20 January 2016

GN4-1 White Paper: Issues and Solutions for SAML Identity Federation Statistics

Supporting Documentation

Work Package/Activity:	09/SA5
Task Item:	Task 1
Dissemination Level:	PU (Public)
Lead Partner:	GÉANT Association
Document Code:	GN4-1-16-31fb71
Author:	Nicole Harris, GÉANT Association

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 691567 (GN4-1).

Contents

1 The Problem Statement

SAML-based identity federations have been well established in the research and education sector for a number of years and the question of how many authentications are carried out via federations is often raised for a variety of reasons, but typically to show the value of federations via a simple metric. The Research and Education FEDerations (REFEDS) group's REFED MET tool currently tracks 39 different federations, and shows that there are currently 10 610 unique entities registered within these federations, representing 3335 Identity Providers (IdPs) and 7275 Service Providers (SPs). While this shows a clear incentive and desire to register services with federations, it cannot demonstrate actual use based on number of authentications. To date, there is no service that can provide this information and no single way of effectively attempting to provide a figure on the number of authentications made via identity federations.

The reason for this is that most SAML-based identity federations are designed around a "mesh" model, where the federation acts purely as a registrar and publisher of metadata and plays no role in the transaction process between IdP and SP. The federation does not, at any point, see information about the log-in process, so is not able to count the number of authentications. The only place that these logs exist is at the IdP and SP.

Some federations have elected to play a more central role within the workflow for federated access, and have formed as either "hub and spoke" or "centralised" models. These federations are able to collate information about number of authentications and have typically published these either internally within a closed community, via a federation customer portal or publically. The raw data from these federations is typically held in a reusable and fairly simply format, such as MySQL or JSON, but no attempts have been made to collate and share this information to date.

More information about federation types is available on the eduGAIN wiki [eduGAIN]:

2 The Value Proposition

Statistical information has clear management use-cases, particularly when trying to show value for services and funding streams as it proves usage and is easily comparable against other offerings. IdPs and SPs will see benefits of gathering statistical information locally to prove that services are being used and to decide when services are no longer useful. One of the problems faced by federations is convincing providers to aggregate this information upwards for benefits to be shown elsewhere. IdPs, in particular, are concerned about sharing information that could be considered Personally Identifiable Information (PII), and as such, be regulated by EU data protection requirements. Despite assurances of anonymity in recent pilots, it is still difficult to get beyond this convenient argument for maintaining the status quo.

Better arguments and incentives are required to enable federations to move forward in this space. This should include:

- Highlighting the importance of sharing statistical information to ensure that federations and interfederations can continue to receive adequate funding at both local and EC levels.
- Easier for federations to demonstrate and explain the usefulness of a federated approach based on statistical evidence and for audiences without a technical background to understand the scale of usage.
- Highlight the benefits of being able to compare local and aggregated information, including opportunities for "gamification" of the data.

3 What information do we currently have?

The GN4-1 project's SA5 Trust and Identity Harmonisation task (Task 1) carried out a brief survey of federations and received feedback from 15 federations. The results of this survey can be found on the GEANT wiki [GÉANT]. The survey demonstrated that there is a wealth of information currently available from hub and spoke and centralised federations and that several federations with a mesh model of operation have taken steps to gather this information in various ways (see results for Edugate, SWAMID, AAF and Haka).

Looking at publicly available information, some trends and a sense of the traffic within federations can be seen. The table below shows some figures based on user login peaks at the beginning of the academic year, as students return

Federation	Approximate Logins
SURFConext	1.5 million logins per week
WAYF	400,000 logins per week
AAI@EduHr	142,445 logins per week

Table 3.1: Portrait <Style: Caption (Copy/paste label and number)>

4 Other Approaches and Information

4.1 Federation Entity Information

At a basic level, it is possible to track the interest in identity federations via the number of entities registered with federations for mesh federations and scope elements for hub and spoke federations. This may not fully demonstrate usage of federations, but the effort involved in installing and implementing an IdP and SP within this environment means that it is very unlikely that effort would be put in to creating and registering entities without them being used in production.

There are two core tools for monitoring entities within federations:

- Metadata Explorer Tool [MET] developed by REFEDS. This is a simple management interface to metadata published by federations as their "main feed" and includes all federations that have requested to be involved. The tool can be managed by the REFEDS coordinators or by each national federation. MET does not currently show changes over time, only the current snapshot of entities.
- eduGAIN Technical Site [eduGAIN-TS] developed by the eduGAIN OT. This shows detailed information of eduGAIN entities.

These tools provide a number of benefits in terms of providing a user-friendly interface to federation metadata as well as a basic set of information that can be used to promote the usage of federations. However, this information is unlikely to prove a sufficient amount of data to satisfy all parties.

4.2 Aggregating Data from Hub and Spoke / Centralised Federations

As previously mentioned, most hub and spoke / centralised federations are currently able to share basic statistical information. The recent survey carried out by SA5 T1 asked federations if they would be willing and able to share this information centrally to help support the aggregation of statistics across federations. Most federations showed a willingness to share information, if a common data format could be defined. Current statistics are managed in the following formats: MySQL, JSON, F-TICKS or RAPTOR.

4.3 Central Discovery Statistics

Some federations offer a central discovery service to their communities that can provide insight into the number of users actively using a selection of services, however, this data is not considered to be useful or reliable as a source for statistics. As detailed in the REFEDS discovery guide, discovery.refeds.org, central discovery is not a recommended practice. Where central services are implemented, federations often have "fallback" discovery services and do not collate information across these services. In addition, not all discovery services support statistical gathering. Such data could be used to monitor general trends in the usage of central discovery services, which in itself provides a useful insight, but should not be seen as a reliable indicator of resource access statistics.

5 Potential Solutions

5.1 RAPTOR / AMAAIS

RAPTOR (a tool designed as part of a Jisc project) and AMAAIS (a SWITCH development project) are designed to provide an easy-to-view, management-style interface to federation statistics [RAPTOR] [AMAAIS]. They require software to be deployed locally by an IdP.

The advantage of both of these projects are that they are broadly technology independent, and allow for multiple types of software to be monitored. The tools also allow for information to be aggregated at a central point. However, both require effort on behalf of the IdP in terms of installation and integration with existing Identity Providers, and take-up has been slow, as it has been difficult to persuade IdPs of the benefits of engaging in this work. This is part of a growing trend where IdP organisations find it difficult to contribute locally to development work, as implementing federated access is seen as a "completed project" by senior management and is not something that requires ongoing investment in terms of time or money.

Future development could focus on providing a greater part of the RAPTOR service at the federation level to reduce the burden. This model has been successfully deployed by the Edugate federation in Ireland, and is proving successful.

5.2 F-TICKS for SAML

The Federated Ticker System (F-TICKS) specification was designed to support the need for metering and monitoring within eduroam in order to provide centralised statistics across eduroam services, which faced the same problems as shown in for SAML-based identity federations. F-TICKS is a text log format which can be used to communicate the essential aspects of authentication events to log reception software[1]. F-TICKS is a useful approach to defining the standardised messaging format for the logs, but work is still needed within individual services to ensure that implementations are optimised to pass this information. There were challenges achieving this for eduroam, which arguably has an easier infrastructure to apply this concept to (e.g. more convergence on technology approaches with RADIUS and more centralised design).

Two federations have explored an F-TICKS for SAML approach to date: SWAMID and AAI@EduHr. Work is underway to develop this as a standard under the RFC banner, based on work previously used for eduroam.

6 Development Proposals and Solutions

Based on evidence gathered for the federation survey and background research for this paper, the following recommendations are made:

1. Work needs to be done to complete the SAML F-TICKS specification to make it useful.

[1] https://tools.ietf.org/html/draft-johansson-fticks-00.

2. A central service to aggregate statistics from federations in the same model as monitor.eduroam.org should be developed, starting with hub and spoke / centralised federations.

3. Further work should be done to develop RAPTOR based on the Edugate approach (parked NIF from GN4-1). This depends heavily on increasing uptake of the tool by organisations.

4. More should be done to enable IdPs to commit time to developments such as support for aggregated statistics. GÉANT should look at producing a set of recommendations for IdPs' "local development plans" that shows areas where IdPs should be investing time and effort in over the next 1 to 2 years in a visual way, useful to senior management.

5. REFEDS / eduGAIN should look at tracking changes in entity information over time in MET and edugain tools.

References

[AAF]	http://aaf.edu.au/
[AAI@EduHr]	http://www.aaiedu.hr/en
[AMAAIS]	http://www.csg.uzh.ch/research/previous-projects/amaais.html
[eduGAIN]	https://wiki.edugain.org/Federation_Architecture
[eduGAIN-TS]	https://technical.edugain.org/entities.php
[Edugate]	http://www.heanet.ie/services/identity-access/edugate
[F-TICKS]	https://monitor.eduroam.org/f-ticks/
[GÉANT]	https://wiki.geant.org/display/gn41sa5/Current+Federation+Statistics
[Haka]	https://confluence.csc.fi/display/HAKA/In+English
[MET]	https://met.refeds.org
[RAPTOR]	https://iam.cf.ac.uk/trac/RAPTOR
[REFEDS]	https://refeds.org/
[SWAMID]	http://www.swamid.se/

Glossary

F-TICKS	Federated Ticker System
IdP	Identity Provider
MET	Metadata Explorer Tool
PII	Personally Identifiable Information
REFEDS	Research and Education FEDerations group
OT	Operations Team
SP	Service Provider