Capacity: an Abstract Model of Control over Personal Data

Daniel Le Métayer, Pablo Rauzy

To cite this version:

HAL Id: hal-01939847
https://hal.archives-ouvertes.fr/hal-01939847
Submitted on 29 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Capacity: an Abstract Model of Control over Personal Data

Daniel Le Métayer
Inria, Université de Lyon
daniel.le-metayer@inria.fr

Pablo Rauzy∗
Université Paris 8 / LIASD
pablo.rauzy@univ-paris8.fr

ABSTRACT
While the control of individuals over their personal data is increasingly seen as an essential component of their privacy, the word “control” is usually used in a very vague way, both by lawyers and by computer scientists. This lack of precision may lead to misunderstandings and makes it difficult to check compliance. To address this issue, we propose a formal framework based on capacities to specify the notion of control over personal data and to reason about control properties. We illustrate our framework with social network systems and show that it makes it possible to characterize the types of control over personal data that they provide to their users and to compare them in a rigorous way.

KEYWORDS
privacy, control, formal model

ACM Reference Format:

1 INTRODUCTION
Instead of the “right to be let alone”, as originally coined by Samuel Warren and Louis Brandeis in their landmark article [16], privacy is increasingly seen as the ability for individuals to control their personal data1. The current trend is also to recommend the integration of privacy requirements in the earliest stages of the design of a product, following the privacy by design approach. However, even if the notions of privacy as control and privacy by design are predominant in the privacy literature, clear definitions of their meanings are still missing. The word “control” in particular is usually used in a very vague way in this context, both by lawyers and by computer scientists. This lack of precision may lead to misunderstandings and makes it difficult to check compliance. To address this issue, we propose a formal framework to specify the notion of control over personal data and to reason about control properties.

1∗This principle is often called “informational self-determination” after a ruling of the German Federal Constitutional Court related to the 1983 census.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CODASPY ’18, March 19–21, 2018, Tempe, AZ, USA
© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.
ACM ISBN 978-1-4503-5632-9/18/03 ... $15.00
https://doi.org/10.1145/3176258.3176314

2 CAPACITY
The notion of control occurs in different contexts such as “access control” or “usage control” in computer science, but none of these variants really encapsulates the intuition underlying the notion of control over personal data. Previous work [8] has identified three dimensions of control over personal data corresponding to the capacities for an individual

(1) to perform actions on their personal data,
(2) to prevent others from performing actions on their personal data, and
(3) to be informed of actions performed by others on their personal data.

Actions can be of various kinds including, without limitation, consultation, modification, deletion and disclosure. In this paper, we build on this reflection to define a formal model of control and we show its relevance through the description of different options of implementation of a social network system. We show that each option provides a different type of control that can be characterized in a formal way.

Contributions and organization of the paper. In Section 2, we introduce an abstract model, called Capacity, which makes it possible to express, inter alia, the three capacities put forward in [8]. We proceed in Section 3 with the definition of requirements characterizing typical variants of control: action control, observability control, authorization control, and notification control. In Section 4, we introduce a concrete case study, a social network system, with its specification and describe it within the Capacity model. Section 5 presents three implementations of the specifications of the case study corresponding to different architectural choices (respectively centralized, peer to peer, and federated). We prove that they meet different control requirements in the Capacity model. In Section 6, we present previous works on control and compare them with our model before suggesting avenues for further research in Section 7. In Appendix B, we define an order relation which is useful to compare different systems based on the level of control that they provide (but not necessary to follow the body of the paper).
2.1 Objects
The Capacity model is based on four types of atomic objects drawn from finite distinct sets \( \mathcal{A}, \mathcal{R}, \mathcal{O}, \) and \( \mathcal{C} \):
- **Agents**, noted \( a_1, a_2, \ldots \in \mathcal{A} \), represent active entities (typically users or services).
- **Resources**, noted \( r_1, r_2, \ldots \in \mathcal{R} \), typically include personal data.
- **Operations**, noted \( o_1, o_2, \ldots \in \mathcal{O} \), may typically include access, update, deletion, and communication operations.
- **Contexts**, noted \( c_1, c_2, \ldots \in \mathcal{C} \), denote the context in which an agent operates on a resource, i.e., any external factors relevant to the operation. Depending on the application, the context can include information such as location, time, or relationships between agents. Contexts can be used in particular to distinguish successive applications of an operation to the same arguments, or more high-level concepts such as the purpose for which personal data are processed.

2.2 Actions
We call an action the application of an operation to a list of parameters in a given context. By convention, we write\(^2\) \( o(x_1, \ldots, x_n) \) the action consisting of the application of operation \( o \) to arguments \( x_1, \ldots, x_n \) in context \( \Delta \). The arguments \( x_i \) can be resources or agents. When the context is irrelevant, we omit the context and simply write \( o(x_1, \ldots, x_n) \). \( \Delta \) denotes the set of actions.

2.3 Relations
To be able to express privacy requirements, we need to introduce three relations on atomic objects:
- **\( \text{Pers}(r, a) \)** expresses that resource \( r \) is a personal data\(^3\) of agent \( a \),
- **\( \text{Inv}(r, a) \)** means that resource \( r \) is involved\(^4\) in action \( a \), and
- **\( \text{Trust}(a, b) \)** expresses that agent \( a \) trusts agent \( b \).

The last relation can be useful to distinguish situations in which the control of an agent over their personal data depends only on trusted agents from situations in which it depends also on untrusted agents\(^5\).

2.4 Requirements
We define a requirement \( R \) as a relation \( \text{Can}^R(a, a, E, W) \subseteq \mathcal{A} \times \Delta \times \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A}) \) such that \( a \notin E \) and \( a \notin W \). The intuition for \( \text{Can}^R(a, a, E, W) \) is that agent \( a \) can perform action \( a \) only if this action is enabled by all agents in \( E \) (the enablers), while all agents in \( W \) (the witnesses) have to be informed about the performance of this action by \( a \). In other words, agents in \( E \) can prevent \( a \) from doing \( a \).

By convention, we use \( \perp \) to denote an undefined or phantom agent, which never performs nor enables any action. Therefore, \( \text{Can}^R(a, a, (\perp), W) \) expresses the fact that requirement \( R \) prevents \( a \) from performing action\(^6\) \( a \). Conversely, \( \text{Can}^R(a, a, \emptyset, W) \) expresses the fact that requirement \( R \) unconditionally allows \( a \) to perform action \( a \).

Requirements make it possible to express the three capacities mentioned in the introduction, depending on the position of an agent \( x \) in the parameters of \( \text{Can}^R(a, a, E, W) \):
1. a property with \( x = a \) expresses the conditions under which \( x \) has the capacity to perform an action,
2. a property with \( x \in E \) expresses the capacity of \( x \) to prevent others from performing an action, and
3. a property with \( x \in W \) expresses the capacity of \( x \) to be informed of the performance of an action by another agent.

We further elaborate on these options in Section 3 in which we take a systematic approach and describe four types of control.

**Definition 1** (Compact requirements). A requirement \( R \) is said to be compact if \( \forall a \in \mathcal{A}, \forall \alpha \in \Delta, \exists a, \mathcal{R}^R(a, a, E, W) \), and \( \forall a \in \mathcal{A}, \forall \alpha \in \Delta, \forall E, E', \mathcal{R}^R(a, a, E, W) \land \mathcal{R}^R(a, a, E', W') \implies E = E' \land W = W' \).

Without loss of generality, we only consider compact requirements in the rest of this paper, and to improve readability, we introduce the following functions:
- \( \text{Can}^R_{\text{req}}(a, a) = E \implies \exists \mathcal{R}^R(a, a, E, W) \),
- \( \text{Can}^R_{\text{req}}(a, a) = W \implies \exists \mathcal{R}^R(a, a, E, W) \),

which are well-defined because of the restriction to compact requirements.

In order to define the semantics of requirements, we characterize execution traces \( \theta \) in a very abstract way (the type of \( \theta \) remains opaque at this level of abstraction), in terms of the following properties:
1. \( \theta \vdash \text{Requests}(a, a) \) means that, in trace \( \theta \), agent \( a \) attempts to perform action \( a \).
2. \( \theta \vdash \text{Enables}(a, b, a) \) means that, in trace \( \theta \), agent \( a \) enables the performance of action \( a \) by agent \( b \).
3. \( \theta \vdash \text{Does}(a, b, a) \) means that, in trace \( \theta \), agent \( a \) performs action \( a \) on behalf of agent \( b \). Does makes it possible to distinguish the agent actually performing an action from the agent that has initiated this action, which is useful in many situations.
4. \( \theta \vdash \text{Notifies}(a, b, c, a) \) means that, in trace \( \theta \), agent \( a \) notifies agent \( b \) the performance of action \( a \) on behalf of agent \( c \).

At this stage, we are content with an intuitive description of the above properties which are used below to define trace consistency and trace compliance. For each implementation, these properties will be defined precisely in terms of the corresponding execution traces (Section 5). It should be noted that this abstract level does not involve any notion of time or position in a trace; as suggested above, the context can be used to distinguish actions corresponding to different occurrences of application of an operation. At the implementation level, this information can be refined, for example, in terms of the position of an event in the trace.

\(^2\)It should be noted that \( o(x_1, \ldots, x_n) \) is not the result of the application of \( o \) to parameters \( x_1, \ldots, x_n \). Formally speaking, it is just a convenient notation for the tuple \([o, c, x_1, \ldots, x_n] \).

\(^3\)A resource can be the personal data of multiple agents. Therefore, we can have \( \text{Pers}(r, a_1) \) and \( \text{Pers}(r, a_2) \) with \( a_1 \neq a_2 \).

\(^4\)More precisely, \( r \) is a parameter (or is included in one of) the operation of \( a \).

\(^5\)The notion of trust is intentionally left informal at our level of abstraction.

\(^6\)Specifying that an action is not possible amounts to requiring that this action has to be enabled by an agent that does not exist or never enables any action.
Definition 2 (Trace consistency). Trace $\theta$ is said to be consistent if $\theta \vdash \text{Does}(c, a, \alpha)$ $\implies$ $\exists d, \theta \vdash \text{Does}(d, c, a)$.

Definition 2 expresses the fact that a trace is inconsistent if it includes an action performed on behalf of an agent that has not requested it or the notification of an action that has not been performed.

Definition 3 (Trace completeness). Trace $\theta$ is said to be complete with respect to requirement $R$ if $\theta \vdash \text{Requests}(a, \alpha)$ and $\forall b \in \text{Can}^R_{\text{e}}(a, \alpha), \theta \vdash \text{Enables}(b, a, \alpha) \implies \exists c, \theta \vdash \text{Does}(c, a, \alpha)$.

Definition 3 characterizes a complete trace by the fact that a requested action is always performed if all the enabling agents have actually enabled it.

In the rest of this paper, we assume that traces are both complete and consistent.

We can now characterize the notion of compliance of a trace with a requirement.

Definition 4 (Trace compliance). Trace $\theta$ is compliant with requirement $R$, noted $\theta \models R$ if and only if $\forall a \in A, \forall d \in A, \forall \alpha \in A, \theta \vdash \text{Does}(d, a, \alpha) \implies \exists b \in \text{Can}^R_{\text{e}}(a, \alpha), \theta \vdash \text{Enables}(b, a, \alpha)$, and $\forall b \in \text{Can}^R_{\text{a}}(a, \alpha), \exists c \in A, \theta \vdash \text{Notifies}(c, b, a, \alpha)$.

In a nutshell, trace $\theta$ complies with requirement $R$ if all Can constraints are met by $\theta$: no action is performed unless it is enabled by all its enablers and all agents that have to be notified are notified.

3 TYPES OF CONTROL

Capacity requirements can be used to characterize different forms and levels of control which can typically be required or expected by data subjects. In the following, we introduce four types of control:

- **Action control**, characterizing an agent’s control on the actions that it initiates.
- **Observability control**, characterizing an agent’s capacity to perform actions that are not observable by others.
- **Authorization control**, characterizing an agent’s control on the actions initiated by others.
- **Notification control**, characterizing an agent’s capacity to be informed about actions performed by others.

For each type of control, we distinguish two levels: absolute control and relative control (in addition to level zero, or lack of control).

Definition 5 (Action control). Requirement $R$ provides agent $a$ an absolute action control over action $\alpha$, noted $\text{AA}_R(a, \alpha)$, if and only if $\text{Can}^R_{\text{e}}(a, \alpha) = \varnothing$.

Requirement $R$ provides agent $a$ a relative action control over action $\alpha$, noted $\text{RA}_R(a, \alpha)$, if and only if $\forall b \in A, b \in \text{Can}^R_{\text{a}}(a, \alpha) \implies \text{Trust}(a, b)$.

$\text{AA}_R(a, \alpha)$ means that $a$ can perform action $\alpha$ without needing any enabler. In other words, $a$ does not depend on any other agent to perform $\alpha$. In contrast, $\text{RA}_R(a, \alpha)$ means that $a$ depends on other agents to perform $\alpha$ but all these agents are trusted by $a$.

Definition 6 (Observability control). Requirement $R$ provides agent $a$ an absolute observability control over action $\alpha$, which is noted $\text{AO}_R(a, \alpha)$, if and only if $\exists c \in A, \theta \vdash \text{Does}(c, a, \alpha)$.

Requirement $R$ provides agent $a$ a relative observability control over action $\alpha$, which is noted $\text{RO}_R(a, \alpha)$, if and only if $\forall b \in A, b \in \text{Can}^R_{\text{e}}(a, \alpha) \implies \text{Trust}(a, b)$.

$\text{AO}_R(a, \alpha)$ means that $a$ can perform $\alpha$ discreetly, that is to say without being observable by other agents. In contrast, $\text{RO}_R(a, \alpha)$ means that other agents can know that $a$ performs $\alpha$ but all these agents are trusted by $a$.

We should emphasize that we consider only observability of actions here and do not express other forms of observability or notions of implicit information flows for example.

Definition 7 (Authorization control). Requirement $R$ provides agent $a$ an absolute authorization control over action $\alpha$, which is noted $\text{AH}_R(a, \alpha)$.

Requirement $R$ provides agent $a$ a relative authorization control over action $\alpha$, which is noted $\text{RH}_R(a, \alpha)$, if and only if $\forall b \in A$ such that $b \neq a, \text{Can}^R_{\text{e}}(b, a) = \{\alpha\}$.

$\text{AH}_R(a, \alpha)$ means that $a$ can enable the performance of action $\alpha$ by other agents and is the only agent having this power\footnote{We recall that $a \in \text{Can}^R_{\text{e}}(b, a)$ means that $a$ has the capacity to prevent $b$ from performing $\alpha$.}.

In other words, the possibility for another agent to perform $\alpha$ depends only on $a$. In contrast, $\text{RH}_R(a, \alpha)$ means that $a$ is not the only agent having this power. In other words, the possibility for another agent to perform $\alpha$ depends not only on $a$ but also on other agents.

Definition 8 (Notification control). Requirement $R$ provides agent $a$ an absolute notification control over action $\alpha$, noted $\text{AN}_R(a, \alpha)$, if and only if $\forall b \in A$ such that $b \neq a, \text{Can}^R_{\text{e}}(b, a) = \{\alpha\}$.

Requirement $R$ provides agent $a$ a relative notification control over action $\alpha$, noted $\text{RN}_R(a, \alpha)$, if and only if $\forall b \in A$ such that $b \neq a, a \in \text{Can}^R_{\text{a}}(b, a)$.

$\text{AN}_R(a, \alpha)$ means that $a$ is informed about the performance of action $\alpha$ by other agents and is the only agent having this power. In contrast, $\text{RN}_R(a, \alpha)$ means that $a$ is not the only agent having this power.

Extensions. All the above definitions of control can be generalized to personal data and agents. For example:

- Requirement $R$ provides agent $a$ an absolute action control over $r$, noted $\text{AA}_R(a, r)$, if and only if $\forall a \in A, \text{In}(r, a)$, then $\text{AA}_R(a, r)$.

- Requirement $R$ provides agent $a$ an absolute action control over their personal data, noted $\text{AA}_R(a)$, if $\forall r \in R, \text{Pers}(r, a) \implies \text{AA}_R(a, r)$.

It is easy to check that, for each variant, absolute control implies relative control. Considering that action, observability, authorization, and notification are four independant variants of control, the above definitions give rise to a lattice (using the order defined by implication) made of 81 forms of control\footnote{Which is equal to $3^3$ considering that 3 levels are possible for each variant: absolute control, relative control, and lack of control.} (for each action, data, or agent).
4 SOCIAL NETWORK SYSTEM

The goal of the Capacity model presented in the previous sections is to capture the meaning of the notion of control in a very abstract and general way. In the rest of this paper, we describe the application of this model to a specific case study and show its relevance to assess different implementation choices. We choose a social network system (“SNS” in the sequel) to illustrate the model, not only because of the central role of social networks nowadays but also because they raise significant challenges in terms of control. We first introduce generic definitions allowing us to describe an SNS in the Capacity framework in Section 4.1 and define some requirements that have to be met by all SNS implementations in Section 4.2. Then, in Section 5, we respectively describe a centralized SNS implementation, a peer to peer SNS implementation, and a federated SNS implementation, and we show that they provide different types of control.

4.1 Generic Definitions

Due to space considerations, we focus on the following set of core SNS features in this paper: profile update, access to profiles, and connection with other users.

The first step to apply the Capacity framework is to define the sets $\mathcal{A}$ (agents), $\mathcal{R}$ (resources), $\mathcal{O}$ (operations) and $\mathcal{C}$ (contexts) introduced in Section 2, as well as the relations Pers, Trust and In introduced in the same section.

- $\mathcal{A}$ includes the users of the social network, noted $u_i$. Depending on the implementation, $\mathcal{A}$ may also include services such as the SNS agent itself (noted sn).
- $\mathcal{R}$ includes two resources per user $u_i$: their name, noted $n_i$, and their complete profile, noted $p_i$ (which includes $n_i$).
- $\mathcal{O}$ consists of the operations update-profile, access-profile, and connect:
  - $\text{update-profile}_c(u_i)$ is the update of the profile $p_i$ of user $u_i$ in context $c$,
  - $\text{access-profile}_c(u_i)$ is the access to the profile $p_i$ of user $u_i$ in context $c$,
  - $\text{connect}_c(u_i)$ is the connection\(^{10}\) to user $u_i$ in context $c$.

Note that an operation does not refer to the agent performing it. This information is provided by relations (such as Can, Requests, Enables, in which actions involving the operation appear.

- $\mathcal{C}$ is defined by $\mathcal{C} \subseteq \mathbb{N}$. A context is simply a natural number corresponding to an index in an execution trace (allowing for disambiguation of otherwise similar events).
- Pers is defined by $\text{Pers}(r, a) \iff a = u_i \land (r = n_i \lor r = p_i)$, assuming for the sake of simplicity, that agents do not share personal data.
- Trust can take different values depending on the agents. For example, some users may trust the SNS agent while others do not, some users may trust some peers but not all other users, etc.
- In is derived from the definition of $\mathcal{O}$:

\[
\ln(r, a) \iff (a = \text{update-profile}_c(u_i) \land (r = p_i \lor r = n_i)) \\
\lor (a = \text{access-profile}_c(u_i) \land (r = p_i \lor r = n_i)) \\
\lor (a = \text{connect}_c(u_i) \land r = n_i).
\]

4.2 Generic SNS Requirements

Some properties, which can be seen as the control oriented part of the SNS specification, have to be met by any SNS implementation. The most important generic requirements are the following:

1. A user cannot prevent another user to update or access their own profile:

   \[
   \forall u_i, u_j \in \mathcal{A}, \text{User}(u_i) \land \text{User}(u_j) \land u_i \neq u_j \implies u_j \notin \text{Can}^R_c(u_i, \text{update-profile}_c(u_i)) \\
   \lor u_j \notin \text{Can}^R_c(u_i, \text{access-profile}_c(u_i)).
   \]

2. A user cannot update the profile of another user:

   \[
   \forall u_i, u_j \in \mathcal{A}, \text{User}(u_i) \land \text{User}(u_j) \land u_i \neq u_j \implies u_j \in \text{Can}^R_c(u_i, \text{update-profile}_c(u_i)).
   \]

3. A user can always refuse a connection request from another user:

   \[
   \forall u_i, u_j \in \mathcal{A}, \text{User}(u_i) \land \text{User}(u_j) \land u_i \neq u_j \implies \bot \in \text{Can}^R_c(u_i, \text{connect}_c(u_j)).
   \]

4. A user cannot interfere in the action concerning two other users:

   \[
   \forall u_i, u_j, u_k \in \mathcal{A}, u_i \neq u_j \neq u_k \neq u_i, \\
   \land \text{User}(u_i) \land \text{User}(u_j) \land \text{User}(u_k) \implies u_k \notin \text{Can}^R_c(u_i, \text{access-profile}_c(u_j)) \\
   \lor u_k \notin \text{Can}^R_c(u_i, \text{connect}_c(u_j)).
   \]

Other properties which are not used in this paper\(^{11}\) are not included in the above list for the sake of conciseness.

5 COMPARING THREE SNS IMPLEMENTATIONS

In this section we describe three SNS implementations: one centralized, one peer to peer, and one federated, and we show that they provide different types of control.

5.1 Centralized SNS Implementation

As a first example of architectural choice, we consider in this section the most common option, that is a centralized SNS implementation. This implementation involves, in addition to user agents $u_i$, the SNS agent $sn$.

In order to describe this implementation in the Capacity framework, we characterize its execution traces in Section 5.1.1 and define the Requests($a, a$), Enables($a, b, a$), and Does($a, b, a$) trace properties\(^{12}\) in Section 5.1.2. Then, we can establish the types and levels of control provided by this implementation in Section 5.1.3.

5.1.1 Execution Traces. Concrete traces in the centralized implementation are sequences of the following events. As a convention, events starting with “U” are those initiated by a user agent $u_i$ and those starting with “S” are initiated by $sn$.

\(^{10}\)What we mean by connection here is a link in the social network (e.g., following someone on Twitter or adding a friend on Facebook).

\(^{11}\)For example, users cannot refuse access to their profiles to users that are connected to them.

\(^{12}\)Due to lack of space, we chose to focus on certain aspects of control and thus voluntarily omit Notifies($a, b, a$).

\[
\begin{align*}
\ln(r, a) & \iff (a = \text{update-profile}_c(u_i) \land (r = p_i \lor r = n_i)) \\
& \lor (a = \text{access-profile}_c(u_i) \land (r = p_i \lor r = n_i)) \\
& \lor (a = \text{connect}_c(u_i) \land r = n_i).
\end{align*}
\]
Can request in trace

In the following, we note $\theta_n$ the $n$th event of an execution trace $\theta$. The above events cannot occur in any order in a valid trace. Space considerations prevent us from presenting the full definition of valid C-traces\(^{13}\) in the core of the paper. The interested reader can find it in Appendix A (Definition 12). To follow the paper, it is sufficient to understand that in a valid C-trace $sn$ does not act spontaneously, in particular no action can be performed on behalf of an agent if this agent has not previously requested this action.

5.1.2 Trace Properties. In order to establish the control requirements provided by the centralized implementation as defined in Section 3, we first have to define the trace properties $Requests(a, \alpha)$, $Enables(a, b, \alpha)$, and $Does(a, b, \alpha)$, which means that these properties are false in all other cases. The context $n$ associated with an action is defined as the index in the trace when this action was requested. Remark that the same event may implement several properties: e.g., event $5.1.2$–$do-upd-profile(sn, u, \alpha)$ implements both $Enables(sn, u, access-profile, \alpha)$ and $Does(sn, u, access-profile, \alpha)$ at the same time because granting profile access and providing profile is implemented in a single step by $sn$ in this architecture.

5.1.2 Trace Properties. As discussed in Section 2 and 3, control can be considered from different perspectives (action, observability, authorization, and notification) and analyzed for each agent and with respect to each action. For the sake of conciseness, we focus on two types of control here:

- Action control of users $u_i$ over the update of their profiles.
- Authorization control of users $u_i$ over the connections to their profile (requested by other users).

Thus we introduce the following control requirement.

**Definition 9** ($Rc$ Requirement). Requirement $Rc$ is defined by:

1. $Can^{Rc}_{u_i}(u_i, update-profile(u_i)) = \{sn\}$

2. $Can^{Rc}_{u_i}(u_i, connect(u_i)) = \{sn, u_i\}$

All other sets are considered empty, which means that we focus only on these two conditions in $Rc$, i.e., that $Can^{Rc}(a, \alpha, \varnothing, \varnothing)$ holds for all other cases.

It is easy to check that $Rc$ satisfies the generic properties presented in Section 4.2. We can now prove that the centralized implementation meets the $Rc$ requirements.

**Theorem 1** (Consistency and compliance of C-traces). Any valid C-trace is consistent and compliant with $Rc$.

Consistency follows directly from Definition 2 (trace consistency) and Definition 12 (validity).

In order to prove compliance, we consider a valid C-trace $\theta$ and show that it complies with the two conditions of Definition 9. From Definition 4 (compliance):

- For the first condition, we have to show:
  \[ \forall u_i \in A, \forall d \in A, \theta \models Does(d, u_i, update-profile(u_i)) \]
  \[ \theta \models Enables(sn, u_i, update-profile(u_i)) \]
  From the definitions of $Does(a, b, \alpha)$ and $Enables(a, b, \alpha)$, this property can be expanded into:
  \[ \exists p \in R, \exists m \in N, \theta_m = S-do-upd-profile(u_i, p) \wedge \Omega(\theta_m, u_i, p) \]
  \[ \exists p \in R, \exists m \in N, \theta_m = S-do-upd-profile(u_i, p) \wedge \Omega(\theta_m, u_i, p) \]

\(^{13}\)Valid traces for the centralized implementation.
\[ \theta_m = S\text{-do-upd-profile}(u_i, p) \land \Omega(n, m, U\text{-req-upd-profile}(u_i, p)) \]

which is obviously true.

- For the second condition, we have to show:
  \[ \forall u_i \in A, \forall d_i \in A, \theta \vdash \text{Does}(d, u_i, \text{connect}_a(u_j)) \implies \exists u_j \in A, \forall \text{Requests}(u_i, u_j) \]

From the definitions (Section 5.1.2) of \text{Does}(a, b, a) and \text{Enables}(a, b, a) the above property can be expanded into:

\[ \exists m \in N, \theta_m = S\text{-do-cnn}(u_i, u_j) \implies \exists m' \in N, \theta_{m'} = S\text{-transfer-req-cnn}(u_i, u_j) \land \Omega(n, m, U\text{-req-cnn}(u_i, u_j)) \land \exists m'' \in N, \theta_{m''} = U\text{-accept-req-cnn}(u_j, u_i) \land \Omega(n, m'', U\text{-req-cnn}(u_i, u_j)) \]

This property follows from the third item of Definition 12 (valid C-traces).

**Theorem 2** (Control under centralized SNS). The centralized implementation provides:

- **Relative action control** on update-profile\(\alpha(u_i)\) to agents \(u_i\) such that Trust\((u_i, s_n)\).
- **No action control** on update-profile\(\alpha(u_i)\) to agents \(u_i\) such that \neg\text{Trust}(u_i, s_n).
- **Relative authorization control** to agents \(u_i\) on connect\(\alpha(u_i)\).

Theorem 2 follows directly from the definitions of relative action control and relative authorization control in Section 3 (Definition 5 and Definition 7 respectively), the definition of \text{Req} (Definition 9) and Theorem 1. It expresses the fact that:

- Agents \(u_i\) may consider that they control the updates of their profile only if they trust the social network \(s_n\).
- Agents \(u_i\) can forbid connections to their profiles but they are not the only actors with this ability.

5.2 **Peer to Peer SNS Implementation**

In this section, we consider a fully decentralized implementation of the social network described in Section 4, in which each agent manages their profile on their own node. In contrast with the previous one, this implementation does not involve any dedicated sn agent.

As was done in the previous section, we first characterize the execution traces of the peer to peer implementation in Section 5.2.1 before defining the trace properties Requests\((a, a)\), Enables\((a, b, a)\), and Does\((a, b, a)\) in Section 5.2.2. Then, we can establish the types and levels of control provided by the peer to peer implementation in Section 5.2.3.

5.2.1 **Execution Traces.** Concrete traces in the peer to peer implementation are sequences of the following events:

- **U\text{-do-upd-profile}(u_i, p):** \(u_i\) sets their profile to \(p\).
- **U\text{-req-acc-profile}(u_j, u_i):** \(u_j\) sends to \(u_i\) a request to access their profile.
- **U\text{-do-acc-profile}(u_i, u_j):** \(u_i\) grants \(u_j\)'s request to access their profile.
- **U\text{-req-cnn}(u_j, u_i):** \(u_i\) sends to \(u_j\) a request to be connected to \(u_j\).
- **U\text{-accept-req-cnn}(u_i, u_j):** \(u_i\) accepts \(u_j\)'s connection request.
- **U\text{-reject-req-cnn}(u_i, u_j):** \(u_i\) rejects \(u_j\)'s connection request.

We remark that all event names start with "U" as users are assimilated to their node of the social network in this fully decentralized model.

A definition of valid P-traces is given in Appendix A (Definition 13). To follow the paper, it is sufficient to understand that in a valid P-trace a user does not address requests that have not been emitted by another agent.

5.2.2 **Trace Properties.** In order to establish the control requirements provided by the peer to peer implementation as defined in Section 3, we must first define the trace properties Requests\((a, a)\), Enables\((a, b, a)\), and Does\((a, b, a)\) in terms of execution traces.

- **\(\theta \vdash \text{Requests}(u_i, \text{update-profile}\alpha(u_i))\) \iff \exists p \in \mathcal{R}, \exists\theta_\alpha = \text{U\text{-do-upd-profile}(u_i, p)}\)
- **\(\theta \vdash \text{Requests}(u_i, \text{access-profile}\alpha(u_j))\) \iff \exists u_j \in A, \forall\text{Update-profile}(u_j, u_i) \land \Omega(n, m, U\text{-req-acc-profile}(u_i, u_j)) \land \exists u_j \in A, \text{U\text{-req-acc-profile}(u_j, u_i)}\)
- **\(\theta \vdash \text{Requests}(u_i, \text{connect}\alpha(u_j))\) \iff \exists u_j \in A, \forall\text{Update-profile}(u_j, u_i) \land \Omega(n, m, U\text{-req-conn}(u_i, u_j)) \land \exists u_j \in A, \text{U\text{-req-conn}(u_j, u_i)}\)

5.2.3 **Control Properties.** As was done in Section 5.1.3, we focus on two types of control here:

- **Action control** of users \(u_i\) over the update of their profiles.
- **Authorization control** of users \(u_i\) over the connections to their profile (requested by other users).

To express these types of control, we introduce the following requirement.

**Definition 10** (Rp Requirement). Requirement Rp is defined by:

1. \(\text{Can}^{\text{Rp}}_{\alpha}(u_i, \text{update-profile}\alpha(u_i)) = \emptyset\)
2. \(\text{Can}^{\text{Rp}}_{\alpha}(u_i, \text{connect}_a(u_j)) = \{u_j\}\)

All other sets are considered empty, which means that we focus only on these two conditions in Rp, i.e., that Can^{Rp}(a, a, \emptyset, \emptyset) holds for all other cases.

It is easy to check that Rp satisfies the generic properties presented in Section 4.2. We can now prove that the peer to peer implementation meets the Rp requirements.

**Theorem 3** (Consistency and compliance of P-traces). Any valid P-trace is consistent and compliant with Rp.
Consistency follows directly from Definition 2 (trace consistency) and Definition 13 (validity).

In order to prove compliance, we consider a valid P-trace $\theta$ and show that it complies with the two conditions of Definition 10.

• The first condition is straightforward (empty set).
• To prove the second condition, we need (Definition 4):
  $$\forall u_i \in A, \forall d \in A, \theta \vdash \text{Does}(d, u_i, \text{connect}_i(u_i)) \implies \theta \vdash \text{Enables}(u_i, u_i, \text{connect}_i(u_i))$$

From the definitions (Section 5.1.2) of Does($a, b, \alpha$) and Enables($a, b, \alpha$), the above property can be expanded into:

$$\exists m \in \mathbb{N}, \exists m' \in \mathbb{N}, \theta_m = \text{U-accept-req-conn}(u_i, u_i) \land \theta_{m'} = \text{U-req-conn}(u_i, u_i)$$

which is obviously true.

**Theorem 4 (Control under peer to peer SNS).** The peer to peer implementation provides:

• Absolute action control on update-profile$_a(u_i)$ to agents $u_i$.
• Absolute authorization control to agents $u_i$ on connect$_i(u_i)$.

Theorem 4 follows directly from the definitions of absolute action control and absolute authorization control in Section 3 (Definition 5 and Definition 7 respectively), the definition of $R_p$ (Definition 10) and Theorem 3. It expresses the fact that:

• Agents $u_i$ do not depend on others to update their profile.
• Agents $u_i$ can forbid connections to their profiles and they are the only actors with this ability.

### 5.3 Federated SNS Implementation

In this section, we consider a partially decentralized implementation of the social network system described in Section 4, where each agent potentially shares their node with others, and may or may not trust their node.

Again, we characterize the execution traces of the federated implementation in Section 5.3.1 before defining the trace properties Requests($a, \alpha$), Enables($a, b, \alpha$), and Does($a, b, \alpha$) in Section 5.3.2. Then, we can establish the types and levels of control provided by the federated implementation in Section 5.3.3.

In this implementation, there are a number of nodes running an instance of the SNS software, which we model by adding:

• a number of agents $s_0, s_1, \ldots$ to $A$, and
• a $\text{Node}(u_i, s_j)$ relation meaning that $U_i(u_i)$ holds and $u_i$ uses the social network via the node $s_j$.

For the sake of simplicity, each user uses only one node, i.e., we have that $\forall u_i, s_j, s_k \in A, \text{Node}(u_i, s_j) \land \text{Node}(u_i, s_k) \implies s_j = s_k$.

In addition, some users may trust their node (e.g., if they are among the node’s administrator and the implementation they run is open source). In such cases we have $\text{Node}(u_i, s_j) \land \text{Trust}(u_i, s_j)$.

#### 5.3.1 Execution Traces

Concrete traces in the federated implementation are sequences of the following events.

• $\text{U-req-upd-profile}(u_i, p, s_j): u_i$ sends an update $p$ of their profile to $s_j$.
• $\text{S-do-upd-profile}(s_j, u_i, p): s_j$ sets $u_i$’s profile to $p$.

### Proof of Theorem 4

- **U-req-acc-profile($u_i, u_j, s_k$):** $u_i$ requests $u_j$’s profile.
- **S-req-profile($s_k, s_j, u_i, u_j$):** $s_k$ requests $u_j$’s profile from $u_i$.
- **S-send-profile($s_j, s_k, u_i, u_j$):** $s_k$ sends $u_j$’s profile to $s_j$.
- **S-do-acc-profile($s_j, s_k, u_i, u_j$):** $s_k$ gives access to $u_j$’s profile to $s_j$.
- **S-req-conn($u_i, u_j, s_k$):** $u_i$ requests $u_j$ to be connected to $u_j$.
- **S-send-conn($s_k, u_i, u_j$):** $s_k$ forwards to $s_j$ the connection request from $u_i$ to $u_j$.
- **S-transfer-req-conn($s_j, u_i, u_j$):** $s_j$ forwards to $s_k$ the connection request from $u_i$.
- **S-req-acc-profile($u_i, u_j$):** $u_i$ accepts $u_j$’s connection request.
- **S-reject-req-conn($u_i, u_j$):** $u_i$ rejects $u_j$’s connection request.
- **S-do-conn($s_j, s_k, u_i, u_j$):** $s_j$ sets up the connection between $u_i$ and $u_j$ via $s_k$.

A definition of valid F-traces is given in Appendix A (Definition 14). To follow the paper, it is sufficient to understand that in a valid F-trace the nodes do not act spontaneously, in particular no action can be performed on behalf of an agent if this agent has not previously requested this action.

#### 5.3.2 Trace Properties

In order to establish the control requirements provided by the federated implementation as defined in Section 3, we must first define the trace properties Requests($a, \alpha$), Enables($a, b, \alpha$), and Does($a, b, \alpha$) in terms of execution traces.
Theorem 5. It expresses the fact that:

\[ \exists k \in \mathcal{K}, \text{Node}(u_i, s_k), \exists m \in \mathbb{N}, \theta_m = \text{S-transfer-req}
\text{-conn}(s_j, u_i, u_j) \land \Omega_\theta(n, m, \text{U-req}
\text{-conn}(u_i, u_j, s_k)) \]

\[ \implies \exists \exists k \in \mathcal{K}, \text{Node}(u_i, s_k), \exists m \in \mathbb{N}, \theta_m = \text{U-accept}
\text{-req}
\text{-conn}(u_i, u_j) \land \Omega_\theta(n, m, \text{U-req}
\text{-conn}(u_i, u_j, s_k)) \]

\[ \theta \triangleright \text{Nodes}(u_i, u_j, \text{connect}_n(u_j)) \]

\[ \iff \exists \exists k \in \mathcal{K}, \text{Node}(u_i, s_k), \exists m \in \mathbb{N}, \theta_m = \text{S-do}
\text{-upd-profile}(s_j, u_i, p) \land \Omega_\theta(n, m, \text{U-
req-upd-profile}(u_i, p, s_j)) \]

\[ \theta \triangleright \text{Nodes}(u_i, u_j, \text{update-profile}(u_j)) \]

\[ \iff \text{Node}(u_i, s_j), \exists p \in \mathcal{R}, \exists m \in \mathbb{N}, \theta_m = \text{S-do}
\text{-upd-profile}(s_j, u_i, p) \land \Omega_\theta(n, m, \text{U-
req-upd-profile}(u_i, p, s_j)) \]

\[ \theta \triangleright \text{Nodes}(u_i, u_j, \text{access-profile}(u_j)) \]

\[ \iff \text{Node}(u_i, s_k), \exists m \in \mathbb{N}, \theta_m = \text{S-do}
\text{-acc-profile}(s_k, u_i, u_j) \land \Omega_\theta(n, m, \text{U-
req-acc-profile}(u_i, u_j, s_k)) \]

\[ \theta \triangleright \text{Nodes}(s_i, u_i, \text{connect}(u_i)) \]

\[ \iff \text{Node}(u_i, s_j), \exists k \in \mathcal{K}, \text{Node}(u_i, s_k), \exists m \in \mathbb{N}, \theta_m = \text{S-
req}
\text{-conn}(s_j, s_k, u_i, u_j) \land \Omega_\theta(n, m, \text{U-
req-conn}(u_i, u_j, s_k)) \]

\[ \Omega_\theta(n, m, \text{U-
req-conn}(u_i, u_j, s_k)) \land \exists m \in \mathbb{N}, \theta_m = \text{S-transfer-req}
\text{-conn}(s_j, u_i, u_j) \land \Omega_\theta(n, m, \text{U-
req-conn}(u_i, u_j, s_k)) \land \exists m \in \mathbb{N}, \theta_m = \text{S-do}
\text{-conn}(s_k, s_j, u_i, u_j) \land \Omega_\theta(n, m, \text{U-
req-conn}(u_i, u_j, s_k)) \]

\[ \exists \exists k \in \mathcal{K}, \text{Node}(u_i, s_k), \exists m \in \mathbb{N}, \theta_m = \text{S-do}
\text{-update-profile}(s_j, u_i, p) \land \Omega_\theta(n, m, \text{U-
req-update-profile}(u_i, p, s_j)) \land \exists p \in \mathcal{R}, \exists m \in \mathbb{N}, \theta_m = \text{S-do}
\text{-update-profile}(s_j, u_i, p) \land \Omega_\theta(n, m, \text{U-
req-update-profile}(u_i, p, s_j)) \]

which is obviously true.

Theorem 6 (Control under federated SNS). The federated implementation provides:

- Relative action control on update-profile\(_n\)(u_i) to agents u_i such that Node(u_i, s_j) \land \text{Trust}(u_i, s_j).
- No action control on update-profile\(_n\)(u_i) to agents u_i such that Node(u_i, s_j) \land \neg \text{Trust}(u_i, s_j).
- Relative authorization control to agents u_i on connect\(_n\)(u_i).

This property follows from the third item of Definition 14 (valid F-traces).

5.3.3 Control Properties. As for the centralized and peer to peer implementations, we focus on two types of control here:

- Action control of users u_i over the update of their own profiles.
- Authorization control of users u_i over the connections to their profile (requested by other users).

5.4 Discussion

We have presented three architectural choices for implementing the social network system introduced in Section 4, namely a centralized implementation (Section 5.1), a peer to peer implementation (Section 5.2), and a federated implementation (Section 5.3). The Capacity model makes it possible to highlight the different types of control provided by these architectural choices.

The centralized implementation gives users relative control over the update over their profile and connections to them, provided that they trust the social network. This would argue in favor of free and open source software for example, as transparency and auditable code is likely to be more trustworthy for users.
The peer to peer implementation gives users the best control over their personal data: they have absolute control over the update of their profile and over who can connect to them.

The federated implementation is, in terms of control, very close to the centralized implementation. Note however that the assumption \( \text{Trust}(u_i, s_j) \) when \( \text{Node}(u_i, s_j) \) in the federated case is much more realistic than the assumption \( \text{Trust}(u_i, sn) \) in the centralized case, as nodes may be operated by users themselves or by people they trust (friends, associations, etc.). In practice, the federated implementation is also more likely to be open source, as its developers do not intend to keep users locked-in their own centralized service.

Because the case study used here is very simple, none of these remarks is really surprising. Nevertheless, it confirms that the intuitive notion of control is well captured by \textit{Capacity}. The added-value of the approach is the fact that these results have been obtained formally through a systematic study of the different implementations.

The same approach can be applied to the analysis of more complex and realistic systems. For example, the CNIL \(^{14}\) recently stated \(^{15}\) that biometric access control on smartphones is acceptable because the biometric data processing is performed under the control of the user. It is not clear, however, in what sense users really control their biometric template, what actions they can perform, enable or observe and what actors they have to trust (in addition to the smartphone provider). The same questions hold for many devices in the internet of things.

The fact that the three implementations studied here implement the generic SNS requirement presented in Section 4.2 suggests the potential benefit of defining relation orders on requirements in \textit{Capacity}. This is further studied in Section B.

Due to lack of space, many aspects of \textit{Capacity} have not been illustrated in this paper. For example, the only type of context used in our case study is the index of an action in a trace. Contexts can actually be used to express different types of contextual information, in the spirit of contextual integrity \(^{1}\).

Other possibilities which have not been illustrated here include personal data of multiple agents, such as pictures involving several friends, and more complex information flows. The first situation can be expressed in a natural way in \textit{Capacity} as the \textit{Pers} relation is not exclusive. As far as complex information flows are concerned, we can consider, for instance, a situation where agent \( A \) grants permission to agent \( B \) to access a data \( d \) but prohibits \( C \) to do so. If \( B \) is able to grant access to this data to \( C \), a model of the system in \textit{Capacity} would show that \( A \) can only have relative control on this data (if they trust \( B \)).

6 RELATED WORK

The two main bodies of work related to the \textit{Capacity} model presented in this paper concern respectively formal privacy policy languages and usage control models. We sketch successively these two trends of work before discussing the main points of departure of the approach followed in this paper.

\textit{Policy languages}. Several languages have been proposed for the definition of privacy policies. They differ mostly in terms of scope (general purpose or specific), target (individual privacy policies, corporate rules, legal rules), and semantics. For this kind of tools to be considered as legitimate means to deliver user consent from a legal point of view, they must be able to express unambiguous choices. One of the criticisms raised against early privacy frameworks such as P3P\(^{16}\) was precisely their lack of clarity and the divergent interpretations of privacy policies. An option to solve the ambiguity problem is to resort to a sound, mathematical definition of the semantics of the language. This approach has been followed in several proposals. For example, CI \(^{1}\) is a dedicated linear temporal logic language inspired by the notion of contextual integrity, which makes it possible to express the conditions that have to be met for an agent, acting in a given role and context, to be allowed to transmit a piece of information. CI makes it possible to express both positive and negative norms, and focuses on the transfer of personal data which forms the core of contextual integrity. Another example of temporal logic based privacy policy language is the language proposed in \(^{2}\) which relies on alternating-time temporal logic.

Other languages such as SAP \(^{3}\) and SIMPL \(^{10}\) rely on a trace semantics. For example, in SIMPL, users can express their policies using sentences such as “I consent to disclose my CV to a third party only if their privacy policy includes the following commitments: only use this data for the purpose of human resource management; delete this data within a maximum delay of three months; do not transfer this data to any third party”.

Particular attention has been paid to privacy policy languages in the specific context of social networks. These proposals are welcome considering that social networks generally provide many privacy options or parameters whose combinations are difficult to grasp. For example, the models presented in \(^{6}\) (which generalizes the access control paradigm) and \(^{12}\) (based on epistemic and deontic properties) can be used not only to better understand the effect of a privacy policy but also to investigate alternative options that are not necessarily supported by existing social networks.

Because they are endowed with a formal semantics in a mathematical framework, the above privacy policy languages make it possible to prove certain properties about the policies (e.g., that a given third party may never receive a given piece of data) and to prove that a given implementation is consistent with the semantics — in other words, that the system behaves as expected by the user.

\textit{Access and usage control models}. The other most relevant body of literature concerns access and usage control models \(^{9, 13, 14}\). Many models have been proposed in the computer security area to characterize the conditions under which subjects can access (or use) certain resources. These models include, \textit{inter alia}, Mandatory Access Control, Discretionary Access Control, and Role-Based Access Control. Usage Control models have been proposed to make it possible to define not only the access rules but also conditions on the use of a resource (e.g., obligations, limitations, etc.). One of the most ambitious frameworks for usage control is the UCON\(_{ABC}\) model, which makes it possible to express authorizations, obligations, and conditions (contextual constraints) at different points of time (before, during, and after usage). Other major features of UCON\(_{ABC}\) are mutable attributes (e.g., actions changing the value

\(^{14}\)French data protection authority.

\(^{15}\)https://www.cnil.fr/fr/biometrie-dans-les-smartphones-loi-informatique-et-libertes-exemption-ou-autorisation

\(^{16}\)https://www.w3.org/P3P/
of an attribute) and the continuity of decisions (i.e., rights may be terminated during the usage when attributes have changed). UCON_{ABC} is very general and can be used to define several families of models. Different approaches have been proposed to define its semantics (or the semantics of a subset of its features), including temporal logic frameworks such as TLA [17] or ITL [4], and process algebra [9].

Capacity. The main point of departure between all the above works and the approach followed in this paper is the fact that the objective of Capacity is not to specify privacy policies or usage control policies, but to define the notion of control itself and to characterize different types of control. For example, UCON_{ABC} being a usage control framework it focuses on the users of a system rather than on the data subjects as it is necessary to properly express privacy requirements and properties. Moreover, Capacity can be seen as a metamodel with respect to privacy policy or usage control models in the sense that it makes it possible to talk about (express, manipulate) notions which are hardwired in these frameworks. Let us take some examples to illustrate this difference of points of view:

Firstly, neither in privacy policy languages nor in usage control models is it possible to express that an agent depends (or does not depend) on other agents to exercise their rights. Some dependencies could be expressed by temporal properties in privacy policy languages (such as, for example, action \( a \) performed by agent \( a \) cannot occur unless action \( b \) performed by agent \( b \) – which could be used to express consent for example – has occurred before). But this would have to be done on a case by case basis, for each relevant action and data (the types of control defined in Section 3, for example, cannot be expressed in these frameworks).

Secondly, both UCON_{ABC} and privacy policy languages assume the existence of an underlying system to manage the rights. The fact that an agent depends on this system to exercise its rights cannot therefore be expressed within the framework itself. This point was raised as the “administrative issues” in UCON and left for further work [13]. Similarly, the privacy languages mentioned above do not make it possible to refer explicitly to the social network itself (the administrator, in UCON terminology). Therefore it is not possible to express the fact that a user can be more or less dependent on the social network. We believe that this possibility is essential in the context of privacy because data controllers (such as social network providers) can in many cases represent the main source of risk for the user[17]. It is therefore necessary to be able to include them in any model of control. In Capacity, the social network is an agent and it is possible to express the level of dependency of the users with respect to this agent just like their dependencies with respect to other users.

Finally, because Capacity is a metamodel rather than a model, it does not make sense to talk about an architecture to implement it (akin, for example, to reference monitors, as discussed in [9]). Requirements in Capacity define abstract control constraints on systems and these constraints have to be refined to establish a link with an actual implementation. For example, the notion of an agent enabling an action is not limited to a matter of granting rights. Enabling can be implemented in many different ways (e.g., forwarding a message, modifying a data on behalf of the requestor, etc.). Access or usage rights are just particular cases for the implementation of Capacity requirements.

7 CONCLUSION
The main objective of this paper was to introduce the Capacity framework and to show its relevance to provide a formal account of the notion of control. The goal of the social network example developed in the previous sections was only to make the Capacity framework more concrete and to show its use to compare different implementations of a system according to control criteria. We have considered only simple social network functionalities and two aspects of control in this example (action control and authorization control). Other aspects of control corresponding to more complex privacy policies can be represented in the same spirit. Other dimensions of privacy can also be expressed in the Capacity framework. In particular, the purpose for which personal data are processed is very important with regard to privacy, and contexts of actions in Capacity can be used to specify such information. Contexts can also be more complex, including, for example, time, space, or environment (work, family, medical, etc.).

Going one step further, additional developments allowing refines in the use of the Capacity framework to formally capture the notion of exposure [11] better than using contexts would be beneficial. Indeed, from a legal point of view, for an event to happen in the public space or in a private space depends on the existence of a community of interest (e.g., on a social network, direct contacts of a particular user that this user has manually approved one by one), which can be modeled using relations as defined in Section 2.3. However, from a privacy point of view, this legal definition is not sufficient as for example the size of a community may be as important as its public or private characterization.

The study of non compact requirements, such as what could be implemented with some types of secret sharing techniques [15] (typically with \((t, n)\)-threshold scheme), would also be interesting. A complementary aspect which has not been discussed in this paper is the verification that a given implementation actually meets the validity properties (Definition 12 and 13) and the completeness property (Definition 3), which ensures that a requested action is always performed if all the enabling agents have actually enabled it. This task pertains to traditional code verification techniques [5].

Another perspective would be the use of the classification of control properties presented in this paper to structure the privacy design space and select appropriate strategies according to the objectives in terms of control [7].

An interesting avenue for further research would be to exploit the systematic classification of control properties presented in Section 3 to help data subjects in the definition of their privacy policies. Indeed, provided that they are supported by user-friendly interfaces, the abstract definitions made possible by the Capacity framework could provide a more systematic and intelligible way to grasp user-defined privacy requirements.
REFERENCES


A FORMAL VALIDITY DEFINITIONS

In this appendix, we present the full definitions of valid C-traces, valid P-traces and valid F-traces referred to in Section 5.1.1, Section 5.2.1 and Section 5.3.1 respectively.

Definition 12 (Valid C-trace). A valid trace for the centralized implementation, or valid C-trace, is a sequence of events (as defined in Section 5.1.1) meeting the following properties:

1. ∀m ∈ N, θ_m = S-do-upd-profile(u_i, p)
   => ∀n ∈ N, n < m \ θ_n = U-req-upd-profile(u_i, p)
2. ∀m ∈ N, θ_m = S-do-acc-profile(u_i, u_j)
   => ∀n ∈ N, n < m \ θ_n = U-req-acc-profile(u_i, u_j)
3. ∀m, m' ∈ N, θ_m = S-do-conn(u_i, u_j)
   => ∀n ∈ N, n < m \ θ_n = U-req-conn(u_i, u_j)

Definition 13 (Valid P-trace). A valid trace for the peer to peer implementation, or valid P-trace, is a sequence of events (as defined in Section 5.2.1) meeting the following properties:

1. ∀m ∈ N, θ_m = U-do-acc-profile(u_i, u_j)
   => ∀n ∈ N, n < m \ θ_n = U-req-acc-profile(u_i, u_j)
2. ∀m ∈ N, θ_m = U-accept-req-conn(u_i, u_j)
   => ∀n ∈ N, n < m \ θ_n = U-req-conn(u_i, u_j)
3. ∀m ∈ N, θ_m = U-reject-req-conn(u_i, u_j)
   => ∀n ∈ N, n < m \ θ_n = U-req-conn(u_i, u_j)

Definition 14 (Valid F-trace). A valid trace for the federated implementation, or valid F-trace, is a sequence of events (as defined in Section 5.3.1) meeting the following properties:

1. ∀m ∈ N, θ_m = S-do-upd-profile(s_j, u_i, p)
   => \text{Node}(s_j, u_i) \land \forall n ∈ N, n < m \ \theta_n = S-req-upd-profile(s_j, u_i, p)
2. ∀m ∈ N, θ_m = S-do-acc-profile(s_k, u_i, u_j)
   => \text{Node}(s_k, u_i) \land \forall n ∈ N, n < m \ \theta_n = S-req-acc-profile(s_k, u_i, u_j)
3. ∀m ∈ N, θ_m = S-do-conn(s_k, u_i, u_j)
   => \forall n ∈ N, n < m \ \theta_n = S-req-conn(s_k, u_i, u_j) \land \Omega_0(n, m, U-req-acc-profile(u_i, s_j))

B RELATION ORDERS ON REQUIREMENTS

In order to make it possible to reason about the level of control provided by a requirement, we introduce several order relations between requirements in Section B.1 and notions of trace compliance with respect to personal data in Section B.2.

B.1 Comparing Requirements

We first define a general order between requirements before introducing specific orders capturing the notion of level of control over personal data.

Definition 15 (R_1 \geq R_2). The general order relation between requirements, noted \geq, is defined as follows: R_1 \geq R_2 if and only if \forall a \in A, \forall \alpha \in A, \text{Can}_e^R(a, \alpha) \subseteq \text{Can}_e^R(a, \alpha) \land \text{Can}_w^R(a, \alpha) \subseteq \text{Can}_w^R(a, \alpha)

Intuitively, R_1 \geq R_2 means that R_1 is more permissive than R_2: it requires less enablers and less witnesses to perform an action. This first order is very generic and it does not deal specifically with personal data. The next definitions allow us to distinguish requirements based on their contraints on personal data.

Definition 16 (R_1 \geq^\alpha R_2). The positive preorder relation between requirements, noted \geq^\alpha, is defined as follows: R_1 \geq^\alpha R_2 if and only if \forall a \in A, \forall \alpha \in A, \text{Pers}(a, \alpha) \land \text{Imp}(a, \alpha) \Rightarrow \text{Can}_e^R(a, \alpha) \subseteq \text{Can}_e^R(a, \alpha) \land \text{Can}_w^R(a, \alpha) \subseteq \text{Can}_w^R(a, \alpha)
Intuitively, $R_1 \succeq_R a R_2$ means that $R_1$ is more permissive than $R_2$ for the manipulation of their personal data by agent $a$: it requires less enablers and less witnesses for $a$ to perform an action on their own personal data.

**Definition 17** ($R_1 \succeq_R a R_2$). The negative preorder relation between requirements, noted $\succeq^a_C$, is defined as follows: $R_1 \succeq^a_C R_2$ if and only if $\forall b \in \mathcal{A}, \forall \alpha \in \Delta, \forall r \in \mathcal{R}$,

- $\text{Pers}(r, a) \land \text{In}(r, a) \land a \in \text{Can}_{\text{R}^b}(b, a) \implies a \in \text{Can}_{\text{R}^b}(b, a)$
- $\text{Pers}(r, a) \land \text{In}(r, a) \land a \in \text{Can}_{\text{R}^w}(b, a) \implies a \in \text{Can}_{\text{R}^w}(b, a)$

Intuitively, $R_1 \succeq^a_C R_2$ means that $R_1$ provides a greater scrutiny than $R_2$ to agent $a$ on the manipulation of their personal data by others: $a$ can enable or witness more actions performed by other agents on $a$’s personal data.

**Definition 18** ($R_1 \succeq_R a R_2$). The control preorder relation between requirements, noted $\succeq_R a$, is defined as follows: $R_1 \succeq_R a R_2$ if and only if $R_2 \succeq^a_C R_1$ and $R_1 \succeq_R b R_2$.

Intuitively, $R_1 \succeq_R a R_2$ means that $R_1$ provides a greater control than $R_2$ over their personal data to agent $a$: it is more permissive about what $a$ can do with their personal data and provides $a$ a greater scrutiny over what others can do with $a$’s personal data.

The above preorder can be generalized to all agents in a natural way. For example, $\succeq_R a$ is defined as follows: $R_1 \succeq_R \mathcal{A} R_2$ if and only if $\forall a \in \mathcal{A}, R_1 \succeq_R a R_2$.

**Theorem 7** (Preorders). $\succeq_R$ is an order relation and $\succeq^a_R, \succeq_R$, $\succeq^a_C, \succeq_R$ are preorder relations.

**Proof.** The property of $\succeq_R$ results from the facts that $\subseteq$ is an order relation and the functions $\text{Can}_{\text{R}}$ and $\text{Can}_{\text{w}}$ together define completely $R$. The other relations are only preorders because their definitions involve only subsets of $\Delta$ (actions involving personal data). As usual, it is possible to derive order relations from these preorders using the quotient sets of the associated equivalence relation. $\square$

### B.2 Relative Trace Compliance

Let us now introduce notions of trace compliance relative to the personal data of a given agent $a$.

**Definition 19** ($\triangleright^a_R R$). The positive compliance of a trace $\theta$ with respect to a requirement $R$ and agent $a$ is defined as follows: $\theta \triangleright^a_R R$ if and only if $\forall \alpha \in \Delta, \forall d \in \mathcal{A}$, $\theta \vdash \text{Does}(d, a, \alpha) \land \text{Pers}(r, a) \land \text{In}(r, a)$

- $\Theta \in \text{Can}_{\text{R}}(a, \alpha, \theta \vdash \text{Enables}(b, a, \alpha))$ and
- $\Theta \in \text{Can}_{\text{w}}(a, \alpha, \exists c \in \mathcal{A}, \theta \vdash \text{Notifies}(c, b, a, \alpha))$

Intuitively, $\theta \triangleright^a_R R$ means that in the execution represented by $\theta$, all actions performed by agent $a$ on their personal data have been enabled by the required agents (members of $\text{Can}_{\text{R}}(a, \alpha, \theta)$) and all the necessary witnesses have been notified (members of $\text{Can}_{\text{w}}(a, \alpha)$).

Positive trace compliance can be generalized to all agents as follows: $\theta \triangleright_R R$ if and only if $\forall a \in \mathcal{A}, \theta \triangleright^a_R R$.

**Definition 20** ($\triangleright_R^a R$). The negative compliance of a trace $\theta$ with respect to a requirement $R$ and agent $a$ is defined as follows: $\theta \triangleright_R^a R$

if and only if $\forall b \in \mathcal{A}, b \neq a, \forall \alpha \in \Delta, \forall d \in \mathcal{A}$, $\theta \vdash \text{Does}(d, b, \alpha) \land \text{Pers}(r, a) \land \text{In}(r, a)$

- $\exists \alpha \in \Delta, \forall d \in \mathcal{A}, \theta \vdash \text{Enables}(b, a, \alpha)$, and
- $\exists \alpha \in \Delta, \forall d \in \mathcal{A}, \theta \vdash \text{Notifies}(c, b, a, \alpha)$.

Intuitively, $\theta \triangleright_R^a R$ means that in the execution represented by $\theta$, any action performed by another agent $b$ on the personal data of $a$ has been enabled by $a$ (if $a$ is a member of $\text{Can}_{\text{R}}(b, a)$) and notified to $a$ (if $a$ is a member of $\text{Can}_{\text{w}}(b, a)$).

Negative trace compliance can be generalized to all agents as follows: $\theta \triangleright_R R$ if and only if $\forall a \in \mathcal{A}, \theta \triangleright_R^a R$.

**Definition 21** ($\triangleright_R R$). The control trace compliance to a requirement with regard to an agent, is defined as follows: $\theta \triangleright_R R$ if and only if $\theta \triangleright_R^a R$.

The control trace compliance is defined as the conjunction of positive and negative trace compliances. The intuition is that the control of an agent over their personal data is greater when they have less constraints on their own actions on their data and more powers to prevent actions from other agents on their data. Control trace compliance can be generalized to all agents as follows: $\theta \triangleright_C R$ if and only if $\forall a \in \mathcal{A}, \theta \triangleright_C R$.

**Theorem 8** (Orders and relative trace compliance). For all requirements $R_1, R_2$, trace $\theta$, and agent $a$, we have:

1. $(1)$ If $R_1 \succeq_R \mathcal{A} R_2$ then $\theta \triangleright_R R_1$ $\implies$ $\theta \triangleright_R R_2$
2. $(2)$ If $R_1 \succeq_R a R_2$ then $\theta \triangleright_R^a R_2$ $\implies$ $\theta \triangleright_R^a R_1$
3. $(3)$ If $R_1 \succeq_R a R_2$ then $\theta \triangleright_R^a R_2$ $\implies$ $\theta \triangleright_R^a R_1$
4. $(4)$ If $R_1 \succeq_R a R_2$ then $\theta \triangleright_R^a R_1$ $\implies$ $\theta \triangleright_R^a R_2$
5. $(5)$ If $R_1 \succeq_R a R_2$ then $\theta \triangleright_R^a R_1$ $\implies$ $\theta \triangleright_R^a R_2$
6. $(6)$ If $R_1 \succeq_R a R_2$ then $\theta \triangleright_R^a R_1$ $\implies$ $\theta \triangleright_R^a R_2$
7. $(7)$ If $R_1 \succeq_R a R_2$ then $\theta \triangleright_R^a R_2$ $\implies$ $\theta \triangleright_R^a R_1$ $\implies$ $\theta \triangleright_R^a R_2$

**Proof.** In order to prove the first property, we assume $R_1 \succeq_R \mathcal{A} R_2$ and $\theta \triangleright_R R_1$. In order to show $\theta \triangleright_R R_2$, let us consider $a \in \mathcal{A}, d \in \mathcal{A}$, $\alpha \in \Delta$ such that $\theta \vdash \text{Does}(d, a, \alpha)$. From $\theta \triangleright_R R_1$, we have:

- $\forall b \in \text{Can}_{\text{R}}^a(a, \alpha, \theta \vdash \text{Enables}(b, a, \alpha))$
- $\forall b \in \text{Can}_{\text{w}}^a(a, \alpha, \exists c \in \mathcal{A}, \theta \vdash \text{Notifies}(c, b, a, \alpha))$

$R_1 \succeq_R a R_2$ entails $\text{Can}_{\text{R}}^a(a, \alpha, \theta \vdash \text{Enables}(b, a, \alpha))$, which allows us to derive:

- $\forall b \in \text{Can}_{\text{R}}^a(a, \alpha, \theta \vdash \text{Enables}(b, a, \alpha))$
- $\forall b \in \text{Can}_{\text{w}}^a(a, \alpha, \exists c \in \mathcal{A}, \theta \vdash \text{Notifies}(c, b, a, \alpha))$

and therefore $\theta \triangleright_R R_2$. The proofs of the other properties are similar with the switch of order for properties 2 and 3 due to the definition of $R_1 \succeq_R \mathcal{A} R_2$. As explained above, more privacy for agents means more possibilities to act on their own data and less possibilities for others. $\square$