FedCM and the R&E Federation Community

~~Registration link: https://events.geant.org/event/1379/~~ [CLOSED - at capacity]
Location: https://www.hotelzico.com/
Dates: February 28 - March 1, 2023 (all day, both days)

Food notes: Muffins from Sweet Diplomacy will be available for morning snackage; hotel will provide lunch. Dinner is on your own (financially speaking) but we'll try to find a place that can accommodate the group.

Info sharing and hackathon

The goal of this meeting is to help get us all on the same page when it comes to what federation looks like in research and education (R&E). We need to show commitment to forward progress and how to make our various code bases better. While the immediate concern in the federation space is the loss of third-party cookies, we're also taking the long view of how we'll want to have FedCM integrated early so R&E is in a position to handle future privacy-enhancing changes to the browser space.

Outcomes

We want to have a positive impact on privacy, coming up with new proposals to enhance FedCM. Draft in a proposal template (noting proposals we like will end up in GitHub)

Day One - February 28

09:00 - 09:30 - Introductions (people) (Nicole do this since Heather will be acquiring muffins)

09:30 - 10:30 - Federations

- Different types of federation - https://wiki.geant.org/display/eduGAIN/Federation+Architectures

    - Mesh federation (InCommon) - similar to a certificate authority in model (so much like DNS registration as well)
    - Hub-and-spoke federations (SURF in The Netherlands)
- Power of Proxies
- eduGAIN and interfederation

10:30 - 10:45 - Coffee Break

10:45 - 12:30 - Demos + problem statement

- Shibboleth - IdP/SP
- SimpleSAMLphp - IdP/SP
- SeamlessAccess - account choosing

12:30 - 14:00 - Lunch

- Can FedCM deconstruct the discovery relationship such that the browser can accept the trust in the relationship between the IdP and the RP; the browser would look at the federation metadata?
- Want a meaningful informed relationship that allows the user to drive the relationship creation. The two institutions should know each other enough to be able to go around the browser if necessary (data transfer of metadata).
- When the browser wants to mediate the protocol flow, that's going to be a decade-long effort (and by then we'd all have moved to wallets).
- Just knowing the origin doesn't cut it.
- What the browser needs is generally not the PII components; it needs a subset of the SAML metadata. But is this about identification via fingerprint? We want the smallest possible threat footprint
- The solution to this problem has to not trust the four parties involved (RP, IdP, FedOp, or Browser) - all four parties have part of the trust picture, but should not allow them to collude until the user is informed about the nature of the collusion that has to happen in order for the user to get to what they are trying to get to
- Protocol agnostic/future proof
- exact problem we're trying to solve from a privacy perspective and from a getting the user where they are trying to perspective; don't try to solve for anything else.
- Find the game theory. If we can define the paths that are just not achievable, we need to just drop them. What are the viable paths and can be adjacent to what's possible.

    - The browser will not disintermediate the backchannel; we can rule all that out. We're only worried about session initiation between an IdP and where they are trying to go, IdP discovery, and metadata introspection in the browser to allow certain flows that would otherwise be disallowed due to privacy considerations.
    - Session expiration is a challenging topic as it timeboxed by both the IdP and SP; they don't have to match, but it does make for different behavior.
- FedCM is potentially a new protocol binding.
- FedCM asks "do you want to log into this website with this IdP" - it is a permission prompt with a very precise question. When the user says "yes" the browser is confident that the user made a choice. It is hard to abuse from a tracking perspective.
    - Don't need to make tracking impossible; need to make it economically unviable.
    - If FedCM could manage some version of whitelists, phishing attacks might be economically unviable too (?)
    - The trust bootstrapping has to be independent of the browser. It needs to be the RP tell the browser what trust framework they are members of. That brings in the fourth party to the IdP, and that would be the anti-phishing piece. This is a nice to have and wouldn't be much extra work.
    - The login process must be back in the hands of the user so it matches the experience they expect.
    - if we have a grammar to express trust at the level the browser cares about it, we can engineer the anti-phishing.
    - Would the prompt to browser be names or a list of descriptors? That's the discovery part.
    - Is there a way to build the database of SPs as well? That is part of the federation metadata.

14:00 - 17:00 - Breaking it down (sequence diagrams)

- Proxy service authentication (Judith)
- Local proxy use cases (Nicole)
- [throw a coffee break in here somewhere]

Early proposal discussion

- Trusted/Trustworth sandbox where we can do discovery; then the browser can see the discovery flow, and if it needs to get in the way of the user, it can interrupt the user. But what if we just take the Shib discovery feed (condensed discovery feed
    - a standard that can be trusted by the browser and then the RP can tell the browser what feed it trusts. The browser needs to know how to bootstrap
    - SPs can assert what IdPs they have a relationship; they aren't querying anything and so there is no privacy leakage between SPs (business concerns)
    - MDQ (metadata query) is an IETF draft that helps handle the fact that the metadata file will get huge
    - How can the browser be assured that nothing has fiddled with that data; want to cross reference what the RP is saying and what the Fed OP is saying
    - ~~The RP can provide key value pairs, the key being where to get the file and the value being what's in the file. This introduces a new phishing problem.~~
    - .well-known on the RP, and an index of strings to the stores; want to ping the service daily,
    - How to identify authoritative fed ops?
    - This allows us to classify this for federation vs tracking.
        - But what if FB of Google are in these lists?
    - This enables a ground truth for trust. Federation only works if there is trust, and this is moving some of this trust into the browser.
    - This becomes entirely backwards compatible as the browser will not even need to ask permission for the user to indicate the RP/IdP relationship. This resolves the classification problem of tracking vs federation, but if there were a list of entities that could be trusted, it would have that signal. If the signal is strong enough, it doesn't have to ask for the user's permission. If the browser ingests this and trusts/verifies it like a Certificate Authority. If it's just RP asserted, there would be less trust. Consider doing both. RP could filter it, but everything in it would be verified.
    - Limit the fingerprinting vector re: what is shown.
    - Trust oracle proposal: on a per-sector basis, have authorities like eduGAIN that sign the public keys of all their member federations, so the browsers only have to trust the per-sector oracle. You could do this with certificate transparency. This would allow for other federations outside of edu (e.g., Exostar from the US DOD). While any origin can claim anything, it also has to match what comes from the fed op.

Day Two - March 1

09:00 - 10:30 - Brainstorming Solutions

- allowlists based on trusted federations
- Discovery service offered by the browser

10:30 -10:45 - Coffee Break

10:45 - 12:30 - Brainstorming Solutions

- Evolving Consider supporting federated login without IDP APIs

12:30 - 14:00 - Lunch

14:00 - 15:00 - Brainstorming Solutions

15:00 - 15:15 - Coffee Break

15:15 - 16:30 - Brainstorming Solutions

16:30 - 17:00 - Next Steps

- IIW sessions

Useful Pre-reading

- "How to Study and Learn SAML" (old but the concepts are still correct and relevant)
- Kantara SAML Deployment Profile and Implementation Profile
- SeamlessAcccess Documentation
- SWITCH (Swiss federation) Demo
- SAML 101 - An intro to SAML for sysadmin
- https://met.refeds.org or https://technical.edugain.org (see also Chris Phillips' notes in issue 319)
- Session handling in IdPs (Shib wiki) and possibly Authentication Process
- This is an attempt to make a profile for handling SAML with FedCM https://github.com/nckroy/fedcm-rne-hacking/blob/main/documentation/fedCMDetailSeq.png

- Authentication and Authorisation for Research and Collaboration (AARC) Blueprint Architecture and linked documents
- NIAID Diagram (added 23 Feb): https://app.cloudcraft.co/view/5f106dfe-3845-4da4-90dc-8986bff02b56?key=8c7efbe5-ee9a-49ee-af5a-2897fb8948a8

Parking Lot of Questions

From our conversations at Seamless Access we formulated a set of questions to be explored (now and/or in the workshop), these largely relate to the user experience of this. IDP in this context is identity provider.

- Can we clearly signal purpose of a IDP (e.g. logging in with Google. vs gaining access to article via your University - both using the FedCM dialog)?
- Cross-domain persistence of the selected IDP (e.g. if I move between Springer.com and Elsevier.com) is key to the "seamless" part of our experience. Its unclear if FedCM can provide in this, if it requires going back into the IDP discovery flow, we lose this value (not sure if the wallet will help with this).
- Will browsers implement their own discovery UI? (we are following the spec work on discovery). We are not sure if also exposing a UI is the correct path, we feel it will limit the ability to create optimised UI's for purposes such as science (e.g. where we filter based on availability).
    - If not, can we actually leverage something a-like the cross-origin authentication for something like discovery? This would reduce the need for all publishing places to implement their own discovery UI.

Questions about the specification:

- Accounts endpoint:
    - 'The accounts list endpoint provides the list of accounts the user has at the IDP.'
        - they mean, the list of authenticated sessions the user has at the IdP? Or;
        - the list of registered accounts the user has with the IDM that sits behind the IdP
            - Which is a harder (impossible maybe) question. I assume it means the former.
    - **'Approved_clients:** A list of RPs (that gets matched against the requesting client_id) this account is already registered with'.

        - This does not scale very well when the IdP does not know who is requesting the account list.
    - As said by many, just because an IdP has a previous authenticated session (tied a session cookie), it does not mean it is appropriate for a given SP (e.g. wrong authentication assurance). So without a request context, I am not sure how selecting an account in this way is a good UX —the RP would eventually have to reject it.

Proposals

- **idp-sp-storage-api:** https://docs.google.com/document/d/1UWuw9_9EAnwtsct9fK6j7BN_ybcMSs2p0oxxb-U9O1s/edit?usp=sharing
- **Offloading Trust:** https://docs.google.com/document/d/13eFqxyQ6VdRh13lW1JiLA9YUUrJ_0S2ej5xixUDImQ0/edit#
-