ISSN 2321 -9017
Volume 2, No.5, August - September 2014
International Journal of Bio-Medical Informatics and e-Health
Available Online at http://warse.org/pdfs/ijbmieh01252014.pdf

SECURITY AND PRIVACY IN HEALTHCARE DATA MINING

Dr.E.Kesavulu Reddy
Assistant Professor Department Of Computer Science College Of Commerce Management & Computer Science
S.V.University, Tirupati, India-517502 Ekreddysvu2008@gmail.com

ABSTRACT

Managing privacy and security of healthcare information used to mine data by reviewing their fundamentals, components and principles as well as relevant laws and regulations. It also presents a literature review on technical issues in privacy assurance and a case study illustrating some potential pitfalls in data mining of individually identifiable information. This paper was recommendations for privacy and security good practices for medical data miners.

Keywords- Information security; privacy; confidentiality integrity; availability; risk management; privacy – protection.

1. INTRODUCTION

Health care organizations generally adopt information technology to reduce costs as well as improve efficiency and quality. Medical researchers hope to exploit clinical data to discover knowledge lying implicitly in individual patient health records. These new uses of clinical data potentially affect healthcare because the patient physician relationship depends on very high levels of trust. To operate effectively physicians need complete and accurate information about the patient. Data mining especially when it draws information from multiple sources poses special problems. For example, hospitals and physicians are commonly required to report certain information for a variety of purposes from census to public health to finance. This often includes patient number, ZIP code, race, date of birth, gender, service date, diagnoses codes (ICD9),procedure codes (CPT), as well as physician identification number, physician ZIP code, and total charges. Compilations of this data have been released to industry and researchers. Because such compilations do not contain the patient name, address, telephone number, or social security number, they qualify as de-identified and, therefore, appear to pose little risk to patient privacy. But by cross linking this data with publicly available databases, processes such as data mining may associate an individual with specific diagnoses.

2. PRIVACY AND SECURITY

Inappropriate disclosure, loss of data integrity, or unavailability may each cause harm. Recent laws and regulations such as HIPAA provide patients with legal rights regarding their personally identifiable healthcare information and establish obligations for healthcare organizations to protect and restrict its use or disclosure. Data miners should have a basic understanding of healthcare information privacy and security in order to reduce risk of harm to individuals, their organization or themselves.

2.1. Privacy in Healthcare Information

The term “privacy” bears many meanings depending on the context of use. Common meanings include being able to control release of information about one’s self to others and being free from intrusion or disturbance in one’s personal life. To receive healthcare one must reveal information that is very personal and often sensitive. We control the privacy of our healthcare information by what we reveal to our physicians and others in the healthcare delivery system. Once we share personal information with our caregivers, we no longer have control over its privacy. In this sense, the term “privacy” overlaps with “confidentiality” or the requirement to protect information received from patients from unauthorized access and disclosure. Thus, ethics, laws and regulations provide patients with certain rights and impose obligations on the healthcare industry that should keep patient health information from being disclosed to those who are not authorized to see it.

2.2. Security in Healthcare Information

The Internet has resulted in recognition that information technology security is of major importance to our society. This concern seems relatively new in healthcare, but information technology security is a well established domain. A large body of knowledge exists that can be applied to protect healthcare information. A general understanding of security can be obtained by understanding: • Security Components
• Security Principles
• Threats, Vulnerabilities, Control Me and Information Assurance. *a. Security Components*
• Confidentiality is the property that data or information is not made available or disclosed to unauthorized persons or processes.
• Integrity is the property that data or information have not been altered or destroyed in an unauthorized manner.
• Availability is the property that data or information is accessible and useable upon demand by an

authorized person.

- Accountability is the ability to audit the actions of all parties and processes which interact with the information and to determine if the actions are appropriate.

2.3. Security Principles

* Accountability Principle: The responsibilities and accountability of owners, providers and users of information systems and other parties concerned with the security of information systems should be explicit.

* Awareness Principle : In order to foster confidence in information systems, owners, providers and users of information systems and other parties should readily be able, consistent with maintaining security, to gain appropriate knowledge of and be informed about the existence and general extent of measures, practices and procedures for the security of information systems. *Ethics Principle: Information systems and the security of information systems should be provided and used in such a manner that the rights and legitimate interests of others are respected.

*Multidisciplinary Principle: Measures, practices and procedures for the security of information systems should take account of and address all relevant considerations and viewpoints, including technical, administrative, organizational, operational, commercial, educational and legal.

* Proportionality Principle: Security levels, costs, measures, practices and procedures should be appropriate and proportionate to the value of and degree of reliance on the information systems and to the severity, probability and extent of potential harm, as the requirements for security vary depending upon the particular information systems.

* Integration Principle: Measures, practices and procedures for the security of information systems should be coordinated and integrated with each other and with other measures, practices and procedures of the organization so as to create a coherent system of security.

* Timeliness Principle: Public and private parties, at both national and international levels, should act in a timely coordinated manner to prevent and to respond to breaches of security of information systems.

*Reassessment Principle: The security of information systems should be reassessed periodically, as information systems and the requirements for their security vary over time.

*Equity Principle: The security of information systems should be compatible with the legitimate use and flow of data and information in a democratic society.

2.4. Threats, Vulnerabilities, Control Measures And Information Assurance

Numerous [6] threats exist to computer systems and the information they contain originating from within and outside organizations. Some common threats include malicious code such as viruses, Trojan horses, or worms. Malicious code often takes advantage of vulnerabilities in operating system software but depends, too, upon organizational weaknesses such as the failure to deploy, update or train workers in the use of antivirus software. Malicious code may enable denial of service attacks, impersonation, information theft and other intrusions. Attacks by famous malicious code such as the Melissa or Love bug viruses highlight the threat of "hackers", outsiders with intent to harm specific organizations or network operations in general. Insiders with privileged access to network operations and a grudge against their employer actually wreak the most harm to say nothing of ill trained workers unintentionally making mistakes.

3. 3.HEALTHCARE OPERATIONS

Healthcare operations are the usual business operations of healthcare providers and health plans. Specifically included are: quality assessment and improvement activities, outcomes evaluation, development of clinical guidelines, population-based activities relating to improving health or reducing health care costs, protocol development, case management and care coordination, contacting of health care providers and patients with information about treatment alternatives; and related functions that do not include treatment, reviewing the competence or qualifications of health care professionals, evaluating practitioner, provider performance and health plan performance, conducting training programs in which students, trainees, or practitioners in areas of health care learn under supervision to practice or improve their skills . IV. DATA MINING FOR PRIVACY AND SECURITY As information technology has become commonly used to support the core processes of healthcare, enormous volumes of data have been produced. Numerous organizations desire access to this data to apply techniques of knowledge discovery. Privacy concerns exist for information disclosed without illegal intrusion or theft. A person's identity can be derived from what appears to be innocent information by linking it to other available data. Concerns also exist that such information may be used in ways other than promised at the time of collection. Statistical databases containing individually identifiable information including: conceptual, query restriction, data perturbation, and output perturbation approaches The conceptual model has not been implemented in an on-line environment and the others involve considerable complexity and cost and may obscure medical knowledge

3.1. Query restriction approach

Five methods have been developed to restrict queries:

1. Query-Set-Size Control: A method that returns a result only if its size is sufficient to reduce the chances of identification,

2.Query-Set-Overlap Control: A method that limits the number of overlapping entities among successive queries of a given user.

3. Auditing A method that creates up-to-date logs of all queries made by each user and constantly checks for possible compromise when anew query is issued.

4.Cell Suppression :A method that suppresses cells that might cause confidential information to be disclosed from being released.
5. Partitioning: A method that clusters individual entities into a number of mutually exclusive subsets thus preventing any subset from containing precisely one individual.

3.2. Input

This approach alters the data before permitting access to users. is inserted in the data that seeks to achieve anonymity and at the same time not change the statistical significance of query results. Users do not have access to the original data.

3.3. Output

This approach permits use of the original data, but modifies or renders the output incomplete. Techniques of output perturbation include processing only a random sample of the data in the query, adding or subtracting a random value that will not alter the statistical difference from the result, and rounding up or down to the nearest multiple of a certain base.

3.4. Specific Approaches

Rendering data anonymous assures freedom from identification, surveillance or intrusion for the subjects of medical research or secondary data analysis while allowing data to be shared freely among investigators A number of techniques exemplifying or combining the general approaches described above have been advocated to help address this issue, including:
1. Data aggregation
2. Data de-identification
3. Binning
4. Pseudonymisation
5. Mediated access

3.4.1 Data Aggregation

Providing access only to aggregate data while prohibiting access to records containing data on an individual constitutes one approach commonly advocated to reduce risks to privacy. Although this approach does protect privacy, it critically limits medical research. Clinical research requires prospectively capturing and analyzing data elements associated with individual patients. Outliers are often a major focus of interest. Aggregate data does not support such efforts.

3.4.2 Data de-identification

The [7] HIPAA Privacy Standard excludes de-identified health information from protected health information. De-identified health information may be used and disclosed without authorization. The HIPAA Privacy Standard considers information to have been deidentified by the use of either a statistical verification of deidentification or by removing 18 explicit elements of data. Such data may be used or disclosed without restriction. The details of these approaches are described in the pamphlet, 3.Protecting: A variety of approaches for this issue have been published including successful implementation of policies, procedures, techniques and toolkits that meet academic medical center needs and comply with the Privacy Standard [1],[2].

3 Binning

Binning deploys a technique for generalizing records in a database by grouping like records into a category and eliminating their unique characteristics 4. Pseudonymisation**:** This technique involves replacing the
true identities of the individuals and organizations while retaining a linkage for the data acquired over time that
permits re-identification under controlled circumstances [6]. A trusted third party and process is involved. The trusted third party and process must be strictly independent, adhere to a code of conduct with principles of openness and transparency, have project-specific privacy and security policies and maintain documentation of operating, reporting and auditing systems[4].

3.4.5 Mediated access

Mediated access puts policy, procedure and technology between the user and the data and, thus, illustrates a general point that all medical investigators should bear in mind: sound health information privacy and security programs include a range of controls. The security officer decides to approve, edit, or reject the information. An associated logging subsystem provides both an audit trail for all information that enters or leaves the domain, and provides input to the security officer to aid in evolving the rule set, and increasing the effectiveness of the system.". Public/ private keys and manages role-based access.

O V. 5.OTHER ISSUES IN EMERGING "PRIVACY TECHNOLOGY

Two kinds of privacy issues for computer science research have been identified: those inherent in applications of developing technology and those related to information practices needed in the development of technology. New efforts in "privacy technology" attempt to protect individual privacy. Threats to Homeland Security have made considerable funding available to investigate this topic in order to support bio-terrorism surveillance and protect individual privacy. Techniques have been reported for embedding encrypted digital watermarking and patient identifiers in medical images to protect privacy during use and transmission. Data mining investigators have begun encouraging their colleagues to take a research interest in issues related to protecting the privacy and security of personal information. The techniques of data mining have been used to address the issue of auditing access and use of data as well as for testing devices for intrusion detection and access control. Commercial products exist that automatically correlate and compare suspicious information gathered from

different points in computer systems, draw conclusions, and act on potential attacks and security violations [10].

6. CONCLUSIONS

A formal approach to managing the use and disclosure of personal health information is in the best interests of patients, individual researchers, organizations and society The risks to those who do not adhere to good security and privacy practices are considerable. Future laws and regulations are likely to increase penalties for inappropriate use or disclosure. While much attention has been given to research, organizations should implement the same general processes to support analyses done for the purpose of healthcare operations as for research.

AUTHOR INFORMATION

I am Dr. E.Kesavulu Reddy working as Assistant Professor and in Department of Computer Science, Sri Venkateswara University College of Commerce Management and Computer Science, Tirupati , Andhra Pradesh-India-517502

REFERENCES

[1] Behlen, F.M., Johnson, S.B. "Multicenter Patient Records Research: Security Policies and Tools," J AmMed Inform Assoc.pp 435-43. (1999).

[2] Berman, J.J "Confidentiality Issues for Medical Data Miners," Artif Intell Med. Pp :25-36.,2002.

{3}. Cios, K.J., Moore, G.W. "Uniqueness of Medical Data Mining," Artif Intell Med. 26(1-2), 1-24,2002.

[4]. Claerhout, B., De Moor, G.J., De Meyer, F. "Secure Communication and Management of Clinical and Genomic Data: The Use of Pseudonymisation as Privacy Enhancing Technique," Stud Health Technol Inform. 95:170-5, 2003.

[5] Moore, G.W., Brown, L.A., Miller, R.E. "Set Theory Definition and Algorithm for Medical De-identification," Johns Hopkins Autopsy Resource, 2000.

[6]. Murphy, S.N., Chueh, H.C.. "A Security Architecture for Query Tools Used to Access Large Biomedical Databases," in Proc AMIA Symp. 552-6, 2002.

[7]. Department of Health and Human Services. Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule, (NIH Publication Number 03–5388), 2004.

[8]. Sweeney, L "Guaranteeing Anonymity When Sharing Medical Data, The Datafly System," in Proc AMIA Symp Pp 51-55. . (1997).

[9]. Sweeney, L. "K-anonymity: A Model for Protecting Privacy," International Journal on Uncertainty, Fuzziness, and Knowledge-based Systems, 10(7), pp 557-570, 2002.

[10].Sweeney, L.. "Navigating Computer Science Research through Waves of Privacy Concerns: Discussions among Computer Scientists at Carnegie Mellon University", ACM Computers and Society, 34(1),pp.1-18,2013.