Per-Entity Metadata Working Group Charter

Problem Statement

In its 10+ years, the InCommon federation has grown from serving a very limited number of applications with an equally small number of participants (mostly large research universities), to a federation that now supports thousands of different applications across approximately 650 active participants. Add to this a rapid growth in the size of InCommon metadata, due to InCommon's production support for the eduGAIN interfederation service, and the federation is at risk of becoming a victim of its own success.

Metadata aggregates, that is, metadata made up of more than one SAML entity descriptor element, are static lists of entity descriptors that are aggregated, validated, signed and distributed to consumers of federation metadata. This model is analogous to how hostname resolution was done before DNS existed, using 'hosts' files, and it has reached the end of its sustainability the way that solution did long ago. Aggregates are inherently brittle - an error in a single entity descriptor can cause issues loading an entire aggregate.

Additionally, very large metadata aggregates, as InCommon now distributes on a daily basis, have a number of other major drawbacks:

1. Increased bandwidth use to distribute a large file that consumers almost certainly don't need in its entirety
2. Inefficient use of client bandwidth to download a large aggregate on a regular basis
3. Increased time to canonicalize (XML document normalization) a large XML document so that the signature on it may be verified - thus increased time to start up a SAML deployment consuming a large aggregate
4. Increased memory needed to canonicalize a large XML document - now on the order of gigabytes, and this will only increase over time. A waste of deployer resources.
5. Intentional or unintentional denial-of-service for consumers of an entire aggregate based on malformed entity descriptors imported from other federations.

To address these and other concerns with the aggregate, InCommon's previous Metadata Distribution Working Group[1] recommended a test deployment of the Metadata Query Protocol (MDQ)[2]. For over two years, InCommon has been running an MDQ testbed to gain experience with the technology and this new model. This new working group is charged with items necessary to allow InCommon Ops to move this technology into a production-ready service.

Stakeholders/Influencers/Influences

Different audiences can impact different aspects of this problem:

1. SAML deployers - IdP, SP, AA, Discovery Services, etc. of various implementations: Shibboleth, SimpleSAMLphp, ADFS, Ping, etc.
2. SAML implementers - Latest versions of both Shibboleth and SimpleSAMLphp support the MDQ protocol, but implementation issues may exist that have not been found due to the need for operational exercising of these features. Other implementations such as ADFS may be enabled to participate in the federation in ways they have not been able to previously.
3. InCommon Operations and Internet2 Technical Services Group (TSG) - Running a highly reliable service that responds to requests for entity descriptors in real-time is a service delivery model that is new for InCommon and will require additional resources to support.
4. Participants - what needs do they have for local per-entity metadata installations to allow for local generation and consumption of per-entity site-specific metadata? Do they have a need for a local copy of a cache of per-entity metadata for redundancy reasons? Etc….
5. International community - how will an InCommon per-entity metadata service align with plans that other federation operators may have?

Charter

The Per-Entity Metadata Working Group will:

1. Work based on the premise that InCommon will be moving toward per-entity MDQ[2] protocol-based distribution of metadata.
2. Develop a roadmap for addressing the immediate needs for reduced aggregate size, as well as intermediate milestones along a trajectory to a sustainable future state, to be determined. The first items on this roadmap should include building a production service which allows production SAML deployments to exercise their per-entity metadata capabilities, and include checkpoints to improve the service and software when issues are encountered. If short-term steps such as InCommon producing separate IdP and SP feeds are deemed necessary, these items should be included in this roadmap. This roadmap should also address the issue of continued creation (or eventual decommissioning) of multi-entity aggregates.
3. Address issues and questions that have arisen about the process of moving from where we are to relying on this new model, including but not limited to:
    a. High availability
    b. Performance
    c. Site redundancy
4. Develop requirements, risks, and recommended risk mitigation strategies for a production per-entity metadata service delivered by InCommon, including a firm definition of the scope of the service, aligned with the immediate needs addressed in the roadmap from (1).
5. Advise InCommon staff on implementation of a solution, based on the requirements of the service documented in (4).

6. Compile the outcomes of these investigations into a report to the TAC

Explicitly out-of scope is:

1. A ‘full DNS’ model which would require changes to the MDQ protocol or current software implementations of the protocol.
2. Other items that need to be resolved by the international community. These items should be addressed via appropriate forums such as REFEDS, or better yet, the IETF.
3. Determining a concrete roadmap for ceasing production of multi-entity aggregates. This item is important, but must be the work of a later group, after we build experience using a production-quality per-entity metadata service.
4. A solution to the IdP discovery problem in light of per-entity metadata. Discussion or debate of options is reasonable, as long as the WG's main deliverables are not sidetracked.
5. Choice of a per-entity metadata server application or specific testing of software related to a specific choice of server application. That is the realm of operationalizing a production service and is up to InCommon staff.
6. Support for querying entities by anything other than entityID (already put out-of-scope by previous work: https://spaces.at.internet2.edu/x/BoGDAg)

Membership

Membership in the Working Group is open to all interested parties. Solicitation will take place on lists such as the InCommon Participants list and the REFEDS list, explicitly seeking international participation. Members join the Working Group by subscribing to the mailing list, participating on the phone calls, and otherwise actively engaging in the work of the group.

Work Products

1. September, 2016 - Draft Report to the TAC, Report out at TechExchange
2. November, 2016 - Final Report to the TAC

Related Resources

1. Metadata Distribution Working Group recommendation on pilot study of MDQ
2. MDQ protocol draft
3. Draft call for participation in MDQ testbed (restricted access)