Lossless Text Steganography in Compression Coding

Chin-Feng Lee¹ and Hsing-Ling Chen²

¹ Department of Information Management,
   Chaoyang University of Technology,
   Taichung 41349,
   Taiwan, R.O.C.
   lcf@cyut.edu.tw
² Department of Applied Foreign Languages,
   Chaoyang University of Technology,
   Taichung 41349, Taiwan, R.O.C.
   hsinglingchen@gmail.com

Summary. This work presents a novel text steganography scheme in compression domain using a lossless compression coding which called variable Huffman coding. The secret message is embedded into compression codes. The goal of the proposed scheme is covert communication by means of text files, providing high embedding capacity, improving security of the embedded secret message and reducing transmission cost. Additionally, the original text files can be reconstructed without any distortion after the embedded secret message is extracted. In the proposed scheme, each leaf of variable Huffman tree can be used to convey a secret bit at least. According to the practical application, the embedding rate for each leaf is to be increased, i.e, the embedding capacity of the proposed scheme is scalability. The secret keys are employed to generate the stego-compression-code for each leaf of variable Huffman tree to protect the embedded message. The extracting embedded message will be meaningless without the secret keys being known. Furthermore, the size of stego-compression-code is less than the size of original file plus secret message in the proposed scheme so that the transmission cost can be reduced. Additionally, experimental results show that the performance of our method is indeed superior to Chang et al.’s scheme as well as Chen and Chang’s scheme in terms of embedding capacity and transmission cost.

8.1 Background

The objective of information hiding provides covert communication avoiding unauthorized users attention, attack or tampered. A secret message can be concealed in a cover medium, forming a stego-medium. The secret message can then be covertly delivered successfully because the human eye cannot easily distinguish between the cover medium and stego-medium. Nowadays, digital media, such as videos, images, text files, and sounds, are commonly used as cover media to convey secret messages
in information-hiding methods. Many information-hiding schemes have been developed [1]–[15]. We most use a digital image as a cover medium to conceal a secret message because a digital image has a special property in that a pixel is similar to its neighboring pixels. Thus, secret data can be concealed easily in a digital image by slightly modifying pixel values. Once a secret message is embedded in a cover image, the cover image is called a stego-image. Secret data can then be conveyed for information sharing as stego-image distortion is insignificant. The peak-signal-to-noise ratio (PSNR) is typically utilized to assess the degree of distortion when a digital image is used as a cover medium. Distortion is generally indiscernible to the human eye when the PSNR exceeds 30 dB.

Today, a great number of information-hiding schemes are developed by different research communities. These schemes can be briefly classified into four main types, i.e. covert channels, steganography, anonymity, and copyright marking, as shown in Fig. 8.1 which can be referred to [16]. However, most information-hiding schemes recently focus on technical steganographic schemes [1, 5, 11, 15, 17] or watermarking schemes [2, 3, 18, 19, 20]. These two categories typically work for the media in the spatial domain or frequency domain. The goal of digital watermarking techniques is to protect copyright ownership. Consequently, their robustness, rather than embedding capacity, is the most important issue. Conversely, steganographic schemes are primarily used for sharing secret messages. Hence, the goals of steganography are high embedding capacity and low distortion. Numerous steganographic schemes which belong to technical steganography have been developed, but the linguistic steganography schemes are seldom discussed. Since a slight modification in text content can be easily discovered so that most efforts work on the field of technical steganography rather than linguistic steganography.

Fig. 8.1. A classification of information-hiding schemes [16].