Design of Smoke Screening Techniques for Data Surreptitiousness in Privacy Preserving Data Snooping Using Object Oriented Approach and UML

K.Satheesh Kumar,Indumathi.J *,Dr.G.V.Uma
***Corresponding author**

Department of Computer Science and Engineering, Anna University,
Chennai – 600 025.Tamilnadu,India

"Walking on water and developing software from a specification are easy if both are frozen."

Edward V Berard

"Creativity involves breaking out of established patterns in order to look at things in a different way."

Edward de Bono

Summary

A smoke screen is a release of smoke in order to disguise the movement or location of military units such as infantry, tanks or ships. while smoke screens would formerly have been used to conceal movement from enemies line of sight, modern technology means that they are currently also available in new forms; they can screen in the infrared as well as visible spectrum of light to prevent detection by infrared sensors or viewers, available for vehicles is a super dense form used to prevent laser beams of enemy target designators or range finders on vehicles and to obscure the sensitive data. Surreptitiousness is used in the perspective of amassing a data trickily secret.

In this paper, we focus primarily on solitude issues in Data Snooping, conspicuously when data are shared before mining, the means to shield it with Unified Modelling Language diagrams. The elucidation for competent systematization and sustain of facets, analysis, and diverse user groups are based on a common conceptual privacy model, and a concepts-services-mechanisms-algorithms-data scheme with use of UML or Unified Modelling Language. Tackling privacy preservation is complex since it should also pledge for well-founded Data Snooping results. Both these issues are equal and orthogonal in direction and it should clout a perfect balance between the two mechanisms. Under such state of affairs there arises the serious need for a rethinking mechanism to make obligatory privacy safeguards without trailing behind the gains of knowledge mining. These mechanisms can escort to novel privacy control methods to translate a database into a new one in such a way as to safeguard the main features of the original database for mining.

Conscientiously, we deal with the problem of transforming a database to be shared into a new one that smoke screens clandestine information while preserving the general patterns and trends from the original database. To focus on this exigent problem, we propose an amalgamated scaffold for Privacy Preserving Data Mining that ensures that the mining process will not trespass privacy up to a certain degree of security. The scaffold framework encompasses a family of privacy-preserving data transformation methods and library of algorithms. Our exploration concludes that our Privacy Preserving Data Mining framework is effective, meets privacy requirements, and guarantees well-founded Data Snooping results while shielding vulnerable information (e.g., sensitive knowledge and individuals' privacy).

Key words:

Association Rules, Clustering, Confidence, Data Snooping, Data Sanitization, Privacy, Privacy Preserving Data Mining, Sensitive Data, Smoke Screening, Surreptitiousness, Unified Modelling Language.

1. Introduction

The Sensitive attributes of patient data need to be fully protected when broadcasted across institutional, provincial, countrywide, and possibly international networks. The UML methodology permits the privacy model to be portrayed with use of several diagrams, such as logical diagrams, use case diagrams, scenario and activity diagrams, collaborations and distribution diagrams,

Manuscript received April 5, 2008
Manuscript revised April 20, 2008

class and class structure diagrams, as well as component and sequence diagrams. In isolation, various views of the overall model are made available. Through the analysis of different occurring privacy preserving health care scenarios, we were able to define the use case type, applying the appropriate UML diagrams.

Conventional medical record maintenance poses ample obstacles to intruders, because those seeking to inspect records must have authorization. They can view records only in person. Moreover, because paper records were decentralized – a single patient's records maybe disjointed across a number of places – in the event of a rupture of security, illegitimate access would be restricted.

Electronic information systems are proposed to conquer corporeal barriers to accessing patient records and to assist the wedding of data from diverse sources into an integrated medical dossier. Electronic storage of medical information opens the prospective for access to the cosmos of populace with ability to gain entry to the data bank. Through remote access, patients' records may be viewed incognito, and, once access is gained, complete information may be obtainable. Even greater privacy invasions may effect from the application of scanning programs that utilize artificial intelligence to hunt files in order to identify records of interest. In addition, when medical records are linked to other electronic databases (e.g., social security, vehicle registration, and economic records including credit card purchases and bank accounts) the aptitude of potential intrusion by artificial intelligence is even outsized. These high tech information technologies make it more probable for folks to learn and misuse more about the private lives of others.

The remainder of this paper is organized as follows: Section 2 offers an overview of the privacy preserving Data Snooping. In this section we have also analyzed the different problems in Data Snooping and the existing solutions. Section 3 discusses the problem statement, PPDM techniques for the health care services. Section 4 presents the block diagram, PPDM techniques for the health care services. Section 5 discusses the implementation details using UML diagrams. Section 6 concludes this paper with a brief summary.

2. Literature Survey

From a knowledge discovery perspective, this study is ensnared in a particular shape of data privacy whereby individual, confidential attribute values are not made available to lawful users, but aggregate 'relationships' of the database are made accessible. Privacy is one of the areas of security with trade-offs. In these narrow terms, one group's interest is to keep the information private whilst the other group's interest is to obtain that information.

The target of data mining is to extract or ‘‘mine’’ knowledge from large voluminous data obtained from copious dissimilar sites and is used to shore up both decision-making processes and to endorse social goals. However, the sharing of data has also raised a number of ethical issues as those of privacy, data security, and intellectual property rights. From a knowledge discovery perspective, this study is ensnared in a particular shape of data privacy whereby individual, confidential attribute values are not made available to lawful users.

Multifarious issues, such as those concerned in Privacy Preserving Data Mining (PPDM), cannot simply be addressed by restricting data collection or even by restricting the secondary use of information technology [16, 11,19]. A fairly accurate explanation could be adequate, depending on the relevance since the suitable altitude of privacy can be interpreted in diverse contexts [2,3]. In some applications (e.g., association rules, classification, or clustering), an apt equilibrium between a want for privacy and knowledge discovery should be originated.

Privacy-related issues ought to be pre-identified from the foundation of system development so that equivalent deterrence mechanisms can be integrated into the system design [14]. Abundant approaches have been recommended for early privacy analysis with shifting degrees of automation of the analysis method; the hard-core formal methods community endorses using formal specification languages. Hence forth formal verification and validation can be systematically performed on formal requirements [21, 6, 9]. Some of the approaches try to create certifiable code directly from the verified specifications [10], bypassing the design stage, in order to ensure that the safety properties verified in the requirements are preserved in the implementation.

The use of formal specification will only be a theoretical approach in many application areas, where the level of privacy often needs to be attuned in trading off with suppleness and reusability. The use of a supple requirements specification language is time and again indispensable for flawless transition from requirements to design, particularly in component-based system development, where bendable design architecture is enviable. In such cases, casual or semi-casual requirements can be deciphered into a prescribed specification for reasoning about the privacy aspect of the requirements [15, 22,23]. The appreciative of the object formal specification language requires complicated mathematical skills, which is a major barrier alongside prescribed methods being a routine part of software engineering processes in practice, thereby making these translations very difficult to perform in practice. We are also not able to understand how to incorporate the analysis result from requirements engineering into the early stage of the design process. In this exertion, we recommend an easy-to-use formal approach for privacy analysis in the early stage of requirements engineering in the context of component-based software development. Our approach is to provide privacy analysis capability while maintaining flexibility and usability which is addressed by using semi-formal use cases with templates [8], which can be then be methodically translated into a formal specification language.

Use cases [8] are a fashionable means to specify behavioral desires of software systems, predominantly due to their eagle-eyed style which can be basically tacit by Engineers and users. This scenery of use cases can sanction skilled proclamation in the midst of stakeholders — a principal reason why they are preferred to prearranged specification languages in exercise. Use cases consist of three artifacts; use case diagrams, documentary metaphors of use cases, and use case scenarios. Use case diagrams demonstrate the link amongst use cases and actors, plentiful an outline of the system behavior.

Preserving privacy when data are pooled for mining is an exigent predicament. The usual methods in database security, such as access control and authentication [18, 8, 17] that have been adopted to lucratively handle the access to data present some restrictions in the milieu of data mining. While access control and authentication protections can preserve against direct disclosures, they do not address disclosures based on inferences that can be strained from released data [12, 4, 13]. Preventing this sort of inference discovery is beyond the reach of the existing methods [16, 19].

Evidently, privacy issues pretence new challenges for fresh uses of data mining technology [7, 20, 15]. These technical challenges indicate a grave need to rethink mechanisms to concentrate on some issues of privacy and accuracy when data are either shared or exchanged before mining. Such mechanisms can lead to novel privacy control methods to alter a database into a new one that conceals private information while preserving the general patterns and trends from the original database

In this paper we address the issue of privacy preserving Data Snooping for a scenario in which the parties owning confidential databases wish to run a Data Snooping algorithm on the union of their databases, without revealing any sensitive information.

3. Privacy Preserving Data Snooping

Privacy Preserving Data Snooping Analysis is an amalgamation of the data of heterogeneous users without disclosing the private and susceptible details of the users.

3.1. Problem Statement

Stipulation of a comprehensible but prescribed approach for early privacy preserving analysis in the milieu of component-based software development, in order to evaluate and compare with apiece and all the techniques in a universal platform and to devise, build up and execute functionalities like a User friendly framework, portability etc., which will be the basis for ascertaining the suitable technique for a given type of application.

The new-fangled system as in figure 3.1.will be state of the art and will have a Windows-based desktop interface to allow knowledge workers to enter Domain information, enter heterogeneous data, select the knowledge workers preferences (such as different Data Obscuring methods), and create various reports.

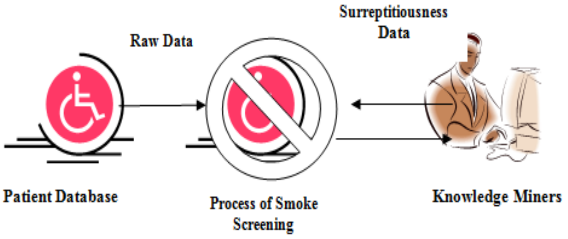


Figure 3.1. Privacy Preserving Data Snooping

3.2. Problem Description

The contributions involved in the Privacy Preserving framework as shown in figure 3. 1 is to delineate the outline description of Privacy Preserving techniques in the milieu of Pattern Analysis, develop a Framework for Privacy Preserving Pattern Analysis and to acquire datasets, policies, rules and regulations as input. It derives information from the datasets, manages the derived information, and provides information to Knowledge consumers with privacy preserved data.

Our approach aims at adopting formal privacy analysis while sustaining suppleness and reliability throughout the development process. To this end, we use semi-formal use cases with templates that can be methodically converted into any formal specification language, whose execution environment integrates automated verification tools. Consistency flanked by use cases and the high-level component design is retained through a methodical changeover, so that the result of the Privacy Preserving analysis can be easily reflected in the design model. Some of the approaches try to generate certifiable code directly from the verified specifications [10].

3.3. Classification of Privacy Preserving Techniques

For privacy preserving Data Snooping the following classification of privacy preserving techniques is discussed below (as in figure 3.2.).

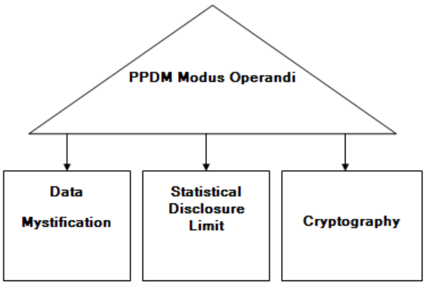


Figure.2. Classification of Privacy Preserving Data Snooping

3.3.1.Cryptography-Based Techniques use different cryptography techniques like secure multiparty computation. Thus, the raw data can be transmitted and received in a secured manner.

3.3.2.Data Mystification Techniques mystify the original values of a database and present the amended database for Data Snooping, thus ensuring privacy. In general, data modification techniques seek at arriving at a proper balance between privacy preservation and knowledge confession.

(i). **Data Swapping techniques** substitute the values in the records of the original database with a new one. The swapped database maintains the same probability distribution as the original one.

(ii). **Data Hiding Techniques** aim at hiding some sensitive information when data are shared for mining. These techniques can be classified either based on data organization or on data nature like the ones shown in figures.

(iii). **Data Perturbation techniques** twist the data to protect individuals' privacy by introducing an error (noise) to the original data. The noise is used to engender the new perturbed database which is subjected to mining. Miners should be able to obtain valid results (e.g., patterns and trends) from the perturbed data.

(iv). **Data Sampling** aims at releasing data for only a sample of a population rather than handing over the full database.

(v). **Aggregation or Merging** which is the combination of several values into a coarser category.

3.3.3. SDL (Statistical Disclosure Limitation) limits the disclosure of statistical data.

Like the chicken-and-egg situation it is pretty confusing to shape out which technique is best for any domain. We decide the best method by taking into account several features and factors.

4. Architecture of the Proposed Work

4.1. Sharing Medical Research Data

Investigating and analyzing the predominance, frequency, and peril factors of diseases is decisive to understanding and treating them. Such analyses have considerable bang on policy decisions. A palpable precondition to (carrying out) such studies is to have the indispensable data available.

First, patient data has to be collected from several health care providers, here hospitals. It has to be subjected to data sanitization and then integrated. The data that is required for pattern evaluation and knowledge mining alone is selected by filtering. These heterogeneous data’s are converted to the desired format. This course of action is tremendously time consuming and toil demanding. Privacy concerns are a major hindrance to streamlining these efforts. Infringing privacy can lead to significant dent to individuals both materially and psychologically. One more predicament is the leeway of discrimination against various sub-groups from outwardly irrefutable statistical results. Correspondingly, health care providers themselves threat trouncing by leaking precise data reflecting their concert and weaknesses. Privacy is addressed nowadays by preventing propagation to a certain extent than integrating privacy constraints into the data sharing process. Privacy-preserving amalgamation and partaking of research data in health sciences has become vital to enabling scientific innovation.

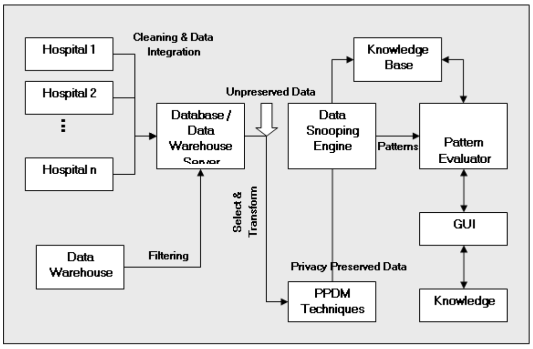


Figure.4.1. System Architecture

4.2. Unpreserved Data Snooping

Data of several patients are collected in hospitals and kept for records. These patient data are collected from several hospitals. The patient data will be in different formats. These data are given as an input to the database server. The data will be converted into a desired format and stored in these database servers. We get one more data input from the data warehouse and send it to the data warehouse servers. From the output of the database servers we select only the desired data and transform to the desired format for which we need to run the Data Snooping engine. We use different Data Snooping techniques like classification, association, clustering etc., on the input unpreserved data. The extracted patterns are sent to the pattern evaluator and the interesting patterns are visually shown for further analysis.

4.3. Privacy Preserved Data Snooping

Data of quite a lot of patients are collected in hospitals and held in reserve for records. These patient data are collected from several hospitals. The patient data will be in dissimilar formats. These facts are keyed in to the database server. This input data will be converted into the preferred format and stored in these database servers. The second data input coming from the data warehouse is sent to the data warehouse servers. The data warehouse server contains a collection about various things. From the database server pool, we choose only the most wanted data and transform it to the desired format. This transformed data is the input data on which we need to run the Data Snooping techniques. Instead, of sending the data directly for Data Snooping we make the data to be obscured by using different privacy preserving Data Snooping techniques with the intention of preserving the sensitive information. This privacy preserved obscured data is the subjected to the various different Data Snooping techniques like classification, association, clustering etc., on the input preserved data. The extracted patterns are sent to the pattern evaluator and the interesting patterns are visually shown for further analysis.

5. Requirements Specification Using Use Cases

Use cases [1] are a trendy means to denote behavioral requirements of software systems, mainly due to their perceptive style which can be effortlessly understood by engineers as well as non-technical stakeholders. This temperament of use cases can promote efficient communication among stakeholders — a chief reason why they are favored to formal specification languages in practice.

Use cases consist of three artifacts; use case diagrams, textual descriptions of use cases, and use case scenarios. Use case diagrams exemplify the rapport among use cases and actors, giving an impression of the system behavior.

Figure 9.1 shows an essential part of the use case diagram specifying the behavior of a PPDM system. An actor exterior to the box characterizes an external entity cooperating with the system. The use cases within the box characterize system functionalities afforded to the external actors, where each use case can include or be extended by other use cases. The use case diagram is complemented by textual use cases with a varying degree of formality — from an informal, casual description to the use of a semi-formal template specifying details of each use case.

Under the current state of affairs of hi-tech developments which has obliterated the distinctions of patient data kept in private and public; we are incapable of shielding the patient privacy. The patient records are kept in private practice, clinics, general hospitals and multi-specialty clinics. In budding health care systems, the habitual physician's responsibility as a custodian of patient privacy is under grave assault. Customary relationships between physicians and patients have been transformed so that physicians may no longer be able to have power over medical information in the manner they once did. Furthermore, new information technologies have enhanced the significance and latent uses of medical data;

as a result, third-party demands for right to use have increased, with attendant risks to patient privacy.

The highly sensitive information of the patients has to be conserved and then mined for effective data dredging. The UML (Unified Modeling Language) methodology allows the PPDM model to be described with use of several diagrams, such as logical diagrams, use case diagrams, object diagrams, component diagrams, deployment diagrams, scenario and activity diagrams, interaction diagrams, composite structure diagrams, state machine diagrams, collaborations and distribution diagrams, class and class structure diagrams. As such, an assortment of views of the overall model is made available. (Refer Appendix).

The scenarios as sequences of obligatory links between objects as instances of classes within corporeal application environments offer to stand for all earth-shattering requirements. They are used to portray the exploit of key mechanisms and to determine the essential ranges of operational cases. Principally, the use case, sequence, and activity diagrams are needed to institute all desired security services and mechanisms. The use case (Refer Appendix) defines a detailed structure for the fitness care information system. Departing from an abstract use case type, fitness care data processes as well as the communication or interaction with health professionals, patients, and related parties are described in the light of medical information protection.

The use case types makes an establishment of contact between the doctor; the diagnosis or treatment by the doctor; the formation of an exact concern and healing plan for the patient; instigation, performance and control of activities; right to use patient data; documentation of health care information; and emblematic situation in the shared hospital environment constitutes the doctor's demand for precise data of an individual patient needing instant care.

The next constitutes the reciprocally obtainable user authentication through cryptographic algorithms. The user's identity certificate is handled by a Trusted Third Party. Sensitive data compilation for recording, meting out, storage or distribution cannot be allowed without the patient permission. The initialization of exchanges involves joint detection and backing of the dissimilar partners, which is established by the certificates, issued by the TTP.

The information precondition use case defines the selection of the lawful patient data for information transfer to the requesting party. The function and communication safekeeping services, which handle the issues of integrity, confidentiality, and accountability, have likewise been included into the model. The business enterprise partners have developed a layered expansion of this security model, which is based on the UML approach. The complex and standard methodology is able to support the abundant study, design and installation of secure hospital information systems.

6. Component Design from Use Cases

Let us see how use case specifications are used to obtain a high level component design in the milieu of the component-based software development approach KobrA [5].

7. UML PPDM Models

Under the current state of affairs of hi-tech developments which has obliterated the distinctions of patient data kept in private and public; we are incapable of shielding the patient privacy. The patient records are kept in private practice, clinics, general hospitals and multi-specialty clinics. In budding health care systems, the habitual physician's responsibility as a custodian of patient privacy is under grave assault. Customary relationships between physicians and patients have been transformed so that physicians may no longer be able to have power over medical information in the manner they once did. Furthermore, new information technologies have enhanced the significance and latent uses of medical data; as a result, third-party demands for right to use have increased, with attendant risks to patient privacy.

The highly sensitive information of the patients has to be conserved and then mined for effective data dredging. The UML (Unified Modeling Language) methodology allows the PPDM model to be described with use of several diagrams, such as logical diagrams, use case diagrams, object diagrams, component diagrams, deployment diagrams, scenario and activity diagrams, interaction diagrams, composite structure diagrams, state machine diagrams, collaborations and distribution diagrams, class and class structure diagrams. As such, an assortment of views of the overall model is made available. (Refer Appendix).

The scenarios as sequences of obligatory links between objects as instances of classes within corporeal application environments offer to stand for all earth-shattering requirements. They are used to portray the exploit of key mechanisms and to determine the essential ranges of operational cases. Principally, the use case, sequence, and activity diagrams are needed to institute all desired security services and mechanisms. The use case (Refer Appendix) defines a detailed structure for the

fitness care information system. Departing from an abstract use case type, fitness care data processes as well as the communication or interaction with health professionals, patients, and related parties are described in the light of medical information protection.

The use case types makes an establishment of contact between the doctor; the diagnosis or treatment by the doctor; the formation of an exact concern and healing plan for the patient; instigation, performance and control of activities; right to use patient data; documentation of health care information; and emblematic situation in the shared hospital environment constitutes the doctor's demand for precise data of an individual patient needing instant care.

The next constitutes the reciprocally obtainable user authentication through cryptographic algorithms. The user's identity certificate is handled by a Trusted Third Party. Sensitive data compilation for recording, meting out, storage or distribution cannot be allowed without the patient permission. The initialization of exchanges involves joint detection and backing of the dissimilar partners, which is established by the certificates, issued by the TTP.

The information precondition use case defines the selection of the lawful patient data for information transfer to the requesting party. The function and communication safekeeping services, which handle the issues of integrity, confidentiality, and accountability, have likewise been included into the model. The business enterprise partners have developed a layered expansion of this security model, which is based on the UML approach. The complex and standard methodology is able to support the abundant study, design and installation of secure hospital information systems.

8. Conclusions

In this paper we are defining privacy preservation in data mining, and the implications of benchmark privacy doctrine in information detection and we are advocating a few policies for PPDM based on these privacy principles. These are vital for the development and deployment of methodological solutions and will let vendors and developers to construct unyielding advances in the upcoming of PPDM.

The moral, lawful, and societal boundaries on Data Snooping relate to privacy and security considerations, trepidation of lawsuits, and the want for a balance of the expected profit of examination against any hassle or possible damage to the person's privacy. Methods of medical Data Snooping must concentrate on the heterogeneity of data sources, data structures, and the incidence of lost values for both scientific and social reasons. The natural history of an ailment affects statistical hypotheses in a mysterious way. Statistical hypothesis tests often take the form of a trap or a contest with a winner and a loser. The application of this model to the natural processes of medicine is problematic. For all its perils, medical Data Snooping is the most rewarding. For a suitably-formulated technical question, trillions of data-elements can be brought to abide on finding a solution. For a fittingly-formulated medical question, discovering an answer could mean prolonging a life, or giving ease to an ill person. These probable plunder more than reimburse for the many bizarre difficulties along the passageway to triumph.

Privacy concerns are a major hindrance to streamlining these efforts. Infringing privacy can lead to significant dent to individuals both materially and psychologically. One more predicament is the leeway of discrimination against various sub-groups from outwardly irrefutable statistical results. Correspondingly, health care providers themselves threat trouncing by leaking precise data reflecting their concert and weaknesses. Privacy is addressed nowadays by preventing propagation to a certain extent than integrating privacy constraints into the data sharing process. Privacy-preserving amalgamation and partaking of research data in health sciences has become vital to enabling scientific innovation.

The Unified Modelling Language is made up of a set of principally graphical portrayal techniques for the condition and citations of object-oriented systems. We describe the experiences gained while using UML for the growth of a small disseminated coding for scheduling of PPDM techniques in hospitals. Our driving force in this case study is not only to assess the techniques given by UML, but also to learn their interrelationships and their meticulous utilization from requirements analysis to execution. Because the instance is comprehensive and self-reliant and provides systematic strategy and hints, it can also be used as a tutorial for UML and for object-oriented improvement in common. It can be used to improve the quality of health-care by preserving the patient's sensitive information and the implementation of internet-based telemedicine system for community health-care is feasible; collaboration with related institutions and parties will play an important role. Our exploration concludes that our Privacy Preserving Data Mining framework is reusable, customizable, and effective, meets privacy requirements, and guarantees well-founded Data Snooping results while shielding vulnerable information (e.g., sensitive knowledge and individuals' privacy).

9.References

[1].Alistair Cockburn. Writing Effective Use Cases. Addison-Wesley Publishing,Company, 2000.

[2].C. Clifton, W. Du, M. Atallah, M. Kantarcio_glu, X. Lin, and J. Vaidya. Distributed Data Mining to Protect Information Privacy. Proposal to the National Science Foundation, December 2001.

[3].C. Clifton. Using Sample Size to Limit Exposure to Data Mining. Journal of Computer Security, 8(4):281-307, November 2000.

[4].C. Farkas and S. Jajodia. The Inference Problem: A Survey. SIGKDD Explorations, 4(2):6{11, December 2002.

[5].Colin Atkinson, Joachim Bayer, and Christian Bunse et al. Component-based Product Line Engineering with UML. Addison-Wesley Publishing Company, 2002.

[6].Constance Heitmeyer, James Kirby Jr., Bruce Labaw, Myla Archer, and Ramesh Bharadwaj. Using abstraction and model checking to detect safety violations in requirements specifications. IEEE Transactions on Software Engineering,24(11):927–948, November 1998.

[7].D. E. O'Leary. Some Privacy Issues in Knowledge Discovery: The OECD Personal Privacy Guidelines. IEEE EXPERT, 10(2):48 -52, April 1995.

[8].D. F. Ferraiolo and R. Kuhn. Role-Based Access Control: Features and Motivations. In Proc. of the 11th Annual Computer Security Applications Conference, pages 241 - 248, New Orleans, LA, USA, Dec. 1995.

[9].Farnam Jahanian and Aloysius K. Mok. Safety analysis of timing properties in real-time systems. IEEE Transactions on Software Engineering, 12(9):890–904, September 1986.

[10].Joachim Thees and Reinhard Gotzhein. The experimental esterel compiler - automatic c generation of implementations from formal specifications. In Proceedings of FMSP ’98, The Second Workshop on Formal Methods In Software Practice., 1998.

[11].L. Brankovic and V. Estivill-Castro. Privacy Issues in Knowledge Discovery and Data Mining. In Proc. of Australian Institute of Computer Ethics Conference (AICEC99), Melbourne, Victoria, Australia, July 1999.

[12].L. Sweeney. k-Anonymity: A Model for Protecting Privacy. International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5):557-570, 2002.

[13].L. Willenborg and T. D. Waal. Statistical Disclosure Control in Practice. Springer- Verlag, 1996.

[14].Nancy G. Leveson. Software safety in embedded computer systems. Communications of the ACM, 34(2):34–46, 1991.

[15].Office of the Information and Privacy Commissioner. Data Mining: Staking a Claim on Your Privacy, Toronto, Ontario, January 1998.

[16].R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic Databases. In Proc. Of the 28th Conference on Very Large Data Bases, Hong Kong, China, August 2002.

[17].R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-Based Access Control Models. IEEE Computer, 20(2):38 - 47, 1996.

[18].S. Castano, M. Fugini, G. Martella, and P. Samarati. Database Security. Addison- Wesley Longman Limited, England, 1995.

[19].S. R. M. Oliveira and O. R. Zaiane. Foundations for an Access Control Model for Privacy Preservation in Multi-Relational Association Rule Mining. In Proc. of the IEEE ICDM Workshop on Privacy, Security, and Data Mining, pages 19 -26, Maebashi City, Japan, December 2002.

[20].W. Kl¨osgen. KDD: Public and Private Concerns. IEEE EXPERT, 10(2):55 -57, April 1995.

[21].William Chan, Richard J. Anderson, Paul Beame, Steve Burns, Fracesmary Modugno, David Notkin, and John D. Reese. Model checking large software specifications. IEEE Transactions on Software Engineering, 24(7):498–520, July 1998.

[22] Woo Jin Lee, Sung Deok Cha, and Yong Rae Kwon. Integration and analysis of use cases using modular petri nets in requirements engineering. IEEE Transactions on Software Engineering, 24(12):1115–1130, 1998.

[23] Wuwei Shen and Shaoying Liu. Formalization, testing and execution of a use case diagram. In 5th International Conference on Formal Engineering Methods, 2003.

10.APPENDIX

This presents us with the a sample of the diagrams for this case study.

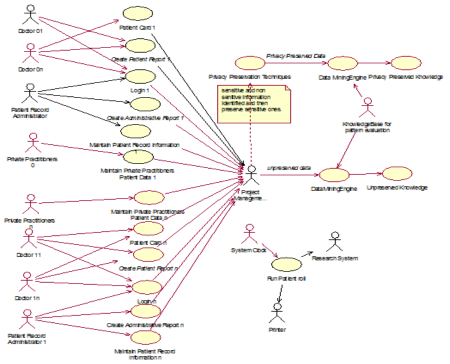
Unified Modeling Language -UML is a modeling language and only specifies semantics and notation but no process is currently defined. Thus, we decided to do the analysis as follows: Use case diagram,Class diagram,Sequence diagram,Collaboration diagram,State diagram

A. Analysis

A.1. Use case diagram

Use case description:

This gives us a comprehensive portrayal of how a system will be used. It endows us with an outline of the projected functionality of the system. PPDM main success scenario (basic flow) that can be extracted from Use Case Diagram Understandable by laymen as well as professionals.

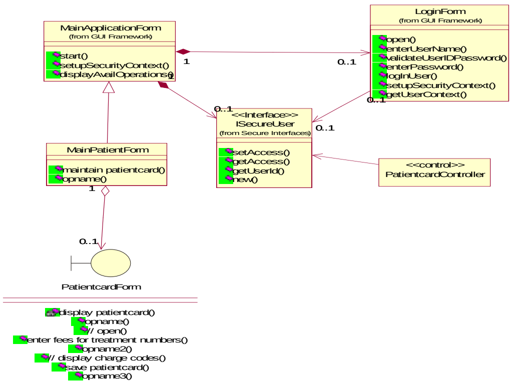

Doctor 01
Patient Card 1
Doctor 0n
Create Patient Report 1
Patient Record Administrator
Login 1
Create Administrative Report 1
Private Practitioners 0
Maintain Patient Record Information
Maintain Private Practitioners Patient Data 1
Privacy Preserved Data
Privacy Preservation Techniques
Data MiningEngine
Privacy Preserved Knowledge
sensitive and non sentive information identified and then preserve sensitive ones.
KnowledgeBase for pattern evaluation
unpreserved data
Project Manageme...
DataMiningEngine
Unpreserved Knowledge
Private Practitioners n
Maintain Private Practitioners Patient Data n
Doctor 11
Patient Card n
Create Patient Report n
Doctor 1n
Login n
Patient Record Administrator 1
Create Administrative Report n
Maintain Patient Record Information n
System Clock
Research System
Run Patient roll
Printer

PPDM System Use Cases

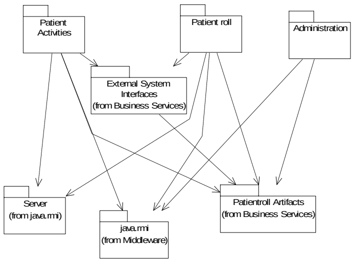
B. Class Diagram

Class diagrams show the static structure of the object, their internal structure, and their relationships.

Class diagram: Patient activities - Main (with Security)



Class diagram: Patient roll – Application/Main



C. State diagram

A state diagram shows the sequences of states an object goes in the course of its life cycle in response to stimuli, together with its responses and actions.

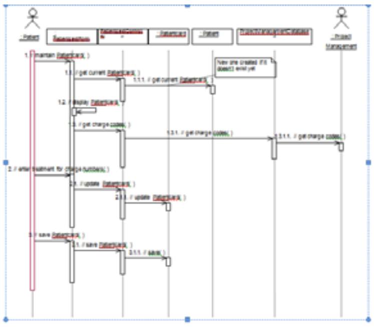
D. Design

The design phase should produce the detailed class diagrams, collaboration diagrams, sequence diagrams, state diagrams, and activity diagram. However, the PPDM problem is too simple for an activity diagram. Thus, we are not using an activity diagram for the PPDM problem.

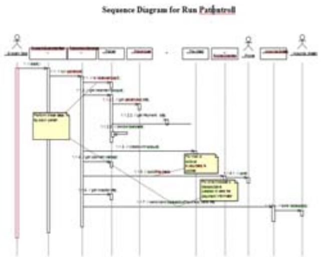
D.1. Sequence Diagram

A sequence diagram and collaboration diagram conveys similar information but expressed in different ways. A Sequence diagram shows the explicit sequence of messages suitable for modeling a real-time system, whereas a collaboration diagram shows the relationships between objects.

Sequence Diagrams: Sequence Diagram for Maintain Patientcard


Patient
Project Management
1: // maintain patientcard( )
1.1: // get current patientcard( )
1.2: // display patientcard( )
2: // enter treatment for charge numbers( )
3: // save patientcard( )

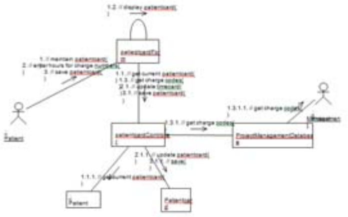
Sequence Diagram for Run Patientroll



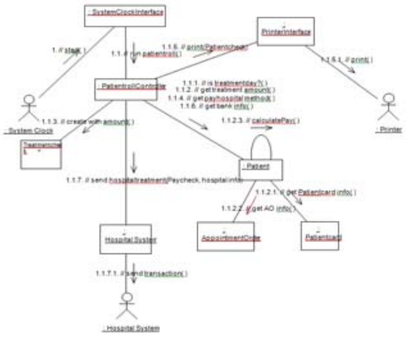
D.2. Collaboration diagram

Describes the set of interactions between classes or types
Shows the relationships among objects

Collaboration diagrams: Maintain Patientcard - Basic Flow


1.2: // display patientcard( )
1: // maintain patientcard( )
2: // enter hours for charge numbers( )
3: // save patientcard( )
Patient
PatientcardController
ProjectManagementDatabase
Patient
Patientca

Run Patientroll -Basic Flow



AUTHORS PROFILE

Satheesh Kumar. K received his Bachelor's Degree in Electrical and Electronics Engineering from JNTU, Hyderabad, India, and his Master's degree in Computer Science from Anna University, Chennai, India. His research interests include Data Mining, Database Security and Information Security. He is currently working for Tata Elexsi organization.

Indumathi.J received her M.E. from Anna University, Chennai, in 1992 and M.B.A from Madurai Kamaraj University, Madurai, India in 1994. She is working for Anna University as a Senior Lecturer. She is currently doing her Ph.D from Anna University, Chennai. Her field of interest spans and is not limited to Computer Science and Financial Management.Her research interests include Security for Data Mining, Databases, Networks, Computers, Software Engineering, Software Testing, Project Management, Biomedical Engineering, Genetic Privacy and Ontology.

G.V.Uma, a Polymath received her M.E. from Bharathidasan University, India in year 1995 and Ph.D. from Anna University, Chennai, India in 2002. She is working for Anna University as a Assistant Professor. Her research interests include Software Engineering, Genetic Privacy, and Ontology.Knowledge Engineering & Management.