Audit Mechanisms for Privacy Protection in Healthcare Environments

Jeremiah Blocki
Carnegie Mellon University

Nicolas Christin
Carnegie Mellon University

Anupam Datta
Carnegie Mellon University

Arunesh Sinha
Carnegie Mellon University

***Abstract*—We take the position that audit mechanisms are essential for privacy protection in healthcare environments. Although audits are used in practice and commercial tools that provide assistance for audits are emerging, we currently lack rigorous models and definitions of properties that can guide the design of appropriate audit mechanisms. We report on our recent result that presents a principled learning-theoretic approach to audits with the goal of stimulating discussion and additional research on this problem.**

I. Introduction

A challenging problem in healthcare environments is to ensure that privacy expectations of patients are respected in the collection, disclosure and use of personal health information. *Access control mechanisms* used to restrict access to medical records have, by design, to be *permissive* since wrongly denying or delaying access to a patient's medical records can hinder effective delivery of healthcare. However, a permissive access control regime opens up the possibility of records being inappropriately accessed. Indeed, recent studies reveal that many policy violations occur in the real world as employees inappropriately access records of celebrities and family members motivated by general curiosity, financial gain and other considerations [1]. To compensate for the permissive nature of their access control mechanisms, medical record systems must, in addition, support *audit mechanisms* that can provide *a posteriori* enforcement of the desired privacy and security properties. This is achieved by recording accesses made by employees in an audit log that is then examined by human auditors to determine if accesses and transmissions were appropriate and to hold individuals accountable for violating policy[1].

The importance of audits has been recognized in the computer security literature. For example, Lampson [2] takes the position that audit logs that record relevant evidence during system execution can be used to detect violations of policy, establish accountability and punish the violators. More recently, Weitzner et al. [3] also recognize the importance of audit and accountability, and the inadequacy of preventive access control mechanisms as the sole basis for privacy protection in today's open information environment. However, while the principles of access control have been extensively studied, there is comparatively little work on the principles of audit.

Our work is aimed at filling this gap. A specific goal is to design audit mechanisms that are guided by pragmatic economic considerations (e.g., budgetary constraints that prevent auditors from examining entire audit logs). Although we seek to develop general mechanisms that are applicable in many different situations, our immediate focus is on audit mechanisms that can provide increased assurance that personal health information is disclosed and used appropriately in healthcare environments.

We begin with the observation that many privacy regulations, including the HIPAA Privacy Rule, include policies about disclosure and use of personal information that cannot be mechanically enforced in their entirety. For example, HIPAA allows transmission of protected health information about an individual from a hospital to a law enforcement agency if the hospital believes that the death of the individual was suspicious. Such beliefs cannot, in general, be checked mechanically either at the time of transmission or in an *a posteriori* audit; the checking process requires human auditors to inspect evidence recorded on audit logs[2].

Specifically, our research aims to answer the following two questions: (1) What is an appropriate mathematical model for studying audit mechanisms and their properties? (2) What kind of experiments should one perform to validate such models and mechanisms in a healthcare environment? In the remainder of this position paper, we detail the progress we have made toward answering the first question, before turning to a discussion of potential research avenues to answer the second question.

II. Regret Minimizing Audits

We present the first principled learning-theoretic foundation for audits of this form [5]. Our contribution is a **repeated game model** that captures the interaction between the defender (e.g., hospital auditors) and the adversary (e.g., hospital employees). The model takes pragmatic considerations into account, in particular, the periodic nature of audits, a budget that limits the number of actions that the defender can inspect, and a loss function that captures the economic impact of detected and missed violations on the

[1] Commercial tools, such as FairWarning, are beginning to emerge to assist in this process.

[2] A related paper [4] presents an algorithm that mechanically enforces objective parts of privacy policies based on evidence recorded in audit logs and outputs subjective predicates (such as beliefs) that have to be checked by human auditors. It also reports on an implementation and application of the algorithm to the entire HIPAA Privacy Rule.

organization. We formulate a desirable property of the audit mechanism in this model by adopting the concept of *regret* in learning theory [6], which naturally accounts for worst-case adversaries. We propose a novel **audit mechanism** that provably minimizes regret for the defender. The mechanism learns from experience and provides operational guidance to the human auditor about which and how many of the accesses to inspect.

Adversary model: In each audit cycle (round of repeated game), the adversary performs a set of actions (e.g., accesses patient records) of which a subset violates policy. Actions are classified into types, e.g., accessing celebrity records could be a different type of action from accessing non-celebrity records. The adversary capabilities are defined by parameters that impose upper bounds on the number of actions of each type that she can perform in any round.

Defender model: In each round, the defender inspects a subset of actions of each type performed by the adversary. The defender has to take two competing factors into account. First, inspections incur cost. The defender has an audit budget that imposes upper bounds on how many actions of each type she can inspect. Second, the defender suffers a loss in reputation for detected violations. The loss is higher for violations that are detected externally than those detected internally, thus incentivizing the defender to inspect more actions. In addition, the loss incurred from a detected violation depends on the type of violation. For example, inappropriate access of celebrities' patient records might cause higher loss to a hospital than inappropriate access of other patients' records. Also, since public memory is short, violations detected in recent rounds cause greater loss than those detected in rounds farther in the past.

Regret property: We require the audit mechanism satisfy the property of low regret studied in learning theory. The idea is to compare the loss incurred when the real defender plays according to the strategy prescribed by the audit mechanism to the loss incurred by a hypothetical defender with knowledge of the number of violations of each type in each round. The hypothetical defender is allowed to pick a fixed strategy (also called an *expert* in the learning literature) to play in each round. For example, one class of experts might prescribe how many actions of each type to inspect. The *regret* of the real defender in hindsight is the difference between the loss of the hypothetical defender and the actual loss of the real defender averaged over all rounds of game play. We require that the regret of the audit mechanism quickly converges to a small value.

Audit mechanism: We develop a new efficient audit mechanism that provably minimizes regret for the defender. In each round, the algorithm prescribes which expert's advice to follow, i.e., how many actions of each type the defender should inspect. It does so by maintaining weights for each possible defender action and picking an action with probability proportional to the weight of that action. The algorithm works by increasing the weights of actions that yielded better payoff than the expected payoff of the current distribution and decreasing the weight otherwise.

III. Research Directions

Enhanced adversary models: The general question of what is an appropriate model for audit merits further research. While our current results hold even if an adversary controls the actions of all the employees in a hospital, it is reasonable to believe that not all employees behave adversarially. We plan to consider an alternative model in which some employees are adversarial, some are selfish and others are well-behaved. Such a model could enable us to develop audit mechanisms that incorporate incentives (e.g., punishment for violations) and possibly prevent violations.

Identifying experts: Our audit mechanism is parametric in the class of experts. While we discussed one class of experts that prescribe in each round the number of actions of each type to inspect, an interesting direction is to identify experts that are suitable for the healthcare domain. In particular, can the experts be learned from audit log data and knowledge of privacy policies?

Experimental evaluation: We also plan to implement and evaluate our audit mechanisms. The design of appropriate experiments to validate these models and mechanisms is itself a challenging problem. There are at least two avenues that one can pursue. First, one can try to obtain existing data (e.g., audits logs, violation records) and, in hindsight, see if the deployment of the mechanisms we prescribe would have done better than the mechanism deployed. The main challenge here is that such data is usually hard to acquire due to privacy concerns. Second, one can try to construct experiments akin to behavioral economics experiments, and test the impact of our proposed algorithms on actual human behavior.

References

[1] G. Hulme, "Steady Bleed: State of HealthCare Data Breaches," September 2010, InformationWeek.

[2] B. W. Lampson, "Computer security in the real world," *IEEE Computer*, vol. 37, no. 6, pp. 37–46, 2004.

[3] D. J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. A. Hendler, and G. J. Sussman, "Information accountability," *Commun. ACM*, vol. 51, no. 6, pp. 82–87, 2008.

[4] D. Garg, L. Jia, and A. Datta, "Policy auditing over incomplete logs: Theory, implementation and applications," in *Proc. ACM CCS*, 2011, To appear.

[5] J. Blocki, N. Christin, A. Datta, and A. Sinha, "Regret minimizing audits: A learning-theoretic basis for privacy protection," in *Proc. IEEE CSF*, 2011.

[6] A. Blum and Y. Mansour, "Learning, regret minimization, and equilibria," *Algorithmic Game Theory*, pp. 79–102, 2007.