IMAGE MODEL AND PRINTED DOCUMENT AUTHENTICATION: A THEORETICAL ANALYSIS

Bao An Hoang Mai, Wadih Sawaya, Patrick Bas

IMAGE MODEL AND PRINTED DOCUMENT AUTHENTICATION: A THEORETICAL ANALYSIS

Bao An Mai Hoang , Wadih Sawaya

Institut Mines-Telecom, Telecom Lille,
LAGIS UMR 8219 CNRS
59650 Villeneuve d'Ascq, France

Patrick Bas

CNRS - LAGIS UMR 8219 CNRS,
Ecole Centrale de Lille
59651 Villeneuve d'Ascq, France

ABSTRACT

This paper combines the principles of statistical estimation and hypothesis testing to analyze the impact of parameter estimation on an authentication system based on graphical codes. The studied authentication system uses the fact that a code, once printed, undergoes a stochastic and non invertible alteration. A statistical test applies a likelihood ratio between the model of the authentic printed and scanned image and the model of the reproduced one, with the particularity here that the later model is unknown. The proposed solution consists in using an optimal estimation of the image model coming from observed fake codes in order to perform the likelihood test. Using a second order expansion, we derive a linear relation between the quadratic error of the estimated parameters and the probability of type II error. We are then able to formulate analytically and practically the error spread region of the Receiver Operating Characteristic (ROC) curves, and to compute the average authentication performance when the receiver has to estimate the opponent print and scan channel.

Index Terms— Authentication, Statistical estimation, Hypothesis testing, Printed documents

1. STUDIED PROBLEM

In the world of global exchanges people can use graphical code as a way to perform authentication of physical products such as documents, goods, drugs,.... Our authentication system is based on printed graphical 2D codes using very high resolution printers (2400dpi). Each printed and scanned set of dots (a dot being a binary element) suffers from a stochastic non-invertible noise which makes the reproduction of the original graphical code impossible [12, 14, 10, 9] (see in Fig. 1). The opponent's goal is then to reproduce a printed and scanned code similar to the original printed one, using a printer that will also generate a non-invertible noise.

The goal of the receiver is to reject copied codes. In previous works [11, 6] we modeled the mentioned authentication system as a hypothesis testing problem and we derived tight bounds on its performance. The authentication system works as follow: a binary authentication image is constructed from a randomly chosen binary sequence x^N in $\{0,1\}^N$ by the legitimate source. It is shared secretly with the legitimate receiver and published as a noisy version y^N modeling the original printed and scanned code (see Fig. 1 on the left). The opponent observes y^N and tries to estimate the original image obtaining $\hat{x}^N$ (see (2) in Fig. 2). He then prints it to create a forged observable noisy image z^N hopping that it will be accepted by the receiver as coming from the legitimate source (see Fig. 1 on the right). The observed images y^N and z^N are 8 bits grey level images. In practice, this attack will be used to create false documents or fake packages that could be considered as authentic.

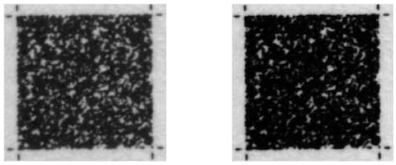

Fig. 1. Left: an original printed and scanned graphical code. Right: a re-printed and scanned (forged) graphical code.

The whole physical process, precisely printing and scanning devices used by the legitimate parts (see (1) in Fig. 2) and by the counterfeiter (see (3) in Fig. 2), are respectively modeled by probability distributions conditioned to the original data $P_{Y/X,\Theta}$ and $P_{Z/X,\bar{\Theta}}$, Θ and $\bar{\Theta}$ are sets of parameters, taken in Ξ, specifying the devices in each case. As pointed in [11], the receiver observes one of the two possible images y^N or z^N and have to decide whether it comes from the legitimate source or not (see (5) in Fig. 2), supposed that the models $P_{Y/X,\Theta}$ and $P_{Z/X,\Theta}$ are known. The print and scan process in this particular setup has been modeled by an AWGN channel or an additive i.i.d. lognormal noise in [2].

Authentication here is based on classical Neyman-Pearson test (NP-test) (see ([7, 4])) in which the receiver considers two hypothesis H_0 and H_1. The former hypothesis attests authenticity, i.e. that the received sequence is generated by $P_{Y/X,\Theta}$ and the latter one unveils a fake code, i.e. that the observed sequence is driven from $P_{Z/X,\bar{\Theta}}$. Performances are evaluated computing the probability of type I error (rejecting

hypothesis H_0 while actually the observed sequence comes from the legitimate source) and the probability of type II error (accepting hypothesis H_0 while it is actually a fake).

In this paper we extend our analysis to the case where the receiver doesn't know the true parameter $\bar{\Theta}$ related to the opponent print and scan process, but establishes a test statistic using estimated ones obtained by a maximum likelihood based algorithm. The estimated parameters are computed from several codes identified previously as fake codes which represent a set of printed and scanned dots driven from $P_{Z/X,\bar{\Theta}}$ (see (4) in Fig. 2).

We derive a linear approximation relating the probability of type II error to the quadratic error on the estimated parameters. This approximation helps us to express analytically the error spread of the authentication performance, and hence to evaluate its average when the receiver doesn't know perfectly the opponent print and scan process.

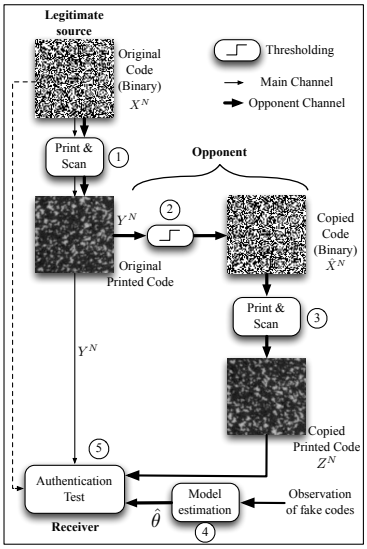

Fig. 2. Principle of authentication using graphical codes.

2. MODELS AND ESTIMATION

2.1. Printing and scanning processes

In our global authentication model, $P_{Y/X,\Theta}$ represents a grey level distribution of the authentic image conditioned to the knowledge of both the authentication dots and the parameters governing the legitimate print and scan process.

When performing a detection to obtain an estimated sequence $\hat{X}^N$, the opponent undergoes errors coming from the facts that he is not able to infer the original code. It is important to note that the opponent will have to print a binary version of its observation because a printing device at this very high resolution can only print binary images.

These errors include probabilities $P_{e,W}$ when confusing an original white dot with a black and $P_{e,B}$ when confusing an original black dot (generated by bit 0) with a white dot (generated by bit 1).

Given $P_{Z/X,\bar{\Theta}}$ a grey level distribution of the forged image conditioned to the knowledge of both the authentication dots and the parameters governing opponent print and scan process $T_{Z/X,\bar{\Theta}}$, we have (equivalently for $X = 0$ or 1):

$$\begin{aligned} P_{Z/X,\bar{\Theta}}(Z = v/X = 0(1), \bar{\Theta}) \\ = (1 - P_{e,B(W)})T_{Z/\hat{X},\bar{\Theta}}(v/\hat{X} = 0(1), \bar{\Theta}) \\ + P_{e,B(W)}T_{Z/\hat{X},\bar{\Theta}}(v/\hat{X} = 1(0), \bar{\Theta}) \end{aligned} \tag{1}$$

2.2. Error spread region for the estimated parameters $\bar{\Theta}$

Although our analysis in this paper does not depend on the estimation method, we consider maximum likelihood estimation (MLE) in order to achieve optimal estimation. In MLE, we have $E[\hat{\bar{\Theta}}] = \bar{\Theta}$ and the estimated set of parameters are jointly normally distributed $\mathcal{N}(\bar{\Theta}; V_{\hat{\bar{\Theta}}})$ where $V_{\hat{\bar{\Theta}}}$ is their covariance matrix which can be computed either practically or analytically using the Fisher information ([1, 3, 13]). Under these assumptions the covariance matrix will help us to provide a measure of how the estimated parameters spread w.r.t the true value. The quadratic form of the error (the variation of the estimation) is chi-squared distributed and:

$$\rho(\hat{\bar{\Theta}}) = (\hat{\bar{\Theta}} - \bar{\Theta})^T V_{\hat{\bar{\Theta}}}^{-1} (\hat{\bar{\Theta}} - \bar{\Theta}) \sim \chi^2_\kappa. \tag{2}$$

where χ^2_κ is the chi-squared distribution with κ degree of freedom ([3, 13, 15]). Herein, κ is the number of parameters that govern the print and scan model. One may observe that $\rho(\hat{\bar{\Theta}}) = cte$ is an ellipsoid in the κ-dimensional space. Using the property of $\rho(\hat{\bar{\Theta}})$ in (2), we can compute the probability that the κ-dimensional error vector lies between two ellipsoids. The error spread region for the estimated parameters is given by:

$$\mathcal{R} = \left\{ \hat{\bar{\Theta}} \,:\, \chi^2_{\kappa,\gamma_1} \leq \rho(\hat{\bar{\Theta}}) \leq \chi^2_{\kappa,\gamma_2} \right\}, \tag{3}$$

where χ^2_{κ,γ_1} and χ^2_{κ,γ_2} are critical levels w.r.t γ_1 and γ_2, i.e, $Pr[\rho(\hat{\bar{\Theta}}) \leq \chi^2_{\kappa,\gamma_1}] = \gamma_1$ and $Pr[\rho(\hat{\bar{\Theta}}) \leq \chi^2_{\kappa,\gamma_2}] = \gamma_2$.

3. AUTHENTICATION PERFORMANCE

In this section we express reliably the probabilities of type I error α and probability of type II error β used to assess the performance of the authentication system in the NP-test setup.

In [11, 6], we indicate several methods to approximate β while keeping α fixed and show that an Asymptotic Expression (AE) provides a reliable approximation for β. We present a brief summary for the AE method.

Given a real number s the Chernoff bounds on α and β may be expressed as:

$$\alpha = \Pr(L \geq \lambda/H_0) \leq e^{-s\lambda} g_{L/H_0}(s) \text{ for any } s > 0, \quad (4)$$

$$\beta = \Pr(L \leq \lambda/H_1) \leq e^{-s\lambda} g_{L/H_1}(s) \text{ for any } s < 0, \quad (5)$$

where L is the log-likelihood ratio of P_{Z^N/X^N} over P_{Y^N/X^N} and λ is a threshold. The function $g_{L/H_j}(s)$ $(j = 0, 1)$ is the moment generating function of L given the hypothesis H_j:

$$g_{L/H_j}(s) = E_{L/H_j}\left[e^{sL}\right]. \quad (6)$$

Without loss of generality, we can use the original code with the same number of bits 0 (N_b) and bits 1 (N_w), i.e., $N_b = N_w = \frac{N}{2}$ with N is the length of the codeword. For large N, (4) and (5) become very tight and specifically with an additional correcting factor we obtain for i.i.d. samples:

$$\alpha \underset{N\to\infty}{\to} \frac{1}{|s_0|\sqrt{N\pi\mu''_\ell(s_0)}} \exp\left\{\frac{N}{2}\left[\mu_\ell(s_0) - s_0\mu'_\ell(s_0)\right]\right\}. \quad (7)$$

$$\beta \underset{N\to\infty}{\to} \frac{1}{|s_1|\sqrt{N\pi\mu''_\ell(s_1)}} \exp\left\{\frac{N}{2}\left[\mu_\ell(s_1) - s_1\mu'_\ell(s_1)\right]\right\}. \quad (8)$$

for $s_0 > 0$ and $s_1 < 0$. Here, given hypothesis H_j, $\mu_\ell(s_j) = \log g_{\ell/H_j}(s_j)$ is the cumulant generating function of $\frac{P_{Z/X}}{P_{Y/X}}$ at the point s_j which is the solution of $\frac{N}{2}\mu'_\ell(s) = \lambda$, and note that one can easily prove that $s_1 = s_0 - 1$.

4. INTERPLAY BETWEEN CHANNEL ESTIMATION AND AUTHENTICATION PERFORMANCE

In this section, we analyze how the set of estimated parameters impacts the performance of the probability of type II error $\beta(\alpha, \hat{\bar{\Theta}})$ for a fixed value of α. Precisely, we relate the error spread region of $\log\, \beta(\alpha, \hat{\bar{\Theta}})$ to the error spread region of $\hat{\bar{\Theta}}$ defined in (3).

For large enough N, the changes of correcting factors $\frac{1}{|s_j|\sqrt{N\pi\mu''_\ell(s_j)}}$ in (7) and (8) are imperceptible and we drop their analysis. Due to the limitation of the paper, we will present our analysis using only one estimated parameter. The extension of this analysis for vectors of estimated parameters will be addressed in future works.

4.1. Analytical Analysis

We want to show the linear tendency of the scatter of $\log \beta(\alpha, \hat{\bar{\Theta}})$ w.r.t $\rho(\hat{\bar{\Theta}})$ by using Taylor expansion. For the case of one parameter, let θ be the estimated version of the true parameter $\theta_t \in \bar{\Theta}$, and $\theta_m \in \Theta$ be the parameter of the legitimate model, and define:

$$\beta^*(\theta) = \frac{2}{N}\log \beta(\alpha, \theta), \quad (9)$$

A Taylor expansion gives:

$$\beta^*(\theta) \cong \beta^*(\theta_t) + \triangle\theta \left.\frac{\partial\beta^*(\theta)}{\partial\theta}\right|_{\theta_t} + \frac{(\triangle\theta)^2}{2} \left.\frac{\partial^2\beta^*(\theta)}{\partial\theta^2}\right|_{\theta_t} + ... \quad (10)$$

where $\triangle\theta = (\theta - \theta_t)$.

We aim now to compute the first and second derivatives of $\beta^*(\theta)$ expressed for $\theta = \theta_t$. Because α is fixed and we choose the same threshold for detection, we have:

$$\frac{E_{P_0}\left[l(\theta)e^{l(\theta)s_0(\theta)}\right]}{E_{P_0}\left[e^{l(\theta)s_0(\theta)}\right]} = \frac{E_{P_1}\left[l(\theta)e^{l(\theta)s_1(\theta)}\right]}{E_{P_1}\left[e^{l(\theta)s_1(\theta)}\right]}, \quad (11)$$

with $P_0 \equiv P_{Y/X,\theta_m}$, $P_1 \equiv P_{Z/X,\theta_t}$, $l \equiv l(\theta) = \log\frac{P_{Z/X,\theta}}{P_{Y/X,\theta_m}}$. This leads to:

$$\frac{\partial\beta^*(\theta)}{\partial\theta} = s_1(\theta)\left\{\frac{E_{P_1}\left[\frac{\partial l}{\partial\theta}e^{l(\theta)s_1(\theta)}\right]}{E_{P_1}\left[e^{l(\theta)s_1(\theta)}\right]} - \frac{E_{P_0}\left[\frac{\partial l}{\partial\theta}e^{l(\theta)s_0(\theta)}\right]}{E_{P_0}\left[e^{l(\theta)s_0(\theta)}\right]}\right\} \quad (12)$$

Notably, when $\theta = \theta_t$, we have $s_1(\theta_t) = s_0(\theta_t) - 1$ and so it can be proved that for every function $f(\theta)$,

$$E_{P_0}\left[f(\theta_t)e^{l(\theta_t)s_0(\theta_t)}\right] = E_{P_1}\left[f(\theta_t)e^{l(\theta_t)s_1(\theta_t)}\right] \quad (13)$$

Therefore:

$$\left.\frac{\partial\beta^*(\theta)}{\partial\theta}\right|_{\theta_t} = 0 \quad (14)$$

The equality (14) is not surprising since the NP-test is known to reach the optimum when applied on the true parameter. Now if we denote:

$$\begin{aligned}
E_1 &= E_{P_1}\left[e^{l(\theta_t)s_1(\theta_t)}\right] \\
E_2 &= E_{P_1}\left[l(\theta_t)e^{l(\theta_t)s_1(\theta_t)}\right] \\
E_3 &= E_{P_1}\left[l^2(\theta_t)e^{l(\theta_t)s_1(\theta_t)}\right] \\
E_4 &= E_{P_1}\left[\left.\frac{\partial l(\theta)}{\partial\theta}\right|_{\theta_t} e^{l(\theta_t)s_1(\theta_t)}\right] \\
E_5 &= E_{P_1}\left[\left(\left.\frac{\partial l(\theta)}{\partial\theta}\right|_{\theta_t}\right)^2 e^{l(\theta_t)s_1(\theta_t)}\right] \\
E_6 &= E_{P_1}\left[l(\theta_t)\left.\frac{\partial l(\theta)}{\partial\theta}\right|_{\theta_t} e^{l(\theta_t)s_1(\theta_t)}\right]
\end{aligned} \quad (15)$$

and given $R_{i1} = E_i/E_1$ for all $i = 2, 3, ..., 6$ we obtain the second derivative of β^* at $\theta = \theta_t$ as

$$\left.\frac{\partial^2\beta^*(\theta)}{\partial\theta^2}\right|_{\theta_t} = \left[R_{61} - R_{41}R_{21}\right]^2\left[R_{31} - R_{21}^2\right]^{-1} + R_{41}^2 - R_{51}, \quad (16)$$

based on the fact that

$$\frac{2}{N}\left.\frac{\partial^2\log\beta(\alpha,\theta)}{\partial\theta^2}\right|_{\theta_t} = \left.\frac{\partial^2\beta^*(\theta)}{\partial\theta^2}\right|_{\theta_t}. \quad (17)$$

If now we call $\gamma(\alpha, \theta_t) = \left.\frac{\partial^2\log\beta(\alpha,\theta)}{\partial\theta^2}\right|_{\theta_t} \times \frac{\mathrm{Var}(\theta)}{2}$ the slope of analytical linear expression, then from (10), (14) and (17), it yields

$$\log \beta(\alpha,\theta) \cong \log \beta(\alpha,\theta_t) + \gamma(\alpha,\theta_t)\rho(\theta) \qquad (18)$$

where $\mathrm{Var}(\theta)$ is the variance of the estimated parameter θ and $\rho(\theta) = \frac{(\triangle\theta)^2}{\mathrm{Var}(\theta)}$ is the variation of the estimation. Moreover, from the property of NP-test, $\beta(\alpha,\theta) \geq \beta(\alpha,\theta_t)$ for all θ so $\gamma(\alpha,\theta_t)$ is always nonnegative. Using (2) and (18), we show that $\log \beta(\alpha,\theta)$ follows a shifted and scaled χ^2 distribution and we are now able to derive a spread error region (see 4.2).

4.2. Numerical results

In order to perform our analysis, we have to construct a MLE scheme for parameter estimation. It is known that the Expectation Maximization (EM) algorithm is an iterative method for finding maximum likelihood. Without loss of generality we assume that $T_{Z/\hat{X}=0,\bar{\Theta}}$ and $T_{Z/\hat{X}=1,\bar{\Theta}}$ are modeled by truncated discrete normal distributions with $\bar{\Theta} = (\bar{\mu}_b, \bar{\sigma}_b^2, \bar{\mu}_w, \bar{\sigma}_w^2)$ such that $P_{Z/X,\bar{\Theta}}$ is a mixture of two truncated Gaussians (see (1)). We then develop an EM algorithm ([5, 8]) for this particular mixture to estimate the set of unknown parameters.

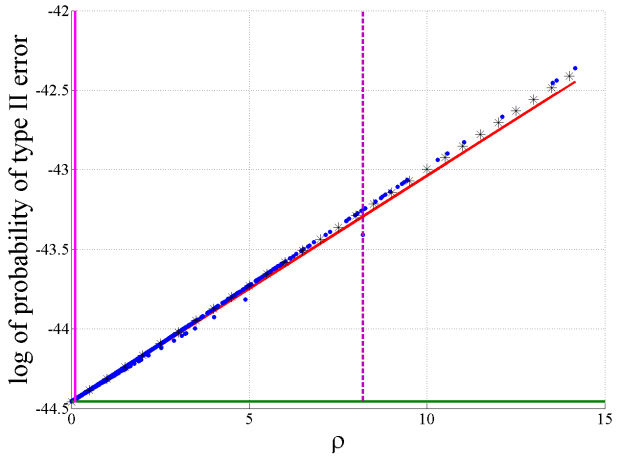

Fig. 3. Comparison between $\log \beta(\rho)$ for the true parameter (horizontal line) and the one (dots) from estimated opponent's parameters. Linear regression (stars); analytical expression (straight line). Critical value $\chi^2_{1,0.025}$ and $\chi^2_{1,0.975}$ are represented by vertical lines. $N_{obs} = 3.10^3$, $\alpha = 10^{-16}$, $N_{iter} = 5.10^3$, $N = 2.10^3$.

In Fig. 3, we suppose that only $\bar{\mu}_w$ is unknown and we run EM algorithm N_{iter} times using each time N_{obs} observations and obtain a set of $\hat{\bar{\mu}}_w$. The scatter plot of Fig. 3 represents the computed values of $\log \beta$ coming from AE method. It is compared with the analytical expression (18) and the statistical linear regression.

In Fig. 4, we analyze the impact of the estimation error on the ROC curves. We select a 95% confidence error region for $\hat{\bar{\mu}}_w$, i.e., $\rho(\hat{\bar{\mu}}_w)$ is bounded by two critical levels $\chi^2_{1,0.025}$ and $\chi^2_{1,0.975}$ such that $Pr[\rho(\hat{\bar{\mu}}_w) \leq \chi^2_{1,0.025}] = 0.025$ and $Pr[\rho(\hat{\bar{\mu}}_w) \leq \chi^2_{1,0.975}] = 0.975$, and we thus obtain a corresponding 95% confidence error region for $\log \beta(\alpha, \hat{\bar{\mu}}_w)$. We then derive two critical ROC curves $C_{min}^{0.025}$ and $C_{max}^{0.975}$ computed analytically from $\chi^2_{1,0.025}$ and $\chi^2_{1,0.975}$ and we choose the mean value for $\rho(\hat{\bar{\mu}}_w)$ to find the mean ROC curve C_{mean}. We then compare $C_{min}^{0.025}$, $C_{max}^{0.975}$ and C_{mean} with the three ones computed from the dataset of $\rho(\hat{\bar{\mu}}_w)$ and we observe that our approximation is accurate.

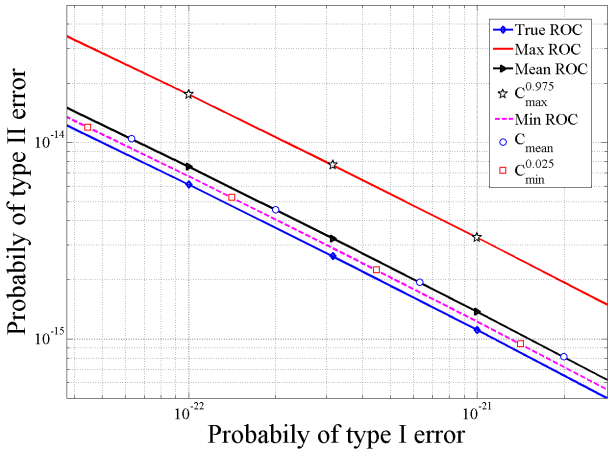

Fig. 4. Comparison between three analytical ROC curves $C_{min}^{0.025}$ (squares), $C_{max}^{0.975}$ (stars) and C_{mean} (circles) with min ROC curve (dash line), max ROC curve (straight line) and mean ROC curve (straight line with triangles) computed from $N_{iter} = 5000$ data of $\hat{\bar{\mu}}_w$.

5. CONCLUSIONS

This paper analyzes the impact of parameters estimation on the performance of document authentication.

- We show experimentally and theoretically an analytical linear relation between the variation of the estimated parameters and the logarithm of the corresponding probabilities of type II error.
- We are able to predict the average authentication loss when performing the NP-test of the estimated distribution for the opponent's channel.
- The proposed analysis is not impacted by the nature of the noise, and can be applied for different memoryless channels that are more realistic to model the printing process, such as the lognormal distribution [2].
- Although we particularly focus on document authentication, our analysis in this paper generally can be applied for a larger class of forensics problems.

6. REFERENCES

[1] E. B. Andersen. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 283–301, 1970.

[2] C. Baras and F. Cayre. Towards a realistic channel model for security analysis of authentication using graphical codes. In *Information Forensics and Security (WIFS), 2013 IEEE International Workshop on*, pages 115–119. IEEE, 2013.

[3] R. Bhattacharya and M. Denker. *Asymptotic statistics*, volume 14. Springer, 1990.

[4] R. Christensen. Testing Fisher, Neyman Pearson, and Bayes. *The American Statistician*, 59(2):121–126, 2005.

[5] A. P Dempster, N. M Laird, and D. B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[6] A. T. Phan Ho, B. A. Mai Hoang, W. Sawaya, and P. Bas. Document authentication using graphical codes: Reliable performance analysis and channel optimization. *EURASIP Journal on Information Security*, 2014(1):9, 2014.

[7] E. L Lehmann. The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association*, 88(424):1242–1249, 1993.

[8] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.

[9] Q. T. Nguyen, Y. Delignon, L. Chagas, and F. Septier. Printer identification from micro-metric scale printing. In *Proc. ICASSP*, pages 6277–6280, 2014.

[10] Q.-T. Nguyen, Y. Delignon, L. Chagas, and F. Septier. Printer technology authentication from micrometric scan of a single printed dot. In *IS&T/SPIE Electronic Imaging*, pages 1–7, 2014.

[11] A. T. Phan Ho, B. A. Mai Hoang, W. Sawaya, and P. Bas. Document authentication using graphical codes: impacts of the channel model. In *Proceedings of the first ACM workshop on Information hiding and multimedia security*, pages 87–94. ACM, 2013.

[12] J. Picard, C. Vielhauer, and N. Thorwirth. Towards fraud-proof id documents using multiple data hiding technologies and biometrics. *SPIE Proceedings–Electronic Imaging, Security and Watermarking of Multimedia Contents VI*, pages 123–234, 2004.

[13] A. W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[14] R. Villan, S. Voloshynovskiy, O. Koval, and T. Pun. Multilevel 2 d bar codes: toward high-capacity storage modules for multimedia security and management. In *Proc. SPIE*, volume 5681, pages 453–464, 2005.

[15] T. H. Wonnacott and R. J. Wonnacott. *Introductory statistics*, volume 19690. Wiley New York, 1972.