Enhancing Authorization of XML File Using Watermark

[1]Arun Thomas K, [2]Mrs. J.Jeysasudha
[1]Student, [2]Assistant Professor
SRM University, Chennai

***Abstract*–Huge amount of data on web of data in cloud are transferred between business applications in the form of XML files. So these XML files need authorization. Security is the main threats in cloud computing and web. Moreover, organizations usually manage XML files via a common XML repository. XML copyright protection and source tracking have become strong requirements for collaborative environments. XML-specific fingerprinting mechanisms have been proposed, inspired by similar work on relational data. However their robustness is impaired by the fact that an XML file can undergo a set of updates that change both the file structure and content. The proposed system has a watermark embedding unit, watermark extraction unit, pattern manager unit, character to special characters translator unit and special characters to character translator unit. This mechanism increases the robustness of XML copyright protection.**

***IndexTerms* - XML,watermarking,authenticity.**

I. INTRODUCTION

XML has an important role in transferring data between business applications. Data exchanged through network are of different level of sensitivity. In current scenario, XML files need authenticity. Some researchers started thinking about the watermarking of XML file. But all these faced several challenges. The first issue is to find a suitable location for the watermark; another one is defining a suitable algorithm for watermark embedding. Another crucial issue is defining a way to evaluate the robustness of the algorithm with respect to alterations and other kind of attacks. Some early attempts were promising but research is still at a preliminary stage.Contribution of [1] targeted embedding technique to choose watermark location according to given user profiles and use fuzzy queries to re-construct data corresponding to locator in XML file in order to tolerate modifications.

In our proposal we use a character to special character translator module to enhance the robustness of the existing system [1]. Each user has unique pattern and character set for transformation. This pattern and character set are generated randomly.

II. EXISTING SYSTEM

Copyright protection has become a major requirement for many applications and digital watermarking is a widely used technique to achieve this goal. In paper [1] contains solutions of three main issues.A first part, which will help choosing the locators, a second issue concerning the watermarking process to use and the last issue is about the use of fuzzy queries to extract the locators

The locators' selection is a big issue in watermarking XML files. Many criterions could be used to choose the locators. One choice is the relevance of the tag for the usability of the data [2]. This requirement means that the removal of this tag from the file will break its meaning. Another possibility is also to define a metric to evaluate the frequency of each tag in the document and then choose the less or most frequent tag, the less frequent because the probability that a malicious user modify this tag is low, as it has not enough importance in the file. And the more frequent tag because the probability that this kind of tags will be completely deleted is very low. Here it use tags that are relevant for the user profiles.

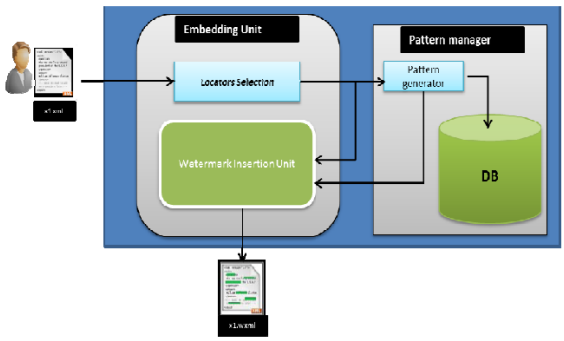

Figure 1. Process of watermarking an XML file

The watermarking process is described in figure 1. It uses a unique pattern for each user. This pattern is randomly generated. A function is used to insert watermark in the selected tags and a reverse function is used to extract the watermark from the tags. Each bit is represented by an invisible character in the locator text. It uses space character to represent "1". Then it insert all the

positions of the pattern in each occurrence of the locator. The redundancy of the watermark increases the robustness. The advantage is that watermark can be recovered from any watermarked tags. The system is robust to reorganization of XML file.

This system consists of three main components, embedding unit, locator selection manager and extraction unit. Figure 2 describes watermark extraction process [1].

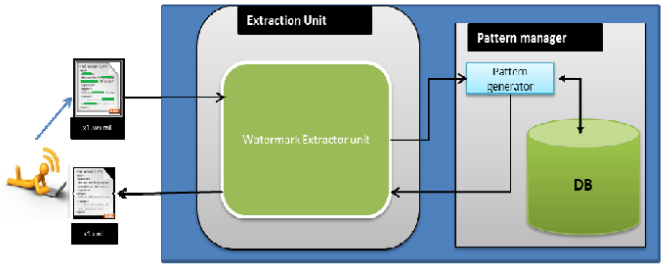

Figure 2. The watermark extraction Process

III. ENSURING AUTHENTICITY OF XML

The watermarking process of XML files includes various units watermark embedding unit, watermark extraction unit, pattern manager unit, character to special character translator, special character to character translator. The following figure shows the watermark embedding in XML file

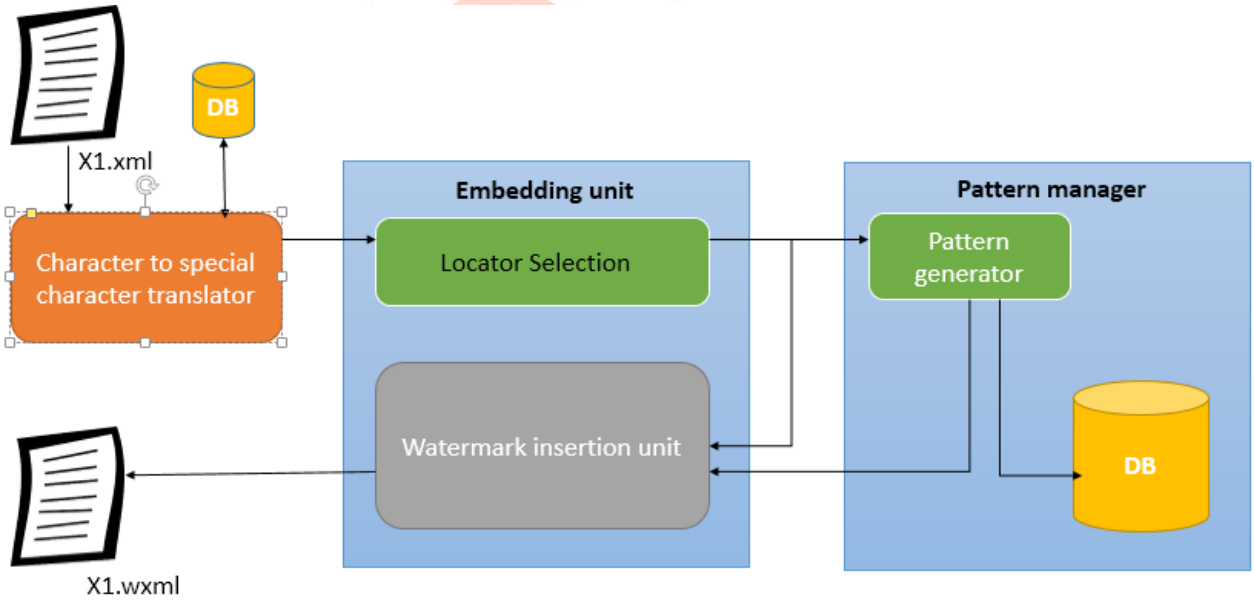

Figure 3: Process of watermarking an XML file

The watermarking process in the figure 1 is also an important issue, mainly in algorithms and information embed. In this model each user has a unique pattern. This pattern is randomly selected sequence of bits. A function is used to embed watermark on the selected locators. Each bit represent an invisible character in the locator text.' *' character represent '0',it can also select randomly. Initially the locator selected manually. All the contents in the locator tag undergoes for character to special character transformation. It will increases the robustness of watermarking. The important components of this model are embedding unit, pattern manager and character to special character translator. Character set for each user select randomly and it is stored in database. Pattern and its user information also are stored in database which is under the control of pattern manager. In character to special character translator each character in the character set has a special character value and this special character is used in the contents of locators instead of the corresponding character. For example character in character set has the following special character values

A -> %
D->&
I -> @
P -> $
T -> #

The watermark extraction unit also similar but in reverse order. The following figure 4 shows the watermark extraction process. The important components of this process are extraction unit, pattern manager and special character to character translator. The special characters are replaced by corresponding characters.

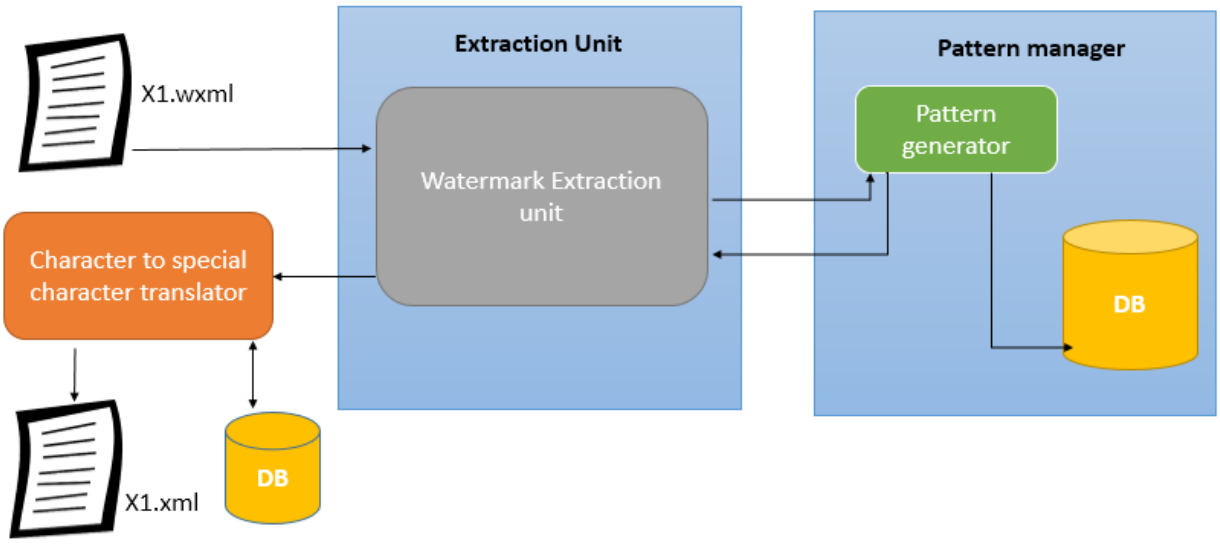

Figure 4 watermark extraction unit

IV. ANALYSIS

The proposed system is assessed with a sample XML file. Robustness of the proposed system can be analyzed with the existing system in terms of different kinds of attack. Here the sample file is first watermarked with the existing method and then it is watermarked with the proposed system. From these results, it is easy to conclude the robustness of proposed system.

In the existing system [1], each user has unique pattern. This pattern is embedded in the selected locators of the XML file. The XML file contains information about books such as the author-name, the authorization of publication-id, the publisher-name. The author-name node is selected as locator and user's unique pattern is embedded in these locators. For example, here we can consider the user pattern as 111001. The watermarked file is shown in the figure 6.

```
<?xml version="1.0" encoding="UTF-8"?>
<bib>
<book year="1994">
        <title-name>TCP/IP Illustrated</title-name>
        <author-name>Stevens W.</author-name>
        <authorization-id>10ab7</authorization-id>
        <publisher-name>Addison-Wesley</publisher-name>
</book>
<book year="1992">
        <title-name>Advanced Programming in the Unix environment</title-name>
        <author-name>Stevens W.</author-name>
        <authorization-id>13cd7</authorization-id>
        <publisher-name>Addison-Wesley</publisher-name>
</book>
<book year="2000">
        <title-name>Data on the Web</title-name>
        <author-name>Abiteboul Serge</author-name>
        <author-name>Buneman Peter</author-name>
        <author-name>Suciu Dan</author-name>
        <authorization-id>19aqd8 </authorization-id>
        <publisher-name>Morgan Kaufmann Publishers</publisher-name>
</book>
<book year="1999">
        <title-name>The Economics of Technology and Content for Digital
        TV</title-name>
        <author-name>Gerbarg Darcy</author-name>
        <authorization-name>11abw</authorization-id>
        <publisher-name>Kluwer Academic Publishers</publisher-name>
</book>
</bib>
```

Figure 5 sample XML file

In the watermark embedding, '0' is indicated by space but in the following figure these 0s replaced by '*' for more viewable.

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<bib>
<book year="1994">
        <title-name>TCP/IP Illustrated</title-name>
        <author-name>Ste**vens** W.</author-name>
        <authorization-id>10ab7</authorization-id>
        <publisher-name>Addison-Wesley</publisher-name>
</book>
<book year="1992">
        <title-name>Advanced Programming in the Unix
        environment</title-name>
        <author-name>Ste**vens** W.</author-name>
        <authorization-id>13cd7</authorization-id>
        <publisher-name>Addison-Wesley</publisher>
</book>
<book year="2000">
        <title-name>Data on the Web</title-name>
        <author-name>Abi**tebo**ul S**erge</author-name>
        <author-name>Bun**eman** Pet**er</author-name>
        <author-name>Suc**iu D**an</author-name>
        <authorization-id>19aqd8 </authorization-id>
        <publisher-name>Morgan Kaufmann Publishers</publisher-name>
</book>
<book year="1999">
        <title-name>The Economics of Technology and Content for
        Digital TV</title-name>
        <author-name>Ger**barg** Dar**cy</author-name>
        <authorization-id>11abw</authorization-name>
        <publisher-name>Kluwer Academic Publishers</publisher-name>
</book>
</bib>
```

Figure 6 watermarked XML file [1]

In the proposed system, a new module is added with the existing system, character to vowel translator. In this module, it translates a random set of characters with special characters. So it will be more robust to the monitoring attacks. The following figure shows the watermarked XML file as per the proposed system.

```
<?xml version="1.0" encoding="UTF-8"?>
<bib>
<book year="1994">
        <title-name>TCP/IP Illustrated</title-name>
        <author-name>St&**v&ns** W.</author-name>
        <authorization-id>10ab7</authorization-id>
        <publisher>Addison-Wesley</publisher-name>
</book>
<book year="1992">
        <title-name>Advanced Programming in the Unix
        environment</title-name>
        <author-name>St&**v&ns** W.</author-name>
        <authorization-id>13cd7</authorization-id>
        <publisher-name>Addison-Wesley</publisher-name>
</book>
<book year="2000">
        <title-name>Data on the Web</title-name>
        <author-name>%b@**t&b$**#l S**erge</author-name>
        <author-name>B#n**&m%n** P&t**er</author-name>
        <author-name>S#c**@# D**%n</author-name>
        <authorization-id>19aqd8 </authorization-id>
        <publisher-name>Morgan Kaufmann Publishers</publisher-name>
</book>
<book year="1999">
        <title-name>The Economics of Technology and Content for
        Digital TV</title-name>
        <author-name>G&r**b%rg** Dar**cy</author-name>
        <authorization-id>11abw</authorization-id>
        <publisher-name>Kluwer Academic Publishers</publisher-name>
</book>
</bib>
```

Figure 7 watermarked XML file

Here random set of characters are a, e, i, o and u and these characters arereplaced by a->%, e->&, i->@, o->$, u->#. The character set and its value is randomly selected and it will be unique for each users. This translation of character set with special character enhances the robustness of the system. From the figure 6 and 7, it is clear that the proposed system is more robust than the existing system.

V. CONCLUSION

The proposed system is more robust than the existing system. It protects XML file from addition of new node, deletion of new node and XML file re-organization. Watermark can be extracted under these attacks. This system also protects the system from file monitoring. Attackers get confused because of the character to special character translation module. So this system increases the robustness of the system.

VI. REFERENCES

[1] Tchokpon Romaric/Ernesto Damiani, Nadia Bennani, "Robust XML Watermarking Using Fuzzy Queries", IEEE 36th International Conference on Computer Software and Applications Workshops 2012

[2] R. Agrawal and J. Kierman. "Watermarking relational databases". In Proceedings of the 28th International Conference on Very large Databases (VLDB), 2002.

[3] Xuan Zhou, HweeHwa Pang and Kian-Lee Tan. "Querying query based watermarking for XML data". In Proceeding ASIACCS '07 the 2nd ACM symposium on Information, computer and communications security ACM New York, NY, USA ©2007.

[4] Xuan Zhou, HweeHwa Pang, Kian-Lee Tan, Dhruv Mangla. "WmXML: A System for Watermarking XML Data". In Proceedings of the 31st VLDB Conference, Trondheim, Norway, 2005.Y.

[5] Shingo, Kyoko, Ichiro, Osamu. "A Proposal on Information Hiding Methods using XML". Mitsubishi Research Institute, Communication Research Laboratory, Yokohama National University and the University of Tokyo, unpublished

[6] Nighat Mir, Sayed Afaq Hussain. "Web Page Watermarking: XML files using Synonyms and Acronyms". In Proceedings of Conference WASET 2011, World Academy of Science, Engeniering and Technology, january 25-27, 2011.