Grid-based Secure Web Service Framework for Bioinformatics

Dawei Sun, Xiaoyu Zhang[†]

Department of Computer Science

California State University San Marcos

San Marcos, CA 92069

ABSTRACT

Although the web-based bioinformatics is very popular after dozens of years' growth, it is inconvenient because biologists may need to access many web sites manually in order to perform a single task. Web-service based bioinformatics was proposed to provide well-defined and program-accessible interfaces. However, security for web services is a very important issue that was not addressed in most web-service based bioinformatics systems. We developed a Grid-based Secure Web Service Framework for Bioinformatics (GSWSF). The GSWSF is designed based on the Open Grid Service Architecture (OGSA) [5] and Grid Security Infrastructure (GSI) [6], which provide two security mechanisms: transport level security and message level security. We can build secure and easy-to-use bioinformatics services using this framework. This paper covers the architecture and some design and implementation details of the framework. A preliminary implementation of the framework can be found at *http://bioinfo.csusm.edu*.

INTRODUCTION

Nowadays the web based bioinformatics is widely used. However, the web pages are designed to be easily read by human beings but it is hard for programs to grab data directly from them. A parser need be written to retrieve useful data from every HTML formatted result of a web query. Furthermore, the formats of results are usually not well documented and subject to change.

Web services are based on XML-formatted SOAP over HTTP. The communications between computers are well-formatted and can be easily understood by programs. Furthermore, it can pass through firewalls or gateways easily. Using web service, users can aggregate biological data and applications simply and reliably. On the other hand, web service based bioinformatics could also provide the familiar human-friendly user interfaces.

Security is always an important issue in web-service applications. the web services are exposed to the public but we want to allow the authorized accesses for some services, for example those changing the states of internal databases. Therefore web services must be secure. Administrators who manage the web services can grant privileges to verified users to a set of secure web services. Authorized users are then authenticated by their credentials before accessing the secure web services.

In order to provide secure and more convenient access to the data and tools in the bioinformatics project, we developed a Grid-based Secure Web Service Framework (GSWSF) at CSUSM. Secure web services and web interfaces are developed for updating databases and accessing computational tools.

GSWSF

The GSWSF framework is based on the concepts of Grid computation. The secure web services in this framework are implemented using the Grid toolkit GLOBUS [1].

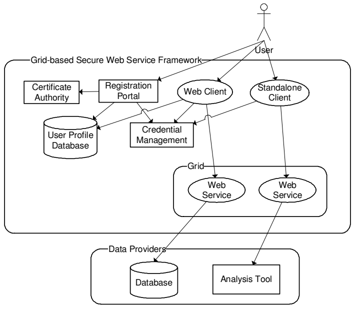


Figure 1. Component diagram of GSWSF

There are five major modules in GSWSF:

- Certificate Authority Module,
- Credential Management Module,
- Registration Portal Module,
- Secure Web Service Module,
- Client Module.

Certificate Authority: This module is responsible for generating and signing user credentials. The GSWSF is designed to hide the complex details of credentials from users. Users even do not need to know what are exactly in their credentials. We choose SimpleCA [3] for the certificate authority module in GSWSF because it is free and simple.

Credential Management: Every registered user must have one credential by which he can access the secure web services. The safety of credentials is critical to keep services secure. Credentials are more securely stored in a credential management system than users' own disks. Users can then access their credentials using a username-and-password pair and never need to know where the credentials actually are. Client programs can automatically retrieve proxies of the credentials when they need to access secure web services. MyProxy [2] is chosen for implementing the credential management module in the GSWSF.

Registration Portal: All secure systems must offer mechanisms of user sign-up, login, and logout. In GSWSF, we designed a specific module called registration portal to deal with registration issues. This module integrates with the certificate authority and credential management modules. The registration portal manages a registration procedure to verify users' identities. After a user is verified, the registration portal module calls certificate authority to generate and sign the user's credential and stores the signed credential in MyProxy. The user's information is also stored in the user profile database. We modified PURSe [4] to be the registration portal in the GSWSF.

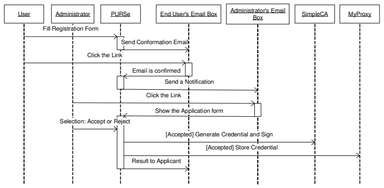


Figure 2. Sequence diagram of a user registration

Secure Web Services: This is the core module of GSWSF, which includes all secure web services published within GSWSF. The interfaces of web services are defined in WSDL files. Web services can be a simple wrapper or computational intensive applications. For example, a secure service can wrap insert, update, and delete operations to a sequence database and other services can provide program-accessible interface to commonly used bioinformatics tools such as BLAST. A secure web service module depends on multiple underlying data providers and/or other existing web services. So the web services in this module can be nested and interconnected. The secure web services developed in the GSWSF are based on the web-service components in GLOBUS [1]. New services can be easily added into the framework.

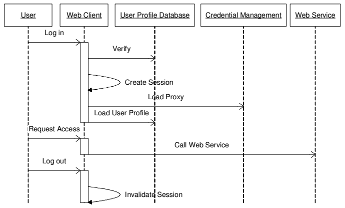
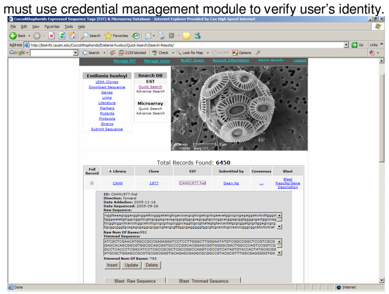


Figure 3. Sequence diagram of web client accessing secure web services.

Clients: Client module is the most diverse and complicated module in the GSWSF. It depends on web service and credential management modules. The responsibility of clients is to provide a user-friendly interface to the web services. A client can call multiple web services. The GSWSF supports two kinds of clients, web clients and standalone clients. The web client can verify the user's identity using user profile database; the standalone client must use credential management module to verify user's identity.


Coccolithophorids Expressed Sequence Tags (EST) & Microarray Database - Internet Explorer Provided by Cox High Speed Internet
File Edit View Favorites Tools Help
Back Search Favorites
Address http://bioinfo.csusm.edu/Coccolithophorids/Emiliania-huxleyi/Quick-Search/Search-Results/
Go Links
Google Search 2155 blocked Check Look for Map AutoFill Options
Manage EST Manage Users BLAST Query Account Information Admin Bioinfo Logout
Emiliania huxleyi
cDNA Clones
Download Sequence
Genes
Links
Literature
Markers
Mutants
Protocols
Strains
Submit Sequence
Search DB
EST
Quick Search
Advance Search
Microarray
Quick Search
Advance Search
Magn 10000x marg 127 2 µm
Insert Update Delete
Blast: Raw Sequence Blast: Trimmed Sequence
Done Internet

Total Records Found: **6450**

Full Record	Library	Clone	EST	Submitted by	Consensus	Blast
⊟	CAHN	1977	CAHN1977.fwd	Dean Ng	...	Blast Results/Gene Description

ID: CAHN1977.fwd
Direction: forward
Date Addedion: 2005-11-16
Date Sequenced: 2005-09-26
Raw Sequence:

```
nggttaaagcggacggtcggattccgggatatcgtcgacccacgcgtccgatcgctcgaacatggccgccgagaggatcctccttgggct
tgggaatatgtcggccggctccgtcgcggagcacagcggcgtggcgcagcggtgccccggcacggagcggtggggcgactggcccag
tccggtcggcctcaccctcggcatcctcgccgcgctcgccggccaggtcgccgtcatagtgtaccactatgcgcggatgcgctggagccgcg
tgcggcgggtgcagagcgaggcgcggccgtacgcgtttggcgaggggggtggcgtcgcacctcgccaacccgggcggcatcctcctcat
```

Raw Num Of Bases: 992
Trimmed Sequence:

```
ATCGCTCGAACATGGCCGCCGAGAGGATCCTCCTTGGGCTTGGGAATATGTCGGCCGGCTCCGTCGCG
GAGCACAGCGGCGTGGCGCAGCGGTGCCCCGGCACGGAGCGGTGGGGCGACTGGCCCAGTCCGGTCG
GCCTCACCCTCGGCATCCTCGCCGCGCTCGCCGGCCAGGTCGCCGTCATAGTGTACCACTATGCGCGG
ATGCGCTGGAGCCGCGTGCGGCGGGTGCAGAGCGAGGCGCGGCCGTACGCGTTTGGCGAGGGGGTGG
```

Trimmed Num Of Bases: 793

Figure 4. GSWSF implementation at http://bioinfo.csusm.edu

Figure 4 shows the web client interface to the secure web services at CSUSM (http://bioinfo.csusm.edu). The insert, delete, update buttons for changing databases are only available if an authorized user logs in.

The authors thank Dr. Read, Dr. Wahlund, Dr. Hadaegh, and Mr. Ranji at CSUSM for useful discussions and assistances.

References

[1] The Globus Alliance and Toolkit, http://www.globus.org/
[2] MyProxy Credential Management, http://grid.ncsa.uiuc.edu/myproxy/
[3] SimpleCA, http://www.vpnc.org/SimpleCA/
[4] Portal-Based User Registration Service (PURSe), http://www.grids-center.org/solutions/purse/
[5] Open Grid Services Architecture, http://www.globus.org/ogsa/
[6] Grid Security Infrastructure http://www.globus.org/security/overview.html

[†] *Email: xiaoyu@csusm.edu*