A Research Roadmap for Healthcare IT Security inspired by the PCAST Health Information Technology Report

Matthew D. Green
Johns Hopkins University
mgreen@jhu.edu

Aviel D. Rubin
Johns Hopkins University
rubin@jhu.edu

1 Introduction

The President's Council of Advisors on Science and Technology (PCAST) recently proposed a set of recommendations for improving the state of Health Information Technology [4]. The report, delivered in December, 2010, comes at a time when the government is spending billions of dollars to modernize the health information technology infrastructure. There is an aggressive push to adopt electronic health records (EHRs) for doctors, hospitals and medical laboratories. The PCAST report identifies many of the challenges posed by widespread EHR adoption. There is a chapter devoted to technology issues as well as a chapter specifically addressing security and privacy. The PCAST report contains several recommendations that we believe provide a useful blueprint for security researchers.

While some of the PCAST security and privacy recommendations can be addressed immediately, others are not as straightforward and will require research. The purpose of our paper is to detail and in some cases expand upon the security and privacy recommendations in the PCAST report and to identify research problems inspired by these recommendations. We hope that our work will provide the security and privacy research community with direction and open problems needing attention, and that it will be a useful resource to US funding agencies when setting their agendas for healthcare information technology.

2 Security Issues identified in PCAST report

Chapter V of the PCAST report addresses Security and Privacy Considerations. The report is critical of HIPAA and the HITECH act as not providing patients, in practice, with fully informed choices for the privacy of their information. Individuals today have little control over the way their health information is collected, stored and shared.

The authors of the PCAST report recommend an exchange language based on tagged data elements as the platform of choice for enabling fine grained individual privacy preferences. We believe that this is the right model. Tagged data elements provide a flexible mechanism for organizing information and for performing various operations on the data. However, as the report observes, the model is only as good as its implementation. Much of the hard work of designing a secure EMR system with fine grained access control comes after picking the data representation framework.

Besides identifying tagged data elements as critical to a security framework, the PCAST report also identifies several security and privacy issues.

Robust User Identity There is a need for identifying and authenticating real world individuals. In many countries there is a national identity number associated with individual citizens. Privacy concerns, perhaps misguided, have resulted in a rejection of such a number in the United States. As such, matching records to physical identities is sometimes fraught with error.

Audit and logging It is important for all events related to EMRs be logged, and that audit capability exists.

Patient Access The report identifies the need for patients to have access to and control of their records. A question remains as to whether patients should also be able to correct their records and what sort of process is needed to resolve disputes that arise about information contained in an individual's EMR.

Cryptographic Keys The report describes an architecture where encryption keys are maintained on remote servers and never stored in the same place as encrypted EMRs. This is a reasonable security policy. Technologies need to be developed to support flexible key management policies such as this one.

De-identification for research The report describes the importance of providing de-identified EMRs for use in medical research. There has been a substantial amount of work in the academic community on how to de-identify data, as well as research on the efficacy of such techniques. The PCAST report also suggests that a digital signature can be used to certify that data has been de-identified securely. We believe that this might be a more challenging task than implied by the report, and that the challenges related to public key infrastructure are non-trivial.

Comparison to security of paper records The PCAST report suggests that EMRs should provide a much stronger level of security than paper records. As discussed in Section 3 of this paper, we believe that security metrics are needed for such an analysis. Such metrics are complicated, and little progress has been made in providing meaningful, generic security metrics, despite many efforts in the research community.

This work was supported in part by the NSF under grant CNS1010928.

The PCAST report identified these important security issues. We agree that to realize a viable and secure health information system these issues will need to be addressed. We further believe that these issues introduce several research problems that need solving. We describe these research problems in the next section.

3 Research Problems Inspired by the PCAST report

Some of the security issues identified in the PCAST report can be dealt with using existing technologies. However, most of the problems listed in the previous section cannot be addressed with off the shelf solutions. In this section, we identify medium and long-term research problems the must be solved to achieve the security level for health information technology described by the authors of the PCAST report.

For example, the report calls for electronic health records with a stronger level of security than is achievable with paper records. To assess security, metrics must be developed. The problem of developing generic security metrics has stumped researchers for years. However, in this limited domain, it is possible that more can be achieved.

In the remainder of this section, we describe specific research problems, inspired by the requirements for security and privacy called out in the PCAST report.

3.1 Meta-Tagging

Meta-tagging is an effective decision-support framework to enable meta-tagging of EHR components. Obviously, meta-tagging alone is an abstract, high-level construct, and does not indicate anything about data semantics. The identification by PCAST of meta-tagging as the method for data representation motivates several interesting and important research problems, described below.

- Develop programmatic means for tagging data based on a particular security policy.
- Research how to utilize automated tagging and access control policies within a defined policy engine. For example, study how to best feed tagged data into an automated policy engine to determine whether a requested access is to be allowed.
- Develop efficient means for parsing and processing tagged data elements.

3.2 Security Metrics

The PCAST report repeatedly suggests that the security of EMRs be compared to paper-based systems. Such comparison requires well-developed, understood and accepted metrics. Security metrics are difficult to define and to implement because of the variety and complexity of threat models, as well as the difficulty of measuring and accounting for the potential flaws in software. Below are specific research problems related to security metrics inspired by the PCAST report.

- Develop threat models for medical records that apply both to paper and to EMR systems.
- Develop techniques for quantifying the level of risk associated with using a software-based system. To date, metrics such as *the number of lines of code* have provided the best measure of the number of flaws. Perhaps in a limited domain, such as working with EMRs, we can do better.

3.3 Cryptographic Access Control for EHRs

The PCAST report describes an approach to securing EHR data in a data sharing environment. The authors recognize that traditional approaches to securing data tend to break down in distributed storage networks, where trust cannot be assumed of the individual storage nodes (e.g., Health Information Exchanges, Cloud services).

In the PCAST report each data record carries with it a detailed policy (“metadata describing its use and access”) that determines which parties may access the data. PCAST proposes that this policy should be enforced through the use of strong encryption and digital signatures, and moreover, that encryption keys should never be stored on the same computer as this data itself. However, beyond this high-level vision, very few concrete engineering details are provided.

Thus, several problems remain open to researchers. To realize the PCAST vision we must identify the appropriate tools and technologies for protecting health care data; deal with issues such as key management and policy enforcement; and propose techniques for developing policies with the input of users. Many of these questions have already been addressed in the literature (e.g. [2]), but developing *practical* systems (that tolerate real design constraints, such as untethered access) is a current research problem.

To extend previous work, we provide a short example here of a possible avenue of exploration. First, we propose the use of *functional encryption* [3]. In a functional encryption scheme, data is encrypted along together with some function input x, while keys embed an input y. Data can be accessed iff $f(x, y) = 1$ for some specific function $f(\cdot)$. This model encompasses schemes such as ciphertext-policy Attribute-Based Encryption [1], in which the encryption input is a detailed *policy* described as a boolean formula over user attributes, and keys embed a list of attributes; the function $f()$ in this case is one that outputs 1 iff the attribute list satisfies the policy. This approach allows for untethered access to record data.

There are many other research problems in the area of applying cryptography to access control to enable role based access control for EMRs, and we believe that this will remain an active area of research for some time.

The following specific research problems relate to cryptography in healthcare IT systems.

- Develop access control specification and enforcement techniques for EMRs based on ABE.
- Develop and experiment with key management solutions for practical implementations of the cryptographic systems that are developed.

- Provide cryptographic mechanisms to properly anonymize records as required by secondary use considerations.

3.4 Managing User Identity

The PCAST report describes the need for identification of principals in any healthcare IT system. Principals include doctors, patients, nurses, and any other entity that requires access to healthcare records. The report describes authentication mechanisms as falling into three categories, namely physical credentials such as smart cards, biometrics, and secrets such as passwords. The report also identifies requiring two of these methods as two-factor authentication.

Authenticating principals in a system has been widely studied, and there are many available solutions. However, managing user identity in the US, where the principals can potentially number in the millions, is problematic. Unlike many other countries, the United States does not associate each citizen with a unique ID number. While most Americans have a social security number, this, by law, cannot be used to index electronic medical records. Thus, a serious challenge exists when records from multiple sources are reconciled. The problem is magnified when some of the records contain slightly different spelling of a principal's name, or when another minor error exists in the record. For example, sometimes the day and month of the birthday are transposed.

One of the problems facing healthcare IT managers is that errors such as those just described can lead to multiple records for the same person being associated with two different identities. Another problem is when the records for two or more different people are mistakenly associated with one person. When two organizations store their records in incompatible formats, and those records need to be unified, the problem of merging them is exacerbated by the lack of a reliable unique identifier.

An obvious step that could ameliorate the problem is to assign a globally unique healthcare identification number to every patient, and every healthcare provider in the country. This concept is fraught with political controversy, and may or may not be feasible. Even if the political hurdles can be overcome, there are technological research problems associated with assigning new and unique identifiers to patients whose records are in multiple legacy systems. Such systems may have no database field available for this purpose, and to the extent that they are extensible, it may be difficult to assign the new IDs accurately.

Without the unique global identifier, managing user identity in a large, complex healthcare system consisting of multiple providers and many incompatible vendors remains an open research problem. The following specific research problems, motivated by the PCAST report, relate to managing user identities in legacy and new healthcare IT systems.

- Develop new biometric techniques and improvements on existing ones for identifying individuals. New research is needed to determine how to best use such biometrics for principal authentication in an end to end system.
- Develop techniques for reconciling data from disparate sources.
- Develop new authentication techniques that are less vulnerable to third party attacks such as the recent attack on RSA's one time authentication tokens.

3.5 Logging

The PCAST report states, "...an audit mechanism records the actions taken by principals in the system along with the information used to authorize those actions. A secure audit mechanism must provide strong protection so that audit records cannot be tampered with or deleted. For example, an audit system must record any access, modification, or deletion of a patients health records and any changes to the associated authorization policies. In case of an error in a patients health record, the audit mechanism would reveal the principal who introduced the mistake and the authorization information for that access. Strong audit mechanisms can be implemented using cryptographic and other techniques. Patients should have the right to review audit records pertaining to their data."

An audit mechanisms involves logging events as they occur. Events to be logged include any modification, addition or deletion in an EMR system. Events such as addition of a principal, changing a principal's access permissions and any other administrative function needs to be logged as well. It is important to log administrative functions in such a way that the administrators cannot modify the logs.

There are numerous challenges to proper logging for the purposes of audit. Logs are only as useful as our ability to extract the information needed when an audit or an incident occurs. While logging every event is technically easy, such an approach yields tremendous volumes of data that may result in an intractable problem when it is necessary to recreate events when an incident occurs. What is needed rather is intelligent logging. The following specific research problems relate to logging and audit.

- Researching new techniques for creating useful logs that can be used to quickly recreate events and assign responsibility when incidents occur, while limiting the volume of information that needs to be stored.
- Developing new approaches for log storage and retrieval that are tailored to the medical record application with access to logs that is user-friendly and intuitive.

3.6 User Interaction with EHR systems

The PCAST report references human factors and user friendly interfaces as important elements for a secure EMR system. These are critical both for administrator and provider interfaces as well as for patient interaction. The following specific research problems relate to patient interaction.

- Develop user friendly mechanisms for dealing with the complexity of user-selected privacy preferences.

- Research how much data to make available to patients and in what format, and study whether it makes sense to provide different amounts of access to different patience based on certain criteria.
- Study issues such as how to enable patients to delegate their access rights.

3.7 Dispute resolution

Among the potential advantages of EMRs for patients are the ability to monitor one's medical records and the potential to correct them. Many health care administrators are wary of placing this capability in the hands of patients both because they do not trust patients to always be honest and correct, and because it is important for the administrators to maintain a history of the information in a principal's medical record at any given time. As EMRs become a reality, we must study the best mechanisms for allowing people to dispute their medical records. The following specific research problems relate to dispute resolution in healthcare IT systems.

- Develop ways for patients to securely and privately monitor their health records.
- Develop mechanisms for patients to dispute details of the EMRs, while preserving the original record.
- Develop conflict resolution techniques when a patient's claim about their EMRs differ from those of a health care provider such as a doctor or a laboratory.

4 Other Research Problems

The PCAST report provides a comprehensive look at EMRs and addresses many important security issues. However, there are some research problems that we feel are important that were not mentioned in the PCAST report. We briefly mention these below.

Implantable devices Research is needed to study communication protocols and safety and security features of implantable devices such as pacemakers and defibrillators.

Home monitoring and technologies for in home care Monitoring patients at home introduces new security challenges such as protecting communication links, and protecting against adversaries in the home. New threat models should be developed.

Formal methods research Many aspects of protecting EMRs lend themselves to formal methods, and we believe that this area poses now opportunities for research advances in formal methods. For example, formal methods could be used to address the problem of parsing meta-tags to enforce complex policies for EMRs.

Legal issues Many of the problems that need to be solved are not purely technological. Dispute resolution is an example of a problem that will require advances in the law as well as in technology.

Social science studies Introducing widespread EMRs and involving the public will require user studies, as well as research into the acceptance and likely adoption of patient-centric solutions.

5 Technology Transfer Challenges

While the PCAST report provides many reasons for optimism, we'd believe that there are sobering challenges when attempting to transfer the type of research described in our paper to practice. As researchers, we have had great difficulty obtaining testbeds and research data. Medical data is very sensitive, and organizations have rightfully placed big restrictions on their use. Consequently, obtaining live, realistic data for testing purposes is onerous. There is a lack of experimental culture in the medical records field. As a result, solutions tend to come from industry and are often closed and non-interoperable.

We believe there is no clear path towards acceptance of research results by the medical community, and we were able to find few examples of successful research projects finding their way into the healthcare IT system.

6 Conclusions

It was our aim to view the PCAST report as providing a framework. Within that framework, we identified issues that are mentioned or implied in the report, and we have provided a list of research problems that we believe need to be addressed for healthcare IT to move forward. We hope that pinpointing these problems will provide a research roadmap to privacy and security researchers interested in working in this area.

References

[1] John Bethencourt, Amit Sahai, and Brent Waters. Ciphertext-policy attribute-based encryption. In *IEEE Symposium on Security and Privacy*, pages 321–334, 2007.

[2] Matt Blaze, Joan Feigenbaum, and Jack Lacy. Decentralized trust management. In *Proceedings 1996 IEEE Symposium on Security and Privacy*, May 1996.

[3] Dan Boneh, Amit Sahai, and Brent Waters. Functional encryption: Definitions and challenges. In *TCC*, pages 253–273, 2011.

[4] The President's Council of Advisors on Science and Technology. Report to the President — Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward. Available at `http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf`, December 2010.