Text Steganography System Using Markov Chain Source Model and DES Algorithm

Weihui Dai
School of Management, Fudan University
Email: whdai@fudan.edu.cn

Yue Yu, Yonghui Dai and Bin Deng
School of Management, Fudan University, School of Software, Fudan University, School of Software, Fudan University
Email: 072025045@fudan.edu.cn, dyh822@163.com, 033053245@fudan.edu.cn

***Abstract*—High transmission efficiency, low resource occupancy and intelligible meaning make text message as the most commonly used type of media in our daily communication. Text steganography, an information hiding technology based on text message, has been a new exploring research in recent years. Due to the restriction of redundant information as well as its alterability in manual operation, text message is difficult to hide secret information effectively and reliably.**

Based on Markov Chain source model and DES algorithm, this paper presents a text steganography system for spelling languages. It can work reliably with the capability of immunity from regular operations, such as formatting, compressing and sometimes manual altering operation in text size, front, color and the space between words. By adopting the technique of heuristic composition, this system can produce cover text close to nature language. It's suitable for hiding short information in online communication, such as E-mail, MSN Messenger, QQ, instant conversation and the short message on mobile phone.

***Index Terms*—Information Hiding; Text Steganography; Markov Chain; DES Algorithm**

I. INTRODUCTION

In recent years, information hiding technology has been a new exploratory field with its superiority by hiding sensitive information secretly into a non-secret cover object to enforce the security of encrypted information as well as to verify the authenticity of its cover object[1][2]. It has been widely applied in digital watermark and probably the secret communication under surveillance [3]. The cover object can be any kind of commonly used media or file, such as image, sound, video, software code or text document [4][5].

Information hiding is different from encrypting with its intention to avoid the intrusion or discovering to the hiding data, other than to restrict the data access [2]. Moreover, it should avoid any damages in the hiding data, with the capability of immunity from those regular operations, such as transformation and compression. Technology about the information hiding in image, sound, video and other no-text format media has been successfully developed [2][4][5], owing to the adequacy of redundant information in those medium. Furthermore, observers are insensitive to the changes in them [1]. High transmission efficiency, low resource occupancy and intelligible meaning make text message as the most commonly used type of media in our daily communication. However, due to the restriction of redundant information as well as its alterability in manual operation, text message is difficult to hide secret information effectively and reliably [2][6][7].

Since 1990s, information hiding technology based on text media has developed from the focus of digital watermark into secret communication. A famous software system designed to hide encrypted data into text document was introduced by M. T. Chapman in 1997 [6]. In 2003, Chinese researchers X. X. Niu and Y.X. Yang presented a new algorithm for the communication of text steganography [7], which hides meaningful information into a stegotext by technical changing in its cover message, such as the character characteristics (size, space, color, front, attribute, intensity, etc.) [8]-[11], sentence structure [12] and other statistical characteristics [13]. Under the circumstances, the invader doesn't know the cover message whether hides other information. Even he knows, it is difficult to distill or wipes off the hidden information.

.Up to now, various approaches and algorithms have already been explored in this field [6]-[18]. Reference [14] summarized the current methods which utilize the space between words, rows and punctuations to realize text steganography. Reference [17] proposed an algorithm to hiding information of the cipher text by the changes in text front and text color. Other approaches and algorithms based on character front [8], character color [9], character intensity [10], structure of the natural language [11] [18], attributes of the HTML markup [12] and the statistical characteristics of characters [13] were presented as recent explorations. On the other side, steganalysis technology has been successfully developed

Manuscript received June 24, 2009; revised September 1, 2009; accepted September 10, 2009.

This research was supported by Shanghai Leading Academic Discipline Project (No.B210).

doi:10.4304/jsw.5.7.785-792

to detect and find the hiding encrypted information in covertext by analyzing its redundancy [2][19].

By a comprehensive analysis of existed researches, we can draw a key problem that the capability of immunity from regular operations, such as formatting, compressing and sometimes manual altering operation is expected to be further explored so far in this field. At the same time, reduce of redundancy in covertext is to be improved to both ensure the transmission efficiency and antagonize the steganalysis.

Markov Chain model is an important probabilistic model to solve sequence representation and statistical problem, and have been applied with success to many related areas [20]-[22]. It can make use of the conditional probability rather than the redundant information to achieve the purpose of text steganography. With the help of Markov Chain source model and DES algorithm, this paper presents a text steganography system to realize information hiding by text message of spelling languages. This system can work reliably with the immunity from regular operations in transformation and communication, and produces cover text close to nature language by adopting the technique of heuristic composition.

II. TEXT STEGANOGRAPHY

A. Information Hiding

Information hiding is a technology that hides meaningful information to a Cover C to get the Stego Cover S. In order to increase the offense difficulty, we can combine encrypting technology with information hiding technology. That is to encrypt the Message M to get the cryptograph information M', and then hide M' to Cover C. In this way, even though the invader wants to get the message, he should first detect the existence of the information, and know how to distill M' from the secret cover S, then decrypt M' to recovered message M.

B. Text Steganography

Text steganography can be applied in the digital makeup format such as PDF, digital watermark or information hiding. Different from the analog signal method based on image, sound or video, the text cover doesn't use signal disposing model. It is more difficult to realize the information hiding based on text. The simplest method of information hiding is to select the cover first, adopt given rules to add the phraseological or spelling mistakes, or replace with synonymy words. For example, Textto [23] setups some sentence structure in advance, fills in the empty location by arranged words, and then the text doesn't have phraseological mistakes, but have some word changes or morphology mistakes. This complex method is to produce secret text according to hiding information, and needn't select covers in advance. A more complex method is to use nature language disposal to make the secret text more nature. TextHide [24] hides the information in the manner of text overwriting and words' selection. NiceText [6] may imitate the given sample's writing style to create the text of approximate nature language, embed hiding information in the creating process. Another method is to read in all the character in its coding mode, these coding numbers exist in integer form, and haven't any redundancy, to express this bunch of number in its it stream, through some transform such as wavelet transform, FFT transform, DCT transform to get the signal that has redundancy, and then to disguise the text in the redundancy space. The current major methods all aim to English text, the research of information hiding technology whose cover is in other language is still very less.

III. MARKOV INFORMATION SOURCE AND THE MODEL FOR NATURE LANGUAGE

A. Markov Chain

Stochastic process analysis is the dynamic description to dynamic relation between a series of random events. It is a powerful tool to research the random phenomena in many fields such as natural science, engineering science and social science. Markov chain is a specific stochastic process, discrete time, discrete status and a non-aftereffect process.

Suppose there is a stochastic variable sequence (always time related), it meets the following conditions: The stochastic variables aren't independent of each other, each stochastic variable is just dependent of the former variable. In many similar systems, we can suppose: We can forecast the future status based on the current status and can ignore the past status. That is, the future stochastic variable has nothing to do with the past, yet conditionally depend on the current status. Such a sequence of variables is called Markov Chains or said to have with the Markov features.

Formally, suppose there is a stochastic variable sequence $X = \{X_1, X_2, ..., X_T\}$, assume the value $S = \{s_1, s_2, ..., s_N\}$, when this sequence of variables has the following attributes:

(i) $P(X_{t+1} = k \mid X_1, X_2, ..., X_t) = P(X_{t+1} = k \mid X_t)$

(ii) $P(X_{t+1} = k \mid X_t) = P(X_2 = k \mid X_1)$

We call such stochastic variable sequence Markov Chains. If the condition probability is independent of the time, it will be called Time Homogeneous Markov Chains.

B. Markov Information Source

In communication system, we can use stochastic variable, stochastic vector and stochastic process to describe the information from its source, or use a sample space and its probability space to describe it.

Some messages from the information source maybe finite or denumerable. We can use one dimension discrete stochastic variable X to describe the output of the information source. We call this kind of source the discrete information source.

The signal sequences have finite dependent relation of some information sources, that is, the probability of the signal at any times is related to some former signal. In order to describe this kind of information source, besides information source signal collect, we should introduce status S. The current output signal is related to the status of the information source.

Suppose the common information source is in the state of $S \in E = \{E_1, E_2, ..., E_J\}$, and in every state may output signal $A = \{a_1, a_2, ..., a_q\}$. And suppose at every time point when the information source sends out a signal, the status of the information source will transfer. Suppose the output stochastic sequence from the information source is $x_1, x_2, ..., x_{l-1}, x_l, ...$, the stochastic status of the information source is $s_1, s_2, ..., s_{l-1}, s_l, ...$. If the signal sequence from the information source and the status of the information source satisfy the following condition:

1) In a certain time point, the output of information source signal is only dependent to its current status, and is independent of the former status and former output signal.

That is:

$$P(x_l = a_k \mid s_l = E_i, x_{l-1} = a_{k_1}, s_{l-1} = E_j, ...) = P(x_l = a_k \mid s_l = E_i)$$

When it is time homogeneous, it has

$$P(x_l = a_k \mid s_l = E_i) = P(a_k \mid E_i) \quad .1$$

, and

$$\sum_{a_k \in A} P(a_k \mid E_i) = 1 \quad . \ .$$

2) The state of the information source in the *i* time point only depends on the current output signals and the former state in the *i-1* time point, that is:

$$P(s_l = E_j \mid x_l = a_k, s_{l-1} = E_i) = \begin{cases} 0, E_i, E_j \in E \\ 1, a_k \in A \end{cases}$$

. .

Then this information source can be called Markov Information Source.

The Markov Information Source Status Collect E is related to the signal sequence. If at any *i* time point, the probability of the signal of information source is only dependent of the former *m* signals, we call it m-rank memory discrete information sources. It is decided by a group of information source signals and a group of condition probability:

$$X : \lfloor P(a_{k_{m+1}} \mid a_{k_1}, a_{k_2}, ..., a_{k_m}) \rfloor \quad . \ .$$

$$k_1, k_2, ..., k_m, k_{m+1} \in \{1, 2, ..., q\}$$

Here,

$$\sum_{k_{m+1}=1} P(a_{k_{m+1}} \mid a_{k_1}, a_{k_2}, ..., a_{k_m}) = 1 \quad . \ .$$

Such information source X can be called m-rank Markov Information Source. In most cases, the sequence of nature languages can be approximated as a signal from Markov Information Source, which has been applied to many areas of language analysis and machine training [20][21].

C. Markov Chain Source Model for Nature Language

The natural spelling languages such as English, French and German which we use in daily life are all information sources that composed by a group of signal collection. These signals are dependent, we can use Markov information source to approach it. If we take English language as the example, the signal collect is selected as letters and spaces, and its output of information is a letter sequence. Because there is not the same probability of the English letters to buildup words, these letters have strict dependent relation. Table I shows the letters' probability [3].

TABLE I. LETTERS' PROBABILITY TABLE

Letter	Probability	Letter	Probability
space	0.1859	N	0.0574
A	0.0642	O	0.0632
B	0.0127	P	0.0152
C	0.0218	Q	0.0008
D	0.0317	R	0.0484
E	0.1031	S	0.0514
F	0.0208	T	0.0796
G	0.0152	U	0.0228
H	0.0467	V	0.0083
I	0.0575	W	0.0175
J	0.0008	X	0.0013
K	0.0049	Y	0.0164
L	0.0321	Z	0.0005
M	0.0198		

If we only consider the appearing frequency of the words and the dependent relation between words, we can use one-rank Markov information source to describe the source of English words. The creating sequence from Markov information source represents an actual English article. So we can use Markov chain to create secret text, thus realize information hiding. Reference [3] presented a concise method that formed approximate 2-rank Markov information source of English word sequence.

Choose a copy of classical English book, turn to a page, randomly choose two connected words, such as 'sunshine on', put these two words as the former two words in the classical series, then randomly choose

another page. If find ‘sunshine on’ on this page, put the word which follows the ‘sunshine on’ as the third word in the classical series. Do the similar way using the third word and in this way, we can get the representative sequence that output by Markov information source which approximates the English information source. By this way, you do not have to calculate the conditional probabilities and avoid complicated computing, and can also generate the classical series which meet Markov approximate information source attribute, for example as Fig.1[3].

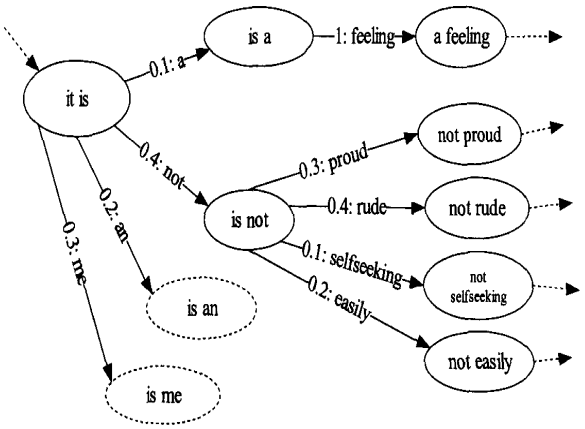

Figure 1. The status conversion figure of Markov approximate information source

If the representative output sequence of the approximate information source appears ‘it is’, then according to different probability, we can select the different embranchment of the status conversion figure to create different representative sequence. Take advantage of this characteristic, we can communicate in secret. If we read ‘it is’ from the secret text, and the next read in character is ‘not’, then we know the probability is 0.4. This creating process of text is the process of status conversion of selecting different embranchment.

We can do further optimization: in order to avoid complex probability calculation, we suppose the probability of a certain status to other status is equal. According to the definition of condition entropy, the information amount of an information source signal from a status of an information source can be expressed as follows:

$$H(X|s=E_i)=-\sum_{k=1}^{n}P(a_k|E_i)\log P(a_k|E_i)$$

Use E_i=(it is) as an example, due to the equal probability, then $P(a_k \mid E_i)=\frac{1}{4}$, the information amount is:

$$H(X|s=E_i)=-\sum_{1}^{4}\frac{1}{4}\log\frac{1}{4}=\log 4=2.$$

Thus

$$H(X|s=E_i)=\log n.$$

We get the conclusion that the information amount of an information source signal from a status that converts to other status is only related to the amount of the information source signal from a status.

According to this conclusion, we can mark the status figure to number the different embranchment of the status conversion, thus get the Fig.2 [3].

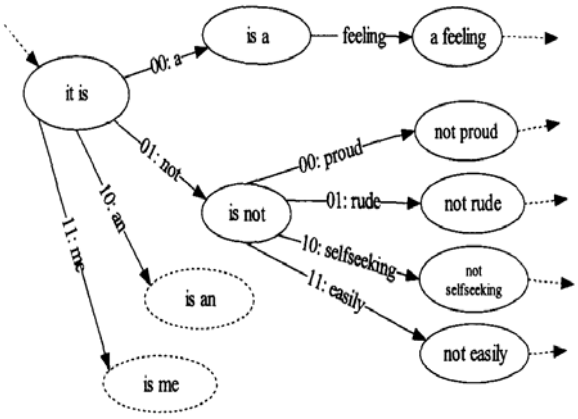

Figure 2. The status conversion figure after marked

From ‘it is’ to ‘not’, we get the secret information ‘01’. From ‘it is’ to ‘selfseeking’, we get the secret information ‘10’. Repeat this process, we can get the secret text ‘it is not selfseeking…’, and get the secret information ‘0110…’. If the receiver has the status conversion figure of this information source, according to the method above, repeat the process until all the signal is read, we can recover the information in the text accurately.

IV. STATUS CONVERSION FIGURE AND INFORMATION CONVERSION IN TEXT STEGANOGRAPHY

A. Sample Text and Status Conversion Figure

There are two steps while using Markov Chain source model to hide information in the text: First, establish this status conversion figure; Second, share this figure by the sender and the receiver. But it is impossible to establish the status conversion figure of the entire Markov chain from a nature language.

We can select some sample text to determine the status conversion figure of the Markov approximate information source, and then fix a statistic algorithm to create the same status conversion figure to the same sample text. The sender uses the status conversion figure to hide information in the text, and after the receiver receives the text, he can recover the secret information by the status conversion figure. They just need to confirm the sample text and algorithm before communication.

Actually the prefix and postfix tables in the creating process of the status conversion figure can be united as a data structure similar to tree during the process of system

realization. The word in the prefix table is the parent node of the tree's child node, and the word in the postfix table is the tree's child node. The probability of the word's combination is the weight value of the tree's node. The probability of a certain node is decided by its parent node and the upper level parent node. For the example of English language, we can get the data structure of the probability figure from the sample text as follows:

```
# define MAX_CHART_SIZE 10000
Typedef struct PTNode
{
 WORDSCHAR word;      /* English words       */
 PROWORDCHAR parent;  /* Pre-words position
*/
 INT       probability;  /* Probability value */
} PTNode;
Typedef struct
{
   PTNODE    nodes[MAX_CHART_SIZE];
   INT       wordnum;  /* All English words amount
*/
}
```

After that, we convert the probability of each English word connected to the following word to the binary data, and different from each other, to meet the need of the subsequent information hiding. We put the high probability prefix and postfix words ahead, and put the low probability ones at the truncation. See the definition of the data structure of the final status conversion figure as follows:

```
# define MAX_CHART_SIZE 10000
Typedef struct PTNode
{
 WORDSCHAR word;      /* English words       */
 PROWORDCHAR parent;  /* Pre-words position
*/
 INT       binavalue;  /* Binary value */
} PTNode;
Typedef struct
{
   PTNODE    nodes[MAX_CHART_SIZE];
   INT       wordnum;  /* All English words amount
*/
}
```

Text steganography and its recovery are all based on the data structure relation of this conversion figure.

B. Information Conversion and DES Algorithm

Before creating the secret text, we need to convert the information that needs to be hidden to binary data, thus can do the subsequent encrypt and information hiding. The current popular computer system can regard the data file as a byte stream, so we just need to solve the problem of the conversion from byte stream to bit stream. We adopt DES (Data Encryption Standard) algorithm to realize the data encryption and decryption.

DES is an algorithm to encrypt the binary data. The grouping data length, secret key length and output secret text length are all 64 bits. DES is a block cipher selected by NBS as an official Federal Information Processing Standard (FIPS) for the United States in 1976 and which has subsequently enjoyed widespread use internationally.

It is based on a Symmetric-key algorithm that uses a 56-bit key (3DES uses 128-bit key). The algorithm was initially controversial with classified design elements, a relatively short key length, and suspicions about a National Security Agency backdoor. DES consequently comes under intense academic scrutiny which motivates the modern understanding of block ciphers and their cryptanalysis.

DES algorithm has been the world range standards of ANSI and ISO [25] for more than 20 years. Up till now, there is no effective method except the exhaustive search to attack this algorithm. So DES (or 3DES for high safety) is still effective to defend most attacks.

V. TEXT STEGANOGRAPHY SYSTEM USING MARKOV CHAIN SOURCE MODEL AND DES ALGORITHM

A. System Framework and Data Processing

The design objective of this system is to provide a software tool for text steganography and its recovery based on spelling languages. Fig.3 shows the system framework and its data processing.

Before the communication, the sender and receiver may select a specific sample text only for this communication. The sample text may come from some public text documents. Selected sample text document is first analyzed with the same Markov Chain source model by both the sender and the receiver; thereof produces the same conversion figure data in both sides. In this communication, sensitive information are converted to binary data (compressed) and encrypted by DES algorithm to form cryptography data by the sender. According to some inherent rules regulated by the conversion figure data, the sender can compose a cover text. This process is operated with the help of heuristic composition, which provides heuristic technique for the composition of cover text and make it close to the nature languages. For example, if the cryptography data are "1A587C13" and you have input "We" as the first word of the cover text, heuristic composition of the following words are suggested as:

"We need drink (get it, buy car)……"

You may choose any two words as the followings of "We" from "need drink", "get it" or "buy car". By some similar steps, you can easily composite a whole sentence as the cover text:

"We need drink after a long travel in mountainous area and have a good rest in the evening."

The cryptography data "1A587C13" have been successfully hid in the above sentence.

After receiving the cover text, the cryptography data can be distilled by the receiver according to the rules in conversion figure data, and then decrypted from DES algorithm. At last, the original information can be recovered from the decrypted binary data.

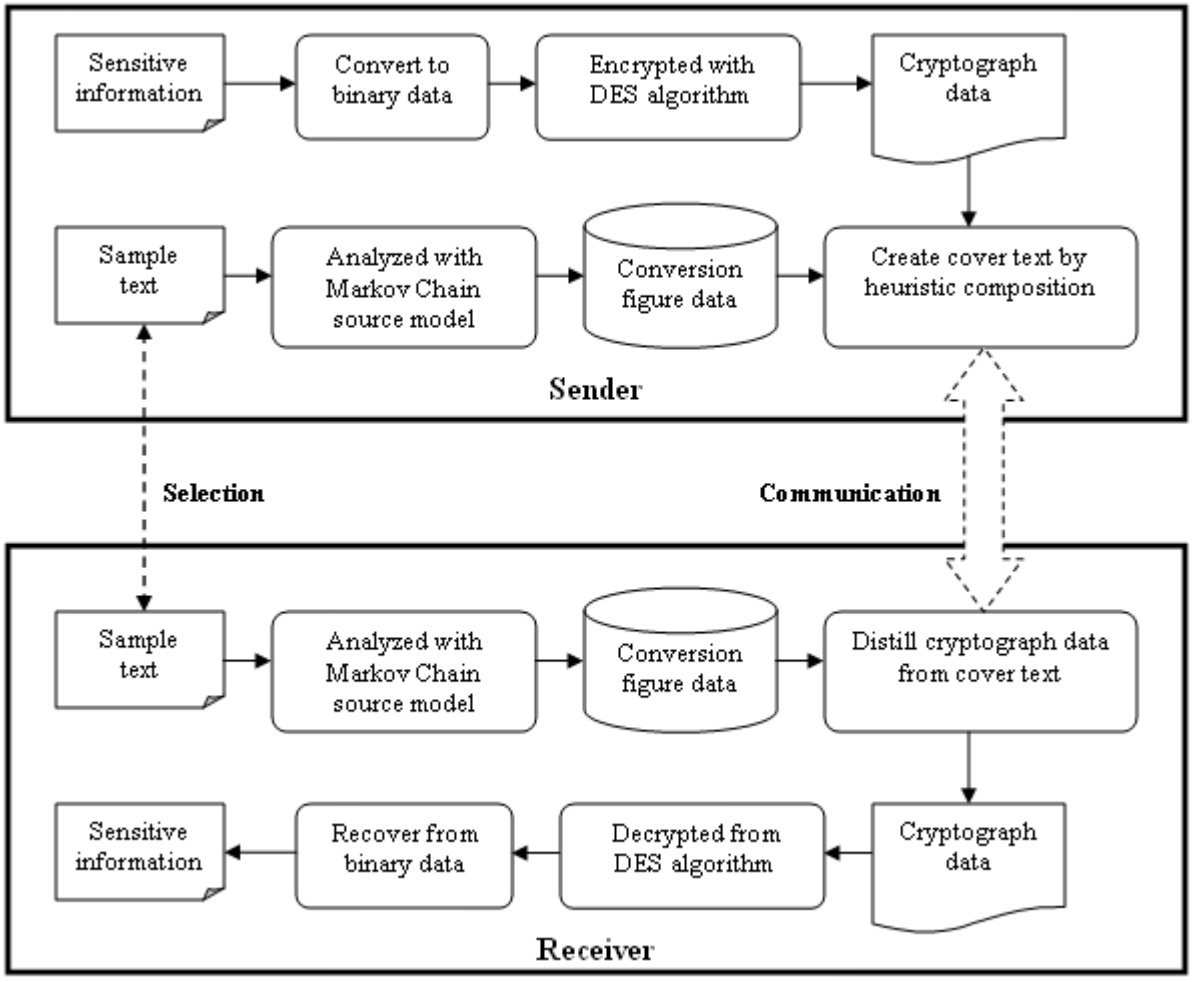

Figure 3. Text steganography system using Markov Chain source model and DES algorithm

B. System Function Modules

This system is composed of six modules: interface, sample analysis, information encryption, information hiding, information recovery, and information decryption. Fig.4 shows its whole structure:

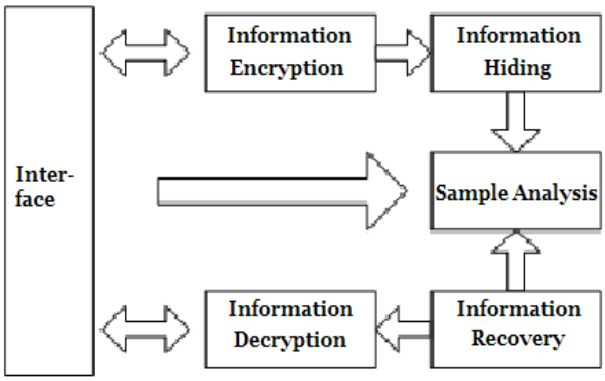

Figure 4. Function modules in text steganography system

The user interface deals with the problem of the alternation of the users which includes the operation such as sample setup, information input and display, encrypting and decrypting, hiding and recovery; The sample analysis module analyzes the sample text, using Markov Chain source model to create the status conversion figure between words. In this process, it determines the weight value of the two words according to their adjacent probability.

With regard to the management of the conversion figure data structure, there are some basic operational function prototypes to be explained:

CreateConverChart(ConverChart &C, Int wordnum): Construct the data structure of the status conversion figure, define the weight value of each word according to the probability.

OrderConverChart(ConverChart &C, Status (*Condition)): Sort the data of the status conversion figure according to the condition, for example, advance the high probability prefix and postfix words.

InsertConverChart(ConverChart &C, PTNode newnode): In the status conversion figure, insert new nodes according to the word and the probability of the former and latter connecting combination.

DeleteConverChart(ConverChart &C, PTNode newnode): In the status conversion figure, delete old

nodes according to the word and the probability of the former and latter connecting combination.

GetRoots(ConverChart &C, PTNode (*Roots): In the status conversion figure, get the initial word data collect.

TraverseConverChart(ConverChart &C, Status (*Visit)): *Visit is an application function operating to the nodes in the data structure.

The information encryption module converts the sensitive information that to be hidden to the binary data and encrypts it with DES algorithm. The information hiding module converts above cryptography data into a cover text of nature language by heuristic composition. The information recovery module distills the cryptography data from cover text also according to the conversion figure. The information decryption module decrypts the binary information from cryptography data, and converts it to original information.

By this system, the communicating parties can reach concealed communication even through a public channel; thereof improve the security of the online communication.

VI. Discussion and Application

This system uses the conditional probability between text words to achieve the purpose of text steganograph, so it has the capability of immunity from regular operations, such as formatting, compressing and sometimes manual altering operation in text size, front, color or the space between words, and can work reliably. But affected by the DES algorithm and the manual operation on the composition of cover text close to nature language, system efficiency probably restricts its application to hide short information. It needs a reasonable trade-off between the efficiency and security, especially between the efficiency and language quality. In order to improve the language quality, we explore a man-machine cooperation approach by adopting the technique of heuristic composition. It is acceptable while compositing a short cover text.

For daily application, this system is suitable for hiding short information in online communication, such as E-mail, MSN Messenger, QQ, instant conversation and the short message on mobile phone. It can embed sensitive information into text message and realize secret communication on public network. As in business or finance areas, we can apply this system to hide the sensitive data in E-business and E-bank trading while communicating online. It can also help to break out the barriers in the transmission of some sensitive information monitored by firewall or restricted by surveillance.

VII. Conclusion

High transmission efficiency, low resource occupancy and intelligible meaning make text message as the most commonly used type of media in our daily communication. Consequently, text steganography has been the mostly applicable information hiding technology in increasing online communication. However, text steganography is difficult to be realized effectively and reliably because of limited redundant information and weak capability of immunity from regular operations.

In this paper, we present a text steganography system based on Markov Chain source model and DES algorithm. This system can work reliably and produces cover text close to nature language. It's suitable for hiding short information in online communication, such as E-mail, MSN Messenger, QQ, instant conversation and the short message on mobile phone.

To spelling language like English, French and German, the system can be applied directly. But to Chinese, it is worth of our further researches.

Acknowledgment

This research is supported by Shanghai Leading Academic Discipline Project (No.B210).

References

[1] W. Bender, D. Gruhl, N. Morimoto, et al., "Techniques for data hiding," *IBM System Journal*, vol.35(4), pp.313-336, 1996.

[2] J. Jin., *Research on Data Hinding in Text Documents*, Shantou: Shantou University, 2008.

[3] S. F. Wu, *Researches on Information Hiding Technology*, Hefei: China Science and Technology University, 2003.

[4] Z. X. Dai, and F. Hong, "Text information hiding based on inverse order of part of speech symbol sequence," *Computer Engineering and Applications*, vol.43(14), pp.160-161, pp.198, May, 2007.

[5] J. X. Huang, "Applying hidden markov chain algorithm to video-stream analysis," *Journal of Huaqiao University (Natural Science)*, vol.25 (3), pp244-246, July, 2004.

[6] M. T. Chapman, *Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text*, Milwaukee: University of Wisconsin-Milwaukee, 1997.

[7] X. X. Niu, and Y.X. Yang, "Research on the algorithm of text steganography," *ACTA Electronica Sinica*, vol.31(3), pp.402-405, March, 2003.

[8] F. Chen, and B. Wang, "An algorithm of text information hiding based on font," *Computer Technology and Development*, vol. 16 (1), pp. 20-22, January, 2006.

[9] P. Chen, S. W. Guo, and H. L. Chen, "Color-based information hiding algorithm for text documents," *Science Technology and Engineering*, vol.7 (14), pp.3544-3546, July, 2007.

[10] L. Ou, X. M. Sun, and Y. L. Liu, "Adaptive algorithm of text information hiding based on character intensity," *Application Research of Computers*, vol. 24(5), pp.130-132, May, 2007.

[11] D. C. Han, *Text Information Hiding Algorithm Based on the Layered Structure of the Natural Language*, Changsha: Hunan Scinece and Technology University, 2008.

[12] S. W. Xu, and D. J. Xu, "New hypertext steganography method based on attribute redundancy," *China Science and Technology Information*, vol. 2007 (19), pp.111-113, 2007.

[13] P. Chen, and F. Zhang, "Research on text information hiding techniques based on statistical characteristics of

characters," *Journal of Pingdingshan Institute of Technology*, vol.16 (4), pp. 16-18, pp.26, July, 2007.

[14] C.Y. Ye, Y. S. Bi, X. S. Zhang, and J. Y. Qi,,"An algorithm of text steganography," China Information Security, vol.2005(11),pp.106-108, November, 2005

[15] J. Bai, Y. Yang., Y.H. Xu, X. X. Niu, and Y.X. Yang ,"An algorithm of text steganography," *Applications of the computer systems*, vol.2005(4), pp.32-35, April, 2005.

[16] X. X. Niu, and Y. X. Yang, "Study on the frame of information steganography and steganalysis," *Acta Electronica Sinica*, vol. 34(12A), pp.2421-2424, December, 2006.

[17] P. Chen, and L. H. Zhang, "Research on information hiding techniques based on text," *Journal of Chongqing University of Science and Technology (Natural Sciences Edition)*, vol. 9(4), pp.107-109, pp.115, December, 2007.

[18] Z. X. DAI, F. Hong, and J. Dong, "Algorithm of Text Information Hiding Based on Huffman Coding," *Computer Engineering*, vol.33(15), pp.147-147, pp.151, August, 2007.

[19] G. Luo, and X. M. Sun, "Steganalysis for stegotext based on text redundancy," *Journal on Communications*, vol. 30(6), pp.19-25, June, 2009.

[20] H. T. Liu, Z. W. Zhao, and G. L. Sheng, "Hidden markov models and its application to natural language process," *Microprocessors*, vol. 2009(3), pp.74-76, June, 2009.

[21] J. D. Yu, X. Z. Fan, and J. H. Yin, "Application of hidden markov model in natural language processing," *Computer Engineering and Design*, vol.28(22), pp.5514-5516, November, 2007.

[22] H. Wang, S. C. Wang, and J. F. Zhang, "Learning hide variables in markov network," *Mini-micro Systems*, vol. 26(3), pp.348-351, March, 2005.

[23] K. Maher. TEXTO. URL: ftp://ftp.funet.fi/pub/crypt/steganography/ texto.tar.gz, May.21, 2008.

[24] P. Grosse. TextHide. URL: http://www.compris.com/TextHide/en/, June 6, 2009.

[25] B.Scheneider, *Applied Cryptography: Protocols, algorithms, and source code in C.*, Beijing: China Machine Press, 2000.

Weihui Dai received his B.S. degree in Automation Engineering in 1987, his Msc. degree in Automobile Electronics in 1992, and his Ph.D. in Biomedical Engineering in1996, all from Zhejiang University, China. Dr. Dai worked as a post-doctor at School of Management, Fudan University from 1997 to 1999, a visiting scholar at Sloan School of Management, M.I.T from 2000 to 2001, and a visiting professor at Chonnam National University, Korea from 2001 to 2002. He is currently an Associate Professor at the Department of Information Management and Information Systems, School of Management, Fudan University, China. He has published more than 100 papers in Software Engineering, Information Management and Information Systems, Complex Adaptive System and Socioeconomic Ecology, Digital Arts and Creative Industry, etc. Dr. Dai became a member of IEEE in 2003, a senior member of China Computer Society in 2004, and a senior member of China Society of Technology Economics in 2004.

Yue Yu received her B.S. degree in Software Engineering in 2007 from Tsinghua University, China. She is current a master student at School of Management, Fudan University, China.

Yonghui Dai received his B.S. degree in Computer Science and Technology in 2002 from Zhejiang University, China, his Master degree in Software Engineering in 2006 from Fudan University, China.

Bin Deng received his Master degree in Software Engineering in 2005 from Fudan University, China.