Experiences in the Logical Specification of the HIPAA and GLBA Privacy Laws

Henry DeYoung
hdeyoung@cs.carnegie.edu

Deepak Garg
dg@cs.cmu.edu

Limin Jia
liminjia@cmu.edu

Dilsun Kaynar
dilsun@cs.cmu.edu

Anupam Datta
danupam@cmu.edu

Carnegie Mellon University, Pittsburgh, PA 15213 USA

ABSTRACT

Despite the wide array of frameworks proposed for the formal specification and analysis of privacy laws, there has been comparatively little work on expressing large fragments of actual privacy laws in these frameworks. We attempt to bridge this gap by giving complete logical formalizations of the transmission-related portions of the Health Insurance Portability and Accountability Act (HIPAA) and the Gramm-Leach-Bliley Act (GLBA). To this end, we develop the PrivacyLFP logic, whose features include support for disclosure purposes, real-time constructs, and self-reference via fixed points. To illustrate these features and demonstrate PrivacyLFP's utility, we present formalizations of a collection of clauses from these laws. Due to their size, our full formalizations of HIPAA and GLBA appear in a companion technical report. We discuss ambiguities in the laws that our formalizations revealed and sketch preliminary ideas for computer-assisted enforcement of such privacy policies.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*; F.4.1 [**Mathematical Logic and Formal Languages**]: Mathematical Logic

General Terms

Security, Legal Aspects

Keywords

Fixed point logic, privacy policy specification, HIPAA, GLBA

1. INTRODUCTION

Privacy is an important concern for organizations that collect and use personal information, such as hospitals, banks, customer support centers, and academic institutions. In fact, designing organizational processes to manage personal data and ensure compliance with privacy regulations such as the Health Insurance Portability and Accountability Act (HIPAA) [24] and the Gramm-Leach-Bliley Act (GLBA) [23] has become one of the greatest challenges facing organizations today (see, for example, a recent survey from Deloitte & Touche and the Ponemon Institute [10]).

Even after a cursory glance through these laws, it is apparent that the legal language is much too dense and intricate for the laws to serve as a day-to-day guide to managers of the regulated organizations. Managers (and the general public) are instead interested in answers to concrete, practical questions, such as "Is the organizational privacy policy of Hospital X consistent with HIPAA?" and "Does GLBA permit Bank Y to disclose Bob's account information to Charlie?"

Recently, researchers have begun to attack the problem of formally expressing privacy laws in various logics [3–5, 12, 14, 16] and languages [2, 9, 15, 17, 19, 21]. The hope is that these logics and languages will permit the construction of interactive tools that can directly answer the kinds of questions that arise in day-to-day business operations.

Despite the wide array of privacy languages and logics, to the best of our knowledge, there has been comparatively little work on expressing *large* fragments of actual privacy laws in these frameworks; instead, the encodings have been limited to small proof-of-concept examples. But this is a significant deficiency if the program of obtaining practical benefits from formal specification of privacy laws is to succeed: we must be confident that the techniques invented for the small examples scale to full privacy laws. The contributions of our work are intended to help bridge this gap.

First, we propose a richly expressive logic and signature for the specification of privacy regulations, which we call PrivacyLFP. A comprehensive examination of HIPAA and GLBA guided the choice of PrivacyLFP's features, including support for self-reference, purposes of uses and disclosures, and real-time provisions and obligations. In Section 2, we motivate these choices using concrete example clauses from the two laws.

From a technical standpoint, PrivacyLFP is least fixed point logic (LFP) [7, 22] with a particular trace-based semantic model, first-order signature, and syntactic sugar for temporal modalities (see Section 3). In this way, we synthesize ideas from LFP and privacy logics, especially LPU [3, 4] from which we inherit an emphasis on traces of actions, such as sending a message, and most of our data model.

Second, we demonstrate the utility of PrivacyLFP by giving what we believe to be the first complete logical formalizations of the transmission-related portions of HIPAA and

WPES'10, October 4, 2010, Chicago, Illinois, USA.

GLBA. Specifically, we have encoded all requirements from §§164.502, 164.506, 164.510, 164.512, 164.514, and 164.524 of HIPAA, and §§6802 and 6803 of GLBA relevant to the transmission of information. We do not formalize the remaining sections because they impose abstract, non-operational demands, such as requiring organizations to develop standards "to ensure the security and confidentiality of customer records" (GLBA §6801).

Unfortunately, we cannot provide our full formalizations in this paper due to their length (about 120 pages when the legal text and explanations are included). Instead, in Section 4, we present encodings of the clauses used to motivate PrivacyLFP's features. We refer the interested reader to the companion technical report [11] for the full formalizations.

Third, we discuss ambiguities in and interesting observations about HIPAA and GLBA that our formalization efforts revealed (see Section 5). These include the range of limits on redisclosures, robustness of purposes, and balancing privacy and utility.

Finally, since our ultimate goal is to enforce privacy regulations using the PrivacyLFP logic directly, we draw on our formalizations of HIPAA and GLBA to sketch ideas for their enforcement using a combination of access control techniques, design-time analysis of organizational processes, and post-hoc audit with human support (see Section 6). Our preliminary results indicate that a majority of the clauses in these laws (12 of 15 GLBA clauses and 47 of 84 HIPAA clauses) can be enforced with little or no audit effort per transmission.

We discuss related work and directions for future work in Sections 7 and 8, respectively.

2. KEY CONCEPTS

Before discussing the formal details of our logical specification of the HIPAA and GLBA privacy laws, it is worthwhile to discuss the key concepts in an informal setting. These concepts are the structure of the privacy laws, including self-referential clauses; abstract data attributes of a transmission; the ability of agents, which we call principals, to dynamically alter the role in which they are active and roles to which they belong; the purpose of a transmission; agents' beliefs about their environment; and temporal conditions for both past provisions and future obligations. This overview simultaneously serves to justify the features of PrivacyLFP, which will be formalized in Section 3.

2.1 Structure of Privacy Laws

2.1.1 Positive and Negative Norms of Transmission

Being based on the philosophical framework of contextual integrity [20] which emphasizes norms of transmission, the fundamental concept of PrivacyLFP, like that of its cousin, LPU [3, 4], is a privacy law's *positive and negative norms*[1].

Positive norms are clauses which state that a transmission may occur *if* a condition is satisfied. For example, §164.506(c)(2) of HIPAA is a positive norm since it allows protected health information to be disclosed *if* the disclosure's purpose is treatment:

> "A covered entity may disclose protected health information for treatment activities of a health care provider."

In this way, positive norms capture the permitting clauses of a regulation.

On the other hand, negative norms are clauses which state that a transmission may occur *only if* a condition is satisfied. For example, the core of HIPAA §164.508(a)(2) is a negative norm since it allows disclosure of psychotherapy notes *only if* it is authorized by the patient (modulo a few exceptions):

> "A covered entity must obtain an authorization for any use or disclosure of psychotherapy notes, except [...]."

Negative norms capture the denying clauses of a regulation because transmissions that do not satisfy the negative norms' conditions are disallowed.

2.1.2 Combining Norms of Transmission

Any privacy law is likely to have more than one norm of transmission, which raises the question of how norms should be combined.

To respect the intuition of positive norms as "ifs," a transmission should be allowed to occur only if it satisfies (at least) *one* of the law's positive norms. Dually, to correctly treat negative norms as "only ifs," a transmission should be allowed to occur only if it also satisfies *all* of the law's negative norms. Stated differently, a transmission should be lawful if and only if it is permitted by at least one of the law's clauses and not denied by any of the law's clauses.

Note that negative norms take precedence over positive norms because of the one-*and*-all structure. If one of the negative norms is unsatisfied, i.e., some clause denies the transmission, then the transmission is not lawful, even if some clause gives permission. In our formalizations of HIPAA and GLBA, we found that this precedence accurately reflects the laws' intent, as conjectured by Barth et al. [3].

2.1.3 Exceptions to Norms of Transmission

In discussing HIPAA §164.508(a)(2) as an example of a negative norm, we conveniently elided the several exceptions to that clause. These include "use [of the notes] by the originator of the psychotherapy notes for treatment." Taking this clause as a canonical example of a negative norm with an exception, we see that exceptions provide a choice: one can either provide evidence of the patient's authorization (thereby satisfying the clause's core), or one can provide evidence that the action is a use by the notes' originator for treatment (thereby satisfying the clause's exception).

Positive norms can also have exceptions, though they have a different flavor. For example, §164.512(c)(1) of HIPAA allows a covered entity to disclose protected health information in reporting cases of abuse or domestic violence, but §164.512(c)(2) makes the exception that such disclosures are allowed only if the covered entity informs the victim of the report. These kind of exceptions simply refine the positive norm to a more specific set of conditions.

2.1.4 Self-Reference

Privacy laws also occasionally reflect on their own definition of lawful transmissions. Specifically, this reflection occurs in §6802(c) of GLBA:

[1]For consistency with LPU's terminology, we use *clause* to refer to a unit of legal text, and use *norm* to refer to a clause's rendering as a logical formula. Note that, in this usage, *norm* does not imply that the clause correctly captures an appropriate social norm.

> "Except as otherwise provided in this subchapter, a nonaffiliated third party that receives from a financial institution nonpublic personal information under this section shall not [...] disclose such information to any other person that is a nonaffiliated third party of both the financial institution and such receiving third party, unless such disclosure would be lawful if made directly to such other person by the financial institution."

In effect, this constitutes a kind of recursive self-reference within a negative norm: GLBA *allows* a disclosure by a nonaffiliated third party to another third party only if GLBA *would allow* a direct disclosure from the financial institution to that other third party.

2.2 Common Concepts in Privacy Laws

2.2.1 Data Attributes

In a physical system, information would be stored as a collection of bytes or strings. But privacy laws do not decide the lawfulness of transmissions at the level of this raw data. Instead, they assume more abstract, structured classes of data, such as "protected health information" and "psychotherapy notes," and regulate according to this classification.

These abstract classes of data also possess a hierarchical structure which the norms of transmission must respect. For example, psychotherapy notes are a particular type of protected health information; a clause that denies certain flows of protected health information should also deny the same flows of psychotherapy notes (unless stated otherwise).

2.2.2 Dynamic Roles

It is impossible for a legal document to give distinct consideration to all possible agents, or principals, by name. Thus, the lawfulness of a transmission is not based on principals' identities, but instead on their roles (e.g., psychiatrist or law enforcement official).

But the roles held by a principal are not static; instead, they evolve over time. For example, §6803(a) of GLBA suggests that principals may become and cease to be customers of a financial institution, i.e., roles are *dynamic*:

> "At the time of establishing a customer relationship with a consumer and not less than annually during the continuation of such relationship, a financial institution shall provide a clear and conspicuous disclosure to such consumer [...], of such financial institution's policies and practices with respect to [disclosing nonpublic personal information]."

Moreover, this clause suggests that roles can be long-standing. The customer relationship is one such typically long-standing role since provisions for annual notices are required. But a principal is not active in the customer role at each moment of the several years during which he is in a customer relationship. Instead, he is variously active in the roles of parent, professor, patient, and, yes, occasionally customer during those years.

2.2.3 Purposes of Uses and Disclosures

In addition to using the transmission's contents, its purpose is often considered when determining the transmission's lawfulness, as in §164.506(c)(2) of HIPAA:

> "A covered entity may disclose protected health information for treatment activities of a health care provider."

Any accurate formalization of this clause must somehow incorporate a transmission's purpose so that it can be checked to match treatment by a provider.

Purposes also obey a hierarchical structure, which must be reflected in PrivacyLFP. For example, the purpose of administering a blood test should be a refinement, or subpurpose, of the treatment purpose.

2.2.4 Principals' Beliefs

Just as a transmission's intended purpose introduces an element of subjectivity, so do principals' beliefs and professional judgment. For example, HIPAA §164.512(f)(4) states:

> "A covered entity may disclose protected health information about an individual who has died to a law enforcement official for the purpose of alerting law enforcement of the death of the individual if the covered entity has a suspicion that such death may have resulted from criminal conduct."

The covered entity's belief that the death may have resulted from criminal conduct is absolutely crucial to the clause's meaning. Without this constraint, §164.512(f)(4) would permit a covered entity to disclose the protected health information of *any* deceased person to law enforcement officials.

2.2.5 Past Provisions and Future Obligations

Allowing a principal to opt-in or opt-out of disclosures are common idioms in privacy regulations. For example, GLBA §6802(b)(1) requires a financial institution to allow opt-out:

> "A financial institution may not disclose nonpublic personal information to a nonaffiliated third party unless— [...] the consumer is given the opportunity, before the time that such information is initially disclosed, to direct that such information not be disclosed to such third party."

In other words, this clause makes the temporal requirement that, at some past time, the consumer was given the opportunity to opt-out of the disclosure and has not since exercised that opportunity. We use the term *provision* to refer to such requirements about past events. Note that this provision implicitly places a bound on the time at which the opportunity was given: it cannot be in the immediate past, for the consumer would not have time to respond. Instead, it must be in the reasonably distant past.

But the temporal conditions present in privacy regulations are not limited to provisions about the past. For example, HIPAA §164.510(a)(2) requires that covered entities provide an opportunity to opt-out of disclosures of directory information, with an exception for cases in which it is not practicable to provide that opportunity (e.g., coma). However, if this exception is used, then §164.510(a)(3)(ii) demands that:

> "The covered health care provider must inform the individual and provide an opportunity to [opt-out of] uses or disclosures for directory purposes as required by paragraph (a)(2) of this section when it becomes practicable to do so."

This imposes a requirement for an event to occur in the future, though there is no concrete time limit since one cannot

predict the time at which it will become practicable to provide an opportunity to opt-out. We use the term *obligation* to refer to such requirements for future events.

In contrast, recall from Section 2.2.2 that §6803(a) of GLBA requires a financial institution to provide a privacy notice to each customer annually throughout the duration of the customer relationship. Although this obligation is unbounded in the sense that there is no concrete limit on the length of the customer relationship, it recurs with a fixed 365-day period, which should make it easier to enforce.

3. THE LOGIC PRIVACYLFP

PrivacyLFP, the logic which we propose for specifying privacy regulations, assigns a particular semantic model and signature to least fixed point logic (LFP) [7, 22].

3.1 A Brief Review of LFP

Assuming a collection of sorts s, first-order terms t, and predicates p, the syntax of LFP formulas is

$$\begin{aligned}\phi, \psi \ ::= \ & p(\vec{t}) \mid \phi \wedge \psi \mid \top \mid \neg\phi \mid \phi \vee \psi \mid \bot \mid \phi \supset \psi \\ \mid \ & \forall x{:}s.\, \phi \mid \exists x{:}s.\, \phi \\ \mid \ & (\mu X(\vec{x}).\, \phi)(\vec{t}) \mid (\nu X(\vec{x}).\, \phi)(\vec{t}) \mid X(\vec{t})\end{aligned}$$

First-order LFP contains all of the usual connectives from first-order logic: conjunction ($\phi \wedge \psi$), truth ($\top$), negation ($\neg\phi$), disjunction ($\phi \vee \psi$), falsehood ($\bot$), implication ($\phi \supset \psi$), and universal and existential quantification over first-order terms of sort s ($\forall x{:}s.\, \phi$ and $\exists x{:}s.\, \phi$). Atomic formulas ($p(\vec{t})$) are built from predicates applied to lists of terms.

The unique feature of LFP is its least and greatest fixed point operators, $(\mu X(\vec{x}).\, \phi)(\vec{t})$ and $(\nu X(\vec{x}).\, \phi)(\vec{t})$, respectively. These define an implicit predicate X as the least and greatest solutions, respectively, of the equation $X(\vec{x}) \triangleq \phi$, and check that the list of terms $\vec{t}$ is in the solution. The predicate variable X and first-order term variables $\vec{x}$ are in scope within ϕ and may be freely renamed. (Since X may appear in ϕ, the atomic formula $X(\vec{t})$ is also included in LFP's syntax.)

For example, we can define a predicate $\mathsf{nat}(n)$, stating that n is a natural number, as the least fixed point of the equation

$$X(x) \triangleq (x = 0) \vee \big(\exists y.\, (x = y + 1) \wedge X(y)\big).$$

So, $\mathsf{nat}(n) \triangleq \big(\mu X(x).\, (x = 0) \vee \big(\exists y.\, (x = y+1) \wedge X(y)\big)\big)(n)$.

Being recursive, this definition of $\mathsf{nat}(n)$ is similar in flavor to the self-referential clauses described in Section 2.1.4; in Section 4.1.2, we will show how to express self-referential clauses of privacy laws using the fixed point operators.

3.2 Signature and Semantic Model

To express the kind of past provisions and future obligations described in Section 2.2.5, PrivacyLFP encodes standard connectives from linear temporal logic (LTL) [18] as syntactic sugar in LFP: $\diamondminus\phi$ and $\diamondsuit\phi$ ("At some time[2] in the past (resp. future), ϕ is true."), $\boxminus\phi$ and $\square\phi$ ("At every time in the past (resp. future), ϕ is true."), $\mathbf{G}\ \phi$ ("At every time (past or future), ϕ is true."), $\phi\, \mathcal{S}\, \psi$ ("ψ is true at some past time and ϕ is true from then to now, i.e., 'since then'."), and $\phi\, \mathcal{W}\, \psi$ ("Either ϕ is true at all future times or ϕ is true from now until ψ is true."). PrivacyLFP also includes as syntactic sugar the "freeze" quantifier $\downarrow x.\, \phi$ from timed propositional temporal logic (TPTL) [1] which binds x in ϕ to the "current" time of the local context. For example, $\downarrow x.\, \diamondminus(\downarrow y.\, (y > x - 10) \wedge \phi)$ means that ϕ was true at some time in the past less than ten units ago.

PrivacyLFP assigns a trace-based semantic model to LFP. Traces are sequences of states in which principals concurrently perform actions, such as sending a message or beginning a new role. As such, traces capture the evolution of the system and its environment.

PrivacyLFP also consists of a first-order signature of sorts, terms, and predicates to express the various concepts that we motivated in Section 2.2. Data attributes and purposes are reflected as term constants of dedicated sorts whereas the attribute hierarchy, structure of purposes, roles, trace actions, and principals' beliefs are accounted for by predicates. Further discussion of these sorts, term constants, and predicates is postponed to the following section, where we give formalizations of the clauses that motivated them.

(For full details of both the encoding of temporal operators in LFP and the semantic model, we refer the interested reader to the companion technical report [11].)

4. FORMALIZING HIPAA AND GLBA

With a brief overview of PrivacyLFP and intuition for its semantics in hand, we now turn to formalizing HIPAA and GLBA in the logic. We return to the structure and common concepts presented in Section 2. This time, we put them on a formal footing by demonstrating their appearance in the logical formalizations.

4.1 Structure

As an overall goal, we would like to have a means, via logical specification, of verifying that only lawful transmissions occur.

In PrivacyLFP, occurrences of transmissions are characterized by the predicate $\mathsf{send}(p_1, p_2, m)$, meaning that principal p_1 is currently sending message m to principal p_2. In practice, we expect the system to set $\mathsf{send}(p_1, p_2, m)$ to true when p_1 sends m to p_2, thereby providing a practical trace-based semantics to the send predicate.

When formalizing a particular law, we will define a predicate $\mathsf{maysend}$ that characterizes the lawful transmissions under that regulation. Specifically, $\mathsf{maysend}(p_1, p_2, m)$ means that the transmission of message m from p_1 to p_2 is lawful. Then, our overall goal is specified by the top-level formula

$$\mathbf{G}\ \big(\forall p_1, p_2, m.\ \big(\mathsf{send}(p_1, p_2, m) \supset \mathsf{maysend}(p_1, p_2, m)\big)\big).$$

That is, at every possible time, we require that if a transmission occurs, it is lawful. Now we must define $\mathsf{maysend}$.

4.1.1 Combining Norms of Transmission

To give a definition of $\mathsf{maysend}(p_1, p_2, m)$, we recall the principles for combining norms of transmission discussed in Section 2.1.1. In summary, these principles define a transmission to be lawful if and only if it satisfies (at least) one of the law's positive norms and all of the law's negative norms. Thus, we arrive at the defining equation

$$\mathsf{maysend}(p_1, p_2, m) \triangleq \Big(\bigvee_i \varphi_i^+\Big) \wedge \Big(\bigwedge_j \varphi_j^-\Big),$$

[2]Readers familiar with LTL will note that we abuse terminology slightly, often using 'time' where 'state' would be more correct. This is done to aid the intuition of readers not familiar with LTL.

where the φ_i^+s and φ_j^-s are the law's positive and negative norms, respectively. This definition faithfully captures the one-all duality between positive and negative norms: the disjunction $\bigvee_i \varphi_i^+$ is satisfied if and only if (at least) one of the φ_i^+s is satisfied, and the conjunction $\bigwedge_j \varphi_j^-$ is satisfied if and only if all of the φ_j^-s are satisfied.

4.1.2 Self-Reference via Fixed Points

When the privacy regulation under consideration does not make any self-references, the maysend predicate will not appear in the formalization of the law's norms. In this case, the defining equation for $\mathsf{maysend}(p_1, p_2, m)$ from Section 4.1.1 has a unique solution, and we may simply treat it as specifying a macro. (This is the case for HIPAA, as we will see in Section 4.2.1.)

However, as discussed in Section 2.1.4, privacy regulations do occasionally make self-references, as exemplified by §6802(c) of GLBA. In such cases, the self-references translate to recursive calls to the maysend predicate. For example, a natural formalization of GLBA §6802(c) would be the following negative norm.

$$\begin{aligned}\varphi^-_{\texttt{6802c}} \triangleq \forall p'.\ \Big(&\neg\mathsf{activerole}(p_1, \mathit{affiliate}(p')) \wedge \\ &\neg\mathsf{activerole}(p_2, \mathit{affiliate}(p')) \wedge \\ &\neg\mathsf{activerole}(p_2, \mathit{affiliate}(p_1)) \wedge \\ &(t \in_{\mathcal{T}} \mathit{npi}) \wedge \\ &\diamondsuit\big(\exists m'.\ \mathsf{hlsend}(p', p_1, m') \wedge \\ &\qquad \mathsf{contains}(m', q, t) \wedge \\ &\qquad \mathsf{activerole}(p', \mathit{institution}) \wedge \\ &\qquad \neg\mathsf{activerole}(p_1, \mathit{affiliate}(p')) \wedge \\ &\qquad \mathsf{belongstorole}(q, \mathit{consumer}(p'))\big)\Big) \\ \supset\ &\\ &\mathsf{maysend}(p', p_2, \{(q, t)\})\end{aligned}$$

(The free variables t and q represent the data being disclosed, and its subject, respectively. The atomic formulas $\mathsf{activerole}(p, r)$ and $\mathsf{belongstorole}(p, r)$ mean that p is active in role r and holds role r, respectively; $(t \in_{\mathcal{T}} \mathit{npi})$ means that attribute t is nonpublic personal information; and $\{(q, t)\}$ is the message composed of the single attribute t for subject q. See the subsequent sections for details of these predicates.)

Once maysend appears in the norms, its defining equation no longer has a unique solution in general. The question, then, is which solution is semantically correct: the least solution, the greatest solution, or something else altogether? We claim that the greatest solution is intuitively the correct interpretation because the formalization should not impose any constraints beyond those required by the law. Stated differently, the formalization should deem lawful all transmissions that are not explicitly ruled out by the law.

Thus, in general, we arrive at a greatest fixed point definition for $\mathsf{maysend}(p_1, p_2, m)$:

$$\left(\nu\, \mathsf{maysend}(p_1, p_2, m).\ \left(\bigvee_i \varphi_i^+\right) \wedge \left(\bigwedge_j \varphi_j^-\right)\right)(p_1, p_2, m).$$

The right hand side of the defining equation of Section 4.1.1 is present in LPU [3], but our factoring into a defining equation and subsequent fixed point definition is new to PrivacyLFP.

4.1.3 Logical Structure of Norms of Transmission

In the privacy regulations that we have examined, positive and negative norms have characteristic logical structures. Because the positive norms specify a list of conditions under which a transmission is allowed, the φ_i^+ syntactic category consists of a conjunction of predicates about the various principals involved. Some of these predicates may be negated or prefixed with temporal operators.

On the other hand, because negative norms specify denials, the φ_j^- syntactic category essentially consists of the form $\phi \supset \psi$, where ψ specifies the *only if* conditions for allowed transmissions satisfying ϕ. A transmission satisfying ϕ but not ψ is denied.

These general logical structures of positive and negative norms are also present in LPU [3].

4.1.4 Exceptions

In Section 2.1.3, we argued that negative norms with exceptions provide a choice between satisfying the core negative norm or one of its exceptions. In the logic, this choice naturally corresponds to a disjunction. For example, if the core of HIPAA §164.508(a)(2) is formalized as the negative norm $\varphi^-_{\texttt{164.508a2}}$ and its exceptions are formalized as $\varphi^e_{\texttt{164.508a2iA}}$, $\varphi^e_{\texttt{164.508a2iB}}$, $\varphi^e_{\texttt{164.508a2iC}}$, and $\varphi^e_{\texttt{164.508a2ii}}$, then, as a whole, the clause can be formalized as

$$\begin{aligned}\varphi^-_{\texttt{164.508a2}'} \triangleq\ &\varphi^-_{\texttt{164.508a2}} \vee \\ &\big((\varphi^e_{\texttt{164.508a2iA}} \vee \varphi^e_{\texttt{164.508a2iB}} \vee \varphi^e_{\texttt{164.508a2iC}}) \vee \\ &\ \varphi^e_{\texttt{164.508a2ii}}\big)\end{aligned}$$

The formula $\varphi^-_{\texttt{164.508a2}'}$ is treated as a negative norm since the core is a negative norm and the exceptions apply only locally to it. In other words, exceptions do not destroy the structure of privacy laws laid out in the definition of the maysend predicate — we simply relax the syntactic category φ_j^-, allowing it to include negative norms that are qualified with exceptions.

As discussed in Section 2.1.3, exceptions to positive norms act as refinements. Logically, refinement is captured by conjunction. For example, if §§164.512(c)(1) and (2) are formalized as the positive norm $\varphi^+_{\texttt{164.512c1}}$ and exception $\varphi^e_{\texttt{164.512c2}}$, then, as a whole, those clauses can be formalized as

$$\varphi^+_{\texttt{164.512c1}'} \triangleq \varphi^+_{\texttt{164.512c1}} \wedge \varphi^e_{\texttt{164.512c2}},$$

with the formula $\varphi^+_{\texttt{164.512c1}'}$ treated as a positive norm.

4.2 Specific Structure of HIPAA and GLBA

Before delving into the expression of common features of privacy laws in PrivacyLFP, it is useful to describe the specific structure of our HIPAA and GLBA formalizations.

4.2.1 HIPAA

To demonstrate PrivacyLFP's utility, we formalized all transmission-related requirements in HIPAA relevant to the disclosure of information. Specifically, we encoded §§164.502, 164.506, 164.510, 164.512, 164.514, and 164.524. HIPAA is primarily composed of positive norms (58), with relatively fewer negative norms (7) and exceptions (19). This reflects HIPAA's general strategy to implicitly forbid all transmissions not explicitly listed as permitted by positive norms.

Most of HIPAA's negative norms are found in §§164.508, 164.510, and 164.524. Since §§164.508 and 164.510 deal with disclosures for which the patient must consent or be provided with an opportunity to opt-out, respectively, this is unsurprising: the disclosure is not denied only if the relevant consent or opt-out opportunity is given. Similarly, the

norms of §164.524 are negative because they obligate a covered entity to respond, within a fixed time limit (typically 30 days), to a patient's requests for access to his medical records.

On the other hand, §§164.506, 164.512, and 164.514 list purposes for which transmissions are permitted, including treatment, public health, law enforcement, and responding to court orders. Because any one of these purposes suffices as permission, the norms from these sections are positive.

Adhering to the above discussion of general structure of privacy laws, we arrive at the following top-level formula for HIPAA. (As we will discuss in Section 4.3.1, $\mathsf{contains}(m, q, t)$ means that message m contains the value of attribute t for subject q.)

$$\mathbf{G}\,\Big(\forall p_1, p_2, m.\ \mathsf{send}(p_1, p_2, m) \supset \big(\forall d, u, q, t.\ (m = \mathrm{info}(d, u)) \wedge \mathsf{contains}(m, q, t) \supset (\textstyle\bigvee_i \varphi_i^+) \wedge (\bigwedge_j \varphi_j^-)\big) \wedge \big(\forall t.\ (m = \mathrm{req\_for\_access}(p_1, t)) \supset \varphi^-_{\texttt{164.524b2i'}} \wedge \varphi^-_{\texttt{164.524b2ii'}}\big)\Big)$$

Because HIPAA contains no self-references, there is a unique solution to the equation defining $\mathsf{maysend}(p_1, p_2, m)$, and so we implicitly expand the definition. Since §164.524 handles access requests, rather than transmissions of information, we also distinguish cases on the shape of message m: it carries either information ($m = \mathrm{info}(d, u)$) or a request for access ($m = \mathrm{req\_for\_access}(p_1, t)$). Consequently, the φ_i^+s and φ_j^-s stand for all HIPAA norms except those from §164.524.

4.2.2 GLBA

The clauses in GLBA relevant to transmissions are found in §6802, which lists requirements with respect to disclosures, and §6803, which relates to annual privacy notices. Because both sections impose demands upon financial institutions, GLBA contains only negative norms (5) and exceptions (10). Although the law lists no positive norms explicitly, it can be seen to have an implicit positive norm, $\top$, which permits all transmissions since they trivially satisfy $\top$. This fits with GLBA's general tactic of deeming lawful all transmissions that are not explicitly denied.

Again, adhering to the principles for general structure of privacy regulations in PrivacyLFP, we arrive at the following top-level formula for GLBA:

$$\mathbf{G}\,\Big(\big(\forall p_1, p_2, m.\ \mathsf{hlsend}(p_1, p_2, m) \supset \big(\nu\,\mathsf{maysend}(p_1, p_2, m).\ \forall d, u, q, t.\ (m = \mathrm{info}(d, u)) \wedge \mathsf{contains}(m, q, t) \supset \top \wedge (\varphi^-_{\texttt{6802ae}} \wedge \varphi^-_{\texttt{6802be}} \wedge \varphi^-_{\texttt{6802c}} \wedge \varphi^-_{\texttt{6802d}})\big)(p_1, p_2, m)\big) \wedge \big(\forall q, r, p.\ \mathsf{beginrole}(q, r) \wedge (r = \mathit{customer}(p)) \supset \varphi^-_{\texttt{6803a}} \vee \varphi^e_{\texttt{6803d1}}\big)\Big)$$

As previously discussed, because GLBA §6802(c) contains a self-reference, the equation defining $\mathsf{maysend}(p_1, p_2, m)$ does not have a unique solution. In keeping with the previous discussion, we therefore use a greatest fixed point operator.

GLBA §6803 handles annual privacy notices, rather than requirements on disclosure of nonpublic personal information. Thus, similarly to HIPAA's top-level formula, GLBA's top-level formula separates out §6803's negative norm and exception from the rest of the norms. Note that these are triggered by a principal q beginning in the role of customer, explaining the presence of $\mathsf{beginrole}(q, r)$.

Finally, note that $\mathsf{hlsend}(p_1, p_2, m)$, which stands for "high-level send," replaces the send predicate in GLBA. This is because §6809(9) defines a consumer to be either the individual or his legal representative. The hlsend macro is therefore used to allow the legal representative to act as a suitable substitute for the sender or recipient by abstracting away from which principal physically sent or received the message m. Although all other clauses in GLBA could be formalized in a way that maintains a direct correspondence between the legal text and logical formulas, this is an example where a clause nonlocally affects the rest of the formalization.

4.3 Common Concepts of Privacy Laws

We now discuss how PrivacyLFP formally expresses the common concepts of privacy laws described in Section 2.2.

4.3.1 Data Attributes

To formally model data attributes, PrivacyLFP follows LPU's [3, 4] lead and introduces a sort attr for attributes. As in LPU, attribute structure is characterized by a set of specialization rules which dictate when the value of a principal's attribute can be inferred from other information, similarly to a subtyping relation. These rules are internalized by the $\in_{\mathcal{T}}$ predicate so that $(t_1 \in_{\mathcal{T}} t_2)$ means that t_1 is a subattribute of attribute t_2. For example, $(\mathit{psychotherapy\text{-}notes} \in_{\mathcal{T}} \mathit{phi})$ is true since psychotherapy notes are a particular type of protected health information.

To connect these attributes to the underlying raw messages, PrivacyLFP uses a contains predicate, again borrowed from LPU, where $\mathsf{contains}(m, q, t)$ means that message m contains the value of subject q's attribute t. Due to the vast complexity of extracting classifications from raw messages, we cannot give an explicit definition for this predicate, and instead rely on human intervention for its semantics.

4.3.2 Dynamic Roles

To formally support a notion of role, PrivacyLFP follows LPU [3, 4] and contains a sort role of roles equipped with a partial order $\preceq_{\mathcal{R}}$ that expresses role specialization. For example, we have $\mathit{psychiatrist} \preceq_{\mathcal{R}} \mathit{provider}$ since a psychiatrist is a particular type of health care provider.

To support *dynamic* roles (not found in LPU), PrivacyLFP distinguishes the roles that a principal holds from the role in which he is currently active via the belongstorole and activerole predicates. The formula $\mathsf{belongstorole}(p, r)$ means that principal p currently holds (but is not necessarily active in) role r, whereas $\mathsf{activerole}(p, r)$ means that p is currently active in role r. To start or finish belonging to a role, the principal must initiate special actions, which are characterized by the predicates $\mathsf{beginrole}(p, r)$ and $\mathsf{endrole}(p, r)$. On the other hand, a principal's active role may freely change over time without constraint.

4.3.3 Purposes of Uses and Disclosures

To formally account for purposes of transmissions, PrivacyLFP includes a sort purp of purposes, equipped with a partial order that models the underlying hierarchy. The predicate $\in_{\mathcal{U}}$ internalizes this hierarchy, so that $(u_1 \in_{\mathcal{U}} u_2)$ is true whenever u_1 is a subpurpose of purpose u_2. For in-

stance, ($\mathit{administer\text{-}blood\text{-}test} \in_{\mathcal{U}} \mathit{treatment}$) is true because administering a blood test is a type of treatment purpose.

With purposes and roles, we can now specify §164.506(c)(2) of HIPAA (see Section 2.2.3) as the positive norm

$$\begin{aligned}\varphi^{+}_{\texttt{164.506c2}} \triangleq\ & \mathsf{activerole}(p_1, \mathit{covered\text{-}entity}) \wedge \\ & \mathsf{activerole}(p_2, \mathit{provider}(q)) \wedge \\ & (t \in_{\mathcal{T}} \mathit{phi}) \wedge \\ & (u \in_{\mathcal{U}} \mathit{treatment}(p_2)).\end{aligned}$$

This formula is satisfied if *1*) the sender p_1 is currently acting as a covered entity; *2*) the recipient p_2 is currently acting as a health care provider for q, the subject of attribute t; *3*) the attribute t is (a subattribute of) protected health information; and *4*) the message's purpose, u, is (a subpurpose of) treatment by provider p_2.

4.3.4 Principals' Beliefs

It is extremely difficult, if not impossible, to give a semantics for professional judgment and beliefs due to their inherent subjectivity. Therefore, PrivacyLFP relegates these concepts to uninterpreted predicates of the form $\mathsf{believes\text{-}}\ldots$, which rely on human intervention for their semantics.

For example, we formalize §164.512(f)(4) of HIPAA (see Section 2.2.4) as the positive norm

$$\begin{aligned}\varphi^{+}_{\texttt{164.512f4}} \triangleq\ & \mathsf{activerole}(p_1, \mathit{covered\text{-}entity}) \wedge \\ & \mathsf{activerole}(p_2, \mathit{law\text{-}enforcement\text{-}official}) \wedge \\ & \mathsf{belongstorole}(q, \mathit{deceased}) \wedge \\ & (t \in_{\mathcal{T}} \mathit{phi}) \wedge \\ & (u \in_{\mathcal{U}} \mathit{death\text{-}notification}(q)) \wedge \\ & \mathsf{believes\text{-}death\text{-}may\text{-}be\text{-}result\text{-}of\text{-}crime}(p_1, q).\end{aligned}$$

This formula is satisfied if *1*) the sender p_1 is currently acting as a covered entity; *2*) the recipient p_2 is currently acting as a law enforcement official; *3*) the subject, q, of the disclosure currently belongs to the (long-term) role of a deceased; *4*) the attribute t is (a subattribute of) protected health information; *5*) the message's purpose is to report a death; and *6*) p_1 believes that q's death may be the result of a crime.

4.3.5 Past Provisions and Future Obligations

Support for provisions and obligations is achieved using the set of temporal modalities adapted from LTL [18] and TPTL [1]. Using the freeze quantifier, $\downarrow x.\, \phi$, in combination with the $\diamondminus$, $\boxminus$, or $\mathcal{S}$ modalities, we can express clauses containing provisions about events in the bounded past. We previously cited GLBA §6802(b)(1) as such a clause, which is formalized as a negative norm:

$$\begin{aligned}\varphi^{-}_{\texttt{6802b1}} \triangleq\ & \downarrow x.\ \mathsf{activerole}(p_1, \mathit{institution}) \wedge \\ & \neg\mathsf{activerole}(p_2, \mathit{affiliate}(p_1)) \wedge \\ & \mathsf{belongstorole}(q, \mathit{consumer}(p_1)) \wedge \\ & (t \in_{\mathcal{T}} \mathit{npi}) \\ & \supset \\ & \Big(\big(\neg\exists m''.\ \mathsf{hlsend}(q, p_1, m'') \wedge \\ & \qquad \mathsf{is\text{-}opt\text{-}out}(m'', p_1, p_2, (q,t), u)\big) \\ & \ \mathcal{S} \\ & \ \big(\downarrow y.\ (x \geq y + c) \wedge \\ & \qquad \exists m'.\ \mathsf{hlsend}(p_1, q, m') \wedge \\ & \qquad\quad \mathsf{is\text{-}notice\text{-}of\text{-}potential} \\ & \qquad\quad\ \mathsf{\text{-}disclosure}(m', p_1, p_2, (q,t), u)\big)\Big)\end{aligned}$$

In other words, if *1*) the current time is x; *2*) the sender p_1 is currently active in the role of financial institution; *3*) the recipient p_2 is not currently active in the role of affiliate of p_1; *4*) the subject q holds the role of consumer of p_1; and *5*) the attribute t is nonpublic personal information; then this clause demands that *1*) at some time y at least c units of time before time x, some notice of potential disclosure, m', was sent from the financial institution to the consumer; and *2*) no opt-out message m'' was sent from the consumer to the financial institution since time y.

PrivacyLFP also includes the $\diamondsuit$, $\square$, and $\mathcal{W}$ modalities to express obligations about future events. For example, using $\mathcal{W}$, we can express the unbounded obligation of HIPAA §164.510(3)(ii) to provide opportunity to opt-out of disclosures of directory information as soon as is practicable:

$$\begin{aligned}\varphi^{e}_{\texttt{164.510a3ii}} \triangleq\ & (\neg\mathsf{practicable\text{-}to\text{-}provide\text{-}opt\text{-}out} \\ & \qquad\qquad \mathsf{\text{-}opportunity}(p_1, p_2, (q,t), u)) \\ & \mathcal{W} \\ & (\exists m'.\ \mathsf{send}(p_1, q, m') \wedge \\ & \qquad \mathsf{is\text{-}opt\text{-}out\text{-}opportunity}(m', p_1, p_2, (q,t), u))\end{aligned}$$

This formula means that either *1*) it is never practicable for the covered entity p_1 to provide an opt-out opportunity to the patient q; or *2*) there is some future time where an opt-out opportunity is provided, and between now and then it is not practicable to provide that opportunity.

By combining the $\diamondsuit$, $\square$, or $\mathcal{W}$ modalities with the freeze quantifier, PrivacyLFP allows us to express obligations about future events with a bounded time limit. For instance, GLBA §6803(a), can be expressed as the negative norm:

$$\begin{aligned}\varphi^{-}_{\texttt{6803a}} \triangleq\ & (\downarrow x.\ \diamondsuit(\downarrow y.\ (y \leq x + 365) \wedge \\ & \qquad (\mathsf{endrole}(q, \mathit{customer}(p_1)) \vee \\ & \qquad\ (\exists m'.\ \mathsf{hlsend}(p_1, q, m') \wedge \\ & \qquad\qquad \mathsf{is\text{-}annual\text{-}notice}(m', p_1, q))))) \\ & \mathcal{W} \\ & \mathsf{endrole}(q, \mathit{customer}(p_1))\end{aligned}$$

In other words, at each time x between now and the time when the subject q ceases to be a customer of the financial institution p_1, ensure that there is some future time y, no later than 365 days after x, where either q is no longer a customer or p_1 has sent an annual notice to q.

5. DISCUSSION OF HIPAA AND GLBA

In formalizing HIPAA and GLBA, the careful reading and precision demanded by a logical encoding revealed ambiguities and potential improprieties in several of the laws' clauses.

5.1 Range of Limits on Redisclosures

Recall from Section 2.1.4 that GLBA §6802(c) places limits on the redisclosure of information: a nonaffiliate that received information from a financial institution may redisclose that information to another nonaffiliate if the institution could lawfully disclose the information to that other nonaffiliate directly. The intent is to constrain flow of nonpublic information among nonaffiliates to those disclosures that are lawful for the originating financial institution to make directly. However, as written, this clause does not fully capture that intent, as seen in the following scenario.

Suppose that financial institution p discloses some information to nonaffiliate p_0, who, satisfying §6802(c), discloses it to another nonaffiliate p_1. Now p_1 wishes to disclose the information to yet another nonaffiliate p_2. As p_1 did not

receive the information *directly* from a financial institution, the norm $\varphi^{-}_{\mathsf{6802c}}$ is trivially satisfied!

In our opinion, this is a loophole in the wording of the law, for it means that all protections over information are lost at a range greater than two hops from its origin. Rather than confining this clause to a sender who is a "nonaffiliated third party that *receives* [information] from a financial institution," it would be better if the clause applied to nonaffiliates that receive information which *originated* at a financial institution. We can use a least-fixed point to encode the concept of 'originated at a financial institution' and express this new clause as:

$$\begin{aligned}\varphi^{-}_{\mathsf{6802c}} \triangleq \forall p'.\ \Big(&\neg\mathsf{activerole}(p_1, \mathit{affiliate}(p')) \wedge \\ &\neg\mathsf{activerole}(p_2, \mathit{affiliate}(p')) \wedge \\ &\neg\mathsf{activerole}(p_2, \mathit{affiliate}(p_1)) \wedge \\ &(t \in_{\mathcal{T}} \mathit{npi}) \wedge \\ &\big(\mu\,\mathsf{info\text{-}originated\text{-}at\text{-}institution}(p', p_1).\\ &\quad \diamondsuit\big(\exists m'.\ \mathsf{hlsend}(p', p_1, m') \wedge \\ &\qquad \mathsf{contains}(m', q, t) \wedge \\ &\qquad \mathsf{activerole}(p', \mathit{institution}) \wedge \\ &\qquad \neg\mathsf{activerole}(p_1, \mathit{affiliate}(p')) \wedge \\ &\qquad \mathsf{belongstorole}(q, \mathit{consumer}(p'))\big) \\ &\quad \vee \big(\exists p_0.\ \diamondsuit\big(\exists m'.\ \mathsf{hlsend}(p_0, p_1, m') \wedge \\ &\qquad\qquad \mathsf{contains}(m', q, t)\big) \\ &\qquad \wedge \mathsf{info\text{-}originated\text{-}at} \\ &\qquad\qquad \mathsf{-institution}(p', p_0)\big) \\ &\big)(p', p_1)\Big) \\ \supset\ &\mathsf{maysend}(p', p_2, \{(q, t)\})\end{aligned}$$

Furthermore, §6802(c) is vague about the relative timing of the transmission from the nonaffiliate to the other third party and the hypothetical transmission from the financial institution to the other third party. Does §6802(c) require the hypothetical transmission to be lawful at the same time as the actual transmission? Or, does it require the hypothetical transmission to be lawful at some strictly earlier time? Formally, these two interpretations would be distinguished by the presence or absence of a strict $\diamondsuit$ modality in front of the recursive call to the maysend predicate.

It is unclear to us which interpretation is intended; since §6802(c) makes no mention of time, we assume that no strict $\diamondsuit$ modality should be used. We would like to point out that, because it concerns a recursive call, this seemingly small difference has subtle effects. By including a strict $\diamondsuit$ modality, the recursive calls would become ordered in time, preventing cyclic dependencies between them. In this case, the least and greatest solutions would coincide, and we could choose either a least or greatest fixed point for defining maysend.

5.2 Enforcing Purposes

Another point worth discussing is the extent to which a transmission's purpose is enforced. As an example, consider HIPAA §164.512(b)(1)(i), which permits a covered entity to disclose information to a public health authority for the purpose of preventing or controlling a disease. Is the public health authority expected to ensure that all future uses or disclosures of that information are for the purpose of preventing or controlling that disease? Or, having received the information, is the public health authority free to use or disclose the information for *any* purpose (provided, of course, that those uses or disclosures are themselves lawful)? Alternatively, does the purpose apply to the underlying information or to only the transmission itself?

In formalizing HIPAA, we found no further discussion of which interpretation of purposes is intended, either for §164.512(b)(1)(i) specifically or in general. It is unclear which interpretation is better from a practical standpoint. For the strongest privacy guarantees, purposes that constrain all future uses or disclosures are certainly preferable. But such strong constraints might become problematic in practice since a covered entity might neglect to include or fail to anticipate all necessary purposes. For example, a doctor might submit blood tests to a third party laboratory for analysis but forget to allow the billing purpose. Since the laboratory is unable to use the information for billing the patient, the blood analysis is not completed, preventing the doctor from providing optimal care.

5.3 Balancing Privacy and Utility in HIPAA

One view of privacy, due to Westin [25], that is commonly advocated emphasizes the primacy of an individual's control over when, how, and to what extent his personal information is disclosed. It is interesting to note that HIPAA is not fully consistent with this view. For example, when the disclosure is for the purpose of treatment (§164.506(c)(2)) or the purpose of controlling disease (§164.512(b)(1)(i)), the disclosure may occur without the individual's consent. Similarly, an individual's right of access to his records is void if the access would endanger himself or another person (§§164.524(a)(2)(ii) and (3)(i)). In these cases, the decision to permit or deny a flow of information is based on whether it supports the goals or purposes of the health care context, not on whether the individual desires it. While these clauses go against the grain of the control-first philosophy of privacy, they are indeed consistent with other philosophical models, such as contextual integrity [20].

6. TOWARD ENFORCEMENT

Besides explaining the meaning of privacy laws, clarifying their ambiguities, and reflecting on their propriety, a primary goal of formalizing HIPAA and GLBA is to provide insights to guide the development of principled computer systems for enforcement of the laws. Based on an analysis of our formalizations, we contend that any reasonable, practical system for enforcement of these laws should have two facets: *1*) execution-time access control mechanisms that optimistically resolve unknown predicates, postponing them to *2*) post-hoc audit of access logs.

Execution-time access control mechanisms alone do not suffice to enforce HIPAA and GLBA because most clauses include constraints that, a priori, cannot be checked mechanically. Examples are the contains predicate for classifying messages' free text contents as attributes, future obligations, principals' beliefs, and transmission purposes. (All HIPAA and GLBA clauses, except §6803(d)(1) of GLBA, depend on the contains predicate; virtually all HIPAA clauses and 1/2 of GLBA clauses involve inherently subjective purposes or beliefs; and 7/8 of HIPAA clauses and 2/3 of GLBA clauses use other uninterpreted predicates, especially those for message formats, such as is-annual-notice.) Because these cannot be decided mechanically, the access control mechanism must either rely on human involvement at the time of access, or optimistically assume the constraints are satisfied and fall back on post-hoc audit of access logs.

Having an access policy that demands human involvement at the time of access may be unacceptable where utility goals are concerned. For example, hospitals' policies may eagerly give health care providers access to *all* patients' records, only because the providers may *possibly* have, in the future, a legitimate purpose (e.g., medical emergency) for accessing the records; demanding human involvement at the time of access would delay patient care. Thus, the access policy should optimistically assume that the constraints hold and fall back on post-hoc audit.

The objective, then, is to devise creative decision procedures for as many of these constraints as possible, thereby reducing and structuring human involvement in post-hoc audit. We envisage two such techniques.

First, some predicates can be mechanized if data formats are standardized by a human expert. For instance, in formalizing GLBA §6803(a) (see Section 4.3.5), we require the is-annual-notice predicate to verify that the annual message is indeed a notice of the financial institution's privacy policies. In practice, the institution's legal experts could craft a single message and verify that it is GLBA-compliant. Since the same message would be sent to all customers, the truth of is-annual-notice would be predetermined en masse for all customers' notices. Similarly, standardized data formats would ease the task of classifying messages' contents, e.g., as protected health information by $\mathsf{contains}(m, q, phi)$.

Second, the truth of some predicates can be guaranteed by a design-time analysis of an organization's business processes. For example, HIPAA §164.510(a)(1)(ii) allows disclosure of directory information for directory purposes. If a hospital's information desk is designed to have access to directory information only and respond only to directory requests, then the constraint $(u \in directory)$ can be guaranteed to be true for the information desk.

Using these techniques, we estimate that 8 of the 15 GLBA clauses and 17 of the 84 HIPAA clauses could be enforced with no audit effort per disclosure. We can classify the remaining clauses as requiring either a small amount of non-expert audit effort or a significant amount of expert audit effort (per disclosure). For example, it seems reasonable for a trained but non-expert human to verify most disclosure purposes with a small amount of effort, such as checking that a message is for the purpose of preventing fraud (GLBA §6802(e)(3)). On the other hand, clauses that involve principals' beliefs or professional judgment (e.g., HIPAA §164.512(f)(4)) or determining whether a disclosure is in accordance with other laws (e.g., GLBA §6802(e)(5)) demand larger amounts of effort from medical or legal experts. Under this classification, we estimate that 12 of the 15 GLBA clauses and 47 of the 84 HIPAA clauses can be enforced with only little or no non-expert audit effort; the interested reader may refer to the companion technical report [11] for a classification of each clause's audit effort.

7. RELATED WORK

We divide the most closely related work into three categories: formalization efforts for HIPAA and GLBA, logics for specifying policies and regulations, and privacy languages.

Formalization Efforts for HIPAA and GLBA.

PrivacyLFP extends Barth et al.'s Logic of Privacy and Utility (LPU) [3, 4] with support for representing real-time, purposes, and self-reference. Although similar in structure to their proof-of-concept formalization of five HIPAA and four GLBA clauses, our formalizations cover a much larger part of both laws. Lam et al. have formalized §§164.502, 164.506, and 164.510 of HIPAA in a fragment of stratified Datalog with one alternation of negation, and built a prototype tool to check the lawfulness of a transmission [16]. However, their formalization is not as complete as ours, partly due to the lack of temporal modalities in Datalog that are necessary to express future obligations. Breaux and Antón have developed a methodology for extracting rights and obligations from natural language privacy laws, and applied it to the entire text of HIPAA [8]. Their approach is quite complementary to ours and could possibly ease the logician's task of translating a privacy law into logical formulas. May et al. [19] presented privacy APIs, which extend the traditional matrix model of access control with constructs for logging, and used them to formalize two versions of HIPAA §164.506.

Logics for Specifying Policies and Regulations.

Hilty et al. [14] have shown how to specify future obligations from data protection policies in Distributed Temporal Logic (DTL). They used distributed event structures to model interactions between multiple parties involved in data access and distribution. Because our semantic model is a flat trace of actions, we plan to investigate incorporating distributed structures. Basin et al. [5] used an extension of LTL, Metric First-Order Temporal Logic (MFOTL) for specifying security properties. MFOTL does not have the freeze quantifier, and therefore cannot encode policies that use explicit times. Choosing deontic and temporal logics as a foundation, Dinesh et al. have developed a logic for reasoning about conditions and exceptions in privacy laws [12]. This approach is advantageous in that it simplifies the task of formalizing the law clause by clause: there is no need to modify previously formalized clauses if exceptions appear in later paragraphs. Further investigation is needed to determine whether their ideas can be adapted to our logic.

Privacy Languages.

Privacy languages such as EPAL [2] and XACML [21] are formulated as access control frameworks. EPAL and XACML do not include temporal modalities as primitives, but instead have a much weaker uninterpreted obligation symbol for representing future requirements (see [3]). Our logic, with its rich temporal and obligation constructs, is therefore more expressive. Role-based access control languages (e.g., [15, 17]) emphasize roles but lack notions of data attributes and temporal modalities needed for rich privacy laws. P3P [9] is a privacy language targeted exclusively to web sites, but is unsuited for expressing privacy laws like HIPAA and GLBA due to its domain-specific design and lack of general temporal modalities.

8. CONCLUSION AND FUTURE WORK

We have presented the PrivacyLFP logic for specifying privacy regulations and used it to formalize HIPAA and GLBA. Our ultimate goal is to enforce privacy regulations in a framework directly based on this logic. PrivacyLFP already eases this process by encoding temporal operators as first-order formulas, thereby allowing an off-the-shelf model checker for first-order LFP [6, 13] to perform any model

checking used in enforcement, and by factoring purposes into the $\in_{\mathcal{U}}$ predicate, thereby making possible mechanisms for enforcing purposes that would apply to all laws.

Nonetheless, enforcement via PrivacyLFP would benefit from future work in two directions. First, extending PrivacyLFP with a semantics of use and disclosure for a purpose and a semantics of de-identified data would reduce audit efforts by mechanizing the enforcement of clauses involving purposes and anonymized data. Second, analysis tools for flagging suspicious disclosures would aid human audit.

Acknowledgments

The authors wish to thank Robert J. Simmons for discussions on fixed points, Helen Nissenbaum and Deborah Peel for discussions on HIPAA regulations, and the anonymous reviewers for helpful comments.

This work was partially supported by the U.S. Army Research Office contract on Perpetually Available and Secure Information Systems (DAAD19-02-1-0389) to Carnegie Mellon CyLab, the NSF Science and Technology Center TRUST, the NSF CyberTrust grant "Privacy, Compliance and Information Risk in Complex Organizational Processes," and the AFOSR MURI "Collaborative Policies and Assured Information Sharing". The first author was also supported by an NSF Graduate Research Fellowship.

References

[1] R. Alur and T. A. Henzinger. A really temporal logic. *Journal of the ACM*, 41(1):181–203, 1994.

[2] M. Backes, B. Pfitzmann, and M. Schunter. A toolkit for managing enterprise privacy policies. In *European Symposium on Research in Computer Security*, LNCS 2808, pages 101–119, 2003.

[3] A. Barth, A. Datta, J. C. Mitchell, and H. Nissenbaum. Privacy and contextual integrity: Framework and applications. In *Proceedings of the 27th IEEE Symposium on Security and Privacy*, pages 184–198, 2006.

[4] A. Barth, A. Datta, J. C. Mitchell, and S. Sundaram. Privacy and utility in business processes. In *Proceedings of the 20th IEEE Computer Security Foundations Symposium*, pages 279–294, 2007.

[5] D. Basin, F. Klaedtke, and S. Müller. Monitoring security policies with metric first-order temporal logic. In *Proceedings of the 15th ACM Symposium on Access Control Models and Technologies*, pages 23–34, 2010.

[6] J. Bradfield and C. Stirling. Local model checking for infinite state spaces. *Theoretical Computer Science*, 96(1):157–174, 1992.

[7] J. Bradfield and C. Stirling. *The Handbook of Modal Logic*, chapter Modal Mu-Calculi, pages 721–756. 2006.

[8] T. Breaux and A. Antón. Analyzing regulatory rules for privacy and security requirements. *IEEE Transactions on Software Engineering*, 34(1):5–20, 2008.

[9] L. F. Cranor. *Web Privacy with P3P*. O'Reilly and Associates, Inc., 2002.

[10] Deloitte & Touche and the Ponemon Institute. Enterprise@Risk: 2007 Privacy and Data Protection Survey. White Paper, December 2007.

[11] H. DeYoung, D. Garg, D. Kaynar, and A. Datta. Logical specification of the GLBA and HIPAA privacy laws. Technical Report CMU-CyLab-10-007, Carnegie Mellon University, 2010.

[12] N. Dinesh, A. K. Joshi, I. Lee, and O. Sokolsky. Reasoning about conditions and exceptions to laws in regulatory conformance checking. In *Proceedings of the Ninth International Conference on Deontic Logic in Computer Science*, pages 110–124, 2008.

[13] L.-Å. Fredlund, D. Gurov, T. Noll, M. Dam, T. Arts, and G. Chugunov. A verification tool for ERLANG. *International Journal of Software Tools for Technology Transfer*, 4(4):405–420, 2003.

[14] M. Hilty, D. A. Basin, and A. Pretschner. On obligations. In *Proceedings of the 10th European Symposium on Research in Computer Security*, pages 98–117, 2005.

[15] S. Jajodia, P. Samarati, M. L. Sapino, and V. S. Subrahmanian. Flexible support for multiple access control policies. *ACM Transactions on Database Systems*, 26(2):214–260, 2001.

[16] P. E. Lam, J. C. Mitchell, and S. Sundaram. A formalization of HIPAA for a medical messaging system. In *Proceedings of the 6th International Conference on Trust, Privacy, and Security in Digital Business*, LNCS 5695, pages 73–85, 2009.

[17] N. Li, J. C. Mitchell, and W. H. Winsborough. Design of a role-based trust management framework. In *Proceedings of the 2002 IEEE Symposium on Security and Privacy*, pages 114–130, 2002.

[18] Z. Manna and A. Pnueli. *Temporal Verification of Reactive Systems: Safety*. Springer-Verlag, 1995.

[19] M. J. May, C. A. Gunter, and I. Lee. Privacy APIs: Access control techniques to analyze and verify legal privacy policies. In *Proceedings of the IEEE Workshop on Computer Security Foundations*, pages 85–97, 2006.

[20] H. Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79(1):119–158, 2004.

[21] OASIS XACML Committee. Extensible access control markup language (XACML) v2.0, 2004. Available at http://www.oasis-open.org/specs/#xacmlv2.0.

[22] T. Räsch. *Automata, Logics, and Infinite Games*, chapter Introduction to Guarded Logics, pages 321–342. LNCS 2500. Springer-Verlag, 2002.

[23] US Congress. Gramm-Leach-Bliley Act, Financial Privacy Rule. 15 USC §6801–§6809, November 1999. Available at http://www.law.cornell.edu/uscode/usc_sup_01_15_10_94_20_I.html.

[24] US Congress. Health Insurance Portability and Accountability Act of 1996, Privacy Rule. 45 CFR 164, August 2002. Available at http://www.access.gpo.gov/nara/cfr/waisidx_07/45cfr164_07.html.

[25] A. F. Westin. *Privacy and Freedom*. Atheneum, 1967.