PROCEEDINGS OF SCIENCE

Modeling the Past and Future of Identity Management for Scientific Collaborations

Robert Cowles[1]
Indiana University
USA
E-mail: bob.cowles@gmail.com

Craig Jackson
Indiana University
USA
E-mail: scjackso@indiana.edu

Von Welch
Indiana University
USA
E-mail: vwelch@iu.edu

Over its three year funding period, the eXtreme Science Identity Management (XSIM) research project collected and analyzed real world data on identity management (IdM) implementations in virtual organizations (VOs) representing the last 15+ years of collaborative DOE science. XSIM conducted over 20 semi-structured interviews of representatives from scientific collaborations and resource providers, both in the US and Europe; the interviewees supported diverse set of scientific collaborations and disciplines. We constructed a descriptive IdM model sufficiently complex to produce accurate, useful descriptions of the observed trust relationships and technical implementations, but still simple enough to explain and use in novel situations. It was important that the model be comprehensible to both scientists and IT/Cyber security experts to support a dialog between stakeholder groups with different lexicons.

In this paper, we summarize the VO IdM model and discuss the experiences and lessons learned of the XSIM project, both in the process of conducting socio-technical research in this interdisciplinary space and in utilizing the model to provide guidance to specific communities. Finally, we describe areas of needed or potentially fruitful research that would enhance the adoption of advanced IdM technologies in future scientific collaborations.

International Symposium on Grids and Clouds 2016
13-18 March 2016
Academia Sinica, Taipei, Taiwan

[1]Speaker

PoS(ISGC 2016)014

1. Introduction

Identity management (IdM) is broadly defined as creating and maintaining identifiers and attributes (digital identities) and conveying them to relying entities in a manner such that there is some level of assurance about whom (or what) is communicating and/or being provided access. Relying entities use IdM to make informed, confident decisions regarding a number of activities, e. g., how to schedule and service requests, log activity, and deal with security incidents.

The eXtreme Scale Identity Management for Scientific Collaborations (XSIM) project had an initial objective to develop an evidence-based, descriptive IdM model that could describe a variety of IdM implementations, and provide insight and heuristics into the factors favoring one implementation over another. XSIM worked towards its goal of providing practical advice on designing and optimizing IdM implementations fit for individual needs of virtual organizations (VOs) and relying parties (RPs).

The XSIM project worked in the context of scientific collaborations where the scientists are distributed among universities, US DOE National Laboratories, and research institutes around the world. The resulting IdM model had to be sufficiently flexible to address how IdM for scientific collaborations could interoperate with existing national (e.g., US Federal PKI) and international (e.g., Interoperable Global Trust Federation) IdM standards.

2. Methodology and Project Timeline

2.1 Research Methods

XSIM conducted interviews with the goal to obtain both subjective and objective information regarding identity management implementations across a broad range of VOs and RPs on which to form our VO IdM Model. A semi-structured interview process was developed [eSci]. The interview results were supplemented with published papers, presentations, and articles about the interviewed VO or RP.

Following the interviews, an iterative process was used to form a descriptive model that best fit the data. The goal was to find a model that allowed for both the easy and clear expression of data from all interviews, and could be leveraged to provide guidance in designing a VO IdM implementation. Based on experience, an initial set of parameters and possible values was selected. Those parameters were compared to the interview data, and then iteratively refined to improve the quality of the matches.

XSIM's initial publication (see Timeline below for references), presented at the 9th IEEE International Conference on eScience in 2013, established a simple VO identity model that expressed the VO-RP relationship in terms of the amount of delegation of responsibility for IdM from the RP to the VO. Subsequent work, presented at the 20th International Conference on Computing in High Energy and Nuclear Physics (CHEP2013), explored the motivations that VOs and RPs have for these delegations. It identified the following factors: the need to provide isolation among users; persistence of user data at the RP; complexity of VO roles; cultural and

historical inertia; scaling in terms of the size of the VO and number of RPs; and the RP's incentive to support the VO.

The paper presented for the International Symposium on Grids and Clouds (ISGC 2014), described additional interviews, refinements to the VO IdM model, influential factors in applying a transitive trust approach, and concluded with a NERSC use case illustrating and applying the refined model. For the 2015 Workshop on Changing Landscapes in HPC Security (CLHS'15), XSIM presented how, particularly for US DOE Labs, existing policies allow the delegation of IdM functions to collaboratories within the context of acceptable risk management. The paper suggested strategies that allow for the incremental increase of trust and delegation of IdM functionality.

In all, a total of four iterations of model development were required. The quality was determined subjectively by the authors based on a combined 60+ years of experience in distributed computational science, cybersecurity and identity management.

2.2 Timeline

2.2.1 Socializing the approach to obtain interviews (2013)

The XSIM team made presentations at various meetings and conferences to explain the goals of XSIM and the methodology being used, to obtain feedback on the approach, and to foster adoption of the model. At the same time, the team used the opportunity to approach knowledgeable people and request interviews as part of our data collection. Meetings included: the Open Science Grid (OSG) All-Hands meeting (VO and resource providers in the US); HEPiX (representatives of resource providers in the US and Europe); EUGridPMA (identity and resource providers in Europe); Vo Architecture and Middleware Planning (VAMP) (virtual organizations in Europe); NGNS-PI (NSF-funded researchers).

2.2.2 Presentations of initial results (2013)

In October 2013, XSIM presented initial interview results and had refereed papers accepted for publication by eScience 2013 [eSci] and CHEP 2013 [CHEP].

2.2.3 Additional interviews and test early model (2014)

The latter part of 2013 and early 2014 saw the development of an initial model and additional interviews. A number of presentations were made to obtain feedback on the results obtained to date. Meetings included: HEPiX (representatives of resource providers in the US and Europe); EUGridPMA (identity and resource providers in Europe); TAGPMA (identity and resource providers in North and South America); LBNL and NERSC; OWASP (application security community); National Labs Information Technology exchange (NLIT) 2014 (technology experts from the DOE national Labs); NGNS-PI (NSF-funded researchers). There were additional discussions at other conferences and meetings such as XSEDE, SC14, Federated Identity Management for Research (FIM4R), Security for Collaborating Infrastructures (SCI), and Terena Networking Conference (TNC14).

2.2.4 Fully developed model (2014-2015)

Through several more iterations of the model, updated presentations were made at OSG, EUGridPMA, CERN, PNNL, The Networking Conference (TNC15), and MAGIC. A whitepaper was developed to address the issues often confronted at DOE Labs [FSC1] when adopting a transitive trust model. The XSIM team engaged in a high-level discussion of those issues at a meeting of the National Labs CIOs (NLCIO). Presentations and refereed papers were developed for the International Symposium on Grids and Clouds (ISGC) 2014 [ISGC], Changing Landscape in HPC Security (CLHS) 2015 [CLHS], and finally ISGC 2016 (this paper).

3. Project Accomplishments

3.1 Model development

As mentiond above, the results of the interviews and the fully developed model were documented in papers published in 2013 and 2014 [eSci], [CHEP], [ISGC]. In general, we used a modified form of the Data Flow Diagrams [DFD1] to describe the producers and consumers of the three types of identity information: (1) user identifier; (2) user contact information,;and (3) VO membership/role. These data elements are used for a variety of functions such as: authentication, authorization, scheduling, accounting, auditing, user support, and incident response. We found a general movement to delegate IdM functions from the RP to the VO when there were sufficient enablers and motivators for the delegation to overcome the barriers [ISGC].

3.2 Actionable Guidance for the DOE Community

With some confidence in the model, white papers were developed aimed at giving advice on identity management to (1) virtual organizations joining OSG [OSG1]; (2) the Dark Energy Science Collaboration (DESC) (associated with LSST) [DESC]; and (3) DOE Labs [FSC1], the latter with particular emphasis on removing perceived barriers to delegating IdM functions -- and promoted that white paper by making presentation to a meeting of the National Labs CIO organization (NLCIO).

4. Project Lessons Learned

4.1 Successful Innovations

4.1.1 Composition of the team

The project team (Cowles, Jackson, Welch) had varied backgrounds that both overlapped and complemented each other in a manner that contributed greatly to its success. Areas of expertise included Open Science Grid and supercomputer facilities; DOE cyber security environment; project management; law; philosophy; social science research methods; LHC Grid Security; academic paper writing, and Global PKI requirements for science. The team had strong connections to resources in both the US and Europe; to educational institutions, NSF funded facilities and DOE Labs; these connections were valuable in having contacts to solicit for interviews.

4.1.2 Emphasis on knowledge rather than software

With a significant amount of technical development in progress around IdM and years of applied experimentation, producing one additional software product did not seem useful. Rather

than produce a final product consisting of software that was likely to have little impact once the resources ran out to further develop or maintain it, it was decided to produce an evidence-based "knowledge product" that could be used to alter the way a broad spectrum of developers think about IdM. The knowledge products were stored in Indiana University ScholarWorks systems, "a digital repository service provided by the IU Libraries for showcasing and preserving IU research." [IUSW]

4.1.3 Comprehensive and comprehensible model

XSIM's early descriptive models either were inadequate to describe the observed variations in the research data or were subjectively too complicated (based on the authors' opinion). The search was for a model that was comprehensive, but still simple enough to explain and use in novel situations. It was important that the model be comprehensible to both researchers and IT/Cyber security experts to support a dialog between stakeholder groups with different lexicons. Success in this regard was defined initially by the authors' subjective opinion in applying it to the research data, and then through feedback obtained in the socialization of the model through presentations. In retrospect, a more formal survey-based approach to evaluation could have been utilized.

4.1.4 Evidence-based research

Rather than developing an idealized model only looking at the future, the XSIM team mapped the course of IdM as it has developed in some of the major science collaborations, taking into account the direction they were moving in the last fifteen years. Using that data, a framework was produced that fit past, present and the projected future of identity management. Grounding the model in real world historical and present data from interviews allowed the authors to have increased confidence in the results and it is believed to have generated more interest in the model at presentations since it was seen as more practical.

4.2 Challenges

4.2.1 Collaboration engagement at the right time

A goal of the XSIM project was to work with a new VO to utilize and validate the model as the VO worked to establish its IdM system. The project was not successful in achieving this goal as it was very difficult to engage with a collaboration at the right point in its lifecycle to create a measurable impact on that project. Too early and a project is typically still awaiting to hear about funding and not yet engaged in technical design. Once design is underway to the point design decisions around IdM have been made, it is very difficult to motivate the revisiting of those decisions.

An engagement with DESC at SLAC was attempted, but the timing was both too early and too late for a profitable engagement: too late in the sense that they already had tentative plans for how their IdM was likely to work using designs and technical staff from previous scientific experiments; too early in the sense that various construction and funding delays meant that the first real data collection for the project was not scheduled for 6-7 years so there was not a feeling of urgency to design the IdM during the timeframe of the XSIM project.

In retrospect, XSIM should have identified a scientific collaboration very early on that would have been entering the window of IdM design toward the end of the XSIM project. While it is difficult, in the authors' experience, to entice projects to collaborate on products that are still to be developed, this would have allowed work in parallel to build the relationship as XSIM built its model and guidance.

4.2.2 Reaching the target audience

The target audience for XSIM's work is technical leadership in scientific collaborations. It was a challenge to find natural meetings where such people congregated (as opposed to their counterparts in the National Laboratories, for whom numerous meetings could be identified). As such, XSIM had to engage with them individually, which required significant time and travel. A regular gathering of technical leadership in scientific collaborations to exchange experiences and hear from projects in programs such as NGNS may benefit that community.

5. Next Steps

5.1 Exploring the promise of and systemic barriers to transitive trust

The interviews revealed a clear historical trend to greater delegation of IdM to VO's by the RPs, and transitive or near-transitive trust implementations are taking hold in the community. Feedback from presentations of the model centered around two issues. First: Use of a transitive trust model where the collaborators provided contact information to the VO and the VO was then responsible for any subsequent contact, including incident response, structurally reduced or eliminated the need for the release of user attributes by identity providers and the logging of personally identifiable information by resource providers, reducing privacy and data protection concerns those provided may have had. Second: Attendees expressed significant concern that only a relatively small number of VOs were actually competent to manage user registration and to follow-up on incidents. In their experience, for the case of the "long tail of science" a large number of scientific collaborations do not have the expertise to perform these functions.

5.2 Integration into collaboration supporting infrastructure

There are a number of efforts in the US, Europe, and Australia to provide virtual environments tailored to particular scientific disciplines and these environments provide almost everything in terms of IT infrastructure (including IdM) a collaboration might need to perform its work. Integration of the XSIM IdM model would allow these virtual environments to operate coherently with DOE Laboratories and other organizations. There is currently no obvious effort in the U.S., as there is in Europe by EGI and Géant, to deploy and operate such an infrastructure, meaning an entity taking this on could provide leadership in this integration.

5.3 Developing a taxonomy of scientific data and their security requirements

Risk and trust are tightly intertwined. Nowhere is this more apparent than in the contemporary fields of information security, privacy, and identity management. There was uncertainty in the level of risk involved in delegation of IdM for access to scientific collaboration, in part due to the uncertainty around the sensitivity of the data involved. This lack of clarity

contributes to conservative decisions around delegation, which in turn may unnecessarily inhibit scientific collaboration and discovery.

Data involved in the science and operation of scientific collaboratories can have varying requirements for confidentiality, integrity, and availability. These requirements come from various sources. Systematic research is needed to develop a comprehensive, comprehensible framework of these requirements that will ease the burden for risk and security decision makers involved in IdM delegation (and other facets of setting up a collaboratory). The XSIM team is not aware of current work in this area. XSIM's structured interview, analysis, and socialization methods would be well-suited to produce an evidence-based, highly usable, and high impact framework.

References

[CHEP] R. Cowles, C. Jackson, V. Welch. *Identity management factors for HEP virtual organizations.* 20th International Conference on Computing in High Energy and Nuclear Physics (CHEP2013), 2013. http://www.vonwelch.com/pubs/CHEP2013.

[CLHS] R. Cowles, C. Jackson, and V. Welch. *Facilitating Scientific Collaborations by Delegating Identity Management: Reducing Barriers & Roadmap for Incremental Implementation*, CLHS '15 Proceedings of the 2015 Workshop on Changing Landscapes in HPC Security, 2015. https://dl.acm.org/citation.cfm?id=2752501

[DESC] B. Cowles, C. Jackson, and V. Welch (PI). *DeSC Identity Management: Analysis and Recommendations*. Unpublished Technical Report, August, 2014.

[DFD1] P. D. Bruza, and Th. P. Van der Weide, *"The Semantics of Data Flow Diagrams"*, University of Nijmegen, 1993.

[eSci] R. Cowles, C. Jackson, V. Welch. *Identity Management for Virtual Organizations: A Survey of Implementations and Model*, 9th IEEE International Conference on eScience, 2013. http://www.computer.org/csdl/proceedings/escience/2013/5083/00/5083a278-abs.html

[FSC1] R. Cowles, C. Jackson, and V. Welch. Facilitating Scientific Collaborations by Delegating Identity Management, CACR/XSIM Technical Report, March, 2015. http://cacr.iu.edu/sites/cacr.iu.edu/files/FSCbyDIM0408.pdf

[ISGC] R. Cowles, C. Jackson, V. Welch, and S. Cholia. A Model for Identity Management in Future Scientific Collaboratories, International Symposium on Grids and Clouds (ISGC) 2014. PoS(ISGC2014)026

[IUSW] IUScholarWorks. https://scholarworks.iu.edu/ Visited March 30, 2016

[OSG1] V. Welch, R. Cowles, and C. Jackson. *XSIM OSG IdM Guidance OSG-doc-1199*, July 2014. http://osg-docdb.opensciencegrid.org/cgi-bin/ShowDocument?docid=1199