# Privacy Assurance and Fraud Detection in Healthcare Engineering

Chandrasekaran Subramaniam, Niveditha Narendhran, and Mohammed Nazim Feroz

*Abstract*—**The objective of the paper is to propose a privacy centered process model for healthcare information system focusing on the trust between the entities, security methods and the resulting privacy assurance in the health information. The work addresses the criticality of the medical data and an adaptable security mechanism to prevent anticipated security risk in disclosing the allowable information. The trust between various covered entities including medical insurance agents is managed by a dynamic trust evaluation method in different process. The collaboration of various objects through inter process communication is achieved to determine the needed security method by an object in a different process. The permitted colored tokens help to detect the intruder or a fraud in submitting the same medical case many times. The cyclic processes and sub processes are simulated not only to determine the utilization of various resources of healthcare information system but also the sensitivity of the system in different scenarios.**

*Index Terms*—**Adaptable security, covered entity, fraud detection, process, resource utilization.**

## I. Introduction

The healthcare process can be modeled according to the role and responsibilities in a multi-disciplinary team where interactions between individuals and trusts are the key issues Healthcare information security refers to the process of prevention, protection and detection of the information stored against unauthorized access, modification, destruction or use. Inside threat and outside threat attacks are events that may occur randomly against the information system [1]. Apart from the interaction between healthcare workers, the state transition diagram which depicts the status of the patient and dataflow diagram between human-machine interactions are not focusing the variation of trust between the stack holders and the overall privacy of the protected healthcare information is not dealt under different scenario [2]. A healthcare process may be thought as a set of one or more linked procedures or activities which collectively realize a business objective or policy goal defining function roles and relationships. Healthcare system will increasingly use technologies to customize processes to improve the operational efficiencies and patients safety but not maintaining an acceptable privacy of the patient healthcare information. In healthcare many of the non-functional constraints which cannot be changed are role related. The process model should identify the role and various laws and regulations before satisfying the non-functional requirements [3]. Juha Puustijarvi et al., presented a new transaction model, called two-phase reservation transaction, which can be used to ensure that the clinical resources are reserved in an atomic way to minimize the effects of the cancelled clinical reservation [4]. In the earlier healthcare process simulations, the hidden process issues like cost, privacy, work time and wait time of the individual activities on the covered entities are not addressed. The risk value associated with each and every information exchanged, the initial and final trust values between each and every entity in the healthcare are not addressed.

Information security management in health using ISO/IEC 27002 provides guidelines supporting the implementation of Information Security Management (ISM) in health organizations. The Information security risk can be considered as a potential threat that will exploit a vulnerability or group of assets and thereby cause harm to the organization. The risk which is a combination of the probability of an event and its consequence is to be identified and the assessment of all the risks in the overall processes is to be analyzed and estimated. The involvement of several actors in a process increases the probability of security and privacy related issues arising during the development of a system and during its use. This is especially important in healthcare systems as they often interact with other organizations such as insurance, financial and billing companies that need detailed information about patients [5]. Alfred C. Weaver had implemented a prototype of the system with various authentication methods and devices, authorization policies and representations, and trust brokering among cooperating trust domains using both direct and indirect trust strategies [6]. As for the sensitivity of the data that is considered, the application domain experts have the best knowledge about which data are sensitive and which are not. They annotate sensitive attributes of domain objects with classification in the domain model [7]. Hence the main focus of the work is to identify the sensitivity type of the data in healthcare information domain and select the needed security strategy so that only the covered entities will get the needed information as per their privileges in the disclosure processes.

The paper is organized as follows: Section II describes the need for modelling the healthcare information system in terms of processes involving trust and security according to the sensitivity of data handled.. Section III illustrates a Coloured Petri Net model for formalizing the interaction

Manuscript received April 10, 2012; revised May, 6. Privacy Assurance and Fraud Detection in Healthcare Engineering.

C. Subramaniam is with the Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamil Nadu, 641006, India (e-mail: chandrasekaran.s@msn.com).

N. Narendhran is with the Department of Computer Science, Polytechnic Institute of NewYork University, Brooklyn, NY 11201, USA. (e-mail: nivy_ganguly@yahoo.co.in).

M. N. Feroz is with the Electrical Engineering Department, Texas Tech University, Lubbock, TX 79409 USA, (e-mail: nazimatr@gmail.com).

between various entities through restricted tokens in applying security measures and trust building stages. Section IV explains the integrated privacy technique and a CPN model to detect a fraudulent case in reclaiming the medical bill through an insurance agent.

## II. Process Modeling of TSP

In any information system, the security has to be viewed in three main dimensions: confidentiality, availability and integrity of the health information. The healthcare organizations need to identify and manage many security activities in order to function effectively and efficiently. A healthcare information process is one where the resources are to be managed to enable the transformation of inputs into outputs using a set of interrelated or interacting activities. The output from one process can directly act as the input to another process and generally this transformation is carried out under planned and controlled conditions. All medical records typically contain sensitive data such as name, date of birth, security number, insurance information and medical history of the patients. Security breaches are the result of unauthorized or inadvertent actions of employees of the organization resulting as inside threats or intentional attacks from outside who may not be the covered entities. A covered entity may also rely on an individual's informal permission to disclose to the family members, relatives, and friends or to other persons. Healthcare IT system should protect patient data and report breaches as per HIPAA and HITECH acts [8]. The privacy rule permits the use and disclosure of protected health information without an individual's authorization or permission in a national emergency situation for the purpose of alerting the citizens. Hence a healthcare process model should represent various application scenarios using well defined process flows which capture the relationship between roles, tasks and data functions of the all the entities. The privacy and security policies must be coordinated and developed openly with abundant public input in order to ensure a higher degree of trust.

First focusing on the trust issues, the trust in the healthcare information system may be considered as a variable which is a function of possible random and non periodic outcomes called reputation of patient's treatment. The trust 'u' may be represented as a function of the time of observation ' t 'and the reputation 'r' of the health organization that can be represented as **u (t, r)** .

A family of all such variables and functions can be denoted as

**U(time,reputation),** which may be a random process **U** in the healthcare domain.

**U (t, r)** is a random process where 'u' gives a specific value of the random variable U. This random process can be considered as a single time function where 't' is a variable when the reputation 'r' is fixed at a specific value. $U_1$ denotes the random variable associated with the process **U (t) at time $t_1$**. The expected value of trust $U_1$ is a mean value of the random process at time $t_1$. When both time and reputation are fixed, this random process represents a number. This random process is a discrete random process corresponding to a random variable 'u' having only discrete values while time is continuous that can be represented as,

$$\mathbf{U_i = U\ (t_i,\ r) = U\ (t_i)}$$

A discrete time random process is a set of random variables denoted by $\{U(t_n)\}$ for sample times

$$t_n = nT_s,\ n = 0, \pm 1, \pm 2$$

where $T_s$ is called the sampling interval,

$$\text{sampling rate} = 1/T_s.$$

Applying specific sampling rates for the collection of trust values with a constant reputation of the health organization, the set of trust values at different time intervals are obtained. From the initial trust, the random variable will follow a deterministic style so that the future trust value may be predicted. The deterministic behavior can also be applied to the security and privacy disclosure stages too. If a process model of the organization is understood then the quality attributes like privacy of the system can be audited and any fraud can be detected. The Trust-Security-Privacy-Disclose cycle of the process model in a health organization is shown in Fig.1 the cycle contains four main processes and four sub-processes in their respective stages. The four main processes are trust evaluation, security analysis, privacy checking and PHI revealing. For each and every main process, it may be assumed to have a sub-process like threat assessment, security policy and management policy and trust modification.

For each and every main process, it may be assumed to have a sub-process like threat assessment, security policy and management policy and trust modification. Based on the evaluation of trust, it is possible from the model to decide whether the request for the data is anonymous or not. In the security analysis phase, the request is authorized for the particular data according to the privilege associated and then the privacy is checked through the observability of the data.

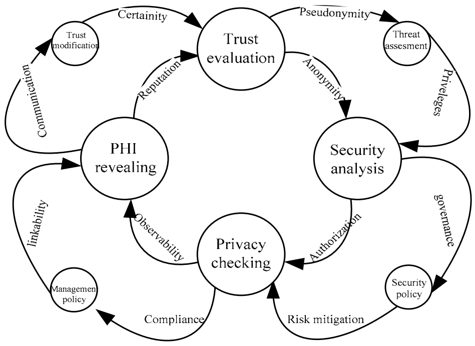

Fig. 1.Trust- security- privacy- disclose cycle

The current disclosure of PHI determines the reputation of the system for the next cycle. A medical record may contain many data fields. In the proposed model, these fields have been classified as patient details, medical details and billing details. These fields are associated with different levels of sensitivity based on which the appropriate security mechanism is applied. For data of low sensitivity, low level of security mechanism is applied which include simple encryption and type mapping. On the contrary, data of higher level sensitivity is enforced with high levels of security that consists of double encryption and type changing. For instance the bill details are highly sensitive and are thus provided with high security mechanisms to enhance the privacy of those information. In other situation, the identification number or the ward or bed number of the patient is treated as low sensitive data and protected with simple encryption but at the same time disclosed only to the authenticated users. The information structure and their sensitivity level are shown in Fig. 2.

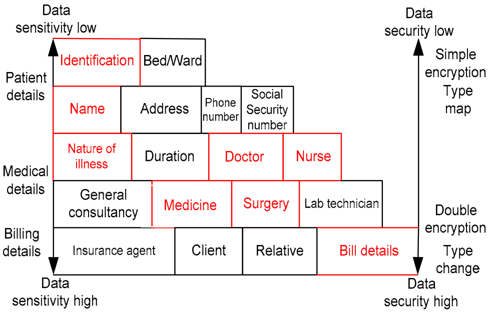

Fig. 2. Information structure and sensitivity

## III. Colored Petrinet Model for Trust and Security

Colored Petri Net (CP-nets or CPNs) is a formal modeling language developed for systems in which communication, synchronization and resource sharing play an important role. CP-nets combine the strengths of ordinary Petri nets with the strengths of a high-level programming language used for the description of discrete distributed systems. Since the healthcare information system is a distributed system and it is domain specific, the CP-nets provide a graphical representation providing an abstract, application-specific view of the current state and activities in the system. With the help of Petri Nets, it is easier to illustrate how the individual processes interact with each other. CP-nets also has a formal, mathematical representation with a well-defined syntax and semantics. This formal verification method is known as state space analysis and invariant analysis. The untimed CPN models are usually used to validate the functional or logical correctness of a system, while timed CPN models are used to evaluate the performance of the system. In the healthcare information system , it is essential to prove the correctness of the system in terms of the full and partial disclosure of correct information to the correct entity in the correct time in the correct mode. Since the system is basically a concurrent and a nondeterministic system, the Petri Net modeling is the correct choice in formally modeling the system for their quality parameters.

For example, in the proposed model, for a *TrustEvaluator* process a binder is developed where the corresponding states $S_t$ are represented as a set

$S_t$={ Diagnosis, Receive report, Send report,
consults_regular patient, Consultancy_Expert,
Total trust_evaluation }

and the permitted transitions and the color sets are given as two different sets , $T_t$ and $C_t$ respectively.

$T_t$={ Directed, Go for, Choose, Decide on }
$C_t$={ P } where
Int.max(n1,n2) is a guard function.

Similarly in *SecurityMechanism* process, another binder with permitted states that are represented as a set $S_s$

$S_s$={Retreive_record,
Identify_record_type,
Receive_nurse, Grant, Request, Store in
Dbase, Encryption, Decryption}

and the permitted transitions and color sets are given as two different sets, $T_s$ and $C_s$ respectively.

$T_s$={ Forward_Information, Transmit, Transfer
to Dbase, Sensitive, Non-Sensitive,
Matched, Pass_shared-public_key}
$C_s$={A, B, C, D, E}

The trust is an aggregated value and it is cumulated with that of each progessing stage by assigning weightages for the various covered entities having unique trust values or weightages. The maximum evaluated trust value, in which only the covered entities with the highest weightage is considered. The trust value may change with varying sequences of paths taken by the data and processing carried out in each step.

The Coloured Petrinet model in Fig 3 depicts the technique for the total trust calculation in a healthcare information system. In the first stage of the model , the trust between the patient and doctor is evaluated with an initial weightage that the patient has at the time of entry. The doctor diagnoses the patient considering the weightage he posseses. As each entity is passed, the corresponding trust value of that particular entity is accumulated with the current trust value. By this process the trust value increases gradually as each subsequent entity is acted upon. The accumulated trust value assigned is carried through each progessing stage in the model until the final trust is calculated. In the given pseudo code, the total number of entities is considered as N out of which the numbers of covered entities are M. These are classified into

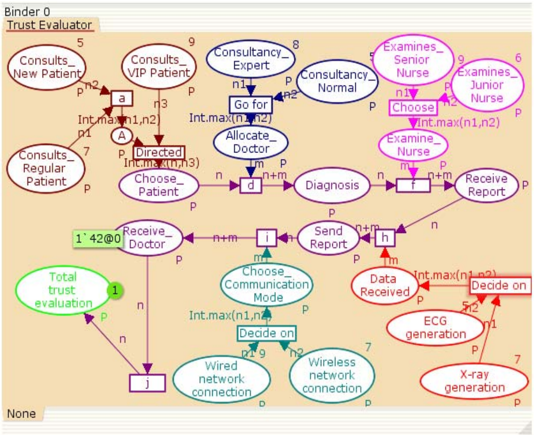

Fig. 3. CPN for trust evaluation.

Three regions R1, R2, R3 which represent entities, device and network respectively. With the initial trust of the system is assumed to be zero, the entity that has the highest priority is chosen and its trust weightage is added to the existing trust value of the relation between those entities in the system. Thus the final trust obtained in the system is the cumulative sum of distributed trust of each entity involved in the different activities in any process.

```
Start : trust evaluator
        total entity : N
        covered entity : M
        $M \subseteq N$
  region R = R1, R2, R3  // R1:entities, R2: device,
                            R3: network //
        $e_g € N$, $e_c € M$
        entities $e_i$, $e_j$
        let initial_$t_{ij}$ = 0
loop:
  read priority$_{i,j}$ $p_i$, $p_j$ // priority of the entity
                            selected//
  read weightage = w // weightage of the
                            relationship//
        max $t_{ij}$ = w * $p_{ij}$
        current_$t_{ij}$ = initial_$t_{ij}$ + max_$t_{ij}$
end loop
total_$t_{ij}$ = $\sum_{R=1}^{3} [ \sum_{i}^{n} \sum_{j}^{m}$ current_$t_{ij}$ ]
end trust_evaluator
```

Fig. 4 represents the Petri Net that models the security mechanism for the system. This ensures that the data is secured by an access control technique thereby accessible only to those who are registered with a username and password. The record that is to be retrieved is identified based on its type and the permitted records are forwarded to concerned recipient. The partial or full record may be sent from the doctor to nurse but not in the other direction. The fields in the record are classified according the structure and its sensitiveness into sensitive or non-sensitive. The non-sensitive data is stored in the database and is accessible to all the entities. The sensitive data is encrypted by the doctor and the nurse in their own way and then stored in the database. The shared public key concept is used for decryption. If the key that the entity possess at that time or session matches with the shared public key then the entity can decrypt the data and access it. Otherwise the sensitive data cannot be accessed. The trust and security values for each instance are considered for the final decision to disclose the information demanded by any entity at that point of time.

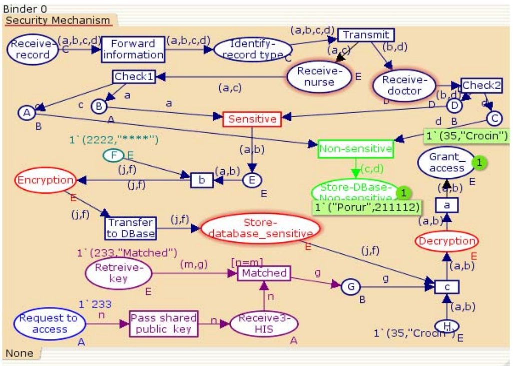

Fig. 4. CPN for security mechanism.

## IV. Integrated Trust Security & Privacy

The privacy of PHI can be considered as the scalar product of allowable uncertainty and the anticipated security risk to permit allowable disclosure. In Figure 5, a covered entity logs into the network by giving personal information and access the network by entering his or her medical information. This medical data is encrypted and checked in accordance with the security compliance as shown in the Security Risk Evaluator stage in the figure. The covered entities who are not authenticated are declared as anonymous and discarded as per the transition shown in figure that collects the number of such anonymous users. The authenticated covered entities are sent to the trust evaluator stage. However, these entities may have a few risks associated with them and based on these risks, the entities or their actions can be declared as pseudonymous or observable for future tracing. But in some situation, an authenticated entity may try to misuse the system for its personal benefits like fraud claim or replicated medical bill claims. Apart from maintaining high level privacy, the information system should be so modeled to prevent such thing to happen. In the proposed *Fradulent_detector* process, another binder with permitted states that are represented as a set $S_p$

$S_p$={ Login_Insuranceagent, Retreive_agentID,
Retreive_agentPWD, Reveal_PHI,
Access_bill, Provide-insurance,
Login_patient} and the permitted transitions and color sets are given as two different sets, $T_p$ and $C_p$ respectively.

$T_p$={Authenticate, Carry PID CID, Check PID,
Transfer CID, Match CID}

$C_p$={T, S, R, W}

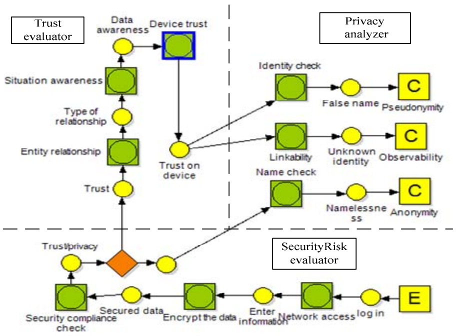

Fig. 5. Integrated model for TSP

The model in Fig. 6 illustrates a scenario in which insurance is provided to a patient after satisfying a few security mechanisms as per the health information standards and business norms. In such a system, once the insurance agent logs on to the system with the identity number and password,

TABLE I: DETECTION COST OF PRIVACY
TOTAL NUMBER OF TOKENS GENERATED=2000.

| Case | No. of Tokens | Wait Time | Cycle Time | Work Time | Cost |
|---|---|---|---|---|---|
| Anonymity | 1011 | 56.81 | 73.121 | 16.31 | 55.71 |
| Pseudonymity | 486 | 60.65 | 88.61 | 27.97 | 100.13 |
| Observability | 503 | 58.03 | 85.54 | 27.5 | 99.96 |

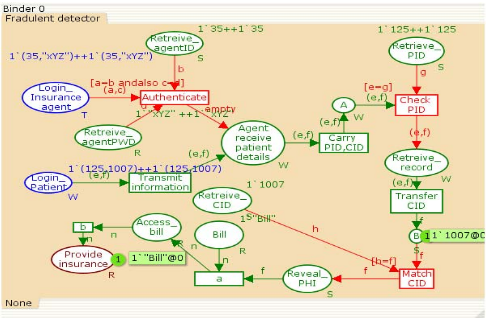

Fig. 6. CPN for fradulent detection.

He will be authenticated to receive patient details. The patient also logs on to the system by entering his credentials at the appropriate entry point. Now the patient provides his medical identity and case identity to the insurance agent. The insurance agent makes use of these details and looks into the medical database to check if the provided details are valid and also if they are eligible for the insurance to be claimed. The database that is searched by the insurance agent is well maintained and has no data redundancy. So, once the agent gets the information from the database he can be sure of the information obtained as it is not redundant. If both identities are matched then insurance is provided. Once insurance has been provided for a particular case then its corresponding case identity is deleted from the database. By this process, Data Redundancy can be avoided which further ensures that the patient will not be able to use the same case identity more than one time and therefore it avoids the situation where the patient can get credit from the insurance agent more than once. For the next time if the patient or the agent tries to claim insurance for the same case, then insurance cannot be provided as the relevant details are not found in the database. The Table I represents the detection cost of privacy for anonymous, pseudonymous and observable users by taking into account the wait time and work time of each of the type of users and the Table II depicts the resource utilization by the various stages in the integrated TSP model.

TABLE II: RESOURCE UTILIZATION

| RoleName | %Busy |
|---|---|
| SecurityRiskEvaluator | 98.91 |
| TrustEvaluator | 50.24 |
| PrivacyAnalyzer | 32.84 |

TABLE III: CASE LIST

| Case | Wait Time | Cycle Time | Work Time | Cost |
|---|---|---|---|---|
| 1 | 0 | 22.43 | 22.43 | 89.42 |
| 2 | 3.93 | 33.94 | 30 | 102.9 |
| 3 | 1.03 | 23.32 | 22.28 | 65.91 |
| 4 | 44.44 | 67.94 | 23.5 | 89.12 |
| 5 | 83.03 | 117.24 | 34.21 | 111.6 |
| 6 | 91.54 | 120.89 | 16.86 | 58.4 |
| 7 | 101.3 | 113.61 | 12.33 | 48.21 |
| 8 | 91.11 | 121 | 29.89 | 105.9 |
| 9 | 95.48 | 116.83 | 21.35 | 86.5 |
| 10 | 10.68 | 33.11 | 22.43 | 87.74 |
| 11 | 29.32 | 58.12 | 28.8 | 99.8 |
| 12 | 33.24 | 51.97 | 18.73 | 59.83 |
| 13 | 32.96 | 66.43 | 33.47 | 114.4 |
| 14 | 43.81 | 73.74 | 29.93 | 103.9 |
| 15 | 29.28 | 41.65 | 12.3 | 46.06 |

Table III discusses about the time taken by the tokens and their relative costs. In this the wait time is the time for which the token has to wait once it is generated until it enters the working phase and the time taken by the token to reach the final transition once it is out of the wait state is known as the work time. The cycle time is the aggregated value of both the wait time and work time, it is generally the time taken by the token to reach the final transition once it is generated

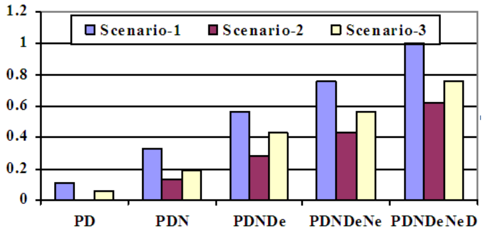

Fig. 7. Normalized trust values.

Fig. 7 shows a comparison between three different scenarios with each scenario has a different sequence of the covered entities involved. The chart illustrates the normalized trust value against the covered entities involved at various stages in the healthcare system where 'P' represents patient, 'D' represents doctor, 'N' represents nurse, 'De' represents device and 'Ne' represents network.

Fig. 8 classifies the total tokens generated into three different categories like anonymity, pseudonymity and Observability. The work time and cost for each category is obtained as a result of the simulation.

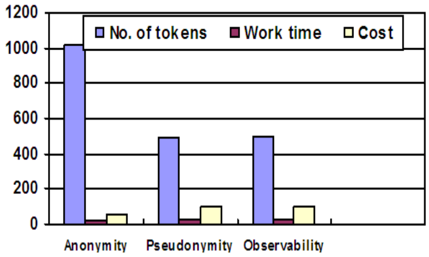

Fig. 8. Classification of generated tokens

## V. CONCLUSION

The privacy assurance is achieved in any healthcare information system by systematically considering the variations of trust between the covered entities and suitably varying the security mechanisms based on the actions performed by the entities. A process cycle named Trust-Security-Privacy-Disclose is proposed with essential sub-processes and the quantitative values for trust and risk values are determined with the help of Colored Petri Net simulator. The performance and the correctness of the integrated model is validated by the time and cost factor for different number of permitted tokens and creating different scenarios respectively. The work identifies that the sensitivity of the critical medical data field in any patient record is responsible for the complexity of the model. The fraudulent action by the patient through reclaiming the medical bill can also be detected in the model. The work can be further enhanced by identifying the behavior patterns of the patient or the agent in disclosing the protected health information in the future work.

## REFERENCES

[1] World Academy of Science, Engineering and Technology (WASET), 2005.

[2] G. T. Jun *et al*, "Healthcare Process Modelling," *International Journal for Quality* in *Healthcare*, vol. 21, no. 3, pp. 214-224, 2009.

[3] P. Dullabh, "Process Transformation in Healthcare Processes," *Tunitas group*, 2003.

[4] J. Puustijarvi *et al*, "Reserving Clinical Resources for Healthcare Processes," *ICDS*, pp. 92-97, 2010.

[5] A. A. Rad *et al*, "An Evaluation Framework for Business Process Modeling Languages in Healthcare," *Journal of Theoretical and Applied Electronic Commerce Research*, Universidad de, Talca-Chile, 2009.

[6] A. C. Weaver, "A Security Architecture for Data Privacy and Security," Department of CSE, University of Virginia, 2003.

[7] Y. Ding *et al*, "Model-Driven Application-level Encryption for the Privacy of E-health Data," *International conference on Availability, Reliability and security, IEEE* 2010, pp.341-346, 2010.

[8] HIPAA and HITECH Acts. [Online]. Available: http://www.sophos.com

**Chandrasekaran Subramaniam** is currently working as Professor in the department of Computer Science and Engineering of Kumaraguru College of Technology, Coimbatore,Tamilnadu, INDIA. He was awarded M.E.,in computer science and engineering and Ph.D degree in the area of Reliability Enhancement of Hybrid Fault Tolerant System by Anna University, Chennai, INDIA. He is an active researcher in the area of context aware computing, Reliability techniques in embedded systems and privacy assurance techniques. He is a member of IEEE, ACM, IET, SEI, WSEAS and ASQ. He published more than 60 papers in reputed international conferences and journals. He is at the age of 54 and have been acting as member in many international program committees of various conferences and acting as a chair cum reviewer in many journals and workshops.