Protection of Semantic Organized Data. Encryption of RDF Graph

Galina Bogdanova[1], Todor Todorov[1,2], Nikolay Noev[1]

[1] Institute of mathematics and informatics, BAS
[2] St. Cyril and St. Methodius University of Veliko Tarnovo
galina@math.bas.bg, todor@math.bas.bg, nickey@math.bas.bg

Abstract. The problems of protection of semantic organized data and encryption of RDF graph are investigated. Technologies used in the construction of a semantic network are presented. XML-encryption and standardization of semantically structured content for XML signature are described. We perform encryption of XML data associated with the semantic representation of data for a kind of percussion instruments.

Keywords: Semantic data, XML, RDF graph, encryption, XML signature.

1 Introduction

With the development of digital technology, most of the information published in the public domain becomes available for fast, easy and high quality copying. This fact gives rise to the problem of protecting information from unauthorized distribution. Research in this area is related to the demand for effective methods for embedding marks, serial numbers, etc. in the original data proving their origin.

Cryptography is the practice and study of techniques for secure communication in the presence of third parties (called adversaries). More generally, it is about constructing and analyzing protocols that overcome the influence of adversaries and which are related to various aspects in information security such as data confidentiality, data integrity, authentication, and non-repudiation. Modern cryptography intersects the disciplines of mathematics, computer science, and electrical engineering.

2 Semantic Web Technologies

The concept of Semantic Web was introduced by Tim Berners-Lee [3], director and creator of the World Wide Web Consortium (W3C) [9], and defines the next generation Internet, which describes the upgrade of conventional network space and expansion of information posted. It is used as a term to describe such information handled by computers and understandable by machines and people. In that sense the computers directly or indirectly interpret and process the meaning and purpose of information resources on the network. In the Semantic Web talk about knowledge bases, such

knowledge will be accessible to computers not only cultivated as databases, but "understandable", which means that the machines will be able to make an intelligent interpretation of knowledge.

Technologies used in the construction of a semantic network is designed to provide a formal description of concepts, terms and relationships within a given knowledge domain.

XML (eXtensible Markup Language) is a meta-language of extended markup of text information, which is the Internet standard for publishing structured textual information. XML is a technology rules for marking text by tags.

RDF (Resource Description Framework) is an XML-based language for presenting information describing Internet resources, providing semantics of a textual content.

Following simplified example that gives short description of the resource http://.../bells/bells.xml#bell1AN. The example uses the URI (Uniform Resource Identifier) to identify the resource. The label <name>, which here is called property, describes the name of the resource. The value of the property is the content between marks <name> and </name> in this case "Bell № 01".

This scheme defines RDF-data model, which consists of three object types: "subjects, predicates and objects". In the example, the subject is http://.../bells/bells.xml#bell1AN (URI, Internet address) predicates are <name>, <location>, <city> and <materials>, and objects are: bell № 01; Cathedral № 01; Sofia; alloy of lead, silver and copper.

```
<RDF>   <description
about="http://.../bells/bells.xml#bell1AN">
    <name> bell № 01 </name>
    <location> Cathedral № 01 </location>
    <city> Sofia </city>
    <materials>alloy of lead, silver and cop-
per</materials>
        </description> </RDF >
```

According to the specification of RDF each predicate defines a binary relation between resources and atomic values that are provided by the definitions of the primitive data types in XML. Object determines what value will accept binary relation. This is called graph-based data model. Its main concepts are resources, properties and statements. Statement is triple: resource-property-value.

URI (Uniform Resource Identifier) is a symbolic identifier that uniquely named and specified a network resource. Identifier can be an Internet address such as: http://.../bells/bells.xml#bell1AN - called URL (Uniform Resource Locator). Identifier can also be the name of a resource in a particular network or area, clearly indicating the resource. Then it is called URN (Uniform Resource Name). For example URN can be “urn: issn 1314-4006” (International Conference on Digital Preservation and Presentation of Cultural and Scientific Heritage).

Namespace is a specific area of information on specific objects with common structure attributes. Namespace is used to specify certain objects to which group or area concerned.

For example, the following code specifies that this object has a bell-described properties and attributes of a particular place and do not give any explanation for the principal impact music device.

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:bell="http://.../bells/bells.xml#">
  <rdf:Description rdf:about=
"http://.../bells/bells.xml#bell1AN">
    <bell:name> bell № 01 </bell:name>
    <bell:location> Cathedral № 01 </bell:location>
  <bell:city> Sofia </bell:city>
  <bell:materials>alloy of lead, silver and cop-
per</bell:materials>
  </rdf:Description>
... </rdf:RDF>
```

3 Encryption of RDF Graph

For encryption of RDF graph we mean electronic signing of text elements or external links contained in semantic structured textual description. The result of electronic encryption is presented as part of the encrypted XML document directly contains coded data or references to them.

3.1 XML-encryption

XML encryption [10] as the process of encryption and decryption of context is determined by using the specified XML syntax and algorithms. The syntax of the electronic signature in the XML document is defined by the standard developed by the W3C: XML-Signature Syntax and Processing (XMLDSIG), which defines the elements and the type of electronic signature embedded in the document content.

The structure of the Signature element in the XML document is as follows:

```
<Signature>
  <SignedInfo>
    <CanonicalizationMethod /><SignatureMethod />
    <Reference><Transforms />
              <DigestMethod /><DigestValue />
    </Reference>
    <Reference /> etc.
 </SignedInfo>
 <SignatureValue /><KeyInfo /><Object />
</Signature>
```

Where <SignedInfo> element refers to the signed data and specifies what algorithms are used. Elements <CanonicalizationMethod /> and <SignatureMethod /> used by <SignatureValue />, to determine the method of signing and algorithms used before signing. In <SignatureValue /> elements are already contained coded results. Elements <KeyInfo /> contain key that can validate the data and is not binding. One or more <Reference> elements may be present, specifying the methods of transformation of resources before they signed. <Object /> Element contains the signed data if it is covering the signature (Enveloping signature).

To validate the signed XML document using a procedure called Core validation. In that procedure, first check all <Reference> elements for their transformations. Then the signature is validated by the methods described in <SignedInfo>.

3.2 Standardization of Semantically Structured Content for XML Signature

Semantic knowledge is inherently a collection of statements and links to other resources. Very rarely all resources and claims to uniquely described in a text document. Therefore, for signing the individual data must meet certain requirements, such as canonization of the content and use of means of unambiguously defining separate statements.

Canonization of content is mandatory if present electronic signature. For example, void spaces not syntactically significant, the element <elem> is on a par with < elem >, but the signing will give different encrypted result. Also in the transmission network or validating XML documents, different systems change characters in the document, like the redundancy of space attributes, etc.

Another major thing is to determine fragment of semantic content to be encrypted or signed. These fragment must be at one place, be uniquely identified using namespaces method and be well formed as single simple statement (message). To do that, first define what is the minimum "standalone" fragment of an RDF model [4]. An RDF statement involves a name if it has that name as subject or object [5]. An RDF graph involves a name, if any of its statements involves that name. Given an RDF statement s, the Minimum Self-contained Graph (MSG) containing that statement, written MSG(s), is the set of RDF statements comprised of the following: The statement in question; Recursively, for all the blank nodes involved by statements included in the description so far, the MSG of all the statements involving such blank nodes.

This definition recursively build the MSG from a particular starting statement. In [8] show that an RDF model has a unique decomposition in MSG s. The MSG definition and properties say that it is possible to sign a MSG attaching the signature information to a single, arbitrary triple composing it. Along with the signature, an indication of the public key to use for verification might be provided. This indication is itself covered by the signing procedure. By "attach" we mean by using a verification procedure.

Example code of signed MSG is shown next, where bold code is original MSG (fragment of an RDF description):

```
<rdf:RDF
```

```
xmlns:bell="http://.../bells/bells.xml#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description rdf:about=
"http://.../bells/bells.xml#bell1AN">
    <bell:bell>
      <bell:name> bell № 01 </bell:name>
      <bell:location> Cathedral № 01 </bell:location>
      <bell:city> Sofia </bell:city>
      <bell:materials>alloy of lead, silver and cop-
per</bell:materials>
    </bell:bell>
  </rdf:Description>
  <rdf:Description rdf:nodeID="bell1AN">
    <rdf:predicate rdf:resource=
"http://.../bells/bells.xml#bell1AN"/>
    <bell:PGPCertificate>
http://.../bells/cert/12345678.asc </bell:PGPCertificate>
    <bell:Base64SigValue>MAq… …B2w</bell:Base64SigValue>
    <rdf:subject
rdf:resource="http://.../bells/bells.xml"/>
    <rdf:object> bell № 01 </rdf:object>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-
rdf-syntax-ns#Statement"/>
  </rdf:Description>
</rdf:RDF>
```

3.3 XML- cryptography Using .NET Framework

XML-cryptography (XML encryption) as the process of encryption and decryption of digital content using XML has XML-specific syntax and algorithms. XML-cryptography provides a standardized means for encrypting structured data and presents the results in XML. XML-cryptography allows you to encrypt any data, whether it is a XML-document, its elements or external data (not necessarily in the format of XML), referenced in the document. The result is represented as an encrypted encryption element in XML, which either directly contains encrypted data or references to them. In both cases, the resulting representation can be used by applications that use the powerful cross-platform XML technology for data access.

Using .NET Framework 4.5 [7] and System.Security.Cryptography.Xml classes we perform encryption of XML data associated with the semantic representation of data for bells.

In XML encryption we replace individual XML elements with <EncryptedData> element that contains the encrypted XML data. The element also contains subelements <EncryptedData> that include information on the keys and the processes used in the encryption. In this type of encryption allows the document to contain multiple encrypted elements.

In the tests we encrypt XML elements with two keys. We generate a pair of RSA public / private key. Then is created a separate session key by using the Advanced Encryption Standard (AES) algorithm, also called Rijndael algorithm. This key is used to encrypt the XML document and then uses RSA public key to encrypt the AES session key. Finally, the encrypted AES key and encrypted XML data is stored in an XML document in a new element <EncryptedData>.

Decrypting XML - with RSA private key is done with the session key, and then it is used to decrypt the data in the XML document.

4 Conclusion

Semantic and encryption technologies are used to complete tasks of protection of semantic organized data. In our development is incorporated signing of semantic description. It is important to verify the copyright against unauthorized interference and change the content. We perform encryption [1, 2, 6] of XML data associated with the semantic representation of ontological data for percussion instruments.

References

1. Bogdanova G., Todorov T., Noev N., Singing individual fragments of an RDF graph of unique Bulgarian bells, ACCT`2010, Twelfth international workshop, pp. 47-52, Academgorodok, Novosibirsk, Russia, 2010, ISBN 978-5-86134-174-5
2. Bogdanova G., Todorov T., Noev N., Digital Repository of Information and Knowledge - Fund "BellKnow", First International Conference “Digital Preservation and Presentation of Cultural and Scientific Heritage” - DiPP`11, pp. 91-98, ISSN: 1314-4006, Veliko Tarnovo, Bulgaria, 2011
3. Berners-Lee T., Fischetti M., Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its inventor., 1999
4. Carroll J., Signing RDF graphs, HP technical report, 2003
5. Carroll J., C. Bizer, P. Hayes and P. Stickler, Named Graphs, Provenance and Trust, HP technical report, 2004
6. Noev N., Organization and Security of the Audio and Video Archive for Unique Bulgarian Bells, Mathematica Balkanica, NewSeries Vol. 24, Fasc.3-4, pp. 285-291, 2010 ISSN 0205-3217
7. Thorsteinson P., Gnana Arun Ganesh G, .NET Security and Cryptography, Prentice Hall PTR, 2003
8. Tummarello G., Ch. Morbidoni, P. Puliti and F. Piazza, Signing individual fragments of an RDF graph, in Proc. International World Wide Web Conference, 2005
9. The World Wide Web Consortium (W3C), http://www.w3.org
10. XML Signature WG, http://www.w3.org/Signature/