Computer-aided Privacy Analysis of Functional Requirements with ProPAn

Rene Meis

paluno - The Ruhr Institute for Software Technology
University of Duisburg-Essen, Germany

February 18, 2016
Research field: Privacy-aware requirements engineering

Research questions:

- How to derive privacy requirements from regulations and standards?
- What kinds of privacy requirements exist?
- How to identify privacy-relevant parts of the software?
- Which additional knowledge is needed for a privacy analysis?
- How to operationalize privacy requirements?
- How to test or verify that a software satisfies its privacy requirements?
Introduction

Research field: Privacy-aware requirements engineering

Research questions:

• How to derive privacy requirements from regulations and standards?
• What kinds of privacy requirements exist?
• How to identify privacy-relevant parts of the software?
• Which additional knowledge is needed for a privacy analysis?
• How to operationalize privacy requirements?
• How to test or verify that a software satisfies its privacy requirements?
The ProPAn method provides a tool supported privacy analysis of a set of functional requirements that are represented as problem diagrams [Jackson, 2001]. ProPAn aims at the identification of:

- **Privacy-relevant domain knowledge** [Meis, 2014, Beckers et al., 2014a]
- **Potential privacy concerns** [Beckers et al., 2014b]
- **Personal data processed by the system-to-be** [Meis and Heisel, 2015]
- **Flow of personal data in the system** [Meis and Heisel, 2015]
- **Privacy requirements** [Meis et al., 2015]
eHealth Scenario
Provided by industrial partners of the EU project NESSoS

Concerns the management and usage of Electronic Health Records (EHRs).

**R1** Doctors shall be able to create and modify EHRs.

**R2** Doctors shall be able to browse EHRs.

**R3** The accounting of patients shall be performed using an insurance application based on the EHRs. If necessary invoices shall be created.

**R4** The billing of patients shall be performed using a financial application based on the invoices.

**R5** Appointments, instructions and alarms shall be sent to the mobile devices of patients based on the EHRs.

**R6** Vital signs of patients sent via their mobile devices shall be recorded in the EHRs.
Problem Frames Approach

[Jackson, 2001]

- **System** consists of the machine and its environment. The environment is structured into domains.
- **Requirements** are represented in problem diagrams.
- Requirements refer to and constrain phenomena of domains.
- Phenomena can be *causal* or *symbolic*.

![Diagram]
All steps make use of one central UML model.

All steps are partly automated by the ProPAn tool\(^1\)

\(^1\)http://www.uml4pf.org/ext-propan/index.html
Overview of our Method

Identification of Personal Data

Privacy Context Elicitation

Functional Requirements
Domain Knowledge

Identification of Personal Data

Functional Requirements
Domain Knowledge
Personal Data of Stakeholders
Detailed Stakeholder Information Flow Graphs

ProPAn Model

«datastore»

Personal Data Flow Analysis

Generate Privacy Requirements

Flow of Personal Data
Personal Data at Domains

Detailed Stakeholder Information Flow Graphs

Personal Data at Domains

Personal Data of Stakeholders
Statements (requirements, facts, and assumptions) imply possible information flows from referred to to constrained domains.
Graph Generation
Aggregation of all Information Flows starting at Patient
Relations that are elicited and documented

- Symbolic phenomena are used to represent personal data
- Biddable domains are persons
Candidates for personal data can be derived from the Detailed Stakeholder Information Flow Graph (DSIFG).

We distinguish two cases:

**Symbolic phenomena** can directly represent personal data.

**Causal phenomena** may transmit/contain personal data. This transmitted/contained personal data is elicited.
Identification of Personal Data

Identified personal data is modelled in **personal information diagrams**.

```plaintext
«personInformationDiagram»
PID Patient

«relatedTo»
collection=[indirect, direct, reused]
origin=[A1]
sensitive=true
linkability=group

«symbolicPhenomenon»
healthStatus

«relatedTo»
collection=[direct, reused]
origin=[A1, A6]
sensitive=true
linkability=group

«symbolicPhenomenon»
vitalSigns

«relatedTo»
collection=[direct]
origin=[A1]
sensitive=true
linkability=group

«symbolicPhenomenon»
mobileDevices

«relatedTo»
collection=[direct]
origin=[A4]
sensitive=true
linkability=group

«symbolicPhenomenon»
patientInsuranceContract

«relatedTo»
collection=[direct, reused]
origin=[A1, A4, A6]
sensitive=true
linkability=single

«symbolicPhenomenon»
patientDemographics

«relatedTo»
collection=[direct]
origin=[A4]
sensitive=true
linkability=group

«symbolicPhenomenon»
healthInformation

«biddableDomain»
Patient
```
Overview of our Method
Identification of Personal Data

Privacy Context Elicitation

Identification of Personal Data

ProPAn Model

Generate Privacy Requirements

Personal Data Flow Analysis

Functional Requirements
Domain Knowledge

Detailed Stakeholder Information Flow Graphs

Personal Data of Stakeholders

Flow of Personal Data
Personal Data at Domains
Personal Data Flow Analysis
Principle idea of a step in the analysis

1. Select a statement in the DSIFG that has an input domain at which personal data is available.
2. Decide which of the available personal data flows to the output domains of the statement.
3. Repeat until all statements have been considered.

Diagram:
- **Assumption** A1
  - Input: {patientInsuranceContract, requestInsuranceContract}
  - Output: {vitalSigns, patientDemographics}
  - {mobileDevices, healthStatus, patientDemographics}

- **Assumption** A4
  - Input: {provideInsuranceContract, insuranceContracts}
  - Output: {createEHR, modifyEHR}

- **Requirement** R1
  - Input: {knowledgeAboutPatient}

- **Requirement** R2
  - Input: {browseEHR}
  - Output: {readEHR}

- **Requirement** R5
  - Input: {sendVitalSigns}
  - Output: {showInstructions, showAlarm, showAppointment}

- **CausalDomain, connection**
  - {healthRecords}

- **LexicalDomain, design...**
  - EHR

- **BiddableDomain**
  - Patient
  - Doctor
  - MobileDevice
Relations that are elicited and documented

- Symbolic phenomena are used to represent personal data
- Biddable domains are persons
View on the resulting personal information diagram for the Patient.
View on the resulting available information diagram for the Insurance Application.
View on the resulting available information diagram for the Insurance Application.
Overview of our Method
Generate Privacy Requirements
Privacy Protection Goals [Hansen et al., 2015]
Generation of Unlinkability Requirements

Undetectability:

*The counterstakeholders shall not be able to sufficiently distinguish whether the personal information phenomena of the stakeholder exists or not.*
Generation of Unlinkability Requirements

Example of an Undetectability Requirement

```
<table>
<thead>
<tr>
<th>stakeholder = Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>counterstakeholder = [Insurance Employee]</td>
</tr>
<tr>
<td>phenomena = [healthStatus, mobileDevices, deviceID, vitalSigns, notes]</td>
</tr>
</tbody>
</table>
```

```
«UndetectabilityRequirement»
Undetectability_Patient_InsuranceEmployee
stakeholder = Patient
counterstakeholder = [Insurance Employee]
phenomena = [healthStatus, mobileDevices, deviceID, vitalSigns, notes]
```
Data Unlinkability:

For each pair of personal information pairs of the stakeholder, the counterstakeholders shall at most be able to link instances of the two elements of the pair to each other with linkability linkability.
Generation of Unlinkability Requirements

Example of a Data Unlinkability Requirement

```plaintext
«DataUnlinkabilityRequirement»
DataUnlinkability_Patient_InsuranceEmployee

stakeholder = Patient
counterstakeholder = [Insurance Employee]
pairs = [(treatment, diagnosis), ..., (patientBillingContact, healthInformation)]
linkability = single
```
Anonymity:

The counterstakeholders shall at most be able to link the personal information phenomena to the stakeholder with linkability linkability.
Generation of Unlinkability Requirements
Example of an Anonymity Requirement

```
AnonymityPatient_InsuranceEmployee
stakeholder = Patient
counterstakeholder = [Insurance Employee]
phenomena = [treatment, diagnosis, ..., patientBillingContact, healthInformation]
linkability = single
```

Diagram showing the relationships and linkability of personal data in an application.
Conclusion

Our contributions are:

- A method that derives flows of personal information and privacy requirements from a requirements model.
- Representation of this information in a UML model.
- An extensible UML profile.
- Tool support that guides the application of the method.

Our future directions include:

- Generation of PIA reports based on the elicited information.
- Identification of privacy threats in the model.
- Integration of PETs that mitigate the privacy threats.
- Validation of our method, the tool support, and the outputs produced by our method.
Overview of the ProPAn Method
Privacy Context Elicitation

Privacy Context Elicitation ➔ Identification of Personal Data
- Functional Requirements ➔ Domain Knowledge ➔ Functional Requirements
- Domain Knowledge ➔ Personal Data of Stakeholders ➔ Detailed Stakeholder Information Flow Graphs

«datastore»
ProPAn Model

Generate Privacy Requirements ➔ Personal Data Flow Analysis
- Privacy Requirements ➔ Flow of Personal Data ➔ Personal Data at Domains
- Flow of Personal Data ➔ Personal Data at Domains ➔ Personal Data of Stakeholders ➔ Detailed Stakeholder Information Flow Graphs

Conclusion
The information provided by a functional requirements model does often not include information about:

- **indirect stakeholders** of whom data is processed
- **information flows outside** of the software-to-be
- **indirect counterstakeholders** who may have access to personal data

To systematically elicit this privacy-relevant information, we developed a method in previous work [Meis, 2014].


