A Heuristic Attack Method to PRH-Based Audio Copy Detectors

Igor Bisio, Carlo Braccini, Alessandro Delfino, Fabio Lavagetto, and Mario Marchese

***Abstract*—Often copyrighted multimedia files are uploaded and shared online. To avoid the unregulated spread of such material many copy detectors have been developed in order to deny the possibility to upload, and consequently make available, copies of copyrighted contents. A widely referenced fingerprint method for content-based audio identification is the Philips Robust Hash (PRH) [1]. This paper introduces a simple but effective attack technique capable to defeat a PRH fingerprint-based audio copy detector without significantly affecting the signal quality. It is a heuristic method that adds a suitable distortion to the original audio signal, so that the modified signal is not detected as a copy of the original one but is perceptively very similar to it. The quality of the modified signal has been evaluated in terms of a distortion measure based on a mathematical model of the human auditory system and of the Peak Signal-to-Noise Ratio (PSNR). The attack method has shown a promising success rate.**

***Index Terms*—Adversarial signal processing, audio fingerprint, copy detection attack, copyright protection.**

I. Introduction

RECENTLY, the availability of multimedia contents has had an exponential growth, facilitated by web services which allow the users to upload and share multimedia contents. User-uploaded contents are often copyrighted material which cannot be spread in unregulated ways. To prevent the uncontrolled diffusion of such material, many content-based copy detection methods have been developed with the aim of checking, every time a new content is uploaded, if it is a copy of a copyrighted content. The most used content-based audio copy detection methods are based on intrinsic characteristics of the audio signal, called audio fingerprint. They are used for different purposes such as: *i*) identifying a song from a short noisy recording [2]; *ii*) clustering recordings according to genre, rhythm, etc… [3]; *iii*) recognizing the live TV channel a user is watching through a short noisy recording of the TV audio [4]; *iv*) identifying copies of the same audio file in a large database [1], [5]. PRH [1] is an important audio fingerprint-based copy detector exploiting the audio signal energy distribution. It is widely referenced and represents a foundation for a set of currently used energy-based copy detectors such as [6], [7].

The aim of this paper is to exploit the weaknesses of PRH-based detector to design an iterative procedure able to deceive them. To do so, we add to the original audio signal a distortion signal that does not significantly affect the perceived audio quality but makes the copy identification system fail. In recent years, much interest has been gathered by the adversarial approach (see [8] for a general framework) applied to forensic detectors, especially by methods aimed at deceiving tampered images detectors. The approaches described in the literature may be classified as: *i*) targeted methods, such as [9] and [10], histogram based; [11], SIFT-based; and [12], BoW-based; *ii*) universal methods, such as [13], [14] (both histogram-based), and [15]. Being based on PRH weaknesses, the targeted attack method introduced here may not be able to tackle other audio fingerprint systems, such as [2], [5], [16]. Actually, this paper is the first step of a wider research aimed at tackling the adversarial problem by defining optimal strategies for attacker and defender. Detailing the features of PRH and evidencing how PRH can be deceived without significantly affecting audio quality can help define both possible countermeasures able to protect the signal and more robust fingerprints suitable to shield the detector from the attach. These important activities are object of ongoing research.

II. Philips Robust Hash (PRH) Copy Detector

PRH computes the audio fingerprint of an audio signal by dividing it in partially overlapped frames of approximately 0.37 s. Each frame is shaped by a Hanning window. The overlap factor is 31/32. The Fourier Transform of each frame is divided into 33 non-overlapped bands equally spaced on a logarithmic scale, and, for each band, the energy of the signal is computed, leading to the energy matrix representation $E(n,m)$, where n is the time frame index and m is the index of the frequency bin. Defining $V(n,m)$ as:

$$\begin{aligned} V(n,m) = [E(n+1,m)+E(n,m+1)] &+ \\ - [E(n+1,m+1)+E(n,m)]& \end{aligned} \tag{1}$$

the computation of the fingerprint $H(n,m)$ is done by coding with one bit the sign of $V(n,m)$ as specified in (2):

$$H(n,m) = \begin{cases} 1 & \text{if } V(n,m) \geq 0, \\ & \qquad\qquad n \in [1,N], m \in [1,M] \\ 0 & \text{if } V(n,m) < 0, \end{cases} \tag{2}$$

Usually copy detection methods work on large databases (e.g. hundreds of thousands of songs), thus making the brute force search (i.e., comparing the input audio with all the database songs) unfeasible. For this reason, in [1] a more efficient search algorithm is proposed, that selects a group of candidates with high probability of being the best matching

Manuscript received May 28, 2014; revised July 17, 2014; accepted July 19, 2014. Date of publication October 15, 2014; date of current version October 30, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Constantine L. Kotropoulos.

The authors are with the Department of Electrical, Electronic, and Telecommunications Engineering, and Naval Architecture (DITEN), Polytechnic School, University of Genova, 16145 Genova, Italy (e-mail: igor.bisio@unige.it; carlo.braccini@unige.it; alessandro.delfino@unige.it; fabio.lavagetto@unige.it; mario.marchese@unige.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2363592

fingerprint in the database through a Look Up Table (LUT) that speeds up the search algorithm. Being $x(k)$ the audio signal stored in the database and $H_x(n,m)$ its fingerprint, $c(k)$ a generic copy and $H_c(n,m)$ its fingerprint, this LUT-based search method selects those fingerprints in the database, which have at least one row, indexed with n_r, equal to the corresponding row of H_c. Denoting such row with $H^{n_r}(m)$, where $H^{n_r}(m) = H(n_r,m), \forall m \in [1,M]$, the condition can be written as $H_x^{n_r} = H_c^{n_r}$ or, equivalently, in terms of Hamming distance, $D_{Ham}(H_x^{n_r}, H_c^{n_r}) = 0$. It has been proven [1] that this condition holds if $c(k)$ is only a "mild" degradation of $x(k)$. Alternatively, this condition can be relaxed by imposing that the closest (in terms of Hamming distance) corresponding rows, indexed with n_c, have Hamming distance below a given threshold δ:

$$D_{Ham}(H_x^{n_c}, H_c^{n_c}) \leq \delta \tag{3}$$

After getting the list of the best candidates' fingerprints, the copy detection system decides that $c(k)$ is the copy of one of the candidates if their fingerprints are similar. The similarity measure employed in [1] is the Bit Error Rate (BER), defined as the number of bits for which the two fingerprints differ divided by the total number of bits:

$$BER(H_x, H_c) = \frac{1}{NM} D_{Ham}(H_x, H_c) \tag{4}$$

with

$$D_{Ham}(H_x, H_c) = \sum_{n=1}^{N} \sum_{m=1}^{M} H_x(n,m) \oplus H_c(n,m) \tag{5}$$

with $\oplus$ indicating the $exclusive\text{–}or$ operator, N the number of rows and M the number of columns of the fingerprints. If the BER value is below a pre-determined threshold γ,

$$BER(H_x, H_c) < \gamma \tag{6}$$

the system identifies the two audio recordings as copies.

III. Definition of the Attack

The proposed attack method relies on the assumption that the used fingerprint based identification system is PRH [1]. There are two possible cases in which this detector fails:

a) when the LUT-based search does not retrieve H_x in the candidate list;
b) when the LUT-based search retrieves H_x in the candidate list but the BER is higher than the threshold γ.

Both cases can be exploited as weaknesses by an attacker to deceive the system. Case *a*), called Fingerprint Not Retrieved (FNR), happens if, for each row n, $H_c(n,m)$ has Hamming distance from $H_x(n,m)$ above δ ($\delta = 0$ when "mild" degradations are assumed), i.e., $D_{Ham}(H_x^n, H_c^n) > \delta, \forall n \in [1,N]$. Case *b*), called BER Above Threshold (BAT), happens when the Hamming distance (and consequently the BER) between the fingerprints is higher than a threshold, i.e., $BER(H_x, H_c) \geq \gamma$. The attack problem can be phrased as finding the distortion signal $\alpha(k)$, within a suitably defined set, such that $x(k)+\alpha(k)$ is classified as different from $x(k)$ by the copy identification system while they are perceptively very similar. Denoting with $d(y,z)$ an objective measure of the perceived distance between two audio signals $y(k)$ and $z(k)$, and $D(H_y, H_z)$ a distance measure between their fingerprints H_y and H_z, the problem can be expressed as:

$$\begin{aligned} \underset{\alpha(k)}{\text{minimize}} \quad & d(x, x+\alpha) \\ \text{subject to} \quad & D(H_x, H_{x+\alpha}) > \theta \end{aligned} \tag{7}$$

In the FNR case $\theta = \delta$ and the distance measure $D(H_y, H_z)$ is the Hamming distance between the closest (already denoted with the index n_c) corresponding fingerprint rows:

$$D(H_x, H_{x+\alpha}) = D_{Ham}(H_x^{n_c}, H_{x+\alpha}^{n_c}) \tag{8}$$

where

$$n_c = \arg\min_{n} \; D_{Ham}(H_x^n, H_{x+\alpha}^n), \forall n \in [1,N] \tag{9}$$

In the BAT case $\theta = \gamma$ and the distance measure is the BER between the two fingerprints defined in (4).

$$D(H_x, H_{x+\alpha}) = BER(H_x, H_{x+\alpha}) \tag{10}$$

The idea is acting iteratively as specified in the next Section, so to modify the signal $x(k)$ at each iteration through the distortion signal $\alpha(k)$ until $D_{Ham}(H_x^{n_c}, H_{x+\alpha}^{n_c}) > \delta \, \forall n_c$ (FNR attack) or until $BER(H_x, H_{x+\alpha}) \geq \gamma$ (BAT attack). The two inequalities are the stopping criteria of the two different attacks. In practice, the FNR attack modifies the signal $x(k)$ at the source, so that the PRH receiving $x(k)$ cannot find H_x in the candidate list deriving from the LUT-based search; the BAT attack modifies the signal $x(k)$ so that the PRH, even if it may retrieve H_x in the candidate list, does not identify any candidate as a copy. It is important to remark that, even if the PRH checks the BER only after retrieving H_x in the candidate list, when the BAT attack is performed it is not known in advance whether or not the LUT-based search of the PRH receiving the modified signal will retrieve H_x in the candidate list. This does not impact the performance because, also in the BAT case, if H_x is not in the candidate list, the PRH is deceived.

It is worth noting how the ratio $\frac{\delta}{\gamma}$ affects the possible mutual implications between FNR and BAT. Specifically, if $\frac{\delta}{\gamma} \geq M$ then FNR $\Rightarrow$ BAT: being $N\delta \geq NM\gamma$ and $D_{Ham}(H_x^{n_c}, H_{x+\alpha}^{n_c}) > \delta \, \forall n_c$, it is true that $BER(H_x, H_c) = \frac{1}{NM}\sum_{n_c=1}^{N} D_{Ham}(H_x^{n_c}, H_{x+\alpha}^{n_c}) > \frac{\delta}{M} \geq \gamma$. Vice versa, if FNR $\Rightarrow$ BAT then $\sum_{n_c=1}^{N} D_{Ham}(H_x^{n_c}, H_{x+\alpha}^{n_c}) > N\delta \geq \gamma NM$, thus $\frac{\delta}{\gamma} \geq M$. If BAT $\Rightarrow$ FNR then $\sum_{n_c=1}^{N} D_{Ham}(H_x^{n_c}, H_{x+\alpha}^{n_c}) \geq \gamma NM > N\delta$, thus $\frac{\delta}{\gamma} < M$. The condition $\frac{\delta}{\gamma} < M$ not necessarily implies BAT $\Rightarrow$ FNR. For the sake of generality, we describe the two attacks separately.

IV. The Iterative Attack Method

The following description refers to the attack through BAT. The aim is to find a solution method of the problem in (7), such that a closed form expression of the functional $d(x, x+\alpha)$ is not required. A heuristic solution is proposed based on the following iterative method: at the j-th iteration, a minimum energy distortion component $\alpha_j(k)$ is added to $x(k)$, such that at least one fingerprint bit change is produced, i.e., $D_{Ham}(H_x, H_{x+\alpha_j}) \geq 1$. The rationale under this heuristic procedure is that the minimum energy distortion signal $\alpha(k)$ found by the attack method is expected to introduce a small perceived distortion. Modelling and explicitly minimizing the perceptual distortion $d(x, x+\alpha)$ are deferred to further research advances. The distortion signal is the sum of all distortion components:

$$\alpha(k) = \sum_j \alpha_j(k), \quad \text{where} \quad \alpha_0(k) = 0, \forall k \tag{11}$$

The j-th iteration starts by finding the fingerprint bit requiring the minimum energy to be changed, i.e., by solving:

$$(n_j, m_j) = \underset{(n,m)\notin Q_j}{\arg\min}\, |{}^{j}V_{x+\alpha}(n,m)|,$$

where

$$\begin{aligned} Q_j = \{(n,m) : H_x(n,m) \neq {}^{j}H_{x+\alpha}(n,m), \\ \forall n \in [1,N], \forall m \in [1,M]\} \end{aligned} \tag{12}$$

${}^{j}V_{x+\alpha}(n,m)$ and ${}^{j}H_{x+\alpha}(n,m)$ are (1) and (2) computed at the j-th iteration on the signal $x(k) + \sum_{l=0}^{j-1} \alpha_l(k)$, and Q_j, computed at step j, is the set of all pairs (n,m) corresponding to the positions of the fingerprint bits already changed in previous steps up to $(j-1)$.

The distortion component α_j is a superposition of sinusoidal signals suitably selected to modify the four energy matrix elements involved in the computation of the fingerprint bit $H(n_j, m_j)$: $E(n_j, m_j)$, $E(n_j+1, m_j)$, $E(n_j, m_j+1)$, and $E(n_j+1, m_j+1)$. The aim of each added sinusoid is to increase the signal energy in order to change the sign of $V(n_j, m_j)$ and, consequently, the value of $H(n_j, m_j)$. Assuming $V(n_j, m_j) \geq 0$ (the extension to the case $V(n_j, m_j) < 0$ is trivial), so that $H(n_j, m_j) = 1$, attacking the PRH-based detector requires to flip the value of this bit. $V(n_j, m_j) \geq 0$ implies, from (1), that:

$$\begin{aligned} E(n_j+1, m_j) + E(n_j, m_j+1) \geq \\ E(n_j+1, m_j+1) + E(n_j, m_j) \end{aligned} \tag{13}$$

To invert the sign of $V(n_j, m_j)$ requires increasing $E(n_j+1, m_j+1)$ and $E(n_j, m_j)$, i.e., the energy of the signal in the (m_j+1)-th frequency bin of the (n_j+1)-th time frame and in the m_j-th frequency bin of the n_j-th frame, respectively. Two consecutive overlapped time frames (n_j and n_j+1) and two adjacent frequency bins are involved in the computation of $H(n_j, m_j)$. In PRH [1], both frames are multiplied by a Hanning window that causes the central samples of the frames to be less attenuated than the peripheral ones. Calling L_f the length of each frame, OL the overlap factor, ranging from 0 (no overlap) to 1 (total overlap), between two consecutive frames and $han(k)$ the L_f-long Hanning window, Fig. 1 shows: $han(k)$; its copy shifted by $L_f(1-OL)$, which is the distance between two consecutive frames; and $han(k) - han(k - L_f(1-OL))$ that illustrates how each sample contributes to the computation of the difference of consecutive time frames. In Fig. 1 the two Hanning functions are scaled by a 10 factor to better fit into the picture. The positions of maximum and minimum of the difference function are identified with k_{max} and k_{min}:

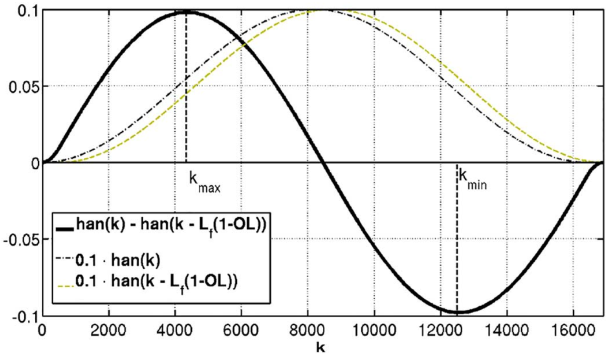

Fig. 1. Effects of the overlap of consecutive Hanning windows.

$$\begin{aligned} k_{max} &= \underset{k}{\arg\max}\,[han(k) - han(k - L_f(1-OL))] \\ k_{min} &= \underset{k}{\arg\min}\,[han(k) - han(k - L_f(1-OL))] \end{aligned} \tag{14}$$

Denoting with $k_h^{n_j}$ the first sample of the n_j-th frame, the additive sinusoidal signal must lay across $k_h^{n_j} + k_{max}$ and $k_h^{n_j} + k_{min}$ to increase the energy of the n_j-th and (n_j+1)-th frames, respectively. Letting T be the duration of this windowed sinusoid, its Fourier Transform is a $sinc$ with main lobe of width $2/T$. Since the widths of the frequency bins follow a logarithmic scale and each sinusoid must have a bandwidth not exceeding that of the bin, the sinusoidal components have different bandwidth and duration. For the m-th bin the sinusoid duration is set to $T_m = 2/W_m$, where W_m is the width of the bin, and its frequency is f_m, the center frequency of the bin. In terms of samples, the duration of the sinusoid is $rd(T_m \cdot f_s)$, where $rd(\cdot)$ is the rounding operator and f_s the sampling frequency. The additive signal $\alpha_j(k)$ is made of two sinusoidal components to modify the n_j-th and (n_j+1)-th frames. The signal that modifies the first frame affects the samples contained in the interval K_1, while the signal that impacts on the second frame lays in K_2. K_1 and K_2 are defined at the bottom of the page in (15). At the j-th iteration the distortion component $\alpha_j(k)$ is:

$$\alpha_j(k) = \begin{cases} r_j \sin\left(\frac{2\pi k f_{m_j}}{f_s}\right), & k \in K_1 \\ r_j \sin\left(\frac{2\pi k f_{m_j+1}}{f_s}\right), & k \in K_2 \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where r_j is a scale factor decided at each j-th iteration as follows: its initial value is set to $r_j = \sqrt{|V(n_j, m_j)|}/10^2$; its effect on the fingerprint bit is checked: if the bit is changed, this r_j value is kept, otherwise it is doubled until the bit is changed. Once this happens, the iteration $j+1$ starts. The iterations stop when $BER(H_x, H_{x+\alpha}) \geq \gamma$ (stopping criterion for the BAT

$$\begin{aligned} K_1 &= \left[k_h^{n_j} + k_{max} - \frac{rd(T_m \cdot f_s)}{2} + 1, k_h^{n_j} + k_{max} + \frac{rd(T_m \cdot f_s)}{2}\right] \\ K_2 &= \left[k_h^{n_j} + k_{min} - \frac{rd(T_{m+1} \cdot f_s)}{2} + 1, k_h^{n_j} + k_{min} + \frac{rd(T_{m+1} \cdot f_s)}{2}\right] \end{aligned} \tag{15}$$

TABLE I
OVERALL SUCCESS RATE OF FNR AND BAT ATTACKS

Attack type	Success Rate [%]	95% Confidence Interval	99.73% Confidence Interval
FNR	69.55	±2.58	±3.95
BAT	50.02	±2.80	±4.29

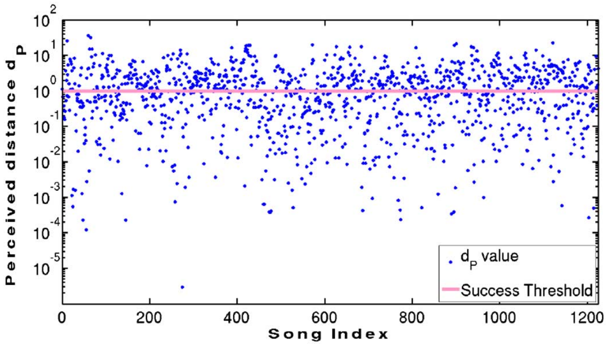

Fig. 2. d_P for each track, BAT attack, $\gamma = 0.1$.

attack). The extension to FNR case is easily carried out by applying the iterative procedure described above to each fingerprint row, $\forall n \in [1, N]$, until $D_{Ham}(H_x^n, H_{x+\alpha}^n) > \delta$ (stopping criterion for the attack through FNR).

V. RESULTS

To evaluate the performance of the proposed attack method, a test set of 1225 mp3 music tracks (also called songs) has been employed, suitably chosen to cover a wide variety of genres. The dimension of the test set assures a high level of confidence of the obtained results, as specified below. Each track has a sample rate of 44.1 kHz. The bit rate of the used tracks ranges from 128 to 320 kbps; the average value is 239.5 kbps. A 3 s segment has been extracted from each track to perform the tests. All segments are modified by using the procedure described in the previous Section up until $BER(H_x, H_{x+\alpha}) \geq \gamma = 0.1$ (BAT) or $D_{Ham}(H_x, H_{x+\alpha}) > \delta = 0$ (FNR). The perceived distortion is measured through two metrics: the Perceived Distance $d_P(x, x+\alpha)$ [17], which exploits an auditory model to measure the perceptual difference between the original signal and the modified one, and the Peak Signal to Noise Ratio (PSNR), as in [18]. In [17] the masking threshold, i.e. the maximum level of $d_P(x, x+\alpha)$ such that α is not detectable if $x+\alpha$ is listened to, has been set to 1. Therefore, we consider successful (as far as the audio quality is concerned) any attack that assures $d_P \leq 1$. Table I shows the success rate and the confidence intervals at 95% and 99.73% of the attack both in FNR and BAT cases. Although the proposed method does not take into account the user perception when it chooses $\alpha(k)$ but carries out the choice on a minimum energy basis, the performance is quite satisfactory. The higher rate of the FNR attack is due to the fact that it modifies less fingerprint bits than the BAT case attack. Fig. 2 shows the Perceived Distance d_P values for each track of the test database for BAT attack. A strictly similar behaviour is obtained for the FNR case. Fig. 3 shows the PSNR achieved in case of BAT attack and the corresponding energy of the inserted α for each song. The more energy is inserted, the lower the PSNR. Indeed, α may be considered as a sort of "controlled" noise. A test on the robustness of the attack with respect to independent noise e has been performed by evaluating the d_P associated with the perceived signal $x+\alpha+e$. Such e, assumed AWGN, may model a purposely inserted noise aimed at decreasing the perceived quality of the received signal and, consequently, the success rate of the attack. Fig. 4 reports the percentage of music tracks with $d_P < 1$ vs. signal-to-noise ratio (SNR) computed between x and e, in FNR and BAT attacks (where such a percentage coincides with the success rate) and in case of "no attack" ($\alpha = 0$). The "no attack" curve shows the impact of the noise e on the signal perceived quality and, in practice, it evaluates the human perception for audio signals when x is disturbed by e: when SNR is below 25 dB, the signal perceived quality is unacceptable; if SNR stays in the range [30–35] dB the quality improves; above 40 dB it is excellent. It is worth noting that the success rate of FNR and BAT attacks follows the same trend: it is not affected when SNR > 40 dB, it begins decreasing in the range [30–35] dB, and it is about 30% when $\mathrm{SNR} = 25$ dB. In short: the attack is successful (i.e., $x+\alpha+e$ has a satisfactory perceived quality) until the signal $x+e$ has a satisfactory perceived quality. The success rate obviously decreases when e is relevant, but exactly as the perceived quality of $x+e$ decreases. In this view, the attack may be considered robust: its effectiveness as a function of the noise e is substantially independent of the signal α, ruling the attack.

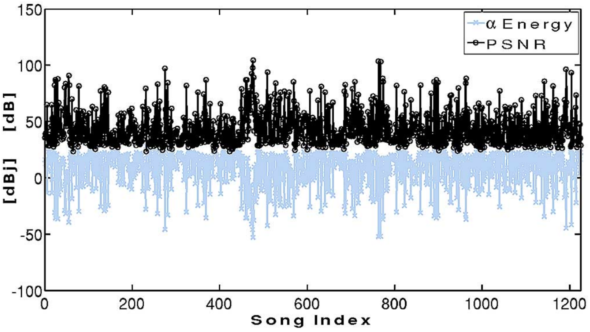

Fig. 3. Energy of the inserted distortion signal α and resulting PSNR for each track in BAT attack.

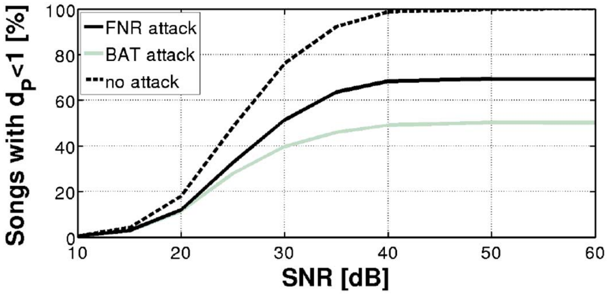

Fig. 4. Percentage of songs with $d_P < 1$ vs. SNR, in presence of AWGN, in case of FNR and BAT attacks, and no attack.

VI. CONCLUSIONS

A heuristic iterative energy-based method has been presented, able to deceive a fingerprint based PRH copy detector by suitably modifying the input audio signal. The results obtained on 1225 music tracks are satisfactory, being the success rates (in terms of perceived quality of the purposely distorted signal that has defeated the detection system) around 50% and 70% depending on the kind of attack. The approach can be considered promising and can be improved by associating a suitable perceptive metric to the employed purely energetic criterion.

REFERENCES

[1] J. Haitsma and T. Kalker, “A highly robust audio fingerprinting system,” in *3rd International Conf. Music Information Retrieval, ISMIR*, 2002.

[2] A. L. C. Wang, “An industrial strength audio search algorithm,” in *4th Int. Conf. Music Information Retrieval, ISMIR*, 2003, pp. 7–13.

[3] W.-H. Tsai and W.-C. Hsieh, “Blind clustering of music recordings based on audio fingerprinting,” in *Fifth Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP ‘09*, 2009, pp. 1086–1089.

[4] I. Bisio, A. Delfino, F. Lavagetto, and M. Marchese, “A television channel real-time detector using smartphones,” *IEEE Trans. Mobile Comput.*, vol. 99, no. PP, p. 1, 2013.

[5] Y. Liu, H. S. Yun, and N. S. Kim, “Audio fingerprinting based on multiple hashing in dct domain,” *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 525–528, 2009.

[6] M. L. Miller, M. A. Rodriguez, and I. J. Cox, “Audio fingerprinting: Nearest neighbor search in high dimensional binary spaces,” *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, vol. 41, no. 3, pp. 285–291, 2005.

[7] Y. Ke, D. Hoiem, and R. Sukthankar, “Computer vision for music identification,” in *in Proc. 2005 IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition (CVPR’05)* , Washington, DC, USA, 2005, vol. 1, pp. 597–604, IEEE Computer Society.

[8] M. Barni and F. Perez-Gonzalez, “Coping with the enemy: Advances in adversary-aware signal processing,” in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2013*, May 2013, pp. 8682–8686.

[9] M. Kirchner and R. Böhme, “Tamper hiding: Defeating image forensics,” in *Proc. 9th Int. Conf. Information Hiding*, 2007, vol. IH’07, pp. 326–341.

[10] M. Fontani and M. Barni, “Hiding traces of median filtering in digital images,” in *Proc. 20th Eur. Signal Processing Conf. (EUSIPCO)*, 2012, pp. 1239–1243.

[11] T.-T. Do, E. Kijak, T. Furon, and L. Amsaleg, “Deluding image recognition in SIFT-based CBIR systems,” in *Proc. 2nd ACM Workshop on Multimedia in Forensics, Security and Intelligence*, New York, NY, USA, 2010, vol. MiFor ‘10, pp. 7–12, ACM.

[12] A. Melloni, P. Bestagini, A. Costanzo, M. Barni, M. Tagliasacchi, and S. Tubaro, “Attacking image classification based on bag-of-visual-words,” in *IEEE Int. Workshop on Information Forensics and Security (WIFS), 2013*, Nov. 2013, pp. 103–108.

[13] P. Comesana-Alfaro and F. Perez-Gonzalez, “Optimal counterforensics for histogram-based forensics,” in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2013*, 2013, pp. 3048–3052.

[14] M. Barni, M. Fontani, and B. Tondi, “A universal attack against histogram-based image forensics,” *Int. J. Digital Crime Forensics*, vol. 5, no. 3, pp. 35–52, Jul. 2013.

[15] M. Barni and B. Tondi, “The source identification game: An information-theoretic perspective,” *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 3, pp. 450–463, 2013.

[16] A. Ramalingam and S. Krishnan, “Gaussian mixture modeling of short-time fourier transform features for audio fingerprinting,” *IEEE Trans. Inf. Forensics Secur.*, vol. 1, no. 4, pp. 457–463, Dec. 2006.

[17] C. H. Taal, R. C. Hendriks, and R. Heusdens, “A low-complexity spectro-temporal distortion measure for audio processing applications,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 5, pp. 1553–1564, 2012.

[18] J.-J. Jiang and C.-M. Pun, “Digital audio watermarking using an improved patchwork method in wavelet domain,” in *6th Int. Conf. Digital Content, Multimedia Technology and its Applications (IDC), 2010*, 2010, pp. 386–389.