Information Security for Secret Data Transmission over Web Services

Lwin Mar Thin, Nang Saing Moon Kham
University of Computer Studies, Yangon
lwinmarthin85@gmail.com, moonkhamucsy@gmail.com

Abstract

Information hiding technique is becoming a research hotspot in information security, and will be widely used to national security, e-commerce, e-government, copyright protection and covert communication. This paper proposes secure, efficient capacity and undetectable hiding algorithm to be used for XML document and the secret message is hidden in cover XML according to the pseudorandom number generator which makes this system more secured. The proposed steganography algorithm, which is different from all steganographic algorithms in XML and text file that uses XOR logical operation and mathematical operation to hide secret data. All other algorithms which hide information by inserting white space in XML tag, rearranging the order of XML elements etc. This paper also developed a secure system in which cryptography and steganography are used as integrated part along with newly developed enhanced security model for web services. In Steganography the secret message is hidden in (Simple Object Access Protocol) SOAP message according to the proposed hiding algorithm. In Cryptography public key (RSA) algorithm is used to encrypt the key for extracting process. This technique is empirically validated to indicate its utility and value.

1. Introduction

Since XML as well as SOAP messages are widely used for data exchange over different networks and exposed to different threats, SOAP message security become a key concern of web service organizations. Information hiding is a field of information security, and it includes methods hiding the existence of information itself, and methods for digital watermarking. These technologies have lately attracted considerable attention as solution to copyright problems and the protecting method for communication privacy [11].

Information hiding can hide secret messages in cover media, such as image, video, audio, document, and webpage [8]. The embedded messages are invisible to general observers. Information hiding has become a new research hotspot of information security and copyright protection of digital multimedia content recently [9]. For security reasoning, many different methods have been implemented and new methods are evolving every day. Cryptography, Steganography and Watermarking are well known ways of securing information but they all work under different mechanisms. Cryptography makes data unreadable by writing into secret code and it ensures confidentiality and integrity [4]. However, it makes the message suspicious enough to attract eavesdropper's attention. This vulnerability can be reduced significantly using steganography, which is a method of convert communication and information security.

Steganography is the art and science of communicating in a way which hides the existence of the communication. Although techniques for image or sounds have mainly been studied, there are few examples of research of the information hiding method on text data and XML [5]. This paper firstly proposed the information hiding algorithm for XML and also applied this algorithm to information security system for web services. This system is presented a new approach of steganography method combined with cryptography and implementation is made by four main steps.

- preprocessing step to create cover SOAP message
- random number generation
- information hiding by using proposed hiding algorithm
- encrypting the stego key

2. Related Work

Most research of steganography was using cover media such as images, videos and sound. However, steganography into the text is usually not preferred because of the difficulty in finding redundant bits in a text document [6]. Some of the methods proposed to solve the problem, such as by line shifting, words shifting, up to whitespaces manipulation into the cover text [2].

Memon et al. [1] designed four steganography techniques to ensure that the confidentiality and integrity of data is maintained in XML documents. Random characters are inserted between XML tags and their values in the first technique. Moreover, the second technique uses the procedure of changing the XML tags in a predetermined sequence. In the third technique, the order of tags is saved in the attributes before the shuffling process. The fourth technique reverses the sequence of characters. All these four methods aim to safeguard the stego XML document

against actual XML content detection rather than against hidden information detection. Therefore, the goal of these methods is totally different from our steganography goals which is undetectable and convert communication.

There is several traditional XML document information hiding algorithms. WU Jing [10] introduced an algorithm of XML information hiding through altering the dimension of some tags, this algorithm is marked as algor I. Guang-hua [3] presented an algorithm of XML information hiding through synonym substitution, this algorithm is marked as algor II. Yang Jie [11] proposed an information hiding algorithm based on equal element. This algorithm gets equal element by permutation and combination of sub-elements. This algorithm is marked as algor III.

In order to overcoming the disadvantages of imperceptibility, robustness and hiding capacity of traditional XML information hiding algorithms, a new XML information hiding algorithm is presented based on Xor logical operations. So, secret data can be embedded does not change the size of XML document and has good imperceptibility without changing the nature of tags.

3. Information Hiding

This section explains the basis of information hiding technologies as background. Information hiding is the technology to hide the secret information into a cover data, and to make the secret information invisible. Figure 1 shows the result of general model of the information hiding [7]. The model consists of three processes, hiding, transmitting, extracting.

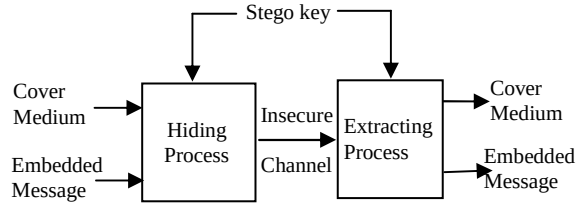

Figure 1.Information Hiding

4. The Proposed Hiding Algorithm for XML

XML are widely used for data exchange, and expected as a language of Web pages and digital contents. To develop the methods of information hiding using XML makes realize the way to establish secret communication channel using XML documents, and the ability to trace the source of unauthorized copies. Figure2 shows the structure of a simple XML document.

The proposed new hiding algorithm is recommended because of its simplicity and efficiency. It can outperform competing algorithms especially in capacity and imperceptibility. The proposed algorithm can be effectively used for embedding secret data without suspicious enough to attract eavesdropper. With the idea of using Xor logical operation technique to hide secret text information into the XML using random number generator to locate where to hide is presented in . The hiding algorithm and extracting algorithm consists of the same operations, only one operation is different – Executed divide operation on the text by 3 in the hiding and executed multiple operation on the text by 3 in the extracting. Divide operation is executed by 3 on the text to release result in the range domain of the ASCII code table at converting the text. The details and working of the proposed algorithm and pseudorandom number generator algorithm are given below.

```
<?xml version="1.0" encoding="UTF-8"?>
<workers>
   <worker>
         <name>JACK</name>
         <telephone>
            <home>66-12</home>
            <office>77-IO</office>
            <mobile>666</mobile>
         </telephone>
   </worker>
   <worker>
         <name>LI</name>
         <telephone>
            <home>11-77</home>
            <office>30-55</office>
            <mobile>888</mobile>
         </telephone>
   </worker>
```

Figure2. XML document

4.1. Random Number Generator Algorithm

Input : prime number(x_0), length of secret text information (m), a ,c
Output : (x_0, x_1,……,x_{m-1})

Process :
step1 : Initialization:input x_0, a, c, m and n←0
step2 : Compute $x_{n+1} \leftarrow (ax_n + c) \bmod m$
step3 : increase n: n ← n+1. If $n \geq m$, then go to step 4, else go to step2
step4 : End

4.2 Proposed Hiding Algorithm

INPUT : Cover XML document (CX) and Secret Message(SM)
OUTPUT : Stego XML document (SX) and Stego key (SK)
1: Input the cover XML document (CX).
2: Read all elements, attributes and value from CX.
3: Extract value (xmlValue) from this SX.
4: Use a psedorandom number generator which

initialized by x_0, m, a, c to generate (x_0,x_1,…, x_{m-1})

5: Get character value (cValue) from xmlValue according to the previous pseudorandom number sequences

6: Convert cValue to ASCII code (hexdata).

7: Foreach(byte b in hexdata)
   a. Convert the previous ASCII code(hexdata) to binary data(bin)
   b. Switch (bin.Length)
      case5: bin = "000"+ bin.
      case6: bin = "00"+ bin.
      case7: bin = "0" + bin
   c. bin2+= bin.

8: Input the Secret Message (SM) to hide.

9: Convert SM to ASCII code(hexdata).

10: Foreach(byte b in hexdata)
   a. Convert the previous ASCII code(hexdata) to binary data(bin)
   b. Switch (bin.Length)
      case5: bin = "000"+ bin.
      case6: bin = "00"+ bin.
      case7: bin = "0" + bin
   c. bin1+= bin.

11: For i from 0 to bin1.Length do the following.
   a. Split(stSplit1)&(stSplit2) 8bits from bin1 and bin2
   b. XOR(strXOR) the binary data stSplit1[i] and stSplit2[i].
   c. Reverse the strXOR and get the ASCII code of it.
   d. Divide the ASCII code by 3 → obtain the ASCII character of the quotient as first character(a)
      the remainder →(b) as a second character.
   e. Write ab in Stego key(SK)

12:Return (SK) and (SX).

4.3. Proposed Extracting Algorithm

INPUT : Stego XML document(SX) and Stego key(Sk)

OUTPUT: Secret Message(SM), Cover XML document (CX)

1: Input the cover XML document (CX).

2: Read all elements, attributes and value from CX.

3: Extract value (xmlValue) from this SX.

4: Use a psedorandom number generator which initialized by x_0, m, a, c to generate (x_0,x_1,…, x_{m-1})

5: Get character value (cValue) from xmlValue according to the previous pseudorandom number sequences

6: Convert cValue to ASCII code(hexdata).

7: Foreach(byte b in hexdata)
   a. Convert the previous ASCII code(hexdata) to binary data(bin)
   b. Switch (bin.Length)
      case5: bin = "000"+ bin.
      case6: bin = "00"+ bin.
      case7: bin = "0" + bin
   c. bin2+= bin.

8: For (i=0; i< Sk.Length; i+=2)
   a. Get the ASCII code of the character at Sk[i].
   b. newascii = (previous ASCII code[i] * 3) + the next digit (remainder) [i+1]

9: Foreach(byte b in newascii)
   a. Convert the previous newascii code to binary data(bin)
   b. Switch(bin.Length)
      case5: bin = "000"+ bin.
      case6: bin = "00"+ bin.
      case7: bin = "0" + bin.
   c. bin1+= bin.

10: Reverse the previous binary data(bin1)

11: For i from 0 to bin1.Length do the following.
   a. Split(stSplit1)&(stSplit2) 8bits from bin2 and bin1
   b. XOR(strXOR) the binary data stSplit1[i] and stSplit2[i].
   c. Convert strXOR to ASCII code
   d. Get the character of previous ASCII code
   e. Write this character value to Secret Message(SM)

12: Return (SM) and (CS).

5. The Proposed Combination System for Web Services

The design for the combining two different techniques over web services is purely based on the idea. At the sender site, the proposed system preprocesses the incoming SOAP message to create Cover SOAP message that would include both the actual and the negative data for hiding process. Then select the content values from Cover SOAP message and the pseudo random number generator is used to generate random number sequences seeded by the prime number according to the length of the selected value. And the proposed system hides the existence of the secret message in the selected value according to random sequences and generates the Stego key. And then encrypts the previous Stego key by RSA public key crypto algorithm. At the receiver site, the opposite process should run to get the back the original secret message. The proposed system designed with three modules-

For security – Security Module

For Steganography – Stego Module

For Cryptography – Crypto Module

The security module are providing make this system more secured. The process flow for the system is as follows.

5.1 Preprocessing step for Cover SOAP message

To create cover SOAP message, the secret data is removed from real input SOAP message and replace counterfeit value. So, the proposed hiding algorithm is used to hide secret data into non-secret cover data. Figure 3 shows creation of the cover SOAP Message.

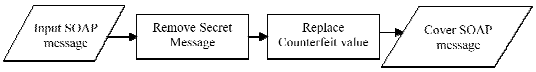

Figure3. Cover generation

5.2 Security Module

The security module provides an enhance security features to the proposed system. This module is used to generate the pseudorandom sequences for hiding process (steganography). Before the hiding process this modules work at first. Figure 4 shows the steps of security module which selects the content values of cover SOAP message and then apply pseudo random number generator to generate random number sequences. The output of this process is random number sequences to hide.

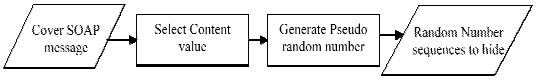

Figure4. Security module

5.3 Stego Module

This module is used to hide secret message according to random number character sequences from above generated Security Module. The hiding process takes secret message and random characters as input and produces stego key and SOAP message as output by using proposed hiding algorithm as shown in Figure5.

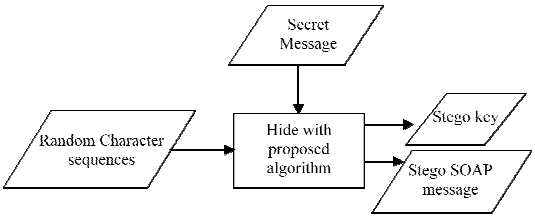

Figure5. Stego module

5.4 Crypto Module

This module is considered for encrypting the key from Stego module. Public key cryptography (RSA) is applied for encrypt the key. Figure 6 shows the process of crypto module.

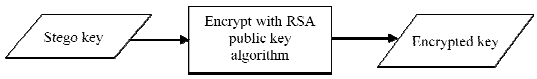

Figure6. Crypto Module

6. Overview of the proposed system architecture

The overall steps of the proposed system framework for sender site and receiver site are discussed below in Figure 7 and Figure 8.

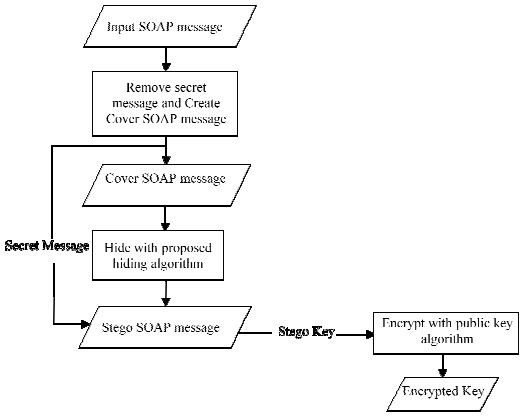

Figure7. The proposed system architecture

(at Sender Site)

The input SOAP message is a real data record including secret data. Firstly, the proposed system removes secret message from the input SOAP message to replace with counterfeit value for security purposes. After replacing, the input SOAP message is written as a cover SOAP message for hiding process. And the secret message is hidden by applying proposed hiding algorithm in cover SOAP message with the help of pseudorandom number generator which makes this system more secured. The outputs of hiding process are Stego key and Stego SOAP message. A public key crypto algorithm is used to encrypt the Stego key.

At receiver site, the proposed system firstly extracts the content values of the Output SOAP message. Then, pseudorandom number generator is used to produce random character sequences for extracting process as an input. Another input is the Stego key which gets by decrypting the Encrypted Key from Embedding process. Secret Message and Cover SOAP message are produced as an output by applying the proposed extracting algorithm. To get actual input SOAP message, we need to replace secret message with counterfeit value. The proposed system architecture for receiver site is shown in Figure 8.

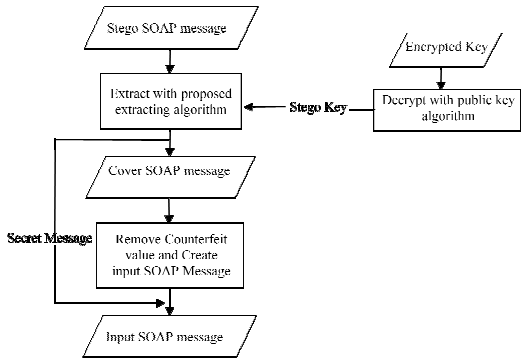

Figure8. The proposed system architecture (at receiver site)

7. Experimental Results and Analysis

There are mainly three aspects that should be taken into account when discussing the results of the proposed algorithm of steganography. They are capacity, robustness and imperceptibility. This method satisfies all security aspects, hiding capacity requirements and robustness. This section shows the results of the experiments conducted to measure the performance of the proposed algorithm and compares with some existing algorithms.

7.1. Evaluation of Steganographic Capacity

The proposed algorithm is compared with algor I, algor II and algor III that discussed in Related Work Section for hiding capacity. The XML document shown in Figure 2 is used as hiding cover-document. The embedding capacity of algor I depends on the amount of elements with content value. Suppose that the number of synonym in each group of algor II is n ($n \geq 2$). Algor III is based on the number of rearranging the order of sub-elements. In order to calculate the maximal hiding capacity of proposed algorithm, the content value of XML document is extracted. The amount of value bytes is the maximal amount of hidden bytes. Table 1 is the result of hiding capacity.

Table1. Compare on Hiding Capacity

maximal hiding capacity(bit)			
algor I	algor II	algor III	proposed algorithm
8	8	12	256

From above table, it can be seen that the maximum hiding capacity of proposed algorithm is more than other algorithms.

7.2. Evaluation of Robustness

One algorithm's robustness can appraise from the integrality of the document after experience of various attacks. The robustness of proposed algorithm is compared with other algorithms. The result is shown in Table2. Word "integral" means secret information has no error or loss, words "not integral" means secret information has error or loss.

Table2: Compare on Robustness

algorithm	Attack Type			
	Format attack	tamper tag	Delete or tamper content	Delete or tamper attribute
algor I	not integral	not integral	integral	integral
algor II	integral	not integral	integral	integral
algor III	integral	integral	not integral	integral
proposed algorithm	Integral	integral	not integral	integral

7.3. Evaluation of Imperceptibility

Secret information "101" embedded in the XML document shown in Figure2 with the use of various algorithms, the hided results of algor I, algor II, algor III and proposed algorithm are shown in Figure 9, Figure 10, Figure 11 and Figure 12 respectively.

```
<?xml version="1.0"?>
  <workers>
    <worker>
      <name>JACK </name> "1"
      <telephone>
        <home>66-12</home> "0"
        <office>77-1O </office> "1"
        <mobile>666<1mobile>
      </telephone>
    </worker>
      |
      |
  </worker >
</workers >
```

Figure9. Stego result of algor I

```
<?xml version="1.0"?>
  <workers>
    <WOker> "1"
      <name>JACK</name> "0"
      <TElephone> "1"
        <home>66-12</home>
        <office>77-IO </office>
        <mobile>666<1mobile>
      </TElephone>
    </WOker>
      |
      |
  </workers >
</workers >
```

Figure10. Stego result of algor II

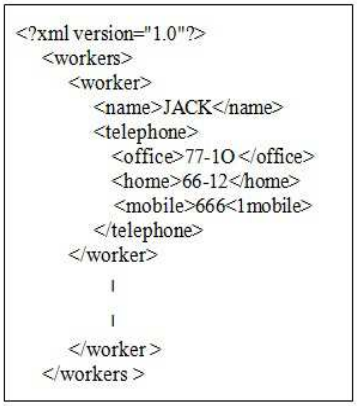

Figure11. Stego result of algor III

From Figure 9 it can be seen that algor I adds sightless blank between tags of selected element. This kind of hiding will be detected due to the change of document's size. Algor II hides information through synonym substitution. We change lowercase to capital as shown in Figure 10. When observers open the XML document by UltraEdit, he or she will detect the unusuality. In Figure11, it can be seen that algor III change the order of sub-elements.

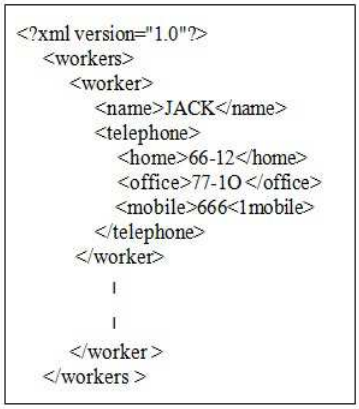

Figure12. Stego result of proposed algorithm

From Figure 12 we can see that the proposed algorithm does not change the size and content of XML document.

8. Discussion

In this paper, communication protocol-based steganography system is also developed a secure system for the real enterprise web services application areas that creates the cover SOAP message from actual SOAP message in the sender endpoint and before it is sent, hiding a secret message accordingly. The proposed data hiding method produces stego SOAP messages that have exactly the same size of the cover message, which makes it undetectable using conventional detecting methods. So, even if an eavesdropper gets both the cover and the stego SOAP message, he can't figure out any difference in the two messages. But if the attacker got the original SOAP message, he/she can easily read the secret message. This is only weakness of the proposed system because the actual SOAP message is used as a cover media for SOAP-based steganography system. Furthermore, this system was developed to overcome the limitations of using cryptographic-based techniques. The first major issue with the traditional security is their impact on the size of SOAP-messages. As a result, the proposed steganography system for web services could be a reasonable solution for transmitted secret data. As a kind of communication security, the process of surely knowing the public key of the other communicating party (on the other end of a channel) is known as Authentication. Thus, RSA public key cryptography secures the authentication portion of the communication which relies on the use of stego key encryption and it provides data confidentiality.

9. Conclusion

In the study of information security for secret data transmission using Steganography combined with Cryptography is a powerful tool which enables people to communicate without possible eavesdroppers even knowing there is a form of communication in the first place. The proposed system have included secure key scheme by using public key encryption to encrypt secret key for extracting process and extra security modules which might highly secured system. A new hiding algorithm has been proposed here for transferring real data record including secret message without altering the constraint of XML features. An intruder can break the key for Cryptographic or can find out Steganographic technique but breaking the combination of both can be nearly impossible. The proposed technique, as demonstrated by example, clearly hide information in XML document and, in such a manner that concerned receiver alone can access the information. This technique does not increase file size and provides all aspects of capacity, robustness and imperceptibility.

References

[1] A.G.Memon, S. Khawaja, and A. Shah, "Steganography: A New Horizon for Safe Communication through XML", Journal of Theoretical Information Technology, Vol. 4, N0.3, pp. 187-2020, 2008.

[2] Carro, F. Incertis, "Methods of invisibly embedding and hiding data into soft-copy text documents", U.S. Patent No. 7240209 B2 July 3rd, 2007.

[3] D. Guang-hua, L. Jia-yong, S. Ke-qiang. "Information Hiding Algorithm Based on XML", in Proceedings of the Computer Engineering, Vol. 34, No. 6, 2008, pp.l55-157.

[4] F.A. P. Petitocolas, R.J. Anderson, M.G. Kuhn. "Information Hiding- A Survey", in Proceedings of the IEEE, Vol.87, No.7, 1999, pp. 1062-1078.

[5] F. Johnson, S. Jajodia, "Steganography: Seeing the Unseen", in Proceedings of the IEEE Computer, Vol. 31, No.2, 1998, pp. 1079-1107

[6] K. Bailey, K. Curran, J.Condell, "Evaluation of Pixel-based Steganography and Stegodetection Methods", The Imaging Science Journal, Vol. 52, No.3, 2004, pp. 131-150.

[7] Matsumoto, Inoue, Kitabayashi, "An information hiding method for Standard MIDI File," Syposium on Cryptography and Information Security, SCIS2000-C03, Jan.2000.

[8] M. Laheen, S. XingMing, "Techniques with Statistics for Web page Watermarking", World Academy of Science, Engineering and Technology, 2005, NSFC No. 60373062, pp.14-17

[9]Singh, Hitesh, P. Kumar Singh and K. Saroha, "A Survey on Text Based Steganography", in Proceedings of the 3rd National Conference Computing For Nation Development, February 26– 27, 2009.

[10]W. Jing, W. Shu-wen, "A Hiding Method Based on eXtensible Markup Language (XML)", China Safty Science Journal, Vol.15, No.12, 2005, pp.78-80.

[11]Y. Jie, "Algorithm of XML Document Information Hiding Based on Equal element", Department of Communication Engineering, Nanjing Institute of Technology, Proceedings of the IEEE, Vol.3, 9-11 July 2010, pp. 250-253