The role of the collaboratory in enabling large-scale identity management for HEP

Robert Cowles, Craig Jackson and Von Welch
Center for Applied Cybersecurity Research, Indiana University, 2719 E. 10th Street, Suite 201, Bloomington, IN 47408, USA

Email: bob.cowles@gmail.com, scjackso@indiana.edu, vwelch@iu.edu

Abstract. The authors are defining a model that describes and guides existing and future scientific collaboratory identity management implementations. Our ultimate goal is to provide guidance to virtual organizations and resource providers in designing an identity management implementation. Our model is captured in previously published work. Here, we substantially extend our analysis in terms of six motivation factors (user isolation, persistence of user data, complexity of virtual organization roles, cultural and historical inertia, scaling, and incentive for collaboration), observed in interviews with community members involved in identity management, that impact implementation decisions. This analysis is a significant step towards our ultimate goal of providing guidance to virtual organizations.

1. Introduction

The Virtual Organization (VO) [1] has emerged as a fundamental way of structuring HEP collaborations and relationships with the resource providers (RPs) that support those collaborations. Prior to the emergence of the VO, RPs had an unmediated relationship with their user communities, and therefore handled all aspects of identity management. The distributed and heterogeneous nature of the computing resources and unique position of the VO in negotiating and managing community relationships has resulted in some identity management (IdM) tasks being delegated from RPs to the VO. The last two decades have seen considerable variety in the ways different VOs and RPs have distributed IdM tasks and responsibilities.

In a paper written by the authors for the eScience 2013 conference [2], we presented two contributions to IdM in the VO context: (1) we summarized 18 semi-structured interviews [3] we conducted with VOs or RPs regarding how IdM was implemented and the bases for key implementation decisions; (2) we distilled common parameters and values to provide a structured VO IdM Model of the different IdM implementations based on our analysis of the notes from those interviews along with existing literature, online presentations and documentation.

The long-term goal of the authors is to extend the VO IdM Model to provide not only description, but also guidance in making implementation choices to best meet trust and risk requirements of VOs and RPs. Our eScience paper was focused on developing the descriptive Model. In this paper we extend our analysis and description of the motivating factors that influence the design (*i.e.*, the parameter values) of the IdM implementations. We believe an understanding of these factors will be critical to making our model useful as a support for decision making.

2. Background: Our virtual organization identity management model

In this section we provide a brief description of our VO IdM Model as presented in [2]. As we described in the Introduction, the emergence of the VO raises the challenging question of how much of the IdM process should be delegated from the RP to the VO. The answer to this question is not all or nothing: An RP's level of trust in a VO's IdM has a range from almost none (*e.g.*, where a VO simply directs users to some identity vetting process operated by an RP), to one where an RP has no ability to identify the individual users of a VO (*e.g.*, the VO enrolls members and handles all identification and direct interactions with those users). We conducted semi-structured interviews with members of the community that represented 18 different VO-RPs relationships and, from analysis of the results of those interviews, we created a Model that allows for expression of their different IdM choices.

The Model is based on the concept of a VO user lifecycle, a concept borrowed from the realm of identity management (*e.g.*, [4]) which expresses a set of discrete steps in the lifecycle from users' enrollment into a VO, through their use of RP resources and eventual departure from the VO. In our Model, each stage represents a possible point where the RP can become involved with the VO's individual users instead of treating the VO as a single entity. The stages of the VO user lifecycle we observe are:

- **Enrollment:** An initial, (typically) one-time process by which the user is admitted into the VO.
- **Provisioning:** Following an enrollment, the one-time creation of any state associated with the user across the VO or RPs.
- **Request:** The process by which the VO makes a request for resources from an RP to provide service to its users. This can be in direct response to a request from a user or can be an *a priori* reservation (*e.g.*, a pilot job).
- **Usage:** Consumption of a RP's resource by a VO to provide service to a user. This can directly follow a request or may occur sometime later.
- **Incident Response:** An exception event that requires response, and typically requires manual interaction with the user to resolve. This includes computer security incident response, a misbehaving user process, or a user support process.
- **Deprovisioning:** Removal of users from the VO, cessation of their right to consume resources on behalf of the VO, and possibly removal of any state or data owned by the user. (In practice, few IdM systems implement deprovisioning and hence we don't have much data with regard to it.)

While varied degrees of RP-to-VO IdM delegation are possible across each stage, and allow for great variety in the exact implementation scheme, a leading indicator of the level of delegation is *at which of the lifecycle stages, if ever, the RP becomes aware of the identity of the user beyond the fact the user is a member of the VO*. In other words, when, if ever, does the VO pass the user's identity to the RP? Based on our observations, any of the above stages (except deprovisioning) are possibilities; or the RP may never learn the user's identity, only knowing them as an anonymous member of the VO. Quantification of this value provides a useful first-order description of the VO-RP's IdM relationship by expressing the degree of delegation of IdM from the RP to the VO. Finer-grained expressions of the delegation details are discussed in [2] and are beyond the scope of this paper.

3. Motivating factors for identity management implementations

We now turn to the new work presented in this paper, which is the result of further analysis of the interviews from [2] and additional subsequent interviews, to understand and articulate the motivations that might steer a VO and RP toward a particular IdM implementation as described by our Model.

The analysis followed a similar trial-and-error approach we used to determine the original model, that is, we greatly leveraged our combined experience in identity management and cybersecurity, developed theoretical models, tested them against our data and refined until we had, in our judgment, a strong fit.

These motivating factors are described subsequently. The factors come under a range of topics from technical to sociological to effort/economics. As we describe subsequently in the Future Work section, the factors and their impact are based on the authors' observations and interview results, and warrant further validation through application with real-world VOs.

The factors, in no particular order, and their impact on IdM decisions follow.

3.1. Isolation Among Users

In a shared environment such as a common supercomputer system, there exists risk that a single user, either maliciously or accidentally, will disrupt the work of others. Web applications, grid computing, cloud computing and virtual machines typically increase the isolation between users by limiting how much they can interact. Different VOs require a range of user interaction or isolation at their RPs. At one end of the spectrum, VO users want to collaborate directly with one another (*e.g.*, in Wikis). At the other end, complete separation is desired. And, there are points in between, such as sharing read access to data.

In order for an RP to provide isolation, it typically requires identity information to distinguish and separate users at the time of servicing a request (if not before). Thus, RPs institute policies for more strongly distinguishing users - *e.g.*, stronger vetting, disallowing shared accounts, and, in many cases, requiring multi-factor authentication. Web applications and then grid computing increase the isolation since access is controlled by web applications or is limited to running batch jobs.

3.2. Persistence of User Data

Some services do not require maintaining data specific to individual users at the RP – *e.g.*, input data is staged in, is common to the VO, or is accessed from a remote location and the results of the computation are sent to a remote storage location. Another example is per-user configuration or personalization. Typically, along with such per-user data, there comes some form of access control requirement, perhaps just to read the data, seemingly always to modify it. Hence, the presence of per-user data at the RP leads to the RP needing identity information. Without persistent per-user data, there is reduced need for the RP to implement access controls other than ensuring running computations are protected from those of other VOs.

This motivating factor shares some similarities with the previously discussed Isolation factor in that they both involve the need for implementing forms of access control at the RP. However isolation between requests is ephemeral, with identity information needed only while the request is handled, and identity information regarding access control to data (*e.g.*, who owns it) is needed for the typically longer lifetime of the data. Hence we conservatively keep these as separate factors for the time being.

3.3. Complexity of VO Roles

VOs may require support of a range of user roles and levels of privilege, with users allowed only to make basic requests through those with privileged administrative roles. Where an RP provides a number of services for a VO, the number of roles for different users can make the access control rules quite complex, meaning there can be a significant range of expectations for the access controls the RP is expected to provide. The expectations can be from "none," in a situation where the VO is isolated from the RP, to "high" where the VO's users have shell accounts and persistent data storage at the RP.

This increasing complexity results in an increasing amount of state to be shared and kept coherent between the VO and the RPs serving it, which can become a challenge. In these situations, in order to reduce the need to share this information, we observe a tendency for more IdM information to be kept at the VO, which takes responsibility for access control (*e.g.*, by web applications brokering access to the RPs).

3.4. Cultural and Historical Inertia

As described in the Introduction, prior to the emergence of VOs, RPs were in complete control of the IdM for their user communities. In our experience, RPs are, in some cases, generally reluctant to delegate IdM to VOs. While this seems to be diminishing over time as the community gains experience and acceptance of VOs, older relationships seem to have more characteristics of the classic model where the RP controls all IdM. These early IdM implementations tended to be more heavily controlled by the RP mapping to an early communication of IdM information in our Model.

We also note that there are economic incentives not to change – in that implementing such change requires effort for which the benefits of changing may not obviously provide sufficient compensation.

3.5. Scaling

There are a number of ways scale comes into play with VOs: The number of users in the VO, the number of different RPs serving the VO, and the number of different services provided by the RPs to the VO are the three primary ways we identified. Similar to the previously described factor of Role Complexity, as the numbers of users, RPs or services increase, this poses an increasing challenge in keeping the IdM information coherent between the VO and RPs. In these cases, we believe there is motivation for an increased delegation of IdM from the RP to the VO for the sake of efficient enrollment of VO users. Meaning, in terms of our Model, IdM information is conveyed later in the lifecycle.

3.6. Incentive for Collaboration

The relative balance of incentives between the VO and RP to make the relationship work seems to effect IdM decisions. For example, some relationships involve an RP willing to share spare computing cycles with a VO, but are not otherwise deeply committed to the VO. On the other end of the spectrum, the VO may be politically important and have a great deal of influence in setting the rules. Most commonly the relationship is very important to both the VO and the RP (*e.g.*, they both receive significant funding from the same agency and are expected to collaborate).

This factor is one not well-understood at this time, but two clear patterns seem to be emerging. We observe that when incentives are balanced, implementations seem to lean toward increased efficiency for both parties, and when the RP is less incentivized (*e.g.*, it contributes free cycles to a VO not critical to its mission), then RP's ease of operations or risk reduction dominate negotiation.

4. Related Work

Prior work by Landau and Moore [6], Brooder, et al [7], and Altunay [8] informed and shaped our interview process. Work by Lin, Vullings, and Dalziel [9] explored factors related to trust in making access control decisions in the context of VOs, but was focused, and admittedly directly applicable, to access control decisions rather than the whole identity management system.

5. Conclusions and Future Work

In this paper we provided an analysis, based on the results of a set of 18 interviews conducted with representatives of virtual organizations and resource providers, of a set of six factors motivating identity management implementation decisions made in the context of virtual organizations. This work builds on our previous description of a VO IdM Model [2] and represents a significant step towards providing guidance to VOs and RPs in making a key decision in their IdM implementation.

As previously described, a goal of the authors is to provide not only a academic model describing different IdM implementations, but usable guidance to VOs and RPs in selecting a IdM implementation that best meets their needs. Hence, our future work includes further validating and refining the Model and the motivating factors by working with VOs to apply them. Finally, the RPs that have been the subject of our interviews to date could be generally described as typical scientific computing centers; we will expand our studies by interviewing RPs more representative of cloud computing and high-performance ("exascale") computing.

6. Acknowledgements.

We thank all those who have taken part in our interviews, who, in addition to those listed in [2], include Shreyas Cholia of NERSC, Romain Wartel of CERN, and Sandy Philpott and Andy Kowalski of JLAB. We also thank the Open Science Grid, whose meetings have provided a good opportunity for interviews and for an ongoing collaboration providing useful feedback.

We thank the US Department of Energy Next-Generation Networks for Science (NGNS) program (Grant No. DE-FG02-12ER26111) for funding this effort. All opinions are those of the authors and do not necessarily reflect those of the Department of Energy or any other organization or person (including the interviewees).

We also thank the anonymous CHEP 2013 reviewers for their feedback on this paper.

References

[1] Ian Foster, Carl Kesselman, and Steven Tuecke. “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”. . *International Journal of Supercomputer Applications, May 10, 2001*. Available: http://toolkit.globus.org/alliance/publications/papers/anatomy.pdf

[2] Robert Cowles, Craig Jackson and Von Welch. “Identity Management for Virtual Organizations: A Survey of Implementations and Model.” 9th IEEE International Conference on eScience, 2013. http://cacr.iu.edu/collab-idm

[3] C. Robson, Real World Research: A Resource for Users of Social Research Methods in Applied Settings, 3rd ed. Chichester, West Sussex, United Kingdom: Wiley, 2011, p. 280.

[4] John K. Walters. “The ABCs of Identity Management.” CSO Online. Accessed January 7, 2014. http://www.csoonline.com/article/205053/the-abcs-of-identity-management

[5] Hacker, T., Athey, B., "A Methodology for Account Management in Grid Computing Environments," Proceedings of the 2nd International Workshop on Grid Computing, November, 2001, Denver Colorado, Lecture Notes in Computer Science, Spinger Verlag Press.

[6] S. Landau and T. Moore, “Economic tussles in federated identity management,” *First Monday*, vol. 17, no. 10, Oct. 2012. Available: http://journals.uic.edu/ojs/index.php/fm/article/view/4254/3340

[7] D. Broeder, et al., “Federated identity management for research collaborations,” CERN-OPEN-2012-006, Apr 23, 2012. Available: http://cds.cern.ch/record/1442597?ln=en

[8] M. Altunay, “How OSG Resource Providers Consume Identity Information”, unpublished presentation to the MAGIC committee, Dec. 4, 2012.

[9] A. Lin, E. Vullings, and J. Dalziel, “A trust-based access control model for virtual organizations,” in *IEEE Proc. Fifth Int. Conf. Grid and Cooperative Computing Workshops (GCCW’06).*