Adding Security Information in XML Documents

Ana I. González-Tablas
Carlos III University of Madrid,
Computer Science Department, Leganés
aigonzal@inf.uc3m.es

Elena Castro
Carlos III University of Madrid,
Computer Science Department, Leganés
ecastro@inf.uc3m.es

Arturo Ribagorda
Carlos III University of Madrid,
Computer Science Department, Leganés
arturo@inf.uc3m.es

Manuel Velasco
Carlos III University of Madrid,
Computer Science Department, Leganés
velasco@inf.uc3m.es

Abstract: *XML's popularity in the last few years has made this mark-up language a de facto standard for the web data interchange. DTD's (or Schemas) definition associated with XML documents introduces data modelling in XML's world, allowing the specification of a hierarchy of concepts or elements that constitute the XML document. Taking into account that the purpose of these data models is the highly structured information exchange among several systems, it is required to incorporate security mechanisms that allow a secure interchange. The World Wide Web Consortium (W3C) is working in the recommendations of several XML security standards. Between them, we emphasize in the XML-Signature Syntax and Processing, which allows the insertion and information processing of authentication and digital signature. Once the XML security standards have been approved as recommendations, the following step will be to include them completely or just certain parts in future or new versions of the DTD's or existing Schemas, but at present many DTD's exists that do not consider these security components within their definition.*

This is the case of the NewsML DTD, standard for the press news electronic interchange. The XML security standards are characterized by high flexibility and extensibility, because of that it is necessary to make an exhaustive study of the domain where it is intended to be applied and define a specific application upon the domain DTD or Schema.

What we propose in this paper is a way to include information of authentication and digital signature in the NewsML DTD. In order to indicate a possible application, we carry on a joint study of XML-Signature Syntax and Processing and NewsML, analysing in what elements and how authentication and digital signature might be included.

Keywords: *Authentication, digital signature, DTD, NewsML, XML, XMLDSig.*

1 INTRODUCTION

XML's popularity in the last few years has made this mark-up language a de facto standard for the web data interchange. One of the main characteristics of this language is the separation between the content and the representation of documents, which allows to select or to reformat data. The thriving evolution of XML will lead to a lot of information written in this language, hand written or automatically generated via web and/or data electronic interchange software.

DTD's definition associated with XML documents introduces data modelling in XML's world. DTDs allow to specify a hierarchy of concepts or elements that constitute the XML document. Taking into account that the purpose of these data models is the highly structured information exchange among several systems, it is required to incorporate security mechanisms that allow a secure interchange. From this point of view, the W3C, representative organization in the standarization of web resources, is working in the recommendations of several XML security standards. Between them, we emphasize in the XML-Signature Syntax and Processing standard which allows to insert and process information of authentication and digital signature. Once the XML security standards have been approved as recommendations, the following step will be to include them completely or just certain parts in future or new versions of the DTD's or existing Schemas. However at present, many DTD's exists that do not consider these security components within their definition. This is the case of the NewsML DTD, standard for the press news electronic interchange.

The NewsML DTD is a specification of the IPTC (International Press Telecommunications Council) [11] to mark-up news in such a way that context could be added to the documents. It is a specific XML vocabulary, that is, a data class. It controls the whole news lifecycle and is independent of the media and the format used to create and edit the news [17].

NewsML consists of a set of interrelated elements that represent the different news items, allowing several representations of the same element. Although the norm does not include any security component within its specification, it points to the advantages it will obtain its integration if there was a standard. One of its requirements (R800) expresses the need of authentication and digital signature in some of the NewsML elements [12]; furthermore, in section 5.14 this need is justified and it is anticipated the use of the XML digital signature standard [6] with that purpose.

In order to understand why it's so important to add authentication information to NewsML DTD we expose then an example that illustrates the fatal consequences of not including it.

Nowadays world moves around information, and this is one of the main basis of business. The so called "fourth power" does and breaks business, contracts, governs, whole countries follow its tendencies. Information makes business. It is that the reason that makes fundamental to assure its reliability and confidence. Information community cannot provide false or tampered information because of its influence on society, and in its credibility. So, it is crucial to know who from information is received, because the credibility that it is granted with depends mainly on the source of the information.

In our example publisher B maintains a trusted relation with content provider A, as each news that B receives from A is granted with high confidence and reliability. If a malicious party C, enemy and competitor of B, sends to B fraudulent news attempting to be A, with the objective of causing discredit and suspicious in B, without doubt, B will trust the reliability of the received news based in that "it comes from A" causing B's misfortune. To prevent these situations authentication information assuring who is the source of messages and that they have not been tampered has to be added to news communicated electronically.

The XMLDSig standard [6] defines a XML syntax and the processing rules to create, represent and validate digital signatures by a XMLDSig application. This way, XMLDSig provides integrity, message authentication and/or signer authentication for data of any type, included in the XML document that includes the signature or in any other place.

The XML security standards are characterized by a high flexibility and extensibility, because of that it is not recommendable applying them directly into a domain. It is necessary to make an exhaustive study of the problem and define a specific application of these upon the domain.

What we propose in this paper is a way to include information of authentication and digital signature in the specific NewsML DTD. In order to indicate a possible application, we carry on a joint study of XML-Signature Syntax and Processing and NewsML, analysing in what elements and how authentication and digital signature might be included.

This work is intended to establish the first features of a framework where security information could be added to particular DTDs or XML Schemas in a rational manner and making use of the XML security standards that have been developed. Also, other of our principal aims is to study the problems and state the lessons learned in the integration of a particular data model – NewsML DTD- with one of the high flexible XML security standards –XMLDSig-.

After the introduction done in section 1, it is presented in section 2 the work relative to our research, it is, the standarization efforts done recently in the area of XML security and related research proposals. After describing some preliminary concepts (section 3), section 4 undertakes the study of NewsML

DTD and section 5 describes XMLDSig main features. Then, authors analyse the scenario of NewsML to select in which elements and how authentication information might be added using XMLDSig in section 6. Also within this section, several future steps that follow this work are presented. Finally, in section 7 , conclusions are explained.

2 BACKGROUND

Although information security in Internet has been widely addressed and several good largely adopted solutions exist, new perspectives to security must be considered. The Web is evolving and new problems and needs emerge. Due to this, security technologies are pushed to join XML. At this point, it may be asked whether it is worthy "reinventing the wheel". Several advantages derive from this union, the most important of them are, first, the fully integration with XML and the Semantic Web, that it is more a necessity than an advantage, and second, the addressing of fine grained security.

By mid 1999 the World Wide Web Consortium (W3C)[1] started to address the standarization of the adequate mechanisms that will provide for implementing security services to/by XML applications. One of the first approaches to XML security standarization was the charter of the join W3C and IETF[2] (Internet Engineering Task Force) XML Digital Signature (XML-DSig) Working Group (WG). Its mission is to develop an XML compliant syntax used for representing the signature of Web resources and portions of protocol messages (anything referenced by a URI) and procedures for computing and verifying such signatures. Recently (1Q2002), they have delivered the *XML Digital Signature Syntax and Processing* [6] W3C Recommendation and IETF Draft Standard.

Several standarization efforts have followed. XML-Digital Signature W3C/IETF WG was followed by the beginning of 2001 with the charter of the XML-Encryption W3C WG, with the objective of developing a process and a XML compliant syntax to encrypt and decrypt digital content (XML documents and portions thereof). This is specified in [7]. Also under W3C coordination, the XML Key Management WG started its activity at mid 2001. Its mission is "to develop a specification of XML application/protocol that allows a simple client to obtain key information (values, certificates, management or trust data) from a web service", the work is based on the *XML Key Management Specification* [9].

In addition of W3C, other organizations are working in XML security standarization. The Organization for the Advancement of Structured Information Standards (OASIS)[3] is one of them. This organization coordinates several Technical Committees (TC) for XML security standardization, most representative of them are the XML-based Security Services TC and the eXtensible Access Control Markup Language TC. The mission of the first one [21] is to define an XML framework for exchanging authentication and authorization information. And the second one is expected to have as its main product a core XML schema, XACML [19], for representing authorization and entitlement policies. This two Technical Committees are working closely.

The European Telecommunications Standards Institute (ETSI)[4] is developing standarization work in the area of electronic signatures and infrastructures. One of its Technical Specifications is the *XML Advanced Electronic Signature* [8], based in XML-DSig and its main aim is to define a XML format for advanced electronic signatures as defined in the European Directive and with long term validity.

Other XML security standarization works are the ones conducted by the XML Common Biometric Format [22] and the Rights Language [20] OASIS Technical Committees, and the W3C Platform for Privacy Preferences (P3P) Project [23].

An excellent survey of security in XML and its standarization, published by mid 2001 and written by A. Selkirk within BT Technologies, can be found in [25] and [26].

By the other hand, several research approaches to XML security and its applications have been developed recently within the scientific community. In [29] there is an updated and complete collection of links to the research, papers and news articles carried out and published about XML security. First proposals to add security information to SGML structured documents are found in

[1] http://www.w3.org/
[2] http://www.ietf.org/
[3] http://www.oasis-open.org
[4] http://www.etsi.org/

SDML [14], a W3C Note where is defined a markup language to sign SGML documents as part of the Electronic Check Project. One of its cited future works is the integration with the emerging XML. Other related work proposed by YH. Chu et al. by mid 1998 and described in [3], was a method of adding extensions to PICS 1.1 labels with the purpose of signing them. This proposals have been put aside looking for solutions integrated with XML, instead with SGML or HTML tags.

Based in this previous works, and already focused in XML documents, R. Anderson and JH. Lee [1] propose by 1999 a markup language and tool named Jikzi for trusted publishing and customisable security policies management. With Jikzi the security policy models together with the authentication and integrity mechanisms are revisited with the new perspective offered by XML.

Most of the XML security research since then has been made in the access control security area. It is represented by the following proposals of, first, Bertino et al., with the Author-X Java-based access control system [2]; second, the complete fine-grained access control system of Damiani et al. [4]; the Kudo et Hada third approach based in provisional authorization [16] and, finally, the XSLT-based proposal of Gabillon et Bruno [10]. Some of them have fed the work carried in [19].

Authentication and digital signatures within XML have been also addressed by other authors. Related to the IETF/W3C XMLDSig standard, Karlinger [13] does a very good prime description of its main features for non XML experts. And Scheibelhofer [24] studies in depth the "What you see is what you sign" topic, already pointed in the security considerations section at XMLDSig, and proposes a method for displaying XML documents in a trustworthy manner.

Other approaches for authenticating XML documents have been presented recently by Devanbu et al. in [5]. Devanbu et al. focus in authenticating the completeness of answers to queries selected by content allowing *untrusted* servers –no need of an on-line private key- to certify them.

Our intention is to apply XMLDSig to NewsML in order to add fine grained authentication information. We can found similarities in the specifications of [7], [8], [9], [19] and [21], which base somehow in XMLDSig. But our work approaches more to Anderson's [1] or Devanbu's [5], as what we aim is to add security information to a particular data domain model, this is NewsML DTD, applying XMLDSig. These point of view has not been considered yet. Furthermore, [1] and [5] focus in authenticating published information, the first one, and the second in authenticating queries over this information, a different scenario from what we get with NewsML. They consider semi-static structures of long-lived data, medical catalogs and books the first one, and general databases the second, considering both trees of hashes. Opposite, we treat mainly with a continuous stream of news not necessarily related. Besides, we improve these approaches adopting a standard, and obtaining the considerable advantages derived from this.

3 PRELIMINARY CONCEPTS

3.1 XML AND DTDS

XML is a markup language for describing semi-structured information [28]. An XML document is composed of a sequence of nested elements, each delimited either by a pair of start and end tags. In order to classificate and to structure XML documents into several domains it is needed to create schemas that allow the generation of documents collection. The DTDs (Document Type Definition) are the mechanism by which the various domains of the application can be represented.

A DTD is a XML schema that models a data class for a specific purpose. XML documents can be classified into two categories: well-formed and valid. An XML document is well-formed if it obeys the syntax of XML. A well-formed document is valid if it conforms to a proper Document Type Definition (DTD). A DTD is a file (external, included directly in the XML document, or both) which contains a formal definition of a particular type of XML documents.

Due to the restrictions that the DTDs impose on the documents that use them, other type of models are being developed as XML Schema. Nevertheless, in spite of the considerable advantages of the Schemas opposite to the DTDs, these ones have existed for a long time and continue being widely used nowadays; furthermore, in essence the role of DTDs or XML Schema is the same, this is to define a grammar for XML documents.

A DTD purpose is to define legal construction blocks to a XML document or class of XML documents, by including it as part of the document (internal DTD) or by referencing it as an external resource (external DTD).

The reason of using DTDs is the need of interchanging information in a coherent manner, because by using a common DTD sender and receiver might verify that the exchanged data are valid.

Intuitively, each DTD is a schema, and XML documents valid according to that DTD are instances of that schema. A DTD is represented as a labelled tree containing a node for each element, attribute, and value associated with fixed attributes in the DTD. Each XML document is described by a tree with a node for each element, attribute, and value in the document, and with an arc between each element and each of its sub-elements / attributes / values and between each attribute and each of its value(s).

Any DTD models a particular class of documents, it is that the reason because several organizations (W3C, IPTC,…) focus its efforts in the definition of standard DTDs which model several application domains. Nevertheless, any individual, organization or enterprise may define its own DTD's that consider its requirements.

3.2 AUTHENTICATION AND DIGITAL SIGNATURE

Data/messages authentication [27] is defined as the procedure to verify that received messages come from the source they assert to come and that they have not been altered. Digital signature is an authentication mechanism that usually also provides non repudiation. Both mechanisms occur in two levels. At the first one, it has to be generated an authenticator object using a function, this authenticator is a value that will allow to the authentication of the message. In the second level, the function and the authenticator object are used to authenticate the received message.

Three methods may be used to generate the authenticator, these are (1) message encryption, (2) message authentication code (MAC) and (3) hash function. With the first method, message encryption, both symmetric and asymmetric encryption may be considered, and the authenticator is provided by the ciphertext of the entire message. With message authentication code method, a public function and a secret key are used to generate a fixed-length value known as checksum or MAC. Last method, hash function, is a variation on MAC. Symmetric or asymmetric encryption are combined with hash functions to provide authentication. One of the most widespread authentication schema that uses hash functions is the following one. Both parties A and B share the knowledge of a secret key K. A generates an authenticator appending K to the message, and calculating its hash value. This schema is described in [15].

The most widespread authentication by means of digital signature is based in hash functions and asymmetric cryptography. The authenticator, that in this case is the digital signature, is the encryption of the hash of the message using the private key. The digital signature is attached to the message and provides strong authentication of the sender, integrity and non repudiation. However, it is more complex and consumes more computational resources.

4 NEWSML

A NewsML document is a XML document that must be valid to the NewsML DTD. In must be also developed by the use of multiple files referenced through entities references [28] or by the use of pointers inside the NewsML document.

As any DTD, NewsML is structured like a tree, in which the root is the selfsame NewsML document. All NewsML document must hold one *NewsEnvelope* element and one or more *NewsItem* elements. Also it may contain one or more *TopicSet* elements that are containers of *Topics* related to the own NewsML document or to any other news item included by reference. The NewsML DTD may also contain one *Catalog* element which identifies and locates the default vocabularies and indicates where are used certain topics in the NewsML document (figure 1).

```
<!ELEMENT NewsML (Catalog? , TopicSet* , (NewsEnvelope , NewsItem+ ))>
<!ATTLIST NewsML %localid; >
```

```
<?xml version="1.0"?>
<!DOCTYPE NewsML PUBLIC "urn:newsml:iptc.org:20001006:NewsMLv1.0:1"
"http://www.iptc.org/NewsML/DTD/NewsMLv1.0.dtd">

<NewsML>
<Catalog>
...
</ Catalog >
<TopicSet>
...
</TopicSet>
<NewsEnvelope>
...
</NewsEnvelope>
<NewsItem>
...
</NewsItem>
<NewsItem>
...
</NewsItem>
</NewsML>
```

Figure 1: A fragment of NewsML DTD

The *NewsEnvelope* element provides information about how a concrete document is being used within a business workflow or a contractual relationship between the news provider and the receiver, therefore there is a clear need of assuring it. In the same direction it should be added some kind of element that provides authentication in the *NewsItem* element, because of it contains a managed news collection which represents a point of view, at a given time, of some events. Subelements *Identification* and *NewsManagement* provide identification information and manageability.

5 XML Digital Signature Syntax and Processing XMLDSIG

The digital signature standard XMLDSig [6] specifies the syntax and processing rules of a digital signature XML compliant. XML digital signatures provide integrity, messages and/or signer authentication of any kind of data, included in the XML document that contains the signature or located somewhere. XML digital signature applies to an arbitrary digital content or data objects by an indirection. First the hash of data objects is calculated, and this value together with other information is placed in an XML element of *Signature* type. This element is authenticated by using a MAC or a digital signature algorithm. XML digital signature relates with its data objects by means of URIs.

Three kind of digital signature are defined: enveloping (signature is parent element), enveloped (signature is child), and detached (signature and data objects are in sibling elements or in external resources).

```
<Signature ID?>
   <SignedInfo>
      (CanonicalizationMethod)
      (SignatureMethod)
      (<Reference (URI=)? >
           (Transforms)?
           (DigestMethod)
           (DigestValue)
      </Reference>)+
   </SignedInfo>
   <SignatureValue>
   (KeyInfo)?
   (Object (ID=)?)*
</Signature>
```

Figure 2: XMLDSig Signature structure

The root element of an XML digital signature is the *Signature* element and its structure can be seen in figure 2, where *SignedInfo* contains the actual signed data, and *SignatureValue* the value of the digital signature. Optional element *KeyInfo* allows the receiver to obtain the information needed to know which key will validate the signature, and optional element *Object* may contain any kind of data being used to include additional information about the signature or its generation, as well as to construct enveloping signatures. Within *Reference* element a list of several transformations (*Transforms*) may be defined. These transformations will be applied before calculating the hash value and specify how the signer obtained the data object in order to sign it.

The XMLDSig specification provides a high flexible mechanism of digital signature and several important recommendations for developers using it are already pointed in [6]. It considers specifically the security models obtained, the arbitrary algorithms that may be executed and the transformations.

6 ADDING AUTHENTICATION INFORMATION IN NEWSML DTD WITH XMLDSIG

6.1 NEWSML SCENARIO

NewsML is a standard thought for news interchange, without considering explicitly the creation, edition, storage or internal news management processes, such as it is stated in [12]. The actors that we can distinguish are the following: news provider (Content Provider), news consumer (Publisher), news broker, and author of news objects. For the delivery of news, there would be usually some contractual long term relationships (1) -in figure 3- between news consumers and at least a news provider. The news flow is characterized for being of high intensity, continuous, and, in its majority, unidirectional in the provider to consumer direction. Brokers or news agents would act as consumers for the providers (2) and as providers for the customers that require their services (3), probably establishing in both cases long term contractual relationships.

Another kind of relationship is the one between the authors of the news objects and the news providers with which they work (4). This relationship is out of scope of this work, because we only address the interchange of news, as NewsML standard does, but, even under this consideration, we want to point that there is a relationship between news object's author and the object which may be subject to authentication by means of a digital signature, in order to assure the authenticity, integrity and non-repudiation of their creations.

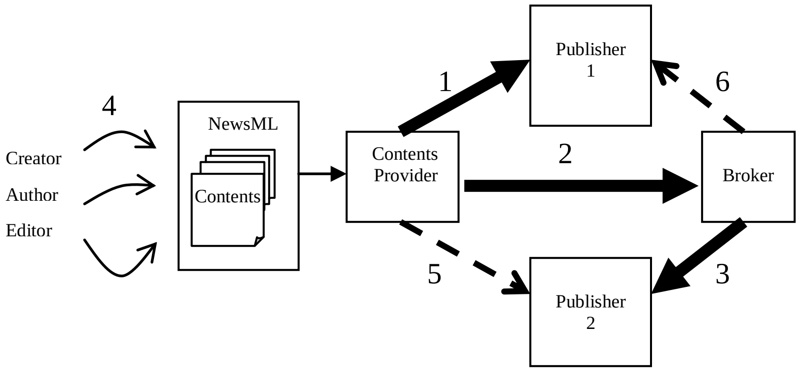


Figure3: NewsML Scenario

Sometimes there would be occasionally relationships without previous contractual relationship between the provider and consumer parties (5), as well as between broker and consumer (6), being less likely an occasional relationship between provider and broker. But once contact has been established, these occasional relations are similar to the previous long term ones. So, the solution developed for relation (1) might be applied with little changes to the other ones except (4).

On the other hand, two NewsML use cases can be distinguished. In the first one, NewsML will represent an independent and autonomous news. Second use case will consider related news collections, in which the complexity, consequence of the probable inclusion and/or reference of a great amount of related and reused *NewsItem* and *NewsComponent* within a *NewsML* element, increases notably. This use case will be very adequate if NewsML elements should be stored and this issue addressed by the specification at some moment. For the present, in this work authors will focus in the first use case, as the proposal might be easily extended to news collections.

6.2 NEWSML AUTHENTICATION EXTENSION PROPOSAL FOR NEWSML

Considering an independent NewsML element, it seems clear that the NewsML elements that should include an authentication token component are:

- *NewsML* because this is the root element and the element that will be transmitted. *NewsEnvelope*, within *NewsML*, since the moment that this element describes the kind of agreement between the news provider and the company in charge of its publication.
- *NewsItem* as this is the main component of a *NewsML* news and one of the reusable ones. More specifically the *NewsComponent* and *NewsManagement* subelements should be considered, because the first includes the metadata required to describe everything relative to author rights, reproduction and use of the news, and the second everything related to news management and its interrelations with other news. It has to be considered also the subelement *Identification*, as it contains metadata to identify the *NewsItem* element.
- *NewsComponent* is also reusable, so it might contain authentication information by its own.

Consequently, our proposal is to add to each one of these selected elements another subelement of *Signature* type. The NewsML DTD will be extended in the following way (figure 4):

```
<!DOCTYPE NewsML PUBLIC …
<!ENTITY % dsig SYSTEM “xmldsig.dtd”>
%dsig;
]>
…
<!ELEMENT NewsML (…, Signature*) >
…
<!ELEMENT NewsItem (…, Signature?) >
…
<!ELEMENT NewsComponent (…, Signature?) >
```

Figure 4: Extension of the NewsML DTD

The authentication information of each *NewsML*, *NewsItem* or *NewsComponent* elements will consider also each of its descendants by means of an enveloped digital signature. In this way, the authentication information will maintain valid despite the authenticated element is exported or referenced externally. To manage this, unique identifiers like the ones provided by NewsML DTD must be used.

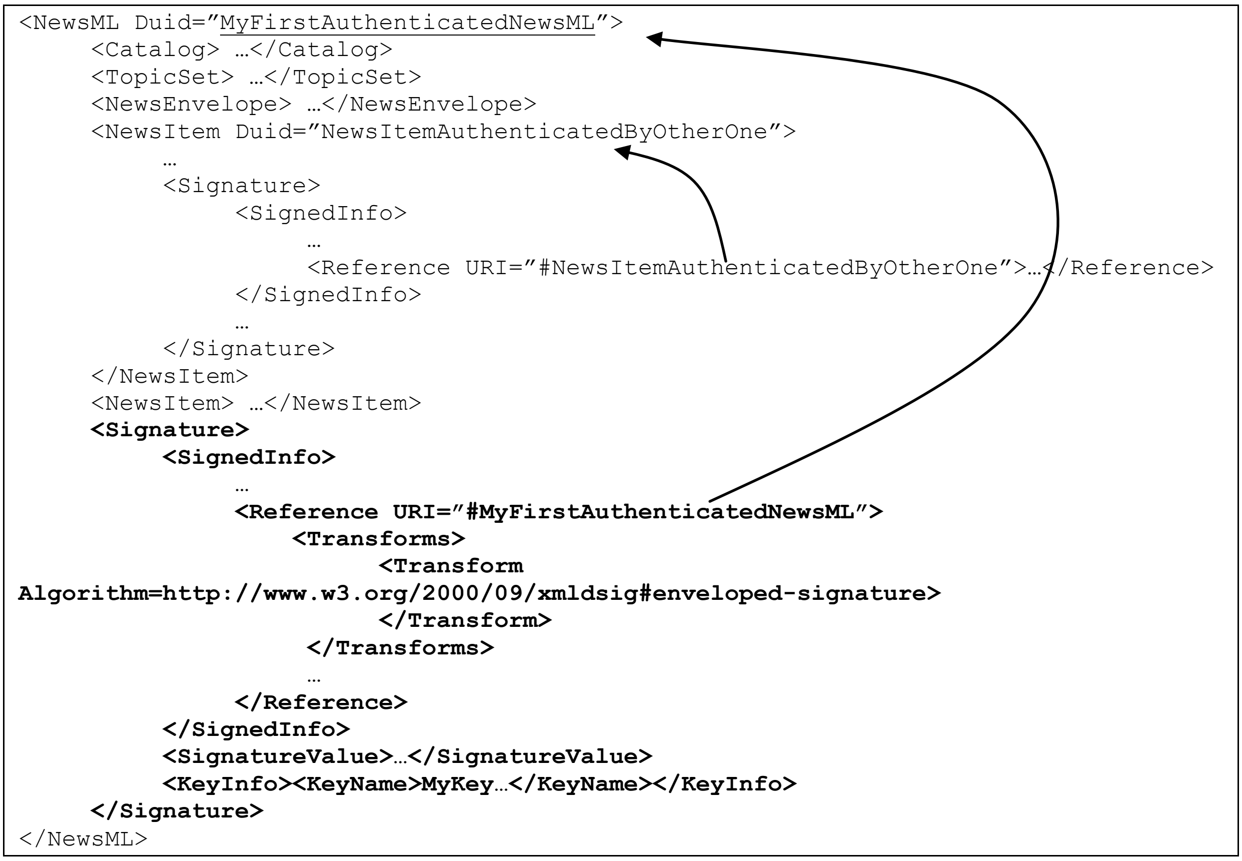
To increase the number of elements to authenticate, besides those mentioned before wards, or not limiting them would provoke greater complexity and increase of processing time. So we think that the proposed ones offer the required authentication granularity.

An example (figure 5) of an authenticated NewsML including an imported NewsItem that has been already authenticated, illustrates the proposal.

On the other hand, digital signature of any *NewsML* subelement might optionally be placed as a child of *NewsML* root element by means of a detached signature. That is why the *Signature* elements cardinality differ being 0 or more in *NewsML* element and 0 or 1 in its child susceptible of containing authentication information.

Since the news traffic is characterized for being very dense, it is an indispensable requisite that the time and computing cost associated to the authentication process be minimal, this feature influences in the selection of the required level of authentication. As there is a contractual long-term relationship between the provider and the consumer, it is proper to use authentication through a keyed hash function [15] implicating that both parties share a secret key S. This key S would be communicated using a secure channel and would have periodic renovation. This authentication is fast and does not imply a big computational cost. Authentication and integrity of the message would be assured.

Digital signature can not be considered the best default option to authenticate an element due to its high cost. However, in some cases it could be required a stronger authentication that the one offered by a keyed hash algorithm, so proper digital signature should be allowed. The possibility of not including authentication information should be allowed too. XMLDSig allows authenticating by using both keyed hash and digital signature algorithms.



```
<NewsML Duid="MyFirstAuthenticatedNewsML">
     <Catalog> …</Catalog>
     <TopicSet> …</TopicSet>
     <NewsEnvelope> …</NewsEnvelope>
     <NewsItem Duid="NewsItemAuthenticatedByOtherOne">
          …
          <Signature>
               <SignedInfo>
                    …
                    <Reference URI="#NewsItemAuthenticatedByOtherOne">…</Reference>
               </SignedInfo>
               …
          </Signature>
     </NewsItem>
     <NewsItem> …</NewsItem>
     <Signature>
          <SignedInfo>
               …
               <Reference URI="#MyFirstAuthenticatedNewsML">
                   <Transforms>
                         <Transform
Algorithm=http://www.w3.org/2000/09/xmldsig#enveloped-signature>
                         </Transform>
                    </Transforms>
                    …
               </Reference>
          </SignedInfo>
          <SignatureValue>…</SignatureValue>
          <KeyInfo><KeyName>MyKey…</KeyName></KeyInfo>
     </Signature>
</NewsML>
```

Figure 5: Example of an authenticated NewsML element.

If message authentication is done by means of a secret key S shared between the entities (e.g. departments within each party), it must be communicated previously through a secure channel. In case of occasional relationships we can not consider both parties sharing a secret key beforehand. This key could be negotiated with some of the existing negotiation algorithms or by means of asymmetric cryptography.

The proposed extension offers several interesting advantages as it considers the authentication information transport and its simple of implementing, besides applying a standard.

6.3 OPEN ISSUES

Some open issues remain. These are listed next:

- Although the authentication of news collections is an extension of present work, it must be properly addressed. Also, collection authentication might consider being compatible with its storage, and in this case, the authentication might be done by means of tree hashes as in [5]. But previously, the storage of NewsML elements has to be addressed by competent entities.
- Strong authentication of authoring rights is clearly derived from our proposal by including a *Signature* element within *ContentItem* element and some other data, but in a similar way than in news storage, it should be contemplated in the NewsML DTD first.
- It should be considered how to address the authentication of the presentation information of News Objects, as it is suggested in the XMLDSig specification [6], for example including a reference to style sheets or to similar information within the *Signature*.
- Including authorization information, as who can access what and what is allowed to do with it, might be addressed by integrating the XACML specification [19] with NewsML, in a similar manner that in this work has been done with XMLDSig.
- The XMLEnc specification [7] might be used to encrypt some NewsML objects or to protect the communication of the used keys in the authentication.

- Finally, the work done in OASIS Right Language TC [20] and within XACML TC [19] should be used to represent properly and respectively the *CopyRight* and *UsageRights* elements, instead of expressing them in natural language.

7 CONCLUSIONS

In this paper, the authors have done a deep study on how to solve the security needs regarding authentication and digital signature of the NewsML standard. To implement the proposed solution, is used the digital signature standard XML-Signature Syntax and Processing [6].

It has been done a specific proposal that addresses in which elements and how authentication information may be added into news items exchanged between the content provider and the publisher. This schema is easily extensible to other types of relationships included in the news interchange scenario, and in news collections too.

This work is the start point of a framework that finally allows the integration of all the XML security standards in the NewsML DTD. A list of such future works appear in the previous section. Moreover, this work is one of the first in such open research, so may be used as an example of how to add authentication information into other domains.

We have learned that the integration of XML security standards into a well-founded DTD in not so immediate. A deep study of the domain security needs must be done in order to apply the standards suitably.

REFERENCES

[1] Anderson, R. J.; Lee, JH. (1999): *Jikzi – A New Framework for Security Policy, Trusted Publishing and Electronic Commerce*. Available from http://www.cl.cam.ac.uk/ftp/users/rja14/jikzi.pdf

[2] Bertino, E.; Castano, S.; Ferrari, E. (2001): *Securing XML Documents with Author-X*. IEEE Internet Computing, Vol. 5, No. 3, May/June 2001.

[3] Chu, YH.; DesAutels, P.; LaMacchia, B.; Lipp, P. (1998): *PICS Signed Labels (DSig) 1.0 Specification*. W3C Recommendation, May 1998. http://www.w3.org/TR/REC-DSig-label/

[4] Damiani, E.; De Capitani, S.; Paraboschi, S.; Samarati, P. (2002): *A Fine-Grained Access Control System for XML Documents*. ACM Transactions on Information and System Security (TISSEC), vol. 5, n. 2, May 2002, pp. 169-202.

[5] Devanbu, P.; Gertz, M.; Kwong, A.; Martel, C.; Nuckolls, G. (2001): *Flexible authentication of XML documents*, Proceedings of 8th ACM Conference on Computer and Communications Security (CCS'01), Nov. 2001, Philadelphia, Pensylvania, USA.

[6] Eastlake, D.; Reagle, J.; Solo, D. (2002): *(Extensible Markup Language) XML-Signature Syntax and Processing RFC 3275*, March 2002.

[7] Eastlake, D.; Reagle, J. (2002): *XML Encryption Syntax and Processing*, W3C Candidate Recommendation, March 2002.

[8] ETSI SEC ESI (Electronic Signatures and Infrastructures) (2002): *XML Advanced Electronic Signatures (XAdES)*. ETSI TS 101 903 V1.1.1 Technical Specification, February 2002.

[9] Ford, W.; Hallam-Baker, P.; Fox, B.; Dillaway, B.; LaMacchia, B.; Epstein, J.; Lapp, J (2002) *XML Key Management Specification (XKMS)* W3C Note, March 2001

[10] Gabillon, A.; Bruno, E. (2001): *Regulating Access to XML documents*. Fifteenth Annual IFIP WG 11.3 Working Conference on Database Security. Niagara on the Lake, Ontario, Canada July 15-18, 2001

[11] IPTC (2001): *NewsML (TM) in Action*. Retrieved from http://www.newsml.org/NewsMLweb/webpage.xml

[12] IPTC (2002): *NewsML Version 1.04. Functional Specification*. 6 March 2002

[13] Karlinger, G. (2001): *XML Electronic Signatures – Application according to the international standard XML Signature Syntax and Processing*. Proceedings of IFIP TC6 / TC11 Fith Joint Working Conference on Communications and Multimedia Security (CMS'01) May 21-22, 2001, Darmstadt, Germany

[14] Kravitz, J. (1998): *SDML – Signed Document Markup Language*. W3C NOTE, Junio 1998. http://www.w3.org/TR/1998/NOTE-SDML-19980619/

[15] Krawczyk, D. et. al. (1997): *HMAC: Keyed-Hashing for Message Authentication.* RFC 2104 February 1997.

[16] Kudo, M.; Hada, S.(2000): *XML Document Security based on Provisional Authorization*, Proceedings of 7th ACM Conference on Computer and Communication Security (CCS'00), Nov. 2000, Athens, Greece.

[17] Marco, M. (2001): *Publicación y gestión del contenido en las publicaciones periódicas en diversos medios*. II Jornadas Españolas de Bibliotecas Digitales JISBD 2001, Almagro, Spain. Retrieved from: http://gaia.dcs.fi.uva.es/~jbidi2001/comunicaciones/mmarco_almagro.pdf

[18] Morrison, M. et. al. (2000): *XML al descubierto*. Prentice Hall, Madrid, Spain.

[19] OASIS eXtensible Access Control Markup Language TC (2002), *OASIS eXtensible Access Control Markup Language (XACML)*, Committee Draft, May 2002. http://www.oasis-open.org/committees/xacml/docs/draft-xacml-v0.13

[20] OASIS Rights Language TC: http://www.oasis-open.org/committees/rights/

[21] OASIS SSTC, *XML-Based Security Services TC (SSTC) Security Assertion Markup Language* Work in Progress http://www.oasis-open.org/committees/security/

[22] OASIS XML Common Biometric Format TC: http://www.oasis-open.org/committees/xcbf/

[23] Platform for Privacy Preferences (P3P) Project: http://www.w3.org/P3P/

[24] Scheibelhofer, K. (2001): *What You See Is What You Sign - Trustworthy display of XML documents for signing and verification.* Proceedings of IFIP TC6 / TC11 Fith Joint Working Conference on Communications and Multimedia Security (CMS'01) May 21-22, 2001, Darmstadt, Germany

[25] Selkirk, A. (2001): *XML and security*. BT Technology Journal, Vol. 19 No. 3 JULY 2001

[26] Selkirk, A. (2001): *Using XML security mechanisms*. BT Technology Journal, Vol. 19 No. 3 JULY 2001

[27] Stallings, W. (1998): *Cryptography and Network Security. Principles and Practice. Second Edition.* William Stallings. Ed. Prentice Hall, 1998.

[28] W3C (1998): *Extensible Markup Language (XML) 1.0. W3C Recommendation.* Bray, T., Paoli, J., Sperberg-McQueen, C.M. February 1998

[29] XML Security Page at the Institute of Data Communications Systems, University of Siegen: http://www.nue.et-inf.uni-siegen.de/~geuer-pollmann/xml_security.html