Game Theoretic Analysis of Camera Source Identification

Hui Zeng¹, Yunwen Jiang¹, Xiangui Kang¹², Li Liu³

¹School of Information Science and Technology, Sun Yat-sen University, Guangzhou, GD, China, 510006
²State Key Laboratory of Information Security (Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093), E-mail: isskxg@mail.sysu.edu.cn
³Marvell Semiconductor Inc., Santa Clara 95054, USA, liuli1@gmail.com

Abstract—Sensor pattern noise (SPN) is recognized as a reliable device fingerprint for camera source identification (CSI). However, source identification method (source test) ignores whether the fingerprint is forged and anti-forensic techniques seldom consider traces they leave behind. Therefore, the performance of above techniques needs to be evaluated again by assuming the existence of both parties of a forensic investigator and an anti-forensic forger. In this paper, we propose a novel counter anti-forensic method based on noise level estimation to detect the possible forgery (forgery test). Furthermore, we evaluate the Nash equilibrium performance when investigator performs both source test and forgery test, and identify the optimal strategies of both parties with the game theory. The experimental results show that our proposed method can achieve good performance without collecting the candidate image set in the existing triangle test method especially when the false alarm rate is held low (e.g. $P_{fa} < 5\%$).

I. INTRODUCTION

Image forensics aims to identify the originality or authenticity of an image based on some intrinsic characteristics of the image, whereas anti-forensics aims to trick forensics by removing or forging the characteristics upon which forensics is based. Therefore, the reliability of forensics is questionable when the anti-forensics exists, which urges researchers to obtain more convincing forensic techniques.

A digital camera source identification method based on SPN was proposed in [1-2] and was further studied in [3-6]. In [1-2], a Weiner filter was used to extract the noise residue from an image in the wavelet domain, and the photo response non-uniformity (PRNU) factor of the camera was estimated using a maximum likelihood approach. A series of post-processing is also performed to suppress artefacts of colour interpolation, on-sensor signal transfer, and sensor design from PRNU [2].

In [7], the authors proposed a fingerprint-copy attack method as an anti-forensic method to trick the forensic methods [1-2]. This attack raises a question that innocent people could be framed and criminals could claim their innocence.

To defeat the fingerprint-copy attack, a counter anti-forensic method was proposed in [8], which is based on the difference between the extracted noise of an original image and a forged (fingerprint-copy attacked) image. The extracted noise from a forged image can be divided into an image-dependent portion and an image-independent portion, whereas that from an original image contains only the image-dependent portion. Another counter anti-forensic method, called the triangle test, was proposed in [9], which is based on a common portion between the extracted noise of a stolen image and a forged image. The experimental results in [9] show that it’s quite reliable to detect the fingerprint-copy attack when the investigator could collect the stolen images accurately. However, the method may become unreliable if the investigator failed to collect the stolen images accurately or the forger could obtain a refined fingerprint [10].

Moreover, the performances of the counter anti-forensic methods mentioned above are heavily related to the attack strength. Therefore, an intelligent forger can make a trade-off in choosing the strength. This cat-and-mouse game raises new questions: What is the optimal strategy that a forger must employ in order to fool the traditional test and to avoid detection of anti-forensics traces? What is the optimal strategy for a forensic investigator to employ in order to identify digital forgeries?

To address above challenges, in this paper, we first propose a novel counter anti-forensic method based on noise level estimation [11]. By considering the fingerprint-copy attack as a procedure of noise addition, we estimate the noise level of a test image to decide whether it is fingerprint-copy attacked. Moreover, by analysing the interplay between the camera source investigator and the forger, we introduce a game theory model to evaluate the performance and identify the optimal strategies for both sides.

The rest of this paper is organized as follows. Section II describes the new counter anti-forensic method and defines both sides in the CSI game. Section III shows how to use the game theory model to evaluate the reasonable performance of both sides. Section IV presents the experimental results and the conclusions are drawn in Section V.

II. INTERPLAY BETWEEN INVESTIGATOR AND FORGER

In this section, we introduce the interplay (game) between the investigator and the forger. Let Alice be the investigator and Eve be the forger. Alice owns a digital camera $A$.

1) Forger. Eve makes a fingerprint-copy attack to frame Alice.

This work was supported by NSFC (Grant nos. 61070167, 61379155, U1135001), the Research Fund for the Doctoral Program of Higher Education of China (Grant no. 20110171110042) and NSF of Guangdong province (Grant no. s2013020012788).
β denotes the strength needs to perform forensic investigator adopts the following measure.

(1).

where \( f \) denotes the fingerprint-copy attack as shown in (1) and choose different overwhelming, we assume that the forger adopts the

between the investigator and the forger being too whereas the source of the test images can be identified accurately when no forger exists (Fig. 1(a)), whereas the source of the test images is hard to be identified when the forger exists (Fig. 1(b)).

Finally, Eve frames Alice as the person who took the image \( J' \). Fig. 1 shows the effectiveness of the fingerprint-copy attack. It can be observed that the source of the test images can be identified accurately when no forger exists (Fig. 1(a)), whereas the source of the test images is hard to be identified when the forger exists (Fig. 1(b)).

In this paper, in order to avoid the analysis of a game between the investigator and the forger being too overwhelming, we assume that the forger adopts the fingerprint-copy attack as shown in (1) and choose different \( r \) for different strength \( β \) since \( β = r × β_{NR} \). In the sequel, we use \( r \) instead of \( β \) to denote strength. We also assume that the forensic investigator adopts the following measure.

2) Investigator. In order to prove her innocence, Alice needs to perform source test firstly to identify the source information and then perform forgery test to detect the forgery.

In this paper, we propose a novel counter anti-forensic method based on noise level estimation [11] to detect the forgery. The noise level estimation (here we use \( σ \) to denote the estimated noise level) can be summarized as the following major steps:

1. The test image is decomposed into overlapping patches. The default patch size is 7 \( × \) 7 pixels.
2. An initial noise level \( σ^{(0)} \) is estimated from the covariance matrix, which is generated using all patches in the test image.
3. The weak textured patches are selected from the test image using a threshold that varies with \( σ^{(0)} \).
4. A new noise level \( σ \) is estimated using the selected patches. The process of step 3 and 4 is iterated until \( σ \) is stable.

The following experiment in Fig. 3 shows that we can distinguish a forged image from an original one by estimating the noise level. This can be explained as follows. PRNU follows Gaussian-like probability density function (pdf) [12] and it is verified by our experiments. Fig. 2(a) and (b) show the pdfs of PRNU and their curve fitting results with Matlab, where SSE is the sum of squares due to error. It’s observed that the pdfs of PRNU fit Gaussian model very well. Hence, it’s reasonable to consider the procedure of the fingerprint-copy attack as Gaussian noise addition (see Appendix for a theoretical proof), which may cause \( σ(J) > σ(J') \).

We calculated the noise level of 300 original images and 300 forged (fingerprint-copy attacked) images. Fig. 3(a) and (b) show the estimated noise level when \( r = 0.5 \) and \( r = 1 \) separately. The ‘+’ represents the estimated noise level of original image, and the ‘*’ represents that of forged image. It can be seen that most of the original images’ noise levels are lower than 0.4, whereas most of the forged images’ noise levels are higher than 0.4. Table 1 presents the detection results when different value \( r \) is chosen. In Table 1, the detection threshold is chosen to maximize the accuracy, where accuracy is set to be (true positive rate + true negative rate)/2.
The true positive rate is the probability of that a forged image is detected as the forged image. The true negative rate is the probability of that an original image is detected as the original image. The accuracy decreases with decreasing $r$. However, it still achieves 91% even when the strength that Eve used is a half of the natural strength ($r = 0.5$). Hence, Alice can use the noise level to detect whether $J'$ is forged.

Table I.

<table>
<thead>
<tr>
<th>$r$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
<th>1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td><strong>91%</strong></td>
<td>92%</td>
<td>93%</td>
<td>93%</td>
<td>94%</td>
<td><strong>94%</strong></td>
<td>95%</td>
</tr>
</tbody>
</table>

Fig. 2 Histogram for PRNU of some cameras. (a) FujiFilm_FinePixJ50, SSE = 1.321e-005 and (b) Kodak_M1063, SSE = 3.036e-005

Fig. 3 Estimated noise level of original images and forged images. (a) $r = 0.5$ and (b) $r = 1$
III. GAME THEORETIC EVALUATION

In this section, we introduce and define the source test and the forgery test in detail. To evaluate the performance with both source test and forgery test, we model the interplay between an investigator and a forger as a zero-sum game.

As mentioned above, Alice needs to perform a two-step test including source test and forgery test.

1) First step. Alice performs source test. She uses a sensor-based camera source identification method $\delta_1$ [2] to identify whether a test image $J'$ is taken by $A$. This can be expressed as a hypothesis testing problem:

$$\delta_1(J') = H_0^{(1)}: J' \text{ is not taken with camera } A \quad \text{and} \quad H_1^{(1)}: J' \text{ is taken with } A$$

The acceptance region of $H_1^{(1)}$ is:

$$J': C(W(J'), K_sJ') < t_1 \quad (3)$$

where $W(J')$ is the noise residue of $J'$, $K_s$ is the fingerprint of camera $A$, and $C(\cdot)$ is a correlation function [2]. The decision threshold $t_1$ is chosen according to the false alarm rate $P_{fa}^1$. $P_{fa}^1$ is defined as:

$$P_{fa}^1 = P(\delta_1(J') = H_1^{(1)} | J' \text{ is taken with } A) \quad (4)$$

2) Second step. Alice uses a counter anti-forensic method $\delta_J$ to perform forgery test. As an example to present, Alice uses our proposed method described in Section II to identify whether the images accepted in $H_0^{(1)}$ are forged or not.

She computes the noise level of the test image $J'$ [11]. If $J'$ is forged, $\sigma(J)$ will be larger than that of original $J$. Thus, the detection problem can be expressed as a hypothesis testing problem:

$$\delta_J(J') = H_0^{(2)}: J' \text{ is original}, \text{that is, } J' \text{ has not been forged.}$$

$$\delta_J(J') = H_1^{(2)}: J' \text{ is forged. In other words, } J' \text{ has been fingerprint-copy attacked.}$$

The acceptance region of $H_1^{(2)}$ is:

$$J': \sigma(J') > t_2 \quad (5)$$

where $t_2$ is chosen according to the false alarm rate $P_{fa}^2$. $P_{fa}^2$ is defined as:

$$P_{fa}^2 = P(\delta_J(J') = H_1^{(2)} | J' \text{ is original}) \quad (6)$$

The total detection rate of the two-step test can be denoted as:

$$P_d = P(\delta(J') = H_1^{(1)} \cup \delta_J(J') = H_1^{(2)} | J' \text{ is forged}) \quad (7)$$

and the total probability of false alarm rate is defined as:

$$P_{fa} = P_{fa}^1 + P_{fa}^2 \quad (8)$$

It is observed that $P_{fa}^1$, $P_{fa}^2 \in [0, P_{fa}]$. $P_{fa}^1 = 0$ corresponds to the case where the investigator only performs forgery test and $P_{fa}^1 = P_{fa}$ corresponds to the case where the investigator only performs source test. For a given $P_{fa}$ and a given $r \in [0, r_m]$ where $r = r_m$ corresponds to the maximum strength, the investigator would choose a $P_{fa}^1$ to achieve the maximum $P_d$ (Fig. 4). For given parameters $P_{fa}$ and $P_{fa}^1$, the forger choose a strength $r$ to achieve the minimum $P_d$ (Fig. 4). The interplay between an investigator and a forger can be modelled as a camera source identification game (CSI game). We adopt the convention that the investigator moves first [13], that is, the investigator chooses $P_{fa}^1$ firstly and then allows the forger to respond.

The utility of the investigator is denoted as:

$$U_1(P_{fa}^1, r) = P_d(P_{fa}^1, r) \quad (9)$$

Fingerprint-copy attack hardly introduces any perceptual distortion [9] and it is verified by our experiments that the PSNR between $J$ and $J'$ is greater than 48 dB for all images. So the CSI game can be assumed to be a zero-sum game, and the utility of the forger is:

$$U_2(P_{fa}^1, r) = -P_d(P_{fa}^1, r) \quad (10)$$

Game theory is used to derive the Nash equilibrium $(P_{fa}^{1*}, r^*)$ [14] where neither of the players has any incentive to change his strategy.

For a given $P_{fa}$, the Nash equilibrium of CSI game can be derived, and the detection ratio of the two-step test can be obtained as $P_d = U_1(P_{fa}^{1*}, r^*)$. By changing $P_{fa}$, a receiver operating characteristic (ROC) curve can be constructed to show the variation of $P_d$ with $P_{fa}$ under the Nash equilibrium (Fig. 5). It is called Nash equilibrium ROC curve [13] in short form.

IV. THE EXPERIMENTAL RESULTS

In this section, we examine the performance of the camera source identification game under Nash equilibrium. All the images involved in our experiments are from the Dresden Image Database (DID) [15] and are cropped from the centre of full size images to a size of 1024 × 768 pixels.
Fig. 4 The detection performance when $P_{fa} = 0.05$

Fig. 5 Nash equilibrium ROC with proposed method
Section IV.A shows the Nash equilibrium performance when Alice adopts the proposed counter anti-forensic method in the forge test. In Section IV.B, by comparison, we show the Nash equilibrium performance when Alice adopts the triangle test [9] in the forge test.

The experiments are performed under the following scenarios. Alice shared many images taken by her camera Kodak M1063 via the Internet. Eve stole $N=50$ of them and extracted a fingerprint from the stolen images. To frame Alice, Eve superimposed this fingerprint on some images taken by other cameras (Casio, FujiFilm, Olympus) to make a fingerprint-copy attack.

A. Result with the proposed method

To evaluate the detection ratio of Alice, we forged 500 images with (1). We find the Nash equilibrium strategies in the experiment by solving the following equation [13]:

$$\left( P_{fa}^*, r^* \right) = \arg \max_{P_{fa}} \min_{r} \beta \left( P_{fa}^0, r^0 \right)$$

(11)

Fig. 4 shows the total detection rate $P_d$ when the total probability of false alarm rate $P_{fa}$ is 0.05. The x-axis represents the strength $r$, and the y-axis denotes the false alarm rate $P_{fa}$. Under this setting, it is observed that the Nash equilibrium $\left( P_{fa}^*, r^* \right)$ is (0.04, 0.6), which corresponds to Alice chooses $P_{fa}^0 = 0.04$ first and Eve responds with $r = 0.6$. The total detection rate $U_c \left( P_{fa}^0, r^0 \right)$ is 90% in this case.

A zoomed version of the Nash equilibrium ROC curve for $P_{fa} \in [0, 0.25]$ is presented in Fig. 5. In general, the forensic analyst is interested with the ROC in the low false alarm rate area. The detection rate $P_d$ is higher than 90% with $P_{fa} = 0.05$, so the CSI results are relatively reliable. Note that every point on the ROC in Fig. 5 is a Nash equilibrium point at different $P_{fa}$, and every detection rate $P_d$ corresponds to a forger’s optimal choice of $r$. The optimal choices of $r$ are between 0.5 and 1 according to our experiments.

B. Result with the triangle test

By comparison, we also investigate the correlation performance when Alice uses the triangle test [9] to perform forge test. In the triangle test, the authors in [9] made an assumption that all stolen images (N=50) are included in the $N_c$ ($N_c \geq N$) candidate images. $N/N_c$ is the accuracy of including the stolen images in the Alice chosen candidate image set.

For the sake of conciseness, in the following the correlation between two images means the correlation between the two noise residue patterns extracted respectively from two images. If $J'$ is an original image, the measured correlation between $J'$ and a not stolen candidate image $I$, is approximately proportional to its theoretical calculated correlation (expected correlation), and they should well fit with a line as the blue line shown in Fig. 6(a). In Fig. 6(a), $J'$ is an original image, the blue points are the measured correlation vs. expected correlation between not stolen images $I$ and $J'$, and they should well fit with a blue line $L$. The red circle points are the measured correlation vs. expected correlation between stolen images $I$ and $J'$. It can be observed that the red points are similar as the blue points.

When $J'$ is a forged image, Fig. 6(b) presents a typical plot of measured correlation vs. expected correlation for stolen image (red circle) and not stolen image $I$ (blue point). It can be observed that the red circles are above deviated from the blue line.

Note that if Alice chooses more images as candidate images, which means, $N_c$ becomes larger, the accuracy $N/N_c$ becomes lower, and the performance of the triangle test would degrade [9]. The Nash equilibrium ROC curves with different $N/N_c$ are presented in Fig. 7, showing that the detection result of the investigator is reliable only when the value of $N/N_c$ is relatively high. For example, the total detection rate $P_d$ is higher than 89% with $P_{fa} = 0.05$ and $N/N_c > 0.2$. However, when the candidate images that Alice collected are not so accurate (Alice has uploaded too many images to her blog, that is, $N_c$ is large), her detection result may not be so reliable. In this case, Alice shall adopt our proposed counter anti-forensic method because our method can achieve good performance without collecting the candidate image set.

It is observed that our proposed method achieves good performance when the false alarm rate is held low (e.g. $P_{fa} < 5\%$). Another advantage of our method is of low computational complexity. The average run time to test a 1024 $\times$ 768 colour image by our method is 1.5s, whereas that is more than 40s ($N_c=100$) by the triangle test. All the tests are performed on a computer with a 3.1 GHz processor and 4 GB RAM.

V. CONCLUSIONS

In this paper, we propose a novel counter anti-forensic method based on noise level estimation and examine the interplay between forensics and anti-forensics. The experimental results show that our proposed method can achieve good performance without collecting the candidate image set and achieve good performance when the false alarm rate is held low (e.g. $P_{fa} < 5\%$). Because the cost associated with false alarm is often high in real world scenarios, these results demonstrate that our proposed technique is more feasible than the existing technique. We introduce a game theory model to evaluate the ultimate result of this interplay and identify the optimal strategies of both sides. It is observed that the optimal strategy for the forger is to use strength between a half of natural strength and a full natural strength $(0.5 < r^* < 1)$.

In this paper, we limit our analysis to a game between a specific anti-forensics and its countermeasures. We assume that the anti-forensic forger adopts the fingerprint-copy attack and the forensic investigator adopts our proposed method or the triangle test. However, the game theory model would be suitable for analyzing similar interplays between forensics and anti-forensics. We also find that many other anti-forensics have a similar Gaussian noise adding procedure [16 - 18] as [7]. Hence, our proposed counter anti-forensic method may be a universal countermeasure to such anti-forensics.
Fig. 6 Measured correlation vs. Expected correlation between $I$ and $J'$. When $I$ is a stolen image, the sample point is red, otherwise, it is blue. (a) $J'$ is original and (b) $J'$ is forged ($r = 1$).
ACKNOWLEDGMENT

The authors would like to thank the authors of [1] [2] and [9] for discussing the problem about the triangle test and sharing their code on the website. The authors would like to thank the authors of [11] for discussing the problem about the noise level estimation and sharing their code on the website.

APPENDIX

Lemma 1. The fingerprint-copy attack (1) can be considered as a procedure of Gaussian noise addition [11].

Proof. For each selected weak textured patch $J_p$ [11], the procedure of fingerprint-copy attack can be denoted as:

$$
\hat{J}_p = \{J_p(1+\beta \hat{K}_p)\} = J_p + J_p \beta \hat{K}_p = J_p + j\beta \hat{K}_p
$$

where $\hat{K}_p$ is the fingerprint $\hat{K}$ of the corresponding location. The last approximately equal sign is because only weak textured patch is selected in the noise level estimation [11]. That is, $J_p \approx j$ is approximately constant on the selected patch. $j\beta \hat{K}_p$ can be modelled as Gaussian noise due to $\hat{K}_p$ is modelled as Gaussian noise. Hence, the fingerprint-copy attack is a procedure of Gaussian noise addition from (12).

QED.

REFERENCES


Fig. 7 Nash equilibrium ROC with triangle test


