Majid Khan, Ali Shahab*, and Zeeshan Asghar

Introduction to Linguistic Steganography

DOI 10.1515/nleng-2015-0013
Received May 10, 2015; revised June 5, 2015; accepted July 18, 2015.

Abstract: The specialty of data covering up has gotten much consideration in the late years as security of data has turn into a major concern in this web time. As sharing of delicate data by means of a typical correspondence station has get to be unavoidable, Steganography – the workmanship and art of concealing data has increased much consideration. We are likewise encompassed by a universe of mystery correspondence, where individuals of numerous types are transmitting data as guiltless as an encoded Visa number to an online store than and as deceptive as a terrorist plot to robbers. Steganography is derived from two Greek words, steganos, meaning covered or secret, and graphia, meaning writing. In simple terms, steganography is the art and science of hiding information in plain sight. Steganography is an innovation where advanced information pressure, data hypothesis, spread range, and cryptography innovations are united to fulfill the requirement for security on the Internet. This paper is an endeavor to examine the different methods utilized as a part of steganography and to recognize zones in which this method can be connected, so that humanity can be profited massively.

Keywords: Security; Computer network; Steganography; Information Hiding

1 Introduction

The risk to security prompted at the early phases of PCs. The PCs were initially associated in around 1969, known as Arpanet. This Arpanet was framed by the division of Defense, United States. The electronic documents were sent along this system. The records exchanged were of significance and consequently required safe exchange. In this way, the development of PC system asked for PC system security. The entire world uses organize as a prime wellspring of correspondence. The utilization of system expanded decade after decade. The utilization of system propelled numerous designers to chip away at system and its building design to make it more security. The intricate system structural engineering got to be inclined to ruptures. Another part of security is in the setting of protected innovation. The unlawful utilization of information in electronic document is enormous issue nowadays and subsequently security means, for example, advanced authentications, computerized marks and so forth are utilized. The paper association contains history of security in system, Steganography.

The PC system security fundamentally contains security gave to both the PC and the system in the middle of PCs. The information ought to be secure while put away in PCs furthermore while getting transmitted by means of systems. Different routines have been found to give wellbeing to the information.

The historical backdrop of system security demonstrates a critical hole between the innovations connected with security and system. The announcement can be demonstrated with the apparent vicinity of International Standardization of Organization's Open System Interconnection model. This model was created with a considered making perfect system model. The model has all around characterized structural planning. It has seven layers with very much characterized capacities. Every layer is free of the other. This in a roundabout way implies that if the usefulness of a layer is enhanced or changed by in future then alternate layers won't be influenced. It additionally had great and all around created type of conventions needed for systems administration. Be that as it may, despite the fact that the model was consummate as far as systems administration, it didn't address any issues of security. Because of the above reason the announcement of hole between innovations gets demonstrated. To extension this crevice between advancements i.e., to give system security, simply giving security to the PC won't work. Case in point give us a chance to accept in a system we simply give security to the PCs , the outcome would be the information would be safe inside of PCs however when the information enters the system connect the wellbeing is altered, it is conceivable that the information would be adjusted before it

***Corresponding Author: Ali Shahab:** Department of Sciences and Humanities, National University of Computer and Emerging Sciences, Islamabad, Pakistan, E-mail: ali.shahab@nu.edu.pk
Majid Khan, Zeeshan Asghar: Department of Sciences and Humanities, National University of Computer and Emerging Sciences, Islamabad, Pakistan

achieves its real destination. The above situation demonstrates the need of security on connections too.

2 Steganography

Since the ascent of the Internet a standout amongst the most critical variables of data innovation and correspondence has been the security of data. Cryptography was made as a strategy for securing the mystery of correspondence and a wide range of systems have been created to encode and unscramble information to keep the message mystery. Sadly it is at times insufficient to keep the substance of a message mystery, it might likewise be important to keep the presence of the message mystery. The system used to execute this, is called steganography.

Steganography is the expertise and learning of shrouded imperceptible correspondence. This is expert through concealing data in other data, therefore concealing the presence of the imparted data. The word steganography is gotten from the Greek words "stegos" signifying "cover " and "grafia" signifying " writing " characterizing it as " covered writing " [1, 7]. The idea and routine of concealing data has a long history. In histories the Greek honorable men if needed to correspond with his child in law some place far they used to shave the head of one of their trusted man. Once shaved the message was tattod on the head's skin and afterward the delivery person would let his hair become once more. Once the hair developed to cover the message totally the emissary would withdraw to the child in law. The child in law would thusly shave the head of the errand person again to recover the message once more.

3 Steganography and its importance

Steganography is an investigation of secured composition. In starting days steganography was done on papers that were utilized to pass on messages from sender to recipient. One such type of steganography was utilization of acidic undetectable ink. This ink would be undetectable on page. The message composed by this ink could just be read if the paper on which it is composed is kept in a specific edge in light. The discharge message was first composed with the imperceptible ink and after that to make the letter additionally persuading some message in the configuration of basic letter was composed in agreement. So this was steganography in its starting stage [3]. Steganography viably shrouds the message however does not conceal the phenomena that two gatherings are imparting. The steganography is a system to conceal the discharge information or data in ordinary information or data. The typical message used to shroud the discharge information is known as transporter. The mystery message is inserted in the transporter to frame the steganography medium. The Steganographic key is utilized to shroud or encode the message. This key will know just to the sender and beneficiary.

At the point when a picture is utilized to conceal an information then after the mystery information is full in the picture the subsequent picture is known as stego-picture. Also when an information is covered up in the feature, the feature containing the concealed information is known as stego-feature. The procedure may be abridged as takes after:

Steganographic medium = hidden information + cover-medium + steganography key.

In an advanced world, Steganography and Cryptography are both longed for shielding data from superfluous gatherings. Both Steganography and Cryptography respect ensure information yet neither innovation alone is impeccable and both can be conked out. It is thus that most specialists would recommend utilizing both as a part of order to include more layers of security. Steganographic advancements are an exceptionally imperative piece of the plausible Internet security and classified on open frameworks, for example, the Internet. Steganographic looks into are fundamentally intending to compliment the cryptography and extension the ruptures or holes or security gaps in cryptography. The steganography centers to give complete security. The analysts have finished up with their studies that the greater part of the legislatures has put impediments on the routines and many-sided quality of the encryption. With these impediments the routines grew by the designers are moderately powerless. These powerless rationales frame the security openings which if assaulted can uncover the shrouded data effectively. Consequently the steganography is an answer for concealment the security openings and subsequently steganography is always under advancements with cryptography.

Steganography is utilized to conceal information inside another information. The sender and recipient both have the learning that the mystery information is covered up in the message. To include different layers of security and to help sub side the "crypto versus law" issues already specified, it is a decent answer for utilization steganogra-

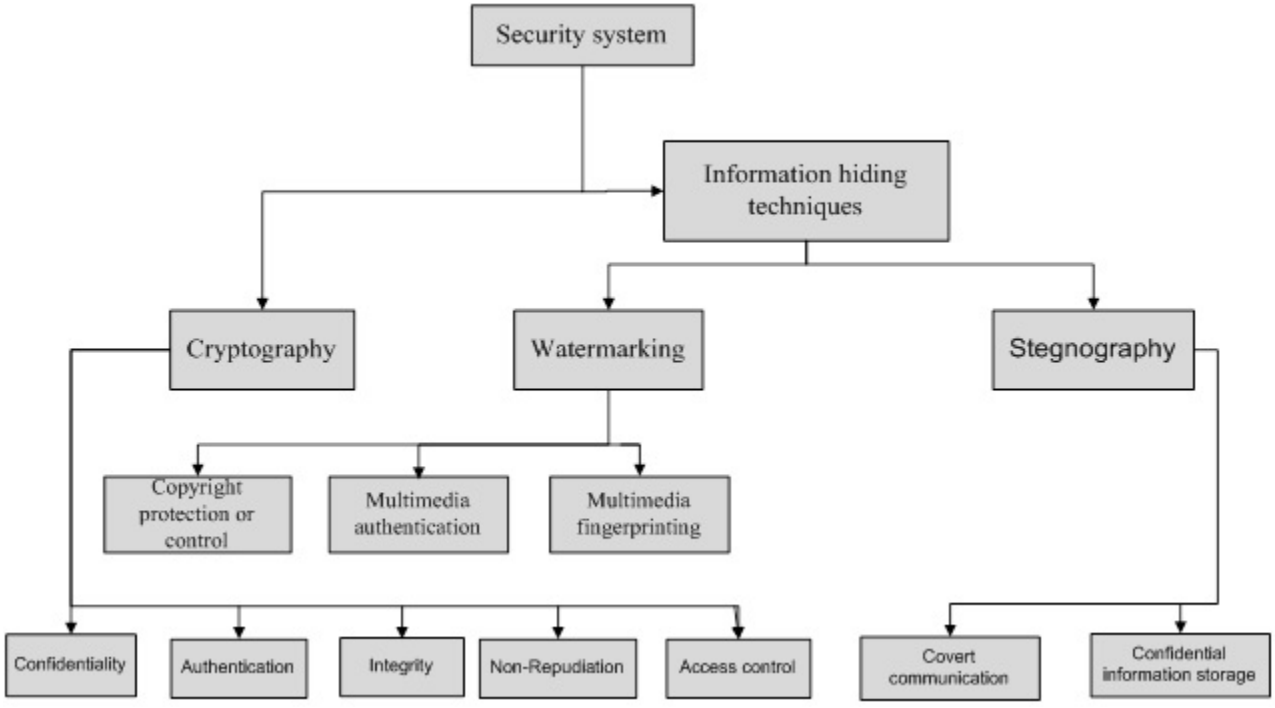

Fig. 1. Classification of information security techniques and its applications.

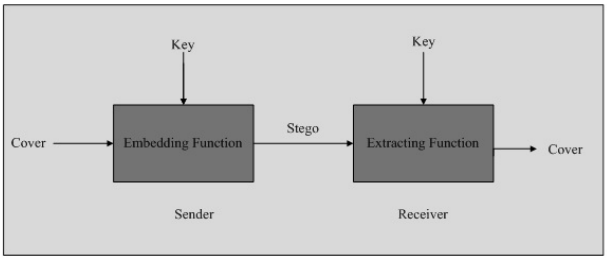

Fig. 2. Basic working algorithm of steganography.

phy in compliment with the cryptographic systems. Both the techniques i.e cryptographic strategies and steganographic routines are having pit-falls yet both when utilized together can discover answers for one another's lacunas. Hence in short we may say that it is helpful to utilize a blend of the two systems [4].

4 Classification of steganography techniques

Steganography is comprehensively classified into two broad categories technical and linguistic. Technical steganography utilizes investigative procedures to conceal a message, for example, the utilization of undetectable ink or microdots and other size-decrease systems. Linguistic steganography shrouds the message in the transporter in some nonobvious traditions and is further arranged as semagrams, spread figures and open figures. Theses classifications are presented in Figure 3. But here we will focused on linguistic steganography instead of technical steganography.

4.1 Linguistic Steganography

Characteristic dialect construct data concealing innovation depends in light of adjusting data in content archives by controlling their lexical, syntactic, and semantic properties while saving the significance however much as could be expected. These systems are more robust than strategies that simply change the presence of content components like text styles, line space, entomb word separation and so forth in accomplishing writings incognito. There are a few cases of linguistic steganography, which can be classified as takes after.

4.1.1 Semagrams

Semagrams are utilized to conceal data through the utilization of signs and images. A visual semagram could identify with an organized code that is transmitted by wav-

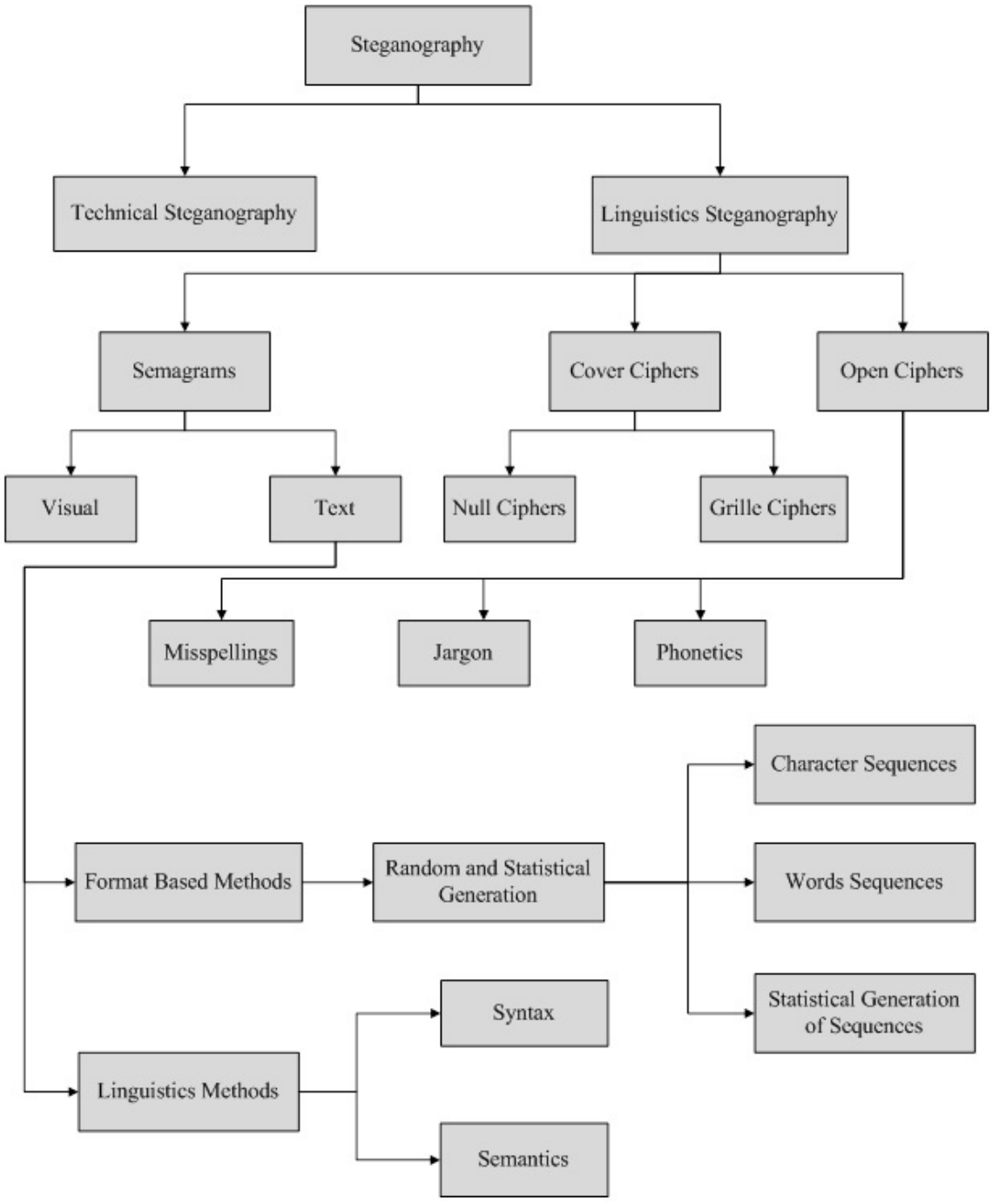

Fig. 3. Classification of Steganography.

ing your hand, setting a thing in a particular area around your work area or changing the look of your site. These signs are hard to recognize and have the benefit of ordinariness in a regular world. Here and there the viable utilization of visual semagrams may be your just strategy for correspondence with your companions and associates, and it is essential to set up and prearrange a few messages that may be transferred in times of threat.

Content semagrams are typical messages encoded through the medium of content. Uppercase letters, highlight, and curious penmanship, clear spaces in the middle of words can all be utilized as signs for a predefined reason. Subliminal messages likewise fall into this class. They are now and then helpful when you wish to convey a little bit of data. Case in point, you could concur with your contacts to trade apparently harmless day by day climate

reports via email. The expression 'the sky is dim' may serve as a ready significance you are in a bad position and they ought to assemble worldwide help.

4.1.2 Open Codes

Open code steganography hides the message in a legitimate piece of text in the ways not immediately obvious to the observer. Computers and humans have different abilities when it comes to steganalysis, or detecting steganographic messages (see below under 'Detection' sub-heading). The following examples may not be applicable to the surveillance carried out by a human steganalyst. They use linguistic variations of the text to fool the common formulas used by electronic filters and surveillance systems. Please bear in mind that these can only be regarded as hints or suggestions to take advantage of the non-intelligent nature of computer systems. They should not be used to communicate important information, but only to test the effectiveness of the filtering system. If you know that certain words in your email will result in its failure to reach the recipient and this information alone will not get you into trouble, you can try out some of the variations below.

4.1.2.1 Jargon

Using jargon in your messages could render its content meaningless to an outside observer. Prearranged meanings or underground terminology can hide the real contents of the message. It is advisable to choose words in such a way that the carrier message remains legible and comprehensible, if not true. The possibilities of the use of jargon are limited only by the stock of the words known to the communicating parties.

4.1.3 Covered Ciphers

Covered ciphers employ a particular method or secret to hide text in an open carrier message. Sometimes these include simple techniques of embedding a message into the words of the carrier. Consider the example below, sent by the German Embassy in Washington DC to the headquarters in Berlin during World War One.

4.1.3.1 Null Ciphers

Null ciphers apply a series of characters and words intended to confuse a cryptanalyst. The communicated appears as obvious nonsense, but can be decoded to a meaningful message. This is an ancient form of encrypted communication in which a message is surrounded by a large amount of redundant characters (known as null ciphers). This form of communication is, in fact, known to have been used by the German army during WW II. The following is an example of a null cipher form of steganography:

"Morning Ali Jamshed, I Drove Karachi Highway At Noon in Speed. Meeting Your Friend Regarding Islamiat Exam Next Day"

"Language of Nurses Did Offend Nursing Ideals So Nurses Opt Their Smiling Attitude For Everyone"

Decoding these messages by extracting the first letter in each word reveals:

"Majid Khan is My Friend"

"London is Not Safe"

The drawback of this form of steganography is that the message sender is forced to make a text cover according to a preset procedure, hence defeats the purpose of steganography. Also, applying a 'brute force' approach to decoding will reveal the message.

There are various types of specialized steganography like content, picture, sound, feature, convention and so on. All advanced document organizations can be utilized for steganography; however designs with high level of repetition are suitable. Excess here can be characterized as the data which is more than once show in the information, and the information could have been effortlessly comprehended without these rehashed substances. The repetitive bits are the bits which when changed are not recognized effortlessly [5]. Picture and sound documents are broadly utilized for steganography yet this doesn't imply that other document groups can't be utilized. Generally, concealing information into content was the critical method for steganography i.e. concealing mystery note in every nth letter of each expression of an instant message. However, since the start of the advanced period, Text, Images, Audio/feature, Protocol, Internet and all the distinctive computerized document organizations has reduced its significance. Content steganography utilizing advanced records is not utilized regularly on the grounds that the content does not have much repetitive information. The picture comprises of numerous pixels which are spoken to as bits. The greater part of these bits is found as needlessly. Also,

this element makes the pictures an essential configuration for steganography.

Comparative system is utilized to stow away discharge information in sound records as that of utilized for picture documents. Concealing is one of the one of a kind system which conceals the information in sound document in such which is imperceptible to human ears. A light fine sound gets to be unperceivable before louder sound. This element makes a space where we can shroud the data. Despite the fact that the aforementioned procedure is on a par with picture steganography yet the increment in size of stego-sound give an insight to gatecrasher about the concealed information [6].

5 Conclusion

The field of Linguistic Steganography is very interesting as it conceals the very existence of secret message from intruder, which is not achievable by cryptography. In this article, we mainly study some basic schemes of Linguistic Steganography and review some standards techniques of Linguistic Steganography. In future, we will definitely classify some other techniques of Linguistic Steganography.

References

[1] K. Bennett, Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text, Tech. Rep. TR 2004-13, Purdue CERIAS, May 2004.

[2] M. T. Chapman, Hiding the hidden: A software system for concealing ciphertext as innocuous text, Master‘s thesis, University of Wisconsin-Milwaukee, May 1997.

[3] M. T. Chapman and G. I. Davida, Hiding the hidden: A software system for concealing ciphertext as innocuous text, in Information and Communications Security: First International Conference, O. S. Q. Yongfei Han Tatsuaki, ed., Lecture Notes in Computer Science 1334, Springer, August 1997.

[4] M. T. Chapman, G. I. Davida, and M. Rennhard, A practical and effective approach to large-scale automated linguistic steganography,? in Information Security: Fourth International Conference, G. I. Davida and Y. Frankel, eds., Lecture Notes in Computer Science 2200, p. 156ff, Springer, October 2001.

[5] R. Bergmair, Towards linguistic steganography: A systematic investigation of approaches, systems, and issues. final year thesis, April 2004. handed in in partial fulfillment of the degree requirements for the degree B.Sc. (Hons.) in Computer Studies to the University of Derby.

[6] I. A. Bolshakov, A method of linguistic steganography based on collocationally verified synonymy., in Information Hiding: 6th International Workshop, J. J. Fridrich, ed., Lecture Notes in Computer Science 3200, pp. 180–191, Springer, May 2004.

[7] K. Winstein, Lexical steganography through adaptive modulation of the word choice hash, January 1999. Was disseminated during secondary education at the Illinois Mathematics and Science Academy. The paper won the third prize in the 2000 Intel Science Talent Search.