Chapter 7
Watermark-Only Security Attack on DM-QIM Watermarking: Vulnerability to Guided Key Guessing

B. R. Matam
Aston University, UK

David Lowe
Aston University, UK

ABSTRACT

This paper addresses the security of a specific class of common watermarking methods based on Dither modulation-quantisation index modulation (DM-QIM) and focusing on watermark-only attacks (WOA). The vulnerabilities of and probable attacks on lattice structure based watermark embedding methods have been presented in the literature. DM-QIM is one of the best known lattice structure based watermarking techniques. In this paper, the authors discuss a watermark-only attack scenario (the attacker has access to a single watermarked content only). In the literature it is an assumption that DM-QIM methods are secure to WOA. However, the authors show that the DM-QIM based embedding method is vulnerable against a guided key guessing attack by exploiting subtle statistical regularities in the feature space embeddings for time series and images. Using a distribution-free algorithm, this paper presents an analysis of the attack and numerical results for multiple examples of image and time series data.

1. INTRODUCTION

Data hiding, embedding information into digital media for the purpose of identification, annotation and copyright is a form of steganography (Bender et al., 1996). Embedding information into digital data (called cover data) is known as watermarking and the embedded information (watermark) signifies invisible information that can be detected and retrieved by authorised personnel or systems designed for that purpose (Lin et al., 2005). The watermarks hence extend the information content of the cover work. They are utilised as a means of securing the rights of the owner of the digital data, authentication of the source or as a tracing mechanism. Kalker (Kalker, 2001) defines the

DOI: 10.4018/978-1-4666-1758-2.ch007

security of a watermarking system as the inability of unauthorised users to remove, detect and estimate, write or modify the raw watermarking bits. Comesaña et al. (Comesaña et al., 2005) take a more restrictive view of the security of a watermarking system linking it to the gaining of knowledge about the secrets of the system in addition to destroying the embedded message. The current paper is consistent with both views since the random key locations will be estimated.

We consider the case where the watermark bits are embedded in selected samples of the cover data wherein the indices of the selected samples are referred to as the secret embedding key. The work presented in (Giakoumaki et al., 2006) states that the watermarks are secure because the key represents a random vector which cannot be easily guessed. Evaluation of the security of watermark embedding methods is still a nascent field though some limited analysis of the security of different lattice based embedding methods exists in the literature.

Based on Diffie-Hellman's Terminology, Cayre et al. (Cayre et al., 2005) grouped the attacks on watermarked content as 1. Watermark only attack (WOA) - wherein the attacker has access to a set of watermarked host data. 2. Known-message attack (KMA) - where the attacker has access to a set of watermarked content, watermarked with the same key and the associate messages. 3. Known-original attack (KOA), where the attacker has access to both the watermarked content and the original un-watermarked content.

Utilising the definition of watermarking security given by Kalker (Kalker, 2001) the work presented in this paper is a WOA analysis of the security of the DM-QIM method explicitly pertaining to the detection of the secret key based on the assumption that the cover work has been identified as watermarked content and the watermark embedding method is known. This paper proposes an efficient distribution-independent approach to attacking watermarks embedded using transform domain DM-QIM. It employs a method to estimate the probable location of the hidden information when only a single copy of the watermarked content is available, an extreme case of the WOA class of attacks, for both discrete wavelet transform (DWT) and independent component analysis (ICA) domain based DM-QIM watermarking methods. DWT and ICA is representative of current state-of-the-art transform domain methods where the signal space is spanned by either fixed orthogonal or data-adaptive non-orthogonal basis functions. Despite the use of a random key for message locations in the transform domain, departures from the natural distribution of the covertext are induced which are amenable to non-parametric density estimation models. The results illustrate the fallibility of DM-QIM against guided key guessing attacks for both image and time series data.

2. KEY SECURITY CLASSES

The concept of key security for QIM based watermarking has been investigated by various authors (Kalker, 2001; Holliman et al., 1999; Bas & Hurri, 2006; Cayre et al., 2005; Pérez-Freire et al., 2006; Cayre & Bas, 2008; Pérez-Freire & Pérez-González, 2007). Overwriting the information in the watermarked cover to partially or completely destroy the original information is possible if the embedding method is known. Even if the exact secret embedding key is not known, it is possible to destroy the embedded message by randomly overwriting the watermarked content to a large extent (although this could be interpreted as robustness rather than security). Cox et. al. in (Cox et al., 1996) claim that $O(\sqrt{l \,/\, ln(l)})$ similar watermarks must be added to the watermarked content to destroy the original watermark, where l represents the number of most perceptually significant frequency components of an image's discrete cosine transform used to embed the original watermark. The watermark used in their