Speech Content Authentication Scheme based on High-Capacity Watermark Embedding

Fang Sun, Xinyang Normal University, College of Computer and Information Technology, Xinyang Henan, China

Zhenghui Liu, Xinyang Normal University, College of Computer and Information Technology, Xinyang Henan, China

Chuanda Qi, Xinyang Normal University, College of Mathematics and Information Science, Xinyang Henan, China

ABSTRACT

The existed content authentication schemes based on digital watermark have some shortcomings. In order to solve the problems, a speech content authentication scheme based on high-capacity watermark embedding is proposed, and the high-capacity embedding method is discussed. Firstly, speech signal is framed and segmented, and the samples of each segment are scrambled. Secondly, DCT is performed on the scrambled signal, and low-frequency coefficients are selected as the watermark embedding domain. Lastly, frame number is mapped to a sequence of integers and embedded into the domain based on the embedding method. Theoretical analysis and experimental evaluation results show that the proposed algorithm is inaudible, robust to desynchronous attacks, enhances the embedding capacity, and improves the security of watermark system.

KEYWORDS

Digital Forensics, Digital Watermark, Discrete Cosine Transform, Speech Authentication, Tamper Location

INTRODUCTION

Digital signal has replaced traditional simulated signal to become the most popular information carrier in communication. However, it's easy to edit, attack and forge the digital information for the increasingly rich of the multimedia editing tools.

In real life, digital speech signals are likely to cause attacker's interest and be maliciously attacked. For attacked signal, the expressed meaning is different to the original one. If recipient regards the attacked signal is an authentic one and acts according to the requirements, it may cause serious consequences (Zhang et al., 2015). Fortunately, the forensic technology based on digital watermarking (Akhaee et al., 2010; Pun & Yuan, 2013; Peng et al., 2013, Lei et al., 2013) gives a method to verify the authenticity of speech signal.

Digital audio watermarking schemes are usually used for protecting audio copyright (Xiang et al., 2006; Wang et al., 2011; Yamamoto & Iwakini, 2009; Salma et al., 2010; Wang, Healy & Timoney, 2011) and have been achieved an outstanding progress in recent years. In (Wang, Shi, Wang & Yang, 2016) authors proposed a robust audio watermarking method based on invariant exponent moments and synchronization code technique. Watermark generated by binary image is embedded into host audio. Experiment results demonstrated that the scheme is resistance against most attacks, and the binary image extracted from watermarked signal after being attacked is similar to the original one. If the scheme is used for authentication, most attacks will not be detected. In Bai et al. (2011), authors

DOI: 10.4018/IJDCF.2017040101

given the audio watermarking scheme based on SVD–DCT with the synchronization code technique. Binary image as watermark is embedded into the high-frequency band of the SVD–DCT block blindly. The scheme is robust against various common signal processing attacks. So, if the schemes (Wang, Shi, Wang & Yang, 2016; Bai et al., 2011) is used for authentication, most attacks will not be detected.

As a carrier to transmit information, the meaning of digital speech signal to express should be intact and authentic. For audiences and users, if they consider the attacked signal as the original one and act according to the instructions of the attacked signal, it may cause serious consequences. So, for digital speech signals, the method used for speech forensics is indispensable, which can be achieved by using digital watermark (Liu et al., 2016). Outstanding progress has been achieved in recently, while they are unsuitable for speech authentication (Liu, Huang, Sun, & Qi, 2016).

By using detection of multiple compression and encoder's identification, Korycki (2014) proposed an authentication scheme for compressed audio recordings. The compressed recordings are authenticated by evaluation of statistical features extracted from MDCT coefficients and other parameters obtained from compressed audio files, used for training selected machine learning algorithms. Although the scheme enhanced the robustness and the effectiveness, it needs a large number of training data.

In Chen and Liu (2007), based on compression technique and codebook-excited linear prediction, authors proposed an authentication scheme. By the features extracted during compression process based on codebook-excite linear prediction, watermark bits are generated and embedded by quantifying the lest significant bits (LSBs). As we known, watermark method based on LSBs is fragile. Signal processing operations is regarded as hostile attack. In (Liu et al., 2016), based on digital watermarking, Liu et al. proposed an authentication and recovery algorithm for speech signal. Speech is compressed firstly, and the compressed signal as watermark is used for authentication and tamper recovery. Experimental evaluation results demonstrated that the scheme proposed improves the security and the accuracy of tamper location.

For watermarked schemes used for digital speech content authentication, there are some shortcomings: 1) Watermark is embedded by quantifying public features, and the features can be got by attackers, which result in that the schemes are vulnerable to feature-analysed substitution attack (Liu & Wang, 2014) 2) For some authentication schemes, they can detect the signal whether is attacked or not, but cannot locate the frames which are attacked (Wang & Fan, 2010). 3) For the schemes based on synchronization code (Wang et al., 2-011; Bai et al., 2011; Vivekananda et al., 2011), on the one hand, synchronization codes are vulnerable to substitution attack (Liu & Wang, 2014). On the other hand, for the schemes, they can detect the watermarked signal, but they cannot verify the authenticity of watermarked signal detected.

Considering above problems, we give the speech content authentication scheme based on high-capacity watermark embedding, in order to improve the security and enhance the tamper location accuracy of authentication scheme. In this paper, speech signal is framed and segmented, and the samples of each segment are scrambled. Then DCT is performed on the scrambled signal, and low-frequency coefficients are selected as the watermark embedding domain. Frame number is mapped to a sequence of integers, which are as watermark and embedded into the low-frequency coefficients. Content is authenticated by comparing with the integers extracted from each frame. For the scheme proposed, watermark embedding domain is secret for attackers, and it is difficult to get the features used to embed watermark. Comparing with the existed watermark algorithms, the scheme proposed enhances the embedding capacity, and improves the security of watermark system.

WATERMARK GENERATION AND EMBEDDING

Suppose the original speech signal is denoted by $A = \left\{a_l, 1 \le l \le L\right\}$.

Step 1: Preprocessing