How Secure are Power Network Signature Based Time Stamps?

Wei-Hong Chuang
University of Maryland
College Park, MD USA
whchuang@umd.edu

Ravi Garg
University of Maryland
College Park, MD USA
ravig@umd.edu

Min Wu
University of Maryland
College Park, MD USA
minwu@umd.edu

ABSTRACT

A time stamp based on the power network signature called the Electrical Network Frequency (ENF) has been used by an emerging class of approaches for authenticating digital audio and video recordings in computer-to-computer communications. However, the presence of adversaries may render the time stamp insecure, and it is crucial to understand the robustness of ENF analysis against anti-forensic operations. This paper investigates possible anti-forensic operations that can remove and alter the ENF signal while trying to preserve the host signal, and develops detection methods targeting these operations. Improvements over anti-forensic operations that can circumvent the detection are also examined, for which various trade-offs are discussed. To develop an understanding of the dynamics between a forensic analyst and an adversary, an evolutionary perspective and a game-theoretical perspective are proposed, which allow for a comprehensive characterization of plausible anti-forensic strategies and countermeasures. Such an understanding has the potential to lead to more secure and reliable time stamp schemes based on ENF analysis.

Categories and Subject Descriptors

I.5.4 [**Applications**]: Signal Processing; K.6.5 [**Security and Protection**]: Authentication

General Terms

Security, Algorithms, Experimentation.

Keywords

Digital Recording Authentication, Time Stamp, Electrical Network Frequency, Information Forensics, Game Theory.

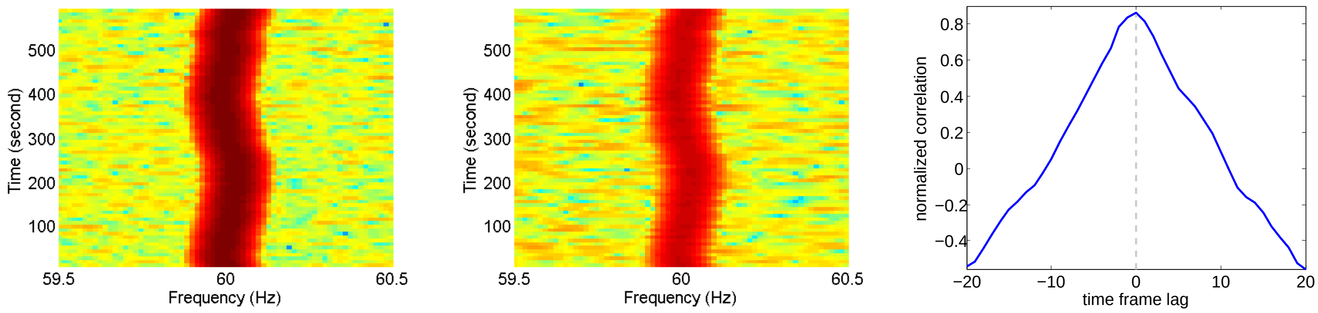
1. INTRODUCTION

The recent decade has witnessed a huge amount of media data, in the form of audio, image, and video, created by various digital recording devices. Once a media document containing important information is created, it can be easily distributed through network and make rapid and broad social impacts through social media infrastructure. Due to their digital nature, these media data can be vulnerable to digital forgeries. Typical examples include digital editing software to cut a clip from one audio/video file and insert into another, or modifying the creation date/time in the metadata field. Given the feasibility of digital forgeries, secure use of media data requires forensic authentication mechanisms that can identify data origin and detect content forgery.

One emerging direction of digital recording authentication is to exploit an potential time stamp originated from the power networks. This time stamp, referred to as the Electrical Network Frequency (ENF), is based on the fluctuation of the supply frequency of a power grid. The nominal value of the ENF is 60Hz in the Americas, Taiwan, Saudi Arabia and Philippines, and is 50Hz in other regions except Japan, which adopts both frequencies. It has been found that digital devices such as audio recorders, CCTV recorders, and camcorders that are plugged into the power systems or are near power sources may pick up the ENF signal due to the interference from electromagnetic fields created by power sources [4]. An important property about the ENF signal is that its frequency is fluctuating around the nominal value because of varying loads on the power grid. For example, in the United States, the ENF usually varies between 59.9Hz and 60.1Hz. It has also been shown that the fluctuations measured at the same time but at two different locations under the same power grid follow basically the same trend [4].

The fluctuation of the ENF has been successfully exploited to authenticate digital recordings [4, 10, 9, 3]. In [4, 10], it is demonstrated that the ENF signal is captured in audio recordings and exhibits a high correlation with the ENF signal measured from the power mains supply at the same time. As such, the ENF signal can be used to indicate the recording time of an audio recording provided that a database of ground-truth ENF signals from the power grid is accessible. An alternative technique in [9] detects the phase discontinuity of the ENF signal, the presence of which suggests where tampering has taken place. Most recently, the work in [3] validated for the first time the presence of the ENF signal in visual recordings. Optical sensors and video cameras are used to demonstrate that the ENF signal can be captured from fluorescent lighting and further picked up by video cameras in an indoor environment. This finding suggests that the same ENF-based time stamp available in audio record-

CCS'12, October 16–18, 2012, Raleigh, North Carolina, USA.



(a) Power mains ENF signal

(b) Audio ENF signal

(c) Normalized correlation

Figure 1: (a) Spectrogram of a power mains signal around the nominal ENF value of 60Hz; (b) spectrogram of an audio signal; (c) normalized correlation between the two extracted ENF signals as a function of their relative frame lag.

ings can be used to authenticate visual data as well. Furthermore, forensic binding can be achieved for visual and audio tracks to verify their temporal synchronization.

The promising potential of ENF analysis in forensic investigations is based on the premise that the ENF signal is present in an audio or video signal in an unaltered manner. This premise ensures that once the ENF signal is successfully extracted, it can be used as a truth-telling evidence to verify time, location, and integrity. However, similar to many other security and forensics tasks, there exist adversaries who have the incentives to perform *anti-forensic operations* to counteract forensic investigations [6, 2]. In order to establish ENF-based analysis as a credible technique, it is of paramount importance to understand its security against anti-forensic operations, namely, whether the ENF signal can be compromised, and to what extent. Further, forensic analysts should understand and address identified vulnerabilities in ENF analysis, and take into consideration possible improvements that an adversary may make. To the best of our knowledge, the current paper is the first work that considers these issues. We apply signal processing techniques to design anti-forensic operations, and then develop detection methods targeting these operations. In response to the detection methods, improvements over the anti-forensic operations are also investigated in this paper, for which various trade-offs are discussed. More fundamentally, we develop a comprehensive understanding of the interplay between the forensic analyst and the adversary, from both an evolutionary perspective and a game-theoretic perspective. We believe that such an understanding can be used to characterize a wide range of actions that may take place, and will contribute to more secure and reliable time stamp schemes based on ENF analysis.

The rest of this paper is organized as follows. Section 2 reviews the mechanism of ENF signal extraction and matching. Section 3 investigates ways to remove and embed ENF signals present in a host signal. Section 4 presents the conditions for anti-forensics detection, which motivate a few concrete methods for anti-forensics detection. In response to the detection, Section 5 studies improvements over the anti-forensic operations, for which various trade-offs are discussed. In view of the dynamic nature of the anti-forensics and the countermeasures, Section 6 provides an evolutionary perspective and a game-theoretic perspective to encompass a wide range of actions and interactions available to a forensic analyst and an adversary. Finally, Section 7 concludes this paper.

2. ENF SIGNAL EXTRACTION AND MATCHING

In this section, we briefly describe our procedure for extracting the ENF fluctuations from a given signal. Two types of signals are considered in this paper for ENF signal extraction and matching. The first is the audio signal that contains speech recordings mixed with music and sporadic sound activities. All audio signals used in this paper have been sampled at 8000Hz with 16-bit quantization precision and a length of 10 minutes. The 10-minute duration ensures that the audio signal as well as the ENF fluctuations are sufficiently long for reliable matching based on the state of the art. Any anti-forensic operations to be investigated in this paper are also assumed to be performed on such audio signals. The second type of signal is the power mains signal that is recorded directly from a power source using a voltage divider device, which is used as ground truth for matching.

Our ENF signal extraction basically follows the procedure described in [3]. The recorded signal (either an audio or power mains signal) is first down-sampled to 500Hz to reduce the computational complexity. A filtering process can then be carried out to only retain the signal component that carries the ENF. The dominant instantaneous frequency in the recorded signal is then estimated to measure the fluctuations in ENF as a function of time, for which we use spectrogram based weighted energy method as in [3]. To obtain the spectrogram of the ENF signal, we divide the signal into overlapping frames of 16 seconds each with an overlap factor of 50%. A high resolution Fast Fourier Transform (FFT) of 8192 points is carried out for each frame. After obtaining the spectrogram, we calculate the weighted average frequency in each time bin of the spectrogram by weighing frequency bins around the nominal values of the ENF with the energy present in the corresponding frequency. From the estimated frequency fluctuations in ENF signals from the audio and power mains recordings, we calculate their normalized correlation for different values of frame lag. The range of the normalized correlation value is between -1 and $+1$. As an example, Fig. 1(a) and 1(b) show the spectrograms around

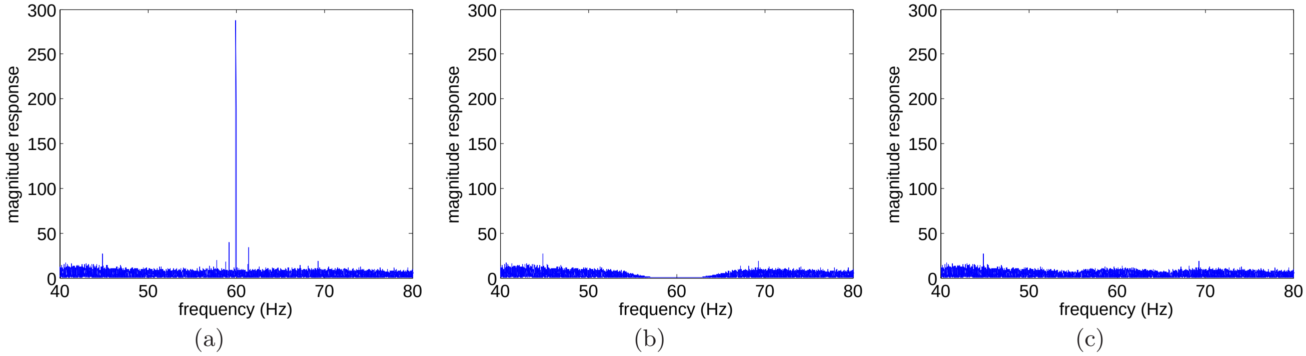


Figure 2: (a) The FFT of an authentic audio clip; (b) the result of bandstop filtering; (c) the result of bandstop filtering followed by noise filling-in.

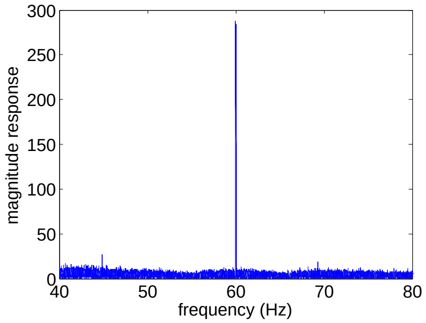


Figure 3: ENF embedding result with peak magnitude matched (see Fig. 2(a) for comparison).

the nominal ENF value of 60Hz of a power mains signal and an audio signal that were recorded at the same time. Their normalized correlation values as a function the frame lag is plotted in Fig. 1(c). We can see that they exhibit consistent fluctuations, which is confirmed by the peak normalized correlation value of 0.86 in Fig. 1(c) when the two recordings are synchronized.

3. ANTI-FORENSIC OPERATIONS AGAINST ENF ANALYSIS

In this section, we investigate anti-forensic operations that can counteract ENF analysis. The general purpose of anti-forensic operations is to alter a host signal so that the traces left in the host signal that pertain to specific forensic investigations are removed or changed. Although plausible anti-forensic operations and countermeasures are domain-specific and may seem ad-hoc at times, exploring these operations and countermeasures is necessary for identifying the available operations of both the forensic analyst and the adversary, leading to a comprehensive understanding of the overall strategy space.

In many anti-forensic tasks against information protection, the adversary has to preserve the quality of the host signal, otherwise the quality degradation in itself will indicate the use of anti-forensics and the host signal will be rejected to be forensic evidence. In our problem, the ENF signal is restricted around narrow neighborhoods of known frequency locations. As such, the ENF signal is less likely to be tightly coupled with the main body of the host signal, making it possible to manipulate the ENF signal while trying to preserve the perceptual quality of the host signal. In this section, we explore two different levels of anti-forensics, starting with the removal of the ENF signal and further considering the embedding of an alien ENF signal.

3.1 ENF Signal Removal by a Bandstop Filter

The first anti-forensic operation that we consider is to remove the ENF signal present in a host signal. Since the ENF signal in nature is restricted in a small frequency region (a.k.a. *narrowband* hereafter), it is reasonable for an adversary to apply a bandstop filter to remove the ENF signal. Bandstop filtering (a.k.a. notch filtering) is a well-studied subject in digital signal processing [8]. A number of design methodologies, such as equiripple filter or Kaiser window filter design, have been proposed and implemented in popular software packages such as MATLAB. To perform bandstop filtering, an adversary selects two main parameters, the stopband bandwidth and the transition bandwidth. The stopband bandwidth controls the frequency range in which the signal is attenuated to the minimum magnitude level. For the task of ENF signal removal, the choice of stopband bandwidth depends on the actual range of ENF variation, and the ENF signal of wider variations may be removed using wider stopbands. The second parameter, the transition bandwidth, is the range in which the signal attenuation varies from maximum to minimum. It has an impact on the filter length and computational complexity; a sharper bandwidth implies a longer filter and more time required to compute the filter output. Since accurate ENF matching requires ENF signals of sufficiently long durations, it is reasonable to assume that audio signals used for anti-forensic operations are also sufficiently long. Therefore, if the adversary can afford the computational cost, he/she has enough signal samples to carry out a bandstop filtering with a reasonably small transition bandwidth. As an example, when the sampling frequency is 8000Hz as common for voice signals, we set the stopband bandwidth as ±1Hz, and the transition bandwidth as 8Hz. If the equiripple linear-phase design is adopted, the filter has a length of 3627 samples, which corresponds to a duration of about half a second.

To illustrate the effect of bandstop filtering, we show in Fig. 2(a) a typical Fourier analysis result on a 10-minute audio recording. There is a salient peak located at 60Hz, which signifies the existence of the ENF signal. The effect of bandstop filtering for the same audio recording is shown

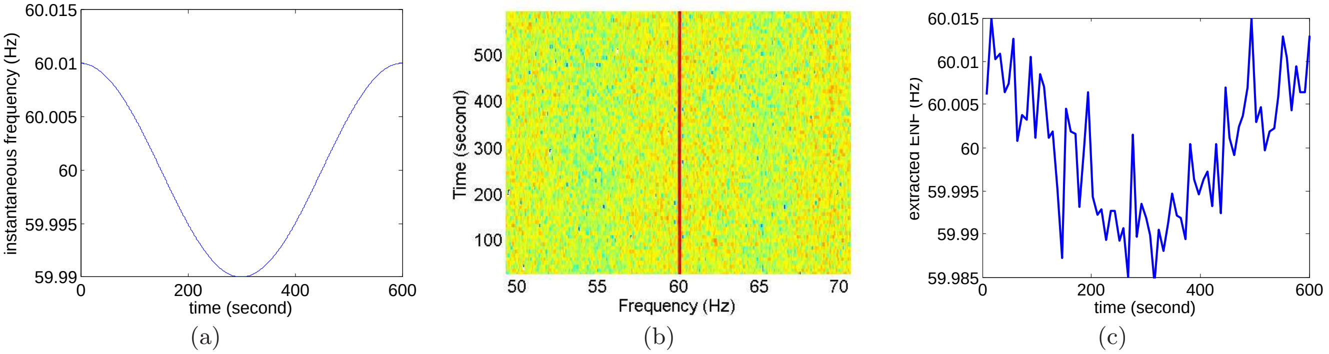


Figure 4: (a) A purely sinusoidal sequence of instantaneous frequencies to be embedded as the ENF signal; (b) the spectrogram around 60Hz where a strong component is present due to the embedding of (a); (c) the corresponding extracted ENF signal.

in Fig. 2(b), wherein the peak at 60Hz disappears, suggesting that the ENF signal has been removed. The removal is further justified by comparing the normalized correlation between the ENF signal extracted from power mains ground truth and the ENF signal extracted from the audio recording. We notice that the normalized correlation reduces from 0.86 to −0.10 due to bandstop filtering, suggesting that the ENF signal has been effectively removed. Furthermore, our subjective tests do not find perceptual audio quality loss, meaning that the ENF signal removal preserves the host signal.

Although bandstop filtering can remove the ENF signal, a notch of very low magnitude around the 60Hz frequency can be noticed in Fig. 2(b). The notch is a strong evidence that suggests the use of bandstop filtering, making the resulting audio recording no longer trustworthy and hence anti-forensics essentially fails. To erase such traces, an option is to "fill in" the frequency region that has been suppressed by bandstop filtering. We design a bandpass filter with passband bandwidth ±1Hz and transition bandwidth 8Hz and pass a white noise signal through the filter to obtain a narrowband signal that is then added to the bandstopped audio recording. The noise power is selected so that the resulting narrowband magnitude equals the average magnitude of neighboring narrowbands, as shown in Fig. 2(c). Since the narrowband now appears smooth and there is no peak at 60Hz, it becomes more difficult for the forensic analyst to determine if there was measurable ENF signal present at 60Hz.

3.2 Embedding Phony ENF Signals

In addition to removing the ENF signal so that the recording time of an audio recording is no longer available, an adversary may further embed a fake ENF signal into a host signal so that ENF analysis conducted over the forged audio signal leads to a wrong estimate for the recording time. This can be done by modulating a carrier sinusoidal signal of a nominal frequency using a given sequence of instantaneous frequencies. In mathematical terms, the carrier signal can be written as

$$c(t) = A\cos(2\pi f_c t). \qquad (1)$$

The modulation is given by

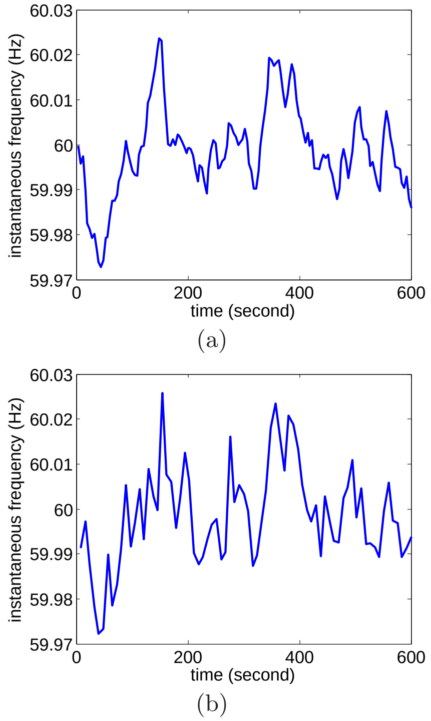
$$e(t) = A\cos\left(2\pi\int_0^t f_m(\tau)d\tau\right), \qquad (2)$$

which is the standard form of Frequency Modulation (FM) synthesis [5]. Indeed, the instantaneous frequency of (2) is given by $\frac{d}{dt}\frac{1}{2\pi}\left(2\pi\int_0^t f_m(\tau)d\tau\right) = f_m(t)$. The magnitude A is a constant to be determined.

Next, we discuss how to embed a modulated signal into a host signal. As in Section 3.1, we first apply a bandstop filter on the host signal and then fill in bandpassed noise whose magnitude is matched to neighborhood regions. The magnitude A in (2) is chosen so that the peak FFT magnitude at the nominal frequency remains the same after the anti-forensic operation, as shown in Fig. 3. This can be achieved using a binary search procedure: starting with a guess of A, each iteration compares the resulting peak FFT magnitude to the targeted value and increases/decreases A accordingly.

We consider two possible types of synthetic ENF signals. If there is no real ENF signal from another time or another power grid available for embedding, one can embed a purely artificial signal such as the sinusoidal variation as shown in Fig. 4(a). The resulting spectrogram has a strong component around 60Hz as shown in Fig. 4(b), and the ENF signal extracted from the forged audio signal is shown in Fig. 4(c), which is a noisy version of Fig. 4(a) since the embedded signal has been mixed into the narrowband. On the other hand, if a real ENF signal originated from a different time or from another power grid is available, then such a ENF signal can also be embedded into the host signal to mislead forensic analysis. Fig. 5(a) shows a power mains ground truth ENF signal, and the corresponding extracted ENF is shown in Fig. 5(b). We can see that the embedded ENF can also be extracted in a more noisy form.

The proposed embedding above is based on the FM synthesis. Alternatively, one can perform a "transplantation" operation to duplicate the ENF signal from one signal into another signal. Specifically, to embed an ENF signal present in a source audio signal into a host signal, we perform bandpass and bandstop filtering upon the source and the host signal, respectively, and then add the bandpassed output of the source signal into the bandstopped output of the host signal. In Fig. 6(a), we show the spectrogram of a trans-



(a)

(b)

Figure 5: (a) Ground-truth ENF signal measured from the power mains; (b) the corresponding extracted ENF signal.

plantation result in which the 60Hz narrowband has been replaced. The extracted ENF signals from the source signal and the resulting signal are shown in Fig. 6(b). The observation that they overlap tightly indicates the effectiveness of the transplantation from an anti-forensic point of view.

4. DETECTING ANTI-FORENSICS

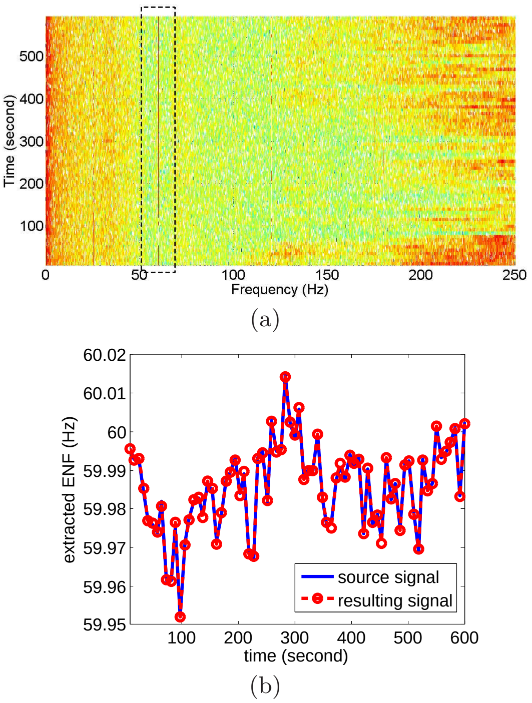
Our study in Section 3 has shown a number of anti-forensic operations that can counteract ENF analysis. In response to these operations, a forensic analyst would devise ways to detect the use of anti-forensic operations, so that a forged audio signal can be identified and rejected as trustworthy evidence. In this section, we first discuss conditions under which the detection is feasible, and then propose effective detection methods.

4.1 Detectability of Anti-Forensic Operations

In order to detect anti-forensic operations, we first provide a mathematical formulation of the anti-forensic operations discussed in Section 3. Without loss of generality, the anti-forensic operations proposed therein create a forged audio signal by mixing a bandstopped input signal and a bandpassed alien signal (either real or synthetic). In the frequency domain, the overall anti-forensic operation can be represented as

$$Y(\omega) = e^{-j\alpha\omega}\left[X(\omega)B_s(\omega) + E(\omega)B_p(\omega)\right], \tag{3}$$

where $X(\omega)$ is the frequency-domain representation of the original audio signal indexed by the frequency ω (in Hz), $Y(\omega)$ is the resulting audio signal, $E(\omega)$ is the alien signal, $B_s(\omega)$ and $B_p(\omega)$ are the frequency responses of the bandstop filter and the bandpass filter, respectively, and $e^{-j\alpha\omega}$ is a phase shift corresponding to a possible time-domain delay of α. The delay is introduced to avoid imperfect boundary conditions due to filtering.



(a)

(b)

Figure 6: (a) Result of narrowband transplantation around 60Hz; (b) ENF signals extracted from the source signal and from the resulting signal.

Consider two mutually exclusive cases. For the frequency outside the narrow passband, $|B_s(\omega)| \approx 1$ and $|B_p(\omega)| \approx 0$, and we have

$$\begin{aligned} |Y(\omega)| &\approx |X(\omega)|, \\ \angle Y(\omega) &\approx -\alpha\omega + \angle X(\omega) + \angle B_s(\omega). \end{aligned} \tag{4}$$

In practice, both the bandstop and the bandpass filters can be designed as zero-phase or linear-phase. As such, the phase term $\angle B_s(\omega)$ is linear outside the narrowband, and by properly selecting the delay α, the two terms $-\alpha\omega$ and $\angle B_s(\omega)$ can be cancelled out, leading to $Y(\omega) \approx X(\omega)$ outside the narrowband. In other words, the anti-forensic operations basically preserve the host signal outside the narrowband. On the other hand, for the frequency inside the narrowband, we have $|B_s(\omega)| \approx 0$ and $|B_p(\omega)| \approx 1$, and

$$\begin{aligned} |Y(\omega)| &\approx |E(\omega)|, \\ \angle Y(\omega) &\approx -\alpha\omega + \angle E(\omega) + \angle B_p(\omega) \\ &\approx \angle E(\omega) + (\beta - \alpha)\omega \end{aligned} \tag{5}$$

provided that the bandpass filter has linear phase in the narrowband. This suggests $Y(\omega) \approx e^{(\beta-\alpha)\omega}E(\omega)$, that is, the output signal inside the narrowband resembles the alien signal inside the narrowband with a possible phase shift. If the bandstop and bandpass filters are designed using the same methods, then α and β are similar and thus the phase shift is close to zero. To summarize, overall the proposed anti-forensic operations from Section 3 only alter the narrowband and leave no substantial influence outside the narrowband.

In order to detect anti-forensic operations, a forensic analyst can carry out a likelihood ratio (LR) test to compare the likelihoods of a forged audio signal and an unforged audio signal. Specifically, the analyst evaluates the following likelihood ratio:

$$\begin{aligned} LR &= \frac{P(Y|\text{forged})}{P(Y|\text{unforged})} \\ &= \frac{P(O=o, I=i|\text{forged})}{P(O=o, I=i|\text{unforged})} \qquad (6) \\ &= \frac{P(I=i|\text{forged}, O=o)}{P(I=i|\text{unforged}, O=o)}, \qquad (7) \end{aligned}$$

where we decompose Y into a pair of (I, O) in (6), standing for the inside-narrowband and outside-narrowband components, respectively, and the terms $P(O = o|\text{forged})$ and $P(O = o|\text{unforged})$ are cancelled out in (7) since the anti-forensic operations do not affect the host signal outside the narrowband.

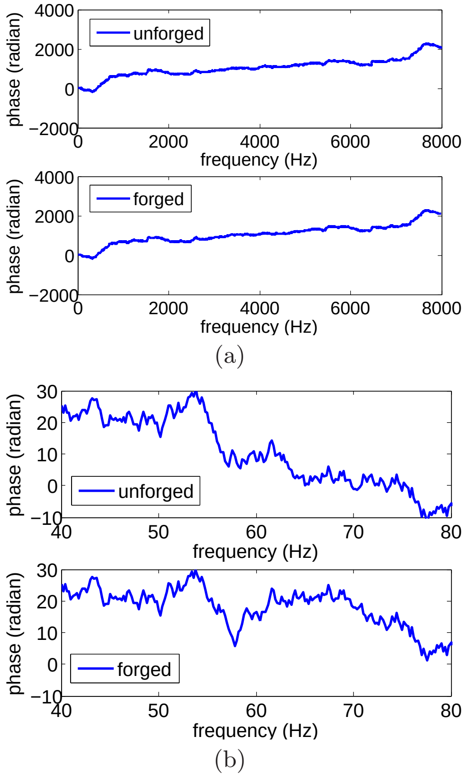
For the anti-forensic operations proposed in Section 3, the forged narrowband is independent of the signal outside the narrowband. Therefore, the numerator in (7) can be written as $P_{EI}(i)$, standing for the likelihood of observing a narrowband i conditioned that the narrowband is from an alien signal. The denominator, on the other hand, has to account for the dependence of the narrowband on the signal outside the narrowband. Specifically, the denominator can be denoted as $P_{XI,o}(i)$, which is the likelihood of a narrowband i given that the narrowband is native (i.e., not from another signal) and the signal outside the narrowband is o. In summary, the likelihood ratio is given by $P_{EI}(i)/P_{XI,o}(i)$.

From such an analysis, we see that a distinction has to be made between the original audio signal X and the alien signal E in the narrowband, in order to detect anti-forensics operations. This is, however, a challenging task, since the adversary can design the bandstop filter to make the narrowband very "narrow", especially compared to the wide frequency range associated with the much higher sampling frequency. As a result, the characteristics of the original audio signal X and the alien signal E cannot be easily distinguished in the narrowband. To illustrate such a difficulty for the forensic analyst, Fig. 7(a) shows the overall phase of an unforged audio signal as well as its forged version, and their difference is hardly noticeable. Zooming into the narrowband as shown in Fig. 7(b), we observe that the two versions differ in the narrowband, but it is not straightforward to characterize their statistical difference and to determine which one is forged.

4.2 Inter-Frequency Consistency Check

Section 4.1 shows that anti-forensic operations can be detected if one can distinguish the two distributions $P_{EI}(i)$ and $P_{XI,o}(i)$ in the likelihood ratio. Motivated by this finding, we propose a few ways toward this end.

We have considered so far the scenario that a forensic analyst only extracts ENF signals from a given frequency (e.g., the fundamental frequency of 60Hz). In this case, it is reasonable for an adversary to focus on tackling this frequency as well. However, due to the non-linear behavior of electrical circuits, the ENF signal is often present not only at the fundamental frequency, but also at the harmonic frequencies (120Hz, 180Hz, etc) [1]. As such, in order to detect anti-forensic operations, the forensic analyst can perform ENF extraction at more than one frequency, and examine the consistency of multiple ENF estimates. To illustrate this idea, we extract ENF signals from an audio signal at 60Hz and 180Hz, respectively, and the results are shown in Fig. 8. Note that these two signals have been normalized with respect to their average values. It can be seen that the two extracted ENF signals highly overlap with each other, and their normalized correlation is 0.66. This is significantly higher than the average normalized correlation value of 0.02 when the 60Hz and 180Hz ENF signals are extracted from two different audio signals, respectively. An issue with this detection method is that the magnitude of ENF signal at higher harmonic frequencies is usually lower, and the host audio signal may have higher magnitude at these frequencies. Hence, the ENF extraction quality is lower at these harmonic frequencies, and it is likely that no ENF signals can be extracted for reliable consistency check.



(a)

(b)

Figure 7: (a) Comparison of overall phase associated with unforged and forged audio signals; (b) comparison of phase around 60Hz associated with unforged and forged audio signals.

4.3 Spectrogram Consistency Check

As an adversary performs the anti-forensic operations proposed in Section 3, the resulting narrowband often exhibits some kind of inconsistency with the signal outside the narrowband, especially the abrupt boundaries that are easily noticeable around 120Hz. Mathematically, this means the value of $P_{XI,o}(i)$ is small, which serves as a strong indicator of the existence of anti-forensics. As an example, consider an adversary that alters the ENF at 120Hz. A typical result-

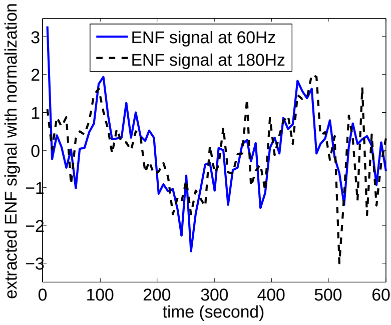


Figure 8: Consistency of ENF signals extracted at the fundamental frequency of 60Hz and a harmonic frequency of 180Hz.

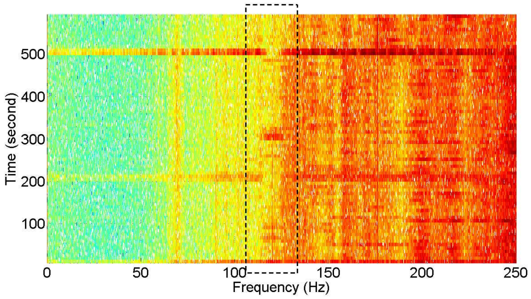
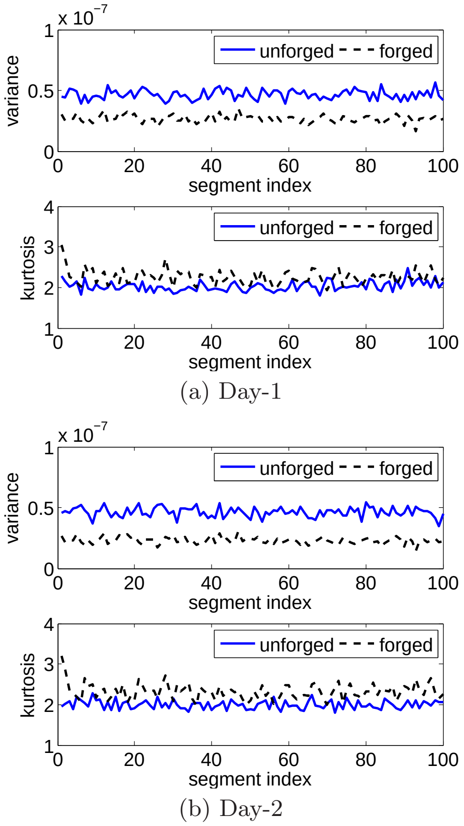


Figure 9: Spectrogram consistency check for a signal with its 120Hz narrowband forged; the obvious inconsistency around 120Hz is highlighted by the dashed box.



(a) Day-1

(b) Day-2

Figure 10: Variance and kurtosis statistics calculated over 5-second segments on (a) Day 1 and (b) Day 2.

ing spectrogram is shown in Fig. 9, where discontinuity at the narrowband boundaries centered at 120Hz can be clearly noticed. Such inconsistency occurs if the host audio and the alien audio signal exhibit strong but unsynchronized temporal variations.

Although the spectrogram consistency check is powerful, automating this check is non-trivial as in reality, a forensic analyst has no *a priori* knowledge about the narrowband range. In order to detect the boundary discontinuity the analyst has to scan the entire frequency range at a fine resolution, which demands a high computational complexity.

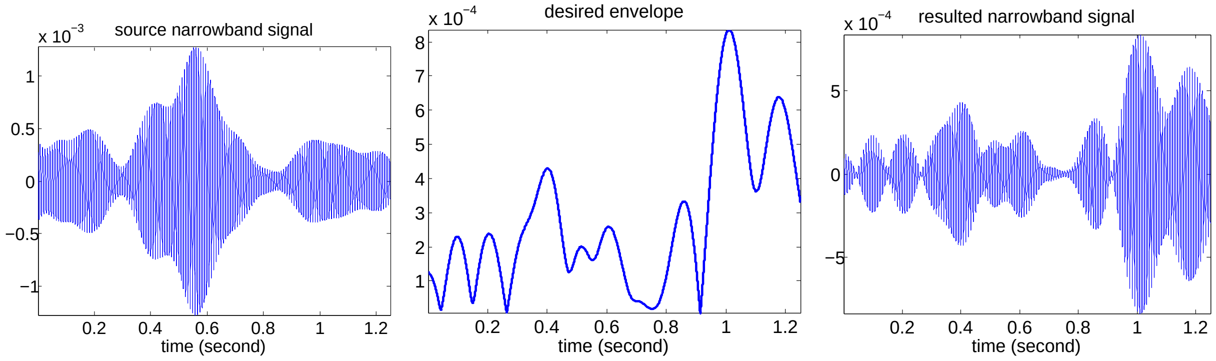
4.4 Reference-based Detection

In Section 4.1, we have seen conditions under which anti-forensic operations can be detected. In particular, a forged and an unforged audio signal can be distinguished if their narrowband characteristics are available. Here we consider a special setting called *reference-based anti-forensics detection*, in which it is assumed that when a query signal is present whose ENF signal remains to be authenticated, a reference signal with similar ENF sensing conditions is also accessible. Note that this is in contrast to the blind detection method that we have discussed previously. The reference-based setting is possible in many practical scenarios. For example, if the adversary presents multiple pieces of audio recordings among which some have forged ENF signals, then the remaining unforged audio recordings can serve as the reference signals. As another example, consider an audio that is used as forensic evidence whose authenticity remains to be determined. A forensic analyst can replicate the recording environment so that the ENF sensing conditions are replicated as well. The reference-based anti-forensics detection can be seen as a resource-augmented detection, and as we know, this has not been exploited previously.

In the reference-based anti-forensics detection setting, since the reference signal contains an authentic ENF signal, information about $P_{XI,o}(i)$ can be learnt from the statistics of the reference signal. Specifically, by writing $P_{XI,o}(i) = P_{XI}(i)\frac{P(o|i,X)}{P(o|X)}$, one can detect an anti-forensic operation upon a query audio signal if it leads to a low $P_{XI}(i)$. To verify this idea, we collect two audio signals recorded on two different days (10 January and 14 January 2012, respectively). The two audio clips were made by playing online streaming via the same speaker and recording using the same microphone. The placement of the microphone and the speaker volume, however, are not exactly the same on the two days. For a given audio file whose narrowband surrounding 60Hz is denoted by $B(n)$, we divide $B(n)$ into segments of a 5-second duration, and calculate sample statistics for each segment. In particular, we examine the variance that measures how each sample spreads out from the average value, and the kurtosis that measures the "peakedness" of each sample, defined as

$$\mathrm{Var}(B) = E[(B(n)-\bar{B})^2], \qquad (8)$$

$$\mathrm{Kur}(B) = \frac{E[(B(n)-\bar{B})^4]}{E^2[(B(n)-\bar{B})^2]}, \qquad (9)$$



(a) (b) (c)

Figure 11: (a) The source narrowband signal in time domain; (b) the envelope of the native narrowband signal; (c) the resulting narrowband signal after envelope matching of (a) to (b).

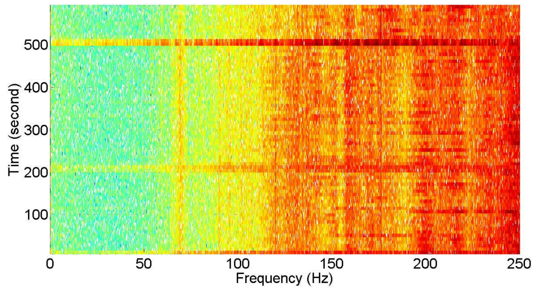


Figure 12: Spectrogram with an envelope-adjusted narrowband. Notice the inconsistency around 120Hz in Fig. 9 is no longer visible.

respectively, where $\bar{B}$ is the average value of $B(n)$ in a segment. We plot the two statistics corresponding to unforged and forged signals for Day 1 and Day 2 in Fig. 10(a) and Fig. 10(b), respectively. We can see that both the unforged and the forged signals have stable statistics on the two days, and unforged and forged signals show noticeably separable statistics values. Therefore, if we are given any of these two unforged recordings as reference, we can detect anti-forensics over the other recording by checking the consistency of the statistics. This idea of reference-based anti-forensics detection can be further augmented by incorporating other useful statistics.

5. IMPROVING ANTI-FORENSIC OPERATIONS

Being aware of the anti-forensics detection methods proposed in Section 4, the adversary will naturally improve the anti-forensic operations. In this section, we examine a few possible methods toward this goal, and discuss their tradeoffs.

To cope with the inter-frequency consistency check, the adversary can alter multiple ENF harmonic frequencies. Two issues have to be addressed by the adversary. First, the alteration has to be performed with regard to possible signal quality degradation. This is because altering the ENF signal at higher harmonics involves applying bandstop filtering by the adversary's anti-forensic operations to the audio signal at higher frequencies, which usually has richer content. Second, from a forensic analyst's point of view, as more ENF frequencies are affected, more traces will be left that may be exploited by the reference-based anti-forensics detection. Nevertheless, as discussed in Section 4.2, ENF signals generally can only be extracted reliably at lower harmonic frequencies. Around these frequencies, host signal quality degradation is barely noticeable according to our subjective perceptual evaluation. As such, the two issues above are not serious in practice.

5.1 Envelope Adjustment

Recall that the anti-forensic operations proposed in Section 3 may result in inconsistency on the spectrogram. This is because the embedded narrowband may have different temporal magnitude variations. To address this issue, an adversary can try to adjust the envelope of the narrowband, so that the adjusted narrowband has similar temporal variation as the native narrowband. Such adjustment can be done by means of the Hilbert Transform [5]. Specifically, the Hilbert Transform of a real-valued narrowband signal in the form of

$$b(t) = A(t)\sin(2\pi f_c t + \phi) \tag{10}$$

is given by

$$\begin{aligned} H\{b(t)\} &= b(t) + jA(t)\sin\left(2\pi f_c t + \phi + \frac{\pi}{2}\right) \\ &= b(t) + jA(t)\cos(2\pi f_c t + \phi), \end{aligned} \tag{11}$$

which includes a purely imaginary part that is $\pi/2$ phase-shifted from $b(t)$. As a result, the amplitude equals to $|H\{b(t)\}| = A(t)$, where the periodical part $\sin(2\pi f_c t + \phi)$ is no longer present. The envelope adjustment is done by matching the envelopes of the native narrowband and the forged narrowband in the following form:

$$\hat{b}_h(t) = \frac{|H\{b_h(t)\}|}{|H\{b_s(t)\}|} b_s(t), \tag{12}$$

where $b_s(t)$ is the alien narrowband signal (source), and $b_h(t)$ is the native narrowband signal (host). Examples of $b_s(t)$ and $|H\{b_h(t)\}|$ are shown in Fig. 11(a) and Fig. 11(b), and the resulting narrowband is given in Fig.11(c). It is clear that the narrowband from the alien signal has been adjusted with a matched envelope. The spectrogram after envelope

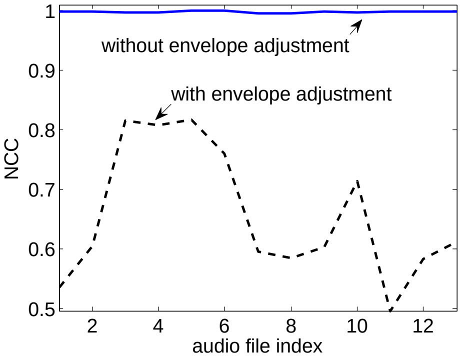


Figure 13: Comparison of normalized correlation values with and without envelope adjustment. Note that the normalized correlation has been substantially reduced when envelope adjustment is applied.

adjustment is given by Fig. 12, which no longer exhibits the spectrogram inconsistency as in Fig. 9.

Envelope adjustment may cause some loss of fidelity in the embedded ENF signal, which can be seen in the following experiment. We perform the narrowband transplantation proposed in Section 3.2 on 13 different audio files. Specifically, for each audio file, we extract the narrowband from another arbitrarily chosen file and transplant the extracted narrowband into the audio file as described in Section 3.2. For these 13 audio files, We first calculate the normalized correlation between the ENF signal present in the alien narrowband and the ENF signal in the forged narrowband. We then perform envelope adjustment and also calculate the normalized correlation between the ENF signal in the alien narrowband and the ENF signal in envelope-adjusted narrowband. As shown in Fig. 13, the normalized correlation reduces from a value close to 1 to about 0.6 as a result of the envelope adjustment. That is, the envelope adjustment introduces distortion to the ENF, which suggests that an adversary only has limited capabilities of preserving the fidelity of the spectrogram and embedded ENF signal at the same time.

5.2 Matching the Statistics

We have seen in Section 4.4 that due to the limited fidelity of ENF embedding, anti-forensic operations may be detectable with the aid of certain statistics from a reference signal. As such, an adversary also has the incentive to match the statistics. We have found that the envelope adjustment technique discussed in Section 5.1 can effectively match the two variance and kurtosis statistics, as shown in Fig. 14. However, while the adversary matches these two statistics, some other statistics may be affected. Fig. 15 shows the peak magnitude at 60Hz on the FFT result with and without envelope adjustment. We can see that, envelope adjustment consistently increases the peak magnitude, which can be exploited accordingly by the forensic analyst to detect anti-forensic operations. This phenomenon is fundamental and indicates that some mismatch always takes place if the adversary only has limited knowledge about how ENF is formed in an audio signal. For both forensic analysts and adversaries, it is therefore crucial to acquire a deeper understanding of ENF's underlying mechanism so as to mimic or to scrutinize the fidelity of ENF embedding. The relations between forensic analysts' and adversaries' actions will be discussed in more depth in the next section.

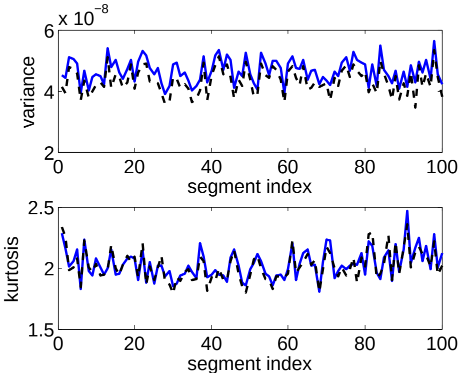


Figure 14: Variance and kurtosis statistics matching via envelope adjustment. Solid and dashed curves represent the statistics associated with authentic data and envelope-adjusted data, respectively.

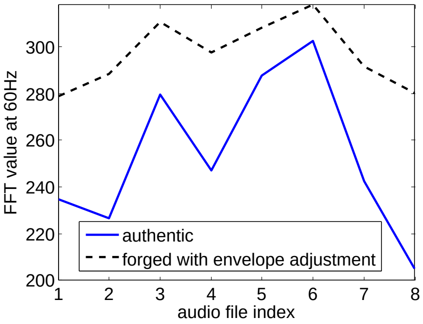


Figure 15: Consistent increase in the peak FFT magnitude due to envelope adjustment.

6. INTERPLAY BETWEEN FORENSIC ANALYST AND ADVERSARY

Summarizing our proposed forensic and anti-forensic operations developed upon empirical data, we can see a highly dynamic interaction between the forensic analyst and the adversary. In this section, we consider such an interaction from two perspectives. The first perspective treats the interaction as an evolutionary process, in which both the forensic analyst and the adversary improve their actions gradually in response to each other's action. We then present a game-theoretic perspective, formulating a game between the forensic analyst and the adversary that highlights the fundamental relation between the players.

6.1 An Evolutionary Perspective

In a security context, system defenders and attackers exploit vulnerabilities in each other's solutions and advance their own solutions. There is always an evolution between the two parties, which has been observed in many practical scenarios such as computer virus v.s. anti-virus competition [7] and the "arms race" for attacking v.s. securing online reputation systems [12]. In a similar spirit, such an evolution can also be observed in ENF analysis, resulting in strategies from simple to complex. Indeed, technical advancements listed below have taken place in this paper:

1. A forensic analyst extracts ENF at the fundamental frequency (e.g., 60Hz). This is sufficient since the ENF signal is stronger compared to the narrowband at the fundamental frequency so ENF extraction is accurate, and the forensic analyst does not examine harmonic frequencies that will incur additional complexity.

2. Given the practice in the previous step, an adversary naturally alters the ENF signal at the fundamental frequency using anti-forensic operations proposed in Section 3 such as ENF signal removal and embedding.

3. In the presence of the adversary, the forensic analyst is now motivated to extract the ENF signal from other harmonic frequencies to examine the inter-frequency consistency, at the cost of higher complexity.

4. In response to the forensic analyst, the adversary also has to make cohesive changes to the ENF signal at higher harmonic frequencies. However, the adversary takes the risk of distorting the host audio signal and has a higher chance of being caught by the reference-based detection.

5. The forensic analyst now has to take into account more advanced detection methods at additional costs, such as checking the spectrogram consistency.

6. In response to the forensic analyst, the adversary can increase the spectrogram consistency via envelope adjustment. However, this may sacrifice the ENF fidelity.

7. Given that the adversary has addressed the blind detection methods, the forensic analyst can resort to non-blind detection such as checking the ENF embedding statistics, which involves the use of reference signals. The means that the forensic analyst can improve his/her capability if more resources are available.

8. In response to the forensic analyst, the adversary now improves the ENF embedding fidelity by matching the statistics at the analyst's disposal. However, we have seen that matching a subset of the statistics may lead to mismatch of other statistics, and it is difficult to perfectly replicate the authentic ENF formation process.

9. Now the forensic analyst has to seek additional anti-forensics detection methods. The interplay continues.

As this paper is the first step regarding anti-forensics and countermeasures of ENF analysis, we expect that such evolution will continue, and increasingly more sophisticated anti-forensic strategies and countermeasures will emerge, pushing forward the research that improves the security of the ENF-based time stamp.

6.2 A Game-Theoretic Perspective

The interplay between the forensic analyst and adversary in the ENF analysis can be further understood under a game-theoretic framework that is extended from the work in [11]. Consider that the forensic analyst (denoted by Player FA) extracts the ENF signal at the fundamental frequency (e.g., 60Hz). Due to the possible presence of the adversary (denoted by Player AD) who would perform anti-forensic operations upon the audio signal, Player FA cannot simply trust the extracted ENF signal until an anti-forensics detector confirms its authenticity. As an illustrative example, assume that Player AD performs ENF embedding proposed in Section 3.2.

A detector can be characterized by its structure and performance metrics. In this paper, we consider a composite construction of anti-forensics detectors. Specifically, consider a total of N individual detectors D_i, $1 \leq i \leq N$, each relying on different signal characteristics to generate a binary output (T/F) with respect to an input audio signal. An overall anti-forensics detector D_{all} can be constructed using a simple OR-rule:

$$D_{\text{all}} = \begin{cases} T, & \text{if } D_i = T \text{ for any } 1 \leq i \leq N, \\ F, & \text{otherwise.} \end{cases} \tag{13}$$

Note that in practice, the detector has constraints on its affordable complexity and the available resources, which determine the individual detectors that can be incorporated into the overall detector. The performance of the detector is measured in terms of its detection probability and false alarm probability. It is well known in detection and decision theory that there is a trade-off between these two probabilities of a given detector: the false alarm probability increases as the detection probability increases. For a total false alarm probability $P_{f,\text{all}}$ allowed for D_{all}, Player FA's strategy selects and configures individual detectors so that the total false alarm probability equals to $P_{f,\text{all}}$.

On the other hand, in response to Player FA's anti-forensics detection, Player AD will seek to hide the traces of anti-forensics. There may also be complexity and resource constraints imposed on Player AD's actions, and Player AD has to select his/her strategy under the constraints so that Player FA's detection capability is minimized while the embedded ENF signal is maximally preserved. Given a pair of Player FA and AD's strategies, the utility that Player FA will maximize is the total detection probability of anti-forensics $P_{d,\text{all}}$. In contrast, Player AD's utility is to minimize $P_{d,\text{all}}$, with additional penalty when distortion is introduced to the ENF signal that Player AD intends to embed.

The specific operations proposed in Section 4 and 5 can be studied under the game-theoretic formulation. In terms of Player FA's detector construction, if more strict constraints on complexity and resources are imposed, then Player FA may only use the low-complexity inter-frequency consistency check as the anti-forensics detector. If a higher complexity is permitted, then the spectrogram consistency detector can be incorporated into the overall detector. Furthermore, if the resources accessible to Player FA are enhanced, for example via the reference signal or via an improved understanding of the ENF mechanism, then Player FA can construct an even more sophisticated detector. On Player AD's side, altering ENF in multiple frequencies is cost-effective against the inter-frequency consistency check, but cannot resist other types of anti-forensics detection. Nonetheless, if higher complexity is allowed for Player AD, he/she can employ envelope adjustment to reduce the anti-forensics detection probability, although at the same time, the embedded ENF signal may suffer from distortion. Similar to Player FA, if more resources, in particular an improved knowledge of the ENF embedding, are available to Player AD, then Player AD can also improve the anti-forensic capability. Our ongoing work builds on the understanding of the strategy space of Player FA and Player AD from this paper, and is evaluating the utilities of both players either analytically or numerically, so that the Nash equilibrium strategies can be determined. This will lead to an understanding of the stable interplay pattern between the two players.

7. CONCLUSIONS

The time stamp based on the Electrical Network Frequency (ENF) has been shown to be a promising tool for authenticating digital audio and video recordings. However, as in many other scenarios of computer and communication security, the existence of adversaries raises a serious concern regarding the security of the ENF-based time stamp and makes it crucial to understand and address possible vulnerabilities in ENF analysis against anti-forensic actions. In this paper, we have investigated anti-forensic operations that can remove and alter the ENF signal present in a host audio signal. We have developed a mathematical framework for ENF modification, which not only entails the effectiveness of ENF modification and challenges of anti-forensics detection, but also motivates detection methods from a forensic analyst's point of view. Improvements over the anti-forensic operations in response to the anti-forensics detection are further proposed and their corresponding trade-offs are discussed. To understand the highly dynamic nature of the forensic analyst-adversary interplay, we have developed an evolutionary perspective and a game-theoretic perspective, which can be used to characterize a wide range of actions that may take place.

Our ongoing work includes experiments that cover a variety of testing conditions, including different geographic areas and recording devices, and the evaluation of utility functions associated with different players in our proposed game formulation. In view of the potential employment of ENF analysis for media data authentication, we envision that its security will receive increasing attention, and research along this direction will contribute to more secure and reliable time stamp schemes based on ENF analysis.

8. ACKNOWLEDGEMENTS

The work was supported in part by the research grant NSF #0824081. We thank Dr. Avinash L. Varna and the anonymous reviewers for their valuable comments.

9. REFERENCES

[1] Wikipedia article on"mains hum". http://en.wikipedia.org/wiki/mains_hum.

[2] W. H. Chuang and M. Wu. Robustness of color interpolation identification against anti-forensic operations. In *Proc. of 14th Information Hiding Conference*, 2012.

[3] R. Garg, A. L. Varna, and M. Wu. "Seeing" ENF: Natural time stamp for digital video via optical sensing and signal processing. In *Proc. of ACM Multimedia*, Nov. 2011.

[4] C. Grigoras. Applications of ENF criterion in forensics: audio, video, computer, and telecommunication analysis. *Forensic Science International*, 167:136–145, Apr. 2007.

[5] S. Haykin. *Communication Systems*. Wiley Publishing, 5th edition, 2009.

[6] M. Kirchner and R. BÃűhme. *Digital Image Forensics*, chapter on Counter-Forensics: Attacking Image Forensics. Springer, 2012.

[7] C. Nachenberg. Computer virus-antivirus coevolution. *Communications on the ACM*, 40:46–51, Jan. 1997.

[8] A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-time Signal Processing*. Prentice-Hall, 2 edition, 1999.

[9] D. P. N. Rodriguez, J. A. Apolinario, and L. W. P. Biscainho. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. *IEEE Trans. on Information Forensics and Security*, 5(3):534 –543, Sep. 2010.

[10] R. W. Sanders. Digital authenticity using the electrical network frequency. In *Proc. of 33rd AES Int. Conf. on Audio Forensics, Theory and Practice*, Jun. 2008.

[11] M. C. Stamm, W. S. Lin, and K. J. R. Liu. Temporal forensics and anti-forensics for motion compensated video. *IEEE Trans. on Information Forensics and Security*, 7(4):1315 –1329, Aug. 2012.

[12] Y. Sun and Y. Liu. Security of online reputation systems: The evolution of attacks and defenses. *IEEE Signal Processing Magazine*, 29(2):87 –97, Mar. 2012.