Chapter 10
Steganography in Thai Text

Natthawut Samphaiboon
Asian Institute of Technology, Thailand

Matthew N. Dailey
Asian Institute of Technology, Thailand

ABSTRACT

Steganography, or communication through covert channels, is desirable when the mere existence of an encrypted message might cause suspicion or provide useful information to eavesdroppers. Text is effective for steganography due to its ubiquity; however, text communication channels do not necessarily provide sufficient redundancy for covert communication. In this paper, the authors propose a novel steganographic embedding scheme for Thai plain text documents that exploits redundancies in the way particular vowel, diacritical, and tonal symbols are composed in TIS-620, the standard Thai character set. This paper provides a Thai text stegosystem following a provably secure construction that guarantees covertness, privacy, and integrity of the hiddentext message under meaningful attacks against computational adversaries. In an experimental evaluation, the authors find that the message embedding scheme allows 203 bytes of embedded hiddentext message per 100KB of covertext on average, and that the document modifications are not readily noticed by observers. The stegosystem is thus a practical and effective secure system for covert communication over Thai plain text channels.

INTRODUCTION

Privacy is a major concern for users of public networks such as the Internet. Traditionally, privacy is among the central concerns of cryptography, which achieves private communication through encryption. One problem with encryption, however, is that although it may hide the *contents* of a message, the mere *transmission* of an encrypted message may cause suspicion.

As a motivating example, consider a scenario in which a government officer from some country is sent to work in another country for purposes of deepening international relationships between the two countries. Any electronic communication with the guest could easily be monitored by the

DOI: 10.4018/978-1-4666-1758-2.ch010

host country's intelligence agency. If the guest is observed sending or receiving messages in encrypted form, he or she might be suspected of espionage. This could be illegal in some countries and could affect the relationship of mutual trust between the countries.

To avoid such suspicion, information hiding techniques collectively called *steganography* have received attention for a long time. Steganography dates at least as far back as ancient Greece (Ryder, 2004). In one story, a messenger had a message tattooed onto his shaved head, waited until his hair grew back, and then was sent to deliver the message. Since the secret message written on the messenger's head could not be read except by those who knew what to look for, his mission was successful. In modern times, steganography, which attempts to hide the existence of message transfer, has become a new major concern for the users of public communication networks.

The vast majority of practical steganographic schemes are instances of *embedding steganography*, in which a secret message is embedded in a given cover document to produce a normal-looking output. The secret message, cover document, and output are called the *hiddentext*, *covertext*, and *stegotext*, respectively. The covertext could be any kind of digital media, such as information in a Web site, Web board, blog, chat session, or email message. As such, steganographic schemes have been proposed for embedding secret messages in images (Lou & Liu, 2002), video (Su, Hartung, & Girod, 1998), audio (Cvejic & Seppänen, 2004), and text in multiple languages (Amano & Misaki, 1999; Brassil, Low, & Maxemchuk, 1999; Huang & Yan, 2001; Kim & Oh, 2004; Kim, Moon, & Oh, 2003; Muhammad, Rahman, & Shakil, 2009; Samphaiboon, in press; Shirali-Shahreza & Shirali-Shahreza, 2006; Shirali-Shahreza & Shirali-Shahreza, 2007; Sun, Luo, & Huang, 2004; Topkara, Taskiran, & Delp, 2005; Yuling, Xingming, Can, & Hong, 2007; Zhang, Zeng, Pu, & Zhu, 2006).

Among the existing practical steganographic schemes, images and video streams are the most popular covertext media, since they contain more redundancy than other types of covertext. However, in some situations, image and video steganography might not be effective, particularly when bandwidth is limited. Text documents are smaller in size than other types of covertext. Moreover, in the real world, people normally do their digital communications through text media such as email messages. Sending and receiving text messages has become normal daily behavior of humans everywhere. Therefore, text is an effective covertext medium for steganography.

It is widely recognized today that steganographic schemes should be secure against both *human observers* and *computational attacks*. There have been some recent attempts to formalize steganographic security notions and to construct stegosystems that are provably secure against computational attacks under those notions (Backes & Cachin, 2005; Cachin, 2004; Dailey, Namprempre, & Samphaiboon, 2010; Hopper, Langford, & Ahn, 2002; Kiayias, Raekow, & Russell, 2005; Mittelholzer, 2000; Moulin & O'Sullivan, 1999; Zollner et al., 1998). In these models, *covertness*, or hiding the existence of the hiddentext message is defined in terms of *statistical* or *computational* indistinguishability against computational attacks. Among the existing models, Dailey et al. (2010) propose a provably secure generic construction that provides not only covertness but also privacy and integrity of the embedded secret message. In the model, covertness requires that polynomial-time computational adversaries with practical resources cannot tell whether the stegotext contains a real hiddentext message or an empty hiddentext. Adversaries are allowed to attack the stegosystem both by observing stegotexts corresponding to chosen covertexts and secret messages and by observing hiddentext messages decoded from chosen stegotexts. Under similar attacks, privacy requires that adversaries cannot tell whether a stegotext contains which of