Analysis and Design of Secure Watermark-Based Authentication Systems

Chuhong Fei, Student Member, IEEE, Deepa Kundur, Senior Member, IEEE, and Raymond H. Kwong, Member, IEEE

IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 1, NO. 1, MARCH 2006
Chuhong Fei (S’04) was born in Zhejiang, China. He received the B.E. and M.E. degrees from Xi’an Jiaotong University, China, in 1994 and 1997, respectively, and the M.A.Sc. degree in electrical and computer engineering from the University of Toronto, Toronto, ON, Canada, in 2001. He is currently pursuing the Ph.D. degree at the University of Toronto.

His research interests include multimedia security, data hiding, multimedia signal processing, and information theory.

Deepa Kundur (S’93–M’99–SM’03) was born in Toronto, ON, Canada. She received the B.A.Sc., M.A.Sc., and Ph.D. degrees, all in electrical and computer engineering, in 1993, 1995 and 1999, respectively, from the University of Toronto.

In January 2003, she joined the Electrical Engineering Department at Texas A&M University, College Station, where she is an Assistant Professor, and a member of the Wireless Communications Laboratory. From September 1999 to December 2002, she was an Assistant Professor at the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, where she was Bell Canada Junior Chair-holder in Multimedia. Her research interests include multimedia and network security for digital rights management, video cryptography, sensor network security, steganography, covert communications, and nonlinear and adaptive information processing algorithms.

Dr. Kundur has been on numerous technical program committees and has given tutorials at ICME 2003 and Globecom 2003 in the area of digital rights management. She was a Guest Editor for the PROCEEDINGS OF THE IEEE Special Issue on Enabling Security Technologies for Digital Rights Management. She was the recipient of the 2002 Gordon Slemon Teaching of Design Award and the 2002 Best Electrical Professor Award awarded by the ECE Department at the University of Toronto.
Raymond H. Kwong (M’75) was born in Hong Kong in 1949. He received the S.B., S.M., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1971, 1972, and 1975, respectively.

From 1975 to 1977, he was a visiting Assistant Professor of Electrical Engineering at McGill University, Montréal, QC, Canada, and a Research Associate at the Centre de Recherches Mathématiques, Université de Montréal. Since August 1977, he has been with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto, Toronto, ON, Canada, where he is now Professor and Associate Chair for Undergraduate Studies. His current research interests include estimation and stochastic control, adaptive signal processing and control, fault diagnosis, discrete event systems, hybrid systems, and multimedia security.
Outline

- Introduction
- Previous works on semi-fragile watermarking
- Requirements for semi-fragile watermarking
- Our approach to semi-fragility: the framework
- An example: Semi-fragility to JPEG compression
- Conclusion
Introduction

- Problem addressed: content authentication using coding-based scheme in which a watermark is used to assist in verifying the integrity of its associated multimedia data.
- Goal: to authenticate the content not its specific format representation
- The embedding of an invisible watermark in a host signal has two main objectives:
  - to alert a party to unacceptable distortions on the host,
  - to authenticate the legitimate source.
Introduction (Cont’d)

- Two groups of possible distortions: legitimate and illegitimate distortions
- Some applications of authentication watermarking: cameras, digital insurance claim evidence, medical image archiving, journalistic photography, and digital rights management systems
Previous approaches on semi-fragile watermarking

- Lin et al. [13] propose a semi-fragile watermarking technique based on extending a simple spread spectrum watermarking method with a modified detector
- Yu et al.[11] use a mean-quantization-based fragile watermark to detect malicious tampering while tolerating some incidental distortions.
- Kundur-Hatzinakos [18]: Telltale tamper-proofing to determine the extent of modification both in the spatial and frequency domains of a signal using a statistics-based tamper assessment function.
Previous approaches on semi-fragile watermarking (Cont’d)

- The above approaches often:
  - tune a robust watermarking scheme to achieve fragility,
  - address specific distortions (usually compression).

- Although these techniques, such as SARI, work well under a class of attacks, their ad hoc design nature focusing on resilience to a specific distortion limits their portability to different applications.
Requirements for semi-fragile watermak-based authentication systems

- Robustness and fragility objectives should be simultaneously addressed
- The authentication system must be secure to intentional tampering
- Embedding must be imperceptible
- Embedding and authentication algorithms must be computationally efficient.
Semi-fragile authentication framework

- Establish framework to quantify tradeoff between:
  - Robustness to legitimate distortions
  - Fragility to illegitimate distortions
Semi-fragile authentication model

- Authentication embedding procedure is described as a function $f$ which takes the host $S$ and the key $k$ as inputs to produce the authenticated signal $X$:
  $$X = f(S, k)$$
  (function used by the transmitter)

- The receiver uses the corresponding binary function $g(Y, k)$ to decide whether the received signal $Y$ is authentic ($g(Y, k) = 1$) or not ($g(Y, k) = 0$)

- The authentication distortion is $D = \frac{1}{nE}\{\|S-X\|\}$
Error detection code

- A subset of signal space is used to communicate with the receiver;

- The receiver authenticates a signal by verifying if it is in the subset;
Error detection code (Cont’d)

- We have two types of errors:
  - Type I error, often called false positive error: application of a legitimate distortion on $X$ results in failure to verify the received signal. (Robustness)
  - Type II error, often called false negative error, occurs when $X$ has been illegitimately tampered but the received signal $Y$ is wrongly verified by the receiver as authentic. (Fragility)
Coding approach

- Based on the verification model because authentication is essentially a detection problem knowing the shared key, whereas robust watermarking is basically a decoding problem for data communication.

- Given a secret $k$ in key space, $C(k)$ is the set of possible authenticated signals generated by Alice:
  $$C(k) = \{ f(S,k) \in \mathbb{R}^n | \forall S \in \mathbb{R}^n \}$$

- And the embedded function $X = f(S,k)$ is as follows:
  $$X = \arg \min_{x \in C(k)} ||S - x||$$
Coding approach (Cont’d)

Fig. 2. (a) Encoding set and verification region. The points marked with + are the encoding set \( C(k) \) for some \( k \). The admissible set \( \Omega \) is the shadowed area, which is a disk in this example. The shadowed region around points marked with + is the verification region \( \mathcal{E}(k) \). (b) Three types of distortions: (i) \( Z_1 \in \Omega \); (ii) \( Z_2 \notin \Omega \) and found to be inauthentic; (iii) \( Z_3 \notin \Omega \) and found authentic, leading to false negative error.
Secure Code Construction
(Lattice code)

- A lattice can be defined as a regular array of points in n-dimensional Euclidean space.

\[ \Lambda = \left\{ \sum_{i=1}^{n} a_i g_i \mid a_i \in \mathbb{Z} \right\} \]
Fig. 3. Lattice codes for semi-fragile authentication. (a) All these points form the lattice $A$. The fundamental Voronoi region $\mathcal{V}_0(A)$ is shown by the dotted shape. The admissible set $\mathcal{S}_2$ is the shadowed area, which is a disk in this example. The fundamental Voronoi region $\mathcal{V}_0(A)$ covers the admissible set $\mathcal{S}_2$. (b) Each encoding set is a subset of the lattice $A$. The points, marked with $+$, corresponds to an encoding set. The shadowed region around points marked with $\dagger$ is its verification region.
Secure code : Nested Lattice based MSB-LSB Scheme

- Given an n-dimensional nested lattice code \((\Lambda_1, \Lambda_2)\)
  - Where \(\Lambda_1\) is a subset of \(\Lambda_2\) \((\Lambda_2 \subset \Lambda_1)\)
  - \(\Lambda_1 = \Lambda_2 + [\Lambda_1 / \Lambda_2]\)

- From this decomposition, for any point \(\lambda_1 \in \Lambda_1\) there exist \(\lambda_2 \in \Lambda_2\) and \(v \in [\Lambda_1 / \Lambda_2]\) such that \(\lambda_1 = \lambda_2 + v\)

- \(\lambda_2\) Correlate to the MSB component and the LSB component. The encoding set is
  \[
  C(k) = \{\lambda_2 + H_k(\lambda_2) | \lambda_2 \in \Lambda_2\}
  \]
MSB-LSB decomposition approach

Fig. 5. Authentication and verification processes for MSB authenticator generation and LSB embedding scheme.
Simulation Results of MSB Approach

\[ \Omega_n = \{ [x_1, x_2, \cdots, x_n] \in \mathbb{R}^n | x_i \in [-\frac{1}{2}, \frac{1}{2}] , \forall i = 1, 2, \cdots, n \} . \]

**Figure 3.** Uniformly distributed noise.
Simulation Results of MSB Approach

Figure 4. Error probability curves: AWGN.
An example: JPEG Compression

- System is designed to be robust to high quality JPEG compression but fragile to low quality compression.

- The admissible set $\Omega = \text{n-dimensional cubic set with edge length } \Delta$,

- The nested lattice code $= (\mathbb{Z}/2\mathbb{Z})\Delta$

- Incorporation of a hashed message authentication code (HMAC) in the MSB-LSB decomposition.
Original and watermarked images
Legitime and illegitime noises

2009/12/30
Multimedia Security ---- Moustapha BANDE
Conclusion

- General coding-type framework which provides useful constructive tools in the analysis and design of semi-fragile watermarked-based authentication systems.
- Authentication is important to protect valuable or legal images;
- Digital watermarking is a successful solution for image authentication;
- Effectiveness of nested lattice codes in achieving design objectives (robustness, fragility, security) has been shown;
- Semi-fragility depends on characterization of allowable distortions by an admissible set $\Omega$. 
Following work


- Future work
  - Channel distortion specification in different applications
  - Incorporate perceptual models
  - More general distortions