A Cryptosystem for XML documents

A. A. Abd EL-Aziz
Research Scholar
Dep. of Information Science & Technology
Anna University
Email: zizoah2003@gmail.com

A.kannan
Professor
Dep. of Information Science & Technology
Anna University
Email: kannan@annauniv.edu

***Abstract*—In this paper, we propose a cryptosystem (encrypting/decryption) for XML data using Vigenere cipher algorithm and EL Gamal cryptosystem. Such a system is designed to achieve some of security aspects such as confidentiality, authentication, integrity, and non-repudiation. We used XML data as an experimental work. Since, we have used Vigenere cipher which is not monoalphabetic, then 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.**

I. Introduction

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[2]. 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 untrusted medium like the Internet. The most effective way of data protection is encryption. A cryptography 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[3]. Cryptography techniques are always employed to protect critical and confidential information against malicious attack from the intruders. There are two main types of cryptography algorithms: symmetric key and asymmetric key cryptography [4]. There have been many cryptographic techniques and algorithms are well-defined in the literature such as DES, AES, and RSA [3]. In this paper, we propose a cryptosystem for Extensible Markup Language (XML) data encryption/decryption by combining the features of both symmetric key and asymmetric key cryptography. We used XML as an experimental work due to the importance of XML in data exchange in distributed systems. XML is being used across the Internet to improve compatibility between disparate Electronic Data Interchange (EDI) systems [1]. 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.

II. Overview of Cryptography

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[3]. Cryptosystems are used to achieve several goals such as:

- Confidentiality, which 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 A mechanism to prove that the sender really sent this message. This is achieved by using a digital signature mechanism.

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[3].

III. TYPES OF 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)[3].

*1) **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 [4]. 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(Plaintetx)=C(Ciphertext)=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 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 keyspace has size exceeding 1.1 x 10^7. This is already large enough to preclude exhaustive key search by hand (but not by computer).

B. Asymmetric key cryptography

This type uses two keys: a private key and a 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[3].

*1) **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 mod 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 α^i=β, 0 $\leq$ i $\leq$ p-2, in a unique way i.e., $\alpha^i \equiv \beta$ (mod)p, i is called the unique **discrete logarithm**.
El Gamal 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 x Z_p, and define K= (p,α, a, β) : $\beta = \alpha^a$ (mod p). The values p, α and β are the public key, and a is the private key(own to the receiver). For K = (p,α, a, β), and for a (secret) random number k $\in Z_{p-i}$, define e_k(x, k) = (y_1,y_2), where y_1=α^k (mod p) and y_2=xβ^k (mod p). For y_1,y_2 $\in$ Z_p, define d_k(y_1,y_2)= y_2 $(y_1{}^a)^{-1}$ mod p. Informally, this is how the El Gamal 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. Clearly the El Gamal Cryptosystem will be insecure if intruder can compute the value a = $\log_\alpha(\beta)$, for then the intruder can decrypt ciphertexts exactly as the receiver does. Hence, a necessary condition for the El Gamal 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. For digital signature, El Gamal signature schema 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, A = Z_p x Z_{p-1}, and define K= (p,α, a, β) : $\beta = \alpha^a$ (mod p). The values p, α and β are the public key, and a is the private key(own to the sender). For K = (p,α, a, β), and for a (secret) random number k $\in Z_{p-i}$, define sig_k(x,k)=(γ,δ), where $\gamma = \alpha^k$ mod p and δ= (x-aγ)k^{-1} mod(p-1). For x,γ $\in$ Z_p and δ $\in$ Z_{p-1}, define Ver_k(x,γ,δ)= true if β^γ $\gamma^\delta \equiv \alpha^x$ mod p.

C. 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 fingerprint of a file, message, or other block of 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. The hash value is appended to the message at the source at a time when the message is assumed or known to be correct. The receiver authenticates that message by recomputing the hash value. Because the hash function itself is not considered to be secret, some means is required to protect the hash value. To be useful for message authentication, a hash function H must have the following properties[3]:

- 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) = 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**.

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.

IV. The Proposed System

The basic idea of our proposed cryptosystem is using the combination of both El Gamal Cryptosystem and vigenere cipher. The proposed system is composed of several modules. The following subsections explain each module.

A. 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. Define K= (p,α, a, β) : $\beta = \alpha^{a} \pmod p$. The values p, α and β are the public key, and a is the private key.

B. Encryption

In the encryption process, the sender does the following:

1) Apply the vigenere cipher for the message.
2) Apply El Gamal cryptosystem to the result of step 1.

C. Decryption

In the Decryption process, the receiver does the following:

1) Use the decryption function of El Gamal cryptosystem to decrypt the message.
2) use the decryption function of the vigenere cipher to decrypt the result of step 1 .

D. Digital Signing

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

1) Decrypt the message to get the signature S and the digest h.
2) Use El Gamal signature schema to generate the digital signature S for the digest h.
3) Use El Gamal cryptosystem to encrypt the Signature S and the digest h, and send the output to the receiver.

E. Verification

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

1) Use the decryption function of El Gamal cryptosystem to decrypt the message and get the signature S and the digest h.
2) Apply the verification process using S and h.
3) If the result is true, then valid signature.

V. 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 XML documents.

References

[1] 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.

[2] B. Schneier and John Wiley & Sons. *Applied Cryptography: Protocols, Algorithms, and Source Code in C*. 1996.

[3] W. Stalling. *Cryptography and Network Security: Principles and Practices*. Prentice Hall, 2006.

[4] Douglas R. Stinson. *CRYPTOGRAPHY Theory and Practice*. 2006.