Guest Editorial Preface

Privacy in Health Informatics

Soon Ae Chun, Columbia University & CUNY-College of Staten Island, USA

Jaideep Vaidya, Rutgers, The State University of New Jersey, USA

INTRODUCTION

Health Information Technology (HIT or Health IT) describes the use of computer information systems to manage the patients' electronic health records. Specifically, Electronic Health Record (EHR) systems allow users in a healthcare organization, such as hospitals, clinic, or a doctor's office, to enter, store, process, access and manage patient healthcare data. Typical data in EHRs include demographics, medical history, medication and allergies, doctor's order entries and comments, immunization status, laboratory test results, radiology images, vital signs, personal stats like age and weight, and billing information. EHRs can support clinicians towards providing better healthcare by granting access to comprehensive patient data. EHRs can also help reduce medical prescription errors with various alerting functions, and can help patients and doctors to manage their treatment and billing records for insurance payments.

Increasingly, Health Information Exchanges (HIE) promote the exchange and sharing of patients' EHRs among healthcare providers, patients, insurance companies, government and other healthcare entities, across organizational boundaries. This sharing and exchanging of EHRs enables more collaborative patient-centered care, and can help public healthcare officials to analyze population level health issues and trends. Instead of implementing enterprise EHR systems, many organizations today utilize EHR systems that are cloud-based Web application services where patient records are stored and made accessible through Web interactions. Given the rapid advances in technology and its adoption, even patient records distributed in various places, such as doctors' offices, hospitals, can be easily linked, aggregated and shared via Health Information Exchanges (HIE) to provide more effective "personalized" treatment and payment. In addition, social media have been widely used to form specialized patient communities to discuss and share their medical and health experiences in near real-time fashion, to reveal trends and patterns of specific diseases, symptoms and drug effectiveness that can advance the healthcare community towards better preparedness and early warnings for health-related risks and emergencies.

These promises, however, may be jeopardized if the privacy concerns of individual patients are not properly addressed. The ultimate success of EHRs and other technology-driven systems built around the EHRs will be at stake without proper mechanisms and policies to preserve the privacy of health data of patients. The malicious or accidental disclosure of a patient's sensitive health records can pose a serious

threat to the privacy of a patient, which in turn can erode the trust in the HIT in general. The privacy infringements can damage the person's reputation, create embarrassment, and can cause denial of opportunities for employment and insurance coverage. The social consequences may include extreme stress, anxiety and even suicides.

As per the HIPAA Privacy Rule and its regulations (Office of Civil Rights, 2003), protected health information (PHI) (Table 1) is defined as any information that can link health and healthcare related data to a specific individual, hence called individually identifiable health information. The rules define high level policies that specify when and for what purpose PHI can be shared and when to notify the patients in case of disclosure or sharing.

Privacy enhancing technologies (PET) that protect PHI include anonymization, pseudonymity, and mix networks, among others. These generic approaches may enhance the EHR privacy control. However, in addition to these PET mechanisms, there are privacy enhancing mechanisms specific for EHR systems and the healthcare environment.

We look at four layers of privacy protection studies and list some of the research issues in each layer as shown in Table 2. The first layer focuses on mandated regulations and laws, such as the HIPAA Privacy Rule, or other government imposed regulations that each healthcare provider and its subsidiaries have to obey. The second type focuses on organizational level privacy policies that emphasize not only the compliance with the governmental laws, but also their organizational access control policies and operational level data guardianship for the patients' privacy. The third level mechanisms address the patients' own privacy preferences that are mostly discretionary but need to be observed to provide the patients with the comfort of controlling the privacy of their own records. Finally, there are data-level privacy studies that address the anonymization of PHI, and privacy preserving mining or fusing techniques.

This special issue is intended to highlight these multi-layer research challenges of preserving the privacy of patients, and computational methodologies and techniques to address them. The four articles we feature in the special issue address some of research issues in each research layer as summarized in Table 2.[1] Three of the articles had their preliminary versions presented at the 2012 ACM SIGHIT International Health Informatics Symposium (IHI, 2012).

PAPERS IN THE SPECIAL ISSUE

Designing a HIPAA compliant EHR system in a sharing environment is a challenging task, especially when the rules and regulations are stated in a natural language. In such a case, the

Table 1. Protected health information defined in HIPPA privacy rule

• Names • All geographical identifiers smaller than a state • Dates (other than year) directly related to an individual • Phone numbers • Fax numbers • Email addresses • Social Security numbers • Medical record numbers • Health insurance beneficiary numbers • Account numbers	• Certificate/license numbers • Vehicle identifiers and serial numbers, including license plate numbers • Device identifiers and serial numbers • Web Uniform Resource Locators (URLs) • Internet Protocol (IP) address numbers • Biometric identifiers, including finger, retinal and voice prints • Full face photographic images and any comparable images • Any other unique identifying number, characteristic, or code except the unique code assigned by the investigator to code the data

Table 2. Multi-layer privacy study areas and research issues in health informatics

Privacy Research Areas	Research Issues	Articles
Regulatory level privacy protection	• Privacy policy, privacy laws and regulations • Privacy policy extraction and modeling • Formal representations for privacy laws and regulations • Privacy policy analysis and implementation • How to make the free-text privacy laws and regulations computer consumable?	Ruoyu Wu, Gail-Joon Ahn and Hongxin Hu, *Towards HIPAA-compliant Healthcare Systems in Cloud Computing*
Organizational/System level privacy protection	• Security and Privacy Enforcement in EHR systems • Auditing health records to detect privacy violations and misuse • Privacy Rule compliance in healthcare systems • Accountability • Access control and consent management • What are the organizational privacy policies and how to implement them • How do the privacy challenges affect software design and process implementation?	Jason King, Ben Smith, and Laurie Williams *Audit Mechanisms in Electronic Health Record Systems: Protected Health Information May Remain Vulnerable to Undetected Misuse*
Personal level privacy protection	• Patient control on the disclosure of their medical data • Personal privacy-preserving health data usage control • Patient-centric monitoring of sensitive data • Patient-controlled security and privacy infrastructure • Personal privacy preference enforcement framework • Usability of privacy enforcement • Health Social network and privacy control mechanisms • What are the patient's discretionary privacy preferences and how can they be represented and enforced. • How to process the social stream data such as tweets in a privacy preserving manner • How can we measure the privacy risks so that the patients can be notified before privacy is infringed?	Thomas Trojer, Basel Katt, Ruth Breu, Thomas Schabetsberger, and Richard Mair *Managing Privacy and Effectiveness of Patient-administered Authorization Policies*
Data level privacy protection	• De-identification and anonymization • Anonymization of health data to counter de-anonymization attacks • How to anonymize the personally and semi-personally identifying data attributes from EHRs • How to link and combine data without leaking privacy • How to mine and analyze health records to reveal interesting patterns and trends in privacy-preserving way	Matt Matlock, Nakeisha Schimke, Liang Kong, Stephen Macke, and John Hale *Systematic Redaction for Neuroimage Data*

automated system that enforces the policies and rules should be able to parse and "understand" the semantics of the natural language text to generate machine enforceable policy rules. Wu, Ahn, and Hu address the first layer challenge on how to make EHR systems of different data providers comply with various HIPAA rules and organizational policies in the shared cloud environment. EHR applications and services in clouds face many security challenges associated with authentication, identity management, access control, policy integration and trust management. In addition to the security issues, EHR systems in the cloud environment face the important issue of providing a proper mechanism to ensure and manage compliance with HIPAA privacy rules and regulations to prevent patients' privacy disclosures, government fines, court representation costs, lost reputation, brand damage, government audits, and additional workforce training costs. The challenges include the fact that the process of compliance checking can be labor-intensive and can result in additional overhead in the healthcare provision; that the regulations are stated in complicated and vague language, thus requiring interpretation and domain knowledge; and that the compliance rules may change or the EHR systems may get upgrades, thus making up-to-date compliance management even more complex.

Compliance management in the cloud environment is even more significant since the lack of compliance with HIPAA privacy rules may cause damage of even larger magnitude due to the interaction of multiple medical providers' systems and services. The authors propose an automated compliance management approach which ensures that EHR systems are compliant with HIPAA regulations in cloud computing environments. Their approach involves a logic-based policy model to perform automated compliance analysis. First, text extraction methods are used to transform both HIPAA regulations and system policies specified in a natural language into a formal policy representation. Next the formal policy representation is transformed into a logic-based representation. Finally, the logic-based reasoning technique of Answer Set Programming (ASP) is leveraged to analyze privacy policy compliance. The evaluation study with the prototype system shows promising results for the feasibility and effectiveness of the proposed approach.

In the second article, the authors, King, Smith, and Williams, investigate the system layer privacy issues: whether EHR systems have proper audit mechanisms that can help record and detect inappropriate access to the Protected Health Information (PHI). The audit mechanisms provide the privacy of PHI in accordance with the HIPAA Privacy Rule and enhance the non-repudiation to mitigate insider attacks on the privacy rules by legitimate users of the EHRs. Since the focus is on insider attacks, the article looks at the authenticated user interaction types, such as storage confidentiality to prevent an access to the log entry, machine-based traces to identify a specific machine for source of log files, application-based non-repudiation to prevent malicious users from creating fake log entries on application software, and transmission confidentiality to ensure the accuracy and integrity of log files during transmission. Auditable events are assembled from the Certification Commission for Health Information Technology (CCHIT), the SysAdmin, Audit, Network, Security (SANS) Institute, the IEEE Standard for Information Technology: Hardcopy Device and System Security, and Chuvakin and Peterson's Web-based system audit list. These lists were filtered into 16 unique events that can be classified as one of these user-based non-repudiation interaction types.

Three EHR systems are tested, including two open source EHR systems, i.e., OpenEMR, Tolven eCHR (Electronic Clinician Health Record system), and one proprietary system, to compare and contrast the audit mechanisms. The authors measure how many of the general auditable events each EHR system provides and measure a percentage of satisfaction of EHR

systems for the 16 general auditable events. The results of the evaluation study show that user events such as viewing, creating, updating and deleting PHIs are better logged in the OpenEMR system than in the other two systems. Finally, the authors strongly advise the EHR system designers to follow general guidelines on building a proper auditing component for an EHR system that handles PHI for thwarting insider attacks by the legitimate authenticated users and for ensuring non-repudiation of user events.

The third article by Trojer et al. addresses the issue of balancing the ability to express patient-controlled privacy policies and shared EHR (SEHR) information system effectiveness. An ordinary patient can neither be considered a security expert, nor do they have the expertise to fully understand typical activities and workflows within the healthcare domain. The privacy of citizens' health-data and the overall effectiveness of a healthcare information system are both at risk if inadequate access control settings are set by a patient. The article develops an authoring and configuration setting system for PHI but also argues that the system should provide an automated analysis tool to analyze and detect the inconsistencies of the patient's privacy policy to warn about misconfigurations. The article then proposes a logic-based policy evaluation framework for automatically deriving analysis rules. The article presents a shared EHR use case in Austria that illustrates the need to find a balance between how to give control of privacy of health data to the patient and how to make the system usable. The privacy policy authoring tool also considers the global privacy settings as well as the individual level preferences related to privacy settings. This study also provides some practical lesson for the development of a usable EHR system that is shared by many stakeholders.

The fourth article by Matlock et al. examines how to protect personally identifiable information (PII) in neuroimage datasets, in accordance with the HIPAA Privacy Rule. There are several challenging issues in providing protection of PII in the data sharing of neuroimages, such as the lack of proper redaction tools to systematically expunge PHI/PII from neuroimage data sets, a difficulty in tracking patient identities in redacted datasets, and the absence of a sanitization workflow. Some structural neuroimages can be classified as identifiable data even without presence of identifying metadata. To protect the images from being personally identifiable, the authors compared the performance of defacing methods, such as MRI Defacer and Quickshear. The article describes the XNAT Redaction Toolkit — an integrated redaction pipeline which extends a popular neuroimage data management toolkit to remove PHI/PII from neuroimages, to establish a standardized and proven workflow for the deidentification of PHI from metadata and structural data. The toolkit relieves the researchers from administrative duties to prepare a plan to protect identifiers from improper use or disclosure. This proposed tool can foster collaboration among researchers from different organizations through secure image data sharing without privacy infringement.

ACKNOWLEDGMENT

This work was done in part while Chun was a visiting scholar at the Network Security Lab at Columbia University's Computer Science Department. We would like to acknowledge the anonymous reviewers for their generous time and effort, and the authors for contributing their articles for the special issue. We appreciate the Editor-in-Chief Professor Aryya Gangopadhyay for giving us this opportunity

Soon Ae Chun
Jaideep Vaidya
Guest Editors
IJCMAM

REFERENCES

Office of Civil Rights. (2003). *Summary of the HIPPA privacy rule*. Washington, DC: Department of Human and Health Services. Retrieved from http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacysummary.pdf

ENDNOTES

[1] We do not claim the research issues listed in each area are exhaustive. In addition, while each article of the special issue falls within a particular research area, it does not cover all of the research issues listed in the area.