2007

Analysis of PKI as a means of securing ODF documents

Gautham Kasinath
Edith Cowan University

Leisa Armstrong
Edith Cowan University

DOI: http://ro.ecu.edu.au/ism/36
Originally published in the Proceedings of 5th Australian Information Security Management Conference, Edith Cowan University, Perth Western Australia, December 4th 2007

Analysis of PKI as a means of securing ODF documents

Gautham Kasinath and Leisa Armstrong
School of Computer and Information Science
Edith Cowan University
Perth, Western Australia
gautham.kasinath@gmail.com
l.armstrong@ecu.edu.au

Abstract

Public Key Infrastructure (PKI) has for the last two decades been a means of securing systems and communication. With the adoption of Open Document Format (ODF) as an ISO standard, the question remains if the unpopular, expensive, complex and unmaintainable PKI can prove to be a viable means of securing ODF documents. This paper analyses the drawbacks of PKI and evaluates the usefulness of PKI in provisioning robust, cheap and maintainable XML security to XML based ODF. This paper also evaluates the existing research on XML security, more specifically fine grained access control.

Keywords

XML security, ODF security, PKI, Public Key Cryptography, Access control, Merkel tree hash scheme.

INTRODUCTION

"Security is a chain: it is only as strong as the weakest link" - (Carl Ellison and Bruce Schneier, 2000)

Extensible Mark-up Language (XML), in recent years has become the de-facto standard for information exchange in various mediums of communication (World Wide Web Consortium, 2006). The popularity of XML is based on its ability of representing data and context in a human readable and organized manner, among many others. XML documents are written in plain text American Standard Code for Information Interchange (ASCII) characters, which makes it easy and robust for not only human understanding, but also for multiple application portability. However, the plain text nature of XML documents has lead to increasing concerns of security risks and implications, where XML based communication and information exchange is critical. Such systems are those that either use XML formatted databases for data storage or XML based Web Services for inter application communication.

XML based technologies have also become the standard in Electronic Document Interchange (EDI). In May 2005, Organization for the Advancement of Structured Information Standard (OASIS) certified Open Document Format (ODF) as a document standard (OASIS-Standard:, 2005). ODF is an XML based format which stores the content of the document, presentation information and document meta-data in a XML format. ODF is not limited to office documents, but is also extended to entire office suites like presentations, spreadsheets, drawings to name a few. Since its adoption as a standard, ODF has now been widely accepted as industry wide standard by technology leaders like IBM and governments like Massachusetts. The influence of the popularity of ODF has also lead to rival office application suite makers such as Microsoft to enable an XML based format for their proprietary office document formats. Microsoft will release the next version of its Microsoft Office Suite with support for Office Open XML, a proprietary XML based format.

In the light of the fast gaining popularity of XML based office document formats, it is critical that an efficient, robust, scalable, inexpensive and usable means of securing these XML documents is discovered. The next section of this paper will discuss the basic working of Public Key Infrastructure (PKI). This paper will, then, briefly highlight some of the existing research and papers on securing XML documents, in the context of Office Documents. In the section after which, this paper will briefly explain the mechanism of Public Key Cryptography, in the context of ODF based EDI. The paper will then, in the following section, evaluate the feasibility of Public Key Cryptography based security mechanism for securing ODF based office documents.

PKI BACKGROUND

PKI is based on the Asymmetric Public Key Cryptography. In Public Key Cryptography, users have one key K. All messages exchanged are first sent through the key function K, in such a way that a unique encrypted text called "cipher" is procured out of it. This cipher is then transmitted to the recipient user. The recipient decrypts the cipher by applying the same key K, to procure the plain text message. The strength of this scheme lies on the

secrecy of the keys (Schneier, 2001). The Symmetric Public Key Cryptography scheme is represented mathematically below in figure 1.

Sender:
Cipher text C = f(K, M).

Recipient:
M = f(K, C)
Where C is the cipher text,
M is the plain text message,
K is the encryption Key and
f is the encryption function.

Hence it can be derived as M = f(K, f(K,M))

Figure 1: Symmetric Public Key Cryptography.

However, in Asymmetric Public Key Cryptography, each user is provided with two keys, Public Key K_{pu} and Private Key K_{pr}. In this scheme, the sender encrypts the plain text message using the public key K_{pu} of the recipient. The recipient decrypts the cipher procured by applying ones own private key K_{pr} (Schneier, 2001). The Asymmetric Public Key Cryptography scheme is represented mathematically below in figure 2.

Sender:
Cipher text C = f(Krcpt,pu, M)

Recipient:
M = f(Krcpt,pr, C)

Where C is the cipher text
M is the plain text message
Krcpt,pu is the recipients' public key
Krcpt,pr is the recipeints' private key
f is the encryption function

Figure 2: Asymmetric Public Key Cryptography.

It is clear from the discussion above, in Asymmetric public key cryptography, it is not required by the users to make public the decryption private keys. Although the scheme described above ensures confidential transmission of messages, it does not solve the problems of message integrity verification and non-repudiation (Schneier, 2001). To solve the same, the encryption function can be modified to use one of the many protocols available, such as zero-knowledge proofs and other arbitrated protocols (Schneier, 2001). However, asymmetric public key cryptography introduces the problem of securely transmitting, storing and distribution of users' public keys. This is because the strength of the scheme lies in the secrecy of the keys (Schneier, 2001). To solve such problems, centralized entities to store, authentically transmit and verify keys and signatures were introduced. These entities are called the Certificate Authority (CA).

PKI is an architectural paradigm, as illustrated in figure 3. In PKI, the CA is a trusted centralized entity that is responsible not only for securely storing users' public keys, but also perform authentication processes to distribute public keys of users to others. In the simplest form of a PKI based scheme, the sender first authenticates with the CA. Upon authentication, the CA sends the sender, the recipients' public key. The sender then applies the encryption function to the plain text message using the recipients' public key retrieved from the CA. The complexity of the PKI lies in the setup of the CA, formation of the user keys, key management, authentication, secure key transmission and secure key storage repository among many others. This complex structure of PKI renders it as an expensive scheme to protect XML document contents. To achieve a low cost and robust solution to protect XML document contents extensive research has been done. (Kolodzinski, 2002; Lioy *et al.*, 2006; Richardson, 2000; Schneier, 2001)

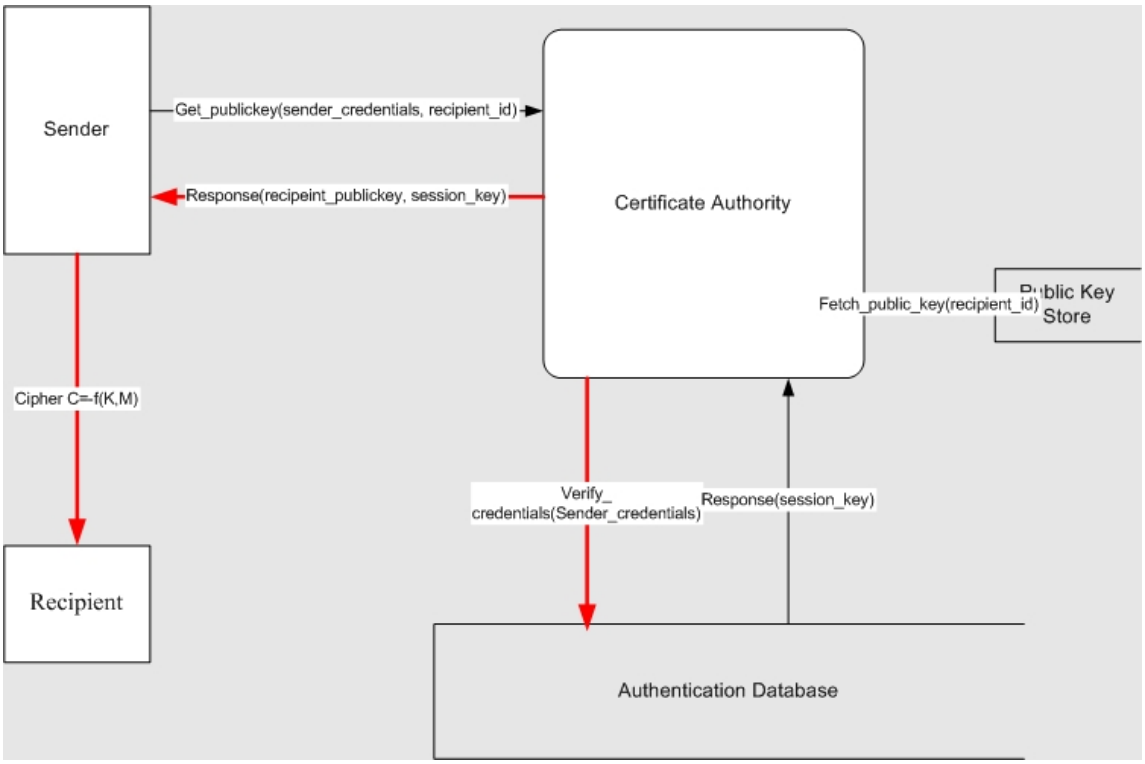

Figure 3: PKI working diagram.

RELATED RESEARCH

Although fast gaining popularity, Open Document Format (ODF), has not been adopted widely in the industry and academia. This is because ODF was recognized as a standard in 2004. Hence there is no documented work available which discusses the security issues related to ODF. However, means of securing XML documents as a whole may be directly applied with minor modifications to ODF, since ODF is XML based. This section will discuss the research related to XML documents in the context of securing their contents using PKI and other authentication and authorization schemes.

As a part of research supported by the National Science Council, Republic of China, (Yang *et al.*, 1998), were one of the earliest to suggest the use of PKI for securing XML documents in EDI. (Yang *et al.*, 1998) modelled a mechanism to exchange document securely using public key cryptography, with the use of session keys encrypted using the recipient public key. This, the authors call a "Digital Envelope". Contained in the Digital Envelope is a digitally signed XML document. The signatures in the XML document, in the model, were procured by applying RSA with SHA based one way hash function. The authors, (Yang *et al.*, 1998), also use a smart card to contain the cryptographic algorithms. The card is secured by the use of a personalized identification number (PIN), set up by each user.

The smart card forms the mechanism of authenticating a user. Upon authentication, the user can access the cryptographic functions in the smart card to sign the document and enclose the same in a digital envelope. Transmission according to (Yang *et al.*, 1998), was by email using the Simple Mail Transport Protocol (SMTP). The use of smart cards not only increases the cost of the model but also introduces the smart card itself as the weakest link in the model (Ellison and Schneier, 2000). This is because the model is only as secure as the security of the smart cards. If the card itself is compromised, then an adversary can impersonate the legitimate smart card user, till its loss is discovered. Further, the computational complexity of the algorithms used, such as SHA, RSA and TripleDES in the model adds performance and complexity overhead.

The authors (Yang *et al.*, 1998), however, in their paper, identify the problem of key distribution and advice the use of PKI to solve the same. As a solution (Yang *et al.*, 1998) propose that the smart cards contain the public key of the recipient, signed by a certifying authority as a correct key. This however aggravates the implications when the smart card is compromised. Whilst the existence of the smart cards as the weakest link is not eliminated, PKI based solutions are also known to be plagued with the problems of cost, complexity and inter-operability of the infrastructure set up and maintenance (Ellison and Schneier, 2000; Kolodzinski, 2002; Lioy *et*

al., 2006; Richardson, 2000). Further, establishing maintaining mutual trust in a PKI context presents another problem.

Kwong and Gertz, (2001), in their paper, propose a solution to the problems presented by mutual trust establishment and verification. Extending their model from databases, (Kwong and Gertz, 2001), apply the Merkel Hash Tree algorithm to compute, distribute and verify the authenticity and integrity of XML documents. In their model, (Kwong and Gertz, 2001), perform a bottom up hash computation using the Merkel Hash Tree scheme. The root of the tree then builds a summary signature of the data.

The summary signature is transmitted by the owner to all recipients. Recipients may perform signature verification of entire documents or sections of a document without further need for communicating with the owner. The verification itself is done with the help of Verification Objects (VO) provided by the data owner, along with the summary signature. Recipients of the document or its sections perform XML queries on the VO to verify the authenticity of the nodes or of the entire document. The VO itself has to meet certain conditions before it can reply to an authenticity query. The VO must ensure that the query against a certain XML sub-tree is available in the document and its path. The VO also must ensure that every sub-tree is available in complete in the document and its path. The document, its query and the path also need to be well ordered for the VO to confirm the authenticity of the same. The authors (Kwong and Gertz, 2001) advice that;

"Proving the completeness and correct ordering, however, is more difficult since one has to provide the client [recipient] with some boundary cases …"

Whilst this model negates the need for a PKI based authentication and allows for the document owner to be offline when recipients require authenticity confirmation, the model employs complex protocols which may require complete understanding of the protocol to set up services that realize the same in a corporate EDI environment. The protocol also does not attempt to solve the non-trivial problem of document modification and re-publication by recipients. In such cases, the original ordering of the tree may be affected, and new VOs need to be propagated, enacting the complete protocol for very change made to the document. In an EDI environment, involving ODF documents, such changes to the documents may happen with more than one entity. Further, documents may be verified and signed by relay stations in EDI. Such changes will nullify the owner submitted VO, and the whole protocol will need to be enacted.

Clark, (2004) in her article discusses the need for identity crisis resolution by the use of technologies alternate to the PKI. She mentions an identity based encryption product that does not use certificates and hence eliminates the use of PKI. The product, according to (Clark, 2004), first converts the recipients email address into a number and encrypts the message. Although there is no mention of the use of the number in the encryption algorithm, it can be assumed that some form of symmetric two-way encryption mechanism is used based on the seed provided by the recipients email address. (Clark, 2004) also brings to light one of the issues in the product architecture revolving around a centralized key repository.

In an article in Information Week, Rosen, (2001) show how two products, in the market make practical implementations of PKI, enhancing user experience and customer security. Predicting a higher adoption of smart cards, Rosen, (2001), in her article mentions one another product that uses smart cards to contain encrypted identification information. This identification according to this article, can unlock credit card, bill-to and shipment addresses to the retail. However, since only the objective of security is shifted, such smart card based solutions only aggregate the security problem, rather than enhance a solution.

Kolodzinski, (2002) in his article describes some of the reasons for PKIs failure to be popular. Kolodzinski, (2002) outlines the inter-operability of PKI verification process between different PKI vendors and CAs. While one CA would use a Light Weight Directory Access protocol (LDAP) directory server to confirm user identity, another may use a Kerberos ticket issued by a Microsoft Active Directory for the same. Further, with various mechanisms of user authentication, varying from username-password pairs to Secure Access Mark-up Language (SAML) tokens, each CA may choose to support only a few. In such cases the preliminary step of verifying user identity stands to be the point of failure for a successful PKI setup. PKI has commonly known to be of complex design issue requiring experienced security professionals to set up.

Ellison and Schneier, (2000) in their paper summarize the drawbacks of PKI as risks of PKI. Ellison and Schneier, (2000) advise that CAs cannot be trusted in a cryptographic context, since it cannot be established how the CA acquired the "trust". Further, the author's advice that risk of security compromise is handed to the certificate verifier instead of the certificate issuer. The second risk that Ellison and Schneier, (2000) describe is with the non-repudiation. If the certifying authorities contain a repository of public keys, the misuse of keys hands the implications to the user whose keys were compromised rather than the CA. The authors also describe the security of the computing environment of the CA itself as a security risk. This is because, although the CA stores only public keys, they contain a "root key". If an adversary is added to the list then the adversary can

issue false certificates which may be cryptographically legal and correct. Hence it is important that the CA computing environment is secure, physically, cryptographically and devoid of software based attacks. Ellison and Schneier, (2000) also describe the certificate format and the information contained in the same as a risk. Certificates issued by CA contain the name of the sender. However, the authors argue that there is no way of ascertaining the identity of the sender as the one the recipient thought it to be. Further, the certificates contain the key holders name and the DNS name of the server. The authors argue that the DNS name in the certificate is not an authoritative statement. This is because it cannot be ascertained how the computer, represented in the DNS name, acquired the authority to perform SSL based communication. In recent years, with the popularity of hacking of DNS servers, it is also impossible to ascertain the authenticity of the DNS server when confronted only with the certificate issued by a CA. The authors, Ellison and Schneier, (2000), describe more risks of using the PKI and finally describe motivation for using a localized CA based PKI.

ALTERNATE MECHANISMS FOR SECURING ODF

Damiani *et al.*, (2001) in their paper describe the limitations of encryption alone for securing XML documents. Damiani *et al.*, (2001) base their paper on the need for a fine grained access control system, to protect information. While the scheme may be sufficient for XML documents, in opinion, it does not suffice for ODF documents. Although ODF documents are also XML documents, the ODF documents may be transmitted as a whole to recipients, instead of views generated based on an access evaluation policy. Hence, it is important that the contents of the XML document be encrypted so that the recipients cannot use read the plain XML to acquire protected information. In other words, in the context of ODF documents, encryption and access control are both mandatory for effective security of documents and its contents.

PKI is known to solve web based authentication, identity and verification problems. But the cost, complexity, inter-operability and risks involved with it hinders its use in corporate EDI. In corporate EDI fine grained authentication to document contents is required. However, some aspects of PKI may be used in combination with other authentication and verification schemes such as Merkel Tree Hash model. The following sections, will describe a hybrid scheme and some of the benefits of the same.

In the last decade, (Dridi and Neumann, 1998), suggest one of the first means of securing structured documents. Highlighting the need for security in EDI, (Dridi and Neumann, 1998) predicted the wide spread use of hyper linked documents in document and information sharing in a corporate network and in the world wide web. Dridi and Neumann, (1998) in their paper describe the solution for unauthorized disclosure of information contained in XML documents using a lattice model. The authors, Dridi and Neumann, (1998), refer to the sensitivity of documents or its components as “classification”. The access privileges of a user is regarded as the users’ “clearance”. The authors, Dridi and Neumann, (1998), construct a lattice of security labels, which comprise of the classification and clearance. Such a lattice may be viewed as an access control matrix. Access control is provided by three forms of mutual exclusive policies, namely, Access Denial, Covert Censure and Censure policy. The Access Denial policy that the authors use defines that the access to the document as a whole must be granted or denied. Covert censure policy covertly modifies the view of the document and does not include the contents denied for a user. The Censure policy on the other hand modifies view in the similar way as that of the Covert Censure policy, but indicates to the user of the access restriction on denied content. The approach of having three schemes to define the response of an access evaluation request allows certain flexibility.

Various research such as those in (Bertino *et al.*, 2002; Bertino *et al.*, 2004; Bertino *et al.*, 2001; Bertino and Ferrari, 2002; Bhatti *et al.*, 2005; Dridi and Neumann, 1998; Fundulaki and Marx, 2004; Johnsten *et al.*, 2003; Kudo and Hada, 2000; Lim *et al.*, 2003) also suggest, with conviction, the efficiency of manipulating the views of the XML documents to achieve data hiding. Kuper *et al.*, (2005) in their paper demonstrated the strength of generating multiple security views of XML documents. However, their model was based on XML documents created from data in database objects. The XML documents, in the model, were governed by access control modifiers in the Document Type Definition (DTD). Since Kuper *et al.*, (2005) worked on a new model, they could freely define DTD that worked for their model. However, in the case of ODF, there already exists the DTD for the storage and presentation format of XML documents that contain the data and presentation logic of ODF documents. Hence, a DTD governed access control policy cannot be directly applied to ODF.

Kuper *et al.*, (2005) however also demonstrated that generating various views of the XML document using access control policies, as the core concept of the model. Generating different views is however, a concept that can be directly applied to ODF documents. To achieve the same, this paper focuses on Extensible Access Control Markup Language (XACML) for creating, managing and applying access control policies (Kay, 2003; Lorch *et al.*, 2003; Sun Microsystems Inc., 2003a, 2003b). A DTD not directly connected to the ODF DTD specification may be used to bind the document contents to its views based on a XACML access control policy. Further, allowing each document creator and modifier the ability to change the access restrictions, the manageability of the access control policies is made easier. However, in ODF, since the entire document is

transmitted as XML, there arises a need for securing the contents of the XML from back door viewing of plain XML.

Prevention of backdoor viewing of plain XML can be achieved, using a simple and manageable cryptographic algorithm or scheme to perform two-way encryption of the select contents of the XML nodes. To perform the same, certain aspects of PKI, namely the asymmetric Public Key Cryptography may be used. Although PKI is plagued with drawbacks, asymmetric public key cryptography is a feasible solution in a limited size corporate environment. However, the use of PKI inherits the key distribution problem. Hence, until a more viable solution is found, a centralized repository of public keys, within a corporate environment, is a feasible solution. Further, the repository may be combined with an LDAP based authentication server to authenticate public key requests. However, the design of such a system binding the LDAP server and key repository will need to consider issues related to the same such as those documented in (Chadwick, 2003).

CONCLUSION

In summary, PKI alone is not a feasible solution to securing XML based ODF documents. However, some of the core concepts of PKI, together with robust fine grained access control system are a feasible attempt at finding a viable solution for securing ODF documents. Because ODF is still in its infancy, no documented research can be found for securing ODF based office documents. However, existing XML security research can be extrapolated to apply to ODF documents. Public Key Cryptography, as the concept behind PKI, can also be applied to securing ODF documents with robust access control mechanism, such as XACML, to achieve a cheap, maintainable and viable solution to securing ODF documents. However it must be noted that security cannot only rely on cryptography alone (Schneier, 1999).

REFERENCE

Bertino E., Carminati B., and Ferrari, E., (2002). A temporal key management scheme for secure broadcasting of XML documents. In *Proceedings of the 9th ACM Conference on Computer and communications security.* pp 31-40. Washington, DC, USA: ACM Press.

Bertino E., Carminati B., Ferrari, E., Thuraisingham, B., and Gupta A., (2004). Selective and authentic third-party distribution of XML documents. *IEEE Transactions on Knowledge and Data Engineering*, 16(10), pp 1263-1278.

Bertino E., Castano S., and Ferrari E., (2001). Securing XML Documents with Author-X. *IEEE Internet Computing*, 5(3), pp 21-31.

Bertino E., and Ferrari, E., (2002). Secure and selective dissemination of XML documents. *ACM Trans. Information and Systems Security*, 5(3), pp 290-331.

Bhatti, R., Ghafoor, A., Bertino E., and Joshi, J. (2005). X-GTRBAC: An XML-Based Policy Specification Framework and Architecture for Enterprise-Wide Access Control. *ACM Transcations on Information and Systems Security,* 8(2), 187-227.

Chadwick D., (2003). Deficiencies in LDAP when used to support PKI. *Communications of the ACM*, 46(3), 99-104.

Clark. A. J. (2004). Encryptions' identity crisis. *Network magazine*, 19, 17-18.

Damiani, E., Samarati De Capitani di Vimercati S., and Paraboschi. S., (2001). Controlling access to XML documents. *IEEE Internet Computing*, 5(6), 18-28.

Dridi F. , and Neumann G., (1998). Towards access control for logical document structures. *Proceedings. Ninth International Workshop on Database and Expert Systems Applications*, 1998., 322 - 327.

Ellison C. and Schneier. B., (2000). Risks of PKI: Electronic Commerce. *Communications of the ACM*, 43(2).

Ellison C., and Schneier. B. (2000). Ten Risks of PKI: What you're not being told about Public Key Infrastructure. *Computer Security Journal*, 16.

Fundulaki I., and Marx. M., (2004). Specifying access control policies for XML documents with XPath. In *Proceedings of the ninth ACM symposium on Access control models and technologies* (pp. 61-69). Yorktown Heights, New York, USA: ACM Press.

Johnsten, T., Sweeney R.B., and Raghavan. V.V. (2003). A methodology for hiding knowledge in XML document collections. *Proceedings of the 27th Annual International Computer Software and Applications Conference*, 2003. COMPSAC 2003.

Kay. R., (2003). XACML. *ComputerWorld*. Retrieved on 24th September 2007 from http://www.computerworld.com/printthis/2003/0,4814,81295,00.html

Kolodzinski. O., (2002). PKI: Commentary and Observations. *The CPA Journal*, 27(11), 10.

Kudo M., and Hada. S., (2000). XML document security based on provisional authorization. In *Proceedings of the 7th ACM conference on Computer and communications security* (pp. 87-96). Athens, Greece: ACM Press.

Kuper, G., Massacci N., and Rassadko. N., (2005). Generalized XML security views. In *Proceedings of the tenth ACM symposium on Access control models and technologies* (pp. 77-84). Stockholm, Sweden: ACM Press.

Kwong A., and Gertz. M., (2001). Authentic publication of XML document data. *Proceedings of the Second International Conference on Web Information Systems Engineering*, 2001. 1, 331 - 340.

Lim, C. H. Park S., and Son. S. H. (2003). Access control of XML documents considering update operations. *In Proceedings of the 2003 ACM workshop on XML security* (pp. 49-59). Fairfax, Virginia: ACM Press.

Lioy, A., Marian, M. Moltchanova, N., and Pala. M. (2006). PKI past, present and future. *International Journal of Information Security*, 5(1), 18.

Lorch M,. Kafura D., and Shah S., (2003). An XACML-based policy Management and Authorization Service for Globus resources. Proceeding of the *Fourth Internationl Workshop on Grid Computing*. p208

OASIS-Standard (2005). Open Document Format for Office Applications (OpenDocument) v1.0, 2005

Richardson. R. (2000, June 2000). Public key infrastructure. *Network Magazine*, 15, 114-120.

Rosen. C. (2001, June 11, 2001). 'Practical PKI' Equals Purchasing Power. *Information Week*, 32.

Schneier. B., (1999). Risks of Relying on Cryptography. *Communications of the ACM*, 42(10), 144.

Schneier. B., (2001). *Applied Cryptography Second Edition: Protocols, Algorithms, and Source Code in C.* In P. Sutherland (Ed.): John Wiley & Sons.

Sun Microsystems Inc. (2003a). *XACML in Use*. In http://www.computerworld.com/computerworld/records/images/story/XACMLinUse_large.gif (Ed.): ComputerWorld.

Sun Microsystems Inc. (2003b). *XACML: A New Standard Protects Content in Enterprise Data Exchange.* Retrieved 30-09-2005, 2005, from http://java.sun.com/developer/technicalArticles/Security/xacml/xacml.html

World Wide Web Consortium. (2006). *XML*. Retrieved 18-March-2006, 2006, from http://www.w3.org/XML

Yang, C.H. Ju S. H. and Rao. T., (1998). A smartcard-based framework for secure document exchange., 1998. *Proceedings of the 32nd Annual 1998 International Carnahan Conference on Security Technology.* p93-96.

COPYRIGHT