Chapter 5.7
The Human Attack in Linguistic Steganography

C. Orhan Orgun
University of California, Davis, USA

Vineeta Chand
University of California, Davis, USA

ABSTRACT

This chapter develops a linguistically robust encryption system, Lunabel, which converts a message into syntactically and semantically innocuous text. Drawing upon linguistic criteria, Lunabel uses word replacement, with substitution classes based on traditional linguistic features (syntactic categories and subcategories), as well as features under-exploited in earlier works: semantic criteria, graphotactic structure, and inflectional class. The original message is further hidden through the use of cover texts—within these, Lunabel retains all function words and targets specific classes of content words for replacement, creating text which preserves the syntactic structure and semantic context of the original cover text. Lunabel takes advantage of cover text styles which are not expected to be necessarily comprehensible to the general public, making any semantic anomalies more opaque. This line of work has the promise of creating encrypted texts which are less detectable to human readers than earlier steganographic efforts.

1. INTRODUCTION

We develop in this chapter Lunabel, a technique for text-based steganography. We refer to our approach more specifically as linguistic steganography, as we take into account certain linguistic criteria that past approaches to text-based steganography have not dealt with (Bergmair, 2007, and references therein). This allows us to more effectively hide information. In particular, our encrypted messages more closely resemble natural text than was possible in past approaches which lack the linguistic sophistication necessary to achieve satisfactory results.

Section 1 introduces the concept of steganography and discusses desiderata for a successful

DOI: 10.4018/978-1-61350-323-2.ch5.7

technique. In section 2, we develop Lunabel and discuss some specific choices that were made in its implementation. Section 3 discusses the details of some of the particularly important choices that were made in developing Lunabel, namely the choice of cover text in which to hide information and the compilation of word substitution classes. In section 4, we compare Lunabel to past approaches to lexical steganography. Section 5 concludes the paper.

1.1 What is Steganography?

"Steganography" means encryption by means of information hiding. It includes hiding information in any form of data, such as images, audio or video files. Our interest in this paper is text-based steganography. This refers to hiding a message in what looks like an ordinary piece of text.

1.2 Linguistic Steganography

Ways of hiding information in text have been used since antiquity. One simple method is the acrostic, in which the initial letters of successive lines of poetry spell a word or words. This method is used more for artistic purposes than for secret information exchange; nonetheless, it provides a useful illustration. Consider the following Edgar Allan Poe poem, in which the first letters of successive lines spell the word *Elizabeth*:

1. Elizabeth it is in vain you say
   "Love not" — thou sayest it in so sweet a way:
   In vain those words from thee or L. E. L.
   Zantippe's talents had enforced so well:
   Ah! if that language from thy heart arise,
   Breathe it less gently forth — and veil thine
   eyes.
   Endymion, recollect, when Luna tried
   To cure his love—was cured of all beside—
   His folly—pride—and passion—for he died.

While this form of steganography may be sufficient for poetic use, a practical system has additional requirements. We would want information hiding to be more effective—the hidden information should not be readily visible to an outside observer. Equally important, the system needs to be algorithmic rather than creative; it should be possible to hide any given message in any desired text. Finally, decryption too needs to be algorithmic: given a text containing a hidden message, the hidden message should be reliably recoverable by a recipient in possession of the required decryption information (the acrostic poem presented satisfies this last requirement, but none of the others).

1.3 Encryption versus Steganography

Today, techniques for encryption have grown sufficiently sophisticated that it is possible to encrypt a message such that its decryption by brute force code cracking is a practical impossibility. However, there will be contexts in which even the fact that two parties are exchanging encrypted messages might be more information than one is willing to reveal to a third party. In such instances, what is needed is a way of hiding information in what appears at least to the casual observer to be an innocuous message. Typical encryption techniques give rise to results that are immediately noticeable as not being natural text. Consider, for example, the following encryption, which was created by a simple key-based technique:

2. XOQRW ISMWK QCBWT MBOFS
   PEMPW VUGQW QHJTO FIOCE KSLUO
   AWOTP RVLDV HLFGH WTDGC
   IAFCT HCLPT OJLQR MQMXD NPNBK
   FUAGV MDRWS ENSES EQJWF SRSVF
   VVRKC TGUDF BQTUS WXTLJ DGQWT
   PUTUG WFKTE GGOOX BQRTH
   FBMVQ LHKRU ULOMS SWUVS
   KOJFU RMRCD XDTMT PREFF DTKND