

Tietoliikenneverkkojen ylläpito tiedon louhinnan haasteena

Vierailuesitelmä Helsingin yliopiston Tietojenkäsittelytieteen laitoksella Tietämyksen muodostamisen -seminaarissa, 9.4.2002.

Tietoliikenneverkko on monimutkainen järjestelmä, jonka tulee toimia 24 tuntia vuorokaudessa jokaisena vuoden päivänä. Verkko tarjoaa samanaikaisesti useita, monella eri tekniikalla toteutettuja palveluita. Palveluiden laatua ja käyttöä valvotaan monella toteutuksen tasolla ja valvontadataa käyttävät useat operaattoreiden toiminnot kuten laskutus, ylläpito, suunnittelu jne.

Matkapuhelinverkko on osa tietoliikenneverkkoa. Tyypillinen valvottava kokonaisuus voi kattaa tuhansia tukiasemia hyvin laajalla maantieteellisellä alueella. Matkapuhelinverkon rakenne on hierarkinen: jokaisella tasolla verkkoelementeistä kerätään useita kymmeniä, jopa satoja aikasarjoja, jotka kootaan yhteen valvontajärjestelmän tietokantaan. Näiden lisäksi verkko saattaa tuottaa päivässä useita tuhansia hälytyksiä, jotka myös päätyvät valvontajärjestelmään.

Operaattorin toimintoja tukeva tiedon erottaminen tästä data tulvasta on haaste, johon viimeisen vuosikymmenen aikana on enenevässä määrin sovellettu erilaisia tiedon louhinnan menetelmiä. Sovellettuihin menetelmiin kuuluu mm. assosiaatiosääntöjen etsintä hälytysten käsittelyyn ja ohjaamatonta luokittelua (unsupervised clustering) verkkoelementtien ja verkon käyttäjien analysointiin. Ongelmaa vaikeuttaa toisaalta verkon laajuus ja monimutkaisuus ja toisaalta yhtä ongelmaa kohti käytettävissä oleva rajallinen aika ja resurssit.

Verrattuna kirjallisuudessa esiteltyihin tietämyksen muodostamisen ja tiedon louhinnan prosesseihin, eroaa tietoliikenneverkkojen ylläpitoa tukeva tiedon louhinta niistä merkittävästi. Verkkojen ylläpidossa kysessä on jatkuvan prosessin tilojen tarkkailu, kun taas kirjallisuuden mallit on suunnattu kerta luonteisiin tieteellisiin tai taloudellisiin sovelluksiin. Tästä seuraa ristiriita, joka helposti vaikeuttaa tiedon louhintasovellusten suunnittelua ja käyttöä huomattavasti.