A Logic of Privacy

Steve Barker, Valerio Genovese

A Logic of Privacy

Steve Barker[1] and Valerio Genovese[2,3]

[1] Dept Computer Science, King's College London
[2] Dept Computer Science University of Luxembourg
[3] Dipartimento di Informatica University of Torino

Abstract. We consider the problem of developing an abstract meta-model of access control in terms of which policies for protecting a principal's private information may be specified. Our concern is with developing the formal foundations of our conceptual model. For both the specific access control models and privacy policies, which may be defined in terms of the meta-model, we adopt a combining approach: we combine access control concepts to form the meta-model and we use a fibred logic for the formal foundations. Our approach enables data subjects to specify flexibly what access controls they wish to apply on their personal data and it provides a formal foundation for policies that are defined in terms of the meta-model.

1 Introduction

For several emerging applications there is a requirement that individual entities be able to choose flexibly what of their data should be accessible to whom, for what purpose, and in what circumstances. As a simple example, an entity may wish to control the release of its history of purchasing, that is held by e-traders, to telemarketers of an e-trader's choosing. The idea of entities having more control over the release of their information has been recognized, in general terms, by Westin [1] and for particular technologies [2] and applications [3].

An important research question applies: how can entities be provided with a formally well-founded framework for defining flexibly the privacy policies that they wish to apply on access to *their* data?

The problem of helping to preserve the privacy of an entity's personal data has recently received attention (see, for example, the work on P3P [4], EPAL [5], Hippocratic databases [6], and XACML [7]). Moreover, researchers in the access control community have proposed various "privacy-aware" access control models (see, for example, the work on P-RBAC [8, 9]). Although these approaches allow data subjects to express some controls on access to their personal information, we argue that, to differing extents, they do not provide sufficient expressive power for individually tailored privacy policies, they fail to accommodate adequately some important concepts (e.g., trust and delegation), and, they lack adequate formal foundations. The work by Barth et al. [10] avoids some of the shortcomings of existing proposals on privacy management. Specifically, Barth et al. adopt a well-defined conceptual basis (*contextual integrity*) and they develop a sound formal basis (from temporal logic and the *Logic of Privacy and*

Utility (LPU)) in terms of which a range of privacy policies may be grounded. We adopt a similar methodological position to Barth et al., but we propose a different conceptual base and different logical foundations, which we will argue have certain attractions.

The principal contribution described in this paper is the proposal of a methodology for specializing access control models for privacy purposes and the development of a formal language and semantics for privacy policies that is based on fibred logic [11]. Specifically, we propose a fibred logic formulation of our meta-model of access control, $\mathcal{M}$, from [12]; we denote the privacy-enhanced form of $\mathcal{M}$ by $\mathcal{M}^P$ (where P stands for "privacy"). We demonstrate how privacy-enhanced access control policies may be derived from $\mathcal{M}^P$ by specializing and combining the relations and logical axioms that we introduce to define this meta-model. Our main objective is to extend the notion of category, first introduced by us in [12], to allow categories to be defined using arbitrary logical formulas that may be expressed in a variety of logic languages (e.g., first-order logic, intuitionistic logic, ... and even SQL). For this, we need to develop formal foundations that are quite different to those described in [12]. Specifically, we describe a variant of predicate Fibred Logic and an enhanced form of our Fibred Security Language (FSL) [13], which enables us to use a range of modalities in representations of privacy policies that are derived from our meta-model.

Although we recognize their importance in privacy-enhanced access control, due to space constraints, we will not consider obligations, audit policies, and hierarchies of objects or of purposes. We assume that data is stored and transmitted securely and that sound methods for the authentication of data subjects, controllers and recipients are employed.

The remainder of the discussion is organized in the following way. In Section 2, we describe the basic syntactic notions on which our approach is based. In Section 3, we present details on our privacy logic. In Section 4, we formally describe the core relations and axioms of the meta-model that we specialize for privacy and which can be specialized in multiple ways for different, specific privacy-enhanced access control models and policies. In Section 5, we show how privacy-enhanced access control policies can be represented in terms of our meta-model, by specializing and combining the relations and axioms that the meta-model includes. In Section 6, we discuss related work. In Section 7, conclusions are drawn and further work is suggested.

2 Language Issues and FSL

In this section, we describe the language for formulating the meta-model and specialized instances of it. We only describe the basic syntax and semantic notions (the minimum details to make the paper self-contained).

The key sets of constants in the universe of discourse that we admit are as follows: -A countable set $\mathcal{C}$ of categories, where c_0, c_1, ... are used to denote arbitrary category identifiers. -A countable set $\mathcal{K}_{ds}$ of data subjects and a countable set of K_{du} of data users (requesters for access) where κ_0, κ_1, ... are used

for (key) identification. -A countable set $\mathcal{A}$ of named atomic *actions*, where a_0, a_1, ... are used to denote arbitrary action identifiers. -A countable set $\mathcal{R}$ of *resource identifiers*, where r_0, r_1, ... denote arbitrary resources, $r(t_1, \ldots, t_n)$ is an arbitrary n-place relation and t_i $(1 \leq i \leq n)$ is a term, a function, a constant or a variable. -A countable set $\mathcal{P}$ of *purposes*, where p_0, p_1, ... are used to denote arbitrary purpose identifiers. -A countable set of meta-policy identifiers; for example, c (for closed policies), o (for open policies), do (for a denials override policy), ... -A countable set $\mathcal{T}$ of *time points*, $\tau_0, \tau_1, \ldots$. -A countable set $\mathcal{E}$ of *event identifiers*, $e_0, e_1, \ldots$

A major difference to the work in [12] is to change the notion of category, the most fundamental element in our ontology. In [12], a category is used as a proper name to simply refer to a category of entities; the term category being interpreted as "being synonymous with, for example, a type, a sort, a class, a division, a domain." In this paper, category is not merely a proper name; it is a viewed as a well-formed logical formula (that may be expressed in first-order, modal, intuitionistic, ... terms) that defines the membership of the category. The reader is referred to [12] for a fuller account of our original notion of category and for its comparison with the work we describe in this paper.

The notion of purpose is also key in privacy; data subjects must be able to specify what of their personal data may be stored by a data controller and for what purposes this personal data may be used by requesters of access to the data. We discuss our interpretation of purpose more fully below. Two special time points will be important in our treatment: 0 denotes the start of time and ∞ is an arbitrary maximal future time. We assume that various comparison operators exist on times $\{<, \leq, \geq, >\}$, with their usual interpretation e.g., $t_1 \leq t_2$ iff time point t_1 is earlier than or the same time point as t_2. Although we refer to time points, the approach that we describe enables various temporal frameworks to be accommodated (by combining temporal logics). Times and events allow us to provide a degree of dynamacy in the framework that we develop.

In the formulation of the rules that we will use to represent access control models and policies, variables will appear in the upper case and constants in the lower case. The only exception to this will be when we use (lower-case) x and y to refer specifically to types of categories.

As the access control logic that we propose is intended for use in distributed scenarios, we need to be able to express delegation among principals. Our logic is therefore centered, like the access control logics of [14] and [15], on formulas such as "A **says** s" where A represents a principal, s represents a statement (a request, a delegation of authority, or some other utterance), and **says** is a modality. It is important to note that it is possible to derive that A **says** s even when A does not directly utter s. For example, when the principal A is a user and one of its programs includes s in a message, then we may have A says s, if the program has been delegated by A. In this case, A says s means that A has caused s to be said, that s has been said on A's behalf, or that A supports s.

We assume that such assertions are understood by a reference monitor in charge of making decisions on access to a resource r. The reference monitor

may implement the policy that a particular data requester A is authorized to perform action a on resource r that contains "private data". This policy may be represented by the formula: $(A \text{ says } do\_on(a, r)) \rightarrow do\_on(a, r)$, which expresses that A **controls** $do\_on(a, r)$.[4] Similarly, a request for the operation a on r from a principal B may be represented by the formula: B says $do\_on(a, r)$. The goal of the reference monitor is to prove that these two formulas imply $do\_on(a, r)$, and to grant access if the implication can be demonstrated. While proving $do\_on(a, r)$, the reference monitor does not need to prove that the principal B controls $do\_on(a, r)$. Rather, it may exploit relations between A and B and certain other facts. For example, the reference monitor may know that B has been delegated by A, and, thus, that B speaks for A as concerns $do\_on(a, r)$:

$$(B \text{ says } do\_on(a, r)) \rightarrow (A \text{ says } do\_on(a, r))$$

3 An Axiomatization of Privacy

Having introduced the basic language details in the previous section, we now describe the details of the logic language that we use for representing privacy-enhanced access control models and policies.

Our logic is based on a variant of the work described in [16] and extends FSL by adding a privacy context modality, where $[p]\varphi$ has the reading:

"φ holds under the purpose p".

P-FSL formulas are expressed in the following way:

Definition 1 (P-FSL).

$$\varphi ::= F(x_1, \ldots, x_n) \mid \neg\varphi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \varphi \rightarrow \varphi \mid \varphi(x) \textit{ says } \varphi \mid [p]\varphi \mid \rho(\varphi(x), \psi(y))$$

The expression, $\varphi(x)$ says ψ should be read as: "The group composed of all of the principals that satisfy $\varphi(x)$ *supports* the assertion *assert* ψ". In this view, the ρ operator describes a general relationship between groups. In line with [17], we write $\varphi(x)$ **controls** ψ as shorthand for $(\varphi(x) \text{ says } \psi) \rightarrow \psi$.

Definition 2 (Axiomatization). *The axiomatization consists of all axioms of intuitionistic propositional logic plus axioms and rules for the* `says` *modality.*

All axioms and rules of First-Order Logic	*(FOL)*
If $\vdash \psi$ *then* $\vdash \varphi(x)$ *says* ψ	*(N)*
$\vdash \varphi(x)$ *says* $(\psi \rightarrow \psi') \rightarrow$ $(\varphi(x)$ *says* $\psi \rightarrow \varphi(x)$ *says* $\psi')$	*(K)*
$\vdash \forall x(\varphi_1(x) \leftrightarrow \varphi_2(x)) \rightarrow$ $(\varphi_1(x)$ *says* $\psi \leftrightarrow \varphi_2(x)$ *says* $\psi)$	*(Ex)*
$\vdash (\psi \rightarrow \psi') \rightarrow \varphi(x)$ *says* $\psi \rightarrow \varphi(x)$ *says* ψ'	*(Md)*
If $\vdash \psi$ *then* $\vdash [p]\psi$	(N^p)
$\vdash [p]\varphi \rightarrow [q][p]\varphi$	$(4^{p,q})$
$\vdash [p](\psi \rightarrow \psi') \rightarrow [p]\psi \rightarrow [p]\psi'$	(K^p)
$\vdash \forall x(\varphi_1(x) \leftrightarrow \varphi_2(x)) \rightarrow \rho(\varphi_1(x), \varphi_2(x))$	(ρ)

[4] In this view. with A **controls** ψ we express that A has a direct permission to do ψ.

Notice that the (ρ) axiom states that the relation ρ is reflexive, i.e., if two formulas φ_1 and φ_2 describe the same group then they are in relation w.r.t. ρ.

Definition 3. *A first-order P-FSL constant domain model is a tuple* $\mathcal{M} = \langle W, N, D, \theta, I\rangle$ *where:*

- D *is a non-empty set, called the* ***domain***[5].
- W *is a set of states.*
- $N : W \times \mathcal{P}(D) \to \mathcal{P}(\mathcal{P}(W))$ *is a neighborhood function that given a state* s *and a set of principals* T*, it associates a family of sets of states (called neighborhoods). The intuition is that at each state* $N(w, T)$ *is the set of propositions, (i.e. the set of states), that are supported by the group of principals* T*.*
- $\theta : P \to \mathcal{P}(W)$ *is a mapping from purposes to a subset of states. We say that that* $P' \in \theta(p)$ *is the set of states that must be considered in the context of purpose* p*.*
- I *is a classical first-order interpretation function where for each* n*-ary predicate symbol* F *and each state* w*,* $I(F, w) \subseteq D^n$*.*

We require the neighborhood function N *to satisfy the following properties:*

(a) *If* $X \in N(w, T)$ *and* $X \subseteq Y$*, then* $Y \in N(w, T)$
(b) *If* $X \in N(w, T)$ *and* $Y \in N(w, T)$*, then* $X \cap Y \in N(w, T)$
(c) $W \in N(w, T)$

Intuitively, given a set of principals $T \subseteq D$, $N(w, T)$ is the set of propositions (i.e., the set of states), that T supports at state w. As shown in [18], conditions (a), (b) and (c) ensure that `says` is a normal modality (i.e., validates axiom K).

The satisfaction relation $\models$ is inductively defined in terms of an *interpretation* $\mathcal{M}, w$ and a valuation σ, which assigns objects to individual variables.

- $\mathcal{M}, w \models_\sigma F(x_1, \ldots, x_n)$ iff $\langle \sigma(x_1), \ldots, \sigma(x_n)\rangle \in I(F, w)$ for each n-place predicate symbol.
- $\mathcal{M}, w \models_\sigma \neg\varphi$ iff $\mathcal{M}, w \not\models_\sigma \varphi$
- $\mathcal{M}, w \models_\sigma \varphi \vee \psi$ iff $\mathcal{M}, w \models_\sigma \varphi$ or $\mathcal{M}, w \models_\sigma \psi$.
- $\mathcal{M}, w \models_\sigma \varphi \wedge \psi$ iff $\mathcal{M}, w \models_\sigma \varphi$ and $\mathcal{M}, w \models_\sigma \psi$
- $\mathcal{M}, w \models_\sigma \varphi \to \psi$ iff $\mathcal{M}, w \models_\sigma \varphi$ implies $\mathcal{M}, w \models_\sigma \psi$
- $\mathcal{M}, w \models_\sigma \forall x\varphi$ iff for every element $d \in D$ we have $\mathcal{M}, w \models_{\sigma[d/x]} \varphi$
- $\mathcal{M}, w \models_\sigma \varphi(x)$ says ψ iff $(\varphi)^{\mathcal{M},\sigma} \in N(w, U)$, where $U = \{d \in D \mid \mathcal{M}, w \models_{\sigma[d/x]} \varphi(x)\}$
- $\mathcal{M}, w \models_\sigma [p]\varphi$ iff for all $t \in \Theta(p)$, $\mathcal{M}, t \models \varphi$
- $\mathcal{M}, w \models \rho(\varphi(x), \psi(y))$ iff $\langle U, U'\rangle \in I(\rho, w)$, where $U = \{d \in D \mid \mathcal{M}, w \models_{\sigma[d/x]} \varphi(x)\}$ and $U' = \{d \in D \mid \mathcal{M}, w \models_{\sigma[d/y]} \psi(y)\}$

[5] For the sake of readability we identify the domain as the set of all principals, if needed P-FSL can be easily extended to cope with different sorts (e.g., principals, time points, purposes, ...).

When it is clear from the context, we will omit σ as index of $\models$. A formula φ is true in a model $\mathcal{M}$ ($\mathcal{M} \models \varphi$) if, for every state w, $\mathcal{M}, w \models \varphi$. A formula is valid ($\models \varphi$) if it is true in all models. A formula φ is a logical consequence of a set of formulae $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ ($\Gamma \models \varphi$), if for every $\mathcal{M}, w$ $\mathcal{M}, w \models \bigwedge_{1 \leq i \leq n} \gamma_i$ implies $\mathcal{M}, w \models \varphi$.

Theorem 1 (Soundness of P-FSL Axiomatization). *Every theorem deducible from the axiomatic proof system of Definition 2 is valid with respect to the semantics.*

$$\textit{If } \vdash \varphi \textit{ then } \models \varphi$$

Proof. By cases on axioms and rules of P-FSL.

It should be noted that P-FSL is, in its full generality, undecidable because it is a extension of first-order modal logic (albeit a conservative one). P-FSL must be understood as providing a general formal framework for studying abstract access control models and, in our case, privacy-enhanced access control models. The undecidability of the logic stems from its high expressive power. However, in relation to access control in practice, many of the features that make first-order logic undecidable are not necessary, like infinite domains, unlimited quantification or an unlimited number of free variables in formulas. As we will see, the expressivity needed to interpret the relations of the abstract access control model $\mathcal{M}^P$ is a restriction of first-order modal logic in which we do not have explicit quantification and formulas have at most one free variable. In [19], the above mentioned restriction is shown to be decidable. More generally, the language of first-order modal logic with two variables (without any restriction on quantification) is decidable with polynomial time complexity with respect to satisfiability.

4 The Model $\mathcal{M}^P$

In the previous section, we established the basic language and axiomatic details. We now consider the specific details that are required for our general meta-model for privacy-enhanced access control and its representation in P-FSL. We wish to accommodate data subjects, data controllers, denials of access, an interpretation of purpose, contextual accessibility criteria and the flexible specification of permitted recipients of a data subject's personal data. For that, the following core relations are included in our meta-model, $\mathcal{M}^P$: -$\mathcal{PCA}$, a 4-ary relation, $\mathcal{K}_{ds} \times \mathcal{K}_{du} \times \mathcal{C} \times \mathcal{P}$. -$\mathcal{ARCA}$, a 5-ary relation, $\mathcal{K}_{ds} \times \mathcal{A} \times \mathcal{R} \times \mathcal{C} \times \mathcal{P}$. -$\mathcal{ARCD}$, a 5-ary relation, $\mathcal{K}_{ds} \times \mathcal{A} \times \mathcal{R} \times \mathcal{C} \times \mathcal{P}$. -$\mathcal{PAR}$, a 3-ary relation, $\mathcal{K}_{du} \times \mathcal{A} \times \mathcal{R}$. -$\mathcal{PRM}$, a 3-ary relation, $\mathcal{K}_{ds} \times \mathcal{R} \times \mathcal{MP}$.

The (informal) semantics of the n-ary tuples in $\mathcal{PCA}$, $\mathcal{ARCA}$, $\mathcal{ARCD}$, $\mathcal{PAR}$, and $\mathcal{PRM}$ are, respectively, defined thus: $(\kappa_{ds}, \kappa_{du}, c, p) \in \mathcal{PCA}$ iff a data user $\kappa_{du} \in \mathcal{K}_{du}$ is assigned to the category $c \in \mathcal{C}$ for the purpose $p \in \mathcal{P}$ according to the data subject κ_{ds}. $(\kappa_{ds}, a, r, c, p) \in \mathcal{ARCA}$ iff the permission (a, r) is assigned

to the category $c \in \mathcal{C}$ for the purpose $p \in \mathcal{P}$ according to the data subject κ_{ds}. $(\kappa_{ds}, a, r, c, p) \in \mathcal{ARCD}$ iff a the permission (a, r) is denied to the category $c \in \mathcal{C}$ for the purpose $p \in \mathcal{P}$ according to the data subject κ_{ds}. $(\kappa_{du}, a, r) \in \mathcal{PAR}$ iff a data user $\kappa_{du} \in \mathcal{K}_{du}$ is authorized to perform the action $a \in \mathcal{A}$ on the resource $r \in \mathcal{R}$. $(\kappa_{ds}, r, m) \in \mathcal{PRM}$ iff the data subject κ_{ds} "controls" access to the resource $r \in \mathcal{R}$ and κ_{ds} asserts that the meta-policy $m \in \mathcal{MP}$ applies to access on the resource r.

The semantics of the *pca*, *arca*, *arcd* and *prm* relations can be formally defined in P-FSL via the "says" operator [17] (where $\leftrightarrow$ is "is equivalent to").

- $\models k_{ds}$ says $([p]\varphi_c(k_{du})) \leftrightarrow (k_{ds}, k_{du}, c, p) \in \mathcal{PCA}$
- $\models k_{ds}$ says $([p](\varphi_c(x)$ **controls** $do\_on(a, r))) \leftrightarrow (k_{ds}, a, r, c, p) \in \mathcal{ARCA}$
- $\models k_{ds}$ says $([p](\neg(\varphi_c(x)$ **controls** $do\_on(a, r)))) \leftrightarrow (k_{ds}, a, r, c, p) \in \mathcal{ARCD}$
- $\models k_{du}$ **controls** $do\_on(a, r) \leftrightarrow (k_{du}, a, r) \in \mathcal{PAR}$
- $\models prm(k_{ds}, r, m) \leftrightarrow (k_{ds}, r, m) \in \mathcal{PRM}$

Here, $\varphi_c(x)$ is intended to be the P-FSL formula with one free variable x that maps category c in the meta-model $\mathcal{M}^P$. In the above mapping, most of the relations of $\mathcal{M}^P$ (with the exception of *prm*) are interpreted over *says* and *controls* operators. In relation to [12], we extend the notion of categories from constants to first-order formulas that identify a collection of principals.[6] Notice also that, by viewing purposes as modal contexts, we can map entire formulas under a specific purpose because in the wff $[p]\varphi$, φ can be *anything*, not just a relation. For instance, delegation may be expressed under a specific purpose $[p](bob$ says $\psi \rightarrow admin$ says $\psi)$.

In what follows, the reader is reminded that variables in rules appear in the upper case and are implicitly universally quantified; constants are in the lower case.

The elements in the set $\mathcal{PAR}$ are defined in terms of $\mathcal{PRM}$, $\mathcal{PCA}$, and a specification of a particular meta-policy m, which itself is defined with respect to $\mathcal{ARCA}$ or $\mathcal{ARCD}$. In P-FSL, the rules defining *par* for different meta-policies (closed (c), open (o), and denials-override (do)) are:

$prm(\kappa_{ds}, R, c) \ \wedge \ (\kappa_{ds}$ says $[P]C(\kappa_{du})) \wedge$
$(\kappa_{ds}$ says $[P]C(x)$ **controls** $do\_on(A, R)) \rightarrow \kappa_{du}$ **controls** $do\_on(A, R)$.
$prm(\kappa_{ds}, R, o) \ \wedge \ (\kappa_{ds}$ says $[P]C(\kappa_{du})) \wedge$
$\neg(\kappa_{ds}$ says $\neg[P](C(x)$ **controls** $do\_on(A, R))) \ \rightarrow \kappa_{du}$ **controls** $do\_on(a, r)$.
$prm(\kappa_{ds}, R, do) \ \wedge (\kappa_{ds}$ says $[P]C(\kappa_{ds})$ **controls** $do\_on(A, R)) \wedge$
$\neg(\kappa_{ds}$ says $\neg[P](C(x)$ **controls** $do\_on(A, R))) \rightarrow \kappa_{du}$ **controls** $do\_on(a, r)$.

The above rules should be read as *axiom schemas* that hold for every formula φ representing a category. In this view the upper case in $C(x)$ stands for a second-order quantification (i.e., over formulas).

[6] As shown in [13], by viewing categories as types we can generalize roles (as in RT) as special instances of categories.

For representing hierarchies of categories in our meta-model, the following definition is included as part of the axiomatization of the model (where '_' denotes an anonymous variable and *dc* is a "directly contains" relation cf. [12]):

$$dc(C, \_) \rightarrow \rho(C, C).$$
$$dc(\_, C) \rightarrow \rho(C, C).$$
$$dc(C', C'') \rightarrow \rho(C', C'').$$
$$dc(C', C''') \wedge \rho(C''', C'') \rightarrow \rho(C', C'').$$

Authorizations may be defined, quite generally, thus (the meta-policy here being closed, as denoted by c where c is short for "closed" policy):

$$prm(K_{ds}, R, c) \wedge K_{ds} \text{ says } C(K_{du}) \wedge \rho(C, C') \wedge$$
$$K_{ds} \text{ says } [C(x) \textbf{ controls } do\_on(A, R)] \rightarrow K_{du} \text{ says } do\_on(A, R)$$

That is, a data user (requester) K_{du} has A access on resource R if a data subject K_{ds}, which controls access to personal data, says that K_{du} is assigned to a category C that inherits the A privilege on R, to which a closed meta-policy on access applies, from a category C' such that $\rho(C, C')$ holds i.e., C is "senior to" C' in a partial ordering of categories.

The careful reader will have noted that what we are defining is a general logic for a family of privacy-enhanced access control models that may be derived from $\mathcal{M}^P$. The meta-model $\mathcal{M}^P$ may be specialized in multiple ways by, for instance, a policy author admitting different or additional sorts (e.g., times) in the relations from our core set, to allow for specific requirements to be met. On this point, it is important to note that, for our definition of $\mathcal{PAR}$, existential quantification on purposes is important; rather than having a purpose sort as part of the definition of authorization, as in the case of purpose-based access control as that term is interpreted in [20], we treat purpose existentially. On this interpretation, purpose specifications are relevant only in terms of the relationship between a data subject and a data controller: the data subject decides what of its data may be released by the data controller for what purpose. A requester K_{du} has A access on R if for *some* purpose K_{du} has A access on R as a consequence of there being a requester-category assignment and a permission-category assignment that implies that this authorization should hold. Of course, a policy author may instead require a data requester to state explicitly the purpose for the access (cf. the notion of intended purpose from [20]). In that case, a purpose parameter may be added to the *par* relation. The explicit specification of purpose implicitly eliminates the existential quantification that the 3-place form of *par* assumes. The different options available to the policy author reflect the different positions the policy author may adopt on, for example, the interpretation of purpose (e.g., whether the purpose of a request is an intention in the mind of the requester that need not be made explicit), what epistemic commitments are required of the requester (e.g., are requester's required to know for what specific purposes a specific data subject has allowed a specific action to be performed on a specific resource that they control access to), etc. Our formulation is based on what we perceive to be a minimal collection of useful relations where by minimal we mean

minimal in terms of the arity of relations as well as their number. As previously stated, a policy author is expected to specialize the meta-model as required. It must also be noted that a data subject may also be free to decide what specific variant of *par*, and other core relations of $\mathcal{M}^P$, are to be used to access their data. Compelling data subjects to use a particular form of *par* runs counter to our intention of allowing data subjects to define the controls applicable to their data and compelling policy authors to use a particular interpretation of $\mathcal{M}^P$ would be counter to the methodological position that we have argued for.

Constraints on categories may also be flexibly specified in terms of the core predicates of our meta-model and are expressed in P-FSL as statements of the following general form (where $\bot$ read as "is inconsistent" and $c \in \mathcal{C}$ and $c' \in \mathcal{C}$ are constants that denote specific categories): $\varphi_c(P) \wedge \varphi_{c'}(P) \rightarrow \bot$. For example, the constraint

$$K_{ds} \text{ says } [P]\varphi_c(K_{du}) \ \wedge \ K_{ds} \text{ says } [P']\varphi_c(K_{du}) \ \wedge \ P \neq P' \rightarrow \bot.$$
$$K_{ds} \text{ says } [P](\varphi_c(x) \ \textbf{controls} \ do\_on(write, r)) \ \wedge$$
$$K_{ds} \text{ says } [P](\varphi_c(x) \ \textbf{controls} \ do\_on(write, r')) \rightarrow \bot.$$

represents that exactly one data user K_{du} may be assigned by a data subject K_{ds} to a category c for a specific purpose (a "separation of categories" constraint) and that *write* privilege on the pair of resources (r, r') is impossible for all categories of data subjects and for all purposes (a "separation of privileges" constraint).

Particular privacy-enhanced access control models can be (and are expected to be) defined within the general axiomatic framework that we have described by specializing predicates and axioms. For example, to accommodate purpose with subject-specified access controls in status-based access control [21], the axioms of $\mathcal{M}^P$ may be simply specialized thus (with the above definition of ρ assumed, with E denoting an event, with C in this case being a category that combines ascribed and action statuses, and with definitions of *pca_init* and *pca_term* omitted):

$$C(P) \wedge \rho(C(x), C'(y)) \ \wedge C'(y) \ \textbf{controls} \ do\_on(A, R) \rightarrow P \ \textbf{controls} \ do\_on(A, R).$$
$$current\_time(T) \wedge happens(E, T_s) \ \wedge agent(E, P) \ \wedge \ act(E, A) \wedge T_s < T \ \wedge$$
$$pca\_init(E, P, A, C, T_s, T) \ \wedge \ \neg ended\_pca(P, C, T_s, T) \rightarrow C(P).$$
$$happens(E', T') \wedge agent(E', P) \ \wedge \ act(E', A') \ \wedge \ T_s < T' \wedge T' \leq T \ \wedge$$
$$pca\_term(E', P, A', C, T_s, T) \ \rightarrow ended\_pca(P, C, T_s, T).$$

5 Privacy Policies in $\mathcal{M}^P$ by P-FSL

In the previous section, we gave an axiomatization of a general class of "privacy enhanced" access control models. In this section, we consider the representation of privacy-enhanced access control policies by specialization and combination of the core relations and axioms of $\mathcal{M}^P$, which can also be multiply interpreted.

We first introduce an additional technical component: annotated rules. An annotated rule φ, which is used by a data controller in the *specification* of a policy, may be annotated with Δ to represent that a data subject is permitted

by the controller to delete or modify φ; the annotation $\neg\Delta$ is used to specify that φ cannot be changed by a data subject in an access policy. In the latter case, a data subject κ_{ds} is still free to insert rules of κ_{ds}'s choosing, but, not surprisingly, only for data that refers to κ_{ds}. Annotations of elements other than rules (e.g., terms) are possible but we omit the details on this. It is also important to note that, as we are concerned about access controls on a data subject's personal data, we assume that the information resource to be accessed by data requesters will contain a personal identifier of a data subject to which the data refers.

The first example that we give is of privacy policy formulation in P-FSL that relates to medical informatics scenario.

Example 1. Consider the following policy of the Virginia Hospital Center (VHC) on the confidentiality of patient data:

> *For the purposes of operating on a patient, the patient's full medical history, which includes the patient's identifier, name, date-of-birth, and history of illnesses, can be seen by any member of the category surgeon (sur) for the purpose of operating (op). The patient's identifier, name, date of birth and diagnosed illnesses in the past six months may be disclosed to the category of non-surgical staff (nss) for the purpose of providing diagnostic support (ds). The pca definitions used by VHC are defined non-locally at υ_1. A closed access control policy is to apply to the release of all data. The access control policy as it relates to data subjects generally is maintained by the VHC administrator denoted by κ_c.*

Suppose that the databases used by VHC include an 8-place relation *pat* (where *pat* is short for patient) that is defined at υ_2 and includes details of the patient's identifier, the patient's name, date of birth, illness, room number (at the hospital), contact number (at the hospital), time of admittance and time of discharge:

$$pat(Id, Name, DoB, Illness, Rm, Pno, Admit, Discharge).$$

To represent their requirements, VHC's policy on the release of patient information may be represented as a privacy policy, which is simply derived from the $\mathcal{M}^P$ model, thus (where *sct* is short for system clock time):

$$\neg\Delta : \upsilon_1 \text{ says } (\kappa_c \text{ says } [P]C(K_{du}) \rightarrow \kappa_c \text{ says } [P]C(K_{du})).$$

$$\neg\Delta : \upsilon_2 \text{ says } pat(K_{ds}, V, W, X, Y, Z, T1, T2) \ \wedge \ sct(T) \ \wedge \ T1 \geq 0 \ \wedge \ T \leq \infty \rightarrow \kappa_c \text{ says } [op](sur(x) \textbf{ controls } do\_on(read, pat(K_{ds}, V, W, X, Y, Z, T1, T2))).$$

$$\neg\Delta : \upsilon_2 \text{ says } pat(K_{ds}, V, W, X, Y, Z, T1, T2) \ \wedge \ sct(T) \ \wedge \ month(T, M) \wedge month(T1, M1) \ \wedge \ M1 \geq M - 6 \rightarrow \kappa_c \text{ says } [ds](nss(x) \textbf{ controls } do\_on(read, pat(K_{ds}, V, W, X, \_, \_, \_, \_))).$$

$$\neg\Delta : prm(K_{ds}, R, c) \ \wedge \ K_{ds} \text{ says } [P]C(K_{du}) \ \wedge \ K_{ds} \text{ says } [P](C(x) \textbf{ controls } do\_on(read, R)) \rightarrow K_{du} \text{ controls } do\_on(read, R).$$

From the example above, it should be noted that κ_c is the controller of VHC's privacy policy. If any data subject were to have the freedom to change VHC's

policy then the data subject could deny access to data users that need to have information on the data subject in order to perform an action of benefit to the data subject (e.g., diagnosing a patient's illness). Nevertheless, the data subject does have the freedom to add to VHC's privacy policy specification in order to represent personal requirements on the release of their data. The next example demonstrates this.

Example 2. Consider the wishes of the individual patient κ_β in relation to VHC's policy on the disclosure of patient information:

> *I agree to the hospital's policy on the release of my personal information for the purpose of operating. However, I also wish some of this information to be accessible to the category of data users that I call* $family$*. Specifically, the category* $family$ *is defined by me (non-locally) at* υ_3 *and I want members of* $family$ *to be able to access (and only access) my name, bedside phone number, and room number for the purpose of contacting me while I am in hospital (a purpose that I denote by* ct*, as shorthand for contact).*

To capture κ_β's individual access control requirements, κ_β adds the following definitions:

$$prm(\kappa_\beta, pat(\kappa_\beta, V, W, X, Y, Z, T1, T2), c).$$
$$\upsilon_2 \text{ says } pat(\kappa_\beta, V, W, X, Y, Z, T1, T2) \rightarrow$$
$$\kappa_\beta \text{ says } [ct] family(x) \textbf{ controls } do\_on(read, pat(\kappa_\beta, V, \_, \_, Y, Z, \_, \_)).$$

κ_β then adds the following *pca* definition to VHC's policy to express his required access controls applicable to his family contacts (where f_mbr is short for "family member"):

$$\upsilon_3 \text{ says } (f\_mbr(\kappa_\beta, K_{du}) \rightarrow \kappa_\beta \text{ says } [ct] family(K_{du}))$$

Consider next an example of our approach for privacy policy formulation in the context of an e-commerce scenario.

Example 3. Suppose that ACo are an on-line trading company that specify the following policy on the confidentiality of customer transaction data that they hold:

> *Our preferred policy is to store a complete history of each customer's purchase transactions (the items bought, the number bought and when); we retain this information indefinitely and make it available at all times to suppliers of our choosing for the purpose of future marketing (*f_mkt*). Any company that we call a supplier is assigned to the category that we call sup. We assign suppliers, for the purpose* f_mkt*, from the time at which the supplier is first approved by us. A closed meta-policy is to apply on all forms of data release by default.*

The databases that are used by *ACo* include a 3-place relation *sp* (short for suppliers), and a history of customer transactions is recorded in a 4-place relation, *tr* (short for transactions):

$$sp(SupId, Name, From).\ tr(CustId, Item, Number, Purchase\_Time).$$

We assume that the definitions of predicates in *sp* and *tr* are, respectively, found at υ_6, and υ_7. The *pca*, *arca* and *prm* definitions are assumed to be stored locally.

To define their access policy on the release of a customer's personal data, *ACo* can express their requirements in P-FSL thus:

$\Delta : \upsilon_6$ says $sp(K_{du}, N, T1) \ \wedge \ sct(T) \ \wedge \ T1 \leq T \rightarrow (\kappa_c$ says $[f\_mkt]sup(K_{du}))$.
$\Delta : \upsilon_7$ says $tr(K_{ds}, X, N, Z) \ \wedge \ sct(T) \ \wedge \ T \geq 0 \ \wedge \ T \leq \infty \rightarrow$
κ_c says $[f\_mkt](sup(x)$ **controls** $do\_on(read, tr(K_{ds}, Y, N, Z)))$.
$\Delta : subject(K_{ds}) \ \wedge \ K_{ds} \neq K_{du} \ \wedge \ \neg prm(K_{ds}, tr(K_{ds}, Y, N, Z), c) \rightarrow$
$prm(\kappa_c, tr(K_{ds}, Y, N, Z), c)$.
$\Delta : prm(K_{ds}, R, c) \ \wedge \ K_{ds}$ *says* $[P]C(K_{du}) \ \wedge$
K_{ds} says $[P](C(x)$ **controls** $do\_on(read, R)) \rightarrow K_{du}$ **controls** $do\_on(read, R)$.

Next, suppose that κ_ϕ is a customer with *ACo* and prefers to define its own access controls on its transaction history. On that, suppose that the following access policy issues arise for κ_ϕ on the confidentiality and use of its data that is held by *ACo*:

> *I will allow my purchase history to be accessed but only by your suppliers that I have recorded as having a status of pr for "premium". My purchase history can only be released to suppliers of yours that satisfy my principals that I categorize as pr, I will only allow access to my transaction data as it relates to the purchase of nuts and the number of nuts bought by me (as I am only interested in nut-related purchases and data users may want to know if I am a "major purchaser"). Moreover, I do not want any release of my transaction data to any supplier for f_mkt purposes if my stock level of nuts, as recorded in stock(item, quantity) at υ_{50}, is greater than 100 units and I will only release my transaction history since 2009/01/01 and only until 2010/03/31 (after which time I will not be making any nut-related purchases so there is no reason for my data to be accessible to any external recipients after this time).*

Assuming that the binary relation *status* is stored at υ_8 (and is used to map users to statuses, like *pr*), to represent the requirements, κ_ϕ's specialization of *ACo*'s privacy-enhanced policy can be represented thus:

κ_c says $[f\_mkt]sup(K_{du}) \ \wedge \ \upsilon_8$ says $st(K_{du}, pr) \rightarrow \kappa_\phi$ says $[f\_mkt]sup(K_{du})$.
υ_7 says $tr(\kappa_\phi, nut, N, T) \ \wedge \ T \geq 20090101 \ \wedge \ T \leq 20100331 \ \wedge$
υ_{50} says $stock(nut, Q) \ \wedge \ Q > 100 \rightarrow$
κ_ϕ says $[f\_mkt](sup(x)$ **controls** $do\_on(read, tr(\kappa_\phi, nut, N, \_)))$.
$prm(\kappa_\phi, tr(\kappa_\phi, Y, N, Z), c)$.

It should be noted from the example above that temporal accessibility constraints and the conditions on access that are defined in terms of notions like stock levels allow for dynamic privacy-enhanced policies to be formulated by κ_ϕ on the release and use of its personal data. Hence, privacy-enhanced policies can change automatically in response to events and without requiring explicit policy modification. Moreover, κ_ϕ freely specifies the sources of access control information *of its choosing* to define allowed forms of access to *its data* (cf. the use of $st/2$ for status).

6 Related Work

The work that we have discussed in this paper is related to that described in [12]. In [12], a formalization of category-based access control is given in terms of identification-based logic programs, which extend the expressive power of the logic programs used in the Flexible Authorization Framework [22] and conceptual notions (e.g., by introducing the notion of category as a generalization of "role" and allowing for distributed trust management). The logic language that we describe in this paper allows for categories that may be defined by formulas in multiple logic languages (including logic programming languages). Our approach differs from [12] in terms of its focus on privacy enhancement in meta-model terms and to both [22] and [12] in that we adopt a combining logic approach and a richer combining model/policy approach.

Issues in privacy policy management have been addressed in the work on P3P [4], EPAL [5], and Hippocratic databases [6]. However, each of these approaches is a particular approach. In contrast, we derive particular cases from the generality of our approach (as we showed by demonstrating how a range of instances of $\mathcal{M}^P$ may be developed as models or policies that can be formulated in P-FSL). P-RBAC [8] also has the attraction of combining access control and privacy as we do. Nevertheless, it is our view that enhancing a *particular* form of access control model for personal data protection, RBAC in the case of P-RBAC, introduces a problem that is common in existing work: the problem of unduly constraining the control that individual data subjects have for managing access to their data. Even though the notion of "role" can be given a quite general interpretation, "role" remains a particular instance of the more general notion of category [12] and category, being more general than "role", offers greater flexibility to data subjects defining access controls on their data. Similarly, although Fischer-Hubner's task-based privacy-oriented access control model [23] is a useful contribution to the literature on access controls on personal data, our approach differs significantly, not least by focusing on a meta-model of access control from which an axiomatic base can be developed that allows for specific models and policies to be derived as particular instances. The work by Byun et al. [20] on Purpose-based Access Control is related to ours in that a formally well-defined framework for privacy protection is described. However, as we previously explained, Byun et al.'s PBAC is a particular interpretation of privacy-based access control whereas our approach is intended to be understood

as a "universal" interpretation that admits multiple particular interpretations, e.g., of authorization (cf. the discussion on treating purpose existentially or explicitly in relation to $\mathcal{PAR}$).

Our proposal has been firmly grounded in fibred logic and specifically P-FSL. Related approaches do not necessarily have the same well-defined foundational semantics that our approach offers. It is, for instance, already well known that the P3P proposal has some troublesome semantic features (so ambiguous and inconsistent P3P policies may be specified) and EPAL has an operational semantics that is dependent on rule order. Moreover, although XACML has a privacy profile, XACML, in its full generality, does not have the type of well-defined semantics on which our approach is grounded.

The work of Barth et al. [10] is related to ours in some important respects. Barth et al. provide an abstract model of privacy that is founded upon a well-defined conceptual basis (*contextual integrity*) and a well-defined formal basis (linear temporal logic and the *Logic of Privacy and Utility (LPU)*) from which a wide range of privacy policies may be formulated. Along similar lines, we have tried to provide a well-defined conceptual base (i.e., $\mathcal{M}^P$) and a well-defined formal basis (fibring and P-FSL) from which a range of privacy policies may be formulated. However, the emphasis in Barth et al.'s work is on protection of the flow of personal information, violations of the normative behaviors that members of a role are expected to adhere to, and a logical formulation of a framework that makes use of LPU. In contrast, our concern is to provide a unified framework in which privacy is treated as an aspect of access control. We base our conceptual framework on the general notion of category and we have also been concerned with actions in general (not just communication actions). For our formal foundations, we use fibring to admit the possibility of formulating models and policies in various logics and for defining categories in various logics. The idea of treating, in our approach, information flows in relation to communications in the context of norms, as Barth et al. propose, is an interesting matter for further work.

P-FSL shares with ABLP [17] the core operators *says* and *controls* and can be seen as an extension of ABLP in various ways. ABLP is a propositional logic whereas P-FSL adopts a first-order language that is more expressive and permits us to embed the abstract meta-model $\mathcal{M}^P$ into P-FSL. Moreover, P-FSL proposes a more fine-grained notion of compound principals, in fact ABLP has ad-hoc operators to combine atomic principals in order to express joint supports (e.g., $A \wedge B$ says ψ means that principal A and principal B *jointly* supports ψ to hold) while in P-FSL groups of principals are described by means of first-order formulas with one free variable. In this view, *every* formula of the language can be used to describe a set of principals.[7] Finally, P-FSL proposes a completely new semantics with respect to existing access control logics, which is grounded on fibring and using neighborhood functions to give semantics to the *says* operator.

[7] In [16] it is shown how this feature can be exploited to represent separation of duties in a compact way, a representation that it is not possible by using ABLP language.

7 Conclusions and Further Work

We have described an approach that provides users with flexible means for defining the access policies that they require to hold on their "private" data. For that, we introduced a general, abstract access control model $\mathcal{M}^P$, which enables data subjects to conceptualize notions. Our meta-model can be specialized by data subjects in multiple ways so that it may be used to represent a range of access control models and privacy-enhanced access control policies. We formally defined the elements of our meta-model and we expressed privacy policies in P-FSL. We provide a general axiomatic framework that may be specialized by users in multiple ways to represent their individual privacy policy requirements. Our use of P-FSL and fibred logic enables us to develop a formal foundation for a range of privacy-enhanced access control models and policies and permits complex categories of subjects to be flexibly defined in various logics.

On specifics, we note that our meta-model is essentially based on the use of just five basic relations (the *pca*, *arca*, *arcd*, *par* and *prm* relations) to which "higher-level" *contains*, *controls* and *says* relations are added. Application-specific predicates and non-logical axioms may also be added to the core sets of meta-model features (which may be variously specialized) in order to enable data subjects to define specific privacy-enhanced access control models and policies to satisfy their particular requirements on the protection and exploitation of their data. Providing data subjects with a simple, high-level, implementation-independent, expressive framework for formulating their individual requirements on releases of their personal data is a start towards addressing the key open question of how to provide means that might enable data subjects "to choose freely under what circumstances and to what extent they will expose themselves, their attitudes, and their behavior, to others" [1].

Future work includes to incorporate the notion of obligations and hierarchies of purposes in our model, to build in auditing procedures, to investigate norm-based interpretations of categories and to investigate the use of standard implementation languages, like SQL, for category definition. The focus of this paper has been on the development of semantic notions. In future work, we intend to investigate relevant proof-theoretic notions, like proving meta-theoretic properties of policies that are expressed in P-FSL.

Acknowledgements: Valerio Genovese is supported by the National Research Fund, Luxembourg. The authors thank the reviewers for their comments, which proved to be helpful for improving the clarity of the paper.

References

1. Westin, A.: Privacy and Freedom. New York: Atheneum (1967)
2. Berners-Lee, T.: The semantic web will build in privacy (2009) http://news.cnet.com.
3. Simons, W., Mandl, K., Kohane, I.: The PING personally controlled electronic medical record system: Technical architecture. Journal of the American Medical Informatics Association **12**(1) (2005) 45–54

4. Cranor, L.F.: P3p: Making privacy policies more useful. IEEE Security & Privacy **1**(6) (2003) 50–55
5. Backes, M., Dürmuth, M., Karjoth, G.: Unification in privacy policy evaluation - translating EPAL into Prolog. In: POLICY. (2004) 185–188
6. LeFevre, K., Agrawal, R., Ercegovac, V., Ramakrishnan, R., Xu, Y., DeWitt, D.J.: Limiting disclosure in hippocratic databases. In: VLDB. (2004) 108–119
7. Anderson, A.H.: A comparison of two privacy policy languages: EPAL and XACMl. In: SWS. (2006) 53–60
8. Ni, Q., Trombetta, A., Bertino, E., Lobo, J.: Privacy-aware role based access control. In: SACMAT. (2007) 41–50
9. Ni, Q., Bertino, E., Lobo, J., Calo, S.B.: Privacy-aware role-based access control. IEEE Security & Privacy **7**(4) (2009) 35–43
10. Barth, A., Datta, A., Mitchell, J.C., Nissenbaum, H.: Privacy and contextual integrity: Framework and applications. In: IEEE Symposium on Security and Privacy. (2006) 184–198
11. Gabbay, D.M.: Fibring logics. Oxford University Press (1999)
12. Barker, S.: The next 700 access control models or a unifying meta-model? In: SACMAT. (2009) 187–196
13. Barker, S., Boella, G., Gabbay, D.M., Genovese, V.: A meta-model of access control in a fibred security language. Studia Logica **92**(3) (2009) 437–477
14. Lampson, B.W., Abadi, M., Burrows, M., Wobber, E.: Authentication in distributed systems: Theory and practice. ACM Trans. Comput. Syst. **10**(4) (1992) 265–310
15. Li, N., Grosof, B.N., Feigenbaum, J.: Delegation logic: A logic-based approach to distributed authorization. ACM Trans. Inf. Syst. Secur. **6**(1) (2003) 128–171
16. Genovese, V., Gabbay, D.M., Boella, G., van der Torre, L.: FSL – fibred security language. In Boella, G., Noriega, P., Pigozzi, G., Verhagen, H., eds.: Normative Multi-Agent Systems. Number 09121 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany (2009)
17. Abadi, M., Burrows, M., Lampson, B.W., Plotkin, G.D.: A calculus for access control in distributed systems. ACM Trans. Program. Lang. Syst. **15**(4) (1993) 706–734
18. Chellas, B.: Modal logic an introduction. Cambridge University Press (1980)
19. Gabbay, D., Kurucz, A., Wolter, F., Zakharyaschev, M.: Many-Dimensional Modal Logics: Theory and Applications. Elsevier - Studies in Logic (2003)
20. Byun, J.W., Bertino, E., Li, N.: Purpose based access control of complex data for privacy protection. In: SACMAT. (2005) 102–110
21. Barker, S., Sergot, M.J., Wijesekera, D.: Status-based access control. ACM Trans. Inf. Syst. Secur. **12**(1) (2008)
22. Jajodia, S., Samarati, P., Sapino, M., Subrahmaninan, V.: Flexible support for multiple access control policies. ACM TODS **26**(2) (2001) 214–260
23. Fischer-Hubner, S.: IT-Security and Privacy. Springer (2001)