A Proposal on Information Hiding Methods using XML

Shingo INOUE[*1]   Kyoko MAKINO[*1]   Ichiro MURASE[*1]   Osamu TAKIZAWA[*2]
Tsutomu MATSUMOTO[*3]   Hiroshi NAKAGAWA[*4]

*1 Mitsubishi Research Institute, Inc.   *2 Communications Research Laboratory
*3 Yokohama National University   *4 The University of Tokyo

Abstract

Since networks progressed remarkably, XML documents created as digital contents or used for the data exchange between companies are becoming general. Resolving copyright problem and realization of communication privacy become more important in the scene using XML documents.

As technics resolving copyright problems in digital contents and communication privacy problems, information hiding methods have attracted considerable attention these days. Although technics for images or sounds have mainly been studied, there are few examples of research of the information hiding method on text data.

In this paper, we propose the information hiding methods using XML.

1. Introduction

With the rapid development of Internet technologies, the amount of information sent and received electronically is increasing greatly. As the technology of transmitting information on network in secure, the importance of information security came to be recognized widely.

Information hiding is a field of information security, and it includes methods creating covert channel where specification of transceiver is difficult, 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.

It is considered that the number of applications will increase which use not only plain-text but formatted data written in markup languages such as SGML or HTML. Nowadays, XML is known as the universal format for structured documents and data, and used as the basic technnology for exchanging information on the Web. The importance of security on XML is growing more.

Although, compared with the information hiding methods intended for images and sounds, there are few methods of hiding information into text. Unfortunately, there is almost no study on the methods for hiding information in structured documents.

In this paper, we introduce some information hiding methods that embed secret messages into XML files.

2. Information Hiding

In this section we explain the basis of information hiding technologies as backgroud.

2.1. Hiding the existence of information

Information hiding is the tecnhology to embed the secret information into a cover data, and to make the secret informaton invisible. Figure 1 shows the general model of the information hiding [2]. The model consists of three processes, embedding, transmitting, extracting.
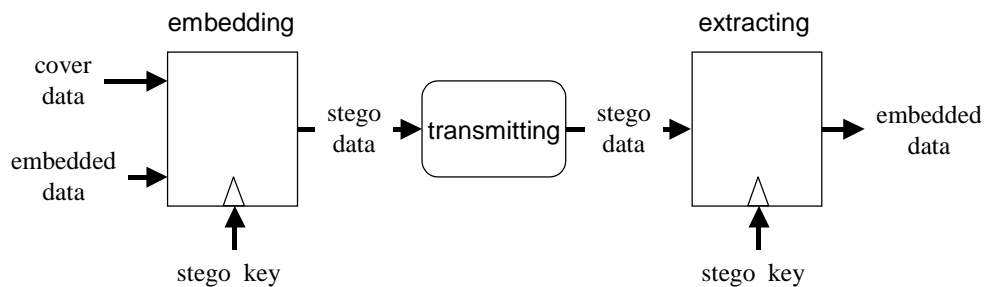

Figure 1: Information Hiding

2.2. Text-based information hiding

Text-based information hiding methods are the technics of information hiding using text as cover data and stego data.

Most of methods for hiding data into text process texts as image essantially. Those methods have a characteristic that the copy of printted matter has the same secret data as original. There are some methods hiding data into text data using character codes. For example, there is the method appending whitespace to the end of lines [7], and the method changing the start position of new lines [5].

Another approach is to handle texts not as paper image but sequence of characters [3]. The method intends that linguistic expressions are altered to hide secrets while the meaning of the text preserved.

For Structured documents, there is a study on PDF and PostScript [4]. In the method, changes to embed secret data are done not to change the final output.

Methods to guard privacy of communication are also developed [1][8]. These methods aim to make communications invisible to mechanical filtering programs. To make mass reasonable texts, these methods use generation of meaningless texts.

3. Model of Information Hiding using XML

XML are used widely 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.

Following is the proposal of the methods of information hiding using XML. In this section, we propose general model of information hiding on XML. The concrete example of the technics will be presented at the next section.

3.1. Supposed XML Components

Document's content, strucuture, and style are handled separately in XML. The texts marked up the content using tags are called XML documents, and the strucuture of the XML documents is defined in DTD. The style is described in stylesheets like CSS or XSL.

Figure 2 shows the supposed model of XML components. It shows the example case in which XML are processed by browser program.
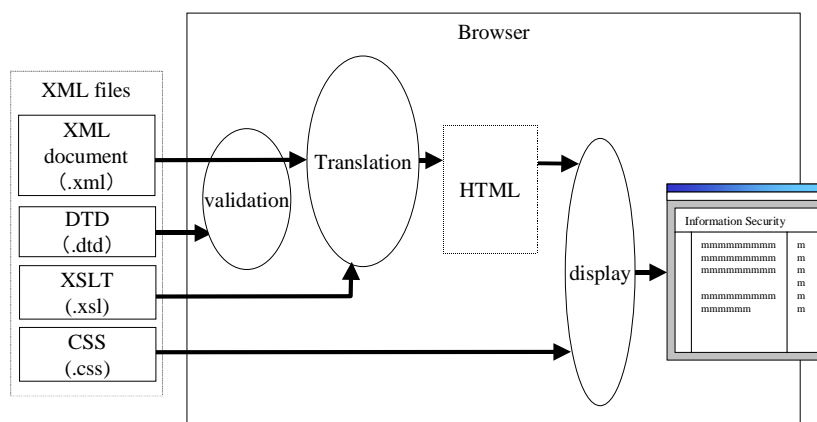

Figure 2: Supposed XML components

Information hiding methods using XML satisfies the following conditions.

> ➤ Cover data is either or all of the XML document, DTD, XSL, CSS.
> ➤ The cover data is changed to embed data while the meaning is preserved in stego data.

Embedder may be the maker of the cover file, or may be the person who gets files and alters them preserving the meaning of original.

The following is the example model of embedding and extracting.

3.2. Tool for data exchange

XML is strongly presented as the tool for data exchange. Generally, fixed and open DTD defined is used when XML data is exchanged. So it is considered that there can be limited oppotunity to embed data into DTD. Figure 3 shows the model of embedding and extracting in XML data exchange.
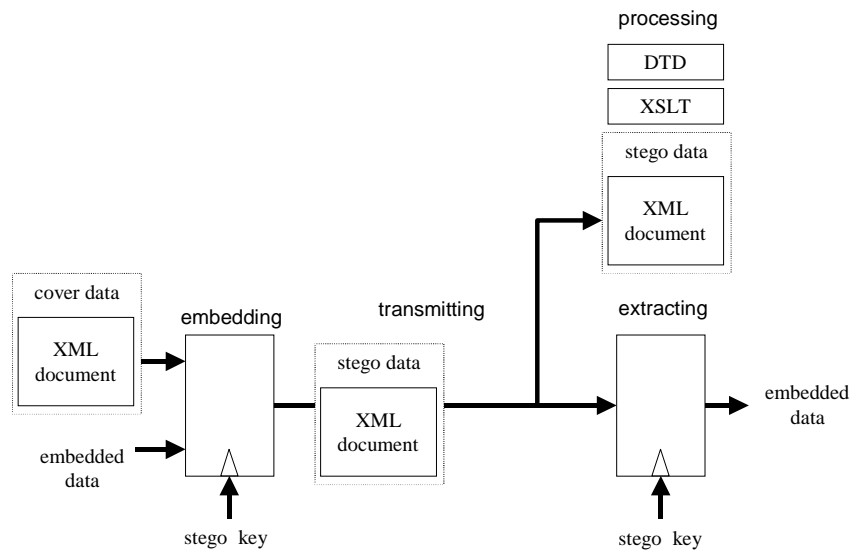
Figure 3: Information hiding in XML data exchange

3.3. XML web pages

Although still HTML has been widely used to describe Web pages, XML pages can be created and browsed by using major browser. In those cases, creaters of the pages can define their own DTD, XSL, and CSS. Therefore, embedder has the chance to use all components of XML as cover data. Figure 4 presents the model of embedding and extracting in XML web pages.
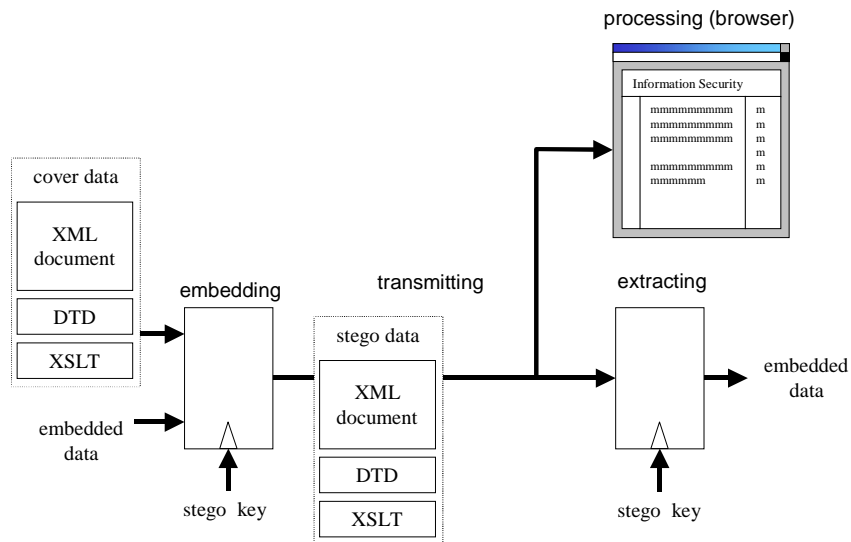

Figure 4: Information hiding in XML web pages

4. Technics

In this section, we explain the some technics to embed the secret data into XML document in detail.

4.1. Representation of empty elements

According to the W3C recommendation, the representation of an empty element is either a start-tag immediately followed by an end-tag, or an empty-element tag [6]. By switching these two forms, we can embed the data preserving all meaning of original document. The following Example 1 shows a method of information hiding by altering img element. This example method can embed one bit of data per an end-tag of empty elements.

Example 1.

```
stego key:
    <img></img>      … 0
    <img/>           … 1
stego data:
    <img src="foo1.jpg"></img>
    <img src="foo2.jpg"/>
    <img src="foo3.jpg"/>
    <img src="foo4.jpg"/>
    <img src="foo5.jpg"></img>
embedded data:
    01110
```

4.2. White spaces in tags

Representation of a tag is eather including some white spaces before close blacket, or no white space [6]. By inserting or deleting spaces, we can embed the data preserving all meaning of original document. The following is an example method of information hiding by inserting or deleting a space. This example method can embed one bit of data per a tag.

Example 2.

```
stego key:
    <tag>, </tag>, or <tag/>        … 0
    <tag >, </tag >, or <tag />    … 1
stego data:
    <user ><name>Alice</name ><id >01</id></user>
    <user><name >Bob</name><id>02</id ></user >
embedded data:
    101100 010011
```

4.3. Appering order of the elements

We can embed secret data in XML documents by exchanging of the appearing order of elements. Following is the example of this method. One bit of data can be hidden in the documents per an exchange of two elements.

Example 3.
```
stego key:
    <user><name>NAME</name><id>ID</id></user>       … 0
    <user><id>ID</id><name>NAME</name></user>       … 1
stego data:
    <user><name>Alice</name><id>01</id></user>
    <user><id>02</id><name>Bob</name></user>
embedded data:
    01
```

There are some conditions to apply this method.

> ➤ No dependance of the order of the elements in the application
>
> ➤ No reorder of the elements before extracting the secret data

4.4. Appering order of the attributes

Secret data can be embedded in XML documents by exchanging of the appearing order of attributes in the element. Following is the example of this method. One bit of data can be hidden per an exchange of the order of attributes.

Example 4.
```
stego key:
    <event month="MONTH" date="DATE">EVENT</event> … 0
    <event date="DATE" month="MONTH">EVENT</event> … 1
stego data:
    <event month="JUL" date="4">Independence day</event>
    <event date="24" month="DEC">Christmas</event>
embedded data:
    01
```

4.5. Elements containing other element

This method uses two or more elements that can contain each other. Example 5 shows the case that exchanging inner-tags and outer-tags are available. One bit of data can be hidden per an exchange.

Example 5.

```
stego key:

    <favorite><fruit>SOMETHING</fruit></favorite> … 0

    <fruit><favorite>SOMETHING</favorite></fruit> … 1

stego data:

    <fruit><favorite>orange</favorite></fruit>

    <favorite><fruit>apple</fruit></favorite>

embedded data:

    10
```

5. Conclusion

In this paper, the model of information hiding using XML has been shown. In the applications of XML data exchnage or XML web pages, the secret data can be embedded in XML modules not changing the content of original. We proposed some methods for hiding data into XML document. These methods can be applied to exiting XML documents easily.

Furture work will include investigating the application and efficiency of these methods, and the methods using DTD and XSL in adddition to XML document.

Acknowledgement

References

[1] Chapman, Davida, "Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text," Financial Cryptography, First International Conference, FC'97, pp.335-345, Feb. 1997.

[2] Matsumoto, Inoue, Kitabayashi, "An information hiding method for Standard MIDI File," Symposium on Cryptography and Information Security, SCIS2000-C03, Jan. 2000 (in Japanese).

[3] Nakagawa, Sanpei, Matsumoto, et al., "MeaningPreserving Information Hiding –Japanese Text Case—," The Transactions of The Information Processing Society of Japan (IPSJ), Vol. 42, No.9, pp.-, Sep.2001. (Accepted) (in Japanese).

[4] Shibuya, Kaji, Kasami, "Digital watermarking for PostScript and PDF documents," Symposium on Cryptography and Information Security, SCIS98-9.2.E, Jan. 1998 (in Japanese).

[5] Takizawa, Yamamura, Nakagawa, et al., "A Proposal of Steganography on Plain Text and XML," The 7th Annual Meeting of the Association for Natural Language Processing, C2-4, pp.135-138, Mar. 2001 (in Japanese).

[6] "Extensible Markup Language (XML) 1.0 (Second Edition)," http://www.w3.org/TR/REC-xml, Feb. 2001.

[7] "The SNOW Home Page," http://www.darkside.com.au/snow/

[8] "spammimic - hide a message in spam," http://www.spammimic.com/index.shtml.