Chapter 1

Introduction

Steganography, the art of information hiding, has been around for thousands of years. The first recorded use was in 450 B.C , where Demaratus sent messages hidden beneath the wax on writing tablets.

Steganography is concerned with hiding the existence of data by encapsulating it within some covertext. The goal is to make the hidden data undetectable, by man or machine. If the existence of hidden data can be proven, then the steganography can be considered to have failed, even if the hidden data itself is not recovered. This is in contrast to cryptography, where the goal is to prevent the data from being revealed, with no attempt to hide the existence of the data (in fact cryptographic data is often easily identified). Cryptography can be used in conjunction with steganography to provide an added layer of security in case the steganography is broken.

Modern steganography can be applied to a number of mediums. Text, images, audio and even video are all commonly used as carriers for secret messages. Text steganography, however, has received less attention in recent years, mainly due to lower capacity to hide information that is associated with it. There are many reasons why text steganography deserves continued research and development. For one, text is still the primary form of communication in many areas of the world where computers and the internet are not widespread. Text is also universally applicable, nobody has images or audio who does not have text.

Perhaps the unique feature of text steganography over steganography which makes use of images or audio is that image and audio steganography uses redundancy in the data. This redundancy can be easily removed during compression, which would destroy any hidden data. On the other hand, while text steganography uses the English language, which still contains redundancy, such as synonyms, this redundancy cannot be removed since the rules of the language are fixed.

In this project I propose an lightweight, robust algorithm for text steganography which uses the idea of synonym substitution i.e. hiding bits by replacing words with words of the same (or similar) meaning. This algorithm will be tested using a simple chat-based application,