SYMMETRIC AND ASYMMETRIC APPROACH FOR XML DOCUMENTS SECURITY

Sowjanya Ragidi[1], B Venkat Suresh Reddy[2]

[1] *M. Tech Scholar (CSE), Nalanda Institute of Tech. (NIT), Siddharth Nagar, Guntur, A.P, (India)*
[2] *Assistant Professor, Nalanda Institute of Technology (NIT), Siddharth Nagar, Guntur, A.P, (India)*

ABSTRACT

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. Then information security is a precondition of e- application systems when communicating over untrusted medium like the Internet. Main effective way of data protection is encryption. Cryptography system which providing two complementing Functions such as encryption and decryption are called cryptosystem. We have three types of cryptography algorithms: Symmetric Key or Secret key Asymmetric Key or Public Key, Cryptography, hash function. In present day's web applications are using this format only. Now we propose a cryptosystem (encrypting/decryption) for XML data using Vigenere cipher algorithm and EL Gamal cryptosystem. So we need a system to acquire some of security aspect like Validation, Verification, Authentication, Privacy, Reliability and Data Redundancy. We used XML data as an experimental work, therefore we have used Vigenere cipher which is not mono alphabetic thus the number of possible keywords of length m in a Vigenere Cipher is 26^m. It is the science of using mathematics to encrypt and decrypt data. Rapid growth of the Internet has made cryptography is more important and critical issue in electronic application systems. Thus that it cannot be read by anyone except the intended recipient. Since many cryptographic techniques and algorithms are well-defined such as RSA, DES and AES. The Cryptosystem for Extensible Markup Language (XML) data Encryption/Decryption by combining the features of both symmetric Key and asymmetric key cryptography. The Proposed technique used XML due to the importance of XML in data exchange in distributed systems. XML designed to achieve the challenges of large-scale electronic publishing and it plays an important role in the exchange of a wide variety of data on the Web.

Keywords— *Vigenere Cipher, Caesar Cipher, XML, XML Granularity, Cryptography*

I. INTRODUCTION

Cryptography is a study of masking information. The term 'Cryptography' is referred from the Greek word "Kryptos" meaning hidden. The proposed technology Cryptography is finest technology in communication security. The growth of the Internet has made cryptography is more important and critical issue in electronic application systems as well as it's become a basic building block for computer security. Unless the system is able to provide some mechanisms to ensure security services, it will have problems to be raised in data sharing. So it is necessary to define most consistent crypto mechanisms have to be proposed and it is become an essential part of today's information systems. Hence it is the science of using mathematics to encrypt and decrypt data. Cryptosystems providing a secure mechanism to store or transmit sensitive information across insecure networks like the internet. Here the original message which is sent by the sender is known as plaintext though

the encoded message is called the cipher text and the process of converting plaintext (XML File) to cipher text (string or decimals) is known as encryption and again at receiver side the process of reconverting the plaintext from the cipher text is called decryption. This system uses encryption algorithms to determine the encryption process, the required software component and the key to encrypt and decrypt the data. This is one of the technological means to provide security to data being transmitted on information and communication systems. It is essential in the cases of financial and personal information systems. Cryptography techniques are always employed to protect critical and confidential information against malicious attack from the intruders/attackers. Here some techniques and algorithms are well defined such as AES, DES and RSA. Here the proposed cryptosystem for Extensible Markup Language (XML) data encryption/decryption by combining the features of both symmetric key and asymmetric key cryptography techniques.

II. LITERATURE SURVEY

2.1 Cryptography

Cryptography is defined as the science of secret writing and aiming at protecting data so that only the intended recipients may decrypt and read the message. Cryptography includes two techniques Encryption and Decryption. We can define the *Encryption* is a mechanism of conversion of data into a form called as cipher text that cannot be understood by unauthorized people based on input key. The *Decryption* is a mechanism that converting encrypted data into its original data, so that it can be readable or understandable by the receiver by using the decryption key. In present systems, cryptographic techniques are used for text files only but not XML files. Algorithms such as RSA, HMAC, AES, DES, and SHA1 are provides less security so that the intruder can break the cipher text by using Brute force techniques. The information which is transferring is having less security, so that receiver cannot get original information Cryptographic algorithms are classified into two types: *Symmetric cryptography* and *asymmetric cryptography.*

Symmetric Cryptography: It is a form of cryptosystem in which encryption and decryption are performed using the same key (Fig 1). This mechanism also known as conventional encryption.

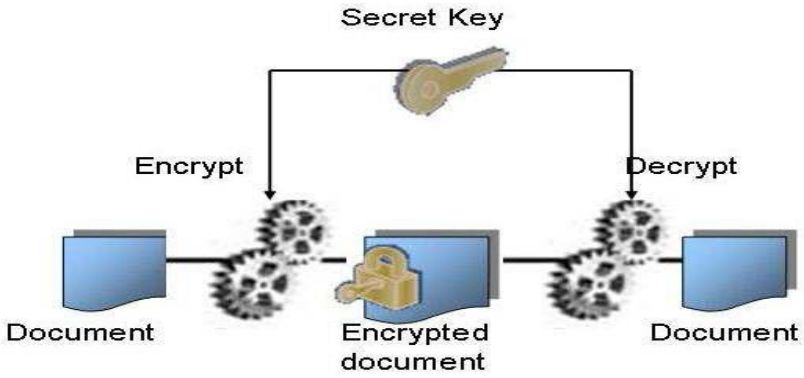

Fig 1: Symmetric Key Encryption

Asymmetric Cryptography: It is a form of cryptosystem in which encryption and decryption are performed using the different keys- one a public key and one a private key (Fig 2). It's also known as public key encryption.

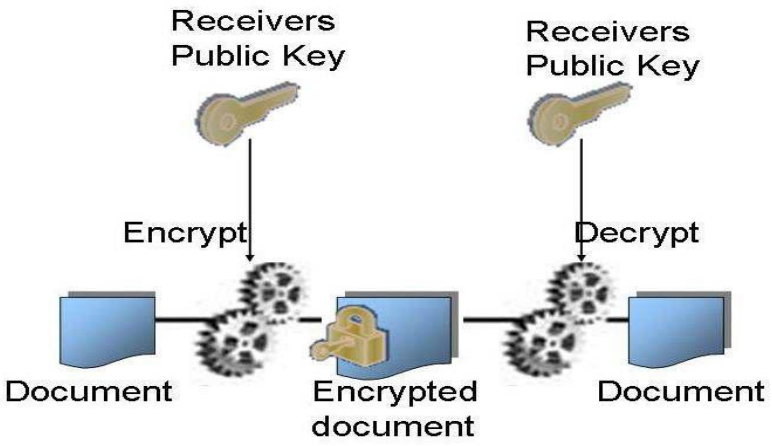

Fig 2: Asymmetric Key Encryption

Cryptosystems having the following aims:

- *Authentication:* Authentication provides identity verification of the sender to the recipient; such that the recipient can be assured that the person sending the information is who and what he or she claims to be.
- *Confidentiality:* Confidentiality is used to maintain information private and secret so that only the intended recipient is able to understand the information.
- *Data Integrity:* Integrity of the data means that the unauthorized alteration of data. For assure data reliability, receiver of the data must have the ability to detect data manipulation by unauthorized parties. It includes such things as modifying, inserting, deleting and substitution of data.
- *Non-Repudiation:* By this non-repudiation process we prove that the sender really sent this message and it is achieved by using a digital signature mechanism.

2.2 Extensible Mark-Up Language (XML)

Extensible Markup Language is an approach to delivering sensitive data over the internet. This is a markup language that is developed by the World Wide Web consortium to conquer the Hypertext Markup Language (HTML) drawbacks. XML is a language for describing data on the web. XML document contains tags called as Mark-ups, it describe the content of the document. It is extensible so that it can be used to create many different applications. It uses human, not computer, language. It is readable and understandable and no more difficult to code than HTML.

XML having the following features:

- XML is completely compatible with Java™ and 100% portable. Any application that can process XML can use your information for any type of platform.
- XML is extendable. Creating own tags, or we can use tags created by others, so that it use the natural language of your domain. XML have the attributes you need, and that provides flexibility to you and users.
- XML is suitable for all applications such as JAVA, and it can be combined with any application which is capable of processing XML irrespective of the platform it is being used on.
- XML is very portable language to the extent that it can be used on large networks with multiple platforms like the internet, and it can be used on handhelds or palmtops or PDAs.

XML having the following advantages:

- It is a platform independent language.
- It can be deployed on any network if it is amicable for usage with the application in use.
- If the application can work along with XML, then XML can work on any platform and has no boundaries.
- It is also merchant independent and system independent. When the data is being exchanged using XML hence there will be no loss of data.

2.3 Vigenere Cipher Algorithm

In this algorithm alphabet, digits and special symbols are included which are used for the key value generation. Sender must give a key value, which should follow this condition, length of XML file = length of Key. Suppose the sender gives the key value less than the length of file then, automatically the remaining value is generated by the algorithm. Then the plain text is converted into cipher text using vignere cipher algorithm.

2.4 El - Gamal Algorithm

The EL-Gamal algorithm is an asymmetric key encryption algorithm for public-key cryptography which is based on the Diffie–Hellman key exchange (Fig 3). For generating the cipher values for the plain text, we need some formulae,

- C1= $e_1^{r \bmod}$ prime
- C2= (P* $e_2^{r) \bmod}$ prime

Where C1, C2 are cipher values and e1, e2 and prime are public keys and D is a private key.

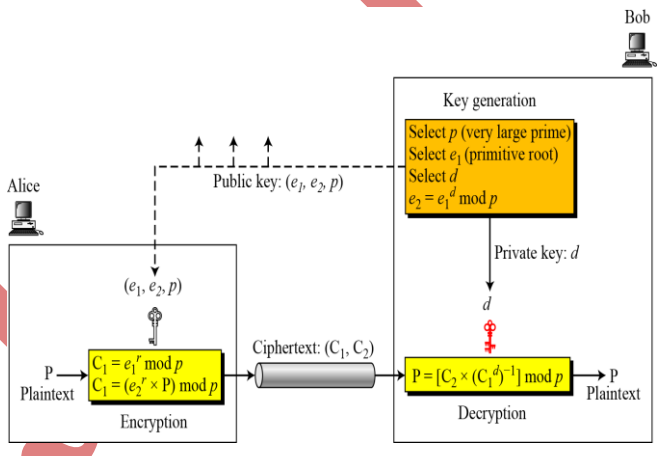

Fig 3: Key Generation, Encryption and Decryption In EL-Gamal

2.5 Digital Signature

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

i. Use El Gamal signature schema to generate the digital signature S for the digest

ii. Use El Gamal cryptosystem to encrypt the Signature S and the digest h and send the output to the receiver.

iii. Decrypt the message to get the signature S and the digest h. Using the signature receiver can identify whether the file is hacked or not.

2.6 Hash Function

One of the fundamental primitives in modem cryptography is the cryptographic hash function. The purpose of a hash function is to produce a blueprint of a file, message, or other block of some data. A hash value h (digest) is

generated by a function H of the form: h = H (M) where M is a variable-length message and H (M) is the fixed Length hash value. Hash value is appended to the message at the source at a time. While the message is assumed or known to be correct. Receiver authenticates that message by recomputing the hash value. Since the hash function itself is not considered to be secret, it means is required to protect the hash value to be useful for message authentication.

Properties of Hash Function H:

- H can be applied to a block of data of any size.
- H produces a fixed-length output.
- H(x) is relatively easy to compute for any given x, making both hardware and software implementations practical.
- For any given value h, it is computationally infeasible to find x such that H(x) is equals to value of h. This is sometimes referred to in the literature as the one-way property.
- For any given block x, it is computationally infeasible to find y x such that H(y) = H(x). This is sometimes referred to as weak collision resistance.
- It is computationally infeasible to find any pair (x, y) such that H(x) = H(y). This is sometimes referred to as strong collision resistance.

III. PROPOSED METHODOLOGIES

In our proposed cryptosystem we are using the combination of both El Gamal Cryptosystem and vigenere cipher.

3.1 El Gamal Cryptosystem Key Generation

Chose p as a prime such that the Discrete Logarithm problem in (Zp) is infeasible, and let $\alpha \in Zp$ be a primitive element. Define K= (p, α, a, β): $\beta = \alpha^a \pmod p$. The values P, α and β are the public key and a is defined as the private key.

3.2 Encryption

In the encryption process, the sender does the following:

i. Apply the vigenere cipher for the message.

ii. Apply El Gamal cryptosystem to the result of step i.

3.3 Decryption

In the Decryption process, the receiver does the following:

i. Use the decryption function of El Gamal cryptosystem to decrypt the message.

ii. Use the decryption function of the vigenere cipher to decrypt the result of step 1.

3.4 Digital Signing

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

i. Decrypt the message to get the signature S and the digest h.

ii. Use El Gamal signature schema to generate the digital signature S for the digest h.

iii. Use El Gamal cryptosystem to encrypt the Signature S and the digest h and it sends output to the receiver.

3.5 Verification

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

i. Use the decryption function of El Gamal cryptosystem to decrypt the message and get the signature S and the digest h.

ii. Apply the verification process using S and h.

iii. If the result is true, then valid signature.

3.6 Vigenere Cipher Algorithms

The vigenere cipher is defined as the following: Let m be a +ve integer. Define P (Plaintext) =C (Cipher text) =K (Keys) = $(Z_{26})^m$. For a key K= (k_1, K_2, k_m), we define e $k(x_1,x_2,...,x_m)=(x_1+k_1,x_2+k_2,..,x_m+k_m)$ and $d_k(y_1,y_2,...,y_m)=(y_1-k_1,y_2-k_2,..., y_m-k_m)$ Where all the operation is performed in Z_{26}. The number of possible keywords of length m in a Vigenere Cipher is 26^m, thus even for relatively small values of m an exact key search would require a long time. For example, if we take m = 5, then the key space has size exceeds 1.1×10^7.

IV. CONCLUSIONS

The techniques like encryption and decryption are time taking for element content and character size is in between 50kb to 500kb. And here security has always been important in electronic applications while transmitting the data. If we use encryption and decryption techniques for information security, while the file size increases the encryption and decryption time increases. The proposed cryptography techniques are employed to protect critical and confidential information against malicious attack from the intruders. Security of a cryptographic system depends heavily on the Strength of its keys. In this, the Cryptosystem for encrypting/decrypting XML documents uses three algorithms to provide more security. So, Intruder takes number of years to break cipher text because Plain text is converted into cipher text in two stages. Receiver easily detects whether the Data is modified or not by using digital signature.

REFERENCES

[1] Abd EL-Aziz Ahmed Abd EL-Aziz and A.kannan “A Cryptosystem for XML documents” on Internatioal Conference on Computer communication and informatics (ICCCI-2012).

[2] Nithin N, Harshitha.K.S, Divyashree K and Shruti.N.Nayak “ Analysis of Symmetric algorithm for XML document security” on International Journal of Innovations in Engineering and Technology (IJIET).

[3] Timothy K. Shih Department of Computer Science and Information Engineering, Tamkang University “Cryptosystem Applications in Mobile Agent Security” on Journal of Security Engineering.

[4] Abdelsalam Almarimi and Uounis Alsahdi. Developing a cryptosystem for xml documents. In Proceedings of the 2nd International Conference on Computer Technology and Development (ICCTD), pages 240 – 244, 2-4 Nov. 2010.

[5] Janailin Warjri, Dr.E.Gerorge Dharma and Prakash Raj, “Analysis of Symmetric Key Algorithms”, *International journal of societal applications of computer science*, Vol. 2, Issue. 9, pp. 454-457.

AUTHOR PROFILE

Sowjanya Ragidi is currently pursuing M.Tech in the Department of Computer Science & Engineering, from Nalanda Institute of Technology (NIT), siddharth Nagar, Kantepudi (V), Sattenapalli (M), Guntur (D), Andhra Pradesh, and Affiliated to JNTU-KAKINADA.

B Venkat Suresh Reddy working as Assistant Professor at Nalanda Institute of Technology (NIT), siddharth Nagar, Kantepudi (V), Sattenapalli (M), Guntur (D), Andhra Pradesh, Affiliated to JNTU-KAKINADA.