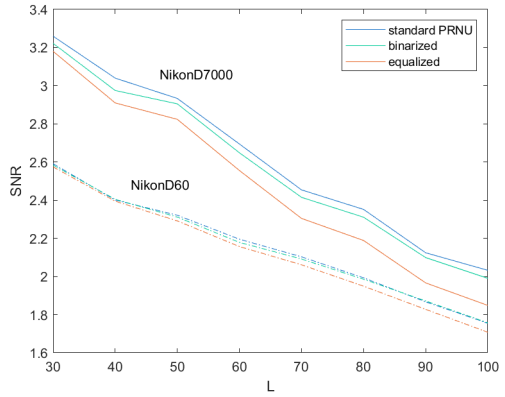


Fig. 2: Detection statistics for the two detectors on a set of 250 images (Nikon D7000 camera). First 50 images correspond to those used in the PRNU estimation.

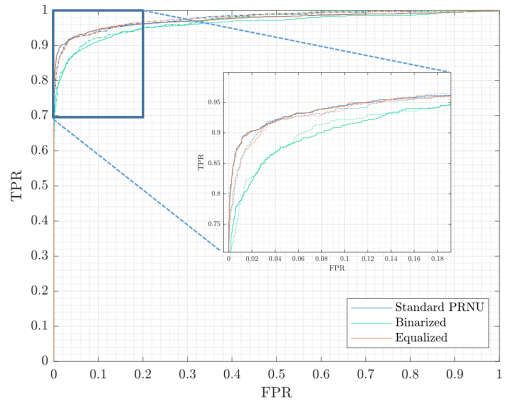


Fig. 3: Receiver operating characteristic for aligned PRNU detection. Solid lines: Nikon D7000; dashed lines: Nikon D60.

equivalent to effectively reducing L by about 10% in the case of the Nikon D7000, where it is most needed. In fact, in Fig. 2 we have not shown the SNRs for the Canon D600 camera which are practically flat with L and below 1, indicating that even for small values of L the MIT does not succeed. This is consistent with the results in Table I which show that for this camera the leakage is the smallest. In this case, binarizing or equalizing the PRNU do not provide any gains.

Fig. 3 represents the ROC for the Nikon D7000 and Nikon D60 cameras in a PRNU detection scenario, where we have cropped out 256×256-pixel patches with random centers in order to put the detectors under a severe scenario. L is set to 50. Correct alignment with respect to the PRNU is assumed for the 200 images corresponding to the H_1 hypothesis; 50 images from each of 15 cameras are considered for the H_0 hypothesis. Probabilities are averaged from 5 runs. As we can see, while binarization reduces the performance in a detection scenario, equalization does not, despite reducing the information leakage. Also notice that even though the D7000 and D60 cameras perform quite differently in terms of the MIT (cf. Fig. 2) this does not translate into significant differences in detection performance.

V. Conclusions

In this paper we complement our results in [3] to provide a more accurate quantification of the leakage from PRNU estimates, together with some simpler but tight approximations. This was made possible by pursuing a novel approach in which we embed a synthetic PRNU in the actual images that is used as a pilot that allows us to estimate some unknown parameters. Another contribution is a simpler way to characterize the performance of membership inference tests that proves to be valuable in estimating the test behavior under changes in the number of images used for PRNU estimation.

Finally, we have proposed two countemeasures for mitigating the leakage. One of them, equalization, reduces the leakage without apparent cost in terms of PRNU detection performance degradation. This success opens the way to other more sophisticated approaches that take into account the contextual leakage $I(\hat{\mathbf{K}}, \mathbf{\Gamma})$ which plays a significant role in the total leakage, as our examples show. In the meantime, we believe that the best practice against leakages is to work with encrypted data at all times, as done in [9], [10], but there may exist hybrid solutions with leakage mitigation that are worth exploring in the future.

References

[1] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor noise," IEEE Trans. Inf. Forensics Security, vol. 3, no. 1, pp. 74–90, Mar. 2008.

[2] P. Korus and J. Huang, "Multi-scale analysis strategies in PRNU-based tampering localization," IEEE Trans. Inf. Forensics Security, vol. 12, no. 4, pp. 809–824, Apr. 2017.

[3] S. Fernandez-Meduiña, F. Pérez-González, "On the Information Leakage of Camera Figerprint Estimates", 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 2020, Submitted. Available at `http://arxiv.org/abs/2002.11162`.

[4] M. Goljan, J. Fridrich, M. Chen, "Defending Against Fingerprint-Copy Attack in Sensor-Based Camera Identification," IEEE Transactions on Information Forensics and Security, vol. 6. pp. 227 - 236, 2011.

[5] M. Masciopinto and F. Pérez-González, "Putting the PRNU Model in Reverse Gear: Findings with Synthetic Signals," 26th European Signal Processing Conference (EUSIPCO), Rome, pp. 1352-1356, 2018.

[6] R. Shokri, M. Stronati, C. Song and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, 2017, pp. 3-18.

[7] M.K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially Adaptive Statistical Modeling of Wavelet Image Coefficients and its Application to Denoising." Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Phoenix, AZ, vol. 6, pp. 3253–3256, March 1999

[8] S. Bayram, H. T. Sencar, and N. Memon, "Efficient sensor fingerprint matching through fingerprint binarization," IEEE Trans. Inf. Forensics Security, vol. 7, no. 4, pp. 1404–1413, Aug. 2012.

[9] M. Mohanty, M. Zhang, M. R. Asghar and G. Russello, "PANDORA: Preserving Privacy in PRNU-Based Source Camera Attribution," 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, 2018, pp. 1202-1207.

[10] A. Pedrouzo-Ulloa, M. Masciopinto, J.R. Troncoso-Pastoriza, F. Pérez-González, "Camera Attribution Forensic Analyzer in the Encrypted Domain," Proc. IEEE International Workshop on Information Forensics and Security (WIFS), Hong-Kong, 2018, pp. 1-7.