Infrastructure for Collaboration and Identifier Mapping

The SNIC perspective

This white paper describes proposed infrastructure for associating identities used in GRIDs, Identify federations and various enterprise authentication mechanisms such as SSH keys or One Time Password (OTP) tokens.

This document is the result of a discussion held at the SNIC (http://www.snic.se) offices 21/1 2010 between Hans Wallberg, Mikael Grønager, Sverker Holmgren and Leif Johansson. The discussion focused on possible solutions to the following challenges facing SNIC:

1. The lack of unified identity management for the HPC centers in Sweden
2. The high cost of computer-security incidents involving lost credentials
3. The need to make SNIC resources available in a user-friendly way
4. The need to support collaboration spanning traditional organizations
5. The desire to quickly establish long-term solutions to these problems

Not all SNIC researchers use the GRID security model to access resources. In fact the typical SNIC researcher often uses SSH to connect to HPC resources. Currently provisioning and de-provisioning identities is a more or less manual process entirely handled by the HPC center.

A researcher that uses GRID has a set of tools available for managing teams of researchers (eg VOMS). The corresponding resources are not available to researchers who do not use the GRID infrastructure.

This document proposes infrastructure which can be used to solve several of these problems. The proposed infrastructure is aligned with current best practice in national research infrastructure and builds upon experience from related projects carried out by (among others) SWITCH, Internet2 and UNINET.

Conventions

This document is intended to be read by anyone in the SNIC community with an interest in unified identity and access control including (but not limited to) users, system administrators and SND directors. The majority of the text is geared towards advanced users and decision makers.

> Here and there there are notes written more for the technically interested readers. Such notes are presented like this and can be safely ignored by those not interested in the gory details.

Collaborative Organization Infrastructure

A Collaborative Organization (CO) is a generalization of the VO concept from the GRID community which can be used to represent – in its most basic form - a group of identities in an identity federation. With most VO software (eg VOMS) the group concept is tied to specific technology for authentication and authorization. A COIP tries to remove as many such dependencies as possible by acting as a gateway between the various naming standards involved (cf below).

Historically Collaboration Infrastructure has been tied to specific use cases (eg a mailing list or LMS service or VOMS instance) or technologies. This has resulted in a proliferation of semantically equivalent but duplicate group definitions.

Users are often confused and irritated by the fact that changing membership in one context doesn't affect other contexts.

By stripping away dependencies on specific use-cases or technologies a CO can be used to drive a much wider range of applications. This is the motivation for introducing the notion of a Collaboration Infrastructure Platform (COIP).

The COIP stores mapping between users and groups without reference to any specific applicability of the groups. In fact group can be reduced to a symbolic *tag* applied to each member of the group. The idea of representing group membership as such a tag (often called an *entitlement* on the user) has wide support in the identity federation community.

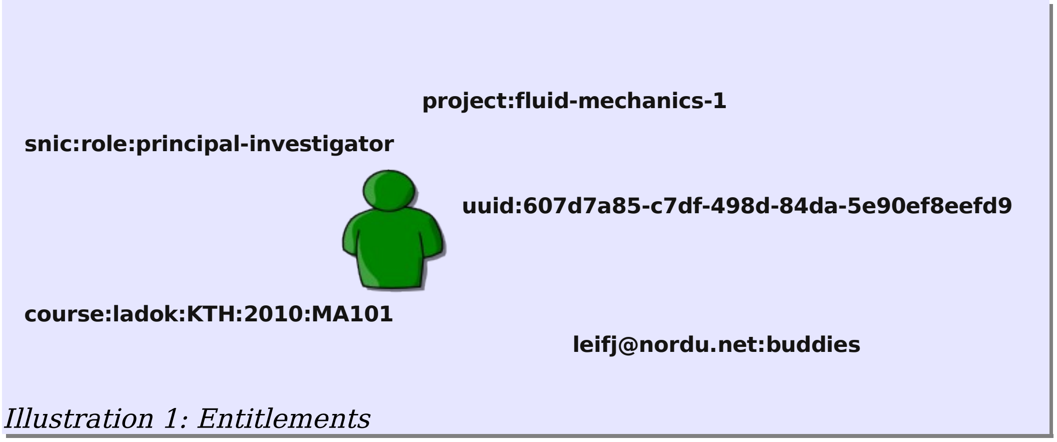
project:fluid-mechanics-1
snic:role:principal-investigator
uuid:607d7a85-c7df-498d-84da-5e90ef8eefd9
course:ladok:KTH:2010:MA101
leifj@nordu.net:buddies

Illustration 1: Entitlements

Data from the COIP (i.e groups in the form of entitlements) can be used to provision groups in a wide range of software including VOMS, mailing list managers and LMS (Learning Management Systems).

Systems that use group definitions from the COIP will often have different ways of representing user identifiers depending on the authentication technologies involved. For instance a VOMS instance will typically represent membership in a VO by listing the GRID Certificate subject name of member users while a mailing list manager would need a list of email addresses.

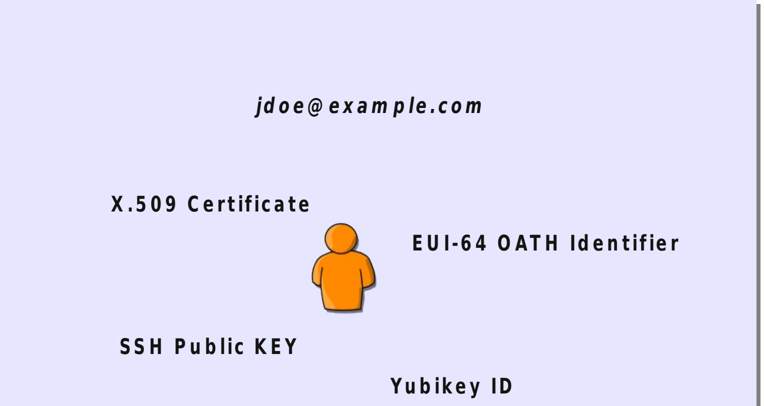

Illustration 2: Multiple Identifiers

Hence in order for (say) a VOMS instance to be able to make use of the group definitions in the COIP a mapping needs to exist between the federated user ID and the GRID Certificate DN.

> Note that this mapping does not need to be sourced in one system: one attribute source may know about certificates and another may know about Kerberos principals and a third about OTP Tokens. It may even be different sources for different users.

Thus in addition to the COIP there is a need for a way to securely map between multiple identifiers associated with an identity. This mapping can be realized at a federation-wide service or at each identity provider through the normal process of attribute release or it could be integrated with the COIP for a simplified user experience.

User workflow

Users who want/need to use services that depend on the COIP must first establish their federated identity at the COIP. This is easily done by a normal federation login process.

The typical way to initiate this process is for an owner of a group to *invite* the user into the group (using an administrative interface of the COIP).

> The invitation will typically take the form of an email containing a so called 'nonce' value. Such email messages are commonly used by mailing list managers and users are quite often comfortable with this type of interaction. The invitation email contains a link with an invitation token. By clicking on that link the users is redirected to the COIP where he/she is prompted to login using her federated identity. The invitation token serves to bind the federated identity with the invitation to the group.

When the user authenticates to the COIP the federation technology associates a set of attributes with the session. These attributes are provided by the federation identity provider (eg a campus identity) and often include basic personal information:

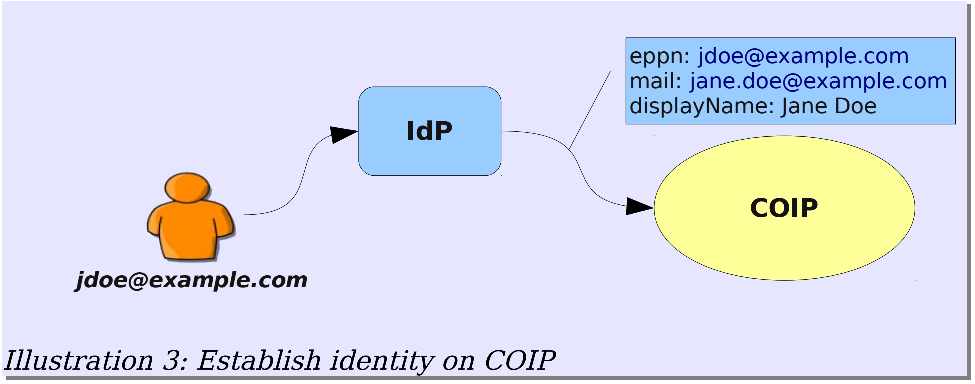

Illustration 3: Establish identity on COIP

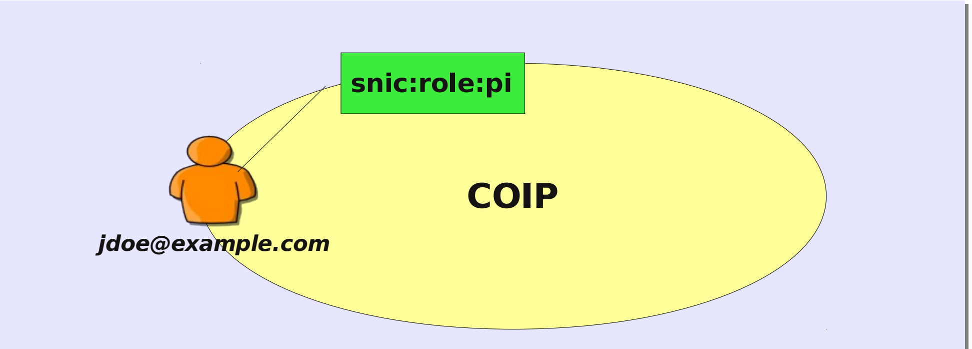

Illustration 4: Entitlement after invitation to COIP

A concrete example could be that of a SNIC manager inviting a PI. By accepting the invitation and logging into the COIP the PI is associated with the **`snic:role:pi`** entitlement in the COIP. The structure of the entitlements doesn't matter for this discussion but they are important in order to maintain separation between multiple uses of the COIP.

Of course jdoe@example.com may already be established on the COIP and could have other entitlements already. For instance Jane Doe may have been registered on courses, the corresponding entitlements of which being imported from LADOK on a regular basis.

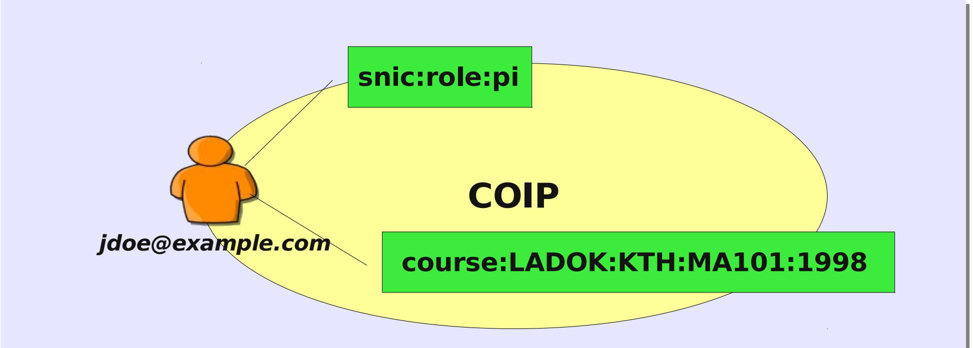

Illustration 5: Entitlement after invitation to COIP

Entitlements are sometimes PII (personal identifiable information) and almost always sensitive in some way. Therefore the COIP will have fine-grained controls for the use of entitlements.

Now jdoe@example.com (as a SNIC PI) may to create a group for her team of scientists. She creates a group in COIP and invites her team members into the group the same way she was invited into the PI role by the SNIC manager. Each team member accepts the invitation thereby establishing a set of identifiers as team members.

> The name of the group(entitlement) isn't arbitrarily chosen. By structuring the namespace of the COIP the semantics of various types of entitlement becomes clear and they can be processed by simple text-based tools such as regular expressions.
>
> Furthermore the difference between a group and a role is only in the eye of the beholder with this model. One way of looking at the difference between a role and a group is that a role is a group associated with a context (often an organizational scope). By managing the namespace the COIP can be used to model roles – for instance by assigning scope to the names. One way of doing this is to use the GMAI model which generalizes the name-structure used for the examples in this text: http://www.swami.se/pub/jsp/polopoly.jsp?d=3619&a=11367. The GMAI model has been adopted by VHS as the basis for authorization in the new “NyA institutionswebb”.

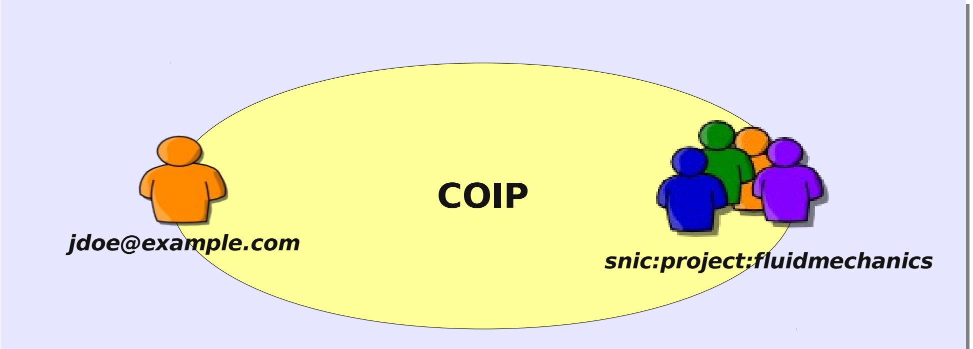

Illustration 6: Jane Doe and her team of scientists

The team is going to be working with both GRID-based batch tools and a visualization server which requires SSH access. Each team member therefore needs to associate at least one SSH key and at least one GRID certificate with his/her identity.

> It might be tempting to allow a user simply to upload X.509 certificates or SSH keys to a web interface but that would not result in a secure binding between the federated identity and the SSH key (or X509 certificate). In order to achieve security we require the user to perform a *secondary* authentication which proves possession of the credential corresponding to the identity

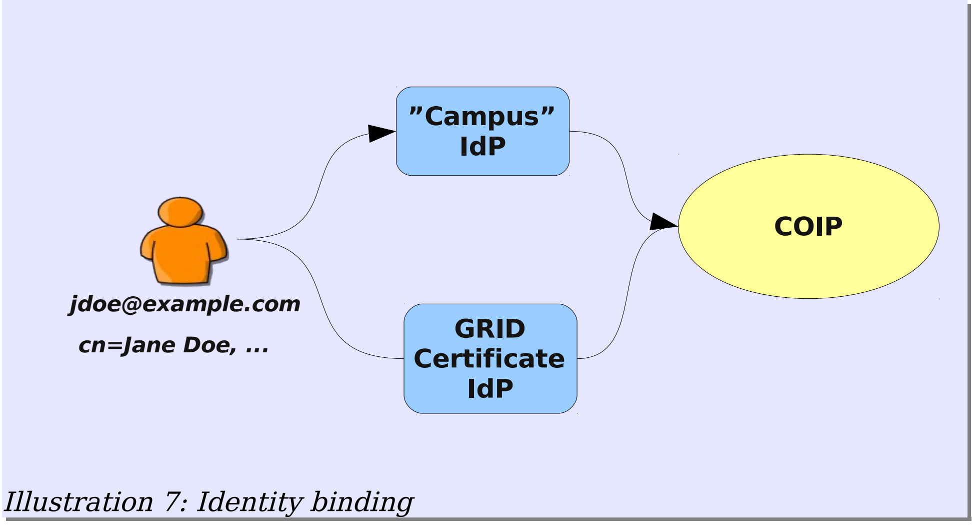

Illustration 7: Identity binding

Jane Doe has both a federated identity (normally the primary identity) issued to her by her employer. In addition Jane has a GRID certificate which is verifiable through the TAG PMA trust anchor list. Jane authenticates to the COIP with her GRID certificate. This gives the COIP a secure binding between Janes federated identity (jdoe@example.com) and her GRID certificate subject name.

> Packaging this verification and that of other authentication mechanisms (for instance an OTP-token) as SAML IdPs gives the COIP a single interface to most secondary authentications.

Certain HPC centers could for instance require the use of other technologies (Kerberos or OTP tokens) and the same pattern could be used to bind the corresponding identifiers into the set of authenticated identifiers for a user.

> There are certain identifiers such as for instance cellular telephone numbers used for SMS-based OTP-mechanisms which need to be validated by external parties but most of the interesting cases can be viewed as just another SAML IdP.

This secure binding of identities is often visible in social applications (facebook, twitter, etc) that allow users to change, or add additional, alias email addresses to their identity. Simply allowing a user to assert an email address would be a security problem and it is common for the user to have to prove possession of the email address using a reply-mail to the list.

When Jane Doe has successfully authenticated herself using her GRID certificate the COIP has a secure association between her campus identity and her GRID identity.

In the future when external systems obtain group membership information involving Jane Doe she can be represented by the identifier best suited to the technologies involved.

> The SSH key case is similar although arguably simpler. In this case Jane uploads an SSH key and is asked to validate the key by using it together with a nonce value (provided by the COIP when she uploads the SSH key) to authenticate to a special SSH server which first requires SSH key authentication and then additionally requires her to provide the correct nonce value. This creates a binding between her SSH key and her campus identity.

At this point the research-team is ready to go to work. Since the team has been allocated SNIC resources the corresponding entitlement (**`snic:project:fluidmechanics`**) is added to a list of groups with authorization to the SNIC resources.

The next section discusses the various integration mechanisms which could be used to implement this authorization. For the remainder of this example we will describe one possible approach – it is not the only way to use COIP groups – chosen for its simplicity.

The resources which Jane Doe and her team is going to use are located at two different HPC centers. Each center has a self-service sign-up web application which allows users of SNIC resources to register for access at the particular center.

Lets assume further that one of the centers uses SSH to provide access to resources. The other uses GRID tools but also Kerberos with an additional requirement of use of an OTP device for increased security.

> OTP devices are identified by EUI-64 identifiers. These are non-personal 64bit numbers uniquely identifying an OTP device. The COIP is used to bind this identifier to the user who is in possession of the device.

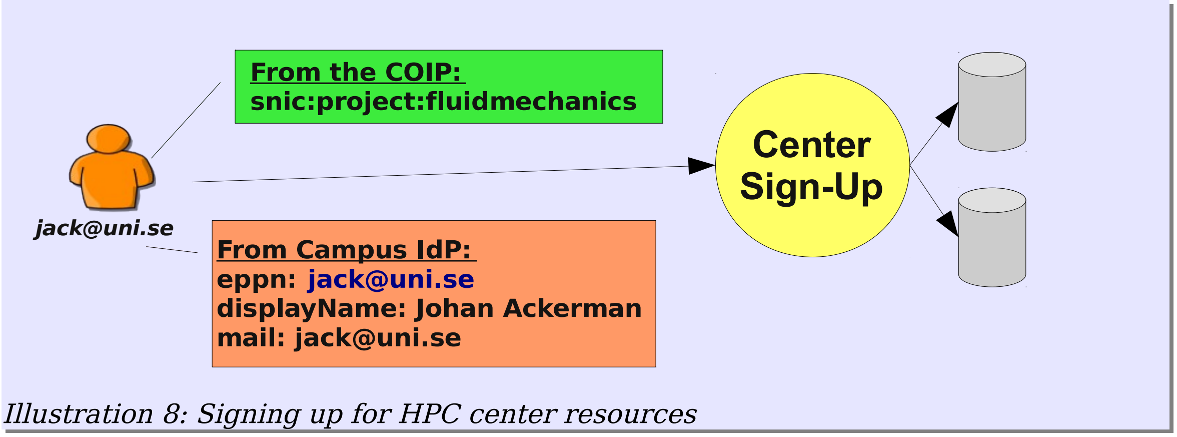

Illustration 8: Signing up for HPC center resources

Each of the team members visit the sign-up application for each of the centers. The page redirects the user to a federation login page. After successfully authenticating with the chosen federated identity the sign-up application obtains basic personal information about the user together with – from the COIP - a set of additional identifiers (GRID certificate names, SSH keys and OTP device identifiers) and a set of entitlements. The sign-up application sees the fluidmechanics project entitlement which is one of the authorized projects for the resources at the HPC center.

The sign-up application uses the information provided to provision the user in the local identity management of the HPC center. This process is entirely dependent on the technologies used by the HPC center and doesn't concern us here.

This may seem like a cumbersome process – each user has to visit each center and click through a registration process. However note that most users have to do this anyway with the difference that the process isn't unified or even (in most cases) automated.

There are several benefits of this simple approach: it is simple, doesn't impose large costs on the centers and scales very well.

> De-provisioning of users needs additional work but can to a large extent be automated. The GN3-JRA3-T2 task of the GEANT3 project is working on simple solutions. By periodically inviting users to re-authorizing their local registration the consequences of a loss of credentials can be limited.
>
> Certain technologies such as Kerberos have built-in federation capabilities (cross realm trust) which could be used instead of, or in conjunction with a manual sign-up application. Such possibilities should be explored.

When the sign-up process is completed the user has access to the resources of the HPC center like a normal user. Note that the sign-up could be managed without the manual intervention of anyone from the HPC center.

> Most parts of the sign-up application will be independent of the local setup at the HPC center and can be produced and maintained as a common software package by SNIC. Integration with the local Identity Management (IdM) can be done with very simple interfaces – eg shell scripts for adding, removing and disabling users.

Application Integration

Collaborative applications such as mailing list managers, VOMS, LMS-systems etc all need access to group and membership information from the COIP. There are three strategies to accomplish this integration: out-of-band (OOB), attribute-oriented or front-channel session-oriented.

In the OOB model the application either pulls data from the COIP or obtains a stream of updates to the information as it changes on the COIP as the result of user activity. This creates a tight coupling between the application and the COIP which might be needed for certain applications. This type of integration however can become expensive to setup and maintain.

One example of this approach is the CoManage system developed by Internet2 and Stanford where an LDAP server is used to represent group membership.

Applications (eg a mailing list manager) is simply configured to use groups from in the LDAP server. While this may seem like a simple approach experience shows that interoperability between LDAP information models are not always good enough.

> Applications often advertise LDAP support but in practice there are many differences in what this actually means. In many cases LDAP support means the ability to authenticate using LDAP which is not what we want in this case.

In most cases it is preferable with a more loose integration between the COIP and applications. Arguably the easiest way is to enable applications for federation login and allow access to data in the COIP is to view the entitlements and identity mappings as normal attributes.

This approach has been show to be easy to deploy using major SAMLv2 implementation in the JRA3-T2 activity of the GN3 project. It has the added benefit of being agnostic wrt the source of the attribute: the campus IdP could *also* be the source of an identifier attribute (SSH keys for instance). This allows participating organizations to differ in how much support they want to provide for their users.

In the case of SSH key distribution the application (eg the SSH key distribution mechanism of an HPC cluster) could either pull keys from the COIP – for instance all keys associated with an entitlement - or could simply put up a web application where the user (by signing in with her federation identity) would transport her SSH keys as normal attributes. This model has the benefit of being “on demand” in the sense that only active users keys need to be handled.

The third model (front-channel session-oriented) uses AJAX technology to pull in attributes from the COIP using the session as a replacement for the user identifier. This approach has some benefits but requires more work in the application.

> This model has been explored in GN3-JRA3-T2 and involves clever uses of OAuth (http://oauth.net) and has an important benefit: it doesn't require that the application has access to the federation identifier (eppn) of the user but can work with so called pseudonomous identifiers. This is important if privacy of researchers is crucial.

Recommendations

- While it is possible for the identity mapping application and the COIP to be implemented separately it is probably easier and cheaper to implement them together in one service.
- SUNET should be given the responsibility for developing the software and service and should coordinate this with both the identity federation community in Sweden and the Geant project aswell.
- SUNET should coordinate with the following communities:
    - SWAMI
    - SWAMID Operations
    - SNIC National HPC centers
    - SNIC
- SUNET should seek involvement from other interested parties from the identity federation community in Sweden and the Nordic countries to broaden support for the service.
- The resulting service should be operated by SUNET as part of the general service offering of SWAMID and if possible should be offered as part of the Kalmar2 inter-federation.