Article in T-110.501 Seminar on Network Security 2001 ISBN 951-22-5807-2
Publications in Telecommunications Software and Multimedia TML-C7 ISSN 1455-9749
Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory

http://www.tml.hut.fi/Studies/T-110.501/2001/papers/index.html

XML security and relevant to E-business

Xiaoxin Sun
Helsinki University of Technology
Computer Science Department
xsun@cc.hut.fi

Abstract

As one of enabling technology for e-commerce, the XML (Extensible Markup Language)'s huge potential has been known more and more widely. XML is known as the new Ascii of the Internet. Especially in E-business messaging, XML is expected to facilitate it because of its flexibility and simplicity. XML is going to be new standard in industry field. And in XML security feature, XML document security and extending XML function are what we will be considered. They are used widely in web data transmission.

1. Introduction

1.1 Brief in XML

XML is a method for putting structured data in a text file and allows that there can be an interaction between different hosts regardless of operation system. XML is extensible because it is not fixed format just like HTML (Hypertext Markup Language). Actually, XML is a 'metalanguage', with which we can describe other languages. This means user can define their own customized language for limitless different types of document. XML can do that because it is simplified from SGML (Standard Generalised Markup Language), which a generic language for writing makeup language.

In some cases, HTML is similar with XML. They are both markup language. They both use tags and attributes. But HTML defines a very simple class of report-style document, it specifies what each tag and attribute means. Instead XML use them only to delimit pieces of data, and leaves the interpretation of the data completely to the application that reads it. HTML appeared when accessing held on computers in faraway countries was still a novel experience for most people and HTML is still the main format for web browse. But world is changing day after day. And people want more from Internet. As some examples of web new application, E-business, E-bank as well as E-education are developing. HTML is not enough. It is limited and clumsy. Instead, extensible and flexible are advantages of XML. As for how XML uses in these new applications, details will be involved later.

1.2 Brief in web security

Nowadays more and more companies use XML to transmit structured data across the Web, the security of documents becomes increasingly important. But why we need security in web? Reason is very simple but important. As we know, Internet is a public network. For instance, in e-commerce field, data transmission is between buyer and seller, if there is no protection against attacking such as eavesdropping and forgery, messages may be stolen or modified during transmission. So that E-commerce messaging is no useful. Seller and buyer do not trust each other but principle of E-business is based on trust. And now people have higher and higher requirement to web application and we can sure that in future almost everything will be based on Internet, for instance, shopping, education, etc. In order to make these applications realised, first security will be paid wide attention to. What kind of security issues needed when we send data across the web? What effects security transmission? Before solving these problems, there are four basic things must be mentioned first.

1. Confidentiality -- No one else can access or copy the data.
2. Integrity -- The data isn't altered as it goes from the sender to the receiver.
3. Authentication -- The document actually came from the purported sender.
4. Nonrepudiability -- The sender of the data cannot deny that they sent it, and they cannot deny the contents of the data.

These four preliminary things are also relevant to XML document security. Even related to XML based E-business. How does it going on? You will find answers in the following chapters. [3]

2. XML security

2.1 Introduction to XML security

Internet is as insecure as ever, and XML itself can’t do anything to improve it. In fact, intercepting and altering an XML document containing important data from one location to another through Internet will be a lure to someone. XML security is to using some methods to improve security of XML file when it is transferred over the Internet. There are two ways to extend XML file security. The first is to improve XML document security itself. Encryption and digital signature are two primary requirements of it. They only take care of the security of XML documents regardless of their use. On the contrary, second category is to seek leverage outside XML documents itself to broaden security function applications. Their primary focus is some other standards, something like SAML (Assertion Markup Language), XKMS (XML Key Management Specification), XTASS (XML Trust Assertion Services Specification), S2ML (Security Services Markup Language). XML is a family-technologies and these are all XML security applications. Now let’s introduce them one by one.

2.2 Security of XML document

Ensuring the security of XML document are very complex. Before any organization designs an application for flow data transmission, primary web security concepts must be considered. We must establish and apply XML document security under confidentiality, integrity, authentication and nonrepuiation. [7]

In order to achieve these security objects. We first should under stand how XML processing and transaction happen. Typical process begins with authentication, which maybe contains multiple transactions. And before submission of any transaction, strict authentication requires. After transaction is submitted, transmission to destination will be established. In this process, sign and encryption happen. When flow data is received or retrieved, signature verification, which is the reverse operation to encryption, also we call it decryption, is provided confidentiality and integrity. Finally, transaction and flow data will be stored. Above is typical process of those four objects achieved. Also same in XML document security transmission, from this process, we can sure that encryption provides the confidentiality, and signatures provide the integrity, authenticity, and, in some cases, nonrepudiation.

2.2.1 XML encryption

Proper encryption is crucial for XML data security. It provides confidentiality for data transmission over Internet. It's easy to think of that in encryption operation data is encrypted on one end. Then decrypted on the other. And in an XML document, there are four basic types of information: encrypted contents, unencrypted, key information and recipient information.

- Encrypted contents contain the actual encrypted data or a reference to the location of this data. There is virtually unlimited flexibility in both the types of data that can be included and methods for logical data collection for encryption.

- Unencrypted contents contain other information that is pertinent to the context of the interaction but isn't encrypted for some reason, perhaps due to performance concerns or because it wasn't deemed private or sensitive enough to warrant encryption.

- Key information is very important thing for decryption. Key information contains information or pointers to information about the keys that perform the encryption. The key information can be maintained elsewhere and replaced by a URL in the XML document.

- Recipient information is not compulsory and it contains information about one or more intended recipients of the encrypted data. Because it is optional, thus allowing situations where the applicable recipient information is known or provided out of band.

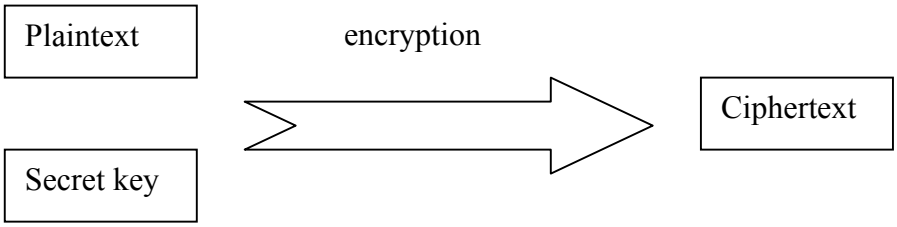

Figure 1 Encryption Principle

Encrypting XML data follows the traditional encryption steps for public key cryptography. This is a procedure, which takes the original message (plaintext) and a small piece of information (key information) arranged in advance between sender and receiver and creates an encoded version of the message (ciphertext). In detail, first the data is encrypted, typically using a randomly created secret key. Then the secret key is encrypted using the intended recipient's public key. This information is packaged to ensure that only the intended recipient can retrieve the key and decrypt the data. Decryption involves applying the private key to decrypt the secret key, then decrypting the data with the secret key. There is possible to have multiple ways to embed encryption elements with an XML document. And W3C (World Wide Web Consortium) is trying to convent an XML encryption work group to create a standard for it. It charges with specifying a data model, syntax and processing for encryption of XML content.

Encryption is not too difficult and its importance is well known. But one thing has to be sure that encryption algorithms should be a popular, well tested and tried one to resistant against plain-text attacks. This is because that tags, which are used in specific XML document, may contain lots of data and it is long that provides enough known plain text for an attack. However well known encryption algorithms can be fairly resistant to such attacks. [12]

2.2.2 Digital signature

In order to control or manage data that is passed and presented in XML transactions, XML signature is used. It is the simplest way to guarantee completeness and integrity in data transmission. Digital signature is totally different from encryption, and signature provides integrity, signature assurance and nonrepudiatability over Web data. “Such features are especially important for documents that represent commitments such as contracts, price lists and manifests. In view of recent Web technology developments, the proposed work will address the digital signing of documents using XML syntax. This capability is critical for a variety of E-commerce applications, including payment tools." This explanation comes from W3C.

Digital signature is not as simple as it seems. There is an example for an ordinary digital signed message transit between person A and person B. Person A is a sender. (1) A creates a mixture message and then encrypting the value with their own private key before sending it to person B. B knows A's public key and can decrypt the value. He knows that only A could have sent it because only A's public key will decrypt it. (2) B then recreates a new value from the message and compares it with the

value sent by A (3). If it is the same, it proves that only A could have sent the message and nobody has tampered with it. (4) The message has been digitally signed and can be considered safe.

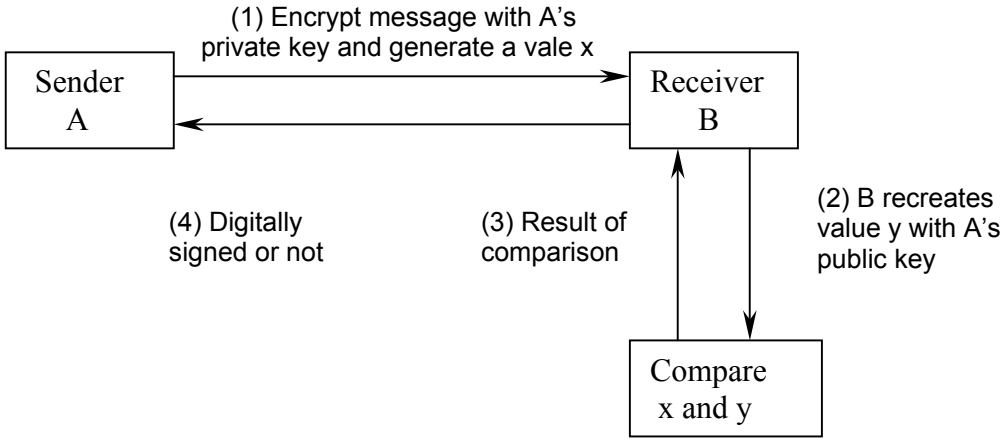

Figure 2 Digital Signature Process

But same principles to an XML document aren't suitable. For example, the XML document concerned could be a form. According to above principle, we have to digital signed whole XML document including form. In this case but user need to be allowed to legitimately tamper with the document by filling it in. Clearly, if we attempted to sign the whole document, any value would be altered by the user's additions. We only can sign parts of the document under this situation. That means an altogether more complex task is needed. So XML signature defines the syntax required to sign all or part of an XML document. XML, with its extensive capabilities and extreme flexibility, doesn't let itself to satisfy the needs of digital signature, where a misplaced space results in a completely different fingerprint that is unverifiable.

Let's first have a look to which information will contain in a digital signed XML instance. This information is used to ensure that the signatures are verified.

1. The canonicalization method, which identifies C14N rule. C14N is used to ensure that instances are structured the same way every time. So stylistic differences will not be confused during digital signatures process. It simplifies and structures an XML instance prior to signature. Verification algorithm will succeed if the data contents haven't been modified.

2. The signature method, is to digital sign the message digest. (Message digest is representation of text in the form of a single string of digits, created using a formula called a one-way hash function.) This method validates digital signed process, authentication and nonrepudiation can be defined.

3. The digest method, which identifies the algorithm that creates the message digest signed by the signature method. When comparing the resulting values, this method can ensure data will be processed with the same algorithm.

4. The digest value, is the fixed, unique and one-way string output through a message digest algorithm. It just likes the fingerprint to the data digital designed contents. The integrity will be defined depending on a valid comparison of digest values.

5. The reference information, which provides information about the data, which will be signed.

6. The optional signature properties, which add data to a signature instance, which may be timestamp or serial number.

How is XML digital signature going on? The principle is combined with two processes. First is called “core generation”. It divides XML document into two parts. One is begun with canonicalization (XML-C14N) to simplify the data content. The rest of digital signing XML document is similar to the typical digital signature process. That means in XML signature digest methods will create digest value of data content instead of whole document digital signed. Second process is for signature verification called “core validation”. There are two steps there: one is to ensure authentication and nonrepudiation with signature validation. Another, the digest value is verified to ensure that the data hasn’t be changed, which is to confirm content integrity. [7] [13]

2.3 XML for security function

Beside XML encryption and digital signature, which exist in XML document inside, seeking new leverage XML’s capabilities to further broaden security function also is very important to XML security compared with XML document security. XACL (XML Access Control Language), SAML, XKMS, XTASS, S2ML are involved here.

2.3.1 XML access control

Access control is the traditional center of gravity of computer security. Its function is to control which principals (person, process, machines, …) have access to which resources in the system – which files they can read, which programs they can execute, how they share data with other principals, and so on. Access control works at several levels. And XML access control works in application level and its aims are to provide XML document with a complex access control model and access control specification language. So XACL appears. XACL is an application for access control process. It is an access control policy specification language that is a primary component of XML Access Control technology.

XACL is developed by IBM and based on a provisional authorization model. In ordinary access control, we define that user can make an access request of a system and the system either authorizes the access request or denies it. But in the provisional authorization model, system is not simply acceptable or deny. It tells the user that his request will be authorized provided he takes certain actions or that his request is denied but the system must still take certain actions. These actions are called provisional actions. Something like auditing, digital signature verification, encryption,

and XSL (Extensible Style Language) transformations in addition to write, create and delete actions. These provisional actions enable us to specify policies such as:

- A user is authorized to access confidential information, but the access must be logged.
- A user is authorized to read sensitive information, but must sign a terms and conditions statement first.
- If unauthorized access is detected, a warning message must be sent to an administrator.

In a word, XACL is centered around a subject-privilege-object oriented security model in the context of a particular XML document. This means, by writing rules in XACL a policy author is able to define who can exercise what access privileges on a particular XML document. [1] [8]

2.3.2 SAML

SAML, Security Assertion Markup Language, is developed under OASIS (Organization for the Advancement Structured of Information Standards)-XML-based security service technical committee. It is created via syntax and semantics of XML. In this case, that means SAML is an XML-based security standard for exchanging authentication and authorization information.

Primary goals of SAML are:

- Let users carry entitlements with them across multiple sites.
- Enable interoperability among authentication and authentication applications in disparate security systems.
- Define SAML bindings to other protocols, such as HTTP, SOAP (Simple Object Access Protocol) and ebXML (Electronic Business eXtensible Markup Language).

SAML "specifies several different types of assertion for different purposes; these are: (1) Authentication Assertion: An authentication assertion asserts that the issuer has authenticated the specified subject. (2) Attribute Assertion: An attribute assertion asserts that the specified subject has the specified attribute(s). Attributes may be specified by means of a URL or through an extension schema that defines structured attributes. (3) Decision Assertion: A decision assertion reports the result of the specified authorization request. (4) Authorization Assertion: An authorization assertion asserts that a subject has been granted specific permissions to access one or more resources." (Above is from specification of SAML.) [16]

2.3.3 S2ML

S2ML, Security Services Markup Language, is the first industry standard for enabling secure e-commerce transactions through XML. S2ML was developed to provide a common language for the sharing of security services between companies

engaged in B2B (Business to Business) and B2C (Business to Consumer) business transactions. [9]

S2ML allows companies to securely exchange authentication, authorization, and profile information between their customers, partners, or suppliers regardless of the security systems or e-commerce platforms that they have in place today. As a result S2ML promotes the interoperability between disparate security systems, providing the framework for secure e-business transactions across company boundaries.

S2ML has following feature:

- Interoperability - With S2ML, e-marketplaces, service providers, and end user companies of all sizes can now securely exchange information about users, Web services, and authorization information without requiring partners to change their current security solutions.

- Open Solution - In the B2B market, S2ML is designed to work with multiple transport protocols such as HTTP, JMS (Java Message Service) and multiple XML document exchange frameworks such as SOAP, and ebXML.

- Single Sign-On Across Sites - In user driven transactions for the B2C market S2ML will enable users to travel across sites with their entitlements so that companies and partners in a trusted relationship can deliver single sign-on across sites

Let's have a deep look to S2ML used in E-commerce. As we know modern business focuses on two different environments. One is B2B environment. Another is B2C environment. In B2B transaction based on XML, current solution is based on XML-document security exchange model. Standard S2ML security information can travel with XML documents based on any agreed-upon vocabulary, enabling secure B2B transactions. In B2C environment, the most important is signal sign-on and access control. That means when a user visits website and to its hyperlink to other related website without re-authenticating. But S2ML can describe authentication information using standard XML language. With this utilization of S2ML, a user can travel across a network and its hyperlink sites without having to log on several times.

2.3.4 Other XML function applications

There are several other languages in the early stages of development and support. One is XKMS, which is set to revolutionize PKI (Public Key Infrastructure) by providing new levels of ease and interoperability to developers implementing secure applications. It simplifies the integration of PKI and digital certificates, the standard methods for securing Internet transactions, with XML applications. With XKMS, trust functions reside in servers accessible via easily programmed XML transactions. XTASS, which is a standardized way to make trust assertions for business partner relationships, is also one of this kind of languages. XMLPay, which provide secure e-commerce payment processing, and EPP (Extensible Provisioning

Protocol), which provide streamline domain name registration, are both new broaden XML security languages. Some of them are widely used in E-commerce.

3. XML E-business

We can see a huge metamorphosis in terms of how businesses are conducted in a wired world. The next generations of browsers are all going to support XML and XML's greatest strength is to allow developers to custom built systems for data exchange. Let's have a look to this new future of E-business.

3.1 Concepts of E-transmission

Adoption of electric technologies to do business over Internet requires same security process as doing business in the real world. That is there shouldn't have some sensitive information be accessed in public. Documents, which should be unalterable, can identify where they come from. (In e-business it is the same as digital signature.) Finally, it should be sure that document has been actually sent. (Doing digital signature again)

Adoption of cryptographic technologies enables the four critical aspects of web security. They are four basic concepts of web security as I mentioned before.

Confidentiality It is much possible to transform E-document so that it is unintelligible to anyone but the intended recipient with encryption technologies. Now, we can transmit sensitive documents through open web much safer without being intercepted and read by unauthorised individuals.

Integrity Digital signatures technology is used here. It can verify that an E-document has arrived intact and unaltered from the moment that the sender signed it. So that a recipient can verify that a document has not been altered, whether deliberately or accidentally, from the time that it was issued.

Authentication Certificates and digital signatures technologies are used here. It is possible to uniquely identify the origin of an E- document so that a recipient can sure to verify from whom a particular message has arrived.

Nonrepudiation With the help of PKI, there is no possibility for signer to revoke an order and an E-document cannot be denied at a later date in an attempt.

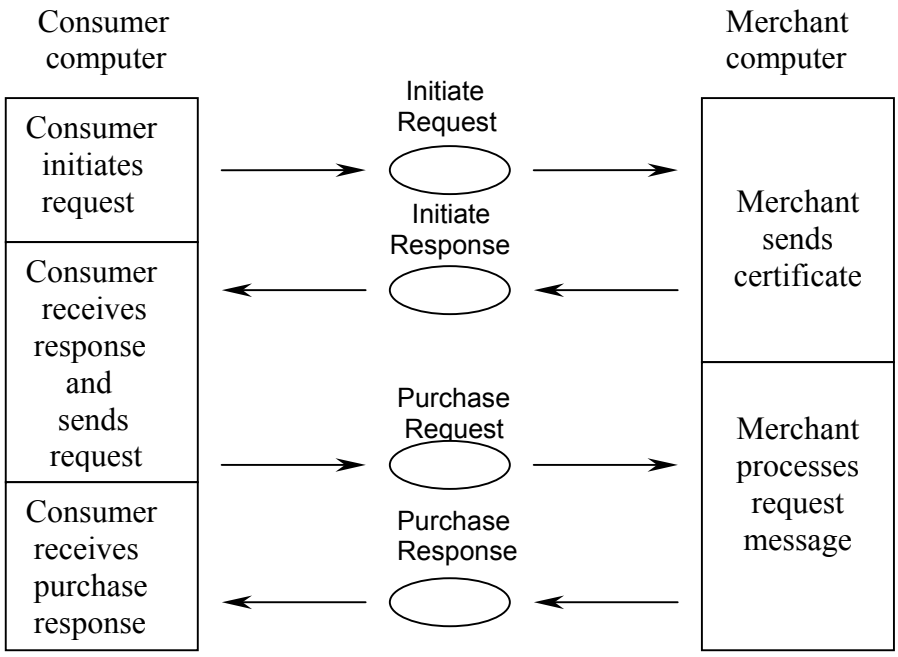

Figure 3 E-purchase Request [14]

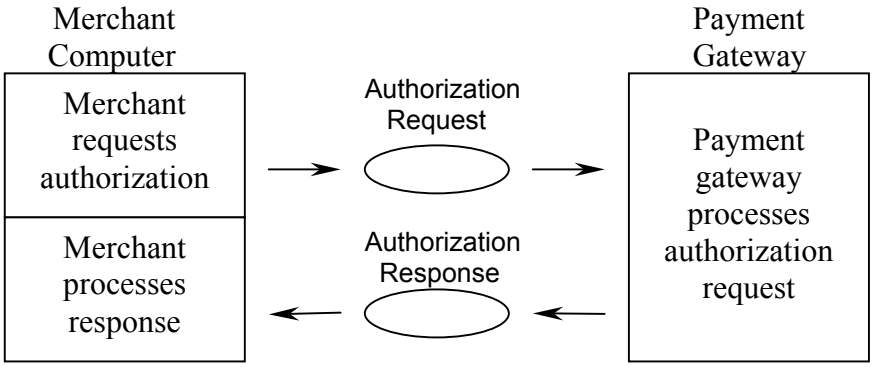

Figure 4 E-payment Authorization [14]

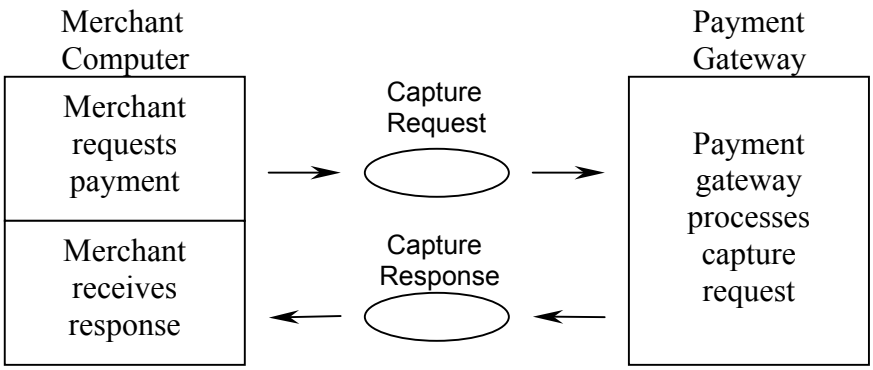

Figure 5 E-payment Capture [14]

3.2 E-business based on XML

E-commerce requires technology that permits disparate systems to communicate and XML does just that. XML promotes a message-oriented view of electronic commerce that isolates business transactions from differences in software, hardware, system architectures, and application programming languages. It is the language of e-commerce. XML can decode a web site and identify individual pieces of information in the page and assign each one of them with a special tag. The new generation of forms will allow a user to fill in his own preferences for an item that he chooses to order online. This order can be validated and processed immediately and at the same time it can check if the wholesaler stocks more of these products and even reorder if necessary. Orders can be downloaded regularly, even on an hourly basis thus enabling timely rescheduling of production to meet and satisfy customer needs.

All trading partners can communicate using a common standard. Encryption and signature standards for XML documents will permit the maximum use of XML capacities in conducing business transactions over the Internet. Beside Improved efficiency as a result of standardization in the representation and transfer of all information, data from XML to outputs in various media is transmitted without the necessity of modifying and duplicating content repeatedly. Also XML can simplify communication (user only need to know common XML vocabulary) and provide lower costs and decreased complexity in E-commerce. XML can access data to many kinds of devices, such as desktop systems, personal digital assistants and cellular phone without making any adjustment at user end. [15]

3.3 Applications of XML in E-business

What are metamorphoses of E-business with the help of XML? [15]

Online Publishing and Web Maintenance

XHTML is one of extended XML language. Data published in XHTML possesses the structure to model itself in any chosen manner. Any XML document can contain a description of its grammar or syntax that can be used to perform structural validations later.

Content and knowledge management

Because of variety of XML-based choice format their back-end applications, different kind of E-business uses itself kind of XML vocabulary. For example, E-banking and financial sectors use FinML (Financial Markup Language) for the exchange and storage of data insurance, real estate and a host of other industries use XML languages. In fields where a suitable vocabulary is not available, a generic one that inclines towards the specific requirements can be used between them.

Data Exchange

Because XML document can be suitable for different system, data can exchange between different institutes based on XML language. For example, health care systems, pharmaceutical agencies and other related organizations will be able to interchange patient records and billing data by replacing their existing disparate systems with a single common XML-based one. Drug testing and research

will happen with greater ease and within a shorter time if a single storage format is adopted by all health care services. A common information format, one for the output and another for the input will enable any system or application in any programming language to exchange data with each other in a simple manner.

Supply Chain Integration

XML format is simple and inexpensive. Before XML used, in the processing of product order, order taking product distribution, services between manufacturers, suppliers, wholesalers, retailers, traditionally used EDI (Electronic Data Format) that is expensive and requires sophisticated networking. Now network is switched to simple, inexpensive XML formats, resulting in the frictionless movement of e-commerce.

Design

XML can be applied in desktop product, mobile phone and so on. This is called multi-platform. So XML-based format can be used over extensive bandwidths and applications will save time in duplication of formats. A real time map can be made available to any user, at his desktop or on his car phone anywhere.

4. Conclusion

Foundation of XML security is based on two things: digital signature and encryption. Encryption standard specifies how to use XML to present digitally encrypted web resource with arbitrary encryption algorithms. Digital signature uses PKI to define its standard and its principle is allowing sender to sign just parts of XML document. These to things increase XML document security itself. With other broadened security function applications, just like S2ML, SAML, etc, no matter inside or outside, we deal with e-transmission security based on XML easily. Of course these applications including advantage of XML, E-business will have a widely usage. S2ML is a good application to do E-business, for example. Extending XML security in E-business, productivity will increase and companies will succeed in cutting costs and broaden business area, but in the long run, the chief winner will be the customer who will see prices decline and standards of living increase. E-business based on XML will have a wide future.

GLOSSARY:

XML	eXtensible Markup Language
HTML	Hypertext Markup Language
SGML	Standard Generalised Markup Language
SAML	Security Assertion Markup Language
XKMS	XML Key Management Specification
XTASS	XML Trust Assertion Services Specification

S2ML	Security Services Markup Language
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
C14N	defined in a W3C draft document called canonical XML
XACL	XML Access Control Language
XSL	Extensible Style Language
SOAP	Simple Object Access Protocol
EbXML	Electronic Business XML
B2B	Business to Business
B2C	Business to consumer
JMS	Java Message Service
PKI	Public Key Infrastructure
EPP	Extensible Provisioning Protocol
FinML	Financial Markup Language
EDI	Electronic Data Format

REFERENCES:

[1] XML Access Control
http://www.trl.ibm.com/projects/xml/xss4j/docs/xacl-readme.html

[2] XML Standards to Give Security Practices a New Face
http://www.advisor.com/Articles.nsf/aid/SMITT129

[3] X/Secure Whitepaper
http://www.baltimore.com/library/whitepapers/xsecure.html

[4] XML Standard To Keep Web Services Secure
http://www.internetweek.com/infrastructure01/infra073001-1.htm

[5] XML security fix in the works
http://www.fcw.com/fcw/articles/2001/0205/tec-xml-02-5-01.asp

[6] Security issues arise regarding XML
http://archive.infoworld.com/cgi-bin/displayArchive.pl?/98/26/i01-26.69.htm

[7] The Language Of XML Security
http://www.networkmagazine.com/article/NMG20010518S0010

[8] XML Access Control Language
http://xml.coverpages.org/xacl.html

[9] Security Services Markup Language
http://xml.coverpages.org/s2ml.html

[10] XML Key Management Specification
http://xml.coverpages.org/xkms.html

[12] XML and Encryption
http://xml.coverpages.org/xmlAndEncryption.html

[13] IT Week: XML and how to secure it, 10:17 Monday 22nd May 2000, KevinTownsend
http://news.zdnet.co.uk/story/0,,s2079073,00.html

[14] secure Electronic Transactions, Last updated 970224
http://www.cdt.luth.se/utbildning.direkt/multimedia/slides/seminars/set/index.html

[15] How XML is Changing the Future of Business,
http://www.stylusinc.net/technology/XML/xml_business.shtml

[16] The XML Cover Pages: Security Assertion Markup Language (SAML), Robin Cover, Last modified: October 23, 2001
http://xml.coverpages.org/xacl.html