Chapter-2

Background and Related work
BACKGROUND AND RELATED WORK

This chapter presents the background and reviews the corresponding literature. Section 2.1 discusses about Health Insurance and its significance. Section 2.2 explains the necessity of security to the Health Insurance data. Section 2.3 to Section 2.5 provides the mathematical background for cryptography, discusses various network, data security and digital signature techniques. Section 2.6 discusses about background of Data mining techniques applied in the Health Insurance System. Section 2.7 presents the factors effecting decision making of consumer of insurance scheme and Section 2.8 deals with summarized findings.

2.1 Health Insurance and its Significance

Material wealth, health and the ability to cope with adverse health events are intimately related. The World Bank defines poverty as "encompassing not only material deprivation but also low achievements in education and health and also vulnerability and exposure to risk". People with low income may be unable to afford preventive care-or curative care in the event of illness-and this may worsen their health.

In the event of serious illness, the poor are particularly vulnerable to financial burden of lost income and out-of-pocket medical expenses, as they have low levels of assets (for example, access to savings and credit or land and belongings for sale/mortgage) necessary to cope. Health Insurance to some extent will help common man to overcome above problems. The term ‘Health Insurance’ relates to a type of insurance that essentially covers your medical expenses. A health insurance policy like other policies is a contract between an insurer and an individual / group in which the insurer agrees to provide specified health insurance cover at a particular "premium" subject to terms and conditions specified in the policy.

Client may be benefitted by getting financial help from Health Insurance Policy from any Health care services company which would normally cover expenses necessarily incurred under the following heads in respect of each insured person subject to overall ceiling of sum insured (for all claims during one policy period).
a) Room, Boarding expenses

b) Nursing expenses, Fees of surgeon, anesthetist, physician, consultants, specialists

c) Anesthesia, blood, oxygen, operation theatre charges, surgical appliances, medicines, drugs, Diagnostic materials, X-ray, Dialysis, chemotherapy, Radio therapy, cost of pacemaker, artificial limbs, cost or organs and similar expenses.[59].

Various insurance schemes are introduced by government authorities like Public health insurance schemes, Micro health insurance schemes, community-based health insurance schemes allow many people's resources to be pooled to cover the costs of unpredictable health-related events. They protect individuals and households from the risk of catastrophic medical expenses in exchange for regular payments of premiums. Prepayment (even in the absence of pooling) can facilitate access to expensive medical care, because it spreads costs over time and prevents people having to pay at the time of treatment. By pooling resources, health insurance schemes can improve equity of and access to health care and can offer financial protection [21]. Following table gives information regarding various Health Insurance companies providing insurance policies along with those of the government.

<table>
<thead>
<tr>
<th>Public / Private Sector</th>
<th>Name of the Insurance Companies</th>
<th>Title of the health insurance policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public sector companies</td>
<td>The Oriental Insurance Company Ltd</td>
<td>1. Mediclaim Policy</td>
</tr>
<tr>
<td></td>
<td>The New India Assurance Company Ltd</td>
<td>1. Mediclaim Policy</td>
</tr>
<tr>
<td></td>
<td>National Insurance Corporation</td>
<td>1. Mediclaim Policy</td>
</tr>
<tr>
<td></td>
<td>United Insurance Company Ltd</td>
<td>1. Mediclaim Policy</td>
</tr>
<tr>
<td></td>
<td>Royal Sundaram Alliance Insurance Company Limited</td>
<td>1. Mediclaim Policy</td>
</tr>
<tr>
<td></td>
<td>Cholamandalam General Insurance Company Limited</td>
<td>1. Mediclaim Policy</td>
</tr>
<tr>
<td></td>
<td>TATA AIG General Insurance Company Ltd</td>
<td>1. Mediclaim Policy</td>
</tr>
<tr>
<td></td>
<td>Bajaj Allianz General Insurance</td>
<td>1. Mediclaim Policy</td>
</tr>
<tr>
<td></td>
<td>Health shield</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic Health Cover</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tata AIG Health First</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Health Guard</td>
<td></td>
</tr>
<tr>
<td>Company Limited</td>
<td>Critical Illness</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>ICICI Lombard General Insurance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Company Limited</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDFC Chubb General Insurance</td>
<td>Group Accident Policy, Hospital</td>
<td></td>
</tr>
<tr>
<td>Company Limited</td>
<td>Cash (Accident only)</td>
<td></td>
</tr>
</tbody>
</table>

**Source:** Insurance Regulatory and Development Authority (IRDA), 2006.

**Table:** 2.1 Showing various policies in different insurance companies

### 2.2 Need for Security in Health Insurance

Now-a-days every field is facing threats due to security. The insurance field is no exception. So it is to study security implications to health insurance field. In health insurance organizations there is a possibility of copying and keeping patients electronic health records (EHRs) through portable storage media, they will encounter new risks to the protection of their private information which should be dealt carefully along with their decisions in this competitive insurance market. Most of the health insurance companies use World Wide Web for patient operations such as opening of patient registration, consultation, prescription, reporting etc. From the security point of view, generally simple password based security is applied in order to identify patients and physicians through a simple window of authentication. The lack of security results in exploitation of insurance companies by some of patients and physicians.

#### 2.2.1 Security Threats of Health Insurance [18]

In the study by Sogukpinar Ibrahim et.al following six threats have been identified:

1. Unauthorized access to physician pages
2. Unauthorized access to patient pages
3. Confidentiality and integrity of Patients insurance data
4. Evil physician: The patient information and may give them to other people in which case patient privacy is violated.
5) Evil Insured person: An insured person may give his card and information to a non-insured person. If a physician is cheated or agreed, a person who has no right to benefit from HIC may use health services.

6) Non-repudiation: A physician may repudiate what he has done about a patient, diagnosis, prescription etc. This has to be prevented in order to avoid misuse of resources of HIC.

In the real world context, patient's health as well as insurance records is distributed around different hospitals and clinics, and the retrieval of this dispersed information when a patient visits a doctor in any particular hospital is a major problem. At present, there are two ways to overcome this problem: either the patient can carry his/her own records manually, or the records can be transmitted through an electronic network. In this Internet generation most of the patient's data is transmitted thorough electronic means which poses threat to data secrecy and accuracy. So stringent measures should be employed to avoid discrepancies by using latest cryptographic techniques either by a trusted Insurance company or by a trusted third party.

2.3 Significance of Cryptography

Computer security encompasses implementing measures for guarding the resources stored on the computer from threats. Network security comprises techniques for protecting the resources like network devices, network transmission media, and the data being transmitted across the network from threats. Cryptography (the science of changing plain text by substituting or transposing characters) is an essential part of providing data confidentiality, integrity, and authentication in the security [1]. Cryptography is best understood by breaking it into four main primitives; the base for designing various cryptosystems. The following primitives will ensure solutions for the three main cryptographic goals Confidentiality, Integrity, and Authentication (CIA),

1. Random number generation: a computational technique designed to generate a sequence of numbers or symbols that lack any pattern, i.e. appear random used in many cryptography algorithms for key generation.
2. *Symmetric encryption*: only a single key is used to both encrypt and decrypt a message.

3. *Asymmetric encryption*: two distinct keys, one public key and one private key are used for encryption and decryption respectively.

4. *Hash functions*: takes a message of any size and computes a smaller, fixed-size message called a hash. Hash functions are used to provide better performance when signing large blocks of data using asymmetric encryption, to provide integrity, in authentication protocols, and to create pseudorandom data.

The earlier cryptographic algorithms are broken by computers, which can try millions of combinations each second. As computers become more powerful, the cryptographic algorithms are to be designed stronger to keep data secure. Since single key is used for both encryption and decryption, the symmetric cryptosystem requires transferring a secret key to both communication parties before secure communication can begin. The shared secret key between the communication parties is a problem as a secure confidential channel is required. So, this problem of shared key is removed in public key cryptosystem which led us to focus on the asymmetric cryptosystem and its functionality.

2.4 Cryptographic Techniques

The cryptographic techniques are classified into two types;

2.4.1 Symmetric Key Cryptography

The idea behind symmetric encryption is that only a single key is used to both encrypt and decrypt a message. The benefit to using symmetric encryption is that it’s very fast and the main drawback to using symmetric encryption: exchanging keys.

2.4.1.1 Stream Cipher

Stream cipher uses a single key to encrypt a message or stream of data. The message is considered to be a stream of data in that each byte is processed with the bytes preceding it, and the order is important. If you were to change the order of any of the bytes in the plain text, the cipher text, from that point forward, would look
different. Fig 2.1 shows what a stream cipher does. Stream ciphers normally do not require any padding of the message. **Padding** is adding extra bits to the message. Because messages are treated as a stream of data they can be of any length and do not need to be padded in any way except to add randomness to common messages.

![Stream Cipher Scheme](image)

*Fig 2.1 Stream Cipher Scheme*

The existing stream ciphers that are used in data and network security are:

- RC4
- SEAL (Software-Optimized Encryption Algorithm)
- ISAAC (stands for indirection, shift, accumulate, add, and count)
- Panama
- A5/1
- A5/2
- FISH (Fibonacci Shrinking)
- Helix

### 2.4.1.2 Block Cipher

Block cipher shown in fig 2.2 is a kind of symmetric encryption algorithm which uses one block at a time and a single key to encrypt a message. A block is considered a certain number of bits and is determined by the algorithm. Each block is processed independently, and there is no correlation between the encrypting of one message block and another.
Block Ciphers is used by applying different encryption modes to prevent the same plain text block always encrypting to the same cipher text block. The encryption modes are described as follows:

- **Electronic code book (ECB):** The message is encrypted one block at a time so that one plain text block maps to one cipher text block. An error in any block only affects the decryption of that block. If an entire block is lost during transmission, none of the other blocks are affected.

- **Cipher-block chaining (CBC):** The output block of the previous encryption is XORed with the next block of plain text before being encrypted. If a bit is changed in the plain text of one block, that change is propagated to all subsequent cipher text blocks. If a cipher text bit is changed, the current block will be corrupted and the changed bit will be inverted in the next block. CBC does not allow blocks to be encrypted in parallel.

- **Propagating cipher-block chaining (PCBC):** Similar to CBC, except that changes to the cipher text are propagated throughout the message. PCBC is the mode of operation used in Kerberos (an authentication protocol).

- **Cipher feedback (CFB):** The previous cipher text block is XORed with the current encrypted text block. This differs from CBC mode in that the XOR occurs after the encryption of the current text block. If a bit in the cipher text is changed, the current block will have that bit inverted and the subsequent block will be corrupted.

- **Output feedback (OFB):** The output of the encryption algorithm is continually fed into the algorithm while the plain text is XORed with this output. This differs from CFB because what is fed into the encryption
algorithm does not include the cipher text. If an error occurs in one block, that 
error is only propagated to those bits that are changed. However, if any of the 
bits are lost, including a whole block, the error is propagated to all of the 
remaining blocks and cannot recover.

DES and AES (discussed previously) are NIST standards which are among the 
block ciphers such as;

✓ DES
✓ Blowfish
✓ Cast
✓ Skipjack
✓ Two fish

Since single key is used for both the encryption and decryption and the secret 
key has to be shared among the users securely before encryption mechanism, the 
problem with this cryptography technique is key distribution problem which will not 
be suitable for Health Insurance system. So, to overcome this we can use either the 
TTP (Trusted Third Party) or Public Key Cryptography (PKC/Asymmetric).

2.4.2 Asymmetric Key Cryptography

In the 1970s, researchers invented asymmetric-key cryptography, a new way 
to securely send keys. This scheme uses two different keys. Although they are related 
to each other—they are partners—they are significantly different. In this type of 
cryptographic algorithm, one key encrypts data and the other key decrypts it. Another 
term for this model is public-key cryptography. Because both keys are needed to lock 
and unlock the data, one of them can be made public without jeopardizing security. 
This key is known as the public key. Its partner is called the private key. You encrypt 
data with the public key and decrypt it with the private key.
The following are some of the public key (asymmetric) cryptographic algorithms;

- Diffie-Hellman Key Exchange Mechanism
- Elgamal Public Key Cryptosystem
- RSA Public Key Cryptosystem

2.4.2.1 Diffie-Hellman Key Exchange Mechanism

Diffie-Hellman is a key agreement algorithm that helps two devices to agree on a shared secret between them without the need to exchange any secret/private information. An overview of the algorithm is given below.

*Algorithm:*

1. Let $a$, $b$ be the private keys of the devices A and B respectively, Private keys are random number less than $p$.

2. Let $g^a \mod p$ and $g^b \mod p$ be the public keys of devices A and B respectively.

3. A and B exchanged their public keys.

4. The end A computes $(g^b \mod p)^a \mod p$ that is equal to $g^{ab} \mod p$.

5. The end B computes $(g^a \mod p)^b \mod p$ that is equal to $g^{ab} \mod p$.

6. Since $K = g^{ab} \mod p = g^{ab} \mod p$, shared secret $= K$.

2.4.2.2 ElGamal public-key Cryptosystem

The Elgamal public-key encryption scheme can be viewed as Diffie-Hellman key agreement in key transfer mode. Its security is based on the intractability of the discrete logarithm problem and the Diffie-Hellman problem.
Algorithm:

1. **Key generation:**
   
   1. Generate a large random prime \( p \) and a generator \( \alpha \) of the multiplicative group \( \mathbb{Z}_p \) of the integers modulo \( p \).
   
   2. Select a random integer \( a \) such that \( 1 \leq a \leq p - 2 \), and compute \( \alpha^a \mod p \).
   
   3. Public key is \((p, \alpha, \alpha^a)\); private key is \( a \).

2. **Encryption:**
   
   (a) Obtain authenticated public key \((p, \alpha, \alpha^a)\).
   
   (b) Represent the message as an integer \( m \) in the range \( \{0, 1, \ldots, p - 1\} \).
   
   (c) Select a random integer \( k \), \( 1 \leq k \leq p - 2 \).
   
   (d) Compute \( \gamma = \alpha^k \mod p \) and \( \delta = m \cdot (\alpha^a)^k \mod p \).
   
   (e) Send the cipher text \( c = (\gamma, \delta) \).

3. **Decryption:**
   
   (a) Use the private key \( a \) to compute \( \gamma^{p-1-a} \mod p \) (note: \( \gamma^{p-1-a} = \gamma^a = \alpha^{-ak} \)).
   
   (b) Recover \( m \) by computing \( (\gamma^{-a}) \cdot \delta \mod p \).

2.4.2.3 RSA Public Key Cryptosystem

The RSA cryptosystem, named after its inventors R. Rivest, A. Shamir, and L. Adleman, is the most widely used public-key cryptosystem. It may be used to provide both secrecy and digital signatures. The RSA encryption method exhibits a specific feature that even an encryption key is publicly revealed it is difficult to trace the corresponding decryption key. A message is encrypted by representing it as a number \( M \), raising \( M \) to a publicly specified power \( e \), and then taking the remainder when the result is divided by the publicly specified product, \( n \), of two large secret prime numbers \( p \) and \( q \). Decryption is similar; only a different, secret, power \( d \) is used, where \( e\cdot d \equiv 1 \pmod{p - 1} \). The security of the system rests in part on the
difficulty of factoring the published divisor n. The RSA Public Key Cryptosystem is depicted in the Fig 2.3.

Fig 2.3 Encryption and Decryption using Public Key Cryptosystem

The security of RSA depends on the following facts:

**Setup:** Let \( p \) and \( q \) be large primes, let \( N = pq \), and let \( e \) and \( c \) be integers.

**Problem:** Solve the congruence \( xe \equiv c \pmod{N} \) for the variable \( x \).

**Easy:** Who knows the values of \( p \) and \( q \), can easily solve for \( x \).

**Hard:** Who does not know the values of \( p \) and \( q \), cannot easily find \( x \).

**Dichotomy:** Solving \( x^e \equiv c \pmod{N} \) is easy for a person who possesses certain extra information, but it is apparently hard for all other people.

**RSA Key Generation Algorithm**

Entities which need to communicate with each other; each entity creates an RSA public key and a corresponding private key. Each entity A should do the following:

1. Generate two large random (and distinct) primes \( p \) and \( q \), each roughly the same size.
2. Compute \( n = p \cdot q \) and \( \phi(n) = (p - 1)(q - 1) \).
3. Select a random integer \( e \), \( 1 < e < \varphi(n) \), such that \( \gcd(e, \varphi(n)) = 1 \).

4. Compute the unique integer \( d \), \( 1 < d < \varphi \), such that \( ed \equiv 1 \pmod{\varphi(n)} \).

5. A's public key is \( (n, e) \); A's private key is \( d \).

**RSA Encryption**

If an entity A has to send the message confidentially to an entity B. Entity A should perform following

(a) Obtain authentic public key \( (n, e) \).

(b) Represent the message as an integer \( M \) in the interval \([0, n - 1]\).

(c) Compute \( C = M^e \mod n \)

(d) Send the cipher text \( C \) to A.

**RSA Decryption**

To recover plaintext \( M \) from \( C \), A should do the following:

(a) Use the private key \( d \) to recover \( M = C^d \mod n \).

During past 20 years of usage of RSA cryptosystem revealed it is also vulnerable to various attacks which is solved from time to time developments. So it is proposed to modify the version of RSA.

**2.5 Digital Signature**

In Digital signature scheme shown in figure 2.4, a message can be signed by a device using its private key to ensure authenticity of the message. Any device that has got the access to the public key of the signed device can verify the signature. Thus the device receiving the message can ensure that the message is indeed signed by the intended device and is not modified during the transit. If any of the data or signature is modified, the signature verification fails.

Private Key = \( P_A \)

Public Key = \( U_A \)
For example, if a device ‘A’ needs to ensure the authenticity of the message, the device A signs the message using the private key \( P_A \). The device A will then send the message ‘\( \text{Msg} \)’ and signature ‘\( \text{Sgn} \)’ to device B. The device B, on receiving the message, can verify the message using A’s public key \( U_A \) and thereby ensuring that the message is indeed sent by A and is also not tampered during the transit. Since only the device A knows its private \( P_A \) key, it is impossible for any other device to forge the signature. The examples of Digital Signature algorithms are RSADSA and ECDSA.

### 2.5.1 RSADSA – RSA Digital Signature Algorithm

**Parameter generation**

1. Choose a 160-bit prime \( q \).
2. For an integer \( z \), choose an \( L \)-bit prime \( p \), such that \( p = qz + 1, 512 \leq L \leq 1024 \), and \( L \) is divisible by 64.
3. Choose \( h \), where \( 1 < h < p - 1 \) such that \( g = hZ \text{ mod } p > 1 \).
4. Choose a random number \( x \), where \( 0 < x < q \).
5. Calculate \( y = gx \text{ mod } p \).
6. Public key is \( (p, q, g, y) \). Private Key is \( x \).

**Signing**

Consider the device A that signs the data \( M \) that it sends to B.

7. Let \( x \) be A’s private key and \( (p, q, g, y) \) be A’s public key.
8. Generate a random per-message value \( k \), where \( 0 < k < q \).
9. Calculate \( r = (gk \text{ mod } p) \text{ mod } q \).
10. Calculate \( s = (k-1(M+x*r)) \mod q \), where \( M \) is the hash SHA1 of the message

11. The signature is \((r, s)\).

**Verification**

12. Let ‘\( M \)’ be the message and \((r, s)\) be the signature received from A.

13. Let \((p, q, g, y)\) be A’s public key. Since \((p, q, g, y)\) is public, B has access to it.

14. Calculate \( w = s-1 \mod q \).

15. Calculate \( u1 = (M \cdot w) \mod q \), where \( M \) is the hash SHA1 of the message.

16. Calculate \( u2 = (r \cdot w) \mod q \).

17. Calculate \( v = ((gu1 * yu2) \mod p) \mod q \).

18. The signature is valid if \( v=r \), invalid otherwise.

A cryptographic system can only be as strong as the encryption algorithms, digital signature algorithms, one-way hash functions, and message authentication codes it relies on. In other words, a break in any of them, and you’ve broken the system. As many applications use public key cryptosystems, we have studied about the possible vulnerabilities of asymmetric key cryptosystems. The public key cryptosystem ensures the security with the pair of public and private keys \( e \) and \( d \) respectively. Various cryptographic algorithms are designed with mathematical illustrations and hence public key cryptosystems are influenced by some run time attacks which obstruct the privacy of users. So, day to day developments suggest the improvements of current security techniques. This is essential to protect data when applied to Health insurance field or any other.

### 2.6 Data Mining Techniques to a Health Insurance Information System

In paper by Scott Judy et.al. [35], they have discussed about the use of knowledge management systems (KMS) in healthcare organizations which states that there is an increase of its usage in health care and insurance systems. Among the
reasons for this trend are pressures to reduce costs, which have been growing at an unsustainable rate and to improve the quality of healthcare.

Data mining is one of the most important steps of the knowledge discovery in databases process and is considered as significant subfield in knowledge management. The need to maximize the use of data for planning and strategic business development in all aspects of management has led many corporations to build comprehensive information systems that record all kinds of operational transactions. As these databases grow larger, with gigabytes sizes becoming quite common, they are overwhelming the traditional query and report-based methods of data analysis [38]. Data mining is the data driven extraction of information from such large databases, a process of automated presentation of patterns, rules, and functions to a knowledgeable user for review and examination. A domain expert plays an essential role in the paradigm because he/she decides whether a pattern, rule, and function is interesting, relevant and useful.

Data mining can enable healthcare organizations or insurance companies to predict trends in the patient conditions and their behaviors, selections regarding insurance policies, hospitals which is accomplished by data analysis from different perspectives and discovering connections and relations from seemingly unrelated information. Healthcare data mining provides countless possibilities for hidden pattern investigation from these data sets for making good decision regarding insurance policies, hospitals etc. Data mining techniques have been applied mostly to ‘database marketing’ through the analysis of customer databases; other applications include analysis and selection of stocks, fraud detection, spending patterns through the study of financial records, and detection of spatial patterns of bit failures in semiconductor memory fabrication [41].

With the steady rise in health-care costs and the growing urgency to control these costs, timely analysis of health-care information has become an issue of great importance. Large corporations, hospitals, health-care maintenance organizations, and insurance companies require expert analysis of their health-care data, a task that is both time consuming and very expensive [8]. Consequently, these organizations are developing automated mechanisms to support the process. For example, a system for
detecting health-care provider fraud in electronically submitted claims has been developed at Travelers Insurance [41].

In addition, HIC has developed and implemented a pattern-recognition technique for classifying the degree of appropriate practice of service providers. By screening with this technology the claiming patterns of all practitioners, a score or class is assigned to each practitioner; those practitioners who have a high likelihood of an inappropriate practice pattern are identified. However, this technique allows only four different classes. The results of this study have demonstrated the potential value of data mining in health insurance information systems, by detecting patterns in the ordering of pathology services, and by classifying the general practitioners into groups reflecting the nature and style of their practices. The approach used led to results which could not have been obtained using conventional techniques.

2.6.1 Data mining techniques

There are several primary data mining techniques discussed in [41] such as

1. Artificial Neural Networks: this is a nonlinear predictive model that learns through training and resembles biological neural networks in structure.

2. Decision trees: tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.

3. Genetic Algorithms: They are optimization techniques that use process such as genetics combination, mutation, and natural selection in a design based on concepts of evolution. It tries to mimic the way nature works. It is an adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetics.

4. Rule Induction: the extraction of useful if-then rules from data based on statistical significance.

5. Regression Methods: this tries to identify the best linear pattern in order to predict the value of one characteristic.

In this paper [41], authors have discussed about applicability of association rules, which is a more sophisticated method of summarization, and neural
segmentation as a particular implementation of clustering. These techniques, whose basic characteristics are summarized next, were selected based on the description of the problem and on our interest on experimenting with two distinct methodologies.

2.6.1.1 Association rules

Association rules over basket data type transactions are a problem defined in [31,36]. The mining of association rules relates to finding intra-transaction patterns and can be defined as follows: given a database of transactions, where each transaction represents a set of items (e.g. services rendered), generate all associations such that the presence of some specific item(s) x in a transaction implies the presence of other item(s) y. The association rule x \( \Rightarrow \) y will hold given a support and a confidence greater than a user-specified minimum support (\( S_{\text{min}} \)) and minimum confidence (\( C_{\text{min}} \)). Wherein support is the number (or fraction) of the transactions that contain a given item set. Confidence measures the frequency that items in a multi-item set are found together.

The database could be a data file, a relational table, or the result of a relational expression $\text{IAS941}$. The strengths of this technique are the capability to handle large databases in an efficient manner, while its execution time scales almost linearly with the size of the data. These characteristics have proven true during the course of this study.

2.6.1.2 Decision tree

Decision tree is a graphical representation of the relations that exist between the data in the database. It is used for data classification. The result is displayed as a tree, hence the name of this technique. Decision trees are mainly used in the classification and prediction. It is a simple and a powerful way of representing knowledge. The models obtained from the decision tree are represented as a tree structure. The instances are classified by sorting them down the tree from the root node to some leaf node [39]. The nodes are branching based on if-then condition. Tree view is a clear and easy to understand, a decision tree.
2.6.1.3 Neural segmentation

Neural segmentation is a pattern detection algorithm, in which the base technology is a self-organizing feature map [38]. Self-organizing feature maps, otherwise known as topological feature maps, loosely preserve the topology of the multi-dimensional space in the two-dimensional map. That is, similar prototypes are near each other on the map.

A self-organizing feature map consists of a two-dimensional array of units; each unit is connected to n input nodes, and contains a n-dimensional vector \( \text{W}_i \) wherein \((i,j)\) identifies the unit at location \( C_{ij} \) of the array. Each neuron computes the Euclidean distance between the input vector \( \text{x} \) and the stored weight vector \( \text{W}_i \). The neuron with the minimum distance is declared the “winner” and the input vector is assigned to this neuron. In addition, each of the weight vectors is modified as follows:

new weight vector \( \text{W}_i' = \text{W}_i + \text{LR} \times \text{NF} \times (\text{x} - \text{W}_i) \) wherein

\( \text{LR} \) is the learning rate, a linearly decreasing scalar which changes after each epoch;

\( \text{NF} \) is the neighborhood function, a Gaussian distribution function in map-space, centered on the winning neuron;

<table>
<thead>
<tr>
<th>Age</th>
<th>lab_id</th>
<th>benefit</th>
<th>Cost_ser1</th>
<th>Cost_ser2</th>
<th>......</th>
<th>Serv1</th>
<th>Serv2</th>
<th>......</th>
<th>provid</th>
<th>Sex</th>
</tr>
</thead>
</table>

*Table 2.2 Record in Episode database*

**Epoch** is the presentation of all input vectors to the system once, that is, one iteration, over the data. Key to the success of this analysis is the representation of behavioral data. Since self-organizing feature maps are a form of clustering, care must be taken into properly balancing the inputs. That is, each vector element intrinsically represents one equally weighted dimension of the subject’s behavior. Balancing the inputs implies that each equally-important aspect of the problem has the same number
of vector elements. This is accomplished through an extensive study of the problem, with attention, to the type of characteristics that are important to the solution, and to a careful review of the variance of each of the inputs. The output of the algorithm is a two dimensional array of segments, each one described by both of its behavioral profile and potentially an overlay of further descriptive data.

2.7 Decision making in Health Insurance

Clients/Patients makes decisions regarding selection of insurance scheme, and to understand consumer behavior in any market is difficult, but health care is particularly challenging for several reasons [32].

- First, the health care sector involves numerous stakeholders, often with complex relationships.
- Second, health status is uncertain because it is generally difficult to predict.
- Third, all health care consumers do not have the same preferences, with some valuing convenience over cost,

Customers typically make health insurance plan purchasing decisions based on six types of data:

1) Information on access to providers (both hospitals, facilities, locations and physicians)
2) Out-of-pocket costs
3) Quality of providers
4) Facilities/Services of Insurance Company
5) Courtesy
6) Administrative burden (i.e., paperwork).

Conjoint analysis is a multivariate technique used specifically to understand how respondents develop preferences for products or services. It is based on the simple premise that customers /client evaluate the value or utility of a policy/service by combining the utilities they associate with each value of each client’s data attribute. But for security purpose some of the attributes should be hidden
cannot be provided, so the existing methods may not be suitable when we consider both security and decision making.

2.8 Summary

The above survey gives a clear idea about Health Insurance systems and problems in existing systems related to security. The study of existing security techniques in Insurance schemes and decision making systems propose a new secure decision making system MASDMA and threshold Selective Secure Field RSA (S²RSA).