CHAPTER 1

INTRODUCTION

This chapter discusses the rationale background of research and the related situation, theories, and concepts used to conceptualize research. In addition, it discusses the variables related to the problem and their relationship to this research paradigm , problem identification, the proposed approach for solving the problem, the assumption that describes the limitation of the input, the scope of work, the delimitation of the output, and the contributions.

1.1   Rationale

Along with the increasing of the exchange of digital data in daily life communication, which are mostly in the forms of texts. Message confidentiality is necessary to guarantee security and to avoid message forgery. One method to guarantee message security (specifically for the authenticity of messages) is text Steganography. In this research, a cover text is used, and it is called Text Steganography. Text Steganography is divided into 2 categories: The Format Based and Linguistic. Some research with Format Based method has many weaknesses, because the suspicion of the stego text is easily detected by using a sensitive text reader such as (MS WORD, PDF READER). Linguistics still defects in terms of lexicality and semantic linkages between sentences that have not based on steganography sentences, which raise the suspicion if the sentences representing the stego text are not natural [6]. Noiseless Linguistic Steganography is a new method of making a stego cover using the Natural Language Generation System to produce a stego cover in the form of natural sentences that are meaningful to a particular domain. Thus, the sentence representing the stego cover is natural and does not raise suspicion from the attacker's point of view. Some of the studies conducted and related to Noiseless Linguistic Steganography are:

1. NORMALS (Normal linguistic steganography). NORMALS employs Natural Language Generation (NLG) techniques to generate noiseless (flawless) and a legitimate text cover by manipulating the non-random series input parameters of an NLG system to camouflage data in the generated text [1].

2. Mature Linguistic Steganography Methodology (Matlist) hides secret messages on documents that have certain domains based on Random Series data. The technique of encoding all letters into Random Series values (e.g., binary, decimal, hexadecimal, octal, alphabetic, alphanumeric, etc.) to generate noiseless textcover [2].

3. INORMALS Improving Using the Modified Baudot-Murray Code hides secret messages in the document as a response to a questionnaire (e-money). The INORMALS encoded the message into quinary codes. The codes are then mapped in an answer to a questionnaire question [3].

The research observes the weakness of the INORMALS which the capacity of embedding and the security level of cover are still low.

1.2   Theoretical Framework

An example of INORMALS implementation is that Bob and Alice are members of secret agents. They communicate on their mission through the public channel. They set the rules on how to communicate with each other securely. Bob is an administrator and he is responsible for managing the questionnaire given to customers of his company, and Alice is one of the custumers who fills out the questionnaire. Suppose Bob wants to send a secret message to Alice, he sends counseling-related documents based on a questionnaire to all intended clients. Hidden messages transmitted in this manner will not look suspicious because Alice is not the sole recipient. However, only Bob and Alice can reveal the hidden message because only these two parties know the rules used in the method. For generating cover text, INORMALS used NLG system by controlling the input set of questions [3]. In this research, the capacity of INORMALS can be increased using the reduction of sentences representing the shifting code. Furthermore, the security level for obtaining the number of codes can be improved. For hiding the number of codes digits, a private key and camouflaging the timestamp are distributed by the company.

1.3   Conceptual Framework

The output of INORMALS is the cover text that consists of sentences and timestamp. For those sentences, the naturalness quality depends on the grammatical aspect and semantic similarity. For timestamp, it is a representation of the number of code digits. The security to represent timestamp depends on the key to generate it. Furthermore, INORMALS must suffice the need for secret message length. The capacity of INORMALS can be determined by the encoding character, the number of questionnaires, and digits of codes-representation. The naturalness of cover's sentences, the security to represent timestamp, and the capacity of message embedding are dependent variables. The diagram of the conceptual framework is shown in Figure 1.1.
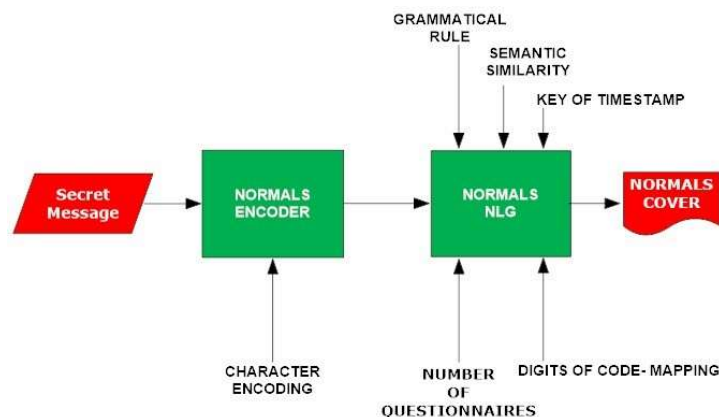

Figure 1.1: Conceptual Framework

1.4   Statement of the Problem

Andre's method is able to embed up to six characters in the provided cover. As there are more than 59% of Indonesian words are longer than six characters, it will be very difficult to conceal a secret message in Indonesian, even for only a one-word message. Hence, possibilities for an updated method that can accommodate a longer secret message should be investigated.

The number of code digits used in the Andre method is basically written in plain text, as part of the timestamp. To reduce the risks of breaking the secret message by the attackers and improve the system's robustness, the number should be concealed.

1.5   Objective and Hypotheses

For improving the embedding capacity, shifting code is represented by one sentence so that the number of sentences in the cover text could be decreased. Sentences representing shifting codes are created manually while maintaining the context of each sentence before and or after the embedding process.

Improving the security level is conducted by hiding the number of code digits of each code using a private key. The private key should be synchronized between the sender and the receiver.

Based on the proposed method that was conducted, increasing the capacity of embedding and improving security level can be conducted.

1.6   Assumption

This research assumes that a secret message is written in the Indonesian and the questionnaire has five possible answers as the input for creating a cover using natural language generation. The cover text consists of sentences in the Indonesian. A set of rules for generating sentences should be agreed upon by both parties.

1.7   Scope and Delimitation

The improved INORMALS method utilizes a more efficient embedding process to reduce the number of sentences, improve the security of hiding the number of code digits using the private key, and generate more variants of sentences in the cover text. The improved INORMALS used Indonesian is to test the naturalness of sentences of the experiment resulted although it is not limited to. The improved INORMALS also uses a new corpus of about 1 thousand words to support the synonyms filtering process.

1.8   Importance of the Study

This research expands the methods in noiseless steganography to conceal secret messages. This research is an example of the implementation strengthening of the INORMALS method that can be used in various areas such as military, diplomatic, personal, and intellectual property applications.