Int. J. Inf. Secur. (2009) 8:247–261
DOI 10.1007/s10207-009-0079-0

REGULAR CONTRIBUTION

Listega: list-based steganography methodology

Abdelrahman Desoky

Published online: 11 April 2009

Abstract The use of textual list of items, e.g., products, subjects, books, etc., is widely popular and linguistically legible. This motivates the development of List-Based Steganography Methodology (Listega). Listega takes advantage of such use of textual list to camouflage data by exploiting itemized data to conceal messages. Simply, it encodes a message then assigns it to legitimate items in order to generate a text-cover in a form of list. The generated list of items, the text-cover, can be embedded among other legitimate noncoded items for more protection based on a predetermined protocol among communicating parties such as read every other item, every fifth item, or any other way than the use of particular sequence. Listega neither hides data in a noise (errors) nor produces noise. Instead, it camouflages data by manipulating noiseless list of legitimate items. Listega establishes a covert channel among communicating parties by employing justifiably reasons based on the common practice of using textual list of items in order to achieve unsuspicious transmission of generated covers. The presented implementation, validation, and steganalysis of Listega demonstrate: the robustness capabilities of achieving the steganographic goal, the adequate room for concealing data, and the superior bitrate of roughly 1.32 up to 3.87% than contemporary linguistic steganography approaches.

Keywords Linguistic steganography · Steganography

A. Desoky (✉)
Department of Computer Science and Electrical Engineering,
University of Maryland, Baltimore County,
Baltimore, MD, USA
e-mail: abd1@umbc.edu; iloveitech@yahoo.com

1 Introduction

Steganography is the science and art of camouflaging the presence of covert communications. The origin of steganography is traced back to early civilizations [1,2]. The ancient Egyptians communicated covertly using the Hieroglyphic language, a series of symbols representing a message. The message looks as if it is a drawing of a picture although it may contain a hidden message that only a specific person who knew what to look for can detect. The Greeks also used steganography, "hidden writing," where the name was derived. Fundamentally, the steganographic goal is not to hinder the adversary from decoding a hidden message, but to prevent an adversary from suspecting the existence of covert communications [3]. When using any steganographic technique if suspicion is raised, the goal of steganography is defeated regardless of whether or not a plaintext is revealed [4,5]. Contemporary approaches are often classified based on the steganographic cover type into image, audio, graph [6,7], or text. When linguistics is employed for hiding data and generating the steganographic cover, an approach is usually categorized as linguistic steganography to distinguish it from nonlinguistic techniques, e.g., image, audio, etc. Linguistic steganography has become more favorable in recent years since the size of nonlinguistic-covers is relatively large and is burdening the traffic of covert communications [5,8,9].

Most of the published steganography approaches hiding data as noise in a cover that is assumed to look innocent. For example, the encoded message can be embedded as an alteration of a digital image or an audio file without noticeable degradation [5,8]. Another example is hiding a message in a text-cover by modifying the format and style of an existing text [3,9,10]. However, such alteration of authenticated covers can raise suspicion and the message is detectable regardless of whether or not a plaintext is reveled [8,9].

The same applies to hiding the data in unused or reserved space for systems software, e.g., the designated storage area of an operating system, the file headers on a hardrive, etc. [11,12], or in the packet headers of communication protocols, e.g., TCP/IP packets transmitted across the Internet [13]. These techniques are vulnerable to distortion attacks [4,8].

On the other hand, a similar argument is made in the literature about linguistic steganography approaches such as null cipher [14], mimic functions [15,16], NICETEXT and SCRAMBLE [17–20], translation-based [21–23], confusing approach [24], and abbreviation-based [25]. The vulnerability and concerns of these linguistic approaches, as explained in Sect. 2, can be summarized as follows. First, the linguistic-cover either introduce detectable flaws (noise), such as incorrect syntax, lexicon, rhetoric, grammar, etc., when generating a text-cover. Obviously, such flaws can raise suspicion about the presence of covert communications. Second, the content of the cover may be meaningless and semantically incoherent, and thus may draw suspicion. Third, the bitrate is very small. Since there is a limit on how many flaws a document may typically have, very large documents will be needed to hide few bytes of data. In fact this applies to nonlinguistics approaches as well. Fourth, the bulk of the efforts have been focused on how to conceal a message and not on how to conceal the transmittal of the hidden message. In other words, the establishment of a covert communication channel has not been an integral part of most approaches found in the literature. Fifth, while these approaches may fool a computer examination, they often fail to pass human inspections. A successful linguistic steganography approach must be capable of passing both computer and human examinations. These concerns have motivated the development of the List-Based Steganography Methodology (Listega), introduced in this paper.

Listega overcomes the issues just mentioned above by manipulating the popular textual list of itemized data to camouflage both a message and its transmittal. Basically, Listega exploits textual itemized data such as books, computer parts, music CD's, movie DVD's, etc., to conceal messages. Such list of items can be fabricated in order to embed data without generating any type of suspicious pattern. Simply, it encodes a message then assigns it to legitimate items in order to generate a text-cover in a form of list.

The main advantages of Listega are as follows. First, the high demand for using textual list of itemized data by a wide variety of people creates a high volume of traffic and averts suspicion in the presence of covert communication channels. Second, Listega does not imply a particular pattern (noise) that an adversary may look for. Third, the concealment process of Listega has no effect on the linguistics of the generated cover (list-cover). Therefore, a list-cover is linguistically legitimate and is thus capable of passing both computer and human examinations. Fourth, Listega can be applied to all languages. Fifth, textual lists have plenty of room for concealing data, as demonstrated later in the paper. The observed average bitrate of the current implementation experiments is superior to all contemporary linguistic steganography approaches found in the literature which roughly 1.32 up to 3.87%. Sixth, Listega is resilient to popular attacks and the hidden message is anti-distortion. Since the reuse and alteration of textual lists are a common practice can pass comparison attacks. The implementation and steganalysis validation demonstrate that Listega methodology is capable of achieving the steganographical goal.

The remainder of this paper is organized as follows. Section 2 discusses the related work and compares Listega to the linguistic steganography techniques found in the literature. Section 3 explains the Listega methodology in detail. Section 4 demonstrates the Listega implementation. Section 5 presents the steganalysis validation of Listega. Finally, Sect. 6 concludes the paper.

2 Related work

The aim of this section is to present the contemporary linguistic and nonlinguistic steganography approaches versus Listega methodology.

2.1 Linguistic steganography

Linguistic steganography approaches conceal data in a linguistic-based textual cover. Linguistic steganography approaches can be categorized as follows.

- **Series of characters and words:** During World War I, the Germans communicated covertly using a series of characters and words known as null-cipher [14]. A null-cipher is a predetermined protocol of character and word sequence that is read according to a set of rules such as: read every seventh word or read every ninth character in a message. Apparently, suspicion is raised because the user is forced to fabricate a text-cover according to a predetermined protocol, which may introduce some peculiarity in the text that draws suspicion and defeats the steganographical goal. In addition, applying a brute force attack may reveal the entire message.
- **Statistical based:** Wayner has introduced the mimic functions approach [15,16] which employs the inverse of the Huffman Code by inputting a data stream of randomly distributed bits to produce text that obeys the statistical profile of a particular normal text. Therefore, the generated text by mimic functions is resilient against statistical attacks. Mimic functions can employ the concept of both context free grammars (CFG) and van Wijnaarden grammars to enhance the output. The output of regular mimic

functions is gibberish rendering it extremely suspicious [8,9]. However, the combination of mimic functions and CFG slightly improved the readability of the text [15,16]. Yet, the text-cover still contains numerous flaws such as incorrect syntax, lexicon, rhetoric, and grammar. In addition, the content of the text-cover is often meaningless and semantically incoherent. These shortcomings may raise suspicion in covert communications.

- **Synonym based:** Chapman and Davida have introduced a steganographic scheme consisting of two functions called NICETEXT and SCRAMBLE that uses a large dictionary [17–20]. NICETEXT uses a piece of text to manipulate the process of embedding a message in a form of synonym substitutions. This process preserves the meaning of text-cover (the original piece of text) every time it is used. The synonyms-based approach attracted the attention of numerous researchers in the last decade: Winstein [26,27], Bolshakov et al. [28,29], Calvo et al. [30], Chand et al. [31], Nakagawa [32], Niimi et al. [33], Bergmair et al. [34–36], Topkara et al. [37], Murphy et al. [38], and Atallah et al. [39,40]. Although the text-cover of synonym-based approach may look legitimate from a linguistics point of view given the adequate accuracy of the chosen synonyms, reusing the same piece of text to hide a message is a steganographical concern. If an adversary intercepts the communications and oversees the same piece of text that has the same meaning over and over again with just different group of synonyms between communicating parties, he will question such use.
- **Noise based:** Grothoff et al. have introduced the translation-based steganographic scheme [21–23] to hide a message in the errors (noise) that are naturally encountered in a machine translation (MT). This approach embeds a message by performing a substitution procedure on the translated text using translation variations of multiple MT systems. In addition, it inserts popular errors of MT systems and also uses synonym substitutions in order to increase the bitrate. Unlike synonyms-based steganography, linguistic flaws in noise-based approach are not a concern unless they appear excessively. However, Grothoff et al. states that one of the concerns is that the continual improvement of MT may narrow the margin of hiding data. In addition, translation-based approach, as pointed out by Grothoff et al., cannot be applied to all languages because of the fundamental structures are radically different. This generates severely incoherent and unreadable text [21–23]. On the contrary, Listega can be applied to all known languages without any exceptions while the generated list-cover is linguistically legitimate. Another noise-based approach has been proposed by Topkara et al. that employs typos and ungrammatical abbreviations in a text, e.g., emails, blogs, forums, etc., for hiding data [24]. Moreover, Shirali-Shahreza et al. have introduced an abbreviation-based scheme [25] to conceal data using the short message service (SMS) of mobile phones. Due to size constraints of SMS and the use of phone keypad instead of the keyboard, a new language called SMS-Texting was defined to make the approach more practical. However, these approaches are sensitive to the amount of noise (errors) that occurs in a human writing. Such shortcoming not only increases the vulnerability of the approach but also narrows the margin of hiding data. Conversely, Listega neither employs errors nor uses noisy text to conceal data.
- **Nostega based:** Recently, the new paradigm in steganography research, namely Noiseless Steganography Paradigm (Nostega) has been introduced [41], in which the message is hidden in the cover as data rather than noise. A number of methodologies have been developed based on the Nostega paradigm. One of these methodologies is the Summarization-Based Steganography Methodology (Sumstega) [42]. Sumstega exploits automatic summarization techniques to camouflage data in the auto-generated summary-cover (text-cover) that looks an ordinary and legitimate summary.

It is worth noting that the presented Listega methodology in this paper follows this new paradigm by exploiting the popular textual list of itemized data to camouflage data without generating any suspicious pattern.

2.2 Nonlinguistic steganography

Nonlinguistic steganography approaches can be categorized based on its file type such as text, image, audio, and graph. Textual steganography, which is based on nonlinguistic techniques, hides data by textual format manipulation (TFM) [8] process. TFM modifies an original text by employing spaces, misspellings, fonts, font size, font style, colors, and noncolor (as invisible ink) to embed an encoded message. However, comparing the original text versus the modified text triggers suspicion and enables an adversary to detect where a message is hidden. In addition, TFM can be distorted and may be discerned by human eyes or detected by a computer [8,9].

On the other hand, image steganography is based on manipulating digital images to conceal a message. Such manipulation often renders the message as noise. In general, image steganography suffers from several issues such as the potential of distortion, the significant size limitation of the messages that can be embedded, and the increased vulnerability to detection through digital image processing techniques [5]. Audio-covers have also been pursued. Example of audio steganography techniques include LSB [43,44], spread spectrum coding [45,46], phase coding [45,47], and echo hiding [47,48]. In general, these techniques are too complex, and like their image-based counterpart, are still subject to distortion and are vulnerable to detection [3,5,8,44]. The hidden message may become to a great extent a foreign body in the cover and thus makes those schemes vulnerable to detection. In addition, contemporary steganography schemes rely on private or restricted access to the original unaltered cover in order to avoid the potential of comparison attacks, which is considered a major threat to the covert communication. Basically, an adversary can detect the presence of a hidden message by comparing a particular image-cover or audio-cover to the original image or audio file and finding out that some alterations have been made.

Hiding information in an unused or reserved space in computer systems [11,12]. For example, Windows 95 operating system has around 31 KB unused hidden space which can be used to hide data. Another example, unused space in file headers of image, audio, etc., can also be used to hide data. This depends on the size of the hardrive used. TCP/IP packets used to transport information across the Internet have unused space in the packet headers [13]. The TCP packet header has six unused (reserved) bits and the IP packet header has two reserved bits. There are tremendous packets are transmitted over the Internet can convey and transmit a secret data. However, these techniques are vulnerable to distortion attacks [3,4,8].

Recently, a Graph Steganography (Graphstega) methodology has been developed [6,7]. Unlike all other schemes, the message is naturally embedded in the cover by simply generating the cover based on the message. Graphstega camouflages a message as data points in a graph, e.g., numerical values that can be plotted in a chart, and thus the message would not be detectable as noise. The approach is shown to be resilient to a wide range of attacks, including a comparison attack by untraceable or authenticated data. Similarly, Chestega [49] exploits popular games, like chess, checkers, crosswords, domino, etc., for concealing messages in an unaltered authenticated data. Graphstega and Chestega represent a new paradigm in steganography research in which the message is hidden in the cover as data rather than noise. Listega follows this new paradigm, namely Nostega [41], by exploiting textual lists of itemized data to camouflage data without generating any suspicious pattern.

3 Listega methodology

To illustrate Listega, consider the following scenario. Bob and Alice are on a spy mission. Bob and Alice run online-business such as ebay.com, craigslist.org, yahoo.com, etc., to buy and sell items such as books, computer parts, music CD's, movie DVD's, etc. Before they went on their mission, which requires them to reside in two different countries, they plot a strategic plan and set the rules for communicating covertly using their online-business as a steganographic umbrella. They basically agree on concealing messages in list of items by naturally manipulating fabricating a list of items to embed data in such a way a text-cover (list-cover) looks unsuspicious. To make this work, Bob and Alice have the right to post or email the list-cover (e.g., list of books, computer parts, music CD's, movie DVD's, etc.) for customers. In addition, both of Bob and Alice have the right to make business with each other such as buying and selling which legitimizes the discernable communications between them for delivering the list-cover in unsuspicious way. Covert messages transmitted in this manner will not look suspicious because the relationship between Bob and Alice is legitimate. Furthermore, Alice is not the sole recipient of Bob's list and vise versa. However, other nonspy customers also receive such list further warding off suspicion. These lists conceal data. However, only Bob and Alice will be able unravel the hidden message because they know the rules of the game. The communications of both Bob and Alice looks legitimate and nothing is suspicious because of the employing of such legitimate relationship between the communicating parties. Alice and Bob are using real data from their business field to make their covert communications legitimate.

The above scenario demonstrates how Listega methodology can be used. Listega methodology is detailed in the remainder of this section.

3.1 Listega architecture

Listega achieves legitimacy by basing the camouflage of both a message and its transmittal on a legitimate list of items. As stated earlier, in the above example of Bob and Alice, using a particular online-business gives legitimacy for camouflaging both a message and its transmittal. The core idea of Listega methodology is basically camouflaging data in the natural and legitimate itemized data. Obviously, such steganographic cover (text-cover) in a form of list of items linguistically and logically is legitimate. The following is an overview of the Listega architecture, which consisted of four modules as shown in Fig. 1.

1 **Domain determination** (Module 1): Determines an appropriate domain(s) for achieving the steganographical goal. One of the major factors for employing a particular domain is the use of list. A domain example, an online business that naturally uses itemized data such as list of items (e.g., books, computer parts, music CD's, movie DVD's, or any other type of items) can be employed by Listega methodology. The process of Module 1 is only involved in the stage of constructing Listega system.
2 **Message encoding** (Module 2): Encodes a message in an appropriate and required form for the camouflaging process (Module 3). The process of generating a list-cover (Module 3) may influence the process of how a message

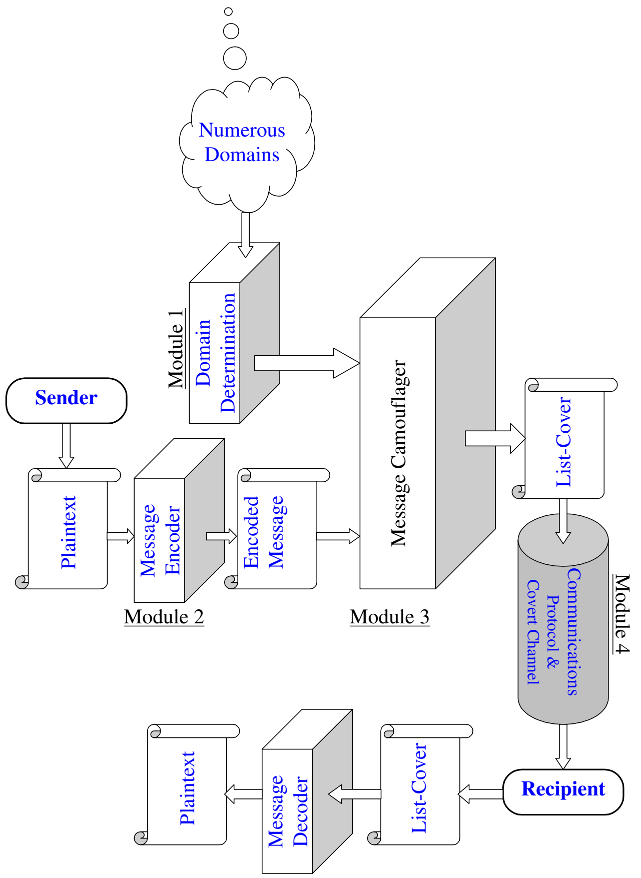


Fig. 1 Illustrates the architecture of Listega and the communications protocol

should be encoded. For example, a message may encode by slicing its binary string into a particular length of bits such as four bits, seven bits, or any required bit length as follows.

Message: "*Stop*"
Convert it to binary:
01010011011101000110111101110000
Then slicing its binary string into a particular length of bits such as four bits:
0101 0011 0111 0100 0110 1111 0111 0000

3 **Message camouflager** (Module 3): Generates the text-cover (list-cover), in which data are embedded by employing the output of Module 2. Simply, the text-cover is a legitimate list of items.
4 **Communications protocol** (Module 4): Configures the basic protocol of how a sender and recipient would communicate covertly. Obviously, it includes the covert channel for delivering a list-cover to the intended recipient and the decoder scheme to unravel a hidden message.

Once the Listega system is implemented, the covert communications will be accomplished in three steps. First, it encodes a message using the predetermined steganographic encoder (Module 2). Second, Module 3 camouflages the steganographic code (encoded message) which is generated by Module 2. Third, it sends a message based on the communications protocol (Module 4). The above modules are detailed in the following sections.

3.2 Domain determination (Module 1)

The chosen domain(s) must be capable of concealing data. In other words, it must allow the process of embedding data without generating noise in order to achieve the steganographical goal. Since Listega mainly manipulates list of items to camouflage messages. Therefore, any domain allows the use of list of items such as books, computer parts, music CD's, movie DVD's, or any other type of items, Listega methodology can be applied. In addition, the chosen domain also has to fit the communicating parties and provide some ground for justifying the communications. For example, an individual, such as student, would not post army aircraft stuff for free or sale on craigslist.org website. Such communications can easily raise suspicion because an individual, such as student, may post his personal stuff for free or sale on an online website such as craigslist.org not army aircraft stuff. Listega naturally camouflages the delivery of a hidden message in a way that makes it appear legitimate and innocent. The scenario discussed in the above (Sect. 3) demonstrates how the communications between Bob and Alice would not be unusual because their interest such as online business plays a role for camouflaging the delivery of list-cover. A legitimate reason, for sending, receiving, accessing, or obtaining a particular material, can legitimize the covert communications among communicating parties. Therefore, selecting the appropriate domain can play an essential role for securing the steganographic communications by establishing an appropriate covert channel for delivering a steganographic cover regardless its type.

3.3 Message encoding (Module 2)

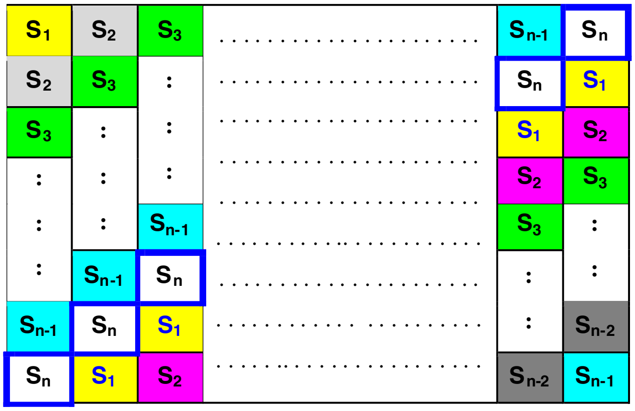
Implementing the message encoder follows a two-steps process. First, determining the encoding parameters in the selected domain by Module 1.Second, defining a steganographic coding, e.g., binary, octal, hexadecimal, etc., based on these parameters. A parameter in this context means some aspects of textual list that can be referred to steganographical values throughout a list-cover. Mainly, the itemized data that forms a list, such as books, computer parts, music CD's, movie DVD's, etc., can be exploited for concealing data.

The definition of the steganographic code would depend on the selected parameters. For example, encoding a message by using a list of books is different from encoding it using a list of computer parts or a list contains a combination of both books and computer parts and so on. The encoding module of Listega exploits theses type of options and determines the parameter(s) that will be employed for concealing messages. The popularity of certain list styles and types is an important factor in the selection. Nonetheless, unusual appearance of certain type of items may draw suspicion. For example, having a list of big airplanes on ebay for sale is an unusual practice. Another concerns is that when certain words are exploited for message encoding, e.g., the use of the word "planner" to mean "0", a set of items will has this word in the list rendering it suspicion. Listega methodology counters all of these concerns by simply imposing on the implementation of Listega system to be made aware of such issues or attacks. In addition, the domain selection is also crucial for justifying the interaction among the communicating parties to establish a covert channel for delivering the steganographic cover (list-cover). One would argue that the encoding parameters may actually influence the selection of a domain for the covert communication and should be done first. While this is a valid concern, the domain selection is crucial for justifying the interaction among the communicating parties and is thus more affected by the criteria for establishing a covert channel.

Listega does not impose any constraint on the message encoder scheme as long as it generates a set of data values that can be embedded in an list-cover. Given the availability of numerous encoding techniques in the literature that fit [3,8], the balance of this section will focus on an example that will be used in Sect. 4 to demonstrate the applicability of Listega. In the presented examples in Sect. 4, the encoding is done as follows. A message is first converted to a binary string. The string can be a binary of cipher text or a compressed representation. The binary string is then partitioned into groups of m bits. The value of m is determined based on the encoding parameters that Listega exploits. In textual list, the items' order, types of item, items' first letter, items' last letter, etc., can be exploited for concealing data. For example, if the list-cover will be four possible items, the binary message is partitioned it into groups of two bits, e.g., 00, 01, 10, and 11, corresponding to the possible items. Again, this encoding scheme is just for illustration and many alternate and more sophisticated schemes can be employed, as demonstrated in Sect. 4.

Countering coding patterns: One of the means for steganalysis is to investigate the steganographic cover, the media that contains hidden message, by looking for unusual patterns. In regards of list-covers an adversary may use the itemized data of a list and look for a pattern that may imply the presence of steganographic code such as using code words. The use of a fixed steganographic coding (encoding values) in multiple covers may create such a pattern. For example, an adversary may correlate itemized data over time. In order to prevent such potential vulnerability, Listega opts to employ some randomness for how the data are embedded in the cover. One possibility is to exploit multiple parameters in the process of encoding messages. An alternate strategy is to use multiple steganographic coding and establishes a protocol for when a particular coding is to be used. In any case, the communicating party ought to preagree on when a particular coding parameter or technique to be used so that a receiver can successfully extract and decode the hidden message. Listega advocates the use of Combinatorics in order to support the desired randomness in list-covers. Unlike non-combinatorics based approaches, the Combinatorics-based coding is predictable to the receiver but quite random to an observer who tries to analyze the steganographic cover (the media that contains hidden message), rendering the steganographic cover more resilient. To illustrate the idea, the following describes how Latin square [50,51] can be employed, by Listega, in defining steganographic coding in a form of table. It should be noted that in this paper a plaintext is concealed for simplicity. In realty a ciphertext is concealed rather than plaintext, which is common practice in steganography.

A Latin square is an $n \times n$ matrix that is filled with n distinct symbols, each occurs only once in each given row or column. An example is shown in Fig. 2. It should be noted that the first row does not have to start from S_1. In other words, the rows can be swapped. A Latin square can be employed in Listega by uniquely mapping a symbol to each value of the steganographic code. The mapping varies each time a message is encoded. Table 1 illustrates the idea through an example. Assume that list of items are pursued for concealing messages. Each item in the list matches the corresponding bit string in the encoded message. Note, the items are selected according to its domain while the concealment and the deeding is achieved according the first letter of each steganographic item. In the first time the encoder



S_1	S_2	S_3	....	S_{n-1}	S_n
S_2	S_3	:	....	S_n	S_1
S_3	:	:	....	S_1	S_2
:	:	:	....	S_2	S_3
:	:	S_{n-1}	....	S_3	:
:	S_{n-1}	S_n	....	:	:
S_{n-1}	S_n	S_1	....	:	S_{n-2}
S_n	S_1	S_2	....	S_{n-2}	S_{n-1}

Fig. 2 In an $n \times n$ Latin square, each row or column is a distinct permutation of n symbols

Table 1 The use of Latin squares introduces randomness in the definition of the steganographic code, yet keeps the code predictable for a receiver to successfully decode the hidden message while it is circulated rendering it unpredictable to an adversary

First letter of items	A	B	C	D
First time used	00	01	10	11
Second time used	01	10	11	00
Third time used	10	11	00	01
Fourth time used	11	00	01	10

Fifth time used will restart as first time and so on

is used, the first row or column of the Latin square will be used to map the first letter of each item such as A, B, C, and D to “00”, “01”, “10”, and “11”, respectively. A list of items, e.g., products, subjects, books, names, combination, etc., can conceal messages using Listega methodology. While concealing a message in the second time, the sender will use the second row (or column) which re-maps (rotates) the code among the items. The third time the sender will use the third row. After the fourth time, the first row will be used again and so on or a new Latin square with different order can be employed for increasing randomness. Note, there is no known close-form formula for the number of $n \times n$ Latin squares with symbols $1, 2, \ldots, n$. The upper and lower bounds that are deemed the most accurate by the technical community are far apart for large n, which makes Latin squares a powerful steganographic coding technique [50]. The receiver is aware of the Latin matrix and can successfully extract the message and decode it. Note that the order of the symbols can be formed differently as long as each element appears only once in each given row and column, e.g., using another Latin square. Table 2 shows an example of concealing the 8-bits ASCII representation of the letter “X”, which is “01011000”, using the matrix in Table 1. It is worth mentioning that the use of empty values, i.e., a null symbol, may also be used. It will then refer to noncoded element and can be manipulated in order to further avert suspicion.

3.4 Message camouflager (Module 3)

As mentioned earlier, the high demand and popularity of using textual itemized data by a wide variety of people render such text an attractive steganographic carrier. Listega takes advantage of such popularity and camouflages data in textual lists by manipulating, mainly, textual itemized data in order to embed messages without generating any suspicious pattern. From a steganographical point of view, reusing or altering an existing text to hide data is not a recommended practice since an adversary can reference the original text and detect the differences. In addition, the reuse of same piece of text more than once may increase vulnerability of the covert communications. If an adversary intercepts the communications and oversees a similar piece of text over and over again between communicating parties, suspicion may be raised because the adversary will wonder of such use. Inversely, this is not a concern in Listega methodology because reusing items or modifying textual list of items are a common practice. For example, an online-business, e.g., online-stores, online-sellers, online free stuff individuals posted by individuals (like craigslist.org), etc., may reuse and modify list of items. Such Listega’s strong feature eases the automation of a text-cover (list-cover). The automation process of list-cover, as illustrated in Fig. 3, is composed of three Submodules:

1. **Bank of textual items** (Submodule 1): This is simply a large database of textual items such as books, scientific subjects, nonscientific subjects, computer parts, music CD’s, movie DVD’s, names, etc., such huge collection of items will be employed in such a way to conceal data. Implementing such bank is accomplished by collecting the required textual items. This initially developed by humans that are experts and capable for doing such task. Thus, a list of items that is generated by such bank of items is often linguistically legitimate given the rigor that the development for the bank of items is made by experts. For example, the wording of items put on a test is often checked multiple times to ensure clarity and accuracy. In addition, the reuse of items, which is a common practice as mentioned above, further strengthens them linguistically given the multiple review cycles that they

Table 2 Demonstrating the effect of randomizing the steganographic code by Latin squares

First time communication letter “X” will be concealed in items that following letters	Second time communication letter “X” will be concealed in items that starts by the following letters	Third time communication letter “X” will be concealed in items that starts by the starts by the following letters
01 = B	01 = A	01 = B
01 = B	01 = A	01 = A
10 = C	10 = B	10 = A
00 = A	00 = D	00 = B

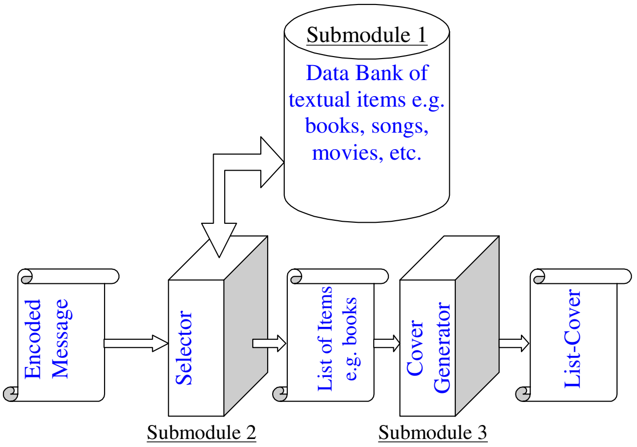


Fig. 3 Illustrates the architecture of message camouflager (Module 3)

go through. An example of such bank is the database of business inventory. The document database does not have to be centralized though. A distributed implementation as peer-to-peer system or web links can also be pursued. As noted earlier, updating such databases is continual and altering a list of items is not unusual and therefore would not draw suspicion. It is also worth noting that such bank can be multiple domains or limited to a particular domain.

2. **Selector** (Submodule 2): This picks the elements from the Bank of items (Submodule 1) that will form the list-cover. The criteria of selection are based on the domain and the message encoding scheme. For example, if the domain is selling books, the scope of the selection will be narrowed to such specific domain. On the other hand, if a list-cover uses books, Listega system will select a list of books that forms the list-cover. The picked books have to enable the concealment of the encoded messages. For example, if a message will be concealed by using book prices, a set of authenticated book prices that matches the symbols (bit string) used in the encoded message have to be picked. The order of these picked book prices in the list-cover is handled by Submodule 3, as explained next.
3. **Cover generator** (Submodule 3): This is responsible for forming a list-cover based on the textual items picked by the Selector (Submodule 2) while embedding the encoded message. For some styles of list-covers, the generator may as simple as enlisting the picked textual items in an order that matches the encoded message. For example, if the encoded message is concealed in a particular set of textual items, the items are then sequenced according to the symbols or the bits of the encoded message, respectively. Some other styles may require a higher level of slight sophistication in order to generate a wrapper. For example, the use of a sample list of computer items as a list-cover may requires special formatting and the inclusion of preamble, header, footer, etc. Since the sender may mix list-cover among other legitimate documents, obviously, the basic configuration of the covert channel should include how a recipient can only decode the right covers. For instance, the Cover Generator (Submodule 3) may putt items of the list-covers among similar items but noncoded items. This can be accomplished by following a particular sequence, such as odd number, even number, every other 3, etc., by placing list-covers in a specific folder, or any other way that is a preagreed upon between sender and receiver.

3.5 Communications protocol (Module 4)

The communicating parties configure the communications protocol of Listega system, as shown in Fig. 1, in order to communicate covertly by predetermining the following. First, the particular specifications of Listega system used including its decoder. Second, the covert channel for transmitting securely list-covers among communicating parties. Once communications protocol is agreed upon, the intended parties are ready to communicate covertly with each other using Listega. First, the particular specifications of Listega system used including its decoder. The first item is addressed by Module 1, 2, and 3 which are discussed in the previous sections. The second item is a particular covert channel that mainly defines how the cover will be delivered to the recipient without raising suspicion. Covert transmittal of the steganographic cover (the material that contains hidden message) is very crucial to the success of steganography. At the core of the cover transmittal issue is how to prevent the association between the sender and recipient from drawing suspicion. For example, exchanging email messages would automatically imply a relationship between the communicating parties. Similarly, downloading files from a web site indicates an interest in the accessed material. With advances in monitoring tools for network and Internet traffic, profiles of user's access pattern can be easily established. An adversary most probably will suspect the presence of a hidden message, even if the content does not look suspicious, because of the observed traffic pattern and the lack of a justification for the interest in the contents of such traffic. For example, if a sender or recipient his pretended profession is an online books-seller and sends or receives other suspicious documents such as list of huge airplanes then suspicious can easily be raised. A books-seller may send or receive only documents that are justifiable to be obtained such as a list of books. Therefore, it is very important to rationalize the sending and receiving of steganographic cover in order to avoid attracting any attention that may trigger an attack. Listega enables an effective solution for the issue of legitimizing a cover transmittal. The use of a particular domain(s) allows establishing a covert channel in a form of legitimizing the association among communicating parties and thus sharing a list-cover would appear an ordinary practice. The use of textual itemized data is highly demanded by a wide variety of people all over the world. Thus, the transmission of the list-covers via e-mail, posting them on web pages, etc., is a natural matter that does not raise suspicion.

4 Listega implementation

This section demonstrates the feasibility of Listega methodology and its distinct capability of achieving the steganographical goal with higher bitrate than contemporary linguistic steganography approaches. It is worth noting that the focus of this section is balanced on showing how Listega achieves the steganographical goal rather than making it difficult for an adversary to decode an encoded message. Employing a hard encoding system or cryptosystem to increase the protection of a message is obviously recommended and straightforward using any contemporary encoder or cryptosystem. Similarly, employing compression to boost the bitrate can easily be accomplished by using the contemporary techniques in the literature. This section shows just few examples of possible implementations following the steps outlined in the previous section.

4.1 Listega configuration

This section first explains how Listega modules are employed and configured to construct the overall Listega system used by the communicating parties.

Determining particular domain(*s*) (*Module* 1): In this paper two domains are employed, namely, the songs and books. Obviously, these domains are just examples and any other domains may apply as stated in Sect. 3. These domains are very popular worldwide among a wide variety of people. Such domains have no constrains for using any combination of items in a list which render these domains suitable to be used by Listega.

Listega encoder (*Module* 2): Listega encodes a message in a form that suits the camouflaging process. To increase the resilience against attacks, Listega introduces some randomness to the steganographic coding used. Therefore, a Latin square is used to define the randomly mapping of symbols to bit strings. The steganographical code in this Listega configuration works as follows:

1. Each item of list-cover conceals particular m bits according to the steganographic code defined. In the presented examples the length of bit string (*m* bits), that can be concealed in a particular item, either four or seven bits. The coding is not dependent of the item though. Instead, the first letter of an item in the list-cover contains a steganographic value according to Table 3. For example, when an item starts with the letter "B" while using the first row it is concluded that the item conceals "0001". However, if using the second row "B" implies "0000", and so on. In other words, an item itself is not encoded. Instead the first letter of each item in the list cover is checked against the Table 3 to find out its steganographic value. On the other hand, the coding that uses seven bits, the length of bit string (*m* bits), camouflage data in authenticated list of items' prices such as legitimate list of prices of books, song CD's, or flowers. The use of this table is illustrated later in this section.
2. Based on the protocol agreed-upon, either a particular row in the entire list-cover or the rows are used in particular order such as one row per item in the list-cover.

Message camouflage (*Module 3*): In this Listega system, that is presented in this paper, list-cover is mainly a list of items from the chosen domains by Module 1, the domains of songs and books. Obviously, these domains are just used as an implementation example and any other domains can be used. The camouflage module generates a text-cover (list-cover) by employing generic and specific (for a particular website) Internet search engines such as google.com, yahoo.com, ebay.com craigslist.org, etc., in order to generate list of items that can conceal data. The selected items, that are generated to camouflage data, are picked based on either the first letter of the item or the price that matches the steganographic code value of an encode message (the bit string of a message). As will be shown in the examples below, the use of first letters or authenticated items' prices does not impose constraints on the employed implementation. Based on the presented Listega configuration each item may conceal four to seven bits.

Communications protocol (*Module* 4): Configures the basic protocol of how a sender and recipient would communicate covertly. The basic communications protocol details the particular specifications of Listega system used that can camouflages data, its decoder that can unravels hidden messages, and the covert channel for transmitting securely list-covers among communicating parties. The chosen domains can play an essential role for legitimizing the discernable communications between sender and recipient such as the scenario of Bob and Alice in Sect. 3. For instance, when a sender and recipient have a related business or profession to the chosen domains of songs or books, then it is a legitimate and common practice to receive, send, obtains, etc., a textual list of items that are related to such domain. Generally, such relationship can justify the discernable association between the communicating parties to legitimize the transmittal of a list-cover. Once the communications protocol is agreed upon,

Table 3 The steganographic code for camouflaging four bits for each item

Code values	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1
	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0
Binary (4 bits) →	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
Decimal →	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7	8	9
↓ Row Index																										
1.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2.	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
3.	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
4.	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
5.	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
6.	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
7.	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
8.	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
9.	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
10.	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
11.	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
12.	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
13.	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
14.	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
15.	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
16.	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
17.	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
18.	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
19.	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
20.	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
21.	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
22.	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
23.	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
24.	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
25.	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
26.	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

The table is based on the properties of Latin square

the intended parties are ready to communicate covertly with each other using Listega. The following demonstrates examples of list-cover.

4.2 Listega examples

This section shows how Listega system can be used to conceal messages. Therefore, the following describes the process of encoding a message, concealing the encoded message in the generated text-cover (list-cover), and it demonstrates the samples of list-cover.

- The plaintext of two messages are: *"get him"* and "*Stop*"
- The Listega Encoder converts the message to a concatenated binary string using the ASCII representation of the individual characters, as follows:

  "get him" → 01100111011001010111010000100000011010000110100101101101
  "Stop" → 01010011011101000110111101110000
- Listega encoder will then divide the above binary message into slices of a particular size that matches the supported the steganographic coding used. It should be noted that the binary string could have been encrypted or compressed prior to this step. Nonetheless, the result is shown below:

  "get him" → 0110 0111 0110 0101 0111 0100 0010 0000 0110 1000 0110 1001 0110 1101
  "Stop" → 0101 0011 0111 0100 0110 1111 0111 0000
- The camouflage module considers the sliced bit string of the encoded message, generated by the encoder, and maps every slice to an item. The item will conceal a part of the message according to Table 3. A slice of 4 bits will

be assigned to each item. The samples of list-cover are shown bellow.

Samples of list-cover: The samples, based on the two domains used, demonstrate the effectiveness and efficiency of Listega. Table 4 demonstrates Sample 1 using first raw of Table 3, while the following is Table 5 which demonstrates Sample 2 according to the second raw of Table 3. As shown, in Tables 4 and 5 each item conceals 4 bites by selecting the item that starts by a particular letter that matches the steganographical code in Table 3. As observed, the use such legitimate list of items, e.g., products, subjects, books, etc., is widely popular, linguistically legible, and unsuspicious. For example, one may consider the list of songs on the Internet as shown in Fig. 4. Note, obviously the list of songs in Fig. 4 does not contain a hidden message and was just an innocent and common practice by people like songs.

4.3 Bitrates

The aim of this section is to compare the bitrate of contemporary linguistic steganography approaches to that achieved by Listega. The bitrate is defined as the size of the hidden message relative to the size of the cover. The average bitrate of the presented Listega system used in this paper is roughly 1.32 up to 3.87%. It is worth noting that the bitrate differs from one item to another, from one topic to another, and from one implementation to another as observed. To put this bitrate figure in perspective, the bitrate of contemporary linguistic steganography approaches has been investigated. The following reports on the findings, categorizing them based on the pursued approaches while Table 6 provides a concise summary of these findings.

1. The statistical-based approach, namely mimic functions: An experiment has been conducted using 30 samples generated using Spam Mimic [55]. An average bitrate of 0.90% is observed.
2. Synonym-based approaches:
   - For the NICETEXT scheme, the samples in [17,20] are used to estimate the bitrate, which is found to be approximately 0.29%.
   - The Winstein's scheme [26,27] roughly hides about 6 bits per sentence, which yields a bitrate of approximately 0.5% based on the sentences listed in the these publications. However, this rate cannot be generalized since not every sentence in the text-cover conceals data. In addition, the size of sentences will affect the bitrate because there are short and long sentences. Nonetheless, the 0.5% figure is assumed given that it is based on the samples developed by the authors.
   - The capability of the scheme of Murphy et al. [38] again is reported as the number of bits per sentence. Based on the samples provided in their publication, the achievable bitrate is roughly 0.30% per sentence.
   - Nakagawa et al. [32] have provided two samples for their scheme. The samples achieve bitrate of 0.06 and 0.12%, respectively. However, it has been noted

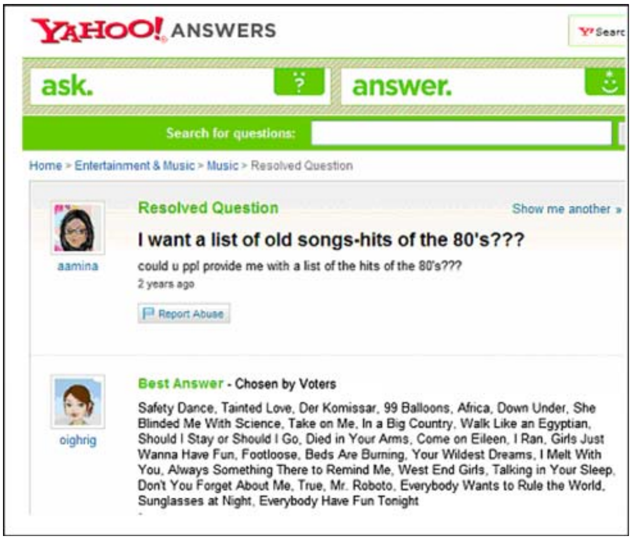
Table 4 Details, Sample 1, the camouflage of the encoded message "*get him*"

Index	Binary strings of ASCII representation for encoded message	First letter of selected item using first row of Table 3	List-cover	
			List of songs with singer names [52,53]	List of songs without singer names
1	0110	G or W →	Wicked Game—1989 Chris Isaak lyrics	Wicked game
2	0111	H or X →	How Do I Live—1997 LeAnn Rimes lyrics	How do i live
3	0110	G or W →	Wonderful Tonight—1978 Eric Clapton lyrics	Wonderful tonight
4	0101	F or V →	Faithfully—1983 Journey lyrics	Faithfully
5	0111	H or X →	Hero—2001 Enrique Iglesias lyrics	Hero
6	0100	E or U →	Endless Love—1981 Diana Ross & Lionel Richie lyrics	Endless love
7	0010	C or S →	Careless Whisper—1984 Wham! Lyrics	Careless whisper
8	0000	A or Q →	And I Love Her—1964 The Beatles lyrics	And i love her
9	0110	G or W →	Wild Thing—1966 The Troggs lyrics	Wild thing
10	1000	I or Y →	Your Song—1971 Elton John lyrics	Your song
11	0110	G or W →	Girl—1965 The Beatles lyrics	Girl
12	1001	J or Z →	Just Fine" is a song by Mary J. Blige	Just fine
13	0110	G or W →	I Want You To Want Me—1979 Cheap Trick lyrics	I want you to want me
14	1101	N →	Nobody Wants To Be Lonely—2000 Ricky Martin lyrics	Nobody wants to be lonely

The binary of this message is: 0110 0111 0110 0101 0111 0100 0010 0000 0110 1000 0110 1001 0110 1101

Table 5 Details the camouflage of the encoded message "*Stop*"

Index	Binary strings of ASCII representation for encoded message	First letter of selected item using second row of Table 3	List-cover	
			List of books with year and author names [52,54]	List of songs without year and author names
1	0101	G or W →	Warrior Heir, (2006). Axelrad, Catherine	Warrior Heir
2	0011	E or U →	Ever (2008). Fitzgerald, M.	Ever
3	0111	I or Y →	Year of Fog, (2008). Scott Sigler	Year of Fog
4	0100	F or V →	Vengeful Virgin, (1958). Benjamin, Ross	Vengeful Virgin
5	0110	H or X →	Hunting Wind, (2002). Smith, Melissa	Hunting Wind
6	1111	Q →	Q is for Quarry (2002). Sue Grafton	Q is for Quarry
7	0111	I or Y →	Inventing the Abbotts (1987). Joss, Morag	Inventing the Abbotts
8	0000	B or R →	Blood Is the Sky (2004). Steve Hamilton	Blood is the sky



Home > Entertainment & Music > Music > Resolved Question

Resolved Question

aamina

I want a list of old songs-hits of the 80's???

could u ppl provide me with a list of the hits of the 80's???

2 years ago

Best Answer - Chosen by Voters

oighrig

Safety Dance, Tainted Love, Der Komissar, 99 Balloons, Africa, Down Under, She Blinded Me With Science, Take on Me, In a Big Country, Walk Like an Egyptian, Should I Stay or Should I Go, Died in Your Arms, Come on Eileen, I Ran, Girls Just Wanna Have Fun, Footloose, Beds Are Burning, Your Wildest Dreams, I Melt With You, Always Something There to Remind Me, West End Girls, Talking in Your Sleep, Don't You Forget About Me, True, Mr. Roboto, Everybody Wants to Rule the World, Sunglasses at Night, Everybody Have Fun Tonight

Fig. 4 Shows the common practice of using textual list of songs

that when tried in a real application, only a bitrate of 0.034% could be reached.

3. Noise-based approaches:
   - The bitrate for the translation-based scheme reported in [23] is roughly 0.33%.
   - Based on the examples in [24], the confusing scheme approximately achieves a bitrate of 0.35%.
   - The linguistic techniques of the SMS-based methodology [25] is said to be capable of hiding few bits in a file of several kilobytes, which yields an extremely low bitrate.

Comparing the achieved bitrate by Listega which is roughly 1.32 up to 3.87% versus the bitrate achieved by the contemporary linguistic approaches in Table 6, it is obvious that Listega achieves much more superior bitrate than all comparable approaches, making it a very effective steganography approach. The high bitrate also enables the use of reasonable cover sizes, a major concern for all steganography approaches linguistic and nonlinguistic.

5 Steganalysis validation

The aim of this section is to show the resilience of Listega to possible attacks. Again the success of steganography is qualified with its ability for avoiding an adversary's suspicion of the presence of a hidden message. It is assumed that an adversary will perform all possible investigations. In addition, the adversary is also aware of Listega, as a public methodology, but he does not know the Listega configuration that the sender and recipient employ for their covert communication.

5.1 Traffic attack

One of the possible attacks an adversary may pursue is to inspect the communications traffic of images, graphs, audio files, etc., in order to detect the existence of covert communications if occurred. For example, the intelligence community has a number of tools at their disposal for analyzing traffic on the internet, tracking access to web sites, monitoring checked out literature from public libraries, etc. The main goal of a traffic attack is to detect unusual or questionable association between a sender and recipient. Traffic analysis intuitively can identify who communicates with whom. The relationship between the communicating parties will be then qualified based on the contents of the message. Traffic attacks can be a threat for most contemporary steganographic techniques regardless of the steganographic cover types (e.g., image, graph, audio file, text, etc.) used. In the context of Listega, the subject of the cover is checked rather than its validity and the consistency of its contents. If someone sends, receives, and accesses some materials without a legitimate reason for doing so, e.g., a pretended deaf

Table 6 The bitrate of contemporary linguistic steganography approaches

Approach	Bitrate (%)	Comment
Mimic functions [15,16]	0.90	Based on 30 samples generated at http://www.spamimc.com
NICETEXT [17,20]	0.29	Based on the samples in the cited papers
Winstein [26,27]	0.5	Based on the samples in the cited papers, and also confirmed in [38]
Murphy et al. [38]	0.30	Average per sentence (as reported in [38])
Nakagawa et al. [32]	0.12	As reported in [32], Bitrate achieved in real application is only 0.034%
Translation-based [23]	0.33	Noted by the authors in the cited papers
Confusing [24]	0.35	Based on the samples in the cited papers

person sends a songs CD to one of his friend, obviously suspicion can be raised and further investigation may be warranted. The additional investigations will involve a thorough analysis of a steganographic cover, as detailed in the next sections.

Traffic analysis is deemed ineffective with Listega. Listega camouflages the transmittal of a hidden message (list-cover) to appear legitimate and thus suspicion is averted. Basically, Listega ensures that the involved parties establish a covert channel by having a well-plotted relationship with each other rendering the communications traffic innocent and to look like any ordinary communications. Analyzing the traffic between them will not reveal any questionable association and will not trigger any further investigation. In addition, Listega imposes on the communicating parties to use innocent domains, e.g., contexts, martial, etc., that retains high demand by a wide variety of people. Such domains create a high volume of traffic that makes it impractical for an adversary to investigate all traffics. The voluminous traffic allows the communicating parties to establish a covert channel in order to transmit a list-cover without drawing attention, rendering Listega an attractive steganographical methodology to be used. Finally, it is noted that if further investigations on an list-cover are triggered by traffic analysis, they would not be successful, as elaborated next. In Listega, differentiating between a list-cover that contains a hidden message and another peer textual list of itemized data without a hidden message is extremely difficult.

5.2 Contrast and comparison attacks

One of the intuitive sources of noise that may alert an adversary is the presence of contradictions in a list-cover. Examples of these contradictions include finding suspicious repetition of an item. Also, if a list-cover contains errors, it is not expected to be numerous. Such contradictions may raise suspicion about the existence of a hidden message, especially when they are present in the same document. Automating the generation of a list-cover through the use of data banks makes the cover very resilient to this type of attacks. As demonstrated in Sect. 4, the use of a tool like Intent search engine, e.g., google.com, allows the selection of appropriate items that not only match the encoded messages but also ensure the validity of textual items and the clarity of the wording and the suitability of the scope. Meanwhile, noise in the context of comparison attacks reflects alteration of authenticated or previously used documents. The goal of the adversary is to find any incorrect and inconsistent data that may imply the manipulation of contents of a list-cover in order to embed a hidden message. However, since reusing and modifying textual lists are common practices, comparison attacks is deemed ineffective.

5.3 Linguistics attacks

Linguistics examination distinguishes the text that is under attack from normal human language. Distinguishing the text from normal human language can be done through the examination of meaning, syntax, lexicon, rhetoric, semantic, coherence, and any other feature that can help in detecting or suspecting the existence of a hidden message. These examinations are used to determine whether or not the text that is under attack is abnormal. Generally, the text used in textual lists is not sophisticated documents and it is easy of such scheme to retain the textual normality of list-cover. In addition, the produced textual itemized data meets the expected properties of a normal human language because it is initially generated by human and any alteration is done is more of cosmetic, e.g., changing the order of itemized data, and thus does not generate any noise (linguistic flaws). As a result, the generated cover as demonstrated in the implementation section is normal text. Furthermore, if there are errors in the list generator engine, it should not be a concern for two reasons. First, it applies to all other textual list of items that contains no hidden messages. Second, nothing is concealed in errors. In addition, an engine error of such list generator is most likely fixable. Therefore, Listega is capable of passing any linguistic attack by both human and machine examinations.

On the other hand, a statistical attack refers to tracking the profile of the used text. A statistical signature (profile) of a text refers to the frequency of words and characters used. An adversary may use the statistical profile of a particular topic of documents that contains no hidden message and compare it to a statistical profile of the suspected list-cover to detect any differences. An alteration in the statistical signature of a particular document may be a possible way of detecting a noise that an adversary would watch for. Unlike image steganography, tracking statistical signatures is an ineffective means for attacking linguistic steganography [21–23]. Nonetheless, Listega is resistant to statistical attacks because it is simply opt to use legitimate text that is generated naturally by human. In addition, the generated textual cover (list-cover) by Listega keep the same profile of its other peer documents that contains no hidden message. Basically, most alterations introduced by Listega are nonlinguistic and do not produce any flaws (noise), as demonstrated in the implementation section, deeming statistical attacks on list-cover ineffective.

6 Conclusion

The presented Listega conceals data in textual list of itemized data. The high demand for textual list of itemized data by a wide variety of people allows the communicating parties to establish a covert channel to transmit hidden messages (list-cover) rendering textual list of items an attractive steganographic carrier. Listega neither hides data in a noise (errors) nor produces noise. Instead, it camouflages data in legitimate list of items by manipulating, mainly the itemized data (e.g., list of books, movie DVD's, music CD's, auto-parts, etc.) in order to embed data without generating any suspicious pattern. The presented implementation achieves bitrate up to 3.87%. Such bitrate is superior to contemporary linguistic steganography approaches found in the literature, confirming the effectiveness of Listega methodology. Furthermore, Listega can be applied to all languages. The steganalysis validation has shown Listega methodology is capable of achieving the steganographic goal.

References

1. Kipper, G.: Investigator's Guide to Steganography, pp. 15–16. CRC Press LLC, Boca Raton (2004)
2. Davern, P., Scott, M.: Steganography its history and its application to computer based data files. Internal Report Working Paper: CA-0795. School of Computing, Dublin City University. http://computing.dcu.ie/research/papers/1995/0795.pdf (1995). Accessed 3 Aug 2006
3. Johnson, N.F., Katzenbeisser, S.: A survey of steganographic techniques. In: Katzenbeisser, S., Petitcolas, F. (eds.) Information Hiding, pp. 43–78. Artech House, Norwood (1999)
4. Kessler, G.C.: An Overview of Steganography for the Computer Forensics Examiner. An edited version, issue of Forensic Science Communications. Technical Report, 6, No. 3 (2004)
5. Martin, A., Sapiro, G., Seroussi, G.: Is image steganography natural? IEEE Trans. Image Process. **14**(12):2040–2050 (2005)
6. Desoky, A.: Graphstega: graph steganography methodology. J. Digit. Forensic Pract. **2**(1), 27–36 (2008). doi:10.1080/15567280701797087
7. Desoky, A., Younis, M.: PSM: Public Steganography Methodology. Technical Report TR-CS-06–07, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County (2006)
8. Petitcolas, F.A.P.: Information hiding—a survey. In: Anderson, R.J., Kuhn, M.G. (eds.) Proceedings of the IEEE, vol. 87(7), pp. 1062–1078 (1999)
9. Bennett, K.: Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text. Technical Report CERIAS Tech Report 2004–13, Purdue University (2004)
10. Shirali-Shahreza, M.H., Shirali-Shahreza, M.: A new approach to Persian/Arabic text steganography. In: The Proceedings of 5th IEEE/ACIS International Conference on Computer and Information Science (ICIS-COMSAR 2006), 10–12 July, pp. 310–315. Honolulu, Hawaii (2006)
11. Anderson, R.J., Needham, R., Shamir, A.: The steganographic file system. In: Proceedings of the Second International Workshop on Information Hiding. Lecture Notes in Computer Science, vol. 1525, pp. 73–82. Springer, Berlin (1998)
12. ScramDisk: Free Hard Drive Encryption For Windows 95 & 98. http://www.scramdisk.clara.net. Accessed 3 Aug 2008
13. Handel, T.G., Sandford, M.T.: Data hiding in the OSI network model. In: Information Hiding: First International Workshop, Proceedings. Lecture Notes in Computer Science, vol. 1174, pp. 23–38. Springer, Berlin (1996)
14. Kahn, D.: The Codebreakers: The Story of Secret Writing. Revised edition. Scribner, New York (1996)
15. Wayner, P.: Mimic functions. Cryptologia **XVI**(3), 193–214 (1992). doi:10.1080/0161-119291866883
16. Wayner, P.: Disappearing Cryptography, 2nd edn, pp. 81–128. Morgan Kaufmann, Menlo Park (2002)
17. Chapman, M., Davida, G.: Hiding the hidden: a software system for concealing ciphertext as innocuous text. In: The Proceedings of the International Conference on Information and Communications Security. Lecture Notes in Computer Science, vol. 1334, pp. 335–345. Springer, Beijing (1997)
18. Chapman, M., Davida, G.I.: Nicetext System Official Home Page. http://www.nicetext.com. Accessed 3 Aug 2007
19. Chapman, M. et al.: A practical and effective approach to large-scale automated linguistic steganography. In: Proceedings of the Information Security Conference (ISC '01), pp. 156–165. Lecture Notes in Computer Science, vol. 2200. Springer, Malaga (2001)
20. Chapman, M., Davida, G.I.: Plausible deniability using automated linguistic steganography. In: Davida, G., Frankel, Y. (eds.) International Conference on Infrastructure Security (InfraSec '02). Lecture Notes in Computer Science, vol. 2437, pp. 276–287. Springer, Berlin (2002)
21. Grothoff, C. et al.: Translation-based steganography. Technical Report CSD TR# 05-009, Purdue University (CERIAS Tech Report 2005-39) (2005)
22. Grothoff, C. et al.: Translation-based steganography. In: Proceedings of Information Hiding Workshop (IH 2005), pp. 213–233. Springer, Barcelona (2005)

23. Stutsman, R. et al.: Lost in just the translation. In: Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC'06). Dijon, France (2006)
24. Topkara, M., Topkara, U., Atallah, M.J.: Information hiding through errors: a confusing approach. In: Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents (2007)
25. Shirali-Shahreza, M. et al.: Text steganography in SMS. In: International Conference on Convergence Information Technology, Issue 21–23, pp. 2260–2265 (2007)
26. Winstein, K.: Lexical steganography through adaptive modulation of the word choice hash. January 1999. Secondary education at the Illinois Mathematics and Science Academy. http://alumni.imsa.edu/~keithw/tlex/lsteg.ps. Accessed 15 April 2008
27. Winstein, K.: Lexical steganography. http://alumni.imsa.edu/~keithw/tlex. Accessed 3 Aug 2008
28. Bolshakov, I.A.: A method of linguistic steganography based on collocationally-verified synonymy. In: Fridrich, J.J. (ed.) Information Hiding: 6th International Workshop. Lecture Notes in Computer Science, vol. 3200, pp. 180–191. Springer, Berlin (2004)
29. Bolshakov, I.A., Gelbukh, A.: Synonymous paraphrasing using wordnet and internet. In: Meziane, F., Metais, E. (eds.) Natural Language Processing and Information Systems: 9th International Conference on Applications of Natural Language to Information Systems, NLDB 2004. Lecture Notes in Computer Science, vol. 3136, pp. 312–323. Springer, Berlin (2004)
30. Calvo, H., Bolshakov, I.A.: Using selectional preferences for extending a synonymous paraphrasing method in steganography. In: Sossa Azuela, J.H. (ed.) Avances en Ciencias de la Computacion e Ingenieria de Computo—CIC'2004: XIII Congreso Internacional de Computacion, pp. 231–242 (2004)
31. Chand, V., Orgun, C.O.: Exploiting linguistic features in lexical steganography: design and proof-of-concept implementation. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS '06), vol. 6, p. 126b. IEEE, New York (2006)
32. Nakagawa, H., Sampei, K., Matsumoto, T., Kawaguchi, S., Makino, K., Murase, I.: Text information hiding with preserved meaning—a case for japanese documents. IPSJ Trans. **42**(9):2339–2350 (2001). Originally published in Japanese. A similar paper By the first author in English. http://www.r.dl.itc.u-tokyo.ac.jp/nakagawa/academic-res/finpri02.pdf. Accessed 4 June 2008
33. Niimi, M., Minewaki, S., Noda, H., Kawaguchi, E.: A framework of text-based steganography using sd-form semantics model. IPSJ J. **44**(8) (2003). http://www.know.comp.kyutech.ac.jp/STEG03/STEG03-PAPERS/papers/12-Niimi.pdf. Accessed 3 June 2008
34. Bergmair, R., Katzenbeisser, S.: Content-aware steganography: about lazy prisoners and narrow-minded wardens. In: Proceedings of the 8th Information Hiding Workshop. Lecture Notes in Computer Science, vol. 4437, pp. 109–123. Springer, Berlin (2007) (in print)
35. Bergmair, R.: Towards linguistic steganography: a systematic investigation of approaches, systems, and issues. final year project, The University of Derby (2004)
36. Bergmair, R., Katzenbeisser, S.: Towards human interactive proofs in the text-domain. In: Proceedings of the 7th Information Security Conference (ISC'04). Lecture Notes in Computer Science. Springer, Berlin (2004)
37. Topkara, U., Topkara, M., Atallah, M.J.: The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions. In: MM&Sec '06: Proceeding of the 8th Workshop on Multimedia and Security, pp. 164–174. ACM Press, New York (2006)
38. Murphy, B., Vogel, C.: The syntax of concealment: reliable methods for plain text information hiding. In: Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents (2007)
39. Atallah, M.J., Raskin, V., Crogan, M., Hempelmann, C., Kerschbaum, F., Mohamed, D., Naik, S.: Natural language watermarking: design, analysis, and a proof-of-concept implementation. In: Moskowitz, I.S. (ed.) Information Hiding: Fourth International Workshop. Lecture Notes in Computer Science, vol. 2137, pp. 185–199. Springer, Berlin (2001)
40. Atallah, M.J., Raskin, V., Hempelmann, C.F., Topkara, M., Sion, R., Topkara, U., Triezenberg, K.E.: Natural language watermarking and tamperproofing. In: Petitcolas, F.A.P. (ed.) Information Hiding: Fifth International Workshop. Lecture Notes in Computer Science, vol. 2578, pp. 196–212. Springer, Berlin (2002)
41. Desoky, A.: Nostega: noiseless steganography paradigm. J. Digit. Forensic Pract. (in press)
42. Desoky, A. et al.: Auto-summarization-based steganography. In: The Proceedings of the 5th IEEE International Conference on Innovations in Information Technology, Al-Ain, UAE (2008) (in press)
43. Cvejic, N., Seppanen, T.: Increasing robustness of LSB audio steganography using a novel embedding method. In: The Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04), pp. 533–537, Las Vegas, Nevada (2004)
44. Cvejic, N., Seppanen, T.: Reduced distortion bit-modification for LSB audio steganography. In: '04. 2004 in the Proceedings of the 7th International Conference on Signal Processing (ICSP 04), vol. 3, pp. 2318–2321, Beijing, China (2004)
45. Bender, W. et al.: Techniques for data hiding. IBM Systems J. **35**(3, 4), 313–336 (1996)
46. Kirovski, D., Malvar, H.: Spread-spectrum audio watermarking: requirements, applications, and limitations. In: The Proceedings of the 4th IEEE Workshop on Multimedia Signal Processing, pp. 219–224. Cannes, France (2001)
47. Ansari, R., Malik, H., Khokhar, A.: Data-hiding in audio using frequency-selective phase alteration. In: The Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '04), 17–21 May, vol. 5, pp. 389–92 (2004)
48. Gruhl, D., Lu, A., Bender, W.: Echo hiding. In: The Proceedings of First International Workshop on Information Hiding. Lecture Notes in Computer Science, vol. 1174, pp. 295–316, Cambridge, UK. Springer, Berlin (1996)
49. Desoky, A., Younis, M.: Chestega: chess steganography methodology. J. Secur. Commun. Netw. (in press)
50. Laywine, C.F., Mullen, G.L.: Discrete Mathematics Using Latin Squares, 1st edn. Wiley-Interscience, London (1998)
51. Dénes, J., Keedwell, A.D.: Latin Squares (Annals of Discrete Mathematics), vol. 46. Elsevier, North-Holland (1991)
52. Google Internet Search Engine: http://www.google.com, used to generate the lest of items. Accessed 26 Sept 2008
53. List of Songs: http://www.advicenators.com/qview.php?q=549142. Accessed 26 Sept 2008
54. Intranet Book List: http://www.iblist.com. Accessed 26 Sept 2008
55. Spam Mimic: http://www.spammimic.com. Accessed 31 July 2007