A New Digital Envelope Approach for Secure Electronic Medical Records

M Gobi
Senior Lecturers, Department of Computer Science and Applications,
PSG College of Arts and Science, Coimbatore – 641 014.

Dr. K. Vivekanandan
Director, BSMED, Bharathiar University, Coimbatore – 641 046.

Summary

In the Metical sector doctors, patient and nurses need to have access to the medical records efficiently and in a secure manner. As information technology gets increasingly deployed in the medical sectors it becomes eminent that the medical records are stored electronically. Secure Electronic Medical Records (SEMR), which aims at providing a set of services which will provide secure and efficient access of the EMRs to the patients, doctors, nurses and insurance agents. The set of services that are provided by SEMR include Authentication, Authorization and Secure communication. In this paper, we suggest a implementation of a digital envelope that combines the hashing algorithm of MD5, the symmetric key algorithm of AES and the asymmetric key algorithm of Hyper Elliptic Curve Cryptography (HECC). The result illustrates that the best alternative digital envelope hybrid cryptosystem for EMR.

Keywords:

Authenticity, Integrity, Non-Reputability, SEMR, Electronic Medical Record, Advanced Encryption Standard (AES), Hyper-Elliptic Curve Cryptosystems (HECC), Message Digest version 5 (MD5), Symmetric Key, Asymmetric Key.

1. Introduction

In the medical sector the EMRs are of utmost importance. It is important that apart from the patients no one else should have complete control over the EMR of a particular patient. Also, such EMRs should be readily accessible to the doctors and patients. In order to create a secure environment for such a sector the minimum set of services that need to be implemented should provide Authentication, Authorization and Secure Communication.

i) Authentication involves identifying whether the user is a valid entity of the system. In order to access any resource in the system the user has to first authenticate himself first. One of the ways to implement the authentication service would be to ask for user identification in terms of username and password. Additional factors like an RSA token or fingerprint reading can be used to improve the security.

ii) Authorization involves identifying the rights of the users. Authenticated users should access only those resources to which they have access to. Also, the type of access (read, write, execute) determine the capability of the user to access that resource. This functionality can be implemented by creating access control list for the resources. These ACLs identify the type of access the user has to the resource.

iii) In a system like the medical sector which is a distributed in nature messages and data needs to be communicated amongst the different services. Thus it is important that the communication takes place in a securely. Such secure communications can be implemented by encrypting the data that is sent over the wire.

Apart from these basic services, auditing of all the events that take place at each of the servers forms an important component of the system. This auditing of events provides the administrators to analyze the past events and make changes to the system if required. They can also be used to trace an user's activity in the case of some intrusive activity performed by the user.

In the SEMR project we have used two-factor authentication method for authenticating the users. Further, each user has roles assigned to him which are used to identify the privileges he has for the resources. We have implemented a role-based access control for implementing the authorization function. For implementing the secure communications amongst the services we are using the OpenSSL library. In section 2 we describe the architecture used in our system. In section 3 we discuss the implementation details.

Manuscript received January 5, 2009
Manuscript revised January 20, 2009

2. Architecture

Our proposed Digital Envelop system has 3 major components –

i. MD5 Hashing Algorithm for secret key integrity checking.

ii. Encrypt the Patient History using AES Algorithm (Symmetric Cryptography)

iii. Encrypt the AES key using HECC Algorithm
(Asymmetric Cryptography)

As can be seen from the diagram, there are basically six interactions that take place in the system for one complete cycle from authentication to accessing an EMR, which is stored as a file. These interactions are

i. Doctor's Machine to generate the secret key for encrypt the Patient history.

ii. Doctor's Machine to digest the secret key using MD5.

iii. Doctor's Machine to encrypt the AES key using the doctor's private key and patient's public key (Asymmetric Encryption). The Doctor's send the Digital Envelop include the Cipher text of patient history, Cipher message of AES key and Message digest of the AES key to the patient.

iv. Patient Machine to decrypt the digital envelop using the patient's private key and Doctor's public key for the AES key.

v. The Patient Machine decrypt the cipher text of patient history using the AES key.

vi. Find the message digest of the AES key and compare to digital envelop message digest, both are equal the patient history is correct otherwise reject.

Additionally, on all the other events are logged into log files which can be further analyzed by the administrator.

3. Implementation

We have used Java for implementing the architecture described above. For the communication channels we have used the Java extensions to provide the encryption and Message digest. Also, for implementing the encryption and decryption of tokens we have used the Java Cryptography Extensions (JCE). Each of the service is described below in more detail .

AES Algorithm

Advanced Encryption Standard (AES) is an approved symmetric key cryptographic algorithm that is a block cipher. The AES algorithm is capable of using cryptographic keys of 128,192, and 256 bits.

The algorithm consists of four stages that make up a round which is iterated 10 times for a 128-bit length key, 12 times for a 192-bit key, and 14 times for a 256-bit key. The first stage "SubBytes" transformation is a non-linear byte substitution for each byte of the block. The second stage "ShiftRows" transformation cyclically shifts (permutes) the bytes within the block. The third stage "MixColumns" transformation groups 4-bytes together forming 4-term polynomials and multiplies the polynomials with a fixed polynomial mod (x^4+1). The fourth stage "AddRoundKey" transformation adds the round key with the block of data.

Details on AES can be had from (FIPS) and (AES 2001).

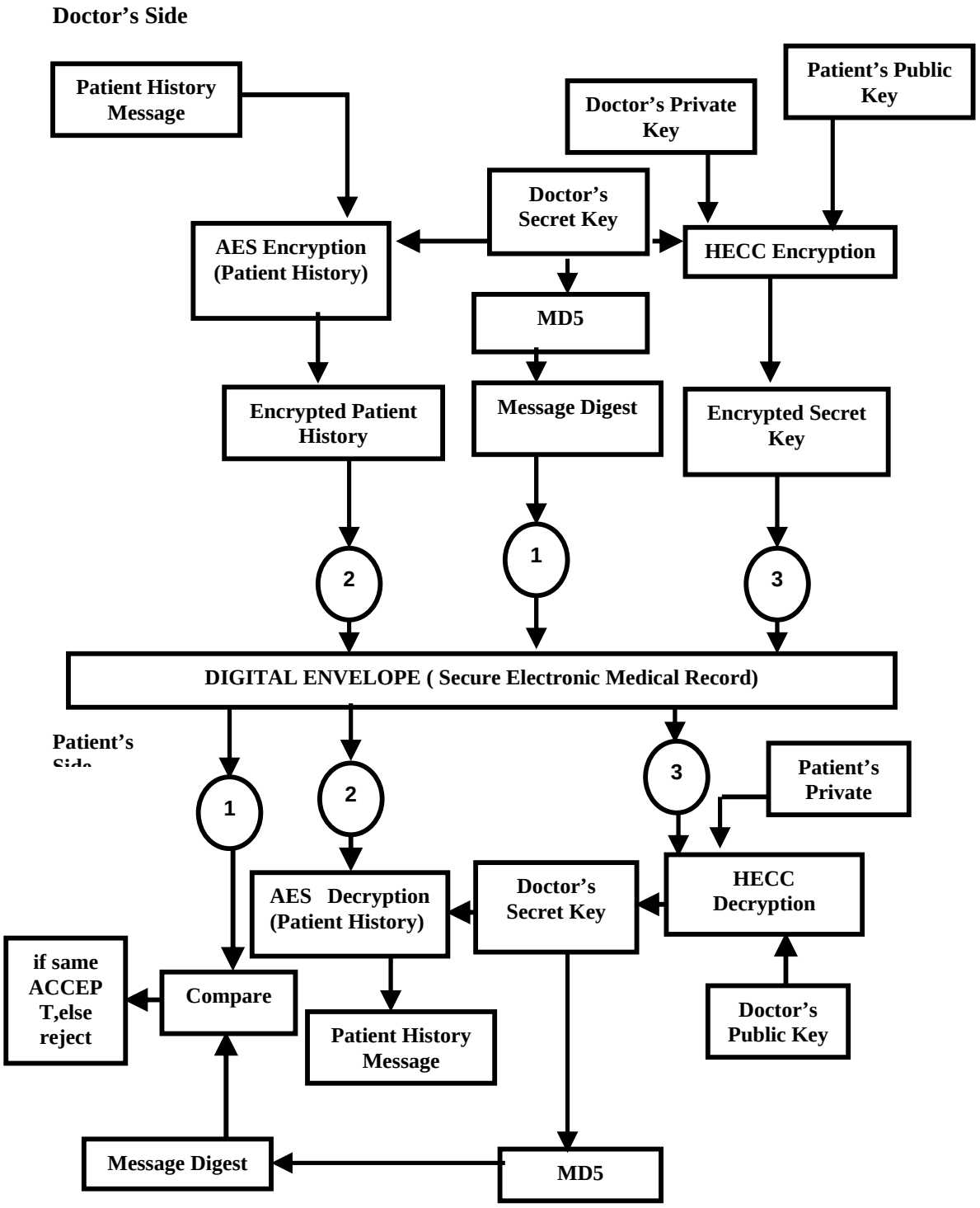
Hashing and Message Digest

A hash function takes a message of any length as input and produces a fixed length string as output, sometimes termed a message digest (Ilya M 2005).
The characteristics of a good hash function are:

Should avoid collisions.
Should try to spread keys evenly in the array.
Should be easy to compute.

The two most-commonly used hash functions are MD5 and SHA-1.

Fig 1: Architecture of SEMR



Basics of MD5

MD5 (**Message-Digest algorithm 5**), is an Internet standard (RFC1321) and is one of the widely used cryptographic hash function with a 128-bit message digest. This has been employed in a wide variety of security applications. The main MD5 algorithm operates on a 128-bit, divided into four 32-bit words. These are initialized to certain fixed constants. The main algorithm then operates on each 512-bit message block in turn, each block modifying the state. The processing of a message block consists of four similar stages, termed rounds; each round is composed of 16 similar operations based on a non-linear function, modular addition, and left rotation.

Hyper-Elliptic Curve Cryptosystems (HECC)

Overview of Hyper-elliptic Curve

The equation for a hyper-elliptic curve (C)is given as (Menezes, Wu & Zuccherato 1996):

$C : y^2 + h(x)y = f(x), h,f \in K[x], deg(f)=2g+1, deg(h) \leq g,$
f *is monic*
where genus g = (deg(f)-1)/2

Unlike elliptic curves, points on hyper-elliptic curves do not form a group. Hence, a group law is defined via the Jacobian variety of C over a field K, which is a finite abelian group.

Thus, a Hyper- Elliptic Curve (HEC) over Finite Field F_p is defined as:

$C: y^2 + h(x)y = f(x) (mod\ p),$
$h,f \in K[x], deg(f)=2g+1, deg(h) \leq g,$
f *is monic,*
where genus (g) = (deg(f)-1)/2

Jacobian of Hyper Elliptic Curve

The Jacobian of the curve C is the quotient group J=D°/P, where D° is the set of divisors of degree zero, and P is the set of divisors of rational functions. The equivalence classes of the Jacobian are represented by a unique reduced divisor (which is represented using Mumford representation) upon which we perform the group law.

Mumford representation

Let g be the genus of a hyper elliptic curve $C: y^2 + h(x)y = f(x)$. Each nontrivial divisor class over the field K can be represented via Mumford representation (u(x), v(x)), where u(x) and v(x), $u,v \in K[x]$, are unique pair of polynomials satisfying the constraints of

u is monic

$\deg v < \deg u \leq g$

$u \mid v^2 + vh - f$

Various mathematical operations can be carried out on these hyper-elliptic curves. Details can be had from can be had from (Duquesne & Lange 2006), (Eigeartaigh) , (Lange 2002) , (Menezes, Wu & Zuccherato 1996) , (Sakai & Sakurai 2000) , (Weng 2003).

Algorithm for a Hyper-Elliptic Curve Cryptosystem (HECC)

The basis for the Hyper-elliptic curve cryptosystem is the Discrete Logarithm Problem which is described as follows:

"Let F_q be a finite field with q elements. Given 2 divisors, D_1 and D_2 in the Jacobian, determine $m \in Z$, such that $D_2 = mD_1$."

The following section describes the proposed HECC algorithm which exploits ElGamal technique for key generation process, encryption and decryption process which is named as HEC-ElG Algorithm (HEC-ElGA).

Algorithm for Public Key & Private Key generation

Input: The public parameters are hyper elliptic curve C, prime p and divisor D

Output: The Public key P_A and Private key a_A

$a_A \in_R N$ *[choose a prime* (a_A) *at random in N]*

$P_A \longleftarrow [a_A] D$

[The form of P_A is (u(x),v(x)) representation which is referred to as Mumford representation]

return P_A *and* a_A

For the random prime number generation in step 1, one can apply the probabilistic test of Robbin-Miller (Stallings 2002) or the deterministic test of AKS (Jin 2005). However, various researches have proved that it takes exponential time to determine the given large number is prime or not using AKS algorithm.

Encryption/Decryption Algorithm

In this section, we present the methodology for encryption and decryption. The message 'm' that is to be sent will be encoded as a series of points represented as (u(x),v(x)). The encoded message is referred as E_m. For the encryption and decryption process using HECC, we have used ElGamal method to design HEC-ElG Algorithm (HEC-ElGA). Details on ElGamal method can be had from (Avanzi & Lange 2006). The algorithm works as follows: To encrypt and send a message to B, A performs the following steps.

$k \in_R N$ (choose k as a random positive prime number in N)

$Q \longleftarrow [k]D$ (D is the Divisor of the HEC & The form of Q is (u(x),v(x)))

$P_k \longleftarrow [k]P_B$ (P_B:(u(x),v(x)) is receiver's(B's) public key)

$C_m \longleftarrow \{ Q , E_m + P_k \}$ (C_m :(u(x),v(x)) is the Cipher Text to be sent)

To decrypt Ciphertext message, the Decryption algorithm works as follows:
To decrypt the Cipher Text C_m , B extracts the first coordinate 'Q' from the cipher text then multiply with its Private Key (a_B) and subtract the result from the second coordinate. This can be written as follows,

$E_m + kP_B - a_B(Q) = E_m + kP_B - a_B(kD) = E_m + kP_B - k(a_B D)$
$= E_m + kP_B - kP_B = E_m$

In the above process, 'A' has masked the message E_m by adding kP_B to it. Nobody but 'A' know the value of k, so even though P_B is a public key, nobody can remove the mask kP_B. For an attacker to remove message, the attacker would have to compute k from the given D and [k]D i.e. Q, which is assumed very hard.

(Avanzi M 2003) has proved that HECC over prime field is satisfactory enough to be considered as a valid alternative to elliptic curves, especially when large point groups are desired. (Fan & Gong 2007) also proved that HECC provides greater efficiency than either integer factorization systems or discrete logarithm systems, in terms of computational overheads, key sizes, and bandwidth. In this work, we have adopted hyper-elliptic curve for genus 2 over GF(p) and have implemented the system.

4. Further Improvements

At this moment, all the certificates that are used in the SEMR system are assumed to be trusted. There is no central Certifying Authority (CA) in
place. One of the further improvements would be to provide a CA so that the certificates of users can also be trusted.

5. Conclusion

The SEMR system implementation provides a secure way to access the EMR. Additional security has been provided by using a two-factor authentication method which also takes into consideration the location of the user. The implementation of the system provided us with an insight in developing a distributed system which is secure, robust and user friendly. It has also provided us a deeper understanding of the role based access control model.

References

[1] AES (2001), U.S. Department of Commerce / National Institute of Standard and Technology, FIPS PUB 197, Specification for the Advanced Encryption Standard (AES), November 2001 from http://csrc.nist.gov/encryption/aes.

[2] Avanzi R M and Tanja Lange (2006), "Introduction to Public key cryptography" from "Handbook of Elliptic and Hyper elliptic curve cryptography" eds. Henri Cohen, Gerhard Frey, Chapman and Hall/CRC, Taylor and Francis, Florida, 2006.

[3] Burton S. Kaliski Jr., Cetin K.Koc and Chrisof Paar, (2002), "Cryptographic Hardware and Embedded systems", 4th International workshop, Bedwood shores, CA, USA, 2002.

[4] Duquesne S and Tanja Lange (2006), "Arithmetic of Hyper elliptic curves" from "Handbook of Elliptic and Hyper elliptic curve cryptography" by Henri Cohen, Gerhard Frey, Chapman and Hall/CRC, Taylor and Francis Group, Florida, 2006.

[5] Eigeartaigh C O, "A comparison of point counting methods for Hyper elliptic curve over prime fields and field ofcharacteristics of 2", Technical report, School of Computing, Dublin City University, Dublin, Ireland.

[6] Fan X and Gong G (2007), "Efficient Explicit Formulae for Genus 2 Hyperelliptic Curves over Prime Fields and Their Implementations", Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, CANADA, 2007.

[7] Ilya M (2005), "Hash functions: Theory, attacks, and applications", Microsoft Research, Silicon Valley Campus, November 14, 2005.

[8] Imad Khaled Salah, Abdullah Darwish and Saleh Oqeili, (2006), "Mathematical attacks on RSA cryptosystem", Journal of Computer Science, August, 2006.

[9] Jin T (2005), "Researching and Implementing on AKS Algorithm", University of Bath, May 2005.

[10] Lange T (2002), Efficient arithmetic on genus 2 hyper-elliptic curves over finite fields via explicit formulae", Cryptology ePrint Archive: Report 2002/121, 2002.
[11] Menezes A J, Yi Hong Wu, Robert J Zuccherato (1996) ,"An elementary introduction to hyper elliptic curves", Technical Report CORR 96-19, University of Waterloo, Ontario, Canada, November 1996.
[12] Sakai Y , Kouichi Sakurai (2000), "On the practical performance of hyper-ellipticcurvecryptosystems in software implementation", IEICE Transaction fundamentals, Vol. E83-A, No. 4 , April – 2000.
[13] Stallings W (2002), "Cryptography and Network Security: Principles and Practice", 2nd Edition, Pearson Education, 2002.
[14] Sun Java, http://java.sun.com /j2se/1.3/docs/guide/ security/ CryptoSpec.html

Gobi M is a Senior Lecturer, Department of Computer Science and Applications in PSG College of Arts & Science, Coimbatore, India. He teaches courses for BSc Computer Science, BCA and Master of Computer Applications (MCA). At present he is pursuing his PhD programme in Computer Science. His research areas of interest include Cryptography, Java, Software Engineering and Information Systems Security

Dr. K. Vivekanandan obtained his PhD in Computer Science from Bharathiar University during 1996. He is with Bharathiar University since 1986. At present he is working as the Faculty Member, Government College of Technology, IBRA, OMAN. He is also a visiting faculty, National University of RWANDA. He has published 9 articles in National/International Journals. He has presented more than 55 research papers in National/International conferences. He has guided 5 PhD's and currently guiding 6 PhD's.