Abstract

This work deals with the evaluation of the performance of a Digital Watermarking system under different attacks, and finding the needed modifications to improve its performance. Digital Watermarking is a copyright protection mechanism. A complete Digital Watermarking system consists of two subsystems - one for embedding and the other for verification. The embedding mechanism enables the owner to mark his media by adding a digital watermark which is a signal that is created by the original media and a unique global key (owner identifier). The embedding process does not reduce the audio quality. This is done by creating a signature which is spectrally shaped according to the human auditory system's masking model. The watermarked signal is the result of the addition of the signature signal to the original. The Dead-lock problem, i.e., multiple ownership claims, is solved by keeping the original media or parts of it by its owner for future ownership claims. The verification mechanism uses the original media and enables the owner to check for the existence of the watermark in a tested media.

The verification mechanism uses the original signal in the following way: First, the Watermark is calculated based on the original signal and the global key in the same way it was calculated in the embedding system. Next, the original signal is used to extract the watermark from the tested signal. If the tested signal is watermarked then the subtraction of the original signal from it will result in the watermark signal itself. Comparing the correlation result, between the two watermark signals, to a given threshold value results in a decision on the existence/non-existence of the claimed watermark in the tested signal.

In this work we implemented such a system and we deal with ownership claims of attackers, i.e., people that duplicate copyright media, attack the signature and claim for their ownership of the media.

During recent years, several standardization committees started working on the definitions of different types of attacks to enable the evaluation of watermark systems. The number of possible attacks is obviously infinite and thus, one could have the impression that this is a “lost case”. However, it seems that the main efforts (as it seems from the standard committees’ proposals) are against attacks that are based on “off the shelf” utilities, such as, Cool-Edit Pro.

Our fundamental assumption is that the attacker is limited in the sense of preserving the audio quality. However, this work does not pretend dealing with all possible attacks, which are uncountable.

On top of dealing with attempted attacks as descried in the previous section, the watermarking system should be able to handle naive attacks, i.e., modification done by an authorized person (who paid for the usage of the media) like a filtering process for personal compliance. The watermark system should be able to verify a signature in non-authorized copies of this filtered media.

The main steps to deal with such actions is to find the characteristics of the attacker's system and then to modify the detection system so that the signature detection probability is maintained under these attacks. As previously mentioned, our fundamental assumption is that the attacker is limited in the sense of preserving the audio quality. We are dealing first with naive attacks (gain, offset, equalization and compression) and then with more sophisticated attacks (fixed and time varying all-pass filters, non-linear distortion, noise and more). For all these attacks we first determined the maximum distortion level by listening tests. We then focused on attacks that do not reduce the audio quality but degrade the performance of the verification system. Then, we defined the attacker's global model and estimated its parameters using system identification methods (including LS and normalized-LMS) that uses both the tested signal and the reference signal.

First, we deal with attacks composed of linear filtering. These attacks include usage of Band-pass filters, All-pass filters, and time varying All-pass filters. Applying system identification (LMS-based) techniques significantly improves the results for these attacks.

The main contribution of the work is in dealing with attacks composed of non-linear filtering. The main assumption is that these attacks are composed of a cascade of linear filters and memory-less soft non-linearities. The non-linearity is approximated by piece-wise linear segments. An adaptive model that composes linear filter estimation, as in ordinary LMS system, together with adaptive estimation of the non-linearity is derived. The proposed method is unique in having to estimate only a small number of coefficients as compared to other estimation methods for non-linear filtering, such as the Volterra method. The algorithm was tested under an attack of linear filtering followed by a memory-less non-linearity and was found to significantly improve the ability of the verification system to verify the signature

under such an attack. LMS update equations are derived also for more complex non-linear filtering models such as Linear - Non-linear - Linear (LNL) or Non-linear - Linear - Non-linear (NLN) models.

The last part of this work deals with the determination of the threshold value used in the verification system. This threshold was set according to the desired False-Alarm rate using histograms that are based on correlation results. We also found that the system performances depend on the length of the segment used for the correlation calculation.