RESEARCH

On the Information Leakage Quantification of Camera Fingerprint Estimates

Samuel Fernández-Menduiña[1] and Fernando Pérez-González[2]

Abstract

Camera fingerprints based on sensor PhotoResponse Non-Uniformity (PRNU) have gained broad popularity in forensic applications due to their ability to univocally identify the camera that captured a certain image. The fingerprint of a given sensor is extracted through some estimation method that requires a few images known to be taken with such sensor. In this paper, we show that the fingerprints extracted in this way leak a considerable amount of information from those images used in the estimation, thus constituting a potential threat to privacy. We propose to quantify the leakage via two measures: one based on the Mutual Information, and another based on the output of a membership inference test. Experiments with practical fingerprint estimators on a real-world image dataset confirm the validity of our measures and highlight the seriousness of the leakage and the importance of implementing techniques to mitigate it. Some of these techniques are presented and briefly discussed.

Keywords: Fingerprint; PRNU; Leakage; Information Theory; Membership Inference.

1 Introduction

The PhotoResponse Non-Uniformity (PRNU) is a multiplicative spatial pattern that is present in every picture taken with a CCD/CMOS imaging device and acts as a unique fingerprint for the sensor itself [1]. The PRNU is due to manufacturing imperfections that cause sensor elements to have minute area differences and thus capture different amounts of energy even under a perfectly uniform light field. The uniqueness of the PRNU has already led to a number of applications in multimedia forensics, both to solve camera identification/attribution problems using images [2] or stabilized videos [3], and to detect inconsistencies that reflect intentional manipulations [4].

Since the PRNU is a very weak signal, its extraction requires the availability of a number (often dozens) of images known to be taken with the camera under analysis. Although several extraction algorithms (both model- and data-driven) exist [1], [5], all of them perform some sort of averaging across the residuals obtained by denoising the available images. The most prevalent method [1] performs a further normalization to take into account the multiplicative nature of the PRNU.

Unfortunately, both the ease with which the PRNU can be extracted and the existence of relatively good theoretical models that explain its contribution lead to attacks that are similar in intention to *digital forgery attacks* in cryptography: the so-called *PRNU copy attack* plants the fingerprint from a desired camera in an image taken by a different device with the purpose of incriminating someone or merely undermining the credibility of PRNU-based forensics [6].

While the PRNU copy attack can be considered a threat to *trust*, in this paper we identify risks to *privacy* by showing that there is substantial information leakage into the PRNU from the images used for its estimation. The existence of this leakage has been already indirectly exploited in the so-called

Correspondence: fperez@gts.uvigo.es.
Signal Processing in Communications Group, Atlanttic Research Center, Campus Universitario Lagoas, 36310 Vigo, Spain.

triangle test [7], which is a countermeasure against the copy attack that in order to detect the forgery relies on the high correlation between the PRNU estimate with any of the image residuals used in the estimation. However, to the best of our knowledge, our work, together with its companion paper [8], constitutes the first attempt at quantifying such leakage by proposing two measures: one based on the mutual information, and another based on the success rate of a membership inference test.

To this end, we provide a detailed derivation of a lower bound for the Mutual Information between a given image and the PRNU, as well as two membership inference tests based on the Neyman-Pearson criterion and the normalized correlation coefficient, respectively. Although we do not explicitly try to recover traces of the images used to extract the PRNU, we show that the leakage is large enough to consider the possibility of recovery a serious threat. In this sense, we remark that images involved in criminal investigations are often of extremely sensitive nature, like in cases involving child abuse and other sexually-oriented crimes, so the mere existence of this leakage calls for the implementation of effective protection mechanisms of the camera fingerprints that ensure privacy is preserved at all times during investigations.

While in an ideal scenario the PRNU of a device can be extracted from flat-field images (e.g., of a cloudy sky or a white wall) in practice this is only feasible when there is access to the camera under investigation. In this scenario, where the estimated PRNU practically leaks little information (as trivially shown by our theory), different law enforcement agencies (LEAs) may share the estimated fingerprints for cross-searching in databases with no privacy risks. However, there is a growing number of investigations where no access to the device is feasible and the PRNU must be estimated from images "in the wild". Cases include images retrieved from hard drives, social networks, and criminal networks in the Dark web. As an example, we discuss the following two cases.

Case 1: During the course of an investigation, police from country A (LEA A) have seized a hard drive containing images from unknown sources involving child abuse. As metadata has been wiped off, LEA A uses some PRNU clustering software to find that the images come from three different cameras, for which the corresponding PRNUs can be extracted. After analyzing the contents of one of the clusters, it is found that some of the pictures taken by camera #1 have been shot in country B. LEA A would like to verify if the police of country B (LEA B) have other images from camera #1 or even the device. Exchanging the highly-sensitive pictures with LEA B is dismissed for privacy reasons; alternatively, LEA A sends the estimated PRNU on the belief that it entails no privacy infringement. This is rooted in the fact that law enforcement agencies are accustomed to sharing hashes in order to search for cross-matches in databases with images of child exploitation. However, as our work shows, contrary to robust hashes, PRNUs may leak considerable amounts of information that should be treated as private as it may identify the victims.

Case 2: Members of a gang have been exchanging pictures over the Dark Web. Some of them involving the gang leader (and third persons) have been taken by the same camera (itself unavailable), as confirmed by the PRNU. The police would be interested in crawling the social networks in search of other pictures captured by the same device. Due to their very limited computational resources, and convinced that nothing can be inferred from an estimated PRNU, the police outsource the search to a web crawling company. However, the leakage from the PRNU allows the company to infer information about people, places and objects contained in the images acquired by the police. In particular, from the PRNU it is possible to read a car license plate.

As our paper concludes, sharing of PRNU fingerprints should be done only after carefully assessing the risks and considering all the possible remedies, some of which are evaluated and discussed in this paper.

As already pointed out and formalized in [8], existing techniques in the literature can mitigate the contextual residues of images on the PRNU. Examples are: 1) compression schemes and binarization

[9–12], which are originally conceived to reduce the computational burden in the estimation process and limit the required storage of the resulting fingerprint; 2) the application of linear filters, as high pass filters (both fixed [13–15] and trainable [16]) and convolutional neural networks for feature extraction [17], which were found to be useful to enforce neural nets to work with noise residuals [5] in both forgery detection [13, 18] and camera attribution [19], and 3) the use of more powerful denoising schemes than the wavelet denoiser. In the present paper, we take a step further in this direction, analyzing empirically the effects of JPEG compression and the use of more powerful denoising schemes, as BM3D [20]. Despite the relative effectiveness of those solutions, we believe that working with encrypted data at all times [21], although yet not entirely practical due to the large amount of computations needed, is the most promising venue in terms of privacy preservation.

Our main contributions in this paper can be summarized as follows.

- We derive a model for the fingerprint estimator in terms of the true PRNU and the estimation noise. This model becomes crucial in our two approaches to quantifying the leakage, and is also assumed (but not derived) in [8].
- We take a step to model and bound the information leakage in camera fingerprints as the PRNU, based on a waterfilling information theoretic approach.
- We propose a membership inference test, which allows to identify the images in a dataset that were used to estimate a given PRNU.
- We propose and test empirically some methods to reduce the leakage in practice.
- We confirm that information leakage is a serious privacy threat that should be properly assessed before sharing camera fingerprints.
- We show that the discovered leakage could be potentially used to detect PRNU copy attacks without resorting to the original images (as is done in the triangle test), since the extracted PRNU will have an underlying structure that will not match that of the host image.

The rest of the paper is organized as follows: in Sect. 2 we review the basic principles of PRNU extraction; in Sect. 3 we propose two metrics to quantify the leakage; Sect. 4 hints at the potential of our discovery to counter injection-based attacks; Sect. 5 briefly discusses several approaches to mitigate the leakage; Sect. 6 contains the results of experiments carried on images taken with popular cameras, and, finally, Sect. 7 presents our conclusions.

1.1 Notation

Matrices, written in boldface font, represent luminance images. All are assumed to be of size $M \times N$. The pixel in position (m, n) of image $\mathbf{X}$ is referred to as $X[m, n]$. Given two matrices, $\mathbf{X}$ and $\mathbf{Y}$, its Hadamard product $\mathbf{Z} = \mathbf{X} \circ \mathbf{Y}$ is such that $Z[m, n] = X[m, n] \cdot Y[m, n]$, for all $m = 1, \dots, M$ and $n = 1, \dots, N$. The Frobenius cross-product of $\mathbf{X}$ and $\mathbf{Y}$ is defined as $\langle \mathbf{X}, \mathbf{Y} \rangle_F \doteq \mathrm{tr}\left(\mathbf{X}^\top \mathbf{Y}\right)$, where $\mathrm{tr}(\cdot)$ denotes trace and T transpose. The all-one matrix is denoted by $\mathbf{1}$. Random variables are written in capital letters, e.g., X, while realizations are in lowercase, e.g., x. Given two random variables X, Y, $X \to Y$ means that X converges to Y in probability.

2 Preliminaries

In this paper, we will use the prevalent simplified sensor output model presented in [1] in matrix form:

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

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.

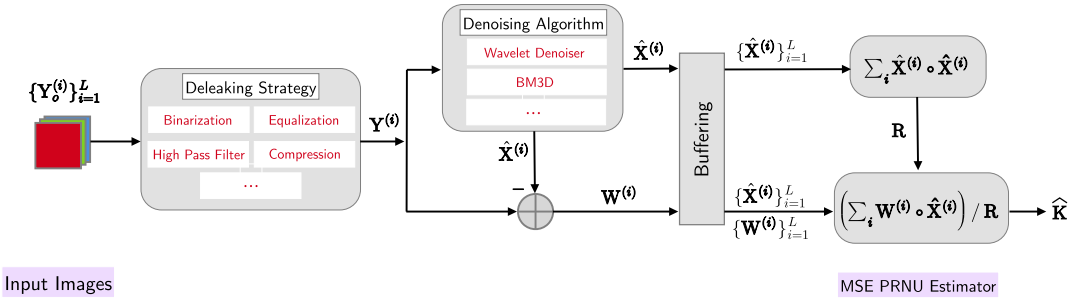


Figure 1 Block diagram of the ML PRNU estimation process from a set of L images $\{\mathbf{Y}_o^{(i)}\}_{i=1}^L$ considered in this paper, jointly with the different variables involved in the process. For each block, all the possible operations are highlighted in red. The deleaking strategy block may not be used in some experiments.

This PRNU term can be estimated from a set of L images $\{\mathbf{Y}^{(i)}\}_{i=1}^L$ coming from the same sensor, as shown in Fig. 1 (no deleaking strategy is used in the conventional estimator). Firstly, the noise-free image $\mathbf{X}^{(i)}$ is estimated using a denoising filter,[1] and this estimate $\hat{\mathbf{X}}^{(i)}$ is used to obtain a residual $\mathbf{W}^{(i)} \doteq \mathbf{Y}^{(i)} - \hat{\mathbf{X}}^{(i)}$. Under the assumption of $\mathbf{N}^{(i)}$ being composed by i.i.d. samples of a Gaussian process, the Maximum Likelihood (ML) estimator of $\mathbf{K}$ reduces to:

$$\hat{\mathbf{K}} = \left(\sum_{i=1}^{L} \mathbf{W}^{(i)} \circ \hat{\mathbf{X}}^{(i)} \right) / \mathbf{R}, \tag{2}$$

where $\mathbf{R} \doteq \sum_{i=1}^{L} \hat{\mathbf{X}}^{(i)} \circ \hat{\mathbf{X}}^{(i)}$, and the division is point-wise. Often, the result of this estimation contains non-unique traces left by color interpolation, compression or other systematic errors, that are removed by post-processing (e.g., zero-meaning and Wiener filtering in the full-DFT domain). Ideally, this PRNU will be a zero-mean white Gaussian process with variance σ_k^2, independent of the location within the matrix.

Unfortunately, the denoising process will not perform perfectly. In fact, the denoised image can be more accurately modeled as:

$$\hat{\mathbf{X}}^{(i)} = \left(\mathbf{X}^{(i)} - \mathbf{\Delta}^{(i)} \right) + \left(\mathbf{1} - \mathbf{\Omega}^{(i)} \right) \circ \mathbf{K} \circ \mathbf{X}^{(i)}, \tag{3}$$

where $\mathbf{\Delta}^{(i)}$ takes into account the traces of the noise-free image that are left out by the denoising and $\left(\mathbf{1} - \mathbf{\Omega}^{(i)} \right)$ models the fraction of the PRNU-dependent component that passes through the denoiser. Then, when subtracted to $\mathbf{Y}^{(i)}$ and applied to the estimator, we have:

$$\hat{\mathbf{K}} = \frac{\sum_{i=1}^{L} \left(\mathbf{\Omega}^{(i)} \circ \mathbf{K} \circ \mathbf{X}^{(i)} + \mathbf{\Delta}^{(i)} + \mathbf{N}^{(i)} \right) \circ \hat{\mathbf{X}}^{(i)}}{\mathbf{R}}. \tag{4}$$

Then, it is easy to show that (4) can be expressed as

$$\hat{\mathbf{K}} = \mathbf{\Omega} \circ \mathbf{K} + \mathbf{N}_k, \tag{5}$$

where $\mathbf{\Omega} \doteq \left(\sum_{i=1}^{L} \mathbf{\Omega}^{(i)} \circ \hat{\mathbf{X}}^{(i)} \circ \mathbf{X}^{(i)} \right) / \mathbf{R}$ is a function of the used images, which takes into account the amount of PRNU removed in the denoising process, and $\mathbf{N}_k$ is estimation noise that depends on

[1] In most of the experiments carried out in this paper, we have used the popular wavelet-based denoiser presented in [22]. Denoising always includes zero-meaning and Wiener filtering in the full-DFT domain, following the approach in [1].

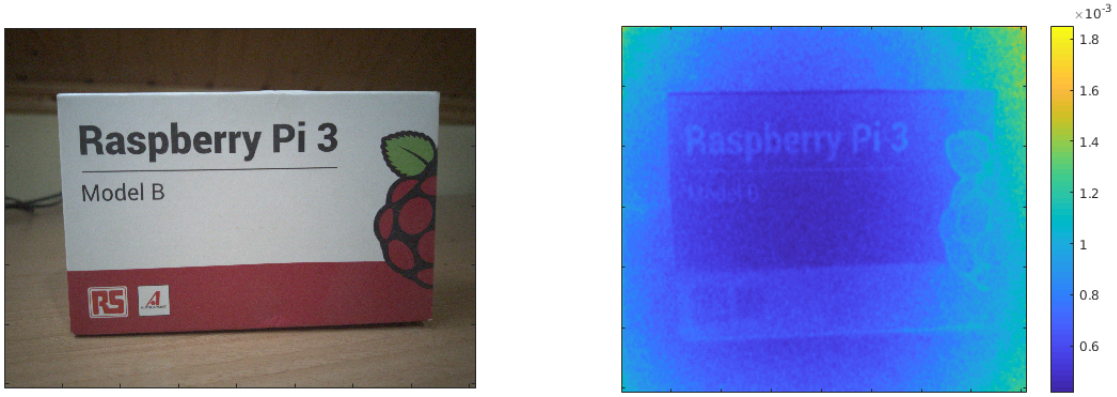


Figure 2 An example of the PRNU leakage problem, where both text and the shape of the elements in the image are preserved in the estimated PRNU. (Left) Sample image containing textual and graphical information; (Right) PRNU extracted from 24 dark images and the image in the left, all coming from the same camera.

both $\{\mathbf{\Delta}^{(i)} \circ \hat{\mathbf{X}}^{(i)}\}_{i=1}^{L}$ and $\{\mathbf{N}^{(i)} \circ \hat{\mathbf{X}}^{(i)}\}_{i=1}^{L}$, which in turn convey contextual information about the images. Experiments reported in [23] show that $\mathbf{N}_k$ can be well-modeled by an independent Gaussian process with variance at the (k,l)th position denoted by $\gamma^2[k,l]$.

Fig. 2 illustrates a rather extreme case of leakage in which the PRNU of a Xiaomi MI5S smartphone camera is estimated from 25 DNG (uncompressed) images: the one on the left panel plus 24 additional dark images. As becomes evident, there is a lot of information leaking from the first image into the estimated PRNU. Although by no means this experiment describes a realistic case, it does expose that such alarming leaks may well occur in smaller areas of the image. A more down-to-earth example is shown in Fig. 3, where the PRNU has been estimated with $L = 25$ images taken with a Nikon D3200 camera (see description of the database in the experimental part), and it visibly contains traces (with semantic meaning) of four images shown in the upper part which were used in the estimation. The bottom panels represent $\log(1+1/\gamma^2[l,k])$, when the local variance $\gamma^2[l,k]$ of $\hat{\mathbf{K}}$ is estimated through a 9×9 window. The division by $\gamma^2[l,k]$ has the purpose of emphasizing the areas with low local variance whereas the logarithm simply enhances the contrast for visualization purposes. Notice that despite the use of the more sophisticated denoising algorithm BM3D [20] (bottom-right panel) as compared to the wavelet-based denoising [22] (bottom-left panel), the leakage is still very conspicuous.

A more systematic approach to quantifying those leaks is presented in the next section.

3 Quantifying the leakage

In this section we discuss the two proposed measures to quantify the leakage into the PRNU estimate of the images used for the estimation.

3.1 Information-theoretic Leakage

The first measure is based on the Mutual Information of the set of images used for the estimation $\{\mathbf{Y}^{(i)}\}_{i=1}^{L}$ and the estimated PRNU $\hat{\mathbf{K}}$, i.e., $I(\{\mathbf{Y}^{(i)}\}_{i=1}^{L}, \hat{\mathbf{K}})$. Since $\mathbf{N}_k$ is a function of $\{\mathbf{Y}^{(i)}\}_{i=1}^{L}$, we can resort to the data processing inequality to show that $I(\{\mathbf{Y}^{(i)}\}_{i=1}^{L}, \hat{\mathbf{K}}) \geq I(\mathbf{N}_k, \hat{\mathbf{K}})$. The right hand side is considerably simpler to manage and produces a lower bound on the leakage.

The main difficulty for the calculation of $I(\mathbf{N}_k, \hat{\mathbf{K}})$ is the lack of a complete statistical characterization for $\mathbf{\Omega}$. It has been proven by Ihara [24] that given a Gaussian process $\mathbf{X}$ with covariance $\mathbf{K}_x$ and a

Figure 3 Several images taken with the NikonD3200 camera from the dataset. Bottom panels: emphasized local variance of the corresponding estimated PRNU computed using a window of size 9×9, (left): extraction using the wavelet denoiser, (right): extraction using the BM3D denoiser.

noise process $\mathbf{Z}$ with covariance $\mathbf{K}_z$, then the mutual information of $\mathbf{X}$ and $\mathbf{X}+\mathbf{Z}$ is minimized when $\mathbf{Z}$ is Gaussian with covariance $\mathbf{K}_z$. Therefore, for a given covariance matrix of $\mathbf{\Omega} \circ \mathbf{K}$, assuming that such process is Gaussian-distributed with the same covariance will produce a lower bound on the mutual information. Now, since $\mathbf{K}$ is assumed to be white, its covariance matrix is $\sigma_k^2 \mathbf{I}_{MN \times MN}$. Hence, the covariance of $\mathbf{\Omega} \circ \mathbf{K}$ will be an $MN \times MN$ diagonal matrix with elements $\omega^2[k,l]\sigma_k^2$. Then, the lower-bounding scenario corresponds to $MN \times MN$ parallel channels, in which the 'desired' signal (i.e., $\mathbf{N}_k$) is transmitted on each subchannel with power $\gamma^2[l,k]$ and there is an additive Gaussian 'disturbance' (corresponding to $\mathbf{\Omega} \circ \mathbf{K}$) with power $\omega^2[k,l]\sigma_k^2$.

Unfortunately, determining $\omega^2[k,l]\sigma_k^2$ turns out to be a difficult problem because even for moderate L, the term $\mathbf{N}_k$ dominates $\mathbf{\Omega} \circ \mathbf{K}$ in (5). One might think of using flat-field images for this purpose, as in this case the contribution of $\mathbf{N}_k$ would be negligible sooner as L increases. However, this path is not advisable because with flat-field images the contribution of $\mathbf{\Omega}$ would be lost. Therefore, we must content

ourselves with estimating the trace of the covariance matrix of $\mathbf{\Omega} \circ \mathbf{K}$, given by $P \doteq \sigma_k^2 \sum_{l,j} \omega^2[l,j]$, and then use it to produce a further lower bound on the mutual information. The value P can be seen as the total disturbance power budget that can be split among the different parallel channels in order to minimize the mutual information. Notice that this represents a worst case because in practice $\sigma_k^2 \omega^2[l,j]$ will deviate at each position (k,l) from such power distribution and the actual leakage will be larger.

The mutual information in this case can be obtained through the use of Lagrange multipliers, which give the following lower bound in nats [25]:

$$I(\mathbf{N}_k, \hat{\mathbf{K}}) \geq \frac{1}{2} \sum_{l,j} \log \left(1 + \frac{2}{\sqrt{1 + 4/(\mu \cdot \gamma^2[l,j])} - 1} \right) \doteq I^-, \tag{6}$$

where μ is the solution to the equation

$$\frac{1}{2} \sum_{k,l} \gamma^2[l,j] (\sqrt{1 + 4/(\mu \cdot \gamma^2[l,j])} - 1) = P. \tag{7}$$

To estimate P, we propose to randomly split the set $\{\mathbf{Y}^{(i)}\}_{i=1}^L$ into two subsets and estimate $\mathbf{K}$ from each. Let $\hat{\mathbf{K}}_1$, $\hat{\mathbf{K}}_2$ be those estimates. Then, P can be estimated as $\hat{P} = \langle \hat{\mathbf{K}}_1, \hat{\mathbf{K}}_2 \rangle_F$. A better estimate can be obtained by repeating several times the splitting of $\{\mathbf{Y}^{(i)}\}_{i=1}^L$ and averaging the resulting values of $\hat{P}$.

In [8] we propose a procedure for the exact computation of the mutual information, based on injecting synthetic signals that serve as pilots for the estimation of $\mathbf{\Omega}$. Unfortunately, the fact discussed above that $\mathbf{N}_k$ dominates $\mathbf{\Omega} \circ \mathbf{K}$ requires synthesizing a huge number of signals which make the procedure rather impractical. However, through experiments reported in [8] we were able to show that the lower bound provided here is tight for real-world images, in the sense that it is very close to its true value and, as we have seen, its computation much more affordable. Thus, even though we cannot claim that the lower bound presented here is always a fine approximation to the the leakage, it is reasonable to employ it to draw conclusions, especially so when comparing scenarios in which only one subsystem or parameter is changed.

We remark here that the leakage that we have quantified through a lower bound corresponds to the complete set of images $\{\mathbf{Y}^{(i)}\}_{i=1}^L$ used for estimating $\hat{\mathbf{K}}$. This means that we are not quantifying the leakage of a specific image, say, $\mathbf{Y}^{(j)}$, $j \in \{1, \cdots, L\}$. Such problem, which is more difficult due to the remaining images acting as a sort of interference, will be the subject of a future work.

From the mutual information formulas above it is interesting to reason about the gain produced by increasing L, which is a possible mitigation strategy. Let us assume that for a certain $L = L_0$ the lower bound in (6) is I_0^- and is achieved when $\mu = \mu_0$ in (7). Now, suppose that we double L to $2L_0$; we are interested in learning by how much the lower bound decreases. First, note that if $\gamma_0^2[l,j]$ denotes the power in the (l,j)th subchannel for L_0, then one would expect that when L is doubled, such power is approximately halved, i.e., $\gamma^2[l,j] = \gamma_0^2[l,j]/2$. This is due to the fact that $\gamma^2[l,j]$ is the variance of the estimation noise $\mathbf{N}_k$, that is expected to go to zero as $1/L$. Now, for small $\gamma_0^2[l,j]$, for all l,j, Eq. (7) is approximately solved as

$$\mu_0 \approx \frac{\left(\sum_{l,j} \gamma_0[l,j] \right)^2}{\left(\frac{1}{2} \sum_{l,j} \gamma_0^2[l,j] + P \right)^2}, \tag{8}$$

and the lower bound in nats approximately becomes

$$I_0^- \approx \frac{1}{2}\sum_{l,j}\log\left(1+\sqrt{\mu_0}\cdot\gamma_0[l,j]\right). \tag{9}$$

If we assume that now $\gamma[l,j]=\gamma_0[l,j]/\sqrt{2}$ for all k,l, it is immediate to prove that the approximate solution μ to (7) satisfies $\mu_0/2\leq\mu\leq 2\mu_0$, where the lower bound is achieved when $P\to\infty$ and the upper bound when $P=0$. Plugging the current $\gamma[l,j]$ and μ into the approximation for the lower bound and taking into account that the logarithm is strictly increasing, we find that

$$\frac{1}{2}\sum_{l,j}\log\left(1+\sqrt{\mu_0}\cdot\frac{\gamma_0[l,k]}{2}\right)\leq\frac{1}{2}\sum_{l,j}\log\left(1+\sqrt{\mu}\cdot\gamma[l,k]\right)\leq\frac{1}{2}\sum_{l,j}\log\left(1+\sqrt{\mu_0}\cdot\gamma_0[l,k]\right). \tag{10}$$

For any $x>0$, from the monotonicity of the logarithm we can write $\log(1+x/2)\geq\log(1+x)-\log(2)$. Then, the decrease in the lower bound when $\gamma[l,j]=\gamma_0[l,j]/\sqrt{2}$, written as $\Delta I^-\doteq I_0^- - I^-$ in nats can be bounded as follows:

$$0\leq\Delta I^-\leq\frac{MN}{2}\log(2). \tag{11}$$

When this change is written in bits per pixel, we arrive at a simple interpretation: whenever L is doubled, the decrease in the leakage is at most 0.5 bits per pixel. As we will confirm in the experimental part, in practice the reduction is more modest, and more so as L keeps increasing (see Fig. 5).

3.2 Membership inference

In the PRNU scenario a membership inference test [26] is a binary hypothesis test that, given a PRNU estimate, classifies a certain image as having been used or not in the estimation. This inference is possible due to the aforementioned leakage: the higher the success rate in the membership inference test, the larger the leakage. It is important to note that the number L of images used in the estimation becomes a key parameter, since as L increases the information provided by the other images will dilute the individual contributions.

The potential recognition of the images used to estimate the PRNU allows any malicious attacker to obtain information about the input database, which may result in privacy risks in certain scenarios. As an example, knowing whether certain images were used to compute the PRNU may aid a convicted criminal in identifying the informant who handed them to law enforcement.

We derive two types of membership detectors: a Neyman-Pearson-based (NP) detector and a normalized-cross-correlation-based (NCC) detector. Even though the former is expected to perform better due to its statistical properties, along its derivation we will find that it requires information that is not readily available to a potential attacker. Therefore, assuming knowledge of such information leads to a 'genie-based' detector which is not practically realizable but is useful as it sets an upper bound on the achievable performance. In contrast, the NCC detector will behave (slightly) worse but is perfectly implementable.

Let $\mathbf{Y}^{(r)}$ be the image whose membership we want to test and which is known to contain the true PRNU $\mathbf{K}$. Note that the available observations to implement the test are $\hat{\mathbf{X}}^{(r)}$, $\mathbf{W}^{(r)}$ and $\hat{\mathbf{K}}$. Then, two hypotheses can be formulated:

$$\mathcal{H}_0 : \hat{\mathbf{K}} = \left(\sum_{i=1}^{L} \mathbf{W}^{(i)} \circ \hat{\mathbf{X}}^{(i)} \right) / \mathbf{R}, \tag{12}$$

$$\mathcal{H}_1 : \hat{\mathbf{K}} = \mathbf{Q} + \left(\sum_{i=1, i \neq r}^{L} \mathbf{W}^{(i)} \circ \hat{\mathbf{X}}^{(i)} \right) / \mathbf{R}, \tag{13}$$

where $\mathbf{Q} \doteq \left(\mathbf{W}^{(r)} \circ \hat{\mathbf{X}}^{(r)} \right) / \mathbf{R}$. The matrix $\hat{\mathbf{K}}$ can be modeled as having independent zero-mean Gaussian elements with variances at position (l, j) denoted by $\lambda_{l,j}^2$ under the hypothesis $\mathcal{H}_0$ and $\theta_{l,j}^2$ under the hypothesis $\mathcal{H}_1$.

Let $\mathbf{P} \doteq \hat{\mathbf{K}} - \mathbf{Q}$. Then, applying the Neyman-Pearson criterion [27], the following test is obtained:

$$J_{\mathsf{NP}} \doteq \sum_{l,j} \left(\log \left(\frac{\lambda_{l,j}}{\theta_{l,j}} \right) - \frac{(P[l,j])^2}{2\theta_{l,j}^2} + \frac{\left(\hat{K}[l,j] \right)^2}{2\lambda_{l,j}^2} \right) > \psi', \tag{14}$$

where ψ' is a threshold selected so that a certain probability of false alarm is attained.

In order to implement the test above, the variances $\lambda_{l,j}^2$ and $\theta_{l,j}^2$ are needed for all l, j. They can be computed as the respective local variances at each position of $\hat{\mathbf{K}}$ and $\mathbf{P}$. Unfortunately, $\mathbf{P}$ is only available through $\mathbf{Q}$ that in turn requires knowledge of $\mathbf{R}$. Since the latter will be in general unknown to an attacker, the NP detector must be considered only of theoretical interest.

When L is large enough, it is reasonable to assume that $\theta_{l,j}^2 \approx \lambda_{l,j}^2$, for all l, j. In such case, the test in (14) simplifies to:

$$\lim_{\mathbf{\Theta} \to \mathbf{\Lambda}} J_{\mathsf{NP}} = \sum_{l,j} \frac{\hat{K}[l,j] Q[l,j]}{\lambda_{l,j}^2} - \frac{(Q[l,j])^2}{2\lambda_{l,j}^2} > \psi'. \tag{15}$$

Notice from (14) that when $L \to \infty$, then $\mathbf{P} \to \hat{\mathbf{K}}$ and $\theta_{l,j}^2 \approx \lambda_{l,j}^2$, for all l, j since the information provided by an individual image is less significant. As a consequence, when $L \to \infty$ the membership test is equivalent to guessing the outcome of (fair) coin tossing.[2]

Assuming J_{NP} is Gaussian distributed under $\mathcal{H}_0$ with mean μ_J and variance σ_J^2, which is reasonable by invoking the Central Limit Theorem, we obtain the following expression for the probability of false alarm in terms of the threshold ψ',

$$P_{\mathsf{FA}} = \mathcal{Q}\left(\frac{\psi' - \mu_J}{\sigma_J} \right) \implies \psi' = \sigma_J \mathcal{Q}^{-1}\left(P_{\mathsf{FA}} \right) + \mu_J, \tag{16}$$

where $\mathcal{Q}(\cdot)$ represents the Q-function, i.e., $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$, and $\mathcal{Q}^{-1}(\cdot)$ its inverse function. Then, using the approximation for large L, we know that under $\mathcal{H}_0$ the mean value is given by

$$\mu_J = -\sum_{l,j} \frac{(Q[l,j])^2}{2\lambda_{l,j}^2}, \tag{17}$$

while, assuming uncorrelation between all pixels, the variance can be approximated by:

$$\sigma_J^2 \approx \sum_{l,j} \frac{(Q[l,j])^2}{\lambda_{i,j}^2}. \tag{18}$$

[2] This should be reflected in ROC curves as following the 'line-of-chance', cf. Sect. 3.2.

As a realizable alternative to the NP detector, it is possible to resort to the NCC of $\hat{\mathbf{K}}$ and $\mathbf{W}^{(r)}$, which has been already employed in camera attribution scenarios [28]. This approach relies on the availability of sample estimates of the respective means ($\hat{\mu}_k$ and $\hat{\mu}_t$) and variances ($\hat{\sigma}_k^2$ and $\hat{\sigma}_t^2$) of $\hat{\mathbf{K}}$ and $\mathbf{W}^{(r)}$. The resulting detection statistic becomes

$$J_{\mathsf{NCC}} \doteq \frac{1}{MN-1} \sum_{l,j} \frac{(\hat{K}[l,j]-\hat{\mu}_k)}{\hat{\sigma}_k} \cdot \frac{(W^{(r)}[l,j]-\hat{\mu}_t)}{\hat{\sigma}_t}. \tag{19}$$

4 Potential in detecting PRNU-copy attacks

One well-known countermeasure against PRNU-copy attacks is the triangle test that assumes the existence of a public set of images from which some have been used to extract the PRNU that is planted in the target image. The test looks for high correlations between the allegedly forged image and the images in the public set. An improved version, the *pooled triangle test* looks for high *joint cross-correlations* between the forged image and some subset of the public set.

The triangle test and more so the pooled one, find some difficulties to get them implemented in practice because the camera owner may lose track of her set of public images. However, the existence of leakage in the case of natural images shown here might be useful for detecting the existence of a planted PRNU, independently from the availability of a public set. Indeed, in the residual computed from the forged image, there will be traces of the planted PRNU with an underlying structure that does not match that of the forged image.

With mere illustrative purposes, we have taken the same PRNU shown in Fig. 3 bottom-left, and planted it in the image of Fig. 4(a). Then, we have computed the residual Fig. 4(b) which shows clear traces of the planted PRNU that obviously do not correspond to Fig. 4(a). For instance, the vehicle from Fig. 3 top-right is still visible in the area of the residual corresponding to the sky. The problem remains when images are JPEG-compressed, because even though the traces of the PRNU may dissipate with compression, the leakage in the estimated PRNU is harder to eliminate (see Sect. 6). This is illustrated in Fig. 4(c), where all intervening images (i.e., those used to extract the PRNU and the host image on which it is planted) are JPEG-compressed with QF=92.

A more systematic approach to exploiting leakage towards PRNU-copy detection is out of the scope of this paper. In any case, the fact that traces of the copied PRNU will be more easily found in flat regions of the target image suggests that a deep neural network trained with residuals coming from both pristine and forged images would be a feasible detector.

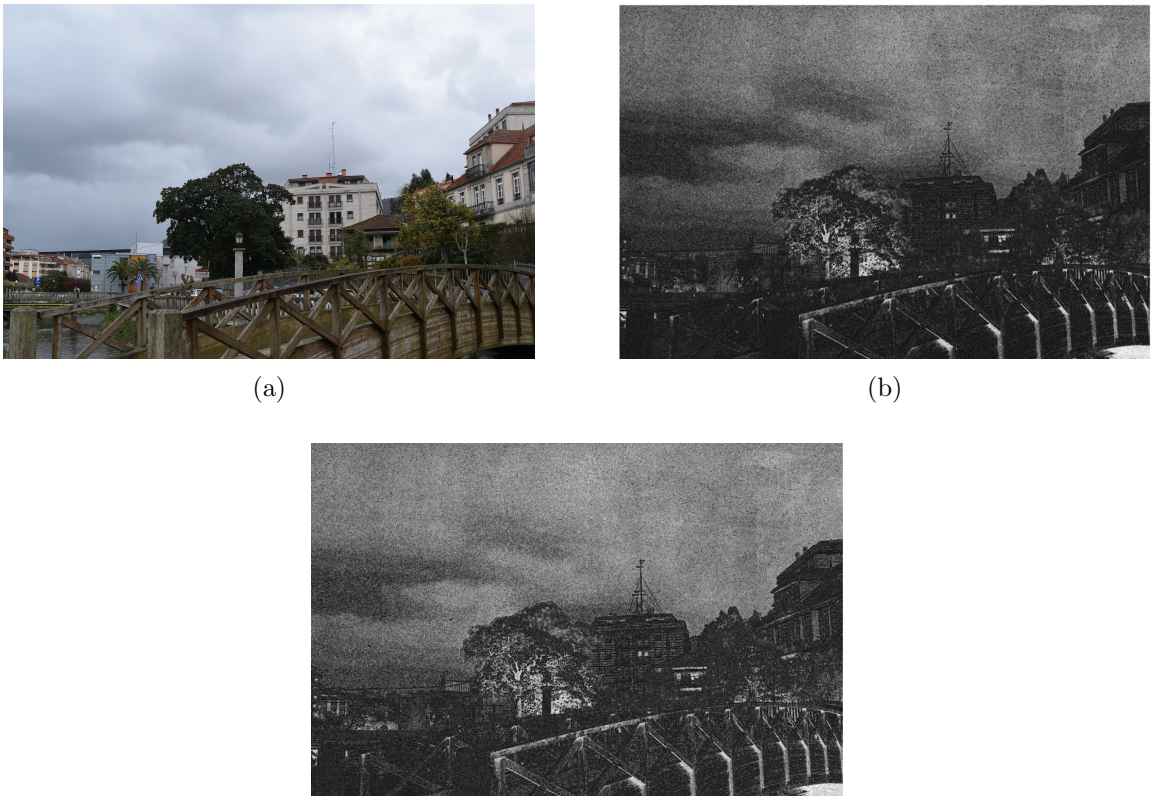
Finally, we remark that leakage mitigation techniques, to be discussed in the following Section should be able to reduce the probability of success of such a detector.

5 Leakage mitigation

Given the privacy risks that PRNU leakage entails, it is worth considering potential mitigation strategies, some of which are discussed here. We refer the reader to [8] for complementary details. We classify countermeasures in three categories: prevention, 'deleaking', and privacy preservation.

Preventive methods aim at conditioning the estimation process so that the resulting PRNU leaks less information. This can be achieved, for instance, by increasing the number of images L whenever possible (see discussion at the end of Sect. 3.1), maximizing the use of flat-field images, or improving denoising algorithms thus reducing $\mathbf{\Delta}^{(i)}$ and, consequently, the leakage, as shown in (4). In Sect. 6 we will present some experimental proof of the leakage reduction afforded by those approaches.

Deleaking methods consist in modifyng the estimated PRNU in a way that has limited loss in the PRNU detection performance, while decreasing the leakage. Examples of this are PRNU compression



(a) (b)

(c)

Figure 4 (a) Image from the database taken with the Nikon D3300 camera; (b) residual after planting the PRNU from Fig. 3 bottom-left, where traces of the car in Fig. 3 top-right are perfectly visible; (c) residual as in (b) where all images used for extracting the PRNU and the target image are JPEG-compressed with QF=92; the traces of the car are still conspicuous.

methods (e.g. [11]), but other possibilities exist, such as high-pass filtering in order to mitigate the pollution introduced by the contextual information of images [29] or whitening the estimated PRNU by normalizing by its local standard deviation (i.e., equalizing) at every spatial position. This PRNU equalization offers practically the same detection performance as using the conventional PRNU but consistently decreases the leakage. A detailed treatment of binarization and equalization as deleaking methods is carried out in [8] and, therefore, is not covered in this work.

Finally, another approach is to limit the exposure of the images and the PRNU in the clear using privacy-preserving techniques. This is possible by carrying out the PRNU estimation with encrypted images (and producing an encrypted PRNU) and detecting the encrypted PRNUs from encrypted query images [21]. This way, PRNU detection can be seen as a zero-knowledge proof mechanism. Although this is a very promising approach, substantial work is still needed to reduce the computational complexity of the underlying methods so that they become practical.

6 Experiments

6.1 Experimental setup and results

We have carried out experiments to validate our measures on a database of images, all in both TIFF and JPEG formats, taken with several commercially available cameras listed in Table 1. The number of images per camera ranges from 122 (Canon1100D#2) to 316 (Canon1100D#1). We discuss the results separately for the mutual information and the membership inference test.

Camera	ILB ($L = 26$)	ILB ($L = 50$)
NikonD60	1.6551	1.3458
Canon1100D#1	1.4007	1.1037
Canon1100D#2	1.7100	1.4092
Canon1100D#3	1.5962	1.2582
NikonD3000	1.4175	1.1147
NikonD3200	1.3827	1.0810
NikonD5100	1.9167	1.5768
Canon600D	0.8013	0.6791
NikonD7000	1.5246	1.2280
XiaomiMI5S	1.3916	1.1428

Table 1 Lower bound (6) in bits per pixel for different cameras and sizes of estimation sets when the wavelet-based denoising filter is employed. The lower bound oscillates for different camera models, ranging from 1.9167 bpp in the best case to 0.8013 bpp (for $L = 26$), which showcases the fact that some camera models may leak more than twice as much information than others when the wavelet denoiser is used.

6.2 Mutual information

In our first experiment, with TIFF images, we have computed the lower bound from (6) (heretofore denoted as Information Leakage Bound, ILB, and measured in bits per pixel, bpp) for two different values of L, namely $L = 26$ and $L = 50$. Denoising is carried out using the wavelet-based denoiser in [22]. The results, shown in Table 1, correspond to the average ILBs of 10 (resp. 5) runs of the experiment with randomly chosen subsets of size $L = 26$ (resp. $L = 50$).

The decreasing trend with L can be explained by the fact that the disturbance power budget P stays approximately constant, while the 'desired' signal $\mathbf{N}_k$ reduces its power with L. In fact, notice that, as $L \to \infty$ the term $\mathbf{N}_k$ is expected to go to zero due to the law of large numbers. The relatively small ILBs observed for the Canon 600D camera are conjectured to be due to the images in the respective dataset being very similar to each other.

Figure 5 (left) better illustrates the decrease of the leakage (as measured by the ILB) with L, as discussed at the end of Sect. 3.1. The plotted values correspond to the average ILBs of 5 runs of the experiment with randomly chosen subsets of size L. As discussed above, increasing L constitutes an advisable leakage mitigation mechanism that adds to the gains achieved in terms of detection performance. Notice, however, the diminishing returns with L: the leakage reduction from, say, doubling L is larger for smaller values of L. There is an important lesson here: as commercially available cameras increase their resolution, an ever smaller L is required to achieve a certain PRNU detection performance. While this fact is valuable from a practical point of view (often the number of available images in forensic cases is very small), it may be detrimental in terms of leakage, and additional measures may be required.

In order to quantify the impact of using flat-field images, in our next experiment we use DNG images taken with a the camera of a Xiaomi MI5S smartphone to build the following: sets `50brt` and `50drk` correspond to $L = 50$ images of respectively white and black cardboard, while in sets `49brt+berry` and `49drk+berry` one of the images is replaced by the one shown in Fig. 2(Left). The corresponding ILBs are given in Table 2.

As we discussed above in connection with the leakage mitigation, by comparing these values with those in Table 1 we can see that the usage of flat-field images tends to reduce leakage substantially. On the other hand, our dark images leak less information than the bright ones. Of course, this leakage does not correspond to perceptually meaningful information. Furthermore, while the inclusion of a non-flat

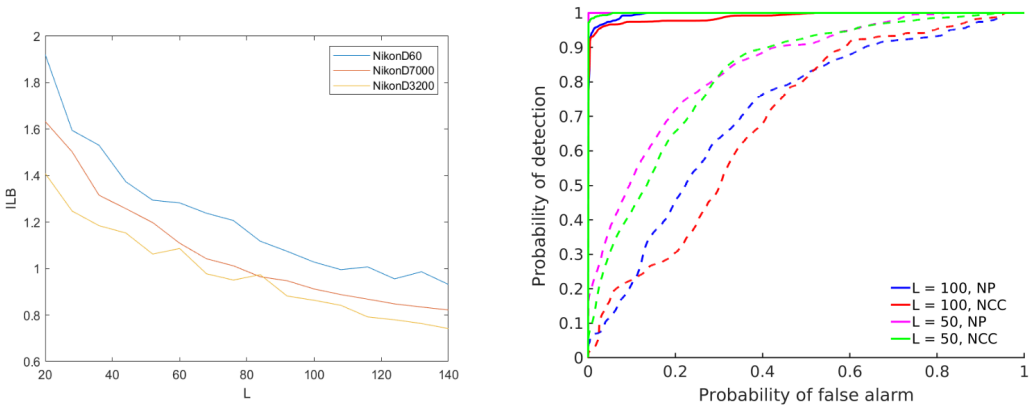


Figure 5 (Left) Information Leakage Bound (in bpp) vs L for three different cameras in the set, showing the expected decrease of the leakage with respect to L. (Right) Receiver operating characteristic for the NP detector and the NCC, for L = 100 and L = 50. Results for the wavelet denoiser. Solid lines: Nikon D7000; dashed lines: Canon 600D. The results indicate that both the NP and the NCC detectors provide a similar detection performance for the Nikon D7000. In contrast, the results for the Canon 600D are less favourable, which is in agreement with the results for the lower bound depicted in Tab. 1. In both cases, the detection performance degrades significantly when L increases.

`50brt`	`50drk`	`49brt+berry`	`49drk+berry`
0.8006	0.4399	0.8074	0.5290

Table 2 Lower bound (6) for flat-field images with and without the image in Fig. 2(Left). The term `drk` stands for dark cardboard images, while `brt` refers to bright cardboard images. The image in Fig. 2(Left) is indicated as `berry`. As seen, leakage is larger for darker images, as predicted by the multiplicative model of the PRNU (5).

Camera	**ILB** ($L = 26$)	**ILB** ($L = 50$)
NikonD60	1.4157	1.0414
Canon1100D#1	1.2263	0.9462
Canon1100D#2	1.3252	1.0502
Canon1100D#3	1.2823	0.9995
NikonD3000	1.1892	0.9397
NikonD3200	1.3060	1.0275
NikonD5100	1.7740	1.3218
Canon600D	0.7126	0.6790
NikonD7000	1.1890	0.9144
XiaomiMI5S	1.3577	1.1285

Table 3 Lower bound (6) in bits per pixel for different cameras and sizes of estimation sets when the BM3D denoising algorithm is employed. The lower bound oscillates for different camera models, ranging from 1.7740 bpp in the best case to 0.7126 bpp (for $L = 26$).

image does not increase the information leakage of bright flat-field images, as the former gets diluted in the latter when averaging, this is not the case for dark images: the new image has a considerable impact on $\mathbf{N}_k$ and thus contributes to a larger leakage. This is consistent with the empirical observation that it is easier to extract traces from the image in Fig. 2(Left) when averaged with dark images (cf. Fig. 2(Right)).

Table 3 contains the results of repeating the experiment shown in Table 1 but using the BM3D denoising algorithm [20] instead of the wavelet-based one. The objective here is to show that a better

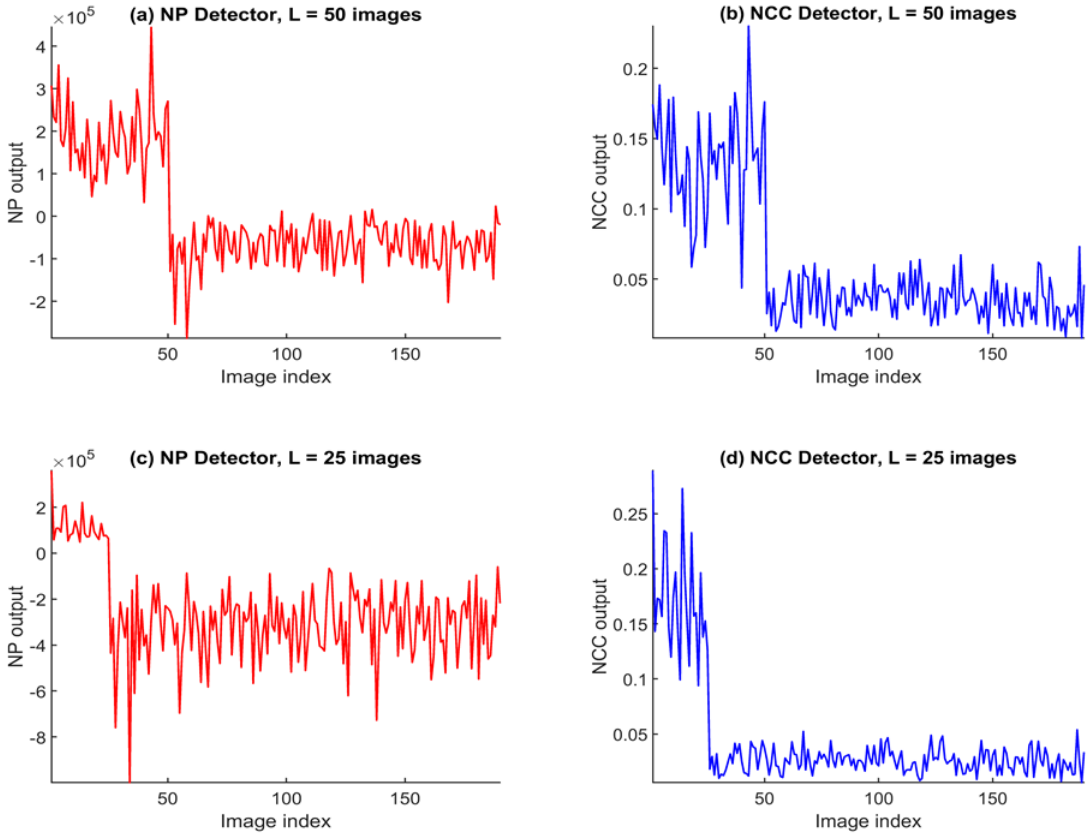


Figure 6 Detection statistics for the Neyman-Pearson detector (14) (images (a) and (c), the corresponding scale is $\times 10^5$) and the Normalized Correlation coefficient detector (19) (images (b) and (d)) on a set of 190 images (Nikon D7000 camera), where the PRNU is estimated from the first 50 images (top row) or the first 25 images (bottom row). From the results, it is clear that the detector is able to identify which images were used to estimate the PRNU, but its performance decreases when more images are employed in the estimation, as the contribution of each individual image gets diluted when larger datasets are considered.

denoising reduces the leakage. Even though all ILBs are smaller for the BM3D algorithm, the reduction with respect to the wavelet-based filter is not as substantial as one would expect, given the additional computational cost that it entails.

6.3 Membership inference

Aiming at testing the ability and accuracy of both NP and NCC membership inference detectors, experiments were performed with PRNUs estimated from subsets of 25 and 50 TIFF images, randomly selected from a set of 190 images captured using the NikonD7000 camera. In Fig. 6 the outputs of the NP and NCC detectors are represented for one such subset. The first 50 samples of the shown sequence correspond to the membership test statistics for those 50 images used to estimate the PRNU. From the results, it is clear that the detector is able to distinguish which images were used to estimate the PRNU in a given dataset.

In the same figure we also show the results of repeating the same process considering PRNUs estimated from randomly chosen sets of 25 TIFF images. As expected, the output of the detectors follows the same trend, but the difference between both levels is now larger, since the individual contributions of each image are less relevant when larger datasets are considered for the estimation.

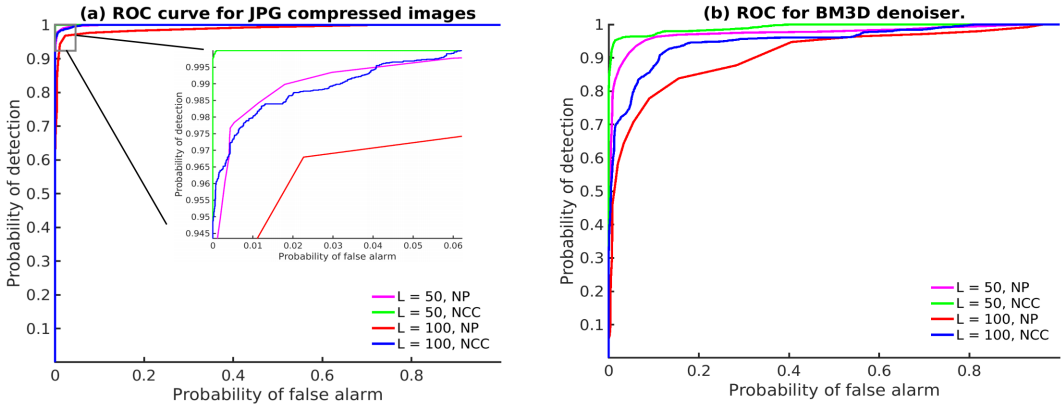


Figure 7 Receiver operating characteristic for the NP detector and the NCC, for L = 100 and L = 50. (Left) Results for the wavelet denoiser, using JPEG compressed images with a Quality Factor of 92, for the camera Nikon D7000. The results indicate that both the NP and the NCC detectors provide a similar detection performance. (Right) Results for BM3D denoiser, for TIFF images obtained from the Nikon D7000. As we can see, the performance of the detectors decreases with respect to the wavelet denoiser, but the test is still able to obtain an acceptable degree of discrimination, showing that the leakage is still present in the images.

These results are confirmed by representing the ROC curves for both detectors in Fig. 7 with $L = 100$ and $L = 50$, generated using 160 different combinations of TIFF images to obtain the PRNU, selected randomly. From this figure the degradation when L increases is again evident. Besides, the NP detector obtains marginally better results, as expected since it was derived from a likelihood ratio. In Fig. 5 (right) the results for the camera Canon600D are also included. From all our set of cameras, this was the only one in which the membership inference method failed systematically. The reasons why are to be fully researched yet. In any case, these results match those depicted in Tab. 1, where the lower bound on the mutual information for this camera is the lowest between all the tested devices. The excellent results (from an attacker's point of view) obtained with the NikonD7000 are also explainable from the ILBs in the table since this particular model exhibits a high ILB. This confirms the existence of a very close relationship between the membership identification and the lower bound expressed in Eq. (6), which we intend to explore in the future.

In Fig. 8 the experiments shown in Fig. 6 were repeated, but considering only $L = 50$ images drawn from a set of 190 JPEG compressed images using a Quality Factor of 92. We focused again on the Nikon D7000. From the results, we can see that both detectors perform similarly in this scenario. These conclusions can be further verified with the ROC curves plotted in Fig. 7(a), obtained following the same experimental setup than for the uncompressed case.

In Fig. 7(b), the ROC curves following exactly the same procedure as in the previous experiments, but considering the BM3D denoiser instead of the wavelet-based approach and TIFF images, are plotted. As we can see from the results, the performance of both detectors decreased, which was expected since the BM3D performs better than the basic wavelet denoiser, removing more contextual information. However, we can see that the test still performs acceptably, showing that improving the denoiser is not the most effective practice to reduce the leakage, and confirming the results obtained with the mutual information.

7 Conclusions

In this paper, the leakage in the PRNU from the database of images used for its estimation is revealed and lower-bounded using a information-theoretic approach. Experimental results show that this leakage

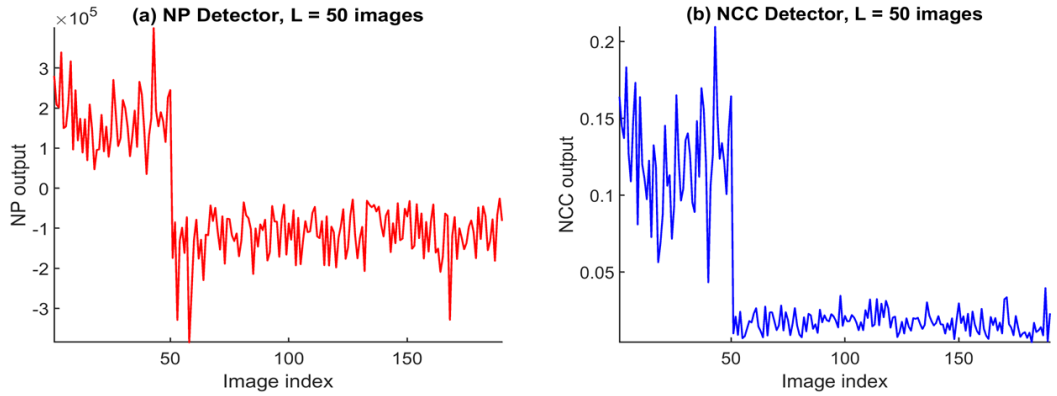


Figure 8 Detection statistics for the Neyman-Pearson detector (14) (a) and the Normalized Correlation coefficient detector (19) (b) on a set of 190 JPEG-compressed images (Nikon D7000 camera), where the PRNU is estimated from the first 50 images. The performance of both detectors decreased slightly, as the compression process enhances the denoising. In both cases, the two levels can still be differentiated.

is substantial and thus can entail significant risks to privacy. As a consequence of this leakage, membership identification based on the PRNU becomes possible using Neyman-Pearson and Correlation based approaches, achieving high accuracy for both detectors. More importantly, the leakage here uncovered calls for a careful risk assessment and additional security and privacy measures when it comes to sharing PRNU-fingerprint databases. Different methods to mitigate the leakage were discussed and experimentally tested. First, we addressed the gain afforded by increasing the number L of images used for the estimation and showed that while effective, this strategy produces diminishing returns. On the one hand, we investigated the option of using JPEG compression as a mean to mitigate this phenomenon, and showed that in practice compression schemes provide few advantages over working with uncompressed images. On the other hand, experiments with the BM3D were also performed. Despite the relative improvement on the obtained results compared with the wavelet denoiser, the results also showed that it is not the most effective way to solve the leakage problem.

This paper is still a first step to model and remove the leakage from the PRNU. Some open problems we expect to tackle in the near future are:

- **Image database reconstruction.** Use machine learning techniques to reconstruct as reliably as possible the image database from the estimated PRNU. This will illustrate even further the threats to privacy and support the use of leakage mitigation techniques.
- **Data-driven PRNU estimators.** Analyze the leakage phenomena on machine learning-based PRNU estimators.
- **Alternative mitigation methods.** Investigate on alternative leakage mitigation techniques, as high pass filters (both fixed and based on learning methods).
- **Compression schemes.** Analyze more aggressive compression schemes, and the trade-off between leakage mitigation and detection performance.

Abbreviations

PRNU: Photo Response Non-Uniformity.
NP: Neyman-Pearson (test/detector).
NCC: Normalized Correlation Coefficient.
ROC: Receiver Operating Characteristic (curve).

BM3D: Block Matching and 3D (filtering).
JPEG: Joint Photographic Experts Group.
TIFF: Tagged Image File Format.
ILB: Information Lower Bound.
DFT: Discrete Fourier Transform.
ML: Maximum Likelihood (estimator).
CCD: Charge Coupled Device (camera sensor).
CMOS: Complementary Metal-Oxide-Semiconductor (camera sensor).

Competing interests
The authors declare that they have no competing interests.

Availability of data and material
Data and material are available under request.

Authors' contributions
All authors contributed equally to this manuscript.

Acknowledgements
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.

Author details
[1]EE Department, Imperial College London, South Kensington Campus, SW7 2AZ London, UK. Email: sf219@ic.ac.uk. [2]Signal Processing in Communications Group, Atlanttic Research Center, Campus Universitario Lagoas, E-36310 Vigo, Spain. Email: fperez@gts.uvigo.es.

References

1. Chen M, Fridrich J, Goljan M, Lukas J. "Determining Image Origin and Integrity Using Sensor Noise". IEEE Transactions on Information Forensics and Security. 2008;3(1):74–90.
2. Rosenfeld K, Sencar HT. A study of the robustness of PRNU-based camera identification. In: Media Forensics and Security. vol. 7254. International Society for Optics and Photonics; 2009. p. 72540M.
3. Taspinar S, Mohanty M, Memon N. Source camera attribution using stabilized video. In: 2016 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE; 2016. p. 1–6.
4. Korus P, Huang J. "Multi-Scale Analysis Strategies in PRNU-Based Tampering Localization". IEEE Transactions on Information Forensics and Security. 2017;12(4):809–824.
5. Cozzolino D, Verdoliva L. Noiseprint: A CNN-Based Camera Model Fingerprint. IEEE Transactions on Information Forensics and Security. 2020;15:144–159.
6. Gloe T, Kirchner M, Winkler A, Bohm R. Can we trust digital image forensics? In: 15th ACM Int. Conf. Multimedia; 2007. p. 78–86.
7. Goljan M, Fridrich J, Chen M. Defending Against Fingerprint-Copy Attack in Sensor-Based Camera Identification. IEEE Transactions on Information Forensics and Security. 2011;6(1):227–236.
8. Pérez-González F, Fernández-Menduiña S. PRNU-leaks: facts and remedies. In: 28th European Signal Processing Conference (EUSIPCO); 2020. .
9. Bondi L, Pérez-González F, Bestagini P, Tubaro S. Design of projection matrices for PRNU compression. In: 2017 IEEE Workshop on Information Forensics and Security (WIFS). IEEE; 2017. p. 1–6.
10. Bondi L, Bestagini P, Perez-Gonzalez F, Tubaro S. Improving prnu compression through preprocessing, quantization, and coding. IEEE Transactions on Information Forensics and Security. 2018;14(3):608–620.
11. Bayram S, Sencar HT, Memon N. Efficient Sensor Fingerprint Matching Through Fingerprint Binarization. IEEE Transactions on Information Forensics and Security. 2012;7(4):1404–1413.
12. Valsesia D, Coluccia G, Bianchi T, Magli E. Compressed fingerprint matching and camera identification via random projections. IEEE Transactions on Information Forensics and Security. 2015;10(7):1472–1485.
13. Cozzolino D, Gragnaniello D, Verdoliva L. Image forgery detection through residual-based local descriptors and block-matching. In: 2014 IEEE international conference on image processing (ICIP). IEEE; 2014. p. 5297–5301.
14. Qian Y, Dong J, Wang W, Tan T. Deep learning for steganalysis via convolutional neural networks. In: Media Watermarking, Security, and Forensics 2015. vol. 9409. International Society for Optics and Photonics; 2015. p. 94090J.
15. Rao Y, Ni J. A deep learning approach to detection of splicing and copy-move forgeries in images. In: 2016 IEEE International Workshop on Information Forensics and Security (WIFS); 2016. p. 1–6.
16. Liu Y, Guan Q, Zhao X, Cao Y. Image forgery localization based on multi-scale convolutional neural networks. In: Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security; 2018. p. 85–90.
17. Bayar B, Stamm MC. A deep learning approach to universal image manipulation detection using a new convolutional layer. In: Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security; 2016. p. 5–10.
18. Verdoliva L, Cozzolino D, Poggi G. A feature-based approach for image tampering detection and localization. In: 2014 IEEE international workshop on information forensics and security (WIFS). IEEE; 2014. p. 149–154.
19. Marra F, Poggi G, Sansone C, Verdoliva L. Evaluation of residual-based local features for camera model identification. In: International Conference on Image Analysis and Processing. Springer; 2015. p. 11–18.
20. Dabov K, Foi A, Katovnik V, Egiazarian K. Image denoising by sparse 3D transform-domain collaborative filtering. IEEE Transactions on Image Processing. 2007;16(8):2080–2095.
21. Pedrouzo-Ulloa A, Masciopinto M, Troncoso-Pastoriza JR, Pérez-González F. Camera Attribution Forensic Analyzer in the Encrypted Domain. In: 2018 IEEE International Workshop on Information Forensics and Security (WIFS); 2018. p. 1–7.

22. Mihcak K, Kozintsev I, Ramchandran K. Spatially Adaptive Statistical and its Application to Denoising of Wavelet Image Coefficients Modeling. In: IEEE Intl. Conf. on Acoustics, Speech and Signal Processing. vol. 6; 1999. p. 3253–3256.
23. Masciopinto M, Pérez-González F. Putting the PRNU Model in Reverse Gear: Findings with Synthetic Signals. In: 2018 26th European Signal Processing Conference (EUSIPCO); 2018. p. 1352–1356.
24. Ihara S. On The Capacity Of Channels with Additive Non-gaussian Noise. Information and Control. 1978;37(1):34–39.
25. Jorswieck EA, Boche H. Performance Analysis of Capacity of MIMO Systems under Multiuser Interference Based on Worst-Case Noise Behavior. EURASIP Journal on Wireless Communications and Networking. 2004 Dec;2004(2):670321.
26. Shokri R, Stronati M, Song C, Shmatikov V. Membership Inference Attacks Against Machine Learning Models. In: 2017 IEEE Symposium on Security and Privacy (SP); 2017. p. 3–18.
27. Kay SM. Detection theory. Prentice Hall PTR; 1998.
28. Goljan M, Fridrich J. Sensor-Fingerprint Based Identification of Images Corrected for Lens Distortion. In: Memon ND, Alattar AM, III EJD, editors. Media Watermarking, Security, and Forensics 2012. vol. 8303. International Society for Optics and Photonics. SPIE; 2012. p. 132–144.
29. Cozzolino D, Poggi G, Verdoliva L. Splicebuster: A new blind image splicing detector. In: 2015 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE; 2015. p. 1–6.