Persistent Identifier Support

Most applications require unique identification of their users. Federation complicates this requirement because of the need for global uniqueness, privacy concerns, and the greater risk associated with the reassignment of identifiers. Reassignment is certainly an issue within a single domain, but the coupling there is tighter and the likelihood of processes to deal with it is higher.

Within higher education, and thus within InCommon, there are two principal mechanisms for user identification profiled for use with SAML: the eduPersonPrincipalName (EPPN) and eduPersonTargetedID attributes.

✓ **Recommended Practice**

- IdPs support the eduPersonPrincipalName and eduPersonTargetedID attributes.
- When SAML 2.0 is used, the "persistent" `<NameID>` format is used to represent the eduPersonTargetedID attribute.
- The release of eduPersonTargetedID is automated for most or all affiliates (save perhaps for students opting out under FERPA) to SPs that are not otherwise subject to user anonymity requirements, such as some library services.

IdPs are encouraged to support both the eduPersonPrincipalName and eduPersonTargetedID attributes, particularly in the case that the institution reassigns EPPN values to different users after periods of disuse. Even for those that do not reassign, there is value in eduPersonTargetedID in contributing to the privacy of one's users as they interact with different services.

eduPersonTargetedID Considerations

eduPersonTargetedID, unlike EPPN, does not have a single, universal string representation. Rather, it's a data "triple" that different deployments will manipulate into forms that are appropriate for their needs. Many applications will also struggle with their length and appearance; such applications are commonly those that also expect to receive personally-identifying attributes such as name and email address, defeating most of the privacy benefits of eduPersonTargetedID in isolation. In such cases, EPPN may be a better choice. In the case that reassignment is allowed, EPPN can be accompanied by eduPersonTargetedID to detect reassignment, but most packaged applications will be unable to rely on such an approach.

There is benefit to IdPs supporting eduPersonTargetedID in order to foster adoption by applications. Releasing it by default when anonymity is not a requirement is a strong step in encouraging its use, given the challenges associated with obtaining attribute release. The privacy risks are minimal, given all the other mechanisms that a single site can use to track repeated access.

There are two general approaches for producing "targeted" values: hashing and storage.

The hashing approach involves hashing fixed values associated with the transaction such as the IdP, SP, and user. Hashing is simple to deploy and doesn't increase the pain of clustering an IdP. However, it is sensitive to changes to any of the hash inputs, and can require additional coding work to allow for such changes, or to allow revising the value independent of the inputs.

The storage approach involves the random or hash-based generation of the value, followed by storage in a database of some kind for reuse. This has the advantage of accommodating lifecycle issues better, and allows for easy lookup/reversing of the identifier, but complicates clustering.

Based on the experiences of deployers, the use of a hashing strategy has proven adequate for most situations and is an easy path to deployment for most IdPs.