Digital Invisible Ink: Revealing True Secrets via Attacking

Chun-Hsiang Huang, Yu-Feng Kuo and Ja-Ling wu
Communications and Multimedia Lab.,
Room. 541, CSIE Building,
National Taiwan University
886-2-23625336 ext 541
bh@cmlab.csie.ntu.edu.tw

ABSTRACT

A novel steganographic approach analogy to the real-world secret communication mechanism, in which secret messages are written on white papers using invisible ink like lemon juice or milk and are revealed only after the papers are heated, is proposed. Carefully-designed informed embedders now play the role of "invisible ink"; some pre-negotiated attacks that can be provided by common content-processing tools correspond to the required "heating" process. Theoretic models and feasible implementations of the proposed digital-invisible-ink watermarking approach based on both blind-detection spread-spectrum watermarking and quantization watermarking schemes are provided. The proposed schemes can prevent the supervisor from interpreting secret messages even when the watermark extractor and decryption tool, as well as session keys, are available to the supervisor. Furthermore, secret communication systems employing the DII watermarking schemes can aggressively mislead the channel supervisor with fake watermarks and transmit genuine secrets at the same time.

Keywords digital invisible ink, steganography, DII watermarking, hiding watermark in watermark

The proposed digital-invisible-ink watermarking system, abbreviated as DII watermarking, is in fact one variant of the informed-embedding watermarking systems since some pre-negotiated attacks are taken into consideration during watermark embedding. However, the content is not actually attacked before or in the transmission channel. Instead, when the intended receiver receives the marked content, the pre-negotiated attacks are then performed via commonly available content processing tools. Since the pre-negotiated attack is applied after the supervised transmission, the quality degradation caused by the attacks can violate the fidelity constraint that most watermarking schemes obey.

In DII watermarking scheme, **the existence of pre-negotiated attack noises is necessary for the successful detection of the hidden payload**. Both DII models of blind-detection spread-spectrum watermarking and quantization watermarking schemes are investigated.

In the case of blind-detection watermarking schemes, the angle between the noise caused by pre-negotiated attack and the pseudo-random watermark vector must be within the range of [90^o, 90^o]. Furthermore, the magnitude of the projection of the pre-negotiated attack noise in the direction of the watermark vector must be larger than the guaranteed amount over the detection threshold. The first condition is naturally satisfied due to the pseudo-randomness of spread-spectrum watermark vectors and the second condition is achieved by an iterative informed-embedding methodology in which scale factors of watermarks are adequately adjusted. The fidelity model of DII spread-spectrum watermarking is also provided.

As for quantization watermarking, the original cover work must be quantized to be within the quantization cell corresponding to the wrong reconstruction point first, and then the pre-negotiated attack should distort the marked work along the direction from the wrong reconstruction point to the correct one. In addition, the magnitude of the pre-negotiated attack must be significant enough so that correct payloads can be extracted. The block-DCT based quantization watermarking scheme utilizing the inequality relationships between magnitudes of DC and AC coefficients, together with certain attacks that either increase or decrease the magnitude of most DCT coefficients, inherently fit the DII quantization watermarking model.

Two important steganographic applications of the proposed DII watermarking methodology are illustrated. First, if the supervisor can check the operational environment of the intended receiver, secret communications may be discovered as long as the steganographic/cryptographic modules are not properly concealed. In this case, if computation resources are limited or strict system policies are enforced, only simple scrambling schemes or proprietary encryption modules can be used and the secret messages may be easily revealed by the supervisor. To make matters worse, the supervisor can pretend to be unaware of the secret communication and then eavesdrop all forthcoming secrets.

The DII spread-spectrum watermarking scheme may be adopted to alleviate the prescribed secret leaking problem. If the necessary pre-negotiated attack (provided by general-purposed content processing tools) is not performed, the watermark will never be correctly extracted. Experimental results also show that adequately introducing more-than-one attacks can further increase the security of steganographic system. Moreover, to avoid the risk that messages may be secretly eavesdropped by supervisors or for scenarios where insensitive watermarks are allowed, the DII quantization watermarking scheme can be utilized to hide genuine secrets in a cover watermark. In this way, the sender can mislead the supervisor and transmit the genuine secrets to the intended receiver at the same time.

In summary, the digital-invisible-ink watermarking approach can improve the capability of existing steganographic architectures. In fact, many other important applications of existing watermarking schemes can also benefit from the DII watermarking approach and will be extensively exploited in our future works.

Acknowledgement This work was supported by NSC-94-2752-E-002-006-PAE, National Science Council, R. O. C.