Code of Conduct pilot 2012-2013

Contents

Pilot report

Approved by the pilot group meeting 12th Apr, 2013

About the Code of Conduct

Service Providers have reported that they have faced difficulties in receiving the necessary attributes from the SAML Identity Provider servers managed by Home Organisations, especially in interfederation setups such as eduGAIN. In European Union, Home Organisations' hesitation to release attributes is believed to follow from the EU Data protection directive, which imposes obligations to the Home Organisation on how they can process their end users' personal data and release them to third parties.

During 2012, REFEDS attribute release workgroup and the eduGAIN task of the GN3 project joined their forces to develop a Code of Conduct, which introduces behavioral rules for Service Providers which want to receive user attributes from the Identity Providers managed by the Home Organisations in EU/EEA. It is expected that Home Organisations are more willing to release attributes to Service Providers who manifest conformance to the Data protection Code of Conduct.

The Code of Conduct is published under a creative commons license and consists of several documents, including

- Code of Conduct for Service Providers in EU/EEA which defines the legal obligations imposed to a Service Provider
- SAML2 metadata profile and Entity Category definition, which introduces the technical mechanisms for Identity and Service Providers to exchange information about the commitment to the Code of Conduct
- informative supportive documents, such as document templates, good practices and guidelines

To get the approach generally approved, the Code of Conduct was exposed to two public consultations and presented in various meetings and workshops in the community during 2012.

About the pilot

To get practical experience on the Code of Conduct, a pilot was organized during the second half of 2012 together with four identity federations (DFN-AAI, Haka, Surffederatie and SWAMID) and CLARIN, the European e-research infrastructure project for language research. The pilot started with a kick-off meeting in Utrecht on the 25th June, 2012. The pilot team had also monthly videoconferences. The pilot was closed in the last videoconference on 12th April, 2013.

The technical development of the pilot systems started in November after the release of the second draft of the Code of Conduct for Service Providers document and the first draft of the SAML2 metadata profile and Entity Category specification. The first Service Provider manifesting compliance to the Code of Conduct was exposed to the eduGAIN SAML2 metadata on the 11th of December and the first CoC-enabled login was done on the 19th of December.

The following persons were involved in the pilot: Pål Axelsson (SWAMID/Uppsala University), Daan Broeder (MPI for Psycholinguistics), Joost van Dijk (Surfnet), Josh Howlett, (Janet), Janne Lauros (CSC), Mikael Linden (CSC), Valter Nordh, (SWAMID/University of Gothenburg), Wolfgang Pempe (DFN), Jussi Piitulainen (University of Helsinki), Aarno Sandvik (University of Helsinki), Brook Schofield (Terena), Sami Silen (CSC), Nadia Sluer (Terena), Ville Tenhunen (University of Helsinki)

This section shortly presents the Federations, Home Organisations and Service Providers in the pilot.

Identity federations

- **DFN-AAI**, the German identity federation, is operated by DFN, the national research and education network (NREN) of Germany. Technically, DFN-AAI is a full-mesh federation whose Members are mostly running the Shibboleth Identity and Service Provider software
- **Haka federation**, the Finnish identity federation, is operated by CSC – It center of Science who runs also Funet, the Finnish NREN. Like DFN-AAI, Haka is a full mesh federation which relies mostly on the Shibboleth software. Haka federation's policy already incorporates a division of responsibility on data protection which is to some extend similar to the Code of Conduct
- **SWAMID federation**, the Swedish identity federation is operated by SUNET, the Swedish NREN. Like DFN-AAI and Haka, SWAMID is a full-mesh federation that makes mostly use of the Shibboleth software
- **SurfFederatie**, the Dutch identity federation, is operated by Surfnet, the Dutch NREN. Unlike the others, Surffederatie has a hub-and-spoke architecture, where Surfnet operates a central Identity Provider on behalf of the Home Organisations. The hub is an integrated combination of custom PHP-based software and PingFederate product.

Service Providers

The Service Providers in the pilot represent mostly the CLARIN community, who has reported problems on retrieving necessary attributes from the Identity Providers in the current identity federations. CLARIN has found that there are language researchers in 176 Home Organisations in Europe, and contacting them all individually to convince them to release attributes does not scale for their purposes. CLARIN had strong incentive to drive a scalable approach to attribute release.

Following Service Provided took part to the pilot and were committed to the Code of Conduct:

Service	Service owner	Federation	SAML SP Software	Required attributes
LAT (Language Archive Technology)	CSC – IT Center for Science	Haka	Shibboleth	ePPN
CLARIN services	Institute for the German language (IDS)	DFN-AAI	Shibboleth	ePPN
IDS repository	Institute for the German language (IDS)	DFN-AAI	Shibboleth	ePPN
Clarin Catalog	MPI in Nijmegen	DFN-AAI	Shibboleth	ePPN, ePTID, mail
MPI Second SP	MPI in Nijmegen	DFN-AAI	Shibboleth	ePPN, cn
Weblicht	Tübingen university	DFN-AAI	Shibboleth	ePTID, ePPN
Funet Filesender	CSC – IT Center for Science	Haka	SimpleSAMLphp	ePPN, cn, displayName, mail

Home Organisations and Identity Providers

Home Organisation	Identity Provider software	Federation	Test report
DFN	Shibboleth	DFN-AAI	Media:Dfn_test_report.pdf
Uppsala University	Shibboleth	SWAMID	
CSC – IT Center for Science	Shibboleth	Haka	Media:Csc_test_report.pdf

Non-technical findings

Service Providers' willingness to commit to the Code of Conduct

In general, the Service Providers in the pilot felt that the requirements the Code of Conduct imposed on them were modest. The Service Providers found that the clauses to which they committed were something that they were already bound to do by the Data protection directive and its national implementations. The Service Providers were comfortable to commit to the Code of Conduct.

In the first draft, the Service Providers were expected to sign the Code of conduct document digitally or with ink. This led to significant administrational burdens and delays in some Service Providers because the document had to go upwards in the organization to the level with people eligible to sign contractual documents. Based on the feedback, the requirement to sign the Code of Conduct was relaxed.

Privacy Policy document

There were some findings in the Code of Conduct text that led to clarification to the Code of Conduct text and supporting documentation, including

- Service Providers were not familiar with the concept of the Privacy Policy and how to develop one for the service. Attention was paid to providing a Privacy Policy template to the Service Providers
- Service Providers had not realized that they have a legal obligation to erase the end users' personal data record record not only if the user him /herself asks for it, but also if the user doesn't show up again for a defined period of time. What the time is depends on the service, but the Service Provider cannot keep the user records forever. More attention was paid to this when developing the Privacy Policy template
- In some cases the list of attributes a Service Provider requests according to its Privacy Policy conflicts with the list of attributes requested according to the SAML 2.0 metadata's RequestedAttributes elements. There needs to be a process in place to check that they are aligned.

Necessary attributes

The Data protection Code of Conduct assumes that Service Providers "only process Attributes of the End User that are necessary for enabling access to the service". Further guidelines are provided in What attributes are relevant for a Service Provider. Interpreting this causes practical problems.

Identifiers. In general, bilateral identifiers (such as SAML 2.0 persistent identifier) are preferred because they preserve user privacy better than shared identifiers (such as eduPersonPrincipalName, ePPN). However, practice has shown that Identity Providers have better support for ePPN than SAML2 persistent identifier. As an outcome, many Service Providers decided to request ePPN because, for them, more important than privacy is that the users can actually access the service. This dilemma remains as long as Identity Providers don't generally support SAML2 persistent identifier.

Common attributes. The question when common attributes like e-mail address and displayName are necessary depends on the service. It is proposed as a rule of thumb that they are necessary, if a trustworthy value for them is needed. In general, the value is trustworthy if it is retrieved from the Identity Provider and not typed in by the user him/herself.

Funet Filesender, the Finnish installation of the popular filesender service, decided that both user's e-mail address and common name are necessary and need to be retrieved from the home organisation, because if the user has an opportunity to type them in by him/herself, Filesender enables the end user to spoof the sender of the file.

In a "Call for action on federated identity" CLARIN-D and DARIAH-DE identified a common set of 6 attributes which are required to enable Web-SSO-based collaboration within both research infrastructures.

Other SAML 2.0 metadata elements

The service providers also introduced names (e.g. “Lux17 Service Provider”) and descriptions (e.g. "Max Planck Institue for Psycholinguistics Lux17 Service Provider") which were not very understandable and useful for common end users.

Home federation's role

Based on the sections above, it appears that there is a need that some external body makes a light sanity check the the service provider's Privacy Policy document, list of requested attributes and other SAML 2.0 metadata elements. In an identity federation, the natural party is the federation operator.

However, a design goal of the Code of Conduct has been to avoid the federation operator becoming liable for the service provider's omissions. The home federation operator cannot perform checks that may expose it to the liability.

Home Organisations’ willingness to release attributes to Service Providers committed to the Code of Conduct

In general, the Home Organisations in the pilot felt comfortable to release attributes to Service Providers who have committed to the Code of Conduct. In some cases this required explaining to the CIO or information security manager that

- the Code of Conduct balances the risks of attribute release leading to legal troubles with the risk of the employees of the organization not being able to do their job effectively, and
- the alternative would require bilateral negotiations with all Service Providers which would mean a lot of extra work for the Identity Provider administration

To further convince the Home Organisation, it was emphasised that

- the Home Organisation can reduce its risks by limiting the maximum set of attributes released to the Service Provider committed to the Code of Conduct. In particular, there is no need to release sensitive personal data
- the Service Providers reside in EU/EEA or similar where the local laws already provide protection similar to the Code of Conduct
- there is not batch release of personal data; instead, in SAML WebSSO the personal data is released only when the user needs to access the service, and the Identity Provider server can inform the user on the release
- the alternative for the Code of Conduct is that every employee of the organization uses his/her Google account or similar to access the Service Provider, and those accounts are outside the employer’s control and will not be closed when the user departs

It is necessary to develop dissemination material on the Code of Conduct and share it among the identity federations.

Technical findings

Deploying a Discovery Service

Although not related directly to the Code of Conduct, the Service Providers were not familiar with how to deploy a discovery service that shows also the Identity Providers in a foreign federation. Training material on UI design (such as REFEDS Discovery guide) and related software modules (such as Shibboleth Embedded Discovery Service or DiscoJuice) is needed.

Approaches to manage attribute release from IdPs

Among the four federations in the pilot, two alternatives were adopted for managing the attribute release to the Service Providers committed to the Code of Conduct.

1. **IdP administrator deploys metadata filtering/scripting**.

In this alternative, the federation operator simply provides the full (eduGAIN interfederation) SAML2 metadata feed to the administrator of the Identity Provider server. It is up to the administrator to deploy local configurations to filter out those Service Providers to which they do not want to release attributes (in this case, those Service Providers that are not committed to the Code of Conduct). This can be done, for instance, by

- a script that preprocesses the incoming metadata stream and produces an output stream that contains only SPs that are committed to the Code of Conduct (this approach is product-independent)
- a filter configuration, that relies on the Identity Provider product's feature to pick up entities based on the Entity Category elements in place in the Service Provider's metadata (this approach depends on the Identity Provider software; see the next section below on Shibboleth)

This approach was deployed by DFN-AAI and SWAMID federations.

2. **Federation operator provides a centralized management UI.**

In this alternative, the federation operator provides a tool for the Home Organisation representatives for picking up those Service Providers to which they want to release attributes (in this case, the Service Providers that are committed to the Code of Condut). In practice, the Home Organisation representative logs in to a management web UI where s/he checks an "I want to release attributes to CoC-enabled Service Providers" checkbox, and the metadata management system

- in a full-mesh-federation, generates to the Identity Provider a tailored SAML2 metadata feed, containing only those Service Providers that have committed to the Code of Conduct, or
- in a hub-and-spoke federation, configures the central Identity Provider to automatically release attributes to the Service Providers that have committed to the Code of Conduct

This approach was deployed by Haka federation (a full mesh) and SurfFederatie (Hub and Spoke).

Alternatives to configure attribute release for a Shibboleth Identity Provider

Three alternatives were found to configure a Shibboleth Identity Provider (2.3.4 or later) to release attributes to a Service Provider that has committed to the Code of Conduct.

1. Release a fixed CoC attribute bundle to CoC SPs

Shibboleth Identity Provider can not handle dynamic attribute release based on requested attributes in SAML2 metadata out of the box. A simple solution is to release a fixed attribute bundle with mostly innocuous attributes to all CoC compliant Service Providers. However, a Home Organisation adopting this approach assumes data protection risks because this solution doesn't meet the data minimalisation principle all the way.

SWAMID has published guidance at http://wiki.swamid.se/display/SWAMID/GEANT+Data+Protection+Code+of+Conduct.

In the pilot, this approach was adopted by Uppsala university.

2. Release a dynamic subset of the CoC attribute bundle to CoC SPs

With the Shibboleth Identity Provider add-on uApprove from SWITCH it's possible to cover the extra mile to fully meet the data minimalisation criteria. You don't have to activate the uApprove consent module in the Identity Provider, just install it. After the installation you can build filters based on requested and required attributes.

SWAMID has published guidance at http://wiki.swamid.se/display/SWAMID/GEANT+Data+Protection+Code+of+Conduct.

In the pilot, this approach was adopted by CSC - It Center for Science and DFN.

3. Generate a separate attribute filter policy for each CoC SPs

For federations that supply a centralised tool to build and distribute attribute filters, it is possible to automatically create and distribute attribute filters for participanting Identity Providers based on CoC entity category and required attributes. For instance, the Identity Provider administrator can log in to the federation's metadata management tool and check a checkbox "Generate me a SAML2 metadata feed and attribute filter policy that contains all SPs committed to the CoC".

This can be also done locally in an Identity Provider with for example a XSLT script.

Service Provider proxies

A Service Provider proxy is a SAML Service Provider that sits between the Identity and Service Provider in the SAML flow and hides several "eventual" Service Provider behind itself. Although it suffers from the well-known problems of any proxy/gateway solutions (e.g. is a potential privacy, security and capacity bottleneck), it is used in some deployments for operational or security reasons.

Usually, the "eventual" Service Providers are managed by the same organisation that manages the proxy, in which case the proxy does not release personal data to 3rd parties and the related section of the Code of Conduct is not applied.

Service Provider proxies can be further split into two categories:

1. Transparent proxies

A Transparent proxy hides several Service Providers with different EntityIDs behind itself and each "eventual" Service Provider is visible to the Identity Provider as a separate SAML2 metadata Entity. In this case, each "eventual" Service Provider has its own EntityID and list of necessary attributes but the SAML endpoints refer to those of the proxy. If only some of the "eventual" Service Providers commit to the Code of Conduct, they can have the appropriate Code of Conduct elements in their SAML2 metadata entity but there can be also Service Providers who don't commit to the Code of Conduct. Obviously the Code of Conduct needs to be respected in the operations of the proxy itself.

2. Non-transparent proxies

A Non-transparent proxy hides several "eventual" Service Providers behind itself and those Service Providers are not visible to the Identity Provider. A single SAML2 metadata entity represents all "eventual" Service Providers towards the Identity Provider. As a consequence, the proxy's list of requested attributes is a union of all the attributes the "eventual" Service Providers require, and the proxy also needs to have the other SAML2 metadata elements the Code of Conduct requires. All the "eventual" Service Providers need to commit to the Code of Conduct together and at the same time with the proxy.

Future work

Privacy Policy webcrawler

The Code of Conduct introduces Federation operator's guidelines suggesting that the home federation operator stores the Privacy Policy document of the Service Provider for future evidence. While this can be done manually, it could also be replaced by a fully automated webcrawler that browses the SAML2 metadata file and stores the Privacy Policy documents automatically. The crawler could be provided and operated e.g. by the eduGAIN interfederation service.

Attribute release GUI implementations

The Code of Conduct covers Notes on Implementation of INFORM/CONSENT GUI Interfaces which are not currently supported by common Identity Provider products. Separate projects are needed to implement and release the recommendations in Identity Provider products/modules such as

- uApprove of Shibboleth
    - for instance, uApprove.jp
- the consent module of SimpleSAMLphp