Présentation Shibboleth

CREPUQ, Journée BRVQ, 8 septembre 2006, 11:45-12:00.

Par Paul-Emil Provost, Directeur, Bureau des systèmes, Direction des bibliothèques, Université de Montréal

Sommaire :

- Concepts de base
- Fonctionnements actuels : description et problématiques
- La gestion fédérée des identités : concepts.
- Comment ca marche : le processus, les acteurs, la notion de fédération
- Implantation : degré d'adoption / pénétration, processus et calendrier d'adoption, transition.
- Conclusion et questions

Contenu d'exposé :

1. **Objectifs : un** bref survol qui permette de savoir un peu de quel genre de service on parle, des enjeux qui y sont associés et donner des pistes pour poursuivre son appropriation personnelle.
2. **Concepts de base :**

Shibboleth est un protocole de gestion fédérée des autorisations i. e. une façon de gérer à plusieurs partenaires, les processus nécessaires au contrôle d'accès à des systèmes appartenant ou gérés par l'un ou l'autre partenaire, ainsi qu'utilisés par des « membres » d'un ou plusieurs autres partenaires. (Voir leur site pour plus de détails : http://shibboleth.internet2.edu/about.html)

Shibboleth est une façon mieux organisée, plus puissante et sophistiquée que nos méthodes actuelles pour donner accès à nos usagers à des systèmes externes.

Pourquoi un nom pareil ? C'est un mot utilisé dans un récit biblique comme mot de passe : les ennemis, parlant un autre dialecte, le prononçaient mal même lorsqu'ils le connaissaient et s'en trouvaient ainsi démasqués…

Qui dit autorisation dit nécessairement identification, voire authentification. Pour nous aider à bien comprendre :

-*Autorisation* : les droits ou permissions d'un usager d'accéder, puis d'utiliser plus ou moins de fonctions d'un service. Un usager peut être une collectivité, un autre système, ou une personne. Par exemple, dans un système de prêt, les services peuvent être de consulter un dossier d'usager, de lui prêter ou non des livres, de lui annuler ou non ses amendes… Ou encore pour un système de diffusion de périodiques électroniques, de voir toute la série ou que les 3 dernières années.

-*Authentification* : l'identification permet de savoir qui demande le service. Mais puisqu'« A beau mentir qui vient de loin », authentifier est le processus de valider cette identification, de s'assurer que celui qui prétend être Pierre est bien Pierre. L'usager nous donne un identifiant (par exemple, un login, une adresse de courriel, un numéro matricule, un code), puis ajoute un authentifiant, par exemple, son mot

de passe, i.e. un mot qu'il est normalement le seul à connaître (dans un monde plus riche, on lira l'empreinte de sa voix, de son iris, la paume de sa main…). Non seulement une carte d'identité est présentée mais on vérifie si l'usager ressemble à la photo. Pour une institution ou un serveur, on se base souvent sur son adresse réseau mais comme cela peut plus ou moins se forger, on a créé les certificats numériques.

On doit aussi distinguer un identifiant unique i.e. un seul identifiant donné par une institution et utilisé par tous ses systèmes, d'un branchement unique (Single Sign ON (SSO)) i.e. un système qui remplace les modules d'identification/authentification de plusieurs systèmes qui acceptent tous alors une identification opérée par un service qui leur est externe et auquel ils font confiance, i.e. que lorsqu'il y a une demande d'accès, elle est routée vers un système central qui authentifie l'usager et retourne la main au service demandé en lui certifiant l'identité du demandeur.

3. Fonctionnements actuels : description et problématiques

a. Deux grands modèles pour contrôler l'accès à des ressources électroniques :

i. La reconnaissance d'adresse IP de provenance.

Les télécommunications informatiques de par leur conception, comprennent toujours l'adresse informatique de l'ordinateur d'origine (il faut bien savoir à qui répondre…). Avant les accès pan-campus, alors que les bibliothécaires faisaient toutes les recherches pour les usagers, et avant les périodiques électroniques, on utilisait des comptes individuels avec mot de passe. Mais lorsque sont arrivés les systèmes où l'usager lui-même interrogeait la base de donnée, et pire encore les périodiques électroniques, le nombre de compte à gérer se serait multiplié par des facteurs de l'ordre des dizaines de millier, ce qui rendait la méthode des comptes individuels impraticable …. On s'est alors retourné vers cette caractéristique existante des télécommunications, qui nous donnait à peu près ce dont on avait besoin, l'adresse IP : on regarde l'adresse IP d'où vient la demande de service et si elle appartient à une institution avec laquelle nous avons une entente de service, on accorde l'accès.

1. C'était facile à mettre en œuvre avec les moyens technologiques ET administratifs dont on disposait alors.
2. Ça recoupe à peu près la notion légale dans les contrats, de « membre de la communauté universitaire ». En effet, si la personne utilise un ordi sur le réseau campus, il y a de fortes probabilités que ce soit un universitaire…
3. Ça offre cependant peu de modulation, par exemple, ça ne permet pas de différencier des types d'usagers envers des types de services…
4. C'est attaché à la topologie des réseaux, qui elle ne procède pas de la logique de l'appartenance à une communauté mais à une logique de câblage et de distribution de signal….
5. Pour des usagers qui sont hors du réseau institutionnel, en permanence ou à l'occasion, on a continué sur cette même lancée en utilisant une autre technologie de réseau, le serveur

mandataire (proxy), qui parle du campus au nom de l'usager et simule donc une provenance reconnue et agrée par le service externe. Noter que c'est généralement un service à accès contrôlé au niveau de l'individu, avec toutes les contraintes que cela comporte..

ii. Identifiant et mot de passe individuels ou collectifs : c'est une méthode difficile entre autres parce qu'il est lourd de distribuer les identifiants et mots de passe aux bonnes personnes et à les sécuriser vraiment, surtout s'ils sont collectifs i.e. pour toute l'institution (les étudiants ne font généralement que peu de cas de refiler le mot de passe de l'institution au copain de Cegep… et on espère qu'il en va autrement pour les profs et chercheurs….).

b. Problèmes et manques :

i. On a déjà dit que le système dominant de contrôle de la provenance IP, ne permet pas d'offre de service différencié pour nos diverses clientèles, ce qui en conséquence ne nous facilite pas la négociation d'une licence dont nous avons besoin que pour une partie de nos usagers.

ii. Pour obtenir des services personnalisés, il est généralement nécessaire dans les systèmes actuels de contrôle d'accès de transmettre des données personnelles. Ce n'est pas si mal si c'est interne à l'institution, ça devient plus délicat si c'est vers une autre entité, encore plus si c'est à l'étranger où les lois de protection des renseignements personnels sont à tout le moins… différentes. Par ailleurs, le vol d'identité informatique devenant un problème majeur dans nos sociétés, nous avons le devoir social, et l'obligation légale chez-nous, de protéger les informations personnelles qui nous sont confiées.

iii. Et ces problèmes s'ajoutent à ceux qui sont généraux aux systèmes d'identification et au contexte mouvant de l'offre de services télématiques:

1. Il y a de plus en plus de systèmes distribués dont les différentes composantes sont payées par une institution elle-même ou en partenariat, utilisées par des usagers d'une ou plusieurs institutions, qui sont d'un ou plusieurs réseaux, sous une ou plusieurs autorités…
2. Il y a pour chaque individu, prolifération de mots de passe et d'identifiants.
3. Il y a, de plus en plus, perte de sécurité sur des systèmes sensibles parce que les usagers, pour s'en sortir, mettent le même mot de passe partout; pour la banque, les notes de cours, l'accès un ordi de l'université…
4. Il devient difficile d'exiger des mots de passe solides (avec majuscules et minuscules, pas de mots du dictionnaire, à changement fréquents, sans lien avec les infos personnelles classiques…) quand il y en a trop.
5. L'héritage de privilèges suite à un changement de statut n'est généralement pas automatisé.

6. Pour donner un accès personnalisé, chaque système ayant son système d'authentification, on doit souvent maintenir des fichiers parallèles d'usagers, donc répliquer les données, les synchroniser, ce qui est lourd et sujet à délais, erreurs, incompréhensions, ...
7. Nos usagers ne se rendent souvent pas compte que l'accès leur est fourni et payé par leur institution...

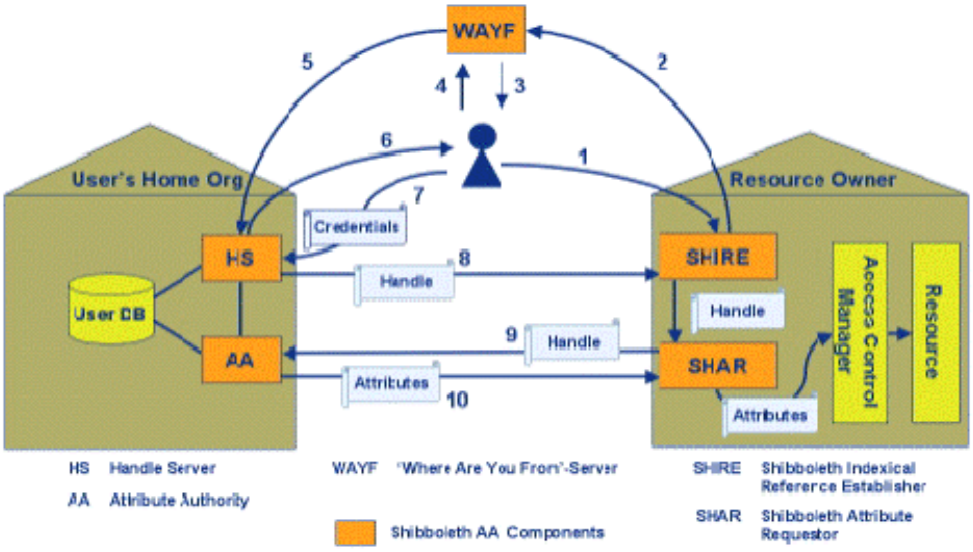
4. La gestion fédérée des identités ; concepts et différentes implantations

Une vraie solution à ces problèmes, un bon outil pour gérer efficacement l'authentification et les autorisations qui en découlent réside dans la gestion fédérée des identités. Quels en sont les principaux éléments :

a. Concepts :
   i. Des usagers identifiés, « nommés » et authentifiés, avec des spécificités reconnues (i.e. des « attributs » : *prof, étudiant, de droit ou d'anthropologie, de telle ou telle université, etc.)*
   ii. Des institutions d'appartenance, conférant des accès / droits / privilèges à des usagers. Dans le jargon, on les appelle souvent des fournisseurs d'identité.
   iii. Des fournisseurs de services (des ressources informationnelles électroniques externes telles Web of science, mais aussi notre Opac, notre dépôt institutionnel, ou un des systèmes distribués, tels Sherlock, etc...)
   iv. Un protocole technique gérant le processus global et les transactions entres les systèmes.
   v. Une entente légale basée sur une relation de confiance permettant la délégation des contrôles d'identité et de statut
b. Alternatives : Shibboleth est la technologie d'authentification distribuée développée par le monde universitaire et qui lui est donc la plus accessible et adaptée. Il y en a d'autres, souvent orientées commerce électronique ou grande entreprise (business to business), telles Liberty, piloté par SUN, etc... On en voit chez les grands fournisseurs publics, par exemple, vous pouvez vous identifier chez Yahoo pour avoir accès à Flickr. Architecturalement, elles se ressemblent beaucoup, c'est surtout la communauté de fournisseurs de services, les services visés et la communauté de fournisseurs d'identité qui sont différents.

5. Comment ca marche (sommairement)

a. Le processus (voir une excellente illustration dans la présentation d'André Béland de la conférence du 6 avril http://www.bibl.ulaval.ca/doelec/crepuq/atelier_06042006/2



i. Un usager de l'UdeM demande l'accès à un service, disons Web of Science.

ii. Celui-ci l'envoie à un service de localisation (WAYF i.e. where are you from) qui lui demande de quelle institution il vient (via sélection dans un menu mais le modèle pourra varier et se perfectionner)

iii. L'usager est redirigé vers son institution d'attache (l'UdeM dans notre exemple) pour y être identifié/authentifié (il reconnaît alors que c'est son institution d'attache qui lui fournit cet accès…), et ce, avec le même identifiant et authentifiant auquel il est habitué sur son campus.

iv. Le service Shibboleth de l'institution d'attache, l'UdeM si on continue notre exemple, authentifie l'usager auprès du fournisseur de service.

v. Au besoin, le fournisseur de service demande les attributs de l'usager, si le service peut s'en trouver modulé.

vi. L'accès est accordé selon le profil de l'usager et les « règles d'utilisation (de prêt…) » de la ressource convenue dans l'entente commerciale.

Ce modèle comporte les avantages suivant :

1. Le fournisseur ne connaît pas l'identité de la personne utilisatrice.
2. Les données personnelles ne sont pas sorties du système Shibboleth de l'institution d'attache.
3. L'usager peut être n'importe où en terme réseau, donc exit le proxy, exit l'impact de la topologie réseau.
4. On PEUT dorénavant donner des accès différenciés SI ON LE VEUT, à différentes clientèles.
5. C'est à large portée, donc une voie vers une identification universelle, au moins dans le monde académique.
6. La gestion des permissions est enfin au plus près de ce qui en décide, le statut universitaire.

b. Les acteurs :

Voyons maintenant qui doit agir et comment pour mettre en place un pareil système :

i. Institution d'attache : doit se donner un service Shibboleth et, au préalable, un répertoire de personnes et une gestion des attributs pertinents. La plupart des campus nord-américains ont du travail à faire au niveau des préalables avant de se donner un serveur Shibboleth pleine puissance, i.e. qui gère aussi la transmission d'attributs qui puissent gérer la qualité du service à offrir…

ii. Fournisseurs de service : doit se doter d'une capacité d'accueil d'une authentification Shibboleth. C'est déjà fait pour la plupart des grands fournisseurs, on le verra plus loin. Sinon, c'est généralement en plan.

iii. Fédérations : on peut imaginer que si chaque institution et chaque fournisseur négocient de un à un, on ne s'en sortira pas. Ça ferait, pour l'UdeM seulement, plusieurs centaines d'ententes à négocier, formaliser, renouveler… Comme il y a forcément des similitudes importantes de besoin et capacité chez plusieurs acteurs, on y gagnera en économie de moyen mais aussi en force de négociation à se regrouper. Les fédérations se créent ici comme ailleurs selon des similitudes d'intérêts, de contexte et de nature. Il se dessine donc principalement deux types de fédérations ayant les principaux rôles suivants:

   1. De fournisseur d'identité : la fédération s'attache à définir des types d'usagers et un contrat type à passer avec les fournisseurs de services, à contrôler les normes au nom de la fédération, etc. C'est elle qui met en place un service WAYF adapté à la nature de ses membres.
   2. De fournisseurs de services : la fédération s'attache à arrêter des niveaux / gammes de service type selon des typologies d'usagers.

c. L'adoption / pénétration.

C'est un standard qui s'impose. En mai 2006, selon le site de Shibboleth (http://shibboleth.internet2.edu/community.html) les fournisseurs sont :

- ArtSTOR
- Blackboard
- Bodington.org
- CSA
- Darwin Streaming Server
- Digitalbrain PLC
- eAcademy
- EBSCO Publishing
- Elsevier ScienceDirect
- ExLibris - SFX
- Fedora
- Higher Markets
- Horde
- Hupnet
- ILIAS
- JSTOR
- Moodle
- Napster
- NSDL
- OCLC
- OLAT
- Ovid Technologies Inc.
- Proquest Information and Learning
- Serials Solutions
- SYMPA
- Thomson Gale
- TWiki
- Useful Utilities - EZproxy
- WebAssign
- WebCT

Et les fédérations existantes sont :

- Australia
- Denmark
- Finland - HAKA
- France - CRU
- Norway - FEIDE (SAML based)
- Switzerland - SWITCH
- UK - SDSS (test)
- UK - (production)
- US - InCommon

d. Autres usages que les ressources de bibliothèques :

On doit noter qu'il y a d'autres usages pour nos campus que celui qui nous intéresse nous. C'est la raison pour laquelle cette implantation doit se faire au niveau institutionnel et non au niveau des bibliothèques. Nommons par exemple :

i. Gestion de privilèges réciproques d'accès : Ex. : Sherlock
ii. Enseignement: 2 profs d'histoire dans 2 universités différentes donnent un cours ensemble à des étudiants de leurs deux institutions et veulent chacun rendre disponibles différentes ressources à l'ensemble des étudiants mais chacun à partir de leur université.
iii. Recherche : une équipe de recherche internationale veut partager des données de recherches
iv. Commerce électronique: contrôle de la diffusion de musique afin que l'accès soit pour la communauté universitaire seulement.
v. Branchement au réseau sans-fil des autres institutions lorsque nous y sommes en visite.
vi. Rédaction et suivi de demandes de subvention, chaque chercheur étant authentifié de son campus mais sur les systèmes de l'organisme subventionnaire.

6. Implantation

C'est bien beau tout ça mais comment le faire, par où commencer et quand est-il pensable d'y arriver est notre prochaine préoccupation.

a. Comment, en théorie :

- Il faut d'abord s'assurer que des conditions préalables soient remplies, i.e.:
  - Que pour chaque campus il y ait une gestion des identités via un système d'authentification compatible à Shibboleth, et que celui-ci comporte les attributs pertinents aux services dont on veut contrôler et/ou moduler l'acès (prof, 1er cycle, programme, etc.)
  - Que les Ressources/systèmes cibles acceptent Shibboleth comme méthode d'authentification.
- Que nous nous donnions ou que nous joignions une fédération, avec les règles qui sont nécessaires pour fonctionner et des définitions de catégories d'usagers communes pertinentes aux services dont on veut contrôler l'accès ensemble.
- Il faut ensuite une entente avec les cibles et/ou via leur fédération quant aux services à offrir à nos usagers des différents types.
- Pour la mise en place technologique, il faut :
  - Installer les modules Shibboleth sur chaque campus
  - Les lier au système d'authentification et aux systèmes campus porteurs des attributs.
- Se donner un scénario de déploiement, où outre les classiques calendriers et mises en place techniques, on prévoit la gestion des problématiques particulièrement cruciales dans ce contexte de :

- La gestion de la transition et de la cohabitation de différents modes,
- De l'information, de la documentation et du soutien à l'usager. Noter toutefois que l'authentification fédérée est souvent plus facile pour l'usager : on lui présente toujours le même écran d'identification et le reste du processus lui est transparent.

b. Quand ce système risque-t-il d'apparaître dans notre environnement ?

Le sous comité des bibliothèques a reçu une recommandation de créer un Groupe de travail CREPUQ comportant des acteurs des bibliothèques et des services informatiques afin de se donner, comme vient notamment de le faire l'Ontario, un plan d'appropriation. Concrètement et au mieux, nous pourrions voir naître des projets pilotes dans 6 mois pour des services tel l'accès sans-fil inter-campus pour des clientèles ciblées. Ce ne sera probablement pas dominant pour l'accès aux ressources documentaires avant quelques années mais la tendance est là, lourde, inévitable, et nous devons nous y préparer.

7. **Références** : site WEB Journée CREPUQ 13 avril 2006.
http://www.bibl.ulaval.ca/doelec/crepuq/atelier_06042006/