Science Gateway Security Recommendations

Jim Basney
National Center for Supercomputing Applications
University of Illinois
Urbana, Illinois, USA
jbasney@illinois.edu

Von Welch
Center for Applied Cybersecurity Research
Indiana University
Bloomington, Indiana, USA
vwelch@indiana.edu

***Abstract*—A science gateway is a web portal that provides a convenient interface to data and applications in support of a research community. Standard security concerns apply to science gateways, including confidentiality of pre-publication research data, integrity of research results, and availability of services provided to researchers. In this paper we identify existing science gateway security recommendations and provide our own perspective.**

Keywords—science gateways; security

I. Introduction

Science gateways must address a range of security issues to provide trustworthy service to researchers, maintain the trust of external resource providers, and properly use external resources in compliance with security policies. In this short paper we reference prior work providing science gateway security recommendations and add our recommendations based on our experience.

II. Science Gateway Models

Science gateways have at least three different deployment models in terms of how they interact with external resources:

1. A science gateway may support only users of the external resources, meaning the science gateway basically provides a different interface to the resource(s) without otherwise changing the user management processes of that resource. An example of this sort of science gateway is the TeraGrid Visualization Gateway [6].

2. A science gateway's resources maybe entirely dedicated to the science gateway and managed by it. In this case the science gateway manages its own users. An example of this sort of science gateway is The Rosetta Online Server That Includes Everyone (ROSIE) [7].

3. A science gateway may have its own user community that runs in a *community account* on the resources that service it using *community credentials* or *robot certificates*. This case, described in more detail in [3], represents a complex trust relationship between the science gateway and the resource provider, with the provider having delegated some security and user management functions to the science gateway. The science gateway may or may not expose user identity information to the resources (for a full discussion, see [5]). Examples of this sort of science gateway include CIPRES [8] and GENIUS [9].

In this paper we discuss security requirements of all three models, highlighting instances where the different models stress those requirements in different ways.

III. Risk Assessment

Each science gateway is unique in the community it serves, the capabilities it provides, the sensitivity of its data, and the underlying software components and infrastructures it builds upon. Understanding these unique aspects helps to assess the security risks associated with a science gateway. Fig. 1 lists some risk factors of particular importance to science gateways.

Science Gateway Risk Factors

- Use of external resources (e.g., supercomputers, data archives, scientific instruments)
- Authentication and vetting of users
- Capabilities provided to users (e.g., types of applications, input data, resource limits)
- Data privacy (e.g., health data, pre-publication data)

Fig. 1. Factors to consider in a science gateway risk assessment.

A science gateway that serves a large, distributed, open user community has greater challenges of user management than a science gateway that serves a small, closely-knit user community. A science gateway that supports a wide range of user capabilities (for example, unvetted scientific codes uploaded by users) has additional risks to manage than a science gateway that provides more limited functionality. A science gateway that analyzes personal health information requires greater privacy protections than a science gateway that analyzes astronomy data. A science gateway that uses external resources (supercomputers, data archives, scientific instruments, etc.) must understand the risks that the science gateway brings to those resources and the risks that those resources bring to the science gateway. These risks can be addressed through security policies and service agreements.

The EGI-InSPIRE Security Policy Group's VO Portal Policy [2] provides an example of an agreement between science gateways and infrastructure operators to address shared risks. The policy provides a classification of science gateway portals according to the capabilities they provide to users. As the level of access provided increases (from "one-click" portals

This material is based upon work supported by the National Science Foundation under grant numbers 1127210 and 1234408.

to “parameter”, “data processing”, and “job management” portals), the requirements on user identification are increased. The policy also includes requirements on limiting the rate of job submissions, keeping audit logs, assisting in security incident investigations, and secure storage of passwords and private keys.

Hazlewood and Woitaszek [1] provide a risk assessment with security recommendations for science gateways using TeraGrid (now XSEDE). Their recommendations include user accounting at the science gateway, limiting access to external resources (using restricted shells and dedicated interfaces), and use of short lived certificates for authentication to external resources. They survey the security policies across 10 resource provider sites and implementations across 20 science gateways, for a broad picture of community practice. In the following sections we build on their recommendations to address additional identity management and operational issues.

IV. Identity Management

Most science gateways support user authentication for personalization, managing user information across sessions, tracking usage, and providing authenticated access to external resources. In some cases resource providers (for example, XSEDE), require science gateways to authenticate users for accounting purposes.

As with any web service, securely managing user passwords in a science gateway brings significant security risks. Science gateways can avoid managing user passwords by supporting federated authentication, via SAML, OpenID, and/or OAuth, so users can log in to the science gateway using their existing credentials. Fig. 2 provides an overview of considerations for using federated identities in science gateways.

Using Federated Identities

- Avoids risk of managing passwords directly
- Benefits from dedicated identity provider security
- Often supported by web application frameworks
- SAML: good for supporting use of campus identities
- OpenID/OAuth: good for users without SAML IDs

Fig. 2. Considerations for using federated identities in science gateways.

The underlying web application frameworks used by science gateways often include built-in support for federated authentication, simply requiring the capability to be configured and enabled or in some cases an add-on package to be installed. The OpenID and OAuth protocols allow users to log in to the science gateway using Google, Facebook, Twitter, or other commonly used credentials. XSEDE also supports an OAuth interface for logging in to science gateways using XSEDE credentials [4]. SAML support enables users to log in using their university credentials via federations such as InCommon in the United States. In some cases users do not have access to a university SAML identity provider, so supporting OpenID/OAuth logins avoids excluding these users. Another benefit of using federated authentication is access to ongoing security advances in the identity management community, such as the recent adoption of two-factor authentication by identity providers such as Google and Twitter.

If a science gateway chooses to manage user passwords itself, e.g., to avoid the reliance on external identity providers, the science gateway should handle the passwords securely. Any input of passwords to the science gateway must be encrypted using HTTPS to protect against eavesdropping attacks, and passwords must be stored in hashed or encrypted form to provide some protection in case the password database is disclosed. Science gateways should use a secure password hashing library provided by their platform, such as PHP’s password_hash() function, which by default uses a per-password random salt and a strong hash algorithm (currently bcrypt, updated over time). Enabling password strength checks (e.g., CrackLib) and online password guessing protection (e.g., Fail2Ban) is also important.

V. Operational Security

In this section we provide a short review of operational security recommendations for science gateways. The goal of these recommendations is to manage the risks related to threats, both in terms of reducing the likihood of those threats being realized and mitigating the impact if they do. Hence, the recommendations address preventing threats, detecting their realization, and enabling an effective response. Fig. 3 provides an operational security checklist for science gateways.

At the top of the list is our recommendation that science gateway developers and operators communicate with operational security staff at their local organization when developing their science gateway. Local security staff can often provide assistance with security services (monitoring, scanning, logging, etc.), security policies, and best practice recommendations tailored to the local environment. Established relationships with local security staff are also critical in the event of a security incident.

Operational Security Checklist

- Software patching
- Controls on administrator access
- Vulnerability scanning
- Centralized logging
- Secure backups
- Firewalls
- Physical security of servers
- File integrity checking
- Intrusion detection
- Log monitoring

Fig. 3. Operational security considerations for science gateways.

For preventing threats, we recommend promptly applying software security updates, restricting access using firewalls, disabling unneeded operating system services, requiring strong authentication for administrative access (two factor authentication, use of a bastian host), and proper management of administrative access (periodically reviewing the list of administrators, removing administrators when they leave). In our experience, compromise of science gateway administrator or developer computers is a common attack vector for compromising science gateway servers and external resources, so maintaining the security of administrator/developer computers and controlling and monitoring access from these systems is particularly important. Lastly, we recommend controlling physical access to science gateway servers (locked racks, access controlled machine rooms) and periodic vulnerability scanning of science gateway servers (e.g., Nmap, OpenVAS).

For detecting security incidents, we recommend file integrity checking (e.g., samhain), host and network intrusion detection (e.g., OSSEC and Bro IDS), and log monitoring and analysis (e.g., sawmill).

To facilitate effective incident response, we recommend maintaining accurate system clocks for good log timestamps (via NTP), logging to a central log collector in case local system logs are modified by an attacker, performing regular secured system backups to enable disaster recovery, and maintaining an incident response plan that includes procedures for handling media inquiries regarding security incidents.

VI. Conclusions

Best practices for science gateway security include the standard recommendations for any online service, plus science gateway-specific concerns such as securely accessing external resources like XSEDE. Risk profiles differ across science gateways, and a risk assessment that considers the unique aspects of each science gateway helps to identify security mechanisms and policies with the proper balance of ensuring confidentiality, availability, and integrity while providing a convenient interface for researchers.

Multiple groups are available to assist science gateway developers and operators with security issues. XSEDE (https://www.xsede.org/gateways) has a vibrant science gateway community where security issues are discussed. Additionally, the Center for Trustworthy Scientific Cyberinfrastructure (http://trustedci.org/) and the Distributed Web Security for Science Gateways (http://www.sciencegatewaysecurity.org/) projects can provide additional advice and recommendations for science gateway security issues on request.

Acknowledgment

The authors gratefully acknowledge the input provided by Terry Fleury, Surresh Marru, Jim Marsteller, and Anand Padmanabhan on the topic of this paper, as well as the comments of the reviewers.

References

[1] V. Hazlewood and M. Woitaszek, “Securing Science Gateways,” TeraGrid Conference, July 2011, Salt Lake City, Utah, USA. http://doi.acm.org/10.1145/2016741.2016781

[2] D. Kelsey, “VO Portal Policy,” EGI-InSPIRE Security Policy Group, July 2010. https://documents.egi.eu/document/80

[3] J. Basney, V. Welch, and N. Wilkins-Diehr, "TeraGrid Science Gateway AAAA Model: Implementation and Lessons Learned," TeraGrid Conference, August 2010, Pittsburgh, PA. http://dx.doi.org/10.1145/1838574.1838576

[4] J. Basney and J. Gaynor, "An OAuth Service for Issuing Certificates to Science Gateways for TeraGrid Users," TeraGrid Conference, July 18-21, 2011, Salt Lake City, UT. http://dx.doi.org/10.1145/2016741.2016776

[5] R. Cowles, C. Jackson and V. Welch, “Identity Management for Virtual Organizations: A Survey of Implementations and Model,” 9th IEEE International Conference on e-Science, October 2013, Beijing, China. http://www.vonwelch.com/pubs/VOIdM13

[6] J. Insley, T. Leggett, and M. Papka. “Using dynamic accounts to enable access to advanced resources through science gateways,” In Proceedings of the 5th Grid Computing Environments Workshop (GCE '09). http://doi.acm.org/10.1145/1658260.1658279

[7] S. Lyskov et al, "Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE)". PLoS One. 2013 May 22;8(5):e63906. doi: 10.1371/journal.pone.0063906

[8] M. Miller, W. Pfeiffer, and T. Schwartz, “The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources,” In Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond (XSEDE '12), July 2012, Chicago, IL, USA http://doi.acm.org/10.1145/2335755.2335836

[9] R. Barbera, G. Andronico, G. Donvito, A. Falzone, J. J. Keijser, G. La Rocca, L. Milanesi, G. P. Maggi, and S. Vicario, “A grid portal with robot certificates for bioinformatics phylogenetic analyses,” Concurrency and Computation: Practice & Experience, Vol. 23, Issue 3, March 2011. http://dx.doi.org/10.1002/cpe.1682