Agenda and Notes - 2016-08-31

Per-Entity Metadata Working Group - 2016-08-31
Agenda and Notes

[Etherpad used to create these notes: Agenda_and_Notes_-_2016-08-31.etherpad]

Dial in from a Phone:
Dial one of the following numbers:
+1.408.740.7256
+1.888.240.2560
+1.408.317.9253
195646158 #
Meeting URL (for VOIP and video): https://bluejeans.com/195646158
Wiki space: https://spaces.at.internet2.edu/x/T4PmBQ

Attendees

- David Walker, InCommon/Internet2
- Scott Koranda, LIGO
- Nick Roy, InCommon/Internet2
- Ian Young
- Regrets for not attending: Chris Phillips /CANARIE
- Regrets for not attending: Rhys Smith, Jisc
- John Kazmerzak, University of Iowa
- Paul Engle (Rice U)
- Tom Scavo, InCommon/Internet2
- IJ Kim, Internet2
- Tom Mitchell, GENI
- Scott Cantor, tOSU
- Paul Caskey, Internet2
- Steve Carmody, Brown
- Ann West, Internet2/InCommon
- Tommy Doan, Southern Methodist University

Agenda and Notes

1. (Discussion of collaboration for the final report before official start of call)
    a. David and Scott will talk about moving final report to Google Docs.
2. NOTE WELL: All Internet2 Activities are governed by the Internet2 Intellectual Property Framework. - http://www.internet2.edu/policies/intellectual-property-framework/
3. NOTE WELL: The call is being recorded.
4. Agenda bash
5. Distributing an IdP-only aggregate
    a. Ops perspective: https://spaces.at.internet2.edu/x/UgAZBg (Tom Scavo)
        i. Nick: No service is permanent; there will always be change. Being conservative about change is warranted, but need to balance it with pragmatism
    b. A single IdP-only aggregate or a pipeline triplet (preview, main, fallback)?
        i. Not clear if there is an Ops recommendation?
        ii. It's nice having the triplet, but there's a cost for each aggregate. Also, there's a potential of confusion due to a large number of aggregates.
        iii. Consensus from last week was that we don't need the triplet. It's still our consensus.
    c. Ops claim: "we only get one chance to migrate deployers to a new metadata configuration." Thoughts?
        i. Scott K disagrees
    d. Getting started on this before final report from working group?
        i. Scott and SteveC will get the issue onto tomorrow's TAC agenda so Ops can start quickly.
6. Commercial CDN latencies
    a. Amazon CloudFront last mile testing: https://media.amazonwebservices.com/FS_WP_AWS_CDN_CloudFront.pdf
    b. Interesting benchmarking exercise: http://goldfirestudios.com/blog/142/Benchmarking-Top-CDN-Providers
        i. It seems we're looking at ~.25 second response times.
        ii. CDNs still seem like the right approach, but we need to have our eyes open.
        iii. A CDN that's connected to the Internet2 backbone is a good idea, although it's not clear how any InCommon participants are Internet2 connected.
    c. Per-entity metadata file size for InCommon (great data!)
        i. Largest (without signature) is 148K (due to embedded logo)
        ii. Smallest (without signature) is 3K
        iii. Median is 5.3K
        iv. Average is 6.3K
        v. Std deviation is 4.7K
        vi. Current overhead of signature is roughly 2.8K
        vii. So most per-entity payloads will be roughly 8.1K
    d. What contribution to the actual user experience does the CDN latency make in a MDQ scenario?
        i. How does it compare to the rest of the SAML flow?
        ii. How does it compare to the rest of the work the IdP or SP must do?
    e. What benchmarking should be part of the roadmap?
    f. What is the requirement for ongoing monitoring?

7. CDN features and MDQ
    a. Push mechanism (scp, sftp, rsync, ...)
    b. Origin pull (instead of push)
    c. Purge (invalidation)
    d. Purge All
    e. HTTPS (custom SSL capability, ie. InCommon can provide X.509 cert)
    f. Access logs
8. SAMLbits CDN (Leif)
    a. Community-driven CDN specifically built for high-trust applications
    b. Can be customized for caching in the CDN flow
    c. Can translate headers, for example, SAML-HTTP
    d. Essentially a varnish cache - wanted to prove that it would be possible to build a community-driven CDN that didn't have to consume a lot of resources at the site
    e. A couple boxes online with I2 operations (TSG)
    f. Doesn't require the network-aware interconnect that a lot of the commercial CDN appliances require
    g. Been running for a couple years and seems to work well
    h. How do we address governance issues as SAMLbits becomes part of our solution?
        i. This a good excuse to start a discussion. We could think of this as governed by the community of InCommon participants or international federation operators.
        ii. If it is to become a part of the solution, it probably needs to go to Steering for a request of some sort +1
    i. It's not overly difficult to deploy a local SAMLbits node, for example at a campus.
9. Solution architecture description for the final report
    a. https://spaces.at.internet2.edu/x/u4EQBg
10. Next call on September 7 at the usual time and place.