TRIPLE PROTECTION FOR WEB DOCUMENTS USING THE COMBINATION OF CRYPTOGRAPHIC AND HASHING TECHNIQUES

CHINNU R, MARIA JOY

Abstract:Accessing information on the global Internet has become an essential requirement of the modern economy. Unless the system is able to provide some mechanisms to ensure security services, the system will have problems to be accepted. More reliable cryptosystems have to be proposed and cryptography is being an essential part of today's information systems. Here a cryptosystem for web documents using Vigenere cipher algorithm and ELGamal cryptosystem together with hashing is proposed ie, combining the features of both symmetric key and asymmetric key cryptography. Also hash value is encrypted to generate the signature and the signature is transmitted together with the message. The receiver extracts the signature and decrypted to obtain the hash. The receiver alsovalue calculate the hash value of the received file. The two hashes are compared. If they are same, no modification takes place to the transmitted file. Such a system is designed to achieve some of security aspects such as confidentiality, authentication, integrity, and non-repudiation.

Keywords: Vigenere cipher, ElGamal encryption, Signature, Symmetric key encryption, Assymetric key encryption.

Introduction:

A web document is similar in concept to a web page, but also satisfies the following broader definition of W3C as 'Every Web document has its own URI. Note that a Web document is not the same as a file:a single Web document can be available in many different formats and languages, and a single file, for example a PHP script, may be responsible for generating a large number of Web documents with different URIs'. A Web document is defined as something that has a URI and can return representations (responses in a format such as HTML or JPEG or JSP) of the identified resource in response to HTTP requests.The growth of the Internet has made cryptography is more important and critical issue in electronic application systems. Unless the system is able to provide some mechanisms to ensure security services, the system will have problems to be accepted. More reliable cryptosystems have to be proposed and, cryptography is being an essential part of today's information systems. Cryptography is the science of using mathematics to encrypt and decrypt data. It enables us to store or transmit sensitive information across insecure networks like the Internet. So that it cannot be read by anyone except the intended recipient [5].

Cryptography is one of the technological means to provide security to data being transmitted on information and communications systems. Cryptography is especially useful in the cases of financial and personal data. Hence, information security is a precondition of e-application systems when communicating over un-trusted medium like the Internet. The most effective way of data protection is encryption. A cryptographic system which provides two complementing functions, encryption and decryption is called cryptosystem.

Cryptosystems use encryption algorithms to determine the encryption process, the necessary software component, and the key to encrypt and decrypt the data. Cryptographic techniques are always employed to protect critical and confidential information against malicious attack from the intruders. There are two main types of cryptographic algorithms: symmetric key and asymmetric key cryptography [4].

In this paper, we propose a cryptosystem for web documents (like xml, html, jsp, php, js etc) data encryption/decryption by combining the features of both symmetric key and asymmetric key [1]. Here we use both Vigenere cipher algorithm and ELGamal cryptosystem together with hashing is proposed. Also hash value is encrypted to generate the signature and the signature is transmitted together with the message. The receiver extracts the signature and decrypted to obtain the hash. The receiver also calculates the hash value of the received file. The two hashes are compared. If they are same, no modification takes place to the transmitted file. Otherwise data is modified during transmission.

The rest of the paper is organized as follows: Section 2 deals with related works, section 3 deals with the motivation,

Section 4 gives the system model, section 5 gives an overview of cryptography fundamentals, followed by an explanation of algorithms used for cryptography to the proposed work in Section 6. Section 7 presents our proposed system, and Section

8 introduces the results, and finally we conclude the paper in section 9.

Related Work:

Since the turn of the millennium, Working Groups of

the W3C have been concentrating on the development of XML based security standards, which are paraphrased as XML Security. XML Security consists of three recommendations:XML(Digital)Signature, XML Encryption and XML Key Management Specification (XKMS), all of them published by the W3C [10]. The eXtensible Markup Language (XML) is a markup language promoted by the World Wide Web consortium (W3C). XML plays an important role in the exchange of wide variety of data over the web.

XML is being used across the Internet to improve compatibility between disparate Electronic Data Interchange (EDI) systems. XML designed to meet the challenges of large-scale electronic publishing. It plays an important role in the exchange of a wide variety of data on the Web..XML overcomes the limitations of hypertext markup language (HTML) and represents an important opportunity to solve the problem of protecting information distributed on the Web, with the definition of access restrictions directly on the structure and content of the document. Unless the system is able to provide some mechanisms to ensure

security services, the system will have problems to be accepted. More reliable cryptosystems have to be proposed and, cryptography is being an essential part of today's information systems.

There were methods to protect the XML documents. One such method used the RSA with some padding scheme [2]; it is extremely difficult to factor large numbers. The property of shift ciphering scheme increases the cost of crypto-analysis. The results are very much satisfactory for securing XML data. They found the estimation required time to break our generated keys is 2502 year s, which is sufficient against any brute-force attacks.

Another one proposed a cryptosystem (encrypting/ decryption) for XML data using Vigenere cipher algorithm and ELGamal cryptosystem. Such a system is designed to achieve some of security aspects such as confidentiality, authentication, integrity, and non-repudiation. They used XML data as an experimental work. Access control techniques for XML provide a simple way to protect confidential information at the same granularity level provided by XML schemas [7]

Motivation:

Data are transferred via Internet through the web documents. So we have to protect these documents from malicious attacks. In our existing systems, the receiver does

not know whether the data is modified or not. They only encrypted the data and transmitted .The receiver decrypt the data with the assumption that they receive the original data. Also these systems provide protection only for XML documents. Also regarding the malicious attacks of xml and

other web related docs [12] , a more secure system is essential. So we need a new method to protect all web documents that carry valuable information.

System Model:

For implementing the cryptographic algorithms, JAVA with SDK version 1.7 is used. For simulation NetBeans IDE 7.0.1 is used. For system evaluation, the web documents are collected from [6] and [9].

Overview Ocryptography:

The word cryptography originated from two Greek words, kryptos which means secret and graphos which means writing; hence it literally means secret writing. In particular, cryptography may be thought of as the science of secret writing, aiming at protecting data so that only the intended recipients may decrypt and read the message.

A cryptosystem is composed of two complementing functions, encryption and decryption. Encryption is the conversion of data into a form, called a ciphertext that cannot be easily understood by unauthorized people based on input key. Decryption is the process of converting encrypted data back into its original form, so it can be understood using the decryption key. Encryption and decryption keys are the same for symmetric cryptosystem and different for asymmetric cryptosystem [1].

Cryptosystems are used to achieve several goals such as:

• Confidentiality is the process of keeping information private and secret so that only the intended recipient is able to understand the information.

• Authentication, which is the process of providing proof of identity of the sender to the recipient, so that the recipient can be

assured that the person sending the information is who and what he or she claims to be.

• Data integrity is a service which addresses the unauthorized alteration of data. To assure data integrity, one must have the ability to detect data manipulation by unauthorized parties. Data manipulation includes such things as insertion, deletion, and substitution.

• Non-repudiation is a mechanism used to prove that the sender really sent this message. This is achieved by using a digital signature mechanism.

A fundamental goal of cryptography is to adequately address these four areas in both theory and practice. These are usually achieved through data encryption mechanism. As cryptography is the science and art of creating secret codes, cryptanalysis is the science and art of breaking those codes, with a view to finding weaknesses in them that will permit retrieval of the original data from the encrypted data, without necessarily knowing the key or the algorithm.

Cryptography and cryptanalysis are two different scientific studies in direct competition with each other, the first attempts to hide a secret and the latter attempts to uncover it [5]. To ensure the security of the message, the original message is transformed to ciphertext using an encryption algorithm by the sender. And the receiver uses a decryption algorithm to transform the ciphertext back into plaintext. Encryption and decryption algorithms are called ciphers. And those algorithms operate on a set of numbers called Key. To encrypt message, we need an encryption algorithm, encryption key and the plain text. These create the ciphertext. Similarly to decrypt a message,
we need a decryption algorithm, decryption key and the ciphertext. These reveal the plaintext.

Algorithms Used For Cryptography:

There are three types of cryptography algorithms: Symmetric key or Secret key Cryptography, Asymmetric key or Public key Cryptography and hash functions.

a) Symmetric Key Cryptography

In Symmetric Key Cryptography, the same key is used by both sender and receiver. To provide privacy, this key needs to be kept secret. The traditional ciphers substitution cipher and transposition cipher. A substitution cipher substitutes one symbol with another. And the transposition cipher does not replace the original text with different text, but moves the original text around. The most popular secret key encryption algorithms are Data Encryption Standard (DES), Triple DES, and Advance Encryption Standard (AES) [5]. Here Vigenere cipher is used for the proposed work due to its simplicity and also we can extend the Vigenere cipher [13].

Vigenere Cipher*:* An example of symmetric key is the Vigenere cipher. It's a symmetric cryptosystem which is not
monoalphabetic, This cipher is named after Blaise de Vigenere, who lived in the sixteenth century [6]. We would
use the Vigenere Cipher (with a modulus of 26) to encrypt ordinary English text by setting up a correspondence between alphabetic characters and residues modulo 26 as follows: A $\leftrightarrow$ 0, B $\leftrightarrow$ 1,..., Z $\leftrightarrow$ 25. The Vigenere cipher is defined as the following:

Let m be a +ve integer. Define P(Plaintext)=C(Ciphertext)=K(Keys) $=(Z_{26})^m$. For a key K=$(k_1, k_2, \ldots\ldots\ldots k_m)$,

define

$e_k(x_1, x_2, \ldots, x_m) = (x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m)$ and

$d_k(y_1, y_2, \ldots, y_m) = (y_1 - k_1, y_2 - k_2, \ldots, y_m - k_m)$

where all the operation are performed in Z_{26}. The number of possible keywords of length m in a Vigenere Cipher is 26^m, so even for relatively small values of m, an exhaustive key search would require a long time. For example, if we take m = 5, then the key space has size exceeding1.1 x10^7. This is already large enough to preclude exhaustive key search by hand (but not by computer).

DES –The Data Encryption Standard Designed at IBM during the 1970's and officially adopted as the NIST standard encryption algorithm for unclassified data. The DES algorithm takes 56 bit and 64 bit plaintext as inputs and outputs 64 bit. In terms of strength its design has stood the test of time very well, but its relatively short key length by modern standards means that it is now considered vulnerable to brute force attacks. It is also, in software, comparatively slow. To encrypt the hash value, DES is used in this work.

B. Asymmetric key cryptography

This type uses two keys namely: a) private key and b) public key. The public key is used to encrypt messages whereas the private key is used to decrypt them. The public encryption key is made available to
who wants to use it, but the private key is kept secret by the key owner. The process is explained below: If A wants to send a message to B, the message is encrypted by a using B's public key. If B receives the message, the message is decrypted by using B's private key. No other recipient can decrypt the message. The most popular public key encryption algorithms are RSA (Rivest, Shamir, and Adleman) and EL Gamal cryptosystem [5].

El Gamal Cryptosystem*:* An example of asymmetric system is El Gamal cryptosystem. Before explaining the system, we will explain the following three definitions [4].

Definition 1: Let a, n are relatively prime (gcd(a,n)=1), then there's at least one integer m that satisfies $a^m \bmod n = 1$. m is referred as the order of a (mod n).

Definition 2: If p is a prime number. An element α having order p-1 is called a primitive element modulo p.

Definition 3: Let p is a prime number and α is a primitive element modulo p. Any element $\beta \in Z_p$ can be written as $\alpha^i = \beta, 0 \le i \le p-2$ in a unique way i.e., $\alpha^i = \beta \bmod p$ where i is called the unique **discrete logarithm**.

ElGamal cryptosyetm is defined as the following: Let p be a prime such that the Discrete Logarithm problem in (Z_p) is infeasible, and let $\alpha \in Z_p$ be a

primitive element. Let $P = Z_p, C = Z_p \text{ x } Z_p$ .Define $K = (p, \alpha, a, \beta) : \beta = \alpha^a \bmod p$ .The values p, α and β are the public key, and a is the private key(own to the receiver). For $K = (p, \alpha, a, \beta)$ and for a random number $k \in Z_p$,define $e^k(x,k) = (y_1, y_2)$,where $y_1 = \alpha^k \pmod p$ and $y_2 = x^{\beta^k} \pmod p$.

For $y_1, y_2 \in Z_p$, define $d_k(y_1, y_2) = y_2 (y_1{}^a)^{-1} \bmod p$ .Informally, this is how the ElGamal Cryptosystem works: The plaintext x is masked by multiplying it by β^k , yielding y_2 .The value a^k is also transmitted as part of the ciphertext. The receiver who knows the private key (a) can compute β^k from a^k . Then he can remove the mask by dividing y_2 by β^k to obtain x. Hence, a necessary condition for the ElGamal Cryptosystem to be secure is that the Discrete Logarithm problem in Z_p is infeasible. This is generally regarded as being the case if p is carefully chosen and a is a primitive element modulo p.

The encryption algorithms like Vigenere, DES, ElGamal etc generally allow the encryption of a whom message by one

person, we call Alice, and the decryption of the message encrypted by another person, with the generic name Bob. But there is an extension of ElGamal is available [11], which is based on the following scheme: Alice is encrypting a message which she is simultaneously sending to the persons Bob1,Bob2, ..., Bob2n+1. The 2n+1 persons (Bob1, Bob2, ..., Bob2n+1) will be able to decrypt the message received from Alice only if they are together, separately this operation is impossible for them.

Hash Functions:One of the fundamental primitives in modern cryptography is the cryptographic hash function; a hash function is a one-way hash is a function (usually mathematical) that takes a variable-length string, a message, and compresses and transforms it into a fixed-length value referred to as a hash value. A hash value is also called a message digest. Here, hashing is used to perform one way encryption. One way means that once the information has been encrypted there is no way to retrieve the original information from the hashed form. The most common cryptographic uses of hash functions are with digital signatures and for data integrity . Hash algorithms that are in common use today include:

Message Digest Service Algorithms (Md) - The message digest kinds of encryption algorithms provide encryption of 128 bits in strength and are designed to be fast and simple. Current standards are MD2, MD4 and MD5. In this paper we use MD5 hashing. Most commonly used present-day message digest algorithm is the 128 bit MD5 algorithm. It was developed by Ron Rivest of the MIT Laboratory for Computer Science and RSA Data Security, Inc.

MD5 is an algorithm which takes an input of any length and outputs a message digest of a fixed length (128-bit, 32 characters).MD5 uses the same algorithm every time. Hence it will always generate the same message digest for the same string (data). MD5 hashes have the advantage of generating completely different looking hashes from seemingly similar inputs. It is a one way hash function.A one way hash means that the message digest which is outputted by the MD5 algorithm is irreversible. These message digest have certain advantages:

1)The generation of a digest is very fast and the digest itself is very small and can easily be encrypted and transmitted over the internet.

2)It is very easy and fast (and therefore cheap) to check some data for validity.

3)The algorithms are well known and implemented in most major programming languages, so they can be used in almost all environments

The Proposed System:

The basic idea of our proposed cryptosystem is using the combination of both ElGamal Cryptosystem and Vigenere cipher. The block diagram is given in Figure 1 and Figure 2 for encryption or decryption for sender side and receiving side respectively.

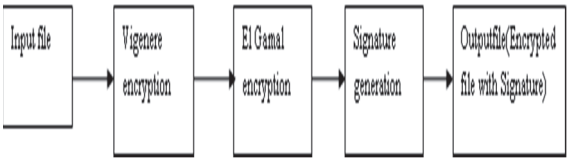

Figure 1: Encryption, Sender side

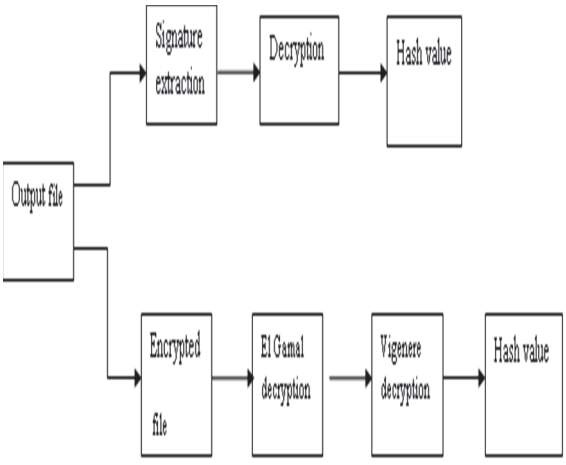

Figure 2:Decryption, Receiving side

The two hash values are generated at the receiving side and are compared. If they are same, then it is concluded that it is valid signature. The proposed system is composed of several modules. The following subsections explain each module.

a) Encryption using Vigenere cipher

The following steps to be followed for encryption using Vigenere cipher using the Vigenere table constructed for all the 256 ASCII characters:

1)Apply the Vigenere cipher for the message. To derive the cipher text, take the intersection of the row given by the corresponding keyword letter and the column given by the plaintext letter.

2)Repeating keyword is used.

b) Encryption using El Gamal

In the encryption process, the sender performs the following steps:

1)Key Generation for El Gamal Cryptosystem.

Chose p as a prime such that the Discrete Logarithm problem in (Z_p) is infeasible, and let $\alpha \in Z_p$ be a primitive element. DefineK= $(p,\alpha,a,\beta):\beta=\alpha^a(\mathrm{mod}\,p)$ The values p, α and β are the public key, and a is the private key.

2) Apply El Gamal cryptosystem to the result of *A*.

c) Digital Signing

In the production of the Digital signature process, the sender generates the key and does the following:

1) Encrypt the hashed file(encryption algorithm used is DES) to get the signature.

2) Signature is transmitted

At the receiver, Signature is extracted using DES decryption process.

d)Decryption

In the Decryption process, the receiver performs the following steps:

1) Use the decryption function of ElGamal cryptosystem to decrypt the message.

2)Use the decryption function of the Vigenere cipher to decrypt the result of step 1.

e) Verification

In the verification process, the receiver uses the public key of the sender and does the following steps:

1) Signature is decrypted to get the hash.

2) Calculate the hash value of received file and compare two hashes.

The sender is sending the hash value. During this process, a new hash value is generated and both the hash values are compared. If both are same, then it is verified that it is valid signature.

Experimental Results:

We have tested our cryptosystem through exchanging encrypted documents. The sender encrypt the file and generates the signature and transmit it to the receiver. The receiver received an encrypted document and the signature, and then the signature is extracted and decrypted it to get the hash value. Decryption of the document got the original document. Receiver also calculates a new hash value for the received document. At the receiver two hashes are compared, if both are same then data transmitted successfully i.e. no modification to the document while transmission.

Also, we have tested the cryptosystem in another way, where we changed some contents of the document after encrypted it, and we transmitted this document into a user in another place (receiver). The receiver decrypted the document, and by hash comparison they detected that document has been altered.

Conclusion:

Security has always been important in electronic applications. Cryptography techniques are employed to protect critical and confidential information against malicious attack from the intruders. The security of a cryptographic system depends heavily on the strength of its keys. In this paper, we have proposed a cryptosystem for encrypting/decrypting web documents .Also hashing is added to the output of the double encryption performed using Vigenere cipher and El Gamal encryption. The proposed system can be used for secure transmission of any kind of web documents through different networks such that any malicious attacks on the documents can be found and the remedial
measures will be taken consequently.

References :

1. *A*. A. Abd EL-Aziz and A.kannan "A Cryptosystem for XML documents". In Proceedings of the 2012 *IEEE International Conference on Computer Communication and Informatics(ICCCI-2012),* January 10-12,2012,Coimbatore, INDIA
2. Abdelsalam Almarimi and Uounis Alsahdi "*Developing a cryptosystem for XML documents*". In Proceedings of the 2nd IEEE International Conference on Computer Technology and Development (ICCTD), pages 240 – 244, 2-4 Nov. 2010ry 2012.
3. Ali Obaid, F. Khalifa, "A Modified One-Time Key Method for Practical Unbreakable Ciphering". *In the Proceedings of the National Conference for IT & Communications*. Pp. 180-184, May 2008 .
4. B. Forouzan, "Cryptography and Network Security", 2008.
5. B. Schneier, John Wiley & Sons, "Applied Cryptography: Protocols, Algorithms, and Source

Code in C", 1996.
6. Childlovskii, Bergholz Andre, "Crawling for domain-specific hidden Web resources", *In the Proceedings of the fourth IEEE International Conference on 10-12 Dec 2003*,Pages 125-133.
7. Damiani E, di Vimercati P S, Paraboschi S, *"Controlling Access to XML Documents",* IEEE Internet Computing *Dec.2001, pages:18-28*
8. Douglas R. Stinson. "*CRYPTOGRAPHY Theory and Practice*.",2006P.
9. De Sousa, Artur Jorge Afonso, Pereira, Jose Luis, Carvalho, Joao Alvaro,"*Querying XML documents"* In the Proceedings of the 22nd IEEE International Conference, 2002.
10. Ekelhart, A. et al., XML security – A comparative literature review, Publisher, Elsevier Science Inc. New York, NY, USA. ISSN: 0164-1212, Vol: 81, 10, October, 2008. pp. 1715-1724
11. Flonta, Stelian, Miclea, Liviu-Cristian Cristian,An extension of the El Gamal encryption algorithm, *In Proceedings of the IEEE International Conference on May 2008,Vol. 3,pages:444-446*
12. Louw, Mike Ter, Venkatakrishnan, V.N, "*Blueprint:robust Prevention of Coss-site Scripting Attacks for Existing Browsers", ",* In the Proceedings of the security and privacy, 2009 30th IEEE conference on 17-20 may 2009.
13. Md. Khalid Imam Rahmani, Neeta Wadhwa1 and Vaibhav Malhotra1,"*alpha-qwerty cipher: an extended vigenère cipher*",An International Journal (ACIJ), Vol.3, No.3, May 2012.

* * *

M.Tech Student, Department of Computer Science & Engineering, KMEA Engineering college
Edathala,Aluva *Email id: chinnuprasenan@gmail.com*
Assistant Professor, Department of Computer Science & Engineering
KMEA Engineering college
Edathala, Aluva
Email id: mariajoy@kmeacollege.ac.in