subject-id - SAML2 General Purpose Subject Identifier in SWAMID

SWAMID och alla andra akademiska identitetsfederationer inom eduGAIN inför en ny långsiktigt unik identifierare, subject-id, som på sikt ersätter eduPersonPrincipalName (ePPN). Denna wikisida beskriver både varför en ny långsiktigt unik identifierare behövs och hur vi under ordnade former byter.

Bakgrund

Sedan SWAMID bildades 2007 har eduPersonPrincipalName (ePPN) rekommenderats som identifierande attribut för att förmedla användarnamn från identitetsutfärdare (IdP) till tjänster (SP). En majoritet av tjänsterna i SWAMID använder ePPN som identifierare av användare. För att säkerställa att organisationen följer svensk lagstiftning avseende åtkomst och hantering av personuppgifter samt känsliga uppgifter, t.ex. Dataskyddsförordningen, kräver SWAMID i sina tillitsprofiler och i teknologiprofilen för SAML att ePPN är globalt unik och aldrig återanvänds till en annan individ:

SWAMID Identity Assurance profiles:

> *5.2.3 Each Subject MUST be represented by one or more globally unique identifiers.*
>
> *Subject identifiers MUST NOT be re-assigned.*

SWAMID SAML WebSSO Technology Profile:

> *5.5.3 An Identity Provider MUST support release of the attribute **eduPersonPrincipalName**. The value of the attribute for a Subject MUST NOT be reassigned to another Subject.*
>
> **Guidance:** The e-mail address of a Subject is not suitable as value for the attribute eduPersonPrincipalName due to name changes and later reassignments to other Subjects.

Ett problem med ePPN är att vissa identitetsfederationer inom eduGAIN tillåter att attributet återanvänds till andra individer. Det gör att attributet fungerar sämre som identifierare internationellt. När en individ slutar vid en organisation så är det högst sannolikt att individens behörigheter finns kvar i federerade tjänster, kopplade till individens användarnamn (ePPN). Om då en ny individ vid samma organisation får samma användarnamn kommer den nya individen med automatik ha samma behörigheter och tillgång till samma data som den föregående innehavaren av användarnamnet. Av denna anledning är det ett absolut krav att användarnamn ej återanvänds till andra individer.

För att hantera detta har en ny identifierare, subject-id, arbetats fram, definerad som General Purpose Subject Identifier (avsnitt 3.3) i SAML V2.0 Subject Identifier Attributes Profile Version 1.0.

Från SWAMID SAML WebSSO Technology Profile:

> *5.5.4 An Identity Provider MUST support the release of the attribute **subject-id**. The value of the attribute for a Subject MUST NOT be reassigned to another Subject.*
>
> **Guidance:** The subject-id is a globally unique identifier identical for all Relying Parties for a given Subject. SWAMID recommends that the value of eduPersonPrincipalName is used for subject-id since it is already defined for all Subjects, widely used as identifier in Relying Parties in SWAMID, unique and non-reassigned for all Identity Providers in SWAMID. The subject-id should not be change d as a result of a change to any other data associated with the Subject (e.g., name, email address, organisational role).

Skillnader mellan ePPN och subject-id

subject-id har samma egenskaper som ePPN har i SWAMID:

- Har ett scope (ex. @org.se)
- Får aldrig kopplas till en annan individ (någonsin)
- Liknar en e-postadress (men ska inte vara en e-postadress)
- Ska behandlas case-insensitive, dvs `AnvandarNamn@ORG.SE` ska räknas som samma värde som `anvandarnamn@org.se`

Det finns dock en viktig skillnad mellan ePPN och subject-id, dess tillåtna tecken:

- ePPN: `A-Z`, `a-z`, `0-9`, bindestreck (`-`), understreck (_), punkt (`.`)
- subject-id: `A-Z`, `a-z`, `0-9`, bindestreck (`-`), likhetstecken (=)

Övrigt att ta hänsyn till

- Det är lämpligt att samma case används för ePPN och subject-id för att minimera risken för mismatch i tjänster.

Planera och genomföra bytet i en identitetsutfärdare (IdP)

Det finns tre uppenbara lösningar på detta:

- Alternativ 1 - Använd ePPN som subject-id (om alla krav uppfylls)
- Alternativ 2 - Byt värde på ePPN (om krav på återanvändning ej uppfylls)
- Alternativ 3 - Översätt ePPN till subject-id på ett bestämt sätt (om ePPN aldrig återanvänds och punkt eller understreck förekommer i ePPN)
- Alternativ 4 - Behåll värde på ePPN, välj ett nytt värde till subject-id

Alternativen har både för- och nackdelar.

Alternativ 1 - Använd ePPN som subject-id om alla krav uppfylls

Om ePPN är valt på ett sätt som säkerställer att de aldrig återanvänds för annan individ och understreck (_) eller punkt (`.`) inte förekommer går det bra att använda samma värde för subject-id.

Fördelar:

- Tjänster kan ersätta ePPN med subject-id utan förändring
- Användarnamnen är väl inarbetade och känns igen av användarna

Nackdelar

- Inga

Alternativ 2 - Byt värde på ePPN (om krav på återanvändning ej uppfylls)

För organisationer som idag använder användarens e-postadress, eller något som baseras på e-postadressen, för ePPN kan det vara lämpligt att passa på att välja ett nytt värde för ePPN som sedan används även för subject-id.

- Välj ett nytt värde på subject-id som inte riskerar att återanvändas (och inte använts som ePPN av någon annan individ tidigare)
- Använd inte punkt (`.`) eller understreck (_) i nya ePPN-värden
- Välj något som är förhållandevis enkelt att hantera av människor
    - Dåligt exempel: `488d2f98-b670-4c13-aedf-c5b4d0783efb@org.se`
    - Bra exempel: `andber01@org.se` (för Anders Bertilsson, notera dock problematik kring namnbyte)
    - Bra exempel: `lusab-babad@org.se` (Proquints, används av eduID och Antagning.se)
- Använd samma värde för subject-id som för nytt ePPN-värde

Fördelar:

- Chans att börja om med unika identifierare som inte riskerar att återanvändas (till skillnad från e-postadresser som används som ePPN av några organisationer inom SWAMID idag)

Nackdelar

- Användare behöver hantera nya användarnamn (vid inloggning eller som visas i gränssnitt)
- Tjänster behöver hantera användarnamnbyte via exempelvis någon av dessa metoder:
    - Byta alla användarnamn i tjänsten samtidigt
    - Skapa upp nya användaridentiteter utan koppling till de gamla

Alternativ 3.1 - Översätt ePPN till subject-id genom att ta bort oönskade tecken

Om ePPN är valt på ett sätt som säkerställer att de aldrig återanvänds för annan individ men understreck (_) eller punkt (`.`) förekommer går det att ta bort oönskade tecken från ePPN för att bilda subject-id.

- Använd ePPN som bas för värde på subject-id
    - Ta bort punkt (`.`) och understreck (_)
    - Exempel:
        - `anna_b@org.se annab@org.se`
        - `fornamn.efternamn1-efternamn2@org.se fornamnefternamn1efternamn2@org.se`
- Säkerställ att ePPN (och subject-id) aldrig återanvänds

    - Detta följer av tillitsprofilen, teknologiprofilen samt GDPR och säkerhetsskäl då det föreligger risk av obehörig tillgång till annans data
    - Om e-postadress används idag, se till att de aldrig återanvänds och skapa en annan identifierare än e-postadress för nya användare (utan punkt och understreck), se Alternativ 2 nedan
    - Det förekommer idag att e-postadress återanvänds, ibland efter explicit rektorsbeslut

Fördelar:

- Tjänster kan ersätta ePPN med subject-id utan förändring, förutom för just de användare där understreck eller punkt förekommer i ePPN, och då antingen göra en översättning med borttagning av understreck och punkt eller kräva nya användaridentiteter i tjänsten för endast dessa användare
- Användarnamnen är väl inarbetade och känns i de flesta fall igen av användarna
- Användarnamnen innehåller inga oväntade teckenkombinationer

Nackdelar

- Det kan uppstå viss förvirring om vilket användarnamn det är som gäller, och det skiljer sig mellan tjänster som använder ePPN som identifierare och de som använder subject-id
- Risk finns för konflikt mellan användarnamn. I Ladok finns drygt 100 000 SWAMID-ePPN varav knappt 1 000 ePPN innehåller punkt eller understreck. Av dessa uppstår sju konflikter, varav fyra sannolikt är resultat av felstavning vid skapande av användaridentiteter som därför inte används.

Implementering i Shibboleth

Ersättning av punkt och understreck i Shibboleth Identity Provider finns beskrivet i Example of a standard attribute resolver for Shibboleth IdP v4 and above . Med mindre justeringar kan i stället tecken plockas bort. Innan dess behöver det säkerställas att två individer inte kan få samma subject-id.

Implementering i ADFS

Stöd för ändring av värde från ePPN till subject-id planeras till version 2.3 av ADFSToolkit.

Alternativ 3.2 - Översätt ePPN till subject-id med ersättningstecken

Om ePPN är valt på ett sätt som säkerställer att de aldrig återanvänds för annan individ men understreck (_) eller punkt (`.`) förekommer går det att ersätta oönskade tecken i ePPN för att bilda subject-id.

- Använd ePPN som bas för värde på subject-id
    - Ersätt punkt (`.`) med =2E
    - Ersätt understreck (_) med =5F
    - Alternativt byta punkter och understreck med bindestreck (`-`) om det ej föreligger risk för konflikter
    - Exempel:
        - `anna_b@org.se` `anna=5Fb@org.se`
        - Alternativt `anna_b@org.se` `anna-b@org.se`
        - `fornamn.efternamn1_efternamn2@org.se` `fornamn=2Eefternamn1=5Fefternamn2@org.se`
        - Alternativt `fornamn.efternamn1_efternamn2@org.se` `fornamn-efternamn1-efternamn2@org.se`
- Säkerställ att ePPN (och subject-id) aldrig återanvänds
    - Detta följer av tillitsprofilen, teknologiprofilen samt GDPR och säkerhetsskäl då det föreligger risk av obehörig tillgång till annans data
    - Om e-postadress används idag, se till att de aldrig återanvänds och skapa en annan identifierare än e-postadress för nya användare (utan punkt och understreck), se Alternativ 2 nedan
    - Det förekommer idag att e-postadress återanvänds, ibland efter explicit rektorsbeslut

Fördelar:

- Tjänster kan ersätta ePPN med subject-id utan förändring, förutom för just de användare där understreck eller punkt förekommer i ePPN, och då antingen göra en översättning med =5F/=2E eller kräva nya användaridentiteter i tjänsten för endast dessa användare
- Användarnamnen är väl inarbetade och känns i de flesta fall igen av användarna

Nackdelar

- De ePPN som innehåller understreck eller punkt får förvirrande nytt värde som subject-id med likhetstecken i sig
- Vissa väl inarbetade användarnamn känns inte längre igen i systemet och skiljer sig från vad som används vid inloggning
- Det kan uppstå viss förvirring om vilket användarnamn det är som gäller, och det skiljer sig mellan tjänster som använder ePPN som identifierare och de som använder subject-id

Implementering i Shibboleth

Ersättning av punkt och understreck i Shibboleth Identity Provider finns beskrivet i Example of a standard attribute resolver for Shibboleth IdP v4 and above.

Implementering i ADFS

Stöd för ändring av värde från ePPN till subject-id planeras till version 2.3 av ADFSToolkit.

Alternativ 4 - Välj nya värden för subject-id

Om ePPN är valt på ett sätt som säkerställer att de aldrig återanvänds för annan individ men understreck (_) eller punkt (`.`) förekommer finns chans att välja nya värden för dessa eller alla användare för att bilda subject-id.

- Välj ett nytt värde på subject-id som saknar direkt koppling till ePPN (och inte använts som ePPN av någon annan individ tidigare)

- Använd inte punkt (`.`) eller understreck (_) i nya subject-id-värden
- Välj något som är förhållandevis enkelt att hantera av människor
    - Dåligt exempel: `488d2f98-b670-4c13-aedf-c5b4d0783efb@org.se`
    - Bra exempel: `andber01@org.se` (för Anders Bertilsson, notera dock problematik kring namnbyte)
    - Bra exempel: `lusab-babad@org.se` (Proquints, används av eduID och Antagning.se)

Fördelar:

- Chans att börja om med unika identifierare som inte riskerar att återanvändas (till skillnad från e-postadresser som används som ePPN av många organisationer inom SWAMID idag)

Nackdelar

- Användare behöver hantera nya användarnamn (vid inloggning eller som visas i gränssnitt)
- Tjänster behöver hantera användarnamnbyte via exempelvis någon av dessa metoder:
    - Använda ePPN men spara undan mottaget subject-id under en övergångsperiod för att sedan gå över till bara subject-id-värden
    - Byta alla användarnamn i tjänsten samtidigt
    - Skapa upp nya användaridentiteter utan koppling till de gamla

Planera och genomföra bytet i en tjänst (SP)

Tänk på att alltid jämföra användaridentifierarna ePPN och subject-id case-insensitive, dvs `AnvandarNamn@ORG.SE` ska räknas som samma värde som `anvandarnamn@org.se`.

Förslag på process för att byta från ePPN som användaridentifierare till subject-id

Tjänster med få användare eller utan möjlighet att automatisera byte av användarnamn

1. Samla in nya användarnamn för användare (subject-id)
    a. Ändra manuellt i systemet
    b. Byt till subject-id som identifierare

Tjänster med många användare och möjlighet att automatisera byte användarnamn

1. Inventera vilka användare som använder tjänsten samt värden för ePPN och subject-id för dessa.
    a. Markera tjänsten med entitetskategorin Personalized för att ta emot subject-id från IdP:er
2. Jämför ePPN med subject-id
    a. Om de är lika för alla användare
        i. Byt till subject-id som identifierare
    b. Om de skiljer sig på ett deterministiskt sätt (om exempelvis alla punkter och understreck är bortplockade)
        i. Utarbeta regler i tjänsten för översättning
        ii. Uppdatera användardatabasen till subject-id-värdena
        iii. Byt till subject-id som identifierare
    c. Om de skiljer sig på ett icke deterministiskt sätt (om exempelvis många användare har helt andra värden på subject-id än ePPN)
        i. Spara undan subject-id för användare under en övergångsperiod (exempelvis via en extra kolumn i användartabellen i databasen eller loggning)
        ii. Upplys användare via en säker kanal om att de behöver logga in senast ett visst datum för att behålla tillgång till systemet
        iii. Byt till subject-id som identifierare