Privacy by Design in Federated Identity Management

Interpreting Legal Privacy Requirements for FIM and Comparing Risk Mitigation Models

EIC 2018, May 16

Rainer Hörbe
Identinetics GmbH, Austria
rh@identinetics.com

Identity/Authorization Requirements vs. Assertions/Attributes

	Demographic Data	Biometric	ABC	Directed Pseudonym	EMail
Legal, Safety	X	(X)			(X)
Group membership			X	(X)	(X)
Payment, delivery	X				(X)
Real names	Name				(X)
Seclusive pseudonym				X	(X)
Repetitive identity				X	(X)

Qualities: provenance, persistence, confidentiality

Models to Share Identity Information

Central - Federated - User-centric

Criteria: Who governs the policy

FIM Usage

Why	**Scalability**: registration cost **Interoperability**: attribute semantics, trust policies **Compliance**: Loss of control across many silos
How	Web/mobile SSO, Bridge-PKI, Trust Status Lists
Where	R&E, airlines, defense supply chains, government extranets, G2C/G2B services, ..
Edge Cases	Mobile SIM, social networks, centralized (single IDP) federations.

FIM-Related Privacy Risks[1]

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 PbD in FIM

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

[1] partially applies to other models of cross-domain IDM (centralized, user-centric) as well

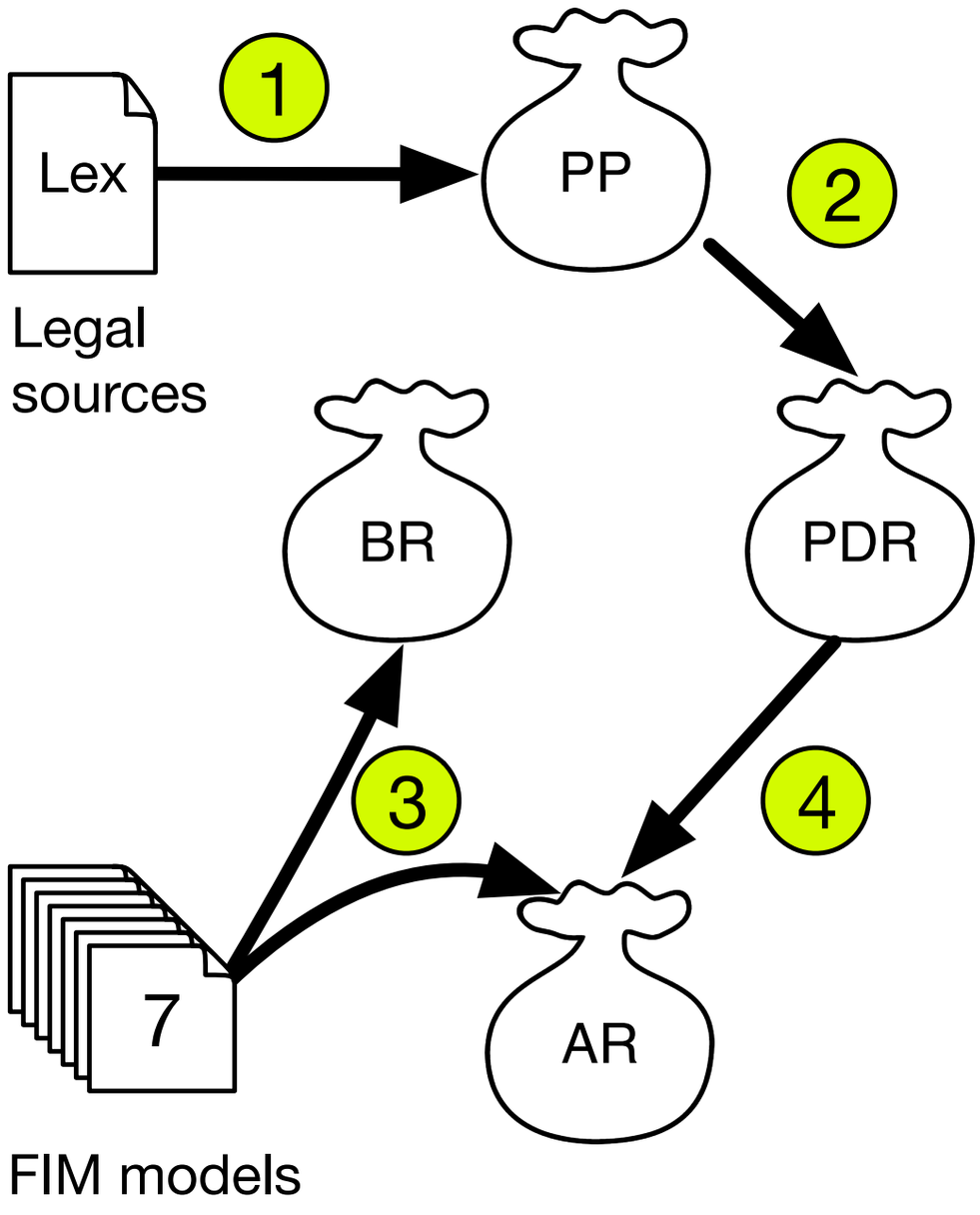
Privacy Risks Unrelated to FIM

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

Motivation and Scope

- 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



PP .. Privacy Principles

PDR .. Privacy by Design Requ.

BR .. Business Requirements

AR .. Architectural Requirements

Privacy Principles

PP1 Fairness + lawfulness		↓	↓	↓	
PP2 Final purpose		↓	↓		
PP3 Proportionality	↓	↓	↓		
PP4 Data quality				↓	
PP5 Information security		↓			↓
PP6 Openness + transparency				↓	
PP7 Individual participation				↓	
PP8 Accountability				↓	

Privacy by Design Rules

PDR1 Minimal identification	X				
PDR2 Disclose/need to know		X			
PDR3 Limited Linkability			X		
PDR4 Transparency + user control				X	
PDR5 Information security					X

Privacy by Design Rules

PDR1 Minimal identification	↓	↓					↓	
PDR2 Disclose/need to know	↓		↓			↓		↓
PDR3 Limited Linkability		↓						
PDR4 Transparency + user control				↓				
PDR5 Information security								↓

Architectural Requirements

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

Existing Implementations

Business Requirements

BR1 Allow limited linking				↑				

Privacy by Design Rules

PDR1 Minimal identification	↓	↓					↓	
PDR2 Disclose/need to know	↓		↓			↓		↓
PDR3 Limited Linkability		↓						
PDR4 Transparency + user control				↓				
PDR5 Information security								↓

Architectural Requirements

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

Existing Implementations

Business Requirements

BR1 Allow limited linking				↑				

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
Polymorphic Pseudonyms
Self-Sovrereign Identity

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) Slow uptake in mainstream products; lack of deployment profiles for SAML and OIDC; (b) IdP business model; (c) performance; (d) ~~Increased complexity~~.

Models for Limited Observability:

(3) Late Binding/Federated Credentials

Credential-only federation (CSPs with brokers, or U2F tokens) separate credential service and attribute assurance. Attributes are registered and vetted by the RP.

Pro: Straightforward architecture that goes well with existing technology based on common SAML profiles.

Con: (a) Less reuse/business value;
(b) Identifying attributes could enable linking.

Models for Limited Observability:
(4) Proxy Pool

Proxies that play RP to an IdP and IdP to an RP can significantly reduce the amount of data collection, if there are many of them operated by independent parties. Each proxy serving only a subset of RPs would not obtain the full profiles of all users.

Pro: This is a use of existing technology, because proxies and gateways for identity management are a well-established technology, e.g., part of the SAML specification (BR2).

Con: (a) Proxies would yield only a very limited improvement on AR1 until the number of proxies is quite large; thus, it would be difficult to overcome the hen-and-egg problem.

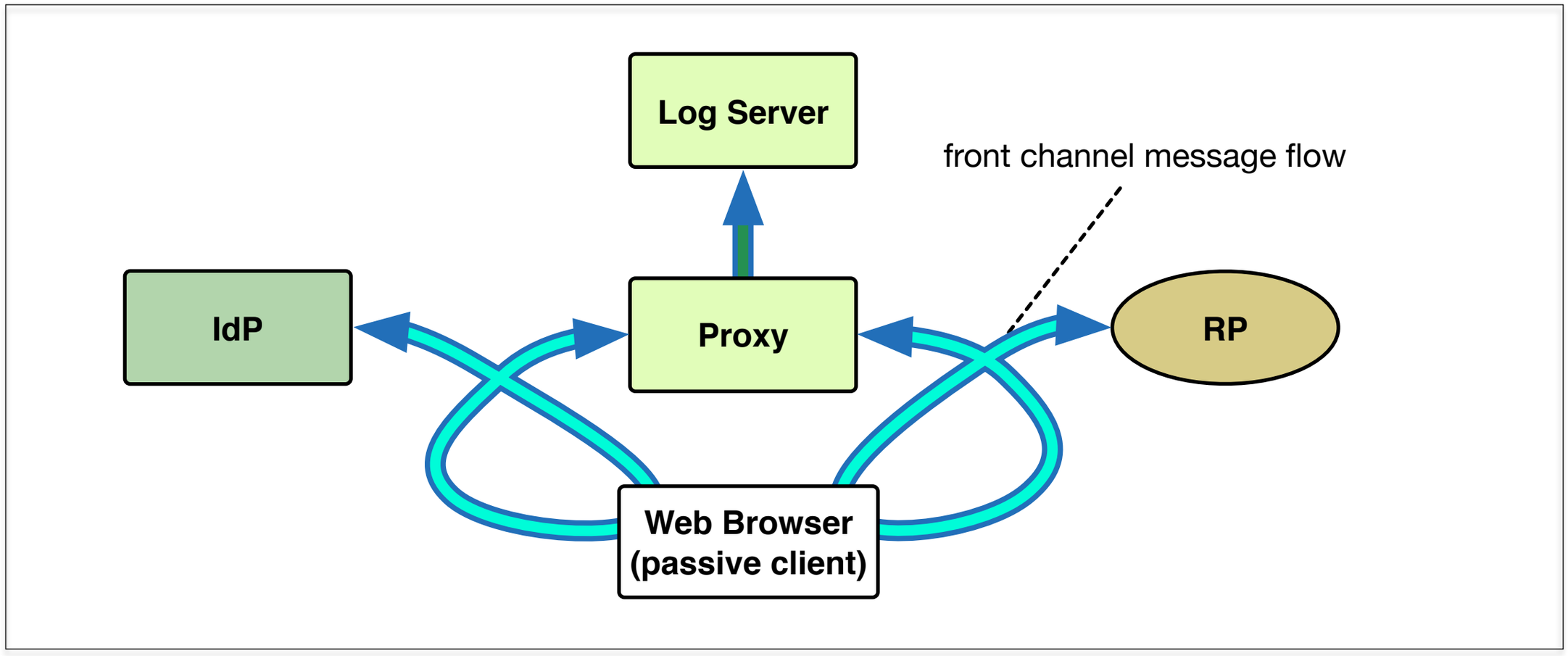
Models for Limited Observability: (5) User-based IdPs

As proposed by IMI [16], the client would be the identity selector and could also hold the credentials locally. A similar concept has been proposed with personal authentication devices [17].

Pro: This architecture provides good support for AR1 with the possible exception of (b) below.

Con: (a) Deployment is hard because it is difficult to enhance web browsers (BR3) and (b) with PKI-based credentials there is still the tracking issue with OCSP responders (AR1). (c) Experience with the “Neuer Personalausweis”, the German national identity card (described in [1]), showed that complex deployment leads to the growth of cloud services that offload some deployment issues but violate AR1 in turn.

Models for Limited Observability:
(6) Constrained Logging Proxy

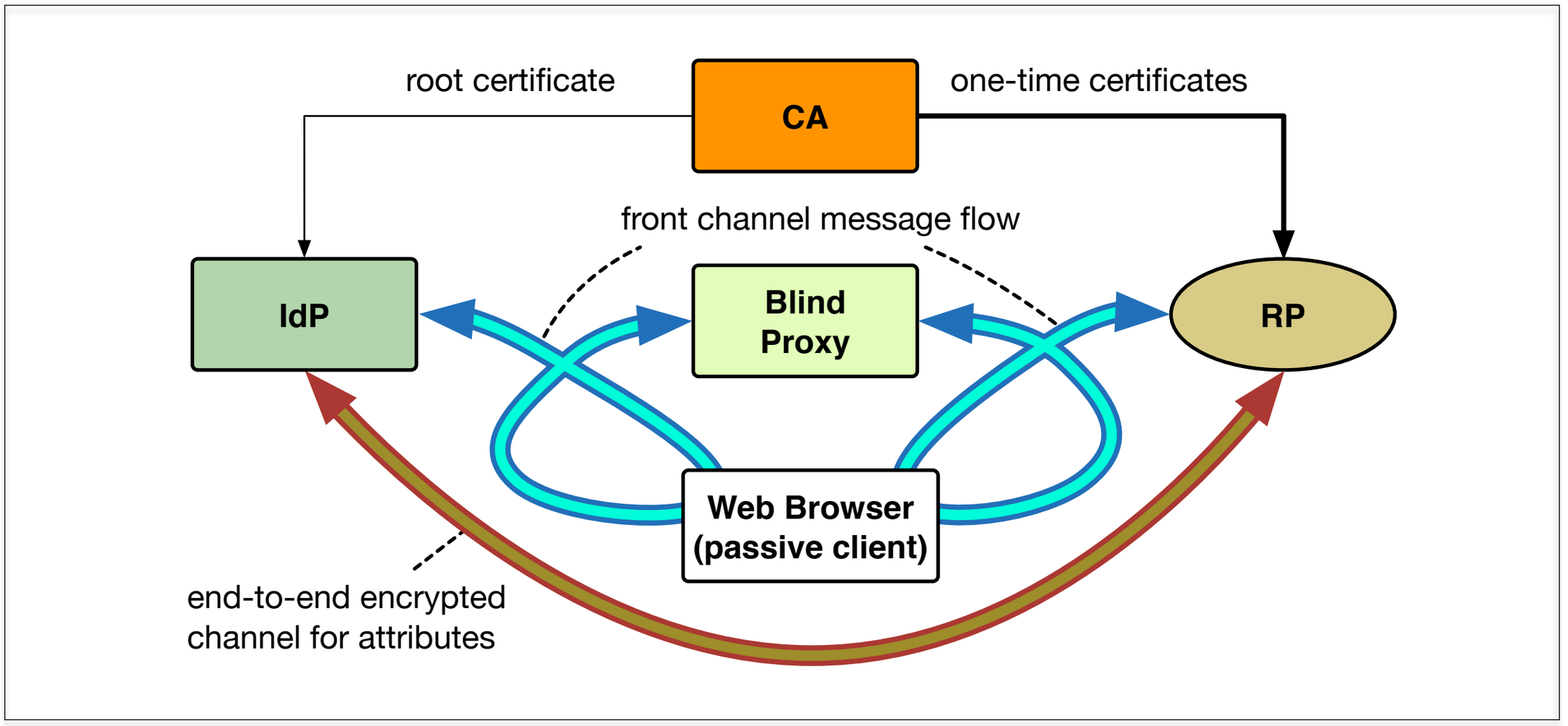


The proxy stores log files in a well-protected system for a short time.

Pro: Has been implemented without changes to FIM protocols.

Con: Fails to withstand a complete take-over of the proxy.

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

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

- 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, messaging
  - User-selected pseudonyms for display names
  - Virtual credit cards, crypto-currencies for payments
  - PO-boxes etc. for physical shipment

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.

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

Minimized Attribute Release & Pseudonymization

- Releasing only attributes required for the purpose of the service is well-established practice (e.g. R&E)
- However, certain attributes, predominantly e-mail, are highly identifying, allowing the RP to link up data from other sources. As a remedy, email-addresses can be directed.
- Same is possible for payment (virtual credit card) and physical delivery (P.O.Box, directed customer Id)

Targeted eMail Address

- Principle: The IDP will release pair-wise email addresses per SP (or privacy domain), e.g. for SAML that could be <persistentNameId@idp.example.org>
- E-mails to this address will be forwarded by the IDP using a reverse mapping scheme (persistent storage, or encryption/decryption scheme)

Conclusions

- The extent of privacy controls depends in a by-case assessment

- Limited observability: effort and strength vary

- Limited linkability: pairwise identifiers are current practice but identifying attributes thwart the effort.

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
  - SATOSA, PYSAML2
    https://github.com/identitypython/satosa
  - Shibboleth SP
  - OpenAM