Design and Implementation of Steganographic Speech Telephone

Zongyuan Deng, Zhen Yang, Xi Shao,
Ning Xu, Chao Wu, and Haiyan Guo

Nanjing University of Posts and Telecommunications,
210003 Nanjing, China
{y050919, yangz, shaoxi, y050921, y050922, y050902}@njupt.edu.cn

Abstract. This paper explains the work to design an information hiding based secure communication system, named covert speech telephone (CST). The overall system is designed over the internet using UDP protocol. Based on a GUI (graphical user interface) software, CST is possible to execute two optional secure modes. It is a completely digital system with high speech quality. Practical effects show that CST can meet the requirement of real-time secure communication. This new technique can effectively guarantee information security in VoIP system.

Keywords: steganography, information hiding, Least Significant Bit (LSB).

1 Introduction

Different from conventional encryption-based secure communication [1], this paper explores the technique of information hiding to guarantee communicating securely by concealing both the contents of information and its existence during the transmission process [2], which ensures that it is not easy to detect the presence of a secret in the mixed message. The main strength of this paper is that we have implemented the whole system. We will discuss the detailed descriptions in the following section.

2 System Architecture

2.1 Proposed Steganographic Scheme

Based on the stegonagraphic and watermarking ideas, we successfully design two approaches optional for real-time CST.

Figure 1 shows the proposed schematic diagram. It can be readily seen from the figure that the first approach is the prevailing LSB-based (Least Significant Bit Substitution) steganographic method [3] combined with modified CELP (Code Excitation Linear Prediction) codec. In addition, CST applies the technique of scrambling to improve security.
To meet the requirement of real-time software implementation, some modifications is done on the basis of ITU G.729A [4]. These modifications do not obviously degrade the speech quality. Instead of applying a pre-selection process, the proposed scheme directly searches the optimum pitch gain by minimizing equation 1.

$$E = \left\| x - g_p y - g_c z \right\|^2 = x^T x + g_p^2 y^T y + g_c^2 z^T z - 2g_p x^T y - 2g_c x^T z + 2g_p g_c y^T z \cdot$$

(1)

Where $x$ is the target vector, $y$ is the filtered adaptive-codebook vector, and $z$ is the filtered fixed codebook vector, $g_p$ is the gain of adaptive-codebook, and $g_c$ is the gain of fixed-codebook [4].

In addition, we also simplify the construct of matrix $\Phi$ in the subroutine of algebraic codebook search.

As viewed above, the most difficulty of the real-time voice information hiding system lies in real-time processing. On the other hand, many transparency and robust watermarking schemes suggest us converting the real-time steganographic message to watermarking information and using robust data hiding scheme to implement the steganographic system. Thus, the second idea is to use speech recognition (so the data rate is reduced dramatically) followed by a watermarking system to add the data to the non-secret speech. From the perspective of information theory, the reason why speech recognition can compress data is that human speech consists of both semanteme and some identity information of the speaker such as tone and emotion. In military secure communication system, these characteristics of speaker are redundant compared to the semanteme which is possible to be discarded. However, the accuracy of speech recognition cannot be guaranteed to be perfectly accurate at present. Therefore, we can make reasonable assumption that for military purpose, the transmitted orders are possible to be limited in definite vocabularies or phrases. In such scenario, the speech recognition algorithm is possible to reach high accuracy. In addition, a synthesized decision-feedback system is applied to avoid the false recognition. This system synthesizes speech according to the former recognition result, followed by a yes-or-no verification procedure. Generally, the recognition