Appunti sui lavori di Science Gateways e Grid Portals svolto a INFN-CATANIA

Premessa

La seguente valutazione non vuole essere esaustiva sulle caratteristiche dei tool sviluppati da INFN-CATANIA, ma solo evidenziare alcune delle scelte tecniche adottate, soprattutto in relazione al contesto e ai requirement di partenza.

Come referenze per la presente valutazione ho preso in considerazione due pubblicazioni molto recenti:

1. "Conjugating Science Gateways and Grid Portals into e-Collaboration environments: the Liferay and GENIUS/EnginFrame use case" (http://goo.gl/hhSto);
2. "The GENIUS Grid Portal and robot certificates: a new tool for e-Science" (http://goo.gl/FwfNQ);

Altri dettagli li ho chiariti personalmente con Giuseppe La Rocca e Roberto Barbera.

Lo scopo di questo lavoro è formulare delle prime valutazioni tra l'esperienza maturata nel contesto dei portali dai catanesi in diversi progetti, nazionali ed internazionali, e le esigenze emerse recentemente nel contesto della collaborazione DUCK (http://www.comput-er.it/), nello sviluppo del progetto WNoDeS (http://web.infn.it/wnodes/) e nelle esperienze in corso della VO superb (http://web.infn.it/superb/).

In particolare al CNAF è iniziata la volutazione di "grid User Support Environment" (http://www.guse.hu/) ovvero WS-pgrade (evoluzione di p-grade, basata su Web Service) basato su Liferay.

Science gateway o Grid portal?

Occorre prima di tutto chiarire i concetti di Science Gateway e Grid Portals. Proporrei di adottare quelli utilizzati negli articoli usati come referenza:

"A science gateway *is a framework of tools that allows scientists to run applications with little concern for where the computation actually takes place. This is similar to cloud computing in which applications run as Web services on remote resources in a manner that is not visible to the end user. However, a science gateway is usually more than a collection of applications. Gateways often let users store, manage, catalog, and share large data collections or rapidly evolving novel applications they cannot find anywhere else. Training and education are also a significant part of some science gateways"*

"**Grid Portals** *are developed to provide an easy and intuitive way to access services offered by a Grid infrastructure. Researchers do not have to waste their time learning several commands or programming language in other to use Grid services and can focus totally on their studies."*

Nelle discussioni spesso si alternano le due diciture, ma penso che ci si riferisca sempre al concetto collegato al Science Gateway, quindi un portale che offre una serie di servizi agli utenti, tra i quali la possibilità di utilizzare servizi e risorse accessibili via Grid. Il lavoro nel quale sono impegnati i catanesi, in particolare nell'ambito del progetto DECIDE (http://www.eu-decide.eu/), è quello di offrire i servizi al momento presenti nel Grid Portal GENIUS anche in un Science Gateway. Come prodotto per la realizzazione del science gateway hanno deciso di adottare ***Liferay***.

e-Collaborative Grid Portal architecture

Uno dei punti di forza di Liferay è l'adozione della tecnologia delle **portlet** (http://en.wikipedia.org/wiki/Portlet) definite da Java Specification Request (JSR 168 e 268), compatibile con le più moderne applicazioni web based.

Nella Figura 1 è mostrata l'architettura a layer prevista dal porting di GENIUS su Liferay, nella quale risulta centrale la componente ***EnginFrame***, sviluppata da Nice. Nel porting in questione, la versione prevista è EnginFrame 2010 che supporta le tecnologie Web 2.0 (non supportate dalla versione 4.1 attualmente utilizzata in GENIUS) e che "*will support portlets very much like liferay or GridSphere* [primo articolo, 3.2]". Le portlet sviluppate fino ad ora da INFN-CATANIA e NICE sono utilizzabili non solo con liferay, ma anche con altri portlet container come GridSphere.

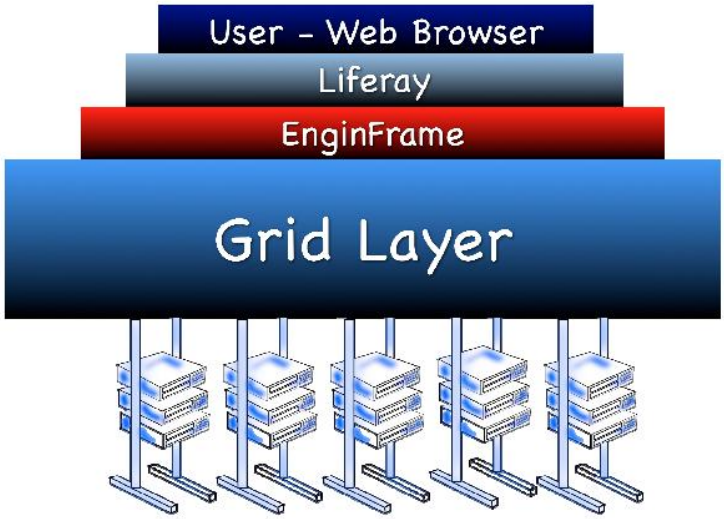

Figura 1. e-Collaborative Grid Portal architecture

Al momento il portale va installato su una UI e EnginFrame si interfaccia con i comandi della UI per la gestione dei job e dei servizi gLite. Il ***Grid Layer*** è pensato per contenere le API parte del middleware della UI per creare applicazioni Grid senza la necessita' di macchine dedicate, ovvero, da quello che capisco, senza la necessita' di installare il portale su una UI. Il Grid Layer è pensato quindi per soddisfare due scenari:

- *"At a higher level, designing a standard interface to the Grid Layer will allow third party applications to access the Grid services even without the need of Liferay and GENIUS/EnginFrame;"*
- *"At a lower layer, the new Grid Layer could be customized for middleware different from gLite such as VDT, used by Teragrid, or others."*

Questa distinzione risulta essere particolarmente interessante per instaurare eventuali collaborazioni nello sviluppo di API nel Grid Layer. Nei Future Plans, gli autori del primo articolo indicano in due mesi il tempo per la realizzazione delle funzionalità previste nel Grid Layer. L'articolo è dell'Agosto 2010, occorre verificare il punto su questa attività.

Single sign on

Riguardo ai metodi di autenticazione al portale, gli articolo non dicono molto dal punto di vista tecnico, ma solo che la single sign on è considerato un requirement.

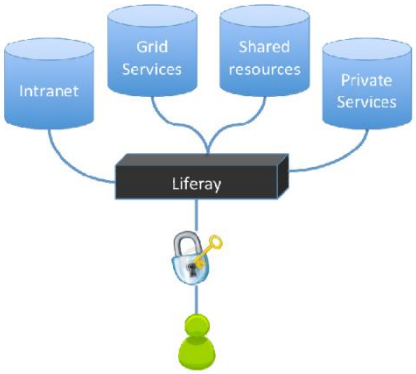

Figura 2. L'architettura del science gateway basato su Liferay e EnginFrame

Giuseppe mi ha riferito che la registrazione al portale sara' obbligatoria e che i privilegi degli utenti registrati saranno gestiti tramite ldap. Un meeting tra di loro per definire le modalita' e le tecniche di registrazione è previsto per Lun 14 Marzo.

Use cases

Utenti senza certificato digitale

Questo è lo use case principale.

1. **L'utente si registra al portale.** In discussioni i dettagli sulle possibili modalita' da adottare per questa fase.
2. **Certificati Robot.** L'utilizzo trasparente dei certificati x509 da parte degli utenti è risolto mediante l'utilizzo di Certificati Robot ("*A personal credential which can perform automated tasks on behalf of the use*" da http://security.fi.infn.it/CA/CPS/CPS-2.3.1.pdf) ufficialmente riconosciuti da IGTF e il cui scopo è proprio quello di permettere agli utenti che non sono familiari con la PKI usata in Grid e il concetto di VO, di utilizzare le risorse dell'infrastruttura Grid diminuendo le barriere iniziali. L'architettura prevede l'associazione di un certificato Robot per ogni applicazione che è possibile sottomettere dal portale. L'utente viene guidato dal portale passo passo per la creazione di un proxy a partire dal certificato Robot legato all'applicazione che vuole utilizzare. L'autorizzazione ad utilizzare o meno le applicazioni presenti sul portale sono definite nel profilo dell'utente, gestito via ldap. Il certificato Robot è stato precedentemente registrato in un VOMS server.
3. **Iterazione con Grid.** L'utente a questo punto viene guidato dal portale e può fare solo determinate operazioni predisposte per lui, come per esempio la sottomissione di una determinata applicazione (magari modificando solo alcuni parametri via web form). Non è previsto che l'utente carichi proprie applicazioni o crei dei workflow diversi da quelli previsti. In questo modo, problemi derivanti da possibili abusi da parte degli utenti che potrebbero portare al banning del certificato sono quasi nulli, poiché' i gradi di libertà sono limitati dal portale stesso.

Utenti con certificato digitale (e registrati in una VO)

Questo use case è considerato secondario, essendo l'obiettivo primario del lavoro offrire la possibilita' di eseguire determinate applicazioni utilizzando le risorse presenti in Grid in maniera trasparente.

1. **L'utente si registra al portale.** (vedi sopra)

2. **L'utente ha un certificato personale.** Nella versione attualmente in produzione del portale GENIUS (quindi senza liferay), era gia' previsto l'uso dei certificati robot, ma nel caso questi non fossero stati per un qualche motivo disponibili, era possibile una autenticazione basata su ***Java applet*** nella quale viene richiesto all'utente di specificare il proprio certificato personale e di inserire la password in modo da generare, localmente, un proxy. Questo proxy viene quindi messo su un myproxy server. Direi che quella della Java applet è la stessa soluzione adottata anche dai pisani. Una volta caricato il proxy, è compito dell'utente selezionare una VO, in modo da integrare nel proxy le estensioni voms. Questo implica che è compito dell'utente essersi precedentemente registrato in una VO.
3. **Iterazione con Grid.** Vedi sopra. Nonostante un utente usi il proprio certificato, può sottomettere solo le applicazioni previste dal portale, ma è pensabile offrire qualche grado di libertà in più permettendo l'upload di propri script e jdl.

Considerazioni finali

Liferay-EnginFrame-Genius vs Liferay-WS-Pgrade

- Il portale Genius è una realtà che si è consolidata nel tempo e che risponde ai requirement delle comunità degli utenti che lo utilizzano. La sua evoluzione basata su Liferay/EnginFrame ha lo scopo di aumentarne la flessibilità e le possibilità di customizzazione e integrazione con alti servizi agli utenti. Il gruppo di persone che ci lavora si è consolidato nel tempo e dispone di fondi, requirement e scadenze legate ai progetti nell'ambito dei quali avviene lo sviluppo.
- La soluzione basata su Liferay e WS-Pgrade, oggetto di analisi al CNAF, dispone a sua volta di un gruppo di lavoro consolidato. P-Grade in particolare ha avuto molta fortuna ed è stato adottato per l'implementazione di portali per diverse comunita' (WS-Pgrade è molto piu recente, rilasciato la scorsa settimana). L'esperienza del CNAF su questo prodotto è minima.

Registrazione al portale

- I dettagli per la registrazione degli utenti nel portale Liferay-EnginFrame-Genius devono essere approfonditi e sono oggetto di una riunione dei catanesi a breve. In Genius e al momento, il riconoscimento e l'approvazione sono a carico dei gestori del portale.
- Una cosa che si vuole evitare nella soluzione analizzata al CNAF è di dover prendersi carico dell'approvazione degli utenti. Per questo motivo la registrazione al portale, comunque necessaria, sara' basata su IDP esterni precedentemente abilitati: quindi, gli utenti di una federazione abilitata, potranno essere registrati in automatico. La registrazione al portale è necessaria per conservare nel portale alcune informazioni sugli utenti non disponibili direttamente dagli attributi necessari al riconoscimento (per esempio, appartenenza ad una o piu' VO, certificato personale, informazioni su attivita' relative a servizi offerti dal portale).

⇨ Possibile campo di cooperazione: diverse modalita' di registrazione ai portali.

Utenti senza certificato

- La soluzione adottata è quella dei Certificati Robot, che sono una realtà e sono nati a questo scopo. Semplificano moltissimo numerosi aspetti di security del portale.
    - Un Certificato Robot per applicazione;
    - Registrazione nella VO a carico di chi gestisce i certificati Robot;
    - Gli utenti possono girare solo le applicazioni previste dal portale;

    - L'accounting per utente non è completamente integrato nel sistema di accounting di EGI.
        - Poiche' i certificati Robot sono una realta' non usata solo dai catanesi, varrebbe la pena collaborare, a prescindere, su come integrare le info di accounting presenti nel portale con quelle di DGAS. Per tenere traccia dell'attivita' degli utenti del portale, i catanesi hanno sviluppato lo Users Tracking System (UTS) che in particolare ha la corrispondenza tra jobid (e quindi anche particolare proxy usato dall'utente) e utente stesso. Queste informazioni potrebbero arrivare a DGAS. Un altro modo potrebbe essere quello di generare i proxy dell'utente, partendo dai Robot, con l'opzione *–label* in modo da aggiungere delle info nel DN del proxy cosi' generato. L'info potrebbe essere il nome e cognome dell'utente. Soluzioni da indagare.
- La soluzione identificata prevede l'utilizzo di una CA Online (utilizzando la funzionalita' dedicata del Myproxy) per il rilascio dei certificati agli utenti che si registrano nel portale.
    - Quando **l'utente si registra al portale**, in maniera trasparente gli viene generato un certificato personale rilasciato dal server Myproxy che fa anche da CA Online.
        - Dal certificato personale viene generato un proxy della durata del certificato e successivamente distrutto il certificato personale.
    - Il certificato viene registrato in una VO catch all predefinita.
    - Quando **l'utente accede al portale**, viene creato un proxy, derivato da quello di lunga durata creato in fase di registrazione, in maniera trasparente.

Il problema maggiore di questa soluzione è che al momento non abbiamo una CA online accreditata, inoltre lo scenario descritto deve essere sviluppato. Da investigare inoltre la possibilita' e il workflow per l'appartenenza a piu' di una VO.

⇨ Possibili campi di cooperazione:
- estensione di DGAS per pieno supporto dei certificati Robot nel caso di user accounting;
- utilizzo della CA Online.

Utenti con certificato

- Questo non è lo use case principale. L'idea è astrarre dai certificati personali mediante l'utilizzo dei Robot. Nel portale Genius era disponibile una Java Applet che permetteva all'utente con certificato di generare localmente un proxy e caricarlo su un server Myproxy. Tale Applet potrebbe essere utilizzata anche nelal soluzione basata su liferay. La richiesta dei certificati personali e la registrazione in una VO sono quindi totalmente manuali e a carico degli utenti.
- Da analizzare la soluzione alla WS-Pgrade per caricare un proxy sul server Myproxy per vedere se è corretta in termini di security. La registrazione in una VO è a carico degli utenti.

⇨ Possibili campi di cooperazione:
- Essendo Liferay un portlet container, si potrebbe sviluppare una soluzione per a creazione e la gestione dei proxy che sia corretta in termini di security (se risultasse che l'operazione alla WS-Pgrade non va bene);