1

Privacy by Design in Federated Identity Management

Interpreting Legal Privacy Requirements for FIM and Comparing Risk Mitigation Models

2015 International Workshop on Privacy Engineering – IWPE'15 - MAY 21, 2015 - SAN JOSE, CA

Rainer Hörbe, presenter
Identinetics GmbH, Austria
rh@identinetics.com

Walter Hötzendorfer, co-author
Centre for Computers and Law
University of Vienna, Austria
walter.hoetzendorfer@univie.ac.at

2

Overview

FIM usage: why, who, where?

FIM-related privacy risks

Motivation for this project

Approach

Findings

3

FIM Usage

Why	**Scalability**: registration cost **Interoperability**: attribute semantics, trust policies **Compliance**: Loss of control across many silos
Who	Independent entities with common interests. (Supply chains, government agencies, R&E institutions, enterprise group members, professional networks, markets with roaming agreements.)
Where	eduGAIN, airlines, defense supply chains, government extranets, G2C/G2B services, ..
Edge Cases	Mobile SIM, social networks, centralized (single IDP) federations.

FIM-Related Privacy Risks

4

Due to FIM:

- **Observability** of behavior by central instances
- **Linkabilty** by introducing common identifiers
- **Impersonation** by Identity/Credential Providers or because of weaknesses in SSO mechanism

Due to the lack of FIM with PbD

- **Linkabilty** by reusing identifying attributes
- **Impersonation** caused by password reuse

Privacy Risks Unrelated to FIM

5

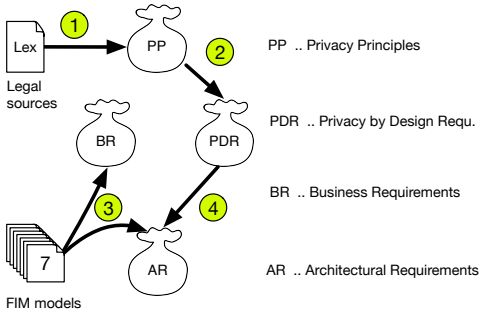
Linkability	Identifying contents across services Services integration/large privacy domains
Observability	Device fingerprinting IP-address
Impersonation	Weak endpoint security Poor crypto

Motivation and Scope

6

- FIM Projects featuring cross-sector federation
  (smart cities, citizen eIDs, B2B across supply chains)
- How to handle the increased privacy risk considering legal requirements, cost, complexity, convenience, feasibility?
- Scope limited on WebSSO use case
  (SAML, OpenID Connect)
- Focus on Observability and Linkability

Approach to Understand Requirements

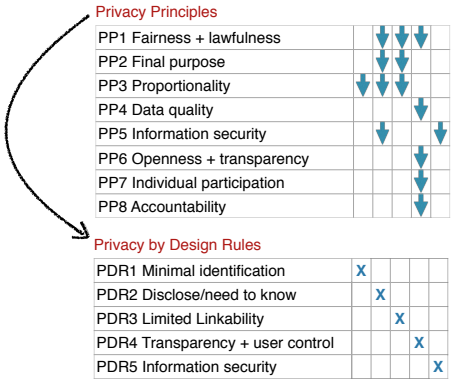


7

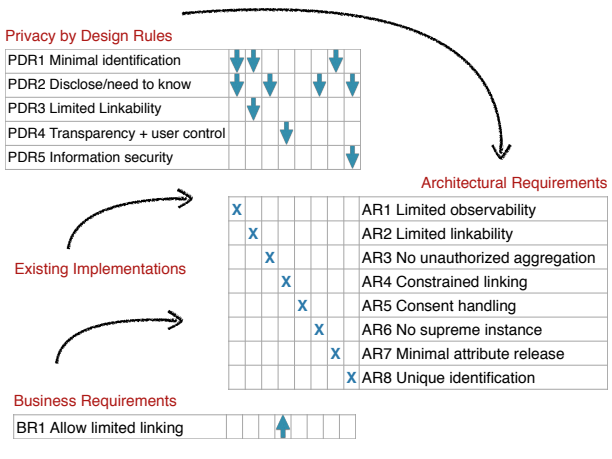
As shown in the diagram, we obtained input from two sides. From the top we took general and abstract privacy principles from privacy legislation and guidelines (1). Based on our knowledge from the field and the literature, we examined what these principles mean for the FIM domain, and how they can be accomplished by design requirements (2). The result was a set of privacy by design requirements that can be realized with technical controls.

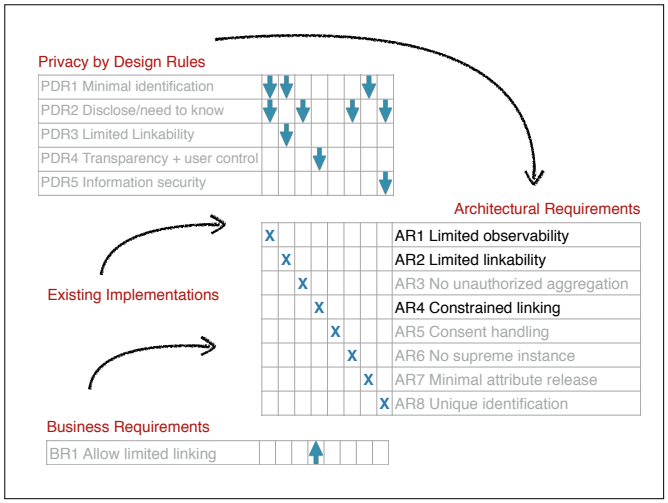
From the bottom we analyzed FIM models claiming improved privacy and recovered their architectural requirements (3). Finally, in an iterative process we joined both sides (4). We also took into account business requirements that are particularly relevant in our context. This resulted in a list of eight architectural privacy-related requirements for FIM systems.

8



9





10

The Problem Children

AR1 Limited observability
AR2 Limited linkability
AR3 No unauthorized aggregation
AR4 Constrained linking
AR5 Consent handling
AR6 No supreme instance
AR7 Minimized attribute release
AR8 Unique identification

- Organizational Controls
- Attribute-Based Credentials
- Late Binding
- Proxy Pool
- User-based IdPs
- Constrained Logging Proxy
- Blind Proxy

11

AR1. Limited observability (PDR1, PDR2). No entity shall be able to aggregate data about the usage of multiple services by users, which will keep it from being able to deduce personal interests or behavior.

AR2. Limited linkability (PDR1, PDR3). Relying parties shall not be able to aggregate personal data used in different privacy domains. Only if it is necessary for a legitimate purpose shall two relying parties processing data of a principal be able to link those data sets. Aside from unique identifiers, this concerns attributes that are identifying with high probability as well.

An important measure is the use of pseudonyms. If the full pseudonymization of user attributes is not feasible, then at least those attributes that identify a user (almost) uniquely shall be pseudonymized. This applies, e.g., to the ubiquitous e-mail address.

AR3. Prevent the unauthorized aggregation of attributes by central intermediaries such as

Models for Limited Observability:
(2) Attribute-Based Credentials

ABCs provide assertions to the RP without the IdP knowing the actual RPs.

Pro: Strong technical control.

Con: (a) No implementation in mainstream products; lack of deployment profiles for SAML or OpenID Connect; (b) IdP business model; (c) performance; (d) Increased complexity.

12

Models for Limited Observability:
(3) Late Binding/Federated Credentials

Credential-only federation (CSPs with brokers, or U2F tokens) rely on the separation between credential service assurance and identity assurance. Attributes are not released by the IdP, but obtained by the RP.

Pro: Straightforward architecture that goes well with existing technology based on common SAML profiles. Credential providers have only a minor privacy risk.

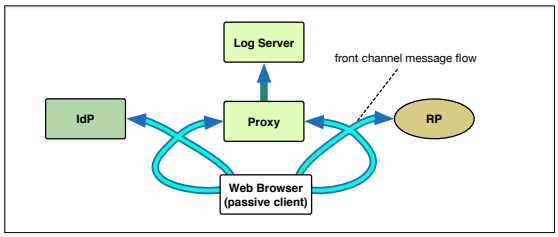
Con: (a) Less business value because attributes are collected per RP;
(b) Identifying attributes like name, residential and e-mail addresses could enable linking.

13

Credential providers provide pseudonymous credentials to users, and RPs will bind attributes to those credentials. This model was proposed by the Government of Canada Cyber-Authentication Architecture. Their separation between credential service assurance and identity assurance implies that attributes are not released by the IdP, but obtained by the RP.

Note: The IdP does store identity attributes for account recovery and non-repudiation.

Models for Limited Observability:
(6) Constrained Logging Proxy



The proxy stores log files only in a separate, well-protected system for a very limited time.

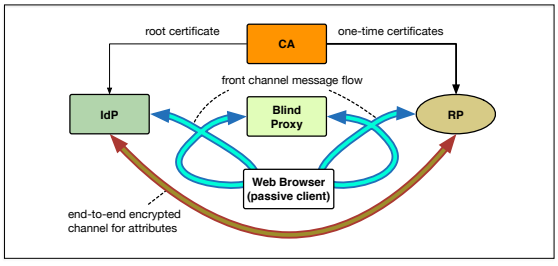
Pro: Has been implemented without changes to FIM protocols.

Con: While an adversary could cause only limited damage with a single data breach, a complete take-over of the proxy would compromise the privacy goal.

14

A proxy (hub) will hide the target RP from the IdP. The hub thus provides limited observability for IdPs, but is violating AR1 itself. To mitigate this, the gateway does not store log data on the local node, but sends it to a remote system where controls such as encryption and deletion after a short term reduce the risk of abuse.

Models for Limited Observability:
(7) Blind Proxy



Pro: It proposes reasonably strong technical control, works with any credential technology and is fairly easy to fit into hub-and-spoke federations.

Con: (a) Requires (small) extension to existing SAML and OIDC implementations. (b) It requires RPs to participate in a considerably large anonymity set.

15

The Privacy-enhanced FIM model introduced by the authors enhances the hub-and-spoke model by offering technical controls that enforce limited observability and enable pseudonymous authentication. Its core property is that attributes are encrypted from the IdP to the RP, but the IdP cannot identify the RP. This is shown in the picture, where a message exchange between RP and IdP is brokered via the blind proxy. As the RP's encryption certificate is issued per transaction, the IdP can only identify groups of RPs. This model claims to have similar properties as attribute-based credentials in option (2), except that it is not resistant against a collusion of RP and IdP.

The Problem Children

16

AR1 Limited observability
AR2 Limited linkability
AR3 No unauthorized aggregation
AR4 Constrained linking
AR5 Consent handling
AR6 No supreme instance
AR7 Minimized attribute release
AR8 Unique identification

Approaches for Limited Linkability Between Privacy Domains

17

- Unique Identifiers limited in scope:
  - Pairwise identifiers (IDP - RP)
  - Group or sector-specific identifiers
- Proxy attributes for identifying attributes:
  - Blind „reverse proxy“ for e-mail and jabber
  - User-selected pseudonyms for display names
  - Virtual credit cards, crypto-currencies for payments
  - PO-boxes etc. for physical shipment

The use of opaque, pairwise identifiers is known as a targeted identifier in research & education federations [13][14], as a sector-specific identifier in government eIDs [19], and called persistent NameId in the SAML specification [20]. This concept is fairly easy to implement and widespread.
However, the problem of linkability using other attributes remains. E-mail address, credit card number, delivery address and name are frequently required and match individuals with high probability. The privacy-enhanced FIM model [18] proposes the use of proxy addresses for email, payment and physical delivery and user-selected pseudonyms for display names.

The Problem Children

18

AR1 Limited observability
AR2 Limited linkability
AR3 No unauthorized aggregation
AR4 Constrained linking
AR5 Consent handling
AR6 No supreme instance
AR7 Minimized attribute release
AR8 Unique identification

Approaches for Constrained Linking (Between Privacy Domains)

- Types of link constraints:
  - A group of privacy domains (>=2)
  - By direction (i.e. unidirectional)
  - Temporal (e.g. until expiry or revocation)
- Examples:
  - Austrian eID with sector-specific identifiers encrypted for another sector's target application
  - Mediated links in a blind proxy model: All access via proxy is encrypted end-to-end, except the identifier that is mapped by the proxy.

19

Unidirectional links have been defined, for example in the Austrian eID using encrypted sector-specific identifiers. This concept uses sector-specific pairwise identifiers encrypted for the target application. On a more general level, constrained links can be direct or mediated. Direct links, as in the Austrian eID system, are durable, whereas mediated links can be established by a broker for a specific transaction only. The latter would allow, for example, that a user consents to use a payment clearing service a single time, without leaving a possibility for either service to link the personal data later on.

Conclusions

- Increased privacy risks introduced by FIM can be mitigated with technical controls.
- Effort to implement controls for limited observability varies with the strength of the controls.
- Limited likability with pairwise identifiers is current practice. However, identifying attributes are left out of the equation. There is room for improvement with moderate effort.

20

Blind Proxy Profiles & Implementations

- SAML PEFIM Profile
  https://kantarainitiative.org/confluence/x/-wIxB
- PEFIM Proxy reference implementation
  http://github.com/its-dirg/pefim-proxy_docker/
- PEFIM IDP & SP implementations
  - PySAML2
    https://github.com/its-dirg/pefim_sp
    https://github.com/its-dirg/pefim_idp
  - Shibboleth
    Will be available soon at shibboleth.net
  - OpenAM
    on request from cryptas.com

21