

Haka-kirjautumistilastot

Haka-resurssirekisterin yhteyteen on toteutettu palvelu, joka kerää kotiorganisaatioilta tilastotietoa tunnistuslähteen suorittamien Haka-kirjautumisten määrästä. Haka-resurssirekisterissä on myös käyttöliittymä, jolla tilastotietoa pääsee kaavioiden avulla tarkastelemaan. Käyttöliittymän kaavioista pääsee pureutumaan syvemmälle kirjautumistapahtumien määriin kutakin tietoa klikkaamalla aina kuukauden, tunnistuslähteen tai palvelun tarkkuudella. Tilastotiedon tarkastelu vaatii Haka-kirjautumisen. Käyttöliittymä on osoitteessa <https://rr.funet.fi/rr/stats.php>

Tilastointi uudistettiin vuoden 2013 lopulla. Uudella menetelmällä kerättävät tiedot käsittävät kirjautumisten yhteismäärän sekä kirjautumisten jakaantumisen palvelukohtaisesti. Sellaista tietoa, josta pystyisi yksilöimään kirjautuneen käyttäjän, ei kerätä.

Hakan kirjautumistilastoja kerätään IdP-palvelimilta haettavien siirtotiedostojen avulla. Haka-operointitiimi lataa siirtotiedostot IdP-ylläpitäjien ilmoittamista https-osoitteista kerran kuussa.

1. Siirtotiedosto-osoitteen lisääminen Resurssirekisteriin

IdP-ylläpitäjä voi lisätä kotiorganisaationsa IdP:n tilastosiirtotiedostoja sisältävän kansion polun suoraan [Resurssirekisteriin](#) IdP:n tietoihin (Muokkaa --> Statistics-välilehti).

2. Tilastosiirtotiedoston luonti IdP:llä

CSC:n käyttämä tilastosiirtotiedosto pohjautuu Shibboleth-ohjelmiston yhteyteen toteutettuun [lokianalyysityökaluun](#). Scriptin avulla voidaan lasketaan IdP:stä eri SP-palvelimille kohdistuneita kirjautumisia IdP:n audit-logeja hyödyntäen.

Tilastosiirtotiedostot tulee pitää jaossa siten, että operaattori pääsee lukemaan niitä https-yhteyden ylitse.

Hakan tilastosiirtotiedoston nimeämiskäytäntö IdP:n luomille tiedostoille on seuraava:

```
HakaStatistics_[vvvv-kk]_[IdP_hostname].txt
esim.
HakaStatistics_2013-10_idp.csc.fi.txt
```

Haka-operaattorin vaatima tilastosiirtotiedostomuoto on seuraava ('date' viittaa kuukauteen, jota tilastot koskevat):

```
IdPEntityId | [IdP EntityId]
date | [yyyy-mm]

logins      | relyingPartyId
-----
[num of logins] | [SP EntityId 1]
[num of logins] | [SP EntityId 2]
[num of logins] | [SP EntityId.....n]
```

Esimerkki Hakan tilastosiirtotiedoston sisällöstä:

```
IdPEntityId | https://idp.csc.fi/idp/shibboleth
date | 2013-08

logins      | relyingPartyId
-----
7890       | https://sui.csc.fi/shibboleth
456        | https://kotaplus.csc.fi
123        | https://aai.csc.fi
```

Scriptiä on muokattu hieman jotta se osuu RelyingPartyId:hen sekä käyttäjään ja msg_profile ei ole käytössä joten se osuu POST kenttään. muokatun scriptin saa tästä [loganalysisCSC.py](#)

Skriptiä voi käyttää testailla komentorivillä alla esiteltyyn tyyliin alustasta riippuen. Esimerkissä IdP on kontissa, jonka johdosta ei käytetä audit logeja, vaan kontin logitusta jossa kaikki logitus tulee yhteen tiedostoon.

```
# zgrep: Purkaa pakatut logitiedostot joiden prefix on idp_[edellisen kuukauden vuosi sekä kuukausi] sekä
nykyisen ei vielä pakatun idp.log tiedoston.
# grep: Filtteröi pois kaikki rivit joilla ei ole kyseisen kuukauden aikaleimaa ja toinen grep ottaa huomioon
vain POST rivit
# cut: Tällä poistamme omassa ympäristössämme tulevat ylimääräiset aikaleimat joita kontitus aiheuttaa.
LOCATION=/tmp;MONTHS=1;zgrep -a "Shibboleth-Audit" logs/idp_$(date -d "$MONTHS month ago" "+%Y%m").tgz logs
/idp.log|grep $(date -d "$MONTHS month ago" "+%Y-%m")|grep "POST"|cut -b 119- > $LOCATION/tmp_audit.log

# Tämä aloittaa logitiedoston tekemisen ja tuottaa kaksi ensimmäistä määrätyn muotoista riviä tiedostoon
LOCATION=/tmp;MONTHS=1;echo -e "IdPEntityId | https://idp.csc.fi/idp/shibboleth\ndate | $(date -d "$MONTHS
month ago" "+%Y-%m")" > $LOCATION/HakaStatistics_$(date -d "$MONTHS month ago" "+%Y-%m")_idp.csc.fi.txt;

# Tämä ajaa loganalysis työkalun ensimmäisellä komennolla tuotettua tiedostoa vasten.
LOCATION=/tmp;./loganalysisCSC.py -n $LOCATION/tmp_audit.log >> $LOCATION/HakaStatistics_$(date -d "$MONTHS
month ago" "+%Y-%m")_idp.csc.fi.txt

## Huomaathan että jos scriptaat tai ajat tätä cronissa tulee sinun käyttää date komennossa \%Y- \%m muotoa
(Escape).
```

Tarvitset siis absoluuttisen polun skriptille, idp-audit.logien sijainnille sekä tulosteelle. Huomaathan myös että tilastosiirtotiedoston tulee sijaita hakemistossa, josta tuo voidaan noutaa https-protokollaa käyttäen.

Lopuksi komento voidaan antaa vaikka cronin ajettavaksi kertaalleen jokaisen kuukauden ensimmäisenä päivänä.