Poster: The Semantics of Purpose Requirements in Privacy Policies

Michael Carl Tschantz
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA, USA
mtschant@cs.cmu.edu

Anupam Datta
CyLab
Carnegie Mellon University
Pittsburgh, PA, USA
danupam@cmu.edu

Jeannette M. Wing
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA, USA
wing@cs.cmu.edu

I. Introduction

Purpose is a key concept for privacy policies. Examples include the privacy policy of Yahoo! Email, which states that "Yahoo!'s practice is *not* to use the content of messages stored in your Yahoo! Mail account *for* marketing purposes" (emphasis added) [1], which prohibits a purpose. Some policies even limit the use of certain information to an explicit list of purposes. For example, the HIPAA Privacy Rule [2] requires that covered entities (e.g., health care providers and business partners) only use or disclose protected health information about a patient with that patient's written authorization or for fixed list of purposes including treatment, payment, health care operations, and research.

These examples show that verifying that an organization obeys a privacy policy requires a semantics of *purpose requirements*. In particular, enforcement requires the ability to determine that the organization under scrutiny obeys at least two classes of purpose requirements. As shown in the example rule from Yahoo!, the first requirement is that the organization does *not* use certain sensitive information *for* a given purpose. The second, as the example rule from HIPAA shows, is that the organization uses certain sensitive information *only for* a given list of purposes. We call the first class of requirements *prohibitive* (not-for) and the second class *restrictive* (only-for). Each class requires determining whether the organization's behavior is *for* a purpose or not, but they differ in whether this indicates a violation or compliance, respectively.

Manual enforcement of these privacy policies is labor intensive and error prone. Thus, to reduce costs and make their operations more trustworthy, organizations would like to automate the enforcement of the privacy policies governing their operations; tool support for this activity is beginning to emerge in the market. For example, Fair Warning offers automated services for the detection of privacy breaches in a hospital setting [3]. Meanwhile, previous research has purposed formal methods to enforce purpose requirements (e.g., [4], [5], [6], [7], [8]).

However, each of these endeavors start by assuming that actions or sequences of actions are labeled with the purposes they are *for*. They avoid analyzing the meaning of *purpose* and provide no method of performing this labeling other than through intuition alone. The absence of a formal semantics to guide this determination has hampered the development of methods for ensuring policy compliance. Such a definition would provide insights into how to develop tools that identify suspicious accesses in need of detailed auditing and algorithms for determining which purposes an action could possibly be for. Such a definition would also show which enforcement approaches are most accurate. More fundamentally, such a definition could frame the scientific basis of a societal and legal understanding of purpose and of privacy policies that use the notion of purpose. Such a foundation can, for example, guide implementers as they codify in software an organization's interpretation of internal and government-imposed privacy policies.

II. Solution Approach

The goal of this work is to study the meaning of *purpose* in the context of enforcing privacy policies and propose formal definitions suitable for automating the enforcement of purpose requirements. Since post-hoc auditing provides the perspective often required to determine the purpose of an action, we focus on automated auditing. If an auditor is concerned with a rational auditee (the person or organization being audited), then we may assume the auditee uses a plan to determine what actions it will perform in its attempt to achieve its purposes. We (as have philosophers [9]) conclude that if an auditee selects to perform an action a while planning to achieve the purpose p, then the auditee's action a is *for the purpose* p. If there exists no plan for achieving the purpose p that calls for an action a to be performed, then the auditor may conclude that the action a was not for the purpose p.

For example, consider a physician accessing a medical record. Under the HIPAA Privacy Rule, the physician may access the record only for certain purposes such as treatment. Thus, for an auditor to determine whether the physician has obeyed the Privacy Rule requires the auditor to determine the purposes for which the physician accessed the record. The auditor's ability to determine the purposes behind actions is limited since the auditor can only observe the behavior of the physician. As a physician may perform the exact

same actions for different purposes, the auditor can never be sure of the purposes behind an action. However, if the auditor determines that the record access could not have possibly been for any of the purposes allowed under the Privacy Rule, then the auditor knows that the physician violated the policy. In particular, the auditor may check the intersection between the set of plans for achieving the allowed purposes and the set of plans that could have given rise to the auditee's behavior. If this intersection is empty, the auditor may conclude that the auditee violated the policy.

The above approach requires formalizing planning. We use the Markov Decision Processes (MDP) to model situations that include stateful actions, probabilistic outcomes, and purposes that can be satisfied to varying degrees. We compare an MDP modeling the environment of auditee and the allowed purposes to the actions of the auditee. If the auditee's actions do not correspond to the actions called for by any of the optimal plans of (solutions to) the MDP, then we conclude that the auditee violated the policy.

III. Overview of Contributions

In a technical report elaborating this work, we show the following results [10].

First, we make our auditing process formal and discuss the ramifications of the auditor only observing the behaviors of the auditee and not the underlying planning process of the auditee that resulted in these behaviors. We show that in some circumstances, the auditor can still acquire enough information to determine that the auditee violated the privacy policy. To do so, the auditor must first use our MDP model to construct all the possible behaviors that the privacy policy allows and then compare it with all the behaviors of the auditee that could have resulted in the observed auditing log. We present an algorithm for auditing based on our formal definitions, illustrating the relevance of our work.

Second, the semantics we introduce is sufficient to put the previous work on enforcing privacy policies on firm semantic ground. We do so and discuss the strengths and weaknesses of each such approach. In particular, we find that each approach may be viewed as a method of enforcing the policy given the set of all possible allowed behaviors, an intermediate result of our analysis. We compare the previous auditing approaches, which differ in their trade-offs between auditing complexity and accuracy of representing this set of behaviors.

Third, most auditees are actually interested in multiple purposes and select plans that simultaneously satisfy as many of the desired purposes as possible. Handling the interactions between purposes complicates our semantics. In particular, actions selected by a single plan may be for different purposes. We present examples showing when our semantics can extend to handle multiple purposes and when difficulties arise in determining which purposes an action is for when an auditee is attempting to satisfy various purposes at once. Currently, the state-of-the-art in the understanding of human planning limits our abilities to improve upon our semantics. However, as this understanding improves, one may replace our MDP-like formalism with more detailed ones while retaining our general framework of defining purpose in terms of planning.

Although motivated by our goal to formalize the notions of *use* and *purpose* prevalently found in privacy policies, our work is more generally applicable to a broad range of policies, such as fiscal policies governing travel reimbursement. In the future, we hope to extend our formalism to handling more complex interactions among multiple purposes and conduct studies showing the accuracy and usefulness of our formalism.

References

[1] Yahoo!, "Privacy policy: Yahoo Mail," 2010. [Online]. Available: http://info.yahoo.com/privacy/us/yahoo/mail/details.html

[2] Office for Civil Rights, U.S. Department of Health and Human Services, "Summary of the HIPAA privacy rule," OCR Privacy Brief, 2003.

[3] FairWarning, "FairWarning: Privacy breach detection for healthcare," accessed Feb. 7, 2011. [Online]. Available: http://fairwarningaudit.com/

[4] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Hippocratic databases," in *VLDB '02: Proceedings of the 28th International Conference on Very Large Data Bases*. VLDB Endowment, 2002, pp. 143–154.

[5] J.-W. Byun, E. Bertino, and N. Li, "Purpose based access control of complex data for privacy protection," in *SACMAT '05: Proceedings of the Tenth ACM Symposium on Access Control Models and Technologies*. New York, NY, USA: ACM, 2005, pp. 102–110.

[6] K. Hayati and M. Abadi, "Language-based enforcement of privacy policies," in *PET 2004: Workshop on Privacy Enhancing Technologies*. Springer-Verlag, 2005, pp. 302–313.

[7] S. S. Al-Fedaghi, "Beyond purpose-based privacy access control," in *ADC '07: Proceedings of the Eighteenth Conference on Australasian Database*. Darlinghurst, Australia: Australian Computer Society, Inc., 2007, pp. 23–32.

[8] M. Jafari, R. Safavi-Naini, and N. P. Sheppard, "Enforcing purpose of use via workflows," in *WPES '09: Proceedings of the 8th ACM Workshop on Privacy in the Electronic Society*. New York, NY, USA: ACM, 2009, pp. 113–116. [Online]. Available: http://doi.acm.org/10.1145/1655188.1655206

[9] R. Taylor, *Action and Purpose*. Prentice-Hall, 1966.

[10] M. C. Tschantz, A. Datta, and J. M. Wing, "On the semantics of purpose requirements in privacy policies," School of Computer Science, Carnegie Mellon University, Tech. Rep. CMU-CS-11-102, Feb. 2011, Also available at http://arxiv.org/abs/1102.4326.