A GAME THEORETIC APPROACH TO SOURCE IDENTIFICATION WITH KNOWN STATISTICS

M. Barni

University of Siena, Dept. of Information Engineering, Via Roma 56, Siena, Italy
barni@dii.unisi.it

ABSTRACT

In the attempt to lay the basis for the construction of a theoretical framework to cast forensics and anti-forensics techniques in, we introduce a game-theoretic model for the source-identification problem with known statistics. The framework is used to derive the Nash equilibrium for an asymptotic version of the game, in which the players' strategies and the payoff are defined in terms of the error exponents of the false positive and false negative probabilities. The payoff at the equilibrium is evaluated and the conditions under which the false negative error probability tends to zero derived.

1. INTRODUCTION

Research in multimedia forensics has recently start focusing on so-called anti-forensics techniques, i.e. techniques aiming at impeding making forensics analysis. As a result, we now know that most early multimedia forensics techniques do not work properly if some simple countermeasures are taken to delete the traces left by the acquisition device or the processing tool that has been used to create a forgery [1, 2]. Researchers have thus started developing tools to detect the traces left by anti-forensic algorithms [3], so to restore the credibility of forensics analysis. While this iterative loop will finally lead to powerful forensics and anti-forensics tools, the need to investigate the ultimate limits of forensics (and anti-forensics) techniques clearly exists. In this paper we move a first step in this direction, by laying the basis for a theoretical analysis of one of the most studied problems in multimedia forensics, namely the source identification problem. By relying on methods typical of game theory and information theory, we propose a rigorous framework that can be used to model the source identification problem and solve it in some particular cases. While we recognize that our analysis does not account for all the subtleties involved in real forensics analysis and that the statistical models adopted in our investigation do not account for the complexity of real signals, we believe the proposed framework to be a fundamental step towards the definition of more complex and realistic models.

2. NOTATION AND DEFINITIONS

In the rest of this work we will use capital letters to indicate scalar random variables, whose specific realizations will be represented by the corresponding lower case letters. Sequences of length n will be indicated either by X^n or x^n according to their random or deterministic nature; X_i, x_i, will indicate the $i-$th element of X^n and x^n respectively. Information sources will also be defined by capital letters. Calligraphic capital letters (e.g. $\mathcal{X}$) will be used to denote the alphabet of information sources. The probability density function (pdf) of a random variable X will be denoted by P_X. The same notation will be used to indicate the probability measure ruling the emission of sequences from a source X, so we will use the expressions $P_X(a)$ and $P_X(x^n)$ to indicate, respectively, the probability of symbol $a \in \mathcal{X}$ and the probability that the source X emits the sequence x^n. Given an event A, we will use the notation $P_X(A)$ to indicate the probability of the event A under P_X. Given two sequences x^n and y^n, asymptotic equality in the logarithmic scale will be indicated as $x^n \doteq y^n$, meaning that:

$$\lim_{m\to\infty} \frac{1}{m} \log\left(\frac{x_m}{y_m}\right) = 0. \tag{1}$$

Throughout the paper we make extensive use of the concept of type class defined as follows [4]. Let $\hat{P}_{x^n}$ indicate the empirical probability distribution induced by the sequence $x^n \in \mathcal{X}^n$, i.e. $\hat{P}_{x^n}(a) = \frac{1}{n}\sum_{i=1}^n \delta(x_i, a)$, with $\delta(x_i, a) = 1$ if $x_i = a$ and zero otherwise. The type class $T(x^n)$ of x^n, is the set of sequences $\tilde{x}^n \in \mathcal{X}^n$ such that $\hat{P}_{\tilde{x}^n} = \hat{P}_{x^n}$. Given a type class T, we will indicate by $\hat{P}_T$ the empirical probability density function induced by the sequences in T. The Kullback-Leibler (KL) divergence between two distributions P and Q defined on the same finite alphabet $\mathcal{X}$ is given by:

$$\mathcal{D}(P||Q) = \sum_{a\in\mathcal{X}} P(a) \log \frac{P(a)}{Q(a)}, \tag{2}$$

where, as usual, $0 \log 0 = 0$ and $p \log p/0 = \infty$ if $p > 0$.

2.1. Game theory

Game theory is a branch of mathematics devoted to the analysis of strategic situations, referred to as games, in which the success of one player depends on the choices made by the other players. Here we are concerned with the class of strategic, 2-players, zero-sum games. In this setup, a game is defined as a 4-uple $G(\mathcal{S}_1, \mathcal{S}_2, u_1, u_2)$, where $\mathcal{S}_1 = \{s_{1,1} \dots s_{1,n_1}\}$ and $\mathcal{S}_2 = \{s_{2,1} \dots s_{2,n_2}\}$ are the set of strategies (actions) the first and the second player can choose from, and $u_l(s_{1,i}, s_{2,j}), l = 1, 2$ is the payoff of the game for player l, when the first player chooses the strategy $s_{1,i}$ and the second chooses $s_{2,j}$. A pair of strategies $s_{1,i}$ and $s_{2,j}$ is called a profile. In a zero-sum competitive game the win of a player is equal to the loss of the other, so we have $u_1(s_{1,i}, s_{2,j}) + u_2(s_{1,i}, s_{2,j}) = 0$. In this case, without loss of generality, we can specify the payoff of the first player (generally indicated by u), with the understanding that the payoff of the second player u_2 is equal to $-u$. A common goal in game theory is to determine the existence of equilibrium points. The most common definition of equilibrium is the one due by Nash [5]. For the particular case of a 2-player game, a profile (s_{1,i^*}, s_{2,j^*}) is a Nash equilibrium if:

$$\begin{array}{ll} u_1((s_{1,i^*}, s_{2,j^*})) \ge u_1((s_{1,i}, s_{2,j^*})) & \forall s_{1,i} \in \mathcal{S}_1 \\ u_2((s_{1,i^*}, s_{2,j^*})) \ge u_2((s_{1,i^*}, s_{2,j})) & \forall s_{2,j} \in \mathcal{S}_2, \end{array} \tag{3}$$

This work was partially supported by the REWIND Project, funded by the Future and Emerging Technologies (FET) programme within the 7FP of the EC, under grant 268478.

where for a zero-sum game $u_2 = -u_1$. In practice, a profile is a Nash equilibrium if each player does not have any interest in changing its choice assuming the other does not change its strategy.

3. SOURCE IDENTIFICATION WITH KNOWN SOURCE

Our definition of the Source Identification (SI) game starts by observing that the task of the Forensics Analyst (FA) is the definition of a test to accept or reject the hypothesis that the sequence under analysis was produced by a certain source X. On the other side, the goal of the adversary (AD) is to take a sequence generated by a different source and modify it in such a way that the FA accepts the hypothesis that the modified sequence has been generated by X. In doing so the AD may want to minimize the amount of modifications it has to introduce to deceive the FA. We assume that both the FA and the AD know the source X. While this may seem an oversimplified assumption, the resulting scenario already contains all the ingredients necessary to define a well-posed multimedia forensics game and represents a starting point for the analysis of more realistic situations. As to the source Y, we initially assume that it is known to both the FA and the AD. In a second step, we will see that, at least in the asymptotic version of the game, the assumption that the FA knows Y may be removed. Let, then, $X \simeq P_X$ and $Y \simeq P_Y$ be two known sources with finite alphabet $\mathcal{X}$. Let y^n be a sequence drawn from Y and let z^n be a modified version of y^n produced by the AD. Let H_0 be the hypothesis that the test sequence has been generated by X, and let H_1 be the opposite hypothesis that the sequence has been generated by Y. We define the source identification game under the known source assumption (SI_{ks}) as follows.

Definition 1. *The $SI_{ks}(\mathcal{S}_{FA}, \mathcal{S}_{AD}, u)$ game is a zero-sum, strategic, game played by the FA and the AD, defined by the following strategies and payoff.*

- *$\mathcal{S}_{FA}$ is the set of acceptance regions for H_0 for which the false positive probability (i.e. the probability of rejecting H_0 when H_0 is true) is below a certain threshold:*

$$\mathcal{S}_{FA} = \{\Lambda_0 : P_X(x^n \notin \Lambda_0) < P_{fp}\}, \tag{4}$$

  where Λ_0 is the acceptance region for H_0 and P_{fp} is a prescribed maximum false positive probability[1].

- *$\mathcal{S}_{AD}$ is formed by all the functions that map a sequence y^n produced by Y into a new sequence z^n subject to a distortion constraint:*

$$\mathcal{S}_{AD} = \{f(y^n) : d(y^n, f(y^n)) \leq nD\}, \tag{5}$$

  where $d(\cdot,\cdot)$ is a proper distance function and D is the maximum allowed per-letter distortion.

- *The payoff function is defined in terms of false negative error probability (P_{fn}), namely:*

$$u(\Lambda_0, f) = -P_{fn} = -\sum_{y^n : f(y^n) \in \Lambda_0} P_Y(y^n). \tag{6}$$

The above definition also clarifies why we had to assume that the FA knows P_Y. In fact, this is a necessary assumption, since for a proper definition of the game it is required that both players have a full knowledge of the payoff for all possible profiles. Solving the SI_{ks} game as stated in definition 1 is a cumbersome task, hence we introduce an asymptotic version of the game that can be solved more easily by resorting to methods typical of information theory. To do so, we replace the error probabilities of the two kinds with the corresponding asymptotic quantities. More specifically, the definition of the new version of the game relies on the error exponents of P_{fp} and P_{fn} defined as:

$$\varepsilon_{fp} = -\lim_{n\to\infty} \frac{1}{n} \log P_{fp}, \;\; \varepsilon_{fn} = -\lim_{n\to\infty} \frac{1}{n} \log P_{fn}. \tag{7}$$

We can now introduce the asymptotic version of the SI_{ks} game.

Definition 2. *The $SI_{ks}^{\infty}(\mathcal{S}_{FA}, \mathcal{S}_{AD}, u)$ game is a game between the FA and the AD defined by the following strategies and payoff:*

$$\mathcal{S}_{FA} = \{\Lambda_0 : \varepsilon_{fp} \geq \lambda\}, \tag{8}$$

$$\mathcal{S}_{AD} = \{f(y^n) : d(y^n, f(y^n)) \leq nD\}, \tag{9}$$

$$u(\Lambda_0, f) = \varepsilon_{fn} \tag{10}$$

where λ is the minimum false error exponent admissible by the FA.

3.1. Equilibrium point of SI_{ks}^{∞} for FA with limited resources

Finding the equilibrium point(s) for the SI_{ks}^{∞} game is not trivial. The error probabilities of the two kinds for a given profile are not easy to calculate and even more difficult to optimize. For this reason we decided to face with this problem by limiting the kind of acceptance regions the FA can choose from. As in [6], we limit the complexity of the analysis carried out by the FA by confining it to depend on a limited set of statistics computed on the test sequence. To simplify the analysis, in the subsequent derivation we assume that the sources X and Y are memoryless, however our arguments can be extended to other source classes as outlined in section 3.4. Given the memoryless nature of X and Y, it makes sense to require that the FA bases its decision by relying only on $\hat{P}_{x^n}$, i.e. on the empirical probability density function induced by the test sequence. Note that $\hat{P}_{x^n}$ is not a sufficient statistics for the FA; in fact, even if Y is a memoryless source, the AD could introduce some memory within the sequence as a result of the application of f. This is the reason why we need to introduce explicitly the requirement that the FA bases its decision only on $\hat{P}_{x^n}$. Our derivation starts with the following lemma.

Lemma 1. *Let Λ_1^* be defined as follows:*

$$\Lambda_1^* = \left\{ x^n : \mathcal{D}(\hat{P}_{x^n} || P_X) \geq \lambda - |\mathcal{X}| \frac{\log(n+1)}{n} \right\} \tag{11}$$

and let Λ_0^* be the corresponding acceptance region. Then we have:

1. *$\varepsilon_{fp} \geq \lambda$,*
2. *for every $\Lambda_0 \in \mathcal{S}_{FA}$ (with $\mathcal{S}_{FA}$ defined as in equation (8)) we have $\Lambda_1 \subseteq \Lambda_1^*$.*

Proof. The proof of the lemma is easily obtained by adopting the same approach used in [6] (section II, Theorem 1), and is omitted here for sake of brevity. □

The first point says that Λ_0^* defines a valid strategy for the FA, while the second one implies the optimality of Λ_1^*. In fact, if for a certain strategy of the AD, $z^n \notin \Lambda_1^*$, *a fortiori* we have that $z^n \notin \Lambda_1$ for any other choice of Λ_1 hence resulting in a higher P_{fn}. An interesting consequence of lemma 1 is that the optimum strategy for the FA does not depend on: i) the strategy chosen by the AD, and ii) P_Y, i.e. the optimum strategy is universally optimal across all the sources Y. As we anticipated, this result makes the assumption

[1] Similarly we will indicate by $\Lambda_1 = \Lambda_0^c$ the acceptance region for H_1.

that the FA knows P_Y un-necessary. We also observe that the strategy expressed by equation (11) has a simple heuristic interpretation: the FA will accept only the sequences whose empirical pdf is close enough (in divergence terms) to P_X.

We now pass to the determination of the optimum strategy for the AD. Since the acceptance region is fixed, the AD can optimize its strategy by assuming that $\Lambda_0 = \Lambda_0^*$. We start by observing that the goal of the AD is to maximize P_{fn}. Such a goal is obtained by trying to bring the sequences produced by Y within Λ_0^*. In doing so the AD must only respect the constraint that $d(y^n, f(y^n)) \le nD$. The optimum strategy for the AD can then be expressed as follows:

$$f^*(y^n) = \arg \min_{z^n : d(z^n, y^m) \le nD} \mathcal{D}(\hat{P}_{z^n} || P_X). \tag{12}$$

Together with lemma 1, the above observation permits to state the first fundamental result of the paper.

Theorem 1. *The profile (Λ_0^*, f^*) defined by lemma 1 and equation (12) defines a Nash equilibrium for the SI_{ks}^{∞} game.*

Proof. Adapting equation (3) to the case at hand yields:

$$u(\Lambda_0^*, f^*) \ge u(\Lambda_0, f^*) \;\; \forall \Lambda_0 \in \mathcal{S}_{FA} \tag{13}$$

$$-u(\Lambda_0^*, f^*) \ge -u(\Lambda_0^*, f) \;\; \forall f \in \mathcal{S}_{AD}. \tag{14}$$

By remembering that for the SI_{ks}^{∞} game u is the false negative error exponent, the inequality (13) derives immediately from lemma 1. In the same way, since f^* maximizes the false negative error probability given Λ_0^*, the inequality (14) is always verified. □

3.2. Payoff at the equilibrium

The next step is the computation of the payoff at the equilibrium, i.e. the best achievable ε_{fn} for the FA. Given the asymptotic nature of the game, it is easy to foresee that P_{fn} will either tend to 0 (strictly positive ε) or to 1 ($\varepsilon = 0$) for $n \to \infty$ depending on the relationship between the maximum allowed distortion and the KL-divergence between P_X and P_Y. In this framework we are interested in understanding the conditions under which P_{fn} tends to 0, and the value of ε_{fn} in this case[2].

Let Γ_{fn} be the set of sequences generated by Y that can be moved into Λ_0^* since they are close enough to such a set, and let Γ_{fn}^c be the complement of Γ_{fn}. We can write:

$$\Gamma_{fn} = \{ y^n : \exists \, z^n \in \Lambda_0^* : d(y^n, z^n) \le nD \}. \tag{15}$$

The false negative error probability is clearly equal to the probability that $y^n \in \Gamma_{fn}$. The problem of determining Γ_{fn} is complicated by the adoption of different distance measures for the definition of Λ_0^* and to specify the distortion constraint. Interesting conclusions can be drawn when the Hamming distance is used to evaluate $d(y^n, z^n)$. To do so we rely on the following lemma.

Lemma 2. *The set:*

$$\Gamma_{fn} = \{ y^n : \exists z^n \in \Lambda_0^* : d_H(y^n, z^n) \le nD_H \}$$

is still a union of type classes. Specifically we have:

$$\Gamma_{fn} = \Gamma^* = \left\{ T : \exists T' \in \Lambda_0^* : ||\hat{P}_T - \hat{P}_{T'}||_{L_1} \le 2D_H \right\}, \tag{16}$$

where the L_1 distance between $\hat{P}_T$ and $\hat{P}_{T'}$ is defined as:

$$d_{L_1}(\hat{P}_T, \hat{P}_{T'}) = ||\hat{P}_T - \hat{P}_{T'}||_{L_1} = \sum_{a \in \mathcal{X}} |\hat{P}_T(a) - \hat{P}_{T'}(a)|. \tag{17}$$

Proof. We start by proving that a sequence whose empirical pdf has a distance larger than $2nD_H$ from the pdf of all the sequences in Λ_0^* can not belong to Γ_{fn}. Let y^n and z^n be two sequences, and let $\hat{P}_{y^n}$ and $\hat{P}_{z^n}$ be their empirical pdf's. The L_1 distance between $\hat{P}_{y^n}$ and $\hat{P}_{z^n}$ can be rewritten as follows:

$$\begin{aligned} ||\hat{P}_{y^n} - \hat{P}_{z^n}||_{L_1} &= \sum_{a \in \mathcal{X}^+} [\hat{P}_{y^n}(a) - \hat{P}_{z^n}(a)] \\ &\quad + \sum_{a \in \mathcal{X}^-} [\hat{P}_{z^n}(a) - \hat{P}_{y^n}(a)] \\ &= 2 \sum_{a \in \mathcal{X}^+} [\hat{P}_{y^n}(a) - \hat{P}_{z^n}(a)], \end{aligned} \tag{18}$$

where $\mathcal{X}^+$ (res. $\mathcal{X}^-$, $\mathcal{X}^=$) indicates the set of a's for which $\hat{P}_{y^n}(a) > \hat{P}_{z^n}(a)$ (res. $\hat{P}_{y^n}(a) < \hat{P}_{z^n}(a)$, $\hat{P}_{y^n}(a) = \hat{P}_{z^n}(a)$), and where the last equality follows from the observation that:

$$\sum_{a \in \mathcal{X}^-} \hat{P}_{y^n}(a) = 1 - \sum_{a \in \mathcal{X}^+} \hat{P}_{y^n}(a) - \sum_{a \in \mathcal{X}^=} \hat{P}_{y^n}(a). \tag{19}$$

Let us consider now the Hamming distance between the sequences y^n and z^n. By considering $\mathcal{X}^+$, we see that $d_H(y^n, z^n)$ is larger than or equal to $\sum_{a \in \mathcal{X}^+} n[\hat{P}_{y^n}(a) - \hat{P}_{z^n}(a)]$. In fact, for each $a \in \mathcal{X}^+$, there must be at least $n[\hat{P}_{y^n}(a) - \hat{P}_{z^n}(a)]$ positions in which the sequences y^n and z^n differ, so to justify the presence of $n[\hat{P}_{y^n}(a) - \hat{P}_{z^n}(a)]$ more a's in y^n than in z^n, thus yielding $n||\hat{P}_{y^n} - \hat{P}_{z^n}||_{L_1} \le 2d_H(y^n, z^n)$. For the sequences y^n that do not satisfy equation (16), we have $||\hat{P}_{y^n} - \hat{P}_{z^n}||_{L_1} > 2D_H$, $\forall z^n \in \Lambda_0^*$, yielding

$$2D_H < ||\hat{P}_{y^n} - \hat{P}_{z^n}||_{L_1} \le \frac{2d_H(y^n, z^n)}{n}, \tag{20}$$

proving that $\Gamma_{fn} \subseteq \Gamma^*$. We now show that $\Gamma^* \subseteq \Gamma_{fn}$. Let y^n be a sequence in Γ^*. Then there exists a type class $T' \in \Lambda_0^*$ whose pdf has an L_1 distance from $\hat{P}_{y^n}$ lower than or equal to $2D_H$. Starting from y^n we can easily build a new sequence z^n whose type is equal to $\hat{P}_{T'}$ by proceeding as follows. Let $\mathcal{X}^+$ be the set of a's for which $\hat{P}_{y^n}(a) > \hat{P}_{T'}(a)$. For each $a \in \mathcal{X}^+$ we take $n[\hat{P}_{y^n}(a) - \hat{P}_{T'}(a)]$ positions where $y_i = a$, and replace a with a value $b \in \mathcal{X}^-$, in such a way that at the end we have $\hat{P}_{z^n}(a) = \hat{P}_{T'}(a) \; \forall a \in \mathcal{X}$. Note that this is possible since we have

$$\sum_{a \in \mathcal{X}^+} [\hat{P}_{y^n}(a) - \hat{P}_{T'}(a)] = \sum_{b \in \mathcal{X}^-} [\hat{P}_{T'}(b) - \hat{P}_{y^n}(b)]. \tag{21}$$

Since to pass from y^n to z^n we modified only $\sum_{a \in \mathcal{X}^+} n[\hat{P}_{y^n}(a) - \hat{P}_{T'}(a)]$ positions of y^n we have:

$$\begin{aligned} d_H(y^n, z^n) &= \sum_{a \in \mathcal{X}^+} n[\hat{P}_{y^n}(a) - \hat{P}_{T'}(a)] \\ &= \frac{n||\hat{P}_y^n - \hat{P}_{T'}||_{L_1}}{2} \le nD_H, \end{aligned} \tag{22}$$

showing that $y^n \in \Gamma_{fn}$, and hence $\Gamma^* \subseteq \Gamma_{fn}$, thus concluding the proof of the lemma. □

Lemma 2 permits to understand if P_{fn} tends to 0 or 1. To do so, we introduce the set Γ_{fn}^{∞} defined as the union of the empirical pdf's contained in Γ_{fn} for all n. We can distinguish two cases: P_Y may either belong to Γ_{fn}^{∞} or not. In the first case P_{fn} tends to 1, otherwise

[2] In the same way we could investigate how fast the probability of a correct decision tends to zero when $\varepsilon_{fn} = 0$. Such an analysis follows exactly the same lines we will use for the computation of ε_{fn} and will not be detailed.

it tends to 0 and the probability that the FA does not distinguish original and fake sequences gets vanishingly small for increasing n. Specifically, the following theorem holds.

Theorem 2. *For the SI_{ks}^{∞} game with an FA with limited resources and Hamming distance, the payoff at the equilibrium is:*

1. $\varepsilon_{fn} = 0$, *if* $P_Y \in \Gamma_{fn}^{\infty}$;
2. $\varepsilon_{fn} = \arg \min\limits_{\hat{P} \in \Gamma_{fn}^{\infty}} \mathcal{D}(\hat{P}||P_Y)$, *if* $P_Y \notin \Gamma_{fn}^{\infty}$.

Proof. Point 1 derives from the law of large numbers. Point 2 is a consequence of lemma 2 and Sanov's theorem ([4], chapt. 12). □

Theorem 2 states that when $P_Y \in \Gamma_{fn}^{\infty}$ the probability of a correct decision by the FA tends to zero. The asymptotic rate by which such a probability tends to zero can again be obtained by invoking Sanov's theorem.

3.3. Bernoulli sources

Let $X \sim \mathcal{B}(p)$ and $Y \sim \mathcal{B}(q)$ be Bernoulli sources with parameters p and q respectively. In this case the acceptance region for H_0 assumes a very simple form. In fact, the KL-divergence between $\hat{P}_{x^n}$ and P_X depends only on the number of 1's in x^n, the divergence being a monotonic increasing function of $|n_x(1)/n - p|$, where we indicated with $n_x(1)$ the number of 1's in x^n. As a consequence the acceptance region may be defined in terms of $n_x(1)$ only:

$$\Lambda_0^* = \{x^n : n_x(1) \in [n_{inf}(\lambda), n_{sup}(\lambda)]\}, \tag{23}$$

with n_{inf} and n_{sup} derive from the equality[3] $\mathcal{D}(\hat{P}_{x^n}||P_X) = \lambda - |\mathcal{X}|\log(n+1)/n$. The optimum strategy of the AD is also easy to define. Given the monotonic nature of the KL-divergence noted above, the AD will increase (decrease) the number of 1's in y^n to make the relative frequency of 1's in z^n as close as possible to p. The AD will succeed in inducing a decision error if the number of ones in z^n belongs to the interval $[n_{inf}, n_{sup}]$. Since the distortion constraint states that $d(y^n, z^n) \leq nD_H$, we have:

$$\Gamma_{fn} = \{y^n : n_y(1) \in [n_{inf}(\lambda) - nD_H, n_{sup}(\lambda) + nD_H]\}, \tag{24}$$

with the boundaries of the interval truncated to 0 or n when needed. For the computation of the payoff of the game at the equilibrium we may distinguish 2 cases:

- $q \in \Gamma_{fn}^{\infty} = [\nu_{inf} - D_H, \nu_{sup} + D_H]$;
- $q \notin \Gamma_{fn}^{\infty} = [\nu_{inf} - D_H, \nu_{sup} + D_H]$,

where ν_{inf} and ν_{sup} are obtained by letting $\mathcal{D}(\hat{P}||P_X) = \lambda$. In the first case $\varepsilon_{fn} = 0$ and P_{fn} tends to 1 for $n \to \infty$, in the second case P_{fn} tends to 0 for $n \to \infty$ and the error exponent can be computed by resorting to Sanov's theorem. Let us suppose for instance that $q > \nu_{sup} + D_H$. The type in Γ_{fn}^{∞} closest to P_Y in divergence is a Bernoulli source with parameter $p^* = \nu_{sup} + D_H$, and hence the payoff of the game will be $\varepsilon_{fn} = \mathcal{D}(\mathcal{B}(p^*)||\mathcal{B}(q))$.

3.4. Non-binary sources and sources with memory

Finding a closed-form solution for the case of multi-valued sources and distances other than the Hamming distance seems a prohibitive task. While the formula defining the optimum acceptance region does not change and can be easily implemented by the FA, the task of the AD is more complex due to the necessity of solving the minimization problem in (12). In this case, the resort to numerical methods appears unavoidable.

The existence of a Nash equilibrium for the SI_{ks}^{∞} game has been proved by assuming that the FA is restricted to base its analysis on the empirical pdf of the test sequence. This assumption makes sense for the class of DMS sources whose characteristics are completely described by first order statistics, but is no more reasonable for sources with memory. A closer inspection of the methods used in sections 3.1, and 3.2, however, reveals that the analysis carried out therein can be extended to sources with memory, as long as the concepts of types and type classes can still be used. This is the case, for instance, of finite-order Markov sources, a model that is commonly used to described a wide variety of sources with memory. In fact, it is known that for this kind of sources, the number of type classes grows polynomially with n [7], hence making the extension of our analysis straightforward. Renewal processes are another class of sources that is amenable to be analyzed by relying on the concept of types. Renewal processes [8] can be used, for instance, to model run length sequences and hence could be of interest in forensics problems dealing with compressed streams adopting run-length coding (e.g. the JPEG coding standard). In [8], it is shown that the number of type classes of renewal processes grows sub-exponentially with n, thus opening the way to the extension of our analysis to this class of sources.

4. CONCLUSIONS

The definition of the SI_{ks} and SI_{ks}^{∞} games, and the derivation of the Nash equilibrium of SI_{ks}^{∞}, represent a first step towards the construction of a rigorous theoretical framework to cast multimedia forensics and anti-forensics in. While the theoretical models will never be able to capture all the details encompassed by real multimedia forensics, we believe that they can highlight the basic trade-offs involved in forensics analysis in the presence of an adversary, and be a useful tool to guide future research in this area. In a future work we will focus on the extension of the results presented in this paper to more realistic models, e.g. Markov sources, or continuous sources, and on the definition of the source identification game when the sources are known only through the availability of training data.

5. REFERENCES

[1] M. Kirchner and R. Bohme, "Hiding traces of resampling in digital images," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 582–292, December 2008.

[2] M. C. Stamm and K. J. R. Liu, "Anti-forensics of digital image compression," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 50–65, September 2011.

[3] M. Goljan, J. Fridrich, and M. Chen, "Sensor noise camera identification: countering counter forensics," in *SPIE Conference on Media Forensics and Security, San Jose, CA*, 2010.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley Interscience, 1991.

[5] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. MIT Press, 1994.

[6] N. Merhav and E. Sabbag, "Optimal watermark embedding and detection strategies under limited detection resources," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 255–274, January 2008.

[7] I. Csiszar, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, October 1998.

[8] I. Csiszar and P. C. Shields, "Redundancy rates for renewal and other processes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 2065–2072, November 1996.

[3] Note that we may have $n_{inf} = 0$ and/or $n_{sup} = n$, since the equality may admit a solution only for $n_x(1) > np$, $n_x(1) < np$, or no solution.