Agents for Preserving Privacy

Pınar Yolum
Email: p.yolum@uu.nl

Department of Information and Computing Sciences
Utrecht University

Privacy

- ”Right to be let alone” (Warren and Brandeis, Harvard Law Review, 1890)
- ”A state in which one is not observed or disturbed by other people.” (Merriam-Webster Dictionary)
- ”Someone’s right to keep their personal matters and relationships secret” (Cambridge Dictionary)
- ”The right and ability of an individual to define and live his or her life in a self-determined fashion” (substantive privacy) (Dennedy, Fox, and Finneran, ”The Privacy Engineer’s Manifesto”, 2014)
  - By the individual
  - By others
  - By using the data about the person

Data Privacy

- What data are private?
  - Name, email
  - Financial information
  - Political opinions
  - Sexual orientation
  - Racial or ethnic origin
  - Medical conditions
- Privacy vs. Security

The Organization for Economic Cooperation and Development (OECD) Guidelines

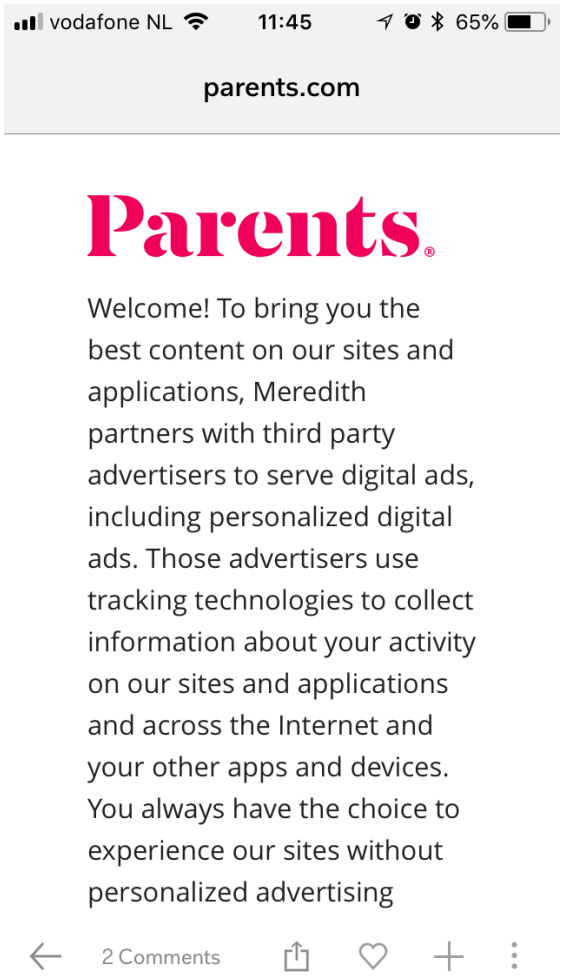
- Collection Limitation: Should be collected lawfully and fairly
- Data Quality Limitation: Should be
  - relevant: Allow appropriate content to be accessed (e.g., age is relevant for checking credit history but phone number is not)
  - accurate: Allow owners to update if necessary
- Purpose Definition Required: Specify explicitly why that information is being shared
- Use Limitation Principle: How it is going to be used (e.g., share with third parties?)
- Accountability Principle: Data sharer will be kept accountable for not abiding with rules

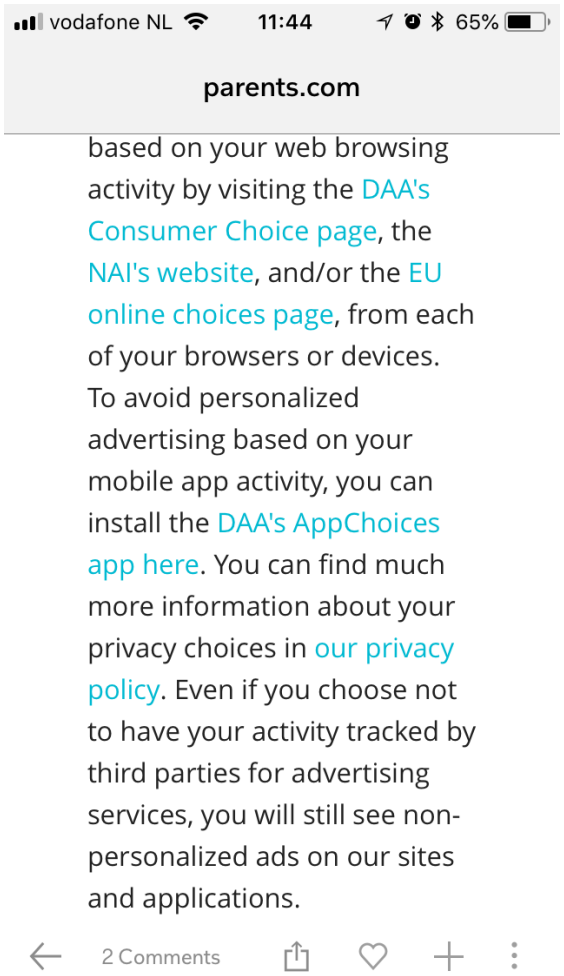
Authorization Types

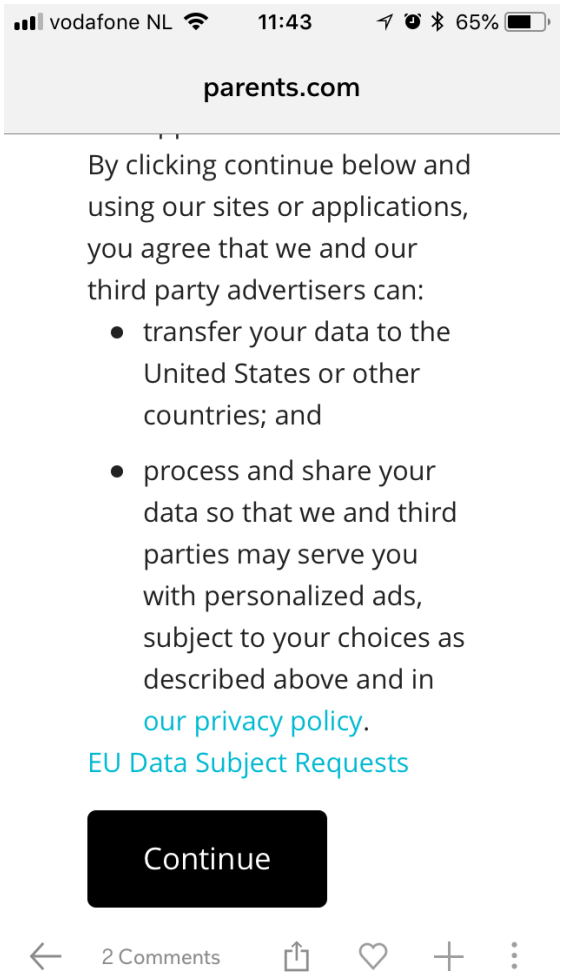
- Opt out (Default is to share)/Opt in (Default is not to share)
- Implied Consent (Your email address appearing on the instructor's list for possible future communication)
- Informed Consent (Explicitly explained how and which information will be used)
- Expressed Consent (Explicitly specified by the user by checking a box or similar)

General Data Protection Regulation (GDPR)

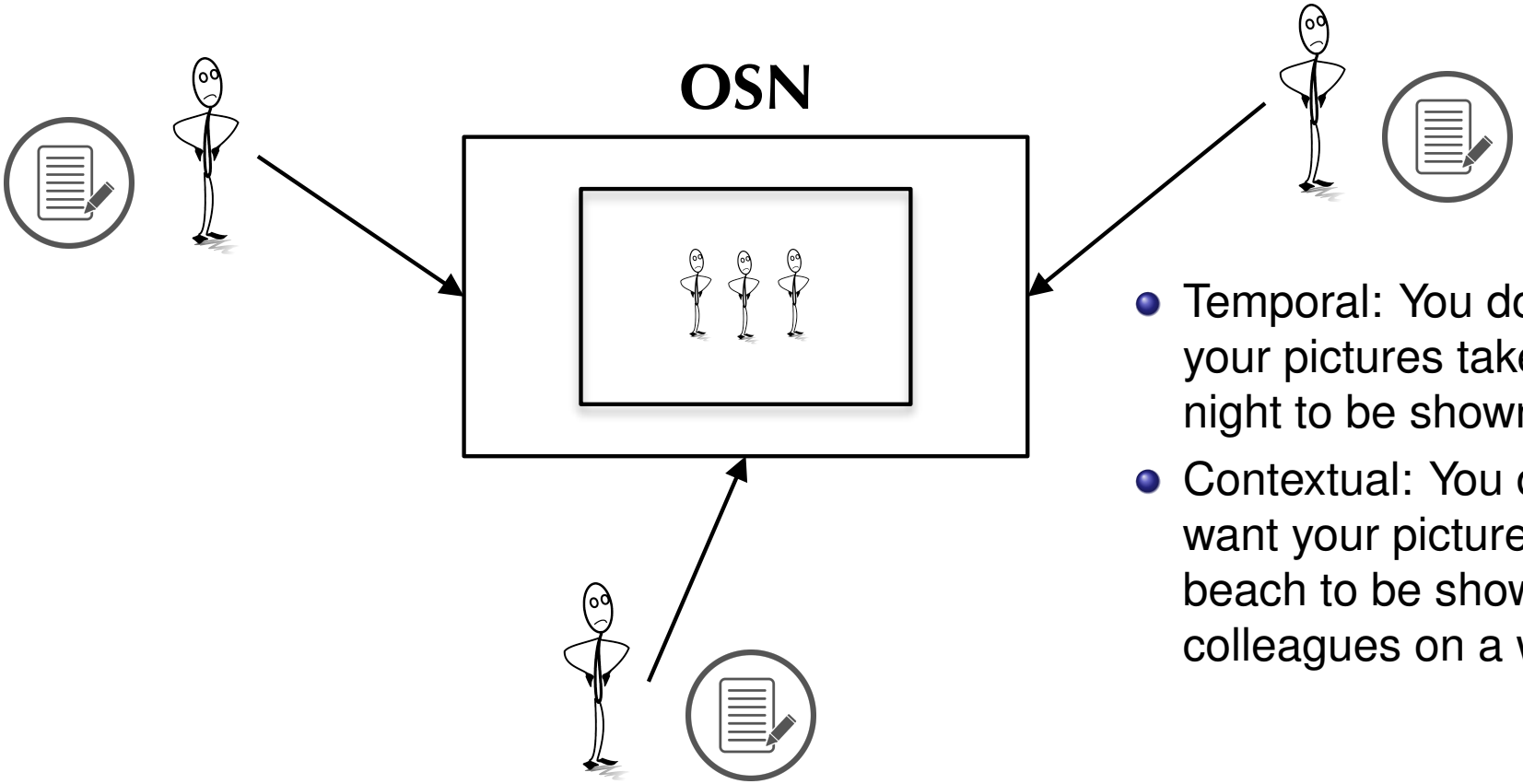
Informed Consent: Explains what and how information is used





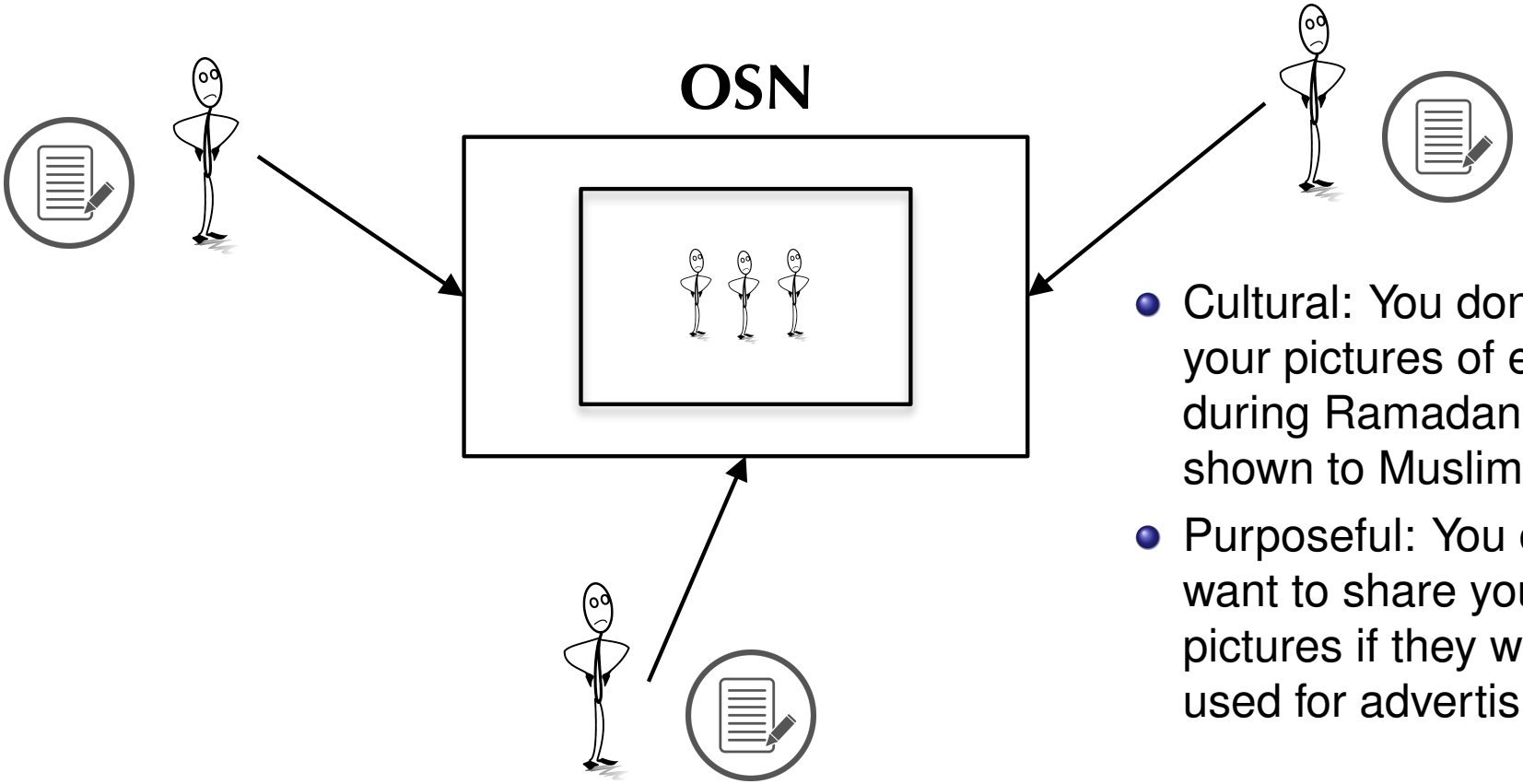


Privacy in Online Social Networks



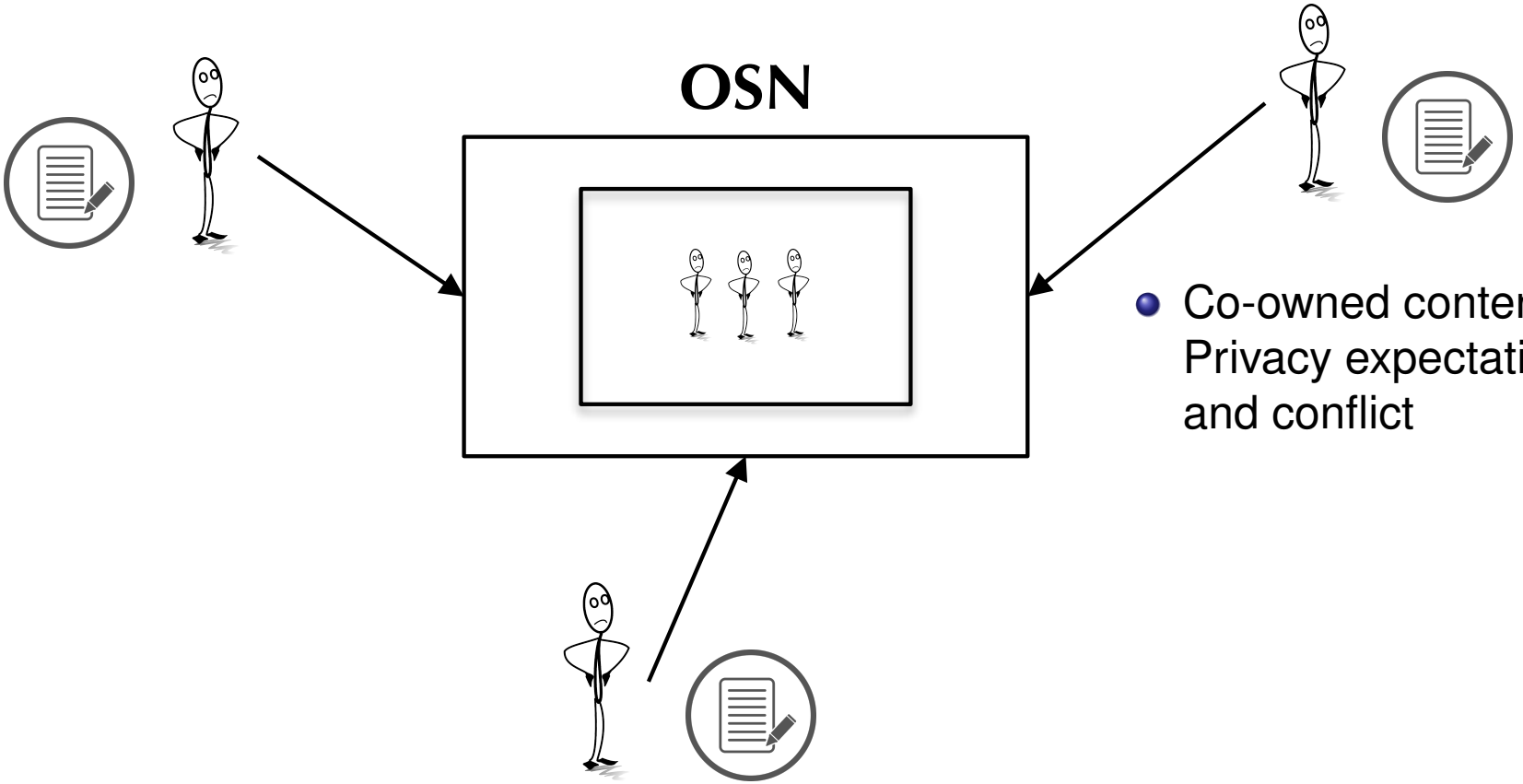
- Temporal: You don't want your pictures taken at night to be shown
- Contextual: You don't want your pictures at the beach to be shown to colleagues on a work day

Privacy in Online Social Networks



- Cultural: You don't want your pictures of eating during Ramadan to be shown to Muslim friends
- Purposeful: You don't want to share your pictures if they will be used for advertising

Privacy in Online Social Networks



- Co-owned content: Privacy expectations vary and conflict

Privacy in Online Social Networks

- Lane v. Facebook: A Class-action lawsuit
  - Sean Lane purchases a diamond ring from Overstock.com.
  - This information shows up on the newsfeed of many of his friends, including his fiancee.
  - This was result of Beacon app, with opt-out privacy options.
  - Facebook ended up paying $9.5M
  - Moral: Information propagates
- Celebrity Stalking (from ABC News)
  - iPhones embed picture locations into the picture (known as geotags)
  - Geotags can easily be deciphered by apps, reveling the location even when not intended
  - Not only bad for celebrities (Craiglist pictures)
  - Moral: Information implies other information

Understanding Privacy Violations

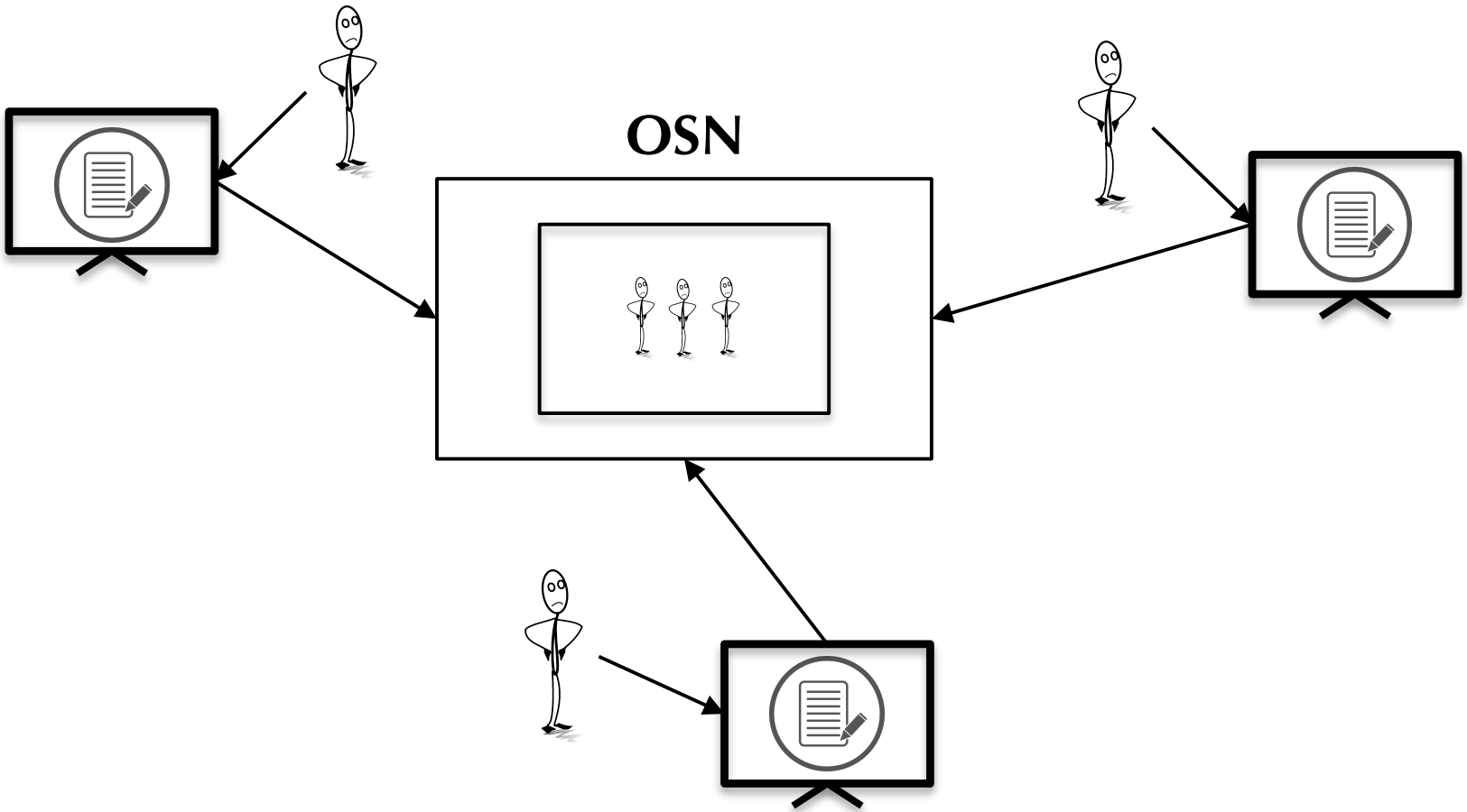
	No inference	**Inference**
User	(i) OSN showing the user's media without consent or user wrongly configuring privacy constraints	(iii) Identifying user's location from a geo-tag in the pictures
Others	(ii) Friend tags the user and makes the picture public where the user did not want to be seen	(iv) Friend tags the user revealing friendship status even when the user had hid her friend list

We have conducted an online survey with 330 participants. More than 96% of the participants face privacy violations that occur through inferences.

Dennis wants his friends to see his pictures but not his location.

	No inference	**Inference**
User	(i) Dennis checks in at a restaurant.	(iii) Dennis shares a picture without declaring his location. It turns out that his picture is geotagged.
Others	(ii) Charlie shares a picture with everyone. He tags Dennis in it as well.	(iv) Charlie checks in at a restaurant. At the same time, Dennis shares a picture of Charlie.

Agent-Based Privacy Management



Do users want agents?

	Internet of Things		Data	PPA		
PID	Opinion	Understanding	Privacy Concern	Notification	Recommendation	Auto
P1	Positive	Low	Concerned	Positive +control	Negative	Negative
P2	Positive	Low	Resigned	Neutral	Negative	Negative
P3	Positive	Low	Resigned	Positive +control	Positive	Automated
P4	Negative	High	Concerned	Positive +control	Positive (education)	Autonomous
P5	Neutral	Average	Concerned	Negative	NA	Autonomous
P6	Both	Average	Resigned	Negative	NA	Autonomous
P7	Both	Low	Unconcerned	Neutral	Positive (education)	Automated
P8	Positive	Average	Neutral	Negative	Positive (education)	Negative
P9	Neutral	Average	Unconcerned	Positive +control	Positive (education)	Automated
P10	Positive	Average	Neutral	Positive +control	No opinion	Autonomous
P11	Both	Average	Neutral	Positive +control	Negative	Negative
P12	Positive	Average	Concerned	Positive +control	NA	NA
P13	Both	High	Concerned	Positive	Negative	Automated
P14	Positive	Average	Unconcerned	Positive +control	[Confused]	Automated
P15	Positive	Average	Unconcerned	Positive	Positive (education)	Autonomous
P16	Positive	Average	Unconcerned	Negative	Positive	Negative
P17	Both	Average	Concerned	Positive +control	Positive	Negative

Table 1. Participant characteristics identified during the interview.

*“Colnago, Jessica, et al. ”Informing the design of a personalized privacy assistant for the internet of things.” Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.”

How to Manage the Privacy of Users?

- How to represent the actual privacy preferences of users?
- How to elicit or learn the privacy preferences from users?
- How to advise the users to take actions that are in line with their privacy preferences?
- How to detect potential privacy violations on a user's side?
- How to agree on how a co-owned content will be shared?

Representations of Privacy Preferences

- Access control: Regulate who can view, edit, use resources
- Role-Based: Users take up roles and act in accordance (RBAC)
- Relation-Based: Capture relations among users
- Attribute-Based: Rules based on values of attributes
- Policy-Based: Enable rules to work in harmony

A Meta-Model for Privacy Aware ABSNs (1)

Definition (Agent)

An agent is a software entity that can share posts (Definition 3) on behalf of a user and can see posts of other agents. $\mathcal{A}$ is the set of agents in the system.

Definition (Content)

C is a set of contents that can be posted in a social network, where $C = \{c_i^t \mid t \in C^{type}\}$. C^{type} is the set of content types.

A Meta-Model for Privacy Aware ABSNs (2)

Definition (Post)

$p_{a,i} = \langle C, x, D \rangle$ denotes a post that is shared by an agent a, where $a \in \mathcal{A}$. A post includes a set of contents C. A post may have a context x. Each post is meant to be seen by a set of agents called its audience D, where $D \subset 2^{\mathcal{A}}$. $\mathcal{P}$ is the set of posts and $\mathcal{P}_a$ is the set of posts shared by agent a.

Definition (Relationship)

r^t_{km} denotes a relationship of type t between two agents k and m, where $k, m \in \mathcal{A}$, $t \in \mathcal{R}^{type}$. R^{type} is the set of relation types, $\mathcal{R}$ is the set of relationships in the system and $\mathcal{R}_k$ is the set of relationships of the agent k.

A Meta-Model for Privacy Aware ABSNs (3)

Definition (OSN Template)

$te_i = \langle R^{type}, C^{type}, \mathcal{N} \rangle$ denotes an OSN template with $te_i \in TE$, where $\mathcal{N}$ is the set of norms.

Definition (Agent-Based Social Network)

ABSN is a three tuple $\langle \mathcal{A}, \mathcal{R}, \mathcal{P} \rangle^{te_i}$, where $te_i \in TE$; $\forall r^{t_1} \in \mathcal{R}, t_1 \in te_i.R^{type}$; $\forall c^{t_2} \in \mathcal{P}.C, t_2 \in te_i.C^{type}$. ABSN is initialized with respect to an OSN template. We assume that ABSN is connected, there is a path between every pair of agents.

A Meta-Model for Privacy Aware ABSNs (4)

Definition (Privacy Requirement)

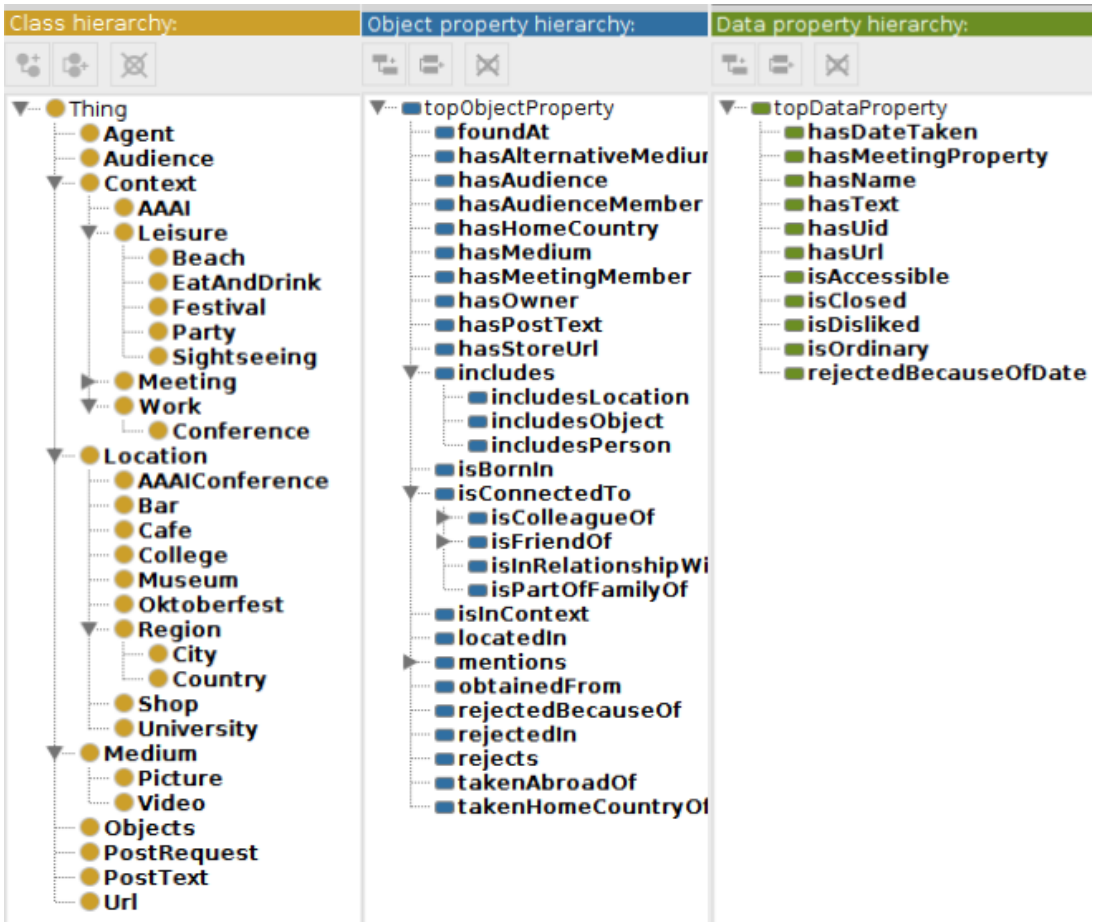
$PR^t_{a,i} = \langle P'_a, I \rangle$ denotes a privacy requirement of the agent a, which is about the set of posts P'_a and affects the set of individuals I, where $P'_a \subset P_a$, $I \subset 2^{\mathcal{A}}$ and $t \in \{+, -\}$. ℓ is a label function that maps the privacy requirement type t to $\{allow, deny\}$, where $\ell(+) = allow$ and $\ell(-) = deny$.

Definition (Privacy Violation)

In a given ABSN, if a privacy requirement $PR^t_{a,i}$ is violated (isViolated($PR^t_{a,i}$, ABSN)), then the following holds: $\exists p \in PR^t_{a,i}.P'_a, \exists a' \in PR^t_{a,i}.I$ and either $t = +$ and $not(canSeePost(a', p))$; or $t = -$ and $canSeePost(a', p)$.

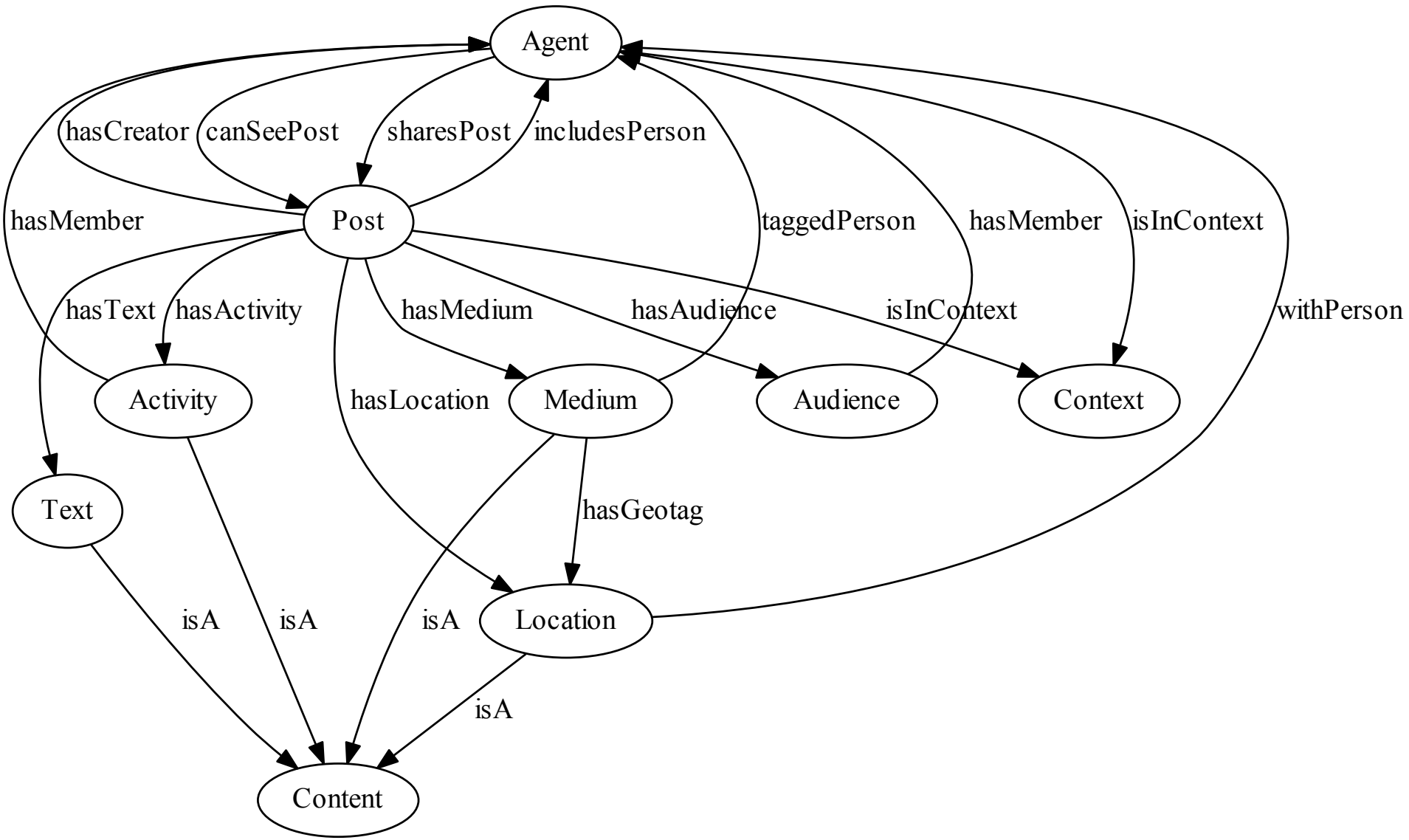
Semantic Representations

Rely on a knowledge representation, such as an ontology, for reasoning on the content.

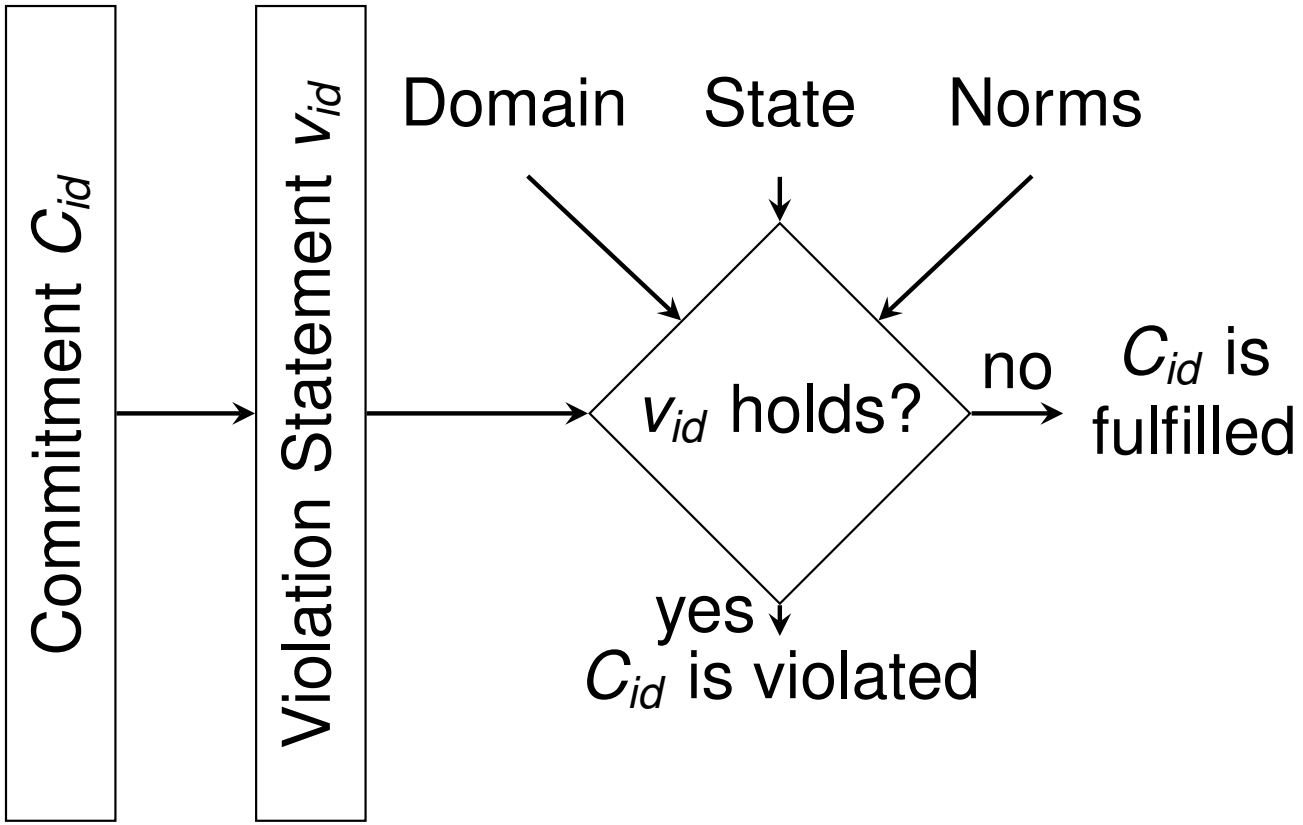


- *Concepts* represent a class of individuals (e.g., wig `:wig` is an instance of *Object*).
- Object properties relate different individuals with a specific relation (e.g., includesObject relates a `:Medium` to a `:wig`).
- Data properties relate data values to individuals (e.g., isOrdinary relates `:wig` to either *true* or *false*).

Content Ontology



Detection Privacy Violations with PriGuard[1]



[1] Nadin Kökciyan and Pınar Yolum. "PriGuard: A Semantic Approach to Detect Privacy Violations in Online Social Networks". In: *IEEE Transactions on Knowledge and Data Engineering* 28.10 (2016), pp. 2724–2737.

Representation of Privacy Requirements

- Commitments are a powerful representation for modeling multiagent interactions.
- Here used to represent the privacy agreement between a user and the OSN.
- A commitment is denoted as a four-place relation: *C(debtor*; *creditor*; *antecedent*; *consequent)*

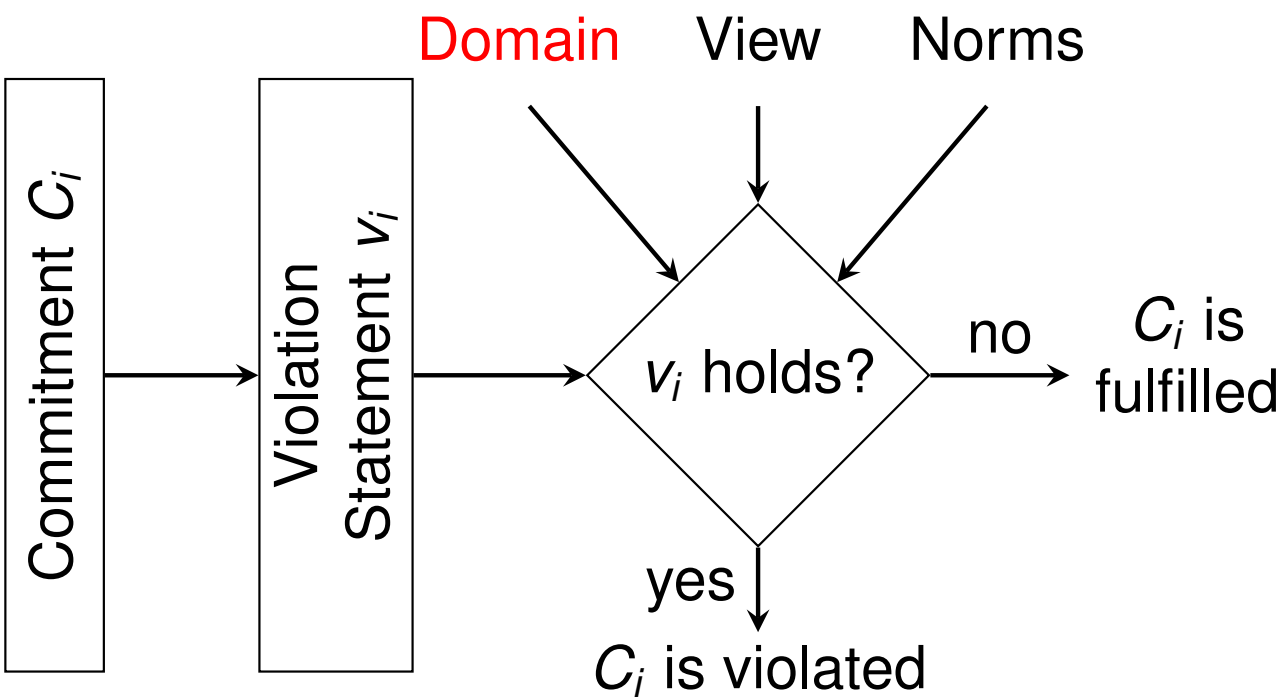
C_1(:osn; :dennis; *isFriendOf*(:dennis,*X*), *sharesPost*(:dennis,*P*), MediumPost(P); *canSeePost*(*X*,*P*))
Friends of Dennis are allowed to see medium posts of Dennis
C_2(:osn; :dennis; *isFriendOf*(:dennis,*X*), *sharesPost*(:dennis,*P*), LocationPost(P); not(*canSeePost*(*X*,*P*)))
Friends of Dennis are not allowed to see location posts of Dennis

Violation Statements

- A violation occurs when the *debtor* fails to bring about the *condition* of a commitment.
- We identify violation statements according to the commitments.
- In a commitment, the *condition* is true if the *antecedent* is true that can be represented as the rule: *precondition* $\rightarrow$ *condition*.
- A violation statement is modeled as the negation of this rule:
  violation: *precondition*,not(*condition*)

C_1(:osn; :dennis; *isFriendOf*(:dennis,*X*), *sharesPost*(:dennis,*P*), MediumPost(P); *canSeePost*(*X*,*P*))
v_1: *isFriendOf*(:dennis,*X*), *sharesPost*(:dennis,*P*), MediumPost(P), not(*canSeePost*(*X*,*P*))

The Social Network Domain



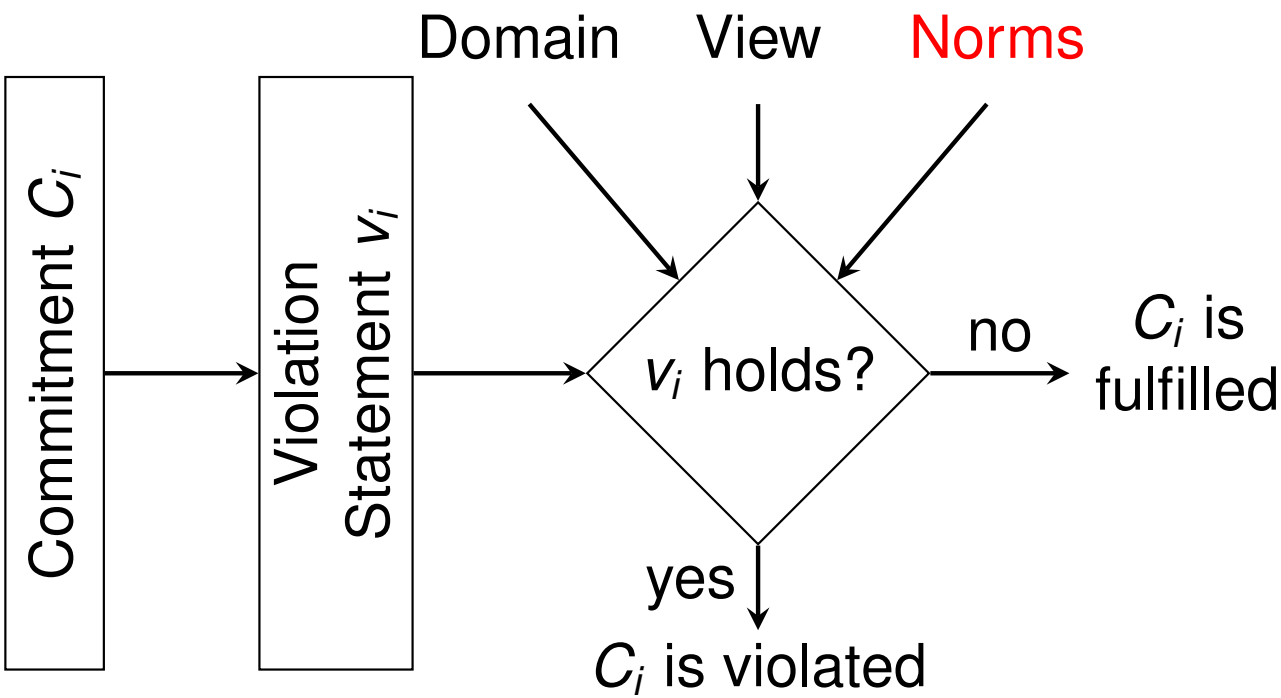
The Social Network Domain: Axioms

$\texttt{Agent}, \texttt{Post}, \texttt{Audience}, \texttt{Context}, \texttt{Content} \sqsubseteq \top$	$\texttt{Leisure}, \texttt{Meeting}, \texttt{Work} \sqsubseteq \texttt{Context}$
$\texttt{Beach}, \texttt{EatAndDrink}, \texttt{Party}, \texttt{Sightseeing} \sqsubseteq \texttt{Leisure}$	$\texttt{Bar}, \texttt{Cafe}, \texttt{College}, \texttt{Museum}, \texttt{University} \sqsubseteq \texttt{Location}$
$\texttt{Picture}, \texttt{Video} \sqsubseteq \texttt{Medium}$	$\texttt{Medium}, \texttt{Text}, \texttt{Location} \sqsubseteq \texttt{Content}$
$\texttt{Post} \sqcap \exists \mathit{sharesPost}^{-}.\texttt{Agent} \equiv \exists \mathit{R\_sharedPost}.\mathit{Self}$	$\texttt{LocationPost} \equiv \exists \mathit{R\_locationPost}.\mathit{Self}$
$\texttt{LocationPost} \equiv \texttt{Post} \sqcap \exists \mathit{hasLocation}.\texttt{Location}$	$\texttt{MediumPost} \equiv \texttt{Post} \sqcap \exists \mathit{hasMedium}.\texttt{Medium}$
$\texttt{TaggedPost} \equiv \texttt{Post} \sqcap \exists \mathit{isAbout}.\texttt{Agent}$	$\texttt{TextPost} \equiv \texttt{Post} \sqcap \exists \mathit{hasText}.\texttt{Text}$

The Social Network Domain: Axioms

Role Inclusions	Role Restrictions
$\mathit{canSeePost} \sqsubseteq U_a$	$\exists \mathit{canSeePost}.\top \sqsubseteq \texttt{Agent}, \top \sqsubseteq \forall \mathit{canSeePost}.\texttt{Post}$
$\mathit{hasAudience} \sqsubseteq U_a$	$\exists \mathit{hasAudience}.\top \sqsubseteq \texttt{Post}, \top \sqsubseteq \forall \mathit{hasAudience}.\texttt{Audience}, \top \sqsubseteq \leq 1\, \mathit{hasAudience}.\top$
$\mathit{hasGeotag} \sqsubseteq U_a$	$\exists \mathit{hasGeotag}.\top \sqsubseteq \texttt{Medium}, \top \sqsubseteq \forall \mathit{hasGeotag}.\texttt{Location}, \top \sqsubseteq \leq 1\, \mathit{hasGeotag}.\top$
$\mathit{hasLocation} \sqsubseteq U_a$	$\exists \mathit{hasLocation}.\top \sqsubseteq \texttt{Post}, \top \sqsubseteq \forall \mathit{hasLocation}.\texttt{Location}, \top \sqsubseteq \leq 1\, \mathit{hasLocation}.\top$
$\mathit{hasMedium} \sqsubseteq U_a$	$\exists \mathit{hasMedium}.\top \sqsubseteq \texttt{Post}, \top \sqsubseteq \forall \mathit{hasMedium}.\texttt{Medium}$
$\mathit{hasMember} \sqsubseteq U_a$	$\exists \mathit{hasMember}.\top \sqsubseteq \texttt{Audience}, \top \sqsubseteq \forall \mathit{hasMember}.\texttt{Agent}$
$\mathit{isAbout} \sqsubseteq U_a$	$\exists \mathit{isAbout}.\top \sqsubseteq \texttt{Post}, \top \sqsubseteq \forall \mathit{isAbout}.\texttt{Agent}$
$\mathit{isConnectedTo} \sqsubseteq U_a$	$\exists \mathit{isConnectedTo}.\top \sqsubseteq \texttt{Agent}, \top \sqsubseteq \forall \mathit{isConnectedTo}.\texttt{Agent}, \mathit{isConnectedTo} \equiv \mathit{isConnectedTo}^-$
$\mathit{isFriendOf} \sqsubseteq \mathit{isConnectedTo}$	$\exists \mathit{isFriendOf}.\top \sqsubseteq \texttt{Agent}, \top \sqsubseteq \forall \mathit{isFriendOf}.\texttt{Agent}, \mathit{isFriendOf} \equiv \mathit{isFriendOf}^-$
$\mathit{taggedPerson} \sqsubseteq U_a$	$\exists \mathit{taggedPerson}.\top \sqsubseteq \texttt{Medium}, \top \sqsubseteq \forall \mathit{taggedPerson}.\texttt{Agent}$

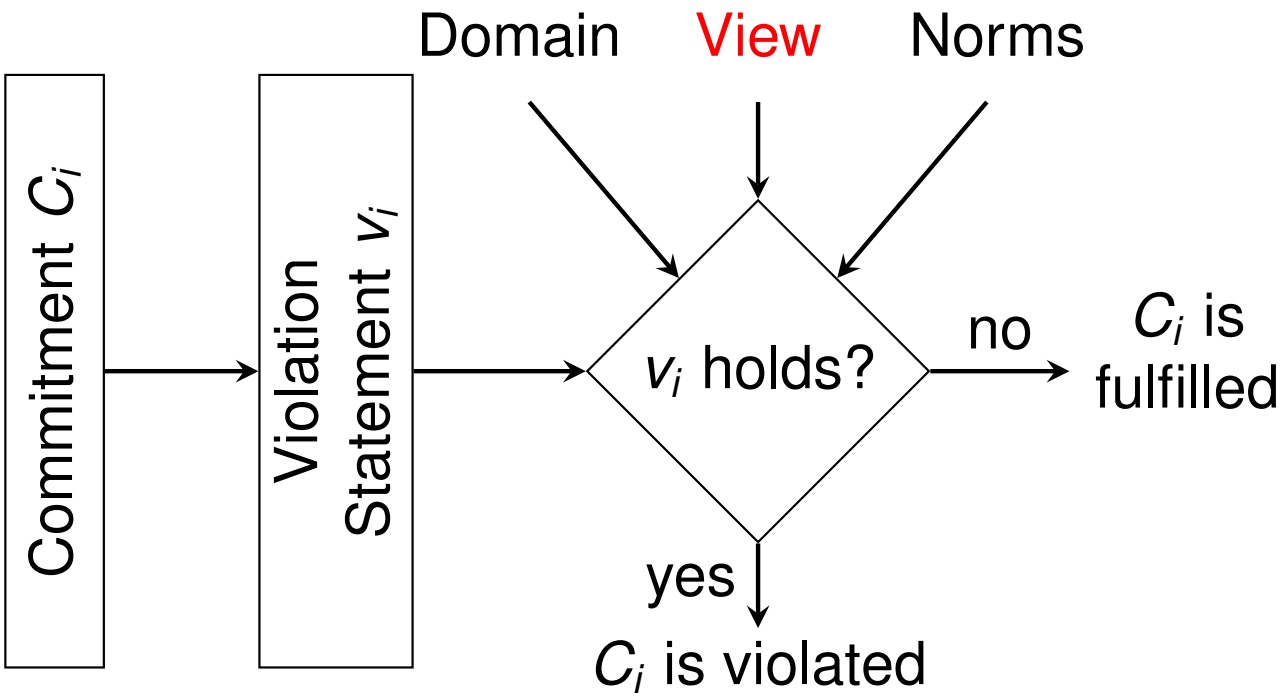
Norms


Domain
View
Norms
Commitment C_i
Violation Statement v_i
v_i holds?
no
C_i is fulfilled
yes
C_i is violated

Norms

N_1: $sharesPost(X,P) \rightarrow canSeePost(X,P)$
[Agent can see the posts that it shares.]
N_2: $sharesPost(X,P) \wedge hasAudience(P,A) \wedge hasMember(A,M) \rightarrow canSeePost(M,P)$
[Audience of a post can see the post.]
N_3: $hasMedium(P,M) \wedge taggedPerson(M,X) \rightarrow isAbout(P,X)$
[Post is about agents tagged in a medium.]
N_4: $\texttt{Post}(P) \wedge hasMedium(P,M) \wedge hasGeotag(M,T) \rightarrow \texttt{LocationPost}(P)$
[Geotagged medium gives away the location.]

View


Domain
View
Norms
Commitment C_i
Violation Statement v_i
v_i holds?
no
C_i is fulfilled
yes
C_i is violated

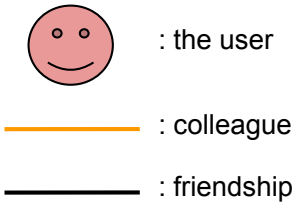
View

- ABSN view captures a given state of the network.

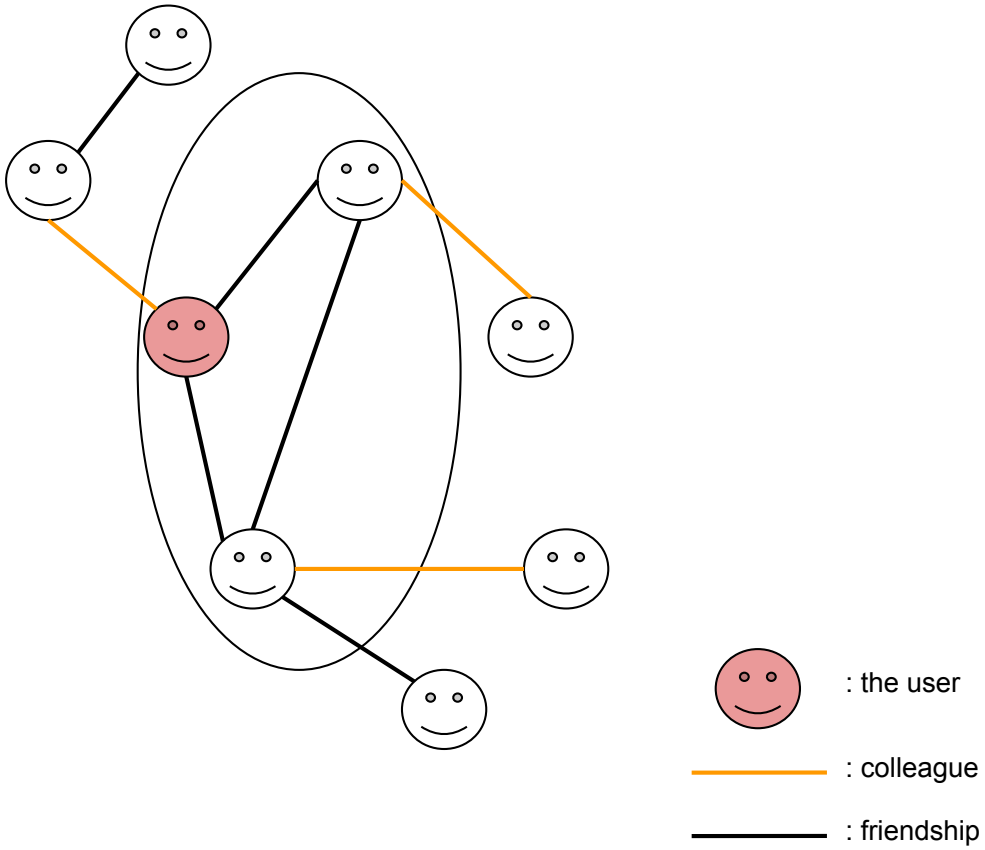
Table: Charlie shares a post `:pc1`

ClassAssertion(`Agent :alice`)	ClassAssertion(`Agent :bob`)
ClassAssertion(`Agent :charlie`)	ClassAssertion(`Agent :dennis`)
ClassAssertion(`Agent :eve`)	ClassAssertion(`Audience :audience`)
ClassAssertion(`Post :pc1`)	ClassAssertion(`Picture :pictureConcert`)
ObjectPropertyAssertion(*isFriendOf* `:alice :bob`)	ObjectPropertyAssertion(*isFriendOf* `:alice :charlie`)
ObjectPropertyAssertion(*isFriendOf* `:bob :charlie`)	ObjectPropertyAssertion(*isFriendOf* `:charlie :dennis`)
ObjectPropertyAssertion(*isFriendOf* `:dennis :eve`)	
ObjectPropertyAssertion(*sharesPost* `:charlie :pc1`)	ObjectPropertyAssertion(*hasAudience* `:pc1 :audience`)
ObjectPropertyAssertion(*hasMedium* `:pc1 :pictureConcert`)	ObjectPropertyAssertion(*taggedPerson* `:pictureConcert :alice`)
ObjectPropertyAssertion(*hasMember* `:audience :alice`)	ObjectPropertyAssertion(*hasMember* `:audience :dennis`)
ObjectPropertyAssertion(*hasMember* `:audience :eve`)	ObjectPropertyAssertion(*hasMember* `:audience :bob`)

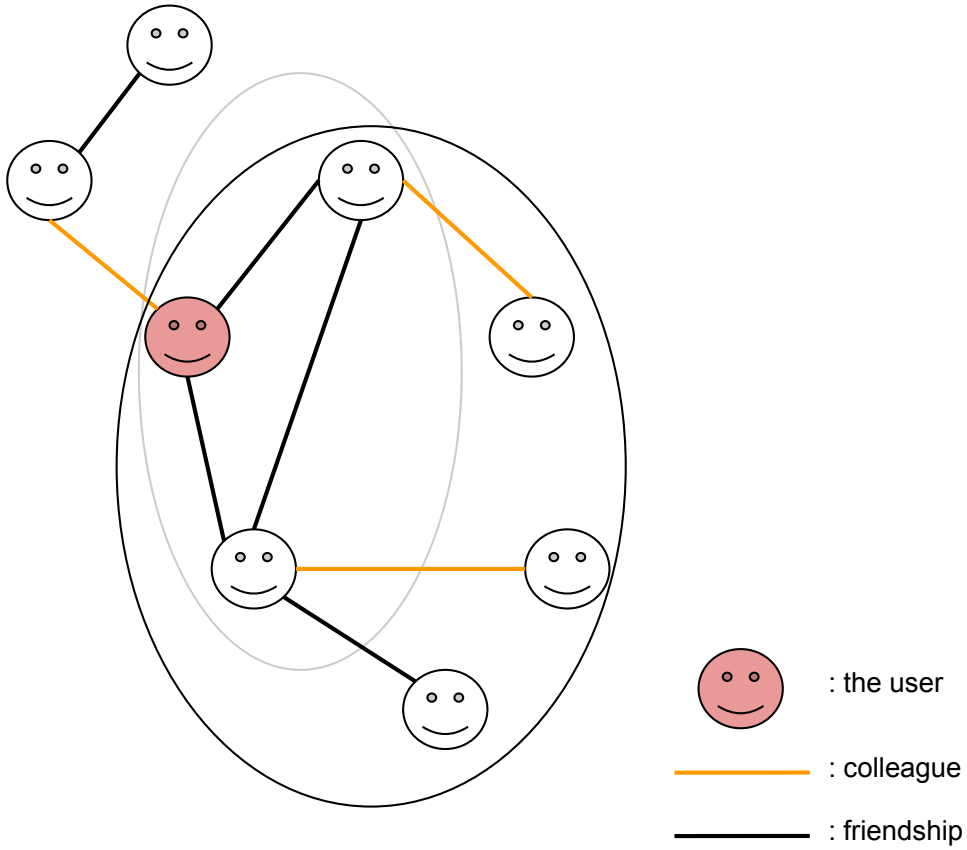
Views



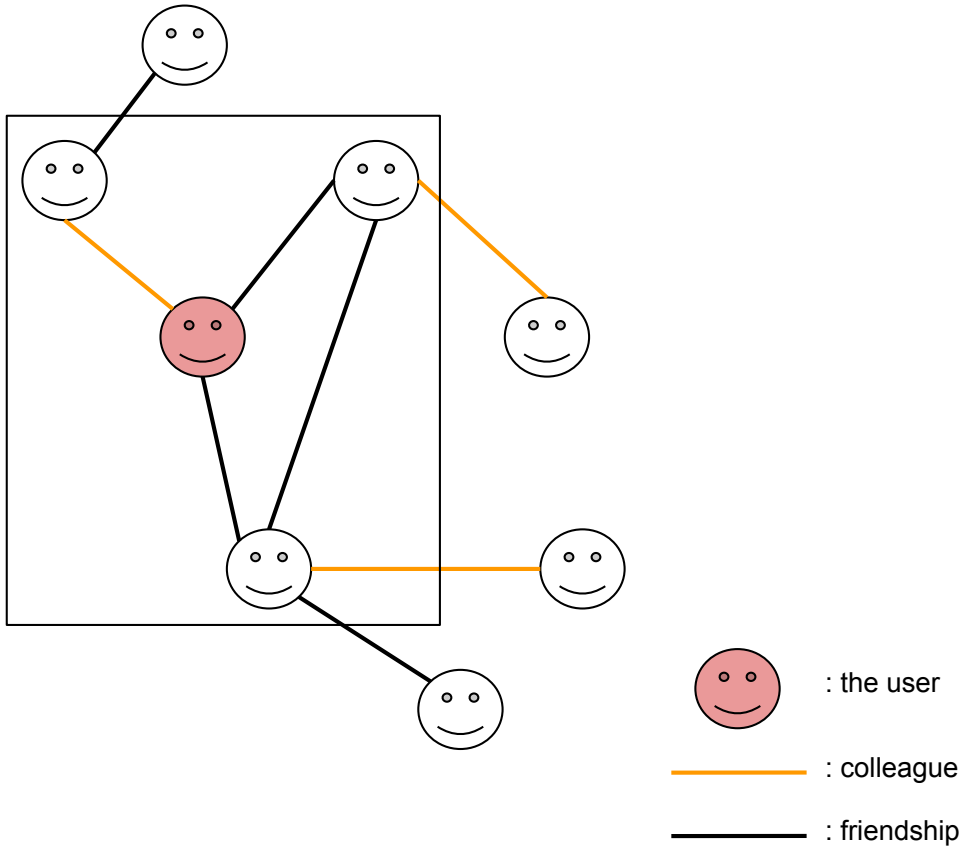
Views



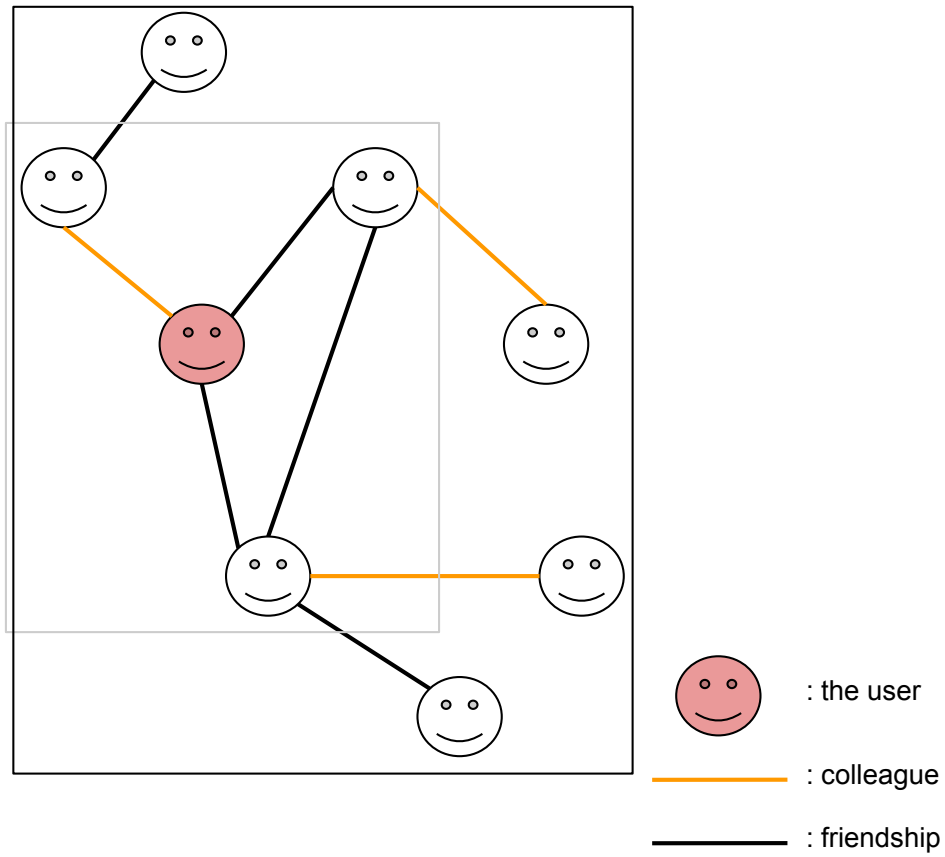
Views



Views



Views



Used to extend the current view. At the final extension, we have the *global view*.

Detection Algorithm

Algorithm 1: DEPTHLIMITEDDETECTION (C, m=MAX)

Input: C, the commitment to be checked
Input: m, the maximum number of iterations
Output: V, the set of privacy violations
Data: KB, the knowledge base (domain + norms)

$S \leftarrow \mathsf{initView}(C.creditor)$;
$V \leftarrow \{\}$, $iterno \leftarrow 0$;
$vstatement \leftarrow C.antecedent, not(C.consequent)$;
while $iterno < m$ **do**
    $KB \leftarrow \mathsf{updateKB}(KB, S)$;
    $V \leftarrow V \cup \mathsf{checkViolations}(KB, vstatement)$;
    $iterno \leftarrow iterno + 1$;
    if $V = \{\}$ **then**
        $S \leftarrow \mathsf{extendView}(S)$;
    else
        return V;
return V;

Theoretical Results

Theorem (Soundness)

Given an ABSN that is correctly represented with a KB, and a commitment C that represents a privacy requirement $PR^t_{a,i}$*, if* DEPTHLIMITEDDETECTION *returns a violation, then* isViolated*(*$PR^t_{a,i}$*,ABSN) holds.*

Theoretical Results

Theorem (Soundness)

Given an ABSN that is correctly represented with a KB, and a commitment C that represents a privacy requirement $PR^t_{a,i}$*, if* DEPTHLIMITEDDETECTION *returns a violation, then* isViolated*(*$PR^t_{a,i}$*,ABSN) holds.*

Proof: Assume that DEPTHLIMITEDDETECTION detects a violation, which is not true. This may occur only if one of the following holds:

- S contains incorrect information.

Theoretical Results

Theorem (Soundness)

Given an ABSN that is correctly represented with a KB, and a commitment C that represents a privacy requirement $PR^t_{a,i}$*, if* DEPTHLIMITEDDETECTION *returns a violation, then* isViolated*(*$PR^t_{a,i}$*,ABSN) holds.*

Proof: Assume that DEPTHLIMITEDDETECTION detects a violation, which is not true. This may occur only if one of the following holds:

- S contains incorrect information.
- KB does not contain the necessary information.

Theoretical Results

Theorem (Soundness)

Given an ABSN that is correctly represented with a KB, and a commitment C that represents a privacy requirement $PR^t_{a,i}$*, if* DEPTHLIMITEDDETECTION *returns a violation, then* isViolated*(*$PR^t_{a,i}$*,ABSN) holds.*

Proof: Assume that DEPTHLIMITEDDETECTION detects a violation, which is not true. This may occur only if one of the following holds:

- S contains incorrect information.
- KB does not contain the necessary information.
- *vstatement* is computed incorrectly so that it does not reflect a privacy violation.

Completeness

Theorem (Completeness)

Given a commitment C, DEPTHLIMITEDDETECTION *always returns a privacy violation, if one exists.*

Completeness

Theorem (Completeness)

Given a commitment C, DEPTHLIMITEDDETECTION *always returns a privacy violation, if one exists.*

Lemma

Given a violation statement of a commitment v_i *and a knowledge base KB, if there is a privacy violation in KB,* checkViolations *returns it.*

Completeness

Theorem (Completeness)

Given a commitment C, DEPTHLIMITEDDETECTION *always returns a privacy violation, if one exists.*

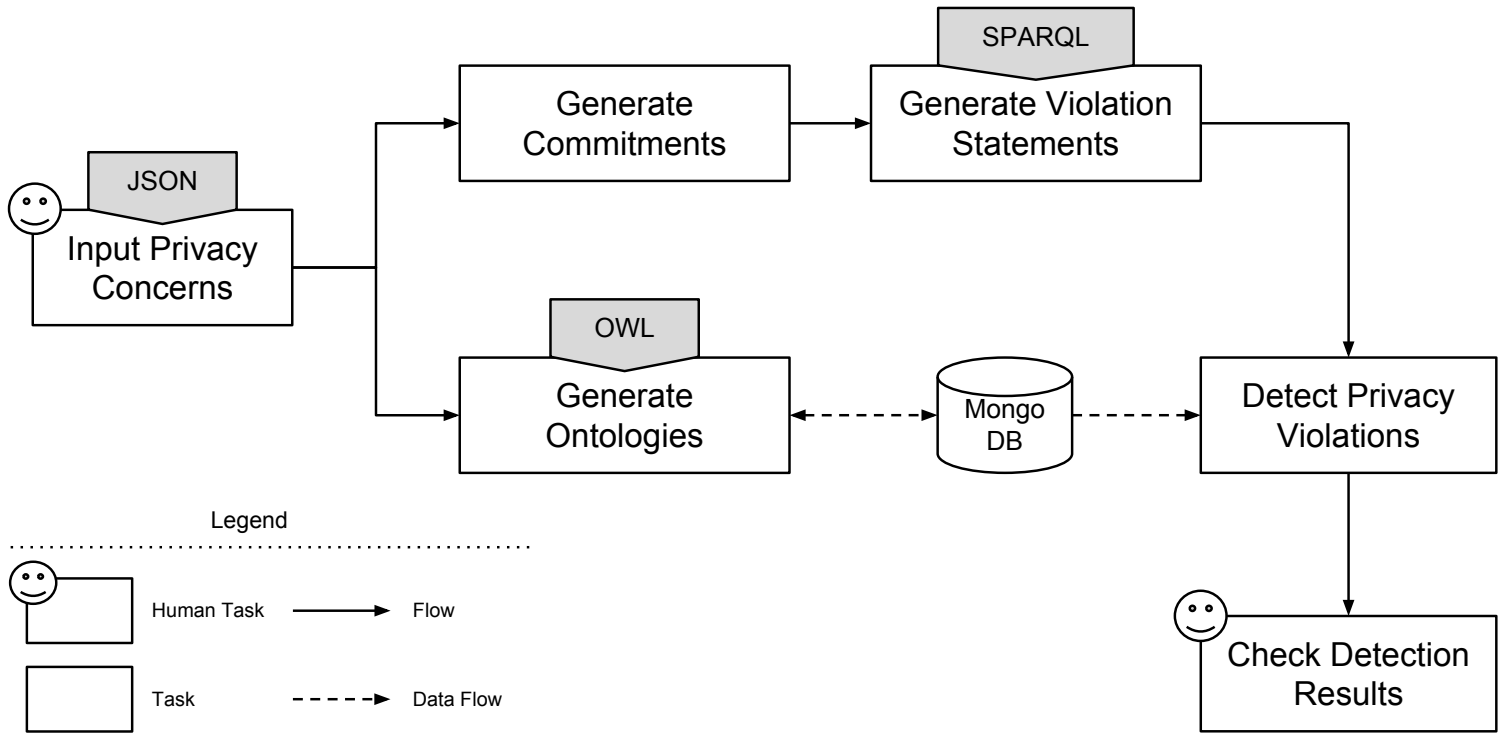
Lemma

Given a violation statement of a commitment v_i *and a knowledge base KB, if there is a privacy violation in KB,* checkViolations *returns it.*

Lemma

extendView *can eventually create the global view.*

A Facebook Application: PriGuardTool[2]



[2] Nadin Kökciyan and Pınar Yolum. "PriGuardTool: A Web-Based Tool to Detect Privacy Violations Semantically". In: *Engineering Multi-Agent Systems: 4th International Workshop, EMAS 2016, Singapore, Singapore, May 9-10, 2016, Revised, Selected, and Invited Papers*. Ed. by Matteo Baldoni et al. Springer International Publishing, 2016, pp. 81–98.

Running Example

Dennis wants his friends to see his pictures but not his location. He posts a picture without declaring his location. However, it turns out that his picture is geotagged.

C_1(:osn, :dennis, isFriendOf(:dennis, X), isAbout(P, :dennis), LocationPost(P), not(canSeePost(X,P)))

V_1 - :osn, :dennis, isFriendOf(:dennis, X), isAbout(P, :dennis), LocationPost(P), canSeePost(X,P))

```
SELECT ?x ?p WHERE {
?x osn:isFriendOf osn:dennis .
?p osn:isAbout osn:dennis .
?p rdf:type osn:LocationPost .
FILTER EXISTS (?x osn:canSeePost ?p) }
```

PRIGUARD: Performance Results

ABSN	depth=0	depth=1	depth=2	G
(#$\mathcal{A}$,#$\mathcal{R}$)	(1,0)	(39,412)	(535,5347)	(535,5347)
G_1: #Axioms	2175	4267	29959	29959
Time	3ms	4.74ms	30.19ms	29.79ms
(#$\mathcal{A}$,#$\mathcal{R}$)	(1,0)	(51,579)	(1035,27783)	(1035,27783)
G_2: #Axioms	2175	5079	125703	125703
Time	2.96ms	5.49ms	123.95ms	122.46ms
(#$\mathcal{A}$,#$\mathcal{R}$)	(1,0)	(123,4199)	(1046,27795)	(4039,88234)
G_3: #Axioms	2175	20423	125883	403555
Time	3.09ms	18.01ms	121.15ms	530.01ms
(#$\mathcal{A}$,#$\mathcal{R}$)	(1,0)	(37,235)	(848,8543)	(60001,728596)
G_4: #Axioms	2175	3535	46463	3636547
Time	3.07ms	4.13ms	47.09ms	18397.26ms
(#$\mathcal{A}$,#$\mathcal{R}$)	(1,0)	(157,2669)	(2787,74217)	(65328,1435168)
G_5: #Axioms	2175	14711	332463	6526759
Time	3.11ms	19.03ms	406.91ms	25890.27ms