

# Suositus verkkopalvelun kävijöiden, sivupyyntöjen ja vierailujen mittaamiseksi

## Dokumentin tila: ehdotus suositukseksi

9.9.1999

### Tekijät

- Hannele Antikainen
- Vesa Kautto
- Markku Kylänpää
- Heikki Nikulin

### Alkusanat

Nämä suositukset on laadittu Suomen Mediatarkastuksen vetämässä hankkeessa, jossa VTT Tietotekniikka on toiminut asiantuntijaorganisaationa. Projektiin osallistuivat:

- Alma Media Oyj
- Grey Interactive Oy
- Mainostajien Liitto
- Mainostoimistojen Liitto
- Nedecon Network Development Consulting Oyj
- Sanoma-WSOY
- Saunalahden Serveri
- Sonera
- Suomen Gallup Web Oy
- Taloustutkimus Oy
- Telia Finland Oy/Netpool Finland Oy

Hankkeessa laadittiin suositus, joiden mukaan verkkopalvelujen mittaamisen keskeiset tunnusluvut voidaan laskea. Suositus on parhaillaan (syksy 1999) Suomen Mediatarkastuksen käsiteltävänä. Verkkopalvelujen mittaamisen keskeiset tunnusluvut ovat kävijä, sivupyyntö ja vierailu. Näiden suositusten toivotaan yhtenäistävän termistöä, jota verkkopalvelujen mittaamisessa käytetään. Ne toimivat myös lähtökohtana verkkopalvelujen mittausjärjestelmien kehittämisessä. Tavoitteena on, että mainostajille voitaisiin esittää eri verkkopalveluista samoilla mittausperiaatteilla saatuja tunnuslukuja. Tämä on yksi verkkomainontakampanjoiden suunnittelun perusedellytyksistä.

### Sisällys

<b>1. JOHDANTO</b>	<b>2</b>
<b>2. KÄSITTEET</b>	<b>2</b>
2.1 KÄVIJÄ	2
2.2 SIVUPYYNTÖ	3
2.3 VIERAILU	6
2.4 MUUT KÄSITTEET	6
<b>LÄHTEET</b>	<b>7</b>
<b>SANASTO</b>	<b>7</b>
<b>LIITE 1. MITTAUSJÄRJESTELMÄT</b>	
<b>LIITE 2. LUETTELO HAKUKONEISTA JA HENKILÖKOHTAISISTA ROBOTEISTA</b>	

# 1. Johdanto

Verkkopalvelun mittaaminen perustuu tietoihin, jotka syntyvät kun käyttäjä selainohjelmansa avulla vierailee verkkopalvelussa. Verkkopalvelujen mittaamiseen tarjolla on useita erilaisia mittaamenetelmiä. Niissä kaikissa on kuitenkin puutteita, minkä vuoksi selainohjelman avulla tehtyjen sivupyynnöiden ja vierailujen määrää ei pystytä tarkasti mittaamaan. Tämän vuoksi mittaustuloksia tarkasteltaessa on aina selvitettävä, millaisella menetelmällä ne on laskettu. Eri menetelmiin perustuvilla mittaajärjestelmillä ja jopa samalla mittaajärjestelmällä, mutta eri selainohjelman asetuksilla, saadaan erilaisia mittaustuloksia. Keskeinen syy erilaisiin mittaustuloksiin on verkkoselailun toimintaympäristö ja erityisesti selainohjelmien suosituksista poikkeavat toteutukset. Mm. proxy-palvelimet, selaimien välimuistit ja muut selainohjelmien asetukset, tiedonsiirtoreitti Internetissä ja tiedonsiirtovirheet vaikuttavat mittaustuloksiin.

Verkkopalvelun mittaamisessa kohteena on verkkopalvelu, joka on antanut sivustonsa mittauksen mittaupalvelua tarjoavalle yritykselle. Mittaustuloksiin vaikuttaa yrityksen käyttämän mittaamenetelmän lisäksi mitattavaksi valittujen sivujen määrä sekä joskus myös se, kuinka mittauselementit sijoitetaan mitattaville sivuille ja millainen on mitattavien sivujen rakenne. Verkkopalvelun ylläpitäjän vastuulla on huolehtia siitä, että sivuston rakenne ja mittauselementtien sijoittelu on tehty tämän suosituksen mukaan. Vastaavasti mittaupalveluja tarjoavan yrityksen vastuulla on se, että mittaajärjestelmä toimii suosituksen mukaisesti ja mittaajärjestelmän rekisteröimät tiedot on jalostettu suosituksen mukaisesti. Mittaustulosten luotettavuus tulee paranemaan sitten, kun Suomeen saadaan auditointipalvelu. Suomen Mediatarkastus suunnittelee parhaillaan sekä mittaajärjestelmien että verkkopalvelujen auditointia.

## 2. Käsitteet

Verkkopalvelun mittaamisessa käytettävät kolme tunnuslukua ovat:

- kävijä
- sivupyyntö
- vierailu

Näiden peruskäsitteiden lisäksi tässä suosituksessa määritellään joukko muita verkkopalvelujen mittaamiseen liittyviä oleellisia käsitteitä.

### 2.1 Kävijä

#### **Määritelmä:**

Kävijä on jollakin menetelmällä tunnistettu käyttäjä.

Tunnistukseen soveltuvia tekniikoita ovat tällä hetkellä käyttäjätunnuksen ja salasanan käyttäminen, sekä cookie-tekniikka, jossa on otettu huomioon cookiet hylkäävien kävijöiden laskenta.

#### **Soveltaminen:**

Käyttäjätunnuksen vaativissa verkkopalveluissa käyttäjä antaa verkkopalvelulle tietoja itsestään ja saa käyttäjätunnuksen ja salasanan, jotka oikeuttavat jatkossa vierailemaan verkkopalvelussa. Käyttäjätunnuksen vaativissa verkkopalveluissa voidaan helposti selvittää yhden kävijän tekemät sivupyynnöt. Käyttäjä pysytään yksilöimään, vaikka hän käyttäisi palvelua eri tietokoneilta ja selainohjelmista. Mutta jos käyttäjä rekisteröityy toiseen kertaan, esimerkiksi salasanan unohtamisen seurauksena, syntyy uusi kävijä.

Käyttäjät voidaan tunnistaa myös cookie-tekniikan avulla. Cookie-tieto lähetetään käyttäjän tietokoneelle, kun hän vierailee verkkopalvelussa ensimmäistä kertaa. Seuraavia sivuja pyytäessä käyttäjän selainohjelma ja tietokone tunnistetaan cookieiden avulla, sillä cookie-tieto palautuu tällöin sen lähettäneelle tietokoneelle. Itse käyttäjää ei siis tunnisteta. Jos sama käyttäjä on yhteydessä samaan verkkopalveluun esimerkiksi kahdelta työpaikan tietokoneelta sekä kotitietokoneensa kahdella eri selainohjelmalla, voi verkkopalveluun syntyä neljä eri kävijää.

## Haamu-cookies on poistettava

Cookie-tietojen käyttämistä kävijän tunnistamisessa haittaa se, että käyttäjä voi itse määrittellä, kuinka verkkopalveluiden lähettämiä cookie-tietoja käsitellään. Jos käyttäjä ei hyväksy cookie-tietojen tallentamista tietokoneelleen, saa hän jokaisen sivupyynnön yhteydessä uuden cookie-tiedon. Jos kävijöiden ja vierailujen mittaaminen perustuu näihin cookieihin, on tuloksena reilu ylimittaus. Tämän vuoksi on saman käyttäjän synnyttämät, ensimmäisen sivupyynnön jälkeen lähetetyt cookie-tiedot eli haamu-cookies on poistettava.

Haamu-cookieiden poistamiseen voidaan käyttää useita menetelmiä. Poistomenetelmissä on eroja, jotka voivat vaikuttaa mitaustuloksiin. Tämä on pidettävä mielessä kun eri menetelmillä saatuja kävijämääriä vertaillaan. Auditoinnissa tarkastetaan, että haamucookieiden synnyttämät kävijät ja vierailut on poistettu.

Cookie-tietojen voimassaoloajaksi suositellaan mahdollisimman pitkää aikaa, vähintään kolmea vuotta. On kuitenkin huomattava, että käyttäjän uusiessa tietokoneensa, saa hän uuden cookieiden ja on WWW-palvelun mittausjärjestelmälle uusi kävijä. Uutta cookieita tarjotaan aina myös silloin, kun käyttäjä on syystä tai toisesta kadottanut voimassa olevat cookies (esim. uusi kovalevy). Joissain tapauksissa uusi cookie voidaan asettaa myös, kun käyttäjä päivittää selainohjelmansa uuteen tai vain muuttaa selainohjelmansa tiettyjä asetuksia.

On arvioitu, että cookies hylkäävien kävijöiden määrä on verkkopalvelusta riippuen muutaman prosentin luokkaa.

Yhteiskäytössä olevan tietokoneen selainohjelmalla voi olla lukuisia käyttäjiä, jotka cookie-pohjaisessa tunnistamisessa tulkitaan yhdeksi käyttäjäksi. Tällaisia tietokoneita on tyypillisesti oppilaitoksissa, kirjastoissa ja kodeissa. On kuitenkin mahdollista antaa eri kävijöille eri cookies myös silloin, kun tietokoneen käyttö perustuu käyttäjätunnukseen.

## 2.2 Sivupyynnöt

### **Määritelmä:**

Sivupyynnöt: Yhden tai useamman tiedoston mahdollisuus tulla näytetyksi käyttäjän selaimen ikkunassa.

Sivupyynnöiden laskennan periaatteena on, että yksi käyttäjän tekemä pyyntö synnyttää yhden sivupyynnön ja että vain inhimillisen käytön seurauksena syntyvät sivupyynnöt lasketaan.

### **Soveltaminen:**

Sivupyynnöt syntyvät, kun käyttäjä kirjoittaa sivun Internet-osoitteen selaimen osoitekenttään, tai kun hän valitsee sivun bookmark/favorites-valikosta tai siirtyä sivulle edelliseltä WWW-sivulta valitsemallaan linkillä. Lisäksi käyttäjä voi ladata sivuja selainohjelman painikkeiden avulla.

Sivupyynnöiksi lasketaan seuraavat käyttäjän toimenpiteet tai selaimen painikkeiden klikkaukset:

- back- ja forward
- reload/refresh
- home
- print
- selainikkunan resize

Jotkut selaimet eivät tuota sivupyyntöä print- ja resize-komennoilla. Tällä hetkellä (1999) Netscapen selainohjelmat tekevät tulostus- ja resize-käskyllä uuden sivupyynnön. Sen sijaan Microsoftin Internet Explorer ei tee sivupyyntöä resize-käskyllä. Explorerin tulostuskäskyllä sivupyyntö syntyy ohjelman versiosta riippuen.

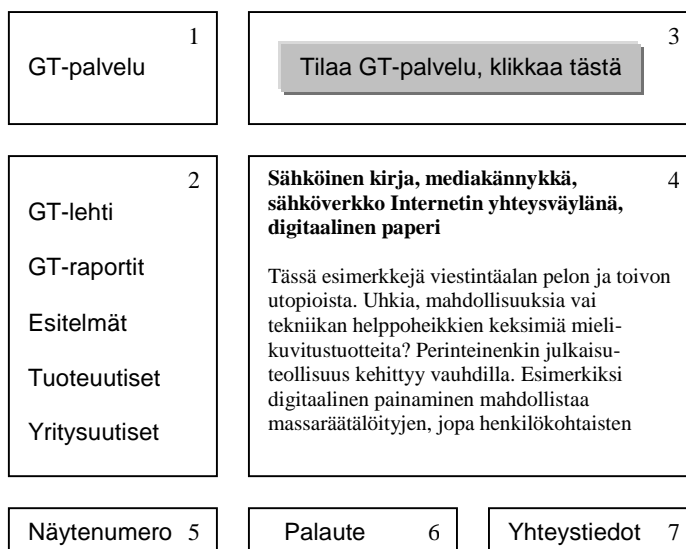
Sivupyyntöjen lopulliseen määrään ei lasketa:

- Automaattisella selainikkunan refresh-toiminnolla synnytettyjä sivupyyntöjä, jossa käyttäjä ei aktiivisesti pyydä sivua.

Jos automaattista refresh-toimintoa käytetään mitattavilla sivuilla, voidaan sivupyynnöksi laskea kävijää kohden vain yksi sivu, ei useampaa. Viime kädessä on auditoinnin asia huolehtia siitä, että automaattisen refresh-toiminnon aiheuttamia sivupyyntöjä ei lasketa tunnuslukuihin.

## Kehyssidut

WWW-sivun osiin jakavien kehyssivujen (framed pages) rakenne suositellaan suunniteltavaksi siten, että yksi käyttäjän yksi klikkaus synnyttää yhden sivupyynnön (kuva 1), ja että klikkauksen seurauksena oleellinen osa selaimen ikkunan sisällöstä muuttuu. Jos käyttäjän tekemä klikkaus kuitenkin synnyttää useampia kuin yhden sivupyynnön, on nämä sivupyynnöt poistettava julkistettavista tunnusluvuista. Kehyssivujen rakenteen muuttaminen voi olla tarpeen, jotta sivupyyntöjen mittaaminen voidaan tehdä suosituksen mukaan. Kehyssivujen rakenteesta vastaa verkkopalvelun ylläpitäjä.



*Kuva 1. Esimerkki kehyssivusta. Jotta mittaus voidaan tehdä suosituksen mukaan, suositellaan sivun rakennetta, jossa yksi käyttäjän suorittama klikkaus tuottaa yhden sivupyynnön. Virheellinen, liian suuri sivupyyntöjen määrä saadaan, jos yhdellä klikkauksella selaimen ikkunaan ladataan jutun lisäksi sivupyyntöjä aiheuttavia navigointisivuja.*

## Kanavasivut

Ns. kanaviin ohjattuja WWW-sivuja ei lasketa sivupyynnöiksi. Tällaiset sivut tulee varustaa tunnisteella, jonka avulla ne voidaan suodattaa pois julkistettavista tunnusluvuista. Kanavasivujen asianmukainen merkintä kuuluu verkkopalvelun ylläpitäjälle.

## Hakukoneet

Hakukoneiden ja henkilökohtaisten hakurobottien aiheuttamia sivupyynnöitä ei lasketa sivupyynnöiksi eikä niitä pidä sisällyttää tunnuslukuihin. Ulkopuoliseen mittauspalvelimeen ja mittauselementtiin perustuvassa mittauksessa hakukoneet eivät aiheuta sivupyynnöitä, jos hakukoneet eivät hae kuvia (mittauselementti on kuva, kts. liite 1, kohta 2). Viime kädessä auditoinnin tehtävänä on huolehtia siitä, että hakurobottien aiheuttamat sivupyynnöt tulevat poistetuksi tunnusluvuista. ABC Interactive julkaisee luetteloja hakukoneista ja henkilökohtaisista roboteista (liite 2).

## Sivupyynnöt omasta verkosta

Yrityksen omasta lähiverkosta tulevat, verkkopalvelun kehittämiseen liittyvät sivupyynnöt sekä verkkopalvelun mittaamisen testaukseen liittyvät sivupyynnöitä ei lasketa sivupyynnöiksi. Vastaavasti on poistettava myös näin syntyneet kävijät ja vierailut. Muut kuin kehittämistarkoituksessa tehdyt sivupyynnöt yrityksen omasta lähiverkosta voidaan laskea sivupyynnöiksi.

## Sivupyynnöt - sivulataus

Sivupyynnössä tiedetään, että sivu on lähtenyt onnistuneesti verkkopalvelun ylläpitäjän palvelimelta. Vastaavasti tiedetään, että mittausjärjestelmän palvelimelta on haettu mittauselementti. Sitä, onko sivupyynnön seurauksena latautunut täydellisesti käyttäjän selaimen ikkunaan, ei varmuudella voida tietää. Ei myöskään voida tietää, onko käyttäjä lukenut sivun. Toki suurin osa ladatuista sivuista päättyy selaimen ikkunaan, mutta käyttäjä voi esimerkiksi tiedonsiirto-ongelmien vuoksi keskeyttää sivun latautumisen tai käyttäjä tekee uuden sivupyynnön esimerkiksi aktivoimalla latautumassa olevan dokumentin linkin.

Verkkopalvelujen mittaamisessa suositellaan käytettäväksi termiä sivupyynnöt, koska se vastaa tällä hetkellä käytännön tilannetta paremmin kuin termi sivulataus. Syy on se, että sivulataus sisältää ajatuksen, että sivu on kokonaisuudessaan latautunut käyttäjän selaimen ikkunaan.

## Eri verkkopalveluiden sivupyynnöitä ei voi vertailla

On huomattava, että useimmissa verkkopalveluissa palvelun tarjoaja valitsee mitattavat sivut ja että mitattavien sivujen määrä vaihtelee merkittävästi verkkopalvelusta toiseen. Tämän vuoksi sivupyynnöiden määrä ei sovellu eri verkkopalveluiden vertailemiseen. Sivupyynnöiden määrä ei sovellu edes saman verkkopalvelun keskinäiseen vertailemiseen eri ajanjaksoina, jos mitattavaksi valittujen sivujen lukumäärä vaihtelee. Sivupyynnöiden kokonaismäärää paremmin vertailuun sopii yksittäisen WWW-sivun tunnusluvut. Tässä suosituksessa esitetyistä tunnusluvuista parhaiten eri verkkopalvelujen vertailuun sopii kävijöiden määrä edellyttäen, että kävijämäärät on laskettu samalla menetelmällä.

## Internet Explorer 3.0 aiheuttaa ongelmia

Internet Explorer 3.0 (IE 3.0) selainohjelma on ongelmallinen useimpien mittausjärjestelmien kannalta, sillä ne eivät kykene mittaamaan kuin osan IE 3.0:lla pyydytyistä sivuista. Syynä on tämän selainohjelman version suosituksista poikkeava toteutustapa. Eräissä mittausjärjestelmissä on lähdetty laskennallisesti korjaamaan sivupyynnöiden määrää siten, että ne vastaisivat paremmin todellisia sivupyynnöitä. Tällainen sivupyynnöiden määrän kasvattaminen ei kuitenkaan ole tämän suosituksen mukaista. Ongelma on pienenee koko ajan, sillä IE 3 -selainohjelman käyttö vähenee koko ajan, kun käyttäjät siirtyvät Internet Explorerin versioihin 4 ja 5.

## 2.3 Vierailu

### **Määritelmä:**

Vierailu on tunnistetun kävijän samasta verkkopalvelusta tekemien sivupyynnöiden sarja, jossa peräkkäisten sivupyynnöiden väli on alle 30 minuuttia.

Jotta tiedetään, että sivupyynnöt ovat yksittäisen käyttäjän tekemiä, on käyttäjä tunnistettava kävijäksi (vrt. luku 2.1).

### **Soveltaminen:**

Vierailun aikana käyttäjä voi tehdä yhden tai useamman sivupyynnön. Jos käyttäjä pitää selailussaan tasan tai yli 30 minuutin pituisen tauon, lasketaan tauon jälkeen alkanut vierailu uudeksi vierailuksi. 30 minuuttia on valittu, koska se yleisimmin käytetty ajanjakso. Syitä käyttäjän yli 30 minuutin selailutaukoon voi olla useita: käyttäjä on poistunut tietokoneensa äärestä tai hän on vierailut muissa verkkopalveluissa. Hän on toki voinut myös tutustua selainikkunan sisältöön koko ajan.

Kävijöiden tunnistamisen perustuessa cookie-tietoihin (luku 2.1) vääristyy myös vierailujen määrä, jos edellä mainittuja saman käyttäjän synnyttämiä haamu-cookieita ei poisteta. Kuten edellä todettiin, haamu-cookieit syntyvät silloin, kun käyttäjä on asettanut selainohjelmansa niin, että hän ei hyväksy cookie-tietojen tallentamista tietokoneensa muistiin. Jos vierailujen laskennassa haamu-cookieita ei poisteta, on tuloksena vierailujen ylimittaus.

## 2.4 Muut käsitteet

### Sivulataus (Page Impression)

Sivulataus tarkoittaa WWW-sivun täydellistä latautumista käyttäjäselaimen ikkunaan. Sivulatausta edeltää käyttäjän tekemä sivupyynnö. Yleensä sivulatauksella tarkoitetaan kuitenkin sivupyynnöä.

### Selain

Selain on tietokoneohjelma, jonka avulla käyttäjä on yhteydessä verkkopalveluun ja lataa näyttöruudulleen WWW-sivuja. Käyttäjä voi itse määrittellä oman selaimensa asetukset. Näissä asetuksissa voidaan sivujen latautumisen nopeuttamiseksi estää kuvien latautuminen. Tällöin mittauspalvelut, joiden mittauselementti perustuu pieneen kuvaan, eivät pysty rekisteröimään sivupyynnöitä.

Selain voi tallentaa käyttäjän viimeksi pyytämät WWW-sivut tietokoneen muistiin (cache). Jos käyttäjä palaa näille sivuille, ladataan sivu - selainohjelmasta riippuen - käyttäjän tietokoneen muistista, ei verkkopalvelun palvelimelta. Tämä nopeuttaa sivun latautumista ja vähentää verkon kuormitusta, mutta tuloksena voi olla sivupyynnöiden alimittaus.

### Cookie

Cookie on tietoa, jonka palvelin lähettää selainohjelmalle ja jonka selainohjelma palauttaa palvelimelle joka kerta, kun selainohjelmalla ollaan uudelleen yhteydessä palvelimeen. Verkkopalvelun mittaamisessa cookie-tietoa käytetään kävijöiden (selainohjelmien) tunnistamiseen.

## Hitti

Hitti on yksi WWW-sivun sisältämä elementti, yleensä teksti- tai kuvatiedosto. Kun käyttäjä pyytää sivun selaimensa ikkunaan, syntyy kaikista WWW-sivun sisältämistä elementeistä - teksti, kuvat ja muut elementit - verkkopalvelun palvelimenlokiin hitti. Koska eri verkkopalveluiden sisältämien WWW-sivujen rakenne vaihtelee, ei hitti kerro sivupyynnöiden, vierailujen tai kävijöiden määrää, vaan verkkopalvelusta ladattujen yksittäisten kuva-, teksti- ja muiden elementtien määrän. Hitti ei sovellu verkkopalvelujen mittaamisen tunnusluvuksi.

## Lähteet

1. Metrics and Methodology. Internet Advertising Bureau's Media Measurement Task Force, 1997.  
<http://www.iab.net/advertise/content/mmtf3.html>
2. Definitions of Web Site Statistics and Terms", Advertising MediaInternet Center,  
<http://www.amic.com/I-trak/webglossary.html>
3. Verkkopalvelun kävijöiden, sivupyynnöiden ja vierailujen mittaamisen käsikirja. (julkaistaan)

## Sanasto

### Cookie

Cookie on tietoa, jonka palvelin lähettää selainohjelmalle ja jonka selainohjelma palauttaa palvelimelle joka kerta, kun selainohjelmalla ollaan uudelleen yhteydessä palvelimeen. Cookie-tietoa voidaan soveltaa monella tavalla. Verkkopalvelun mittaamisessa cookie-tietoa käytetään kävijöiden (selainohjelmien) tunnistamiseen.

### Hitti

Hitti on yksi WWW-sivun sisältämä elementti, yleensä teksti- tai kuvatiedosto. Kun käyttäjä pyytää sivun selaimensa ikkunaan, syntyy kaikista WWW-sivun sisältämistä elementeistä - teksti, kuvat ja muut elementit - verkkopalvelun palvelimenlokiin hitti. Koska eri verkkopalveluiden sisältämien WWW-sivujen rakenne vaihtelee, ei hitti kerro sivupyynnöiden, vierailujen tai kävijöiden määrää, vaan verkkopalvelusta ladattujen yksittäisten kuva-, teksti- ja muiden elementtien määrän. Hitti ei sovellu verkkopalvelujen mittaamisen tunnusluvuksi.

### HTML

Hypertext Markup Language. WWW-dokumenttien koodauskieli.

### HTTP

HTTP eli Hypertext Transfer Protocol on Internetin yhteyskäytäntö, joka siirtää www-sivut palvelimelta käyttäjän selaimen ikkunaan. HTTP-header määrittelee asetukset, joita sivujen siirtämisessä suositellaan noudatettavaksi.

### Java

Sunin kehittämä ohjelmointikieli, jolla tehdyt ohjelmat toimivat periaatteessa kaikilla laitealustoilla. JavaScript on puolestaan Netscapen kehittämä HTML-kielen laajennus.

## **Käyttäjä**

Käyttäjä on WWW-selaimellaan sivupyynnöjä tekevä henkilö. Kun verkkopalvelu tunnistaa käyttäjän jollain tekniikalla, voidaan käyttäjä laskea verkkopalvelun kävijäksi.

## **Kävijä**

Kävijä on jollain menetelmällä tunnistettu käyttäjä. Tunnistukseen soveltuvia tekniikoita ovat cookiet tai käyttäjä tunnuksen ja salasanan käyttö. Cookeja käytettäessä tunnistetaan käyttäjän selain.

## **Loki**

Loki on tiedosto, johon palvelimen tapahtumat kirjautuvat.

## **Mittauselementti**

Usein palvelin pohjaisissa mittausjärjestelmissä mittauselementtinä käytetty pieni kuva, joka ladataan käyttäjän pyytämän WWW-sivun ikkunaan mittauspalvelimelta. Mittauselementti sijoitetaan yleensä WWW-sivun alkuun.

## **Palomuuuri**

Palomuuuri on ratkaisu, jossa yrityksen paikallinen verkko ei turvallisuussyistä näy ulkopuolisille. Palomuurin asetuksista riippuen kaikki organisaation tekemät sivupyynnot voivat olla peräisin yhdestä verkko-osoitteesta riippumatta siitä, kuinka moni käyttäjä sivuja selaimellaan pyytää. Palomuurilla voidaan estää myös Java- ja muiden ohjelmakoodien suorittaminen palomuurin takana olevissa selaimissa.

## **Proxy-palvelin**

Proxy on palvelin, jonka toimii suosittujen WWW-sivujen varastona. Jos käyttäjän selaimelleen pyytämä sivu löytyy proxyn välimuistista (cache), ladataan tämä sivu, mikä tapahtuu nopeammin kuin jos sivu haettaisiin alkuperäiseltä palvelimelta. Tämä johtaa siihen, että alkuperäisellä palvelimella ei ole tietoa näistä sivulatauksista. Proxy-palvelimen asetuksilla voidaan vaikuttaa siihen, mitä tietoa sen kautta voi kulkea. Internet-operaattorien proxy-palvelimia käytetään siten, että asiakkaan verkko-osoite muuttuu käyttökerrasta toiseen. Tämän vuoksi IP-osoitteen avulla käyttäjää ei voida tunnistaa luotettavasti.

## **Refresh**

Sivupyynnöjä voidaan generoida automaattisesti ns. refresh-toiminnon avulla ilman, että käyttäjä tekee sivupyynnöitä. Tämä tapahtuu HTML-dokumentin head-osassa olevien tietojen perusteella. Näissä tiedoissa määritellään mikä sivu käyttäjän selaimen ikkunaan ladataan. Samoin määritellään milloin uuden sivun lataaminen aloitetaan. Tällaisia ratkaisuja voidaan käyttää verkkopalvelun aloitussivulla, jonka latautumisen jälkeen käyttäjä saa automaattisesti selaimensa ikkunaan uuden sivun. Näitä automaattisesti generoituvia sivuja ei lasketa sivupyynnöiksi.

## **Selain**

Selain eli selainohjelma on tietokoneohjelma, jonka avulla käyttäjä on yhteydessä verkkopalveluun ja lataa WWW-sivuja. Käyttäjä voi itse määritellä oman selaimensa asetukset. Hän voi mm. estää kuvien latautumisen sivujen latautumisen nopeuttamiseksi. Selain voi tallentaa käyttäjän viimeksi pyytämät WWW-sivut tietokoneen muistiin. Jos käyttäjä palaa näille sivuille, voidaan sivu ladata joko käyttäjän tietokoneen muistista tai verkkopalvelun palvelimelta. Näiden sivujen selailuun voidaan käyttää selaimen back- ja forward-painikkeita. Reload- ja refresh-painikkeilla käyttäjä voi puolestaan ladata WWW-palvelimelta sivun uudestaan selaimensa ikkunaan.



## **Sivu**

Sivu eli WWW-sivu on HTML-koodattua tekstiä ja muita elementtejä, jotka käytetään käyttäjän selaimen ikkunassa käyttäjän määrittelemillä selainasetuksilla. Verkkopalvelut eli WWW-sivustot ovat kokoelma tällaisia WWW-sivuja. WWW-sivujen mahdollisia elementtejä ovat teksti, kuvat sekä multimediaelementit, kuten äänitiedostot ja videokuvatiedostot sekä selaimessa suoritettava ohjelmakoodi (Java, JavaScript, Active-X) ja tyyliomakkeet. Verkkopalvelun mittaamenetelmästä riippuen WWW-sivun osana voi olla myös mittauselementti. WWW-sivulla on yksikäsitteinen verkko-osoite eli URL (Universal Resource Locator).

## **Sivupyynnö (Page Request)**

Sivupyynnössä (Page Request) yhdellä tai useammalla tiedostolla on mahdollisuus tulla näytetyksi käyttäjän WWW-selaimen ikkunassa. Sivupyynnö rekisteröidään mittausohjelmiston avulla. Sivupyynnö syntyy, kun selaimen käyttäjä pyytää WWW-sivuja selaimensa ikkunaan. Ajatuksena on, että käyttäjän tekemä yksi klikkaus voi synnyttää vain yhden sivupyynnön.

## **Sivulataus (Page Impression)**

Sivulataus tarkoittaa HTML-sivun täydellistä latautumista käyttäjä selaimen ikkunaan. Sivulatausta edeltää käyttäjän tekemä sivupyynnö.

## **Sivunluku aika**

Sivunlukuajan mittaamiseen ei ole syntynyt vakiintunutta käytäntöä. Tällä hetkellä Java-pohjaiseen aktiiviseen mittauselementtiin perustuvalla mittausjärjestelmällä pysytään rekisteröimään sivunlatauksen aloitusajankohta sekä aika, jolloin sivulta siirrytään toiselle sivulle. Tulokseksi saadaan sivunluku aika. Useimmiten sivunluku aika joudutaan määrittelemään peräkkäisten sivupyynnöjen välisenä aikana. Riippuen kävijän verkossa liikkumisen tavasta tämä antaa yleensä käyttökelpoisen arvion sivunluku-ajasta. Ongelmia tuottaa kuitenkin mm. viimeiseksi pyydetyn sivun lukuajan määrittäminen, koska tietoa tämän sivun lukemisen lopettamisesta ei saada. Myös jos kävijä poistuu joksikin ajaksi verkkopalvelusta muille sivustoille ja palaa myöhemmin takaisin, voi poissaolo aika tulla lasketuksi mukaan sivunluku aikaan. Seurauksena siis voivat olla liian suuret sivunlukuajat.

Viimeisen sivun lukuajan määrittämiseksi on ehdotettu useita käytäntöjä, kuten: viimeiselle sivulle ei lasketa lainkaan luku aikkaa, viimeisen sivun luku aika on 5 minuuttia, viimeisen sivun luku ajaksi määritellään muiden sivujen keskimääräinen luku aika. Koska yhtenäistä käytäntöä ei ole, on aina selvitettävä kuinka mahdolliset sivunlukuajat on laskettu ja arvioitava kuinka luotettavia luvut voivat olla.

## **Verkkopalvelu**

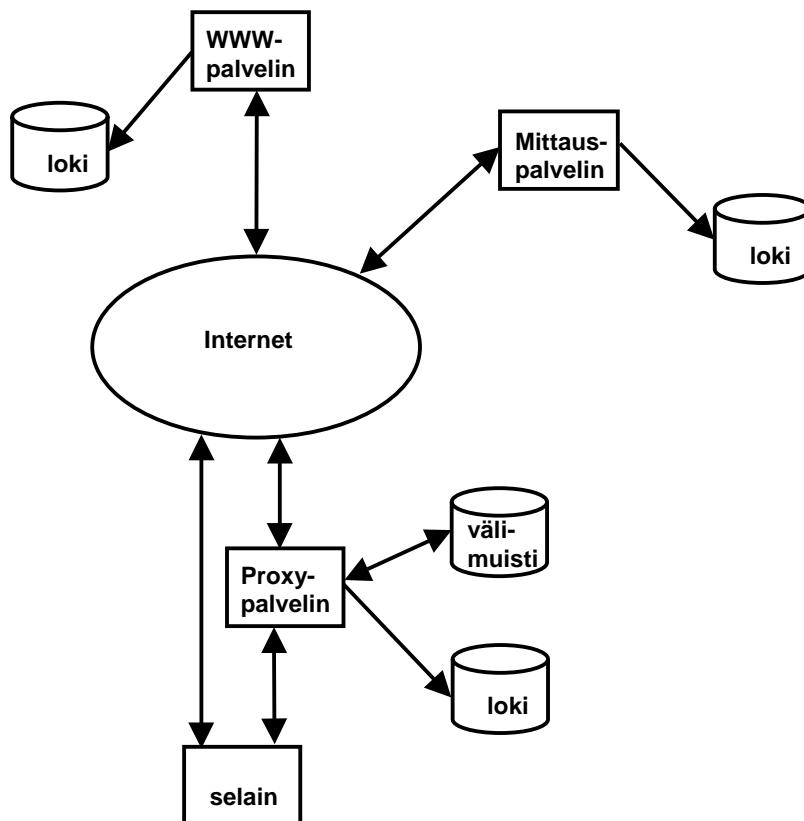
Verkkopalvelu eli WWW-sivusto on koostuu WWW-sivuista.

## **Vierailu**

Vierailu on sivulatauspyynnöjen sarja, jonka käyttäjä tekee edellyttäen, että peräkkäisten sivupyynnöjen väli on alle 30 minuuttia.

## Liite 1. Mittausjärjestelmät

Verkkopalvelun mittaamiseen on tarjolla useita vaihtoehtoja. Mittaaminen ja analysointi voidaan tehdä joko itse ja/tai se voidaan antaa kolmannen osapuolen tehtäväksi. Jos mittaaminen annetaan kolmannen osapuolen hoidettavaksi, kriittiseksi tekijäksi voi muodostua mittausjärjestelmän suorituskyky, jonka pitäisi riittää myös silloin kun kävijämäärät ovat huipussaan. Tällä hetkellä (syksy 1999) Suomessa käytössä olevat kolmansien osapuolien tarjoamat mittausjärjestelmät perustuvat pieneen mittauselementtiin ja siihen sidottuun cookie-tietoon (kts. tämän liitteen luku 2). Markkinoille on tulossa Java-pohjaiseen aktiiviseen mittauselementtiin perustuvia järjestelmiä, joissa mitataan käyttäjän selaimella tekemiä toimenpiteitä. Tämä menetelmä vaatii, että käyttäjä hyväksyy selainohjelmassa Javan käytön (luku 3). Mittaustulokset saadaan analysoimalla joko WWW-palvelimen tai mittauspalvelimen lokiin syntyneitä tietoja.



*Kuva 1. Internet-ympäristö, jossa mittausjärjestelmät toimivat. Lokianalysointiohjelmistoilla analysoidaan WWW-palvelimen lokiin kertyviä tietoja. Ulkopuoliseen mittauspalvelimeen perustuvassa mittausjärjestelmässä mittaustiedot saadaan sivulle sijoitettavan mittauselementin ja siihen sidotun cookie-tiedon avulla. Aktiiviseen mittauselementtiin perustuvissa mittausjärjestelmissä selaimelta lähetetään käyttäjän toimenpiteistä kertovia tietoja mittauspalvelimelle.*

### 1. Lokianalysointiohjelmit

Perinteisin vaihtoehto on analysoida verkkopalvelun palvelimelle syntyneitä lokitietoja. Tällaisella lokianalysointijärjestelmällä muutetaan lokiin kertynyt tieto havainnolliseen muotoon, minkä jälkeen lokitietoja voidaan analysoida verkkopalvelun kehittämiseksi. Lokianalysointi on kuitenkin altis proxy-palvelimille. Sivupyynnöiden määrä riippuu siitä, kuinka mitattavan verkkopalvelun sivujen ja sivun osien määrittelytiedot asetetaan. Näiden tietojen asetuksesta riippuu mm. se, ladataanko sivu tai sivun osa käyttäjän selaimelle mahdollisesti käytössä olevalta proxilta vai suoraan WWW-palvelusta. Riippuen

verkkopalvelusta pelkän lokianalysointiohjelman käyttö ei aina riitä, vaan ilmoittajia varten tarvitaan mittaustietoja myös riippumattomalta kolmannelta osapuolelta.

## **2. Ulkopuoliseen mittauspalvelimeen perustuvat järjestelmät**

Ulkopuoliseen mittauspalvelimeen perustuvissa mittausjärjestelmissä mitattavan palvelun WWW-sivuille lisätään mittauselementtejä. Nämä mittauselementit ovat yleensä pieniä, noin 1 x 1 pikselin kokoisia kuvia, jotka ladataan sivulle mittauspalvelua tarjoavan yrityksen mittauspalvelimelta. Näihin mittauselementteihin sidotaan käyttäjän selainohjelman yksilöivä cookie ja HTTP-header, joka aiheuttaa mittauselementtien vanhenemisen välimuistista (cache), jolloin ne ladataan periaatteessa aina uudelleen. Näillä järjestelmillä mittaus epäonnistuu, jos käyttäjä on kytkenyt selaimestaan kuvien lataamisen pois päältä.

## **3. Aktiiviseen mittauselementtiin perustuvat järjestelmät**

Selainohjelmassa aktiiviseen mittauselementtiin perustuvissa mittausjärjestelmissä mitataan käyttäjän selainohjelmassa suorittamia toimintoja, joiden tiedot lähetetään mittauspalvelimelle. Näissä järjestelmissä WWW-sivuille lisätään aktiivinen elementti, ohjelmakoodi, joka voi olla esimerkiksi Java-appletti tai myös Active-X-kontrolli. Kun tällainen sivu ladataan käyttäjän selaimen ikkunaan, ohjelmakoodi aktivoituu ja lähettää tiedon käyttäjän tekemästä sivupyynnöstä mittauspalvelimelle. Tämä tekniikka antaa mahdollisuuden myös sivunlukuajkojen mittaamiseen rekisteröimällä aika, jonka Java-appletti oli aktiivinen. Tällaiset mittausjärjestelmät eivät kuitenkaan toimi, jos Javan käyttäminen selainohjelmassa on estetty.

## Liite 2. Luettelo hakukoneista ja henkilökohtaisista roboteista

(<http://www.abcinteractiveaudits.com/guide/robots.html>)

### Hakukoneet (tilanne 10.7.1999)

activeagent	infoseek robot	oti_robot
aesir	infoseek Sidewin	oti_spider
alkalinebot	infoseek sidewinder	paperboy
altavista	infoseek_robot	permansurfer
anzwerscrawl	ipswitch_whatsup	ping plus
arachnet	iron33	rbse
architextspider	jumpstation	savvysearch
autonomy	keynote	scooter
backrub	link.alert	scoutget
bbn-webbot	linkbot	slurp
checktime	lycos - birdlite	stuff
compass	lycos monitoring robot	testur1
contentadvisor	lycos spider	tecor robi
crawler@alexa.com	lycos_spider	urlcheck
customcrawl	mailcast	urlck
cyber 411	metabustador netlink	user@domain
cyber411	momspider	verity
cybermapper	ncn-ultraseek	vspider
emailcollector	netrecorder	webcheck
emailsiphon	ninsunspider	webcrawler
excitespider	noasurf	webindex
ezresult	open text index	wget
headend	open text site crawler	wisewire
hotwired	opentext/otwr	worm
ia_archive	opentextsitecrawler	ync

### Henkilökohtaiset hakukoneet (tilanne 10.7.1999)

address rover	cherrypicker	eurospider
adobepepspider	churl	eventconsole
agentname	coast webmaster	excalibur internet spider
ahoy	cognagent	extractorpro
alphaconnect	coldfusion	e-zinedatabaseadd
ananzi	conceptbot	ezinesdatabase
anarchiecrawler	crescent	fetch
anderson consulting	cs-hkust wise	ffagenttechnolog
answerchase	cxwebot	fgc wwwgrab
ant fresco	cyberdog	fido
arach	cyberjack	fischer
arachnoidea	cyberpilotpro	fish
arachnophilia	cyberspyder	fish-search-robot
aretha	datachannel channelmanager	flashnavi
aspider	dbf.spider	flashsite
aturesys	desktop news	fletcher
autoftp	detective	fly
auto-proxy downloader	devsoft	folio_retriever
autoseek	deweb(c) katalog/index	folio4
avantgo	digmarc cgireader	freeloader
backweb	digout4u	fulcrumfind webscanner
big brother	dnfocus_seeder	funnelweb
bigwave libgreg	download	furet
bizbot	downtown	gais robot
black widow	dvorak page retrieval	get
blackwidow	dvorak websearch	get_http_file
bordermanager	ebsco url checker	getbot
bpftp	ecatch	geturl
btg's webtool	echo	gizmo
bullseye	echoping	gnw-web-stat
buzz	ei*net	go-ahead-got-it
cactvs chemistry spide	eirgrabber	gofetch
californiabrownsider	eit link verifier	Googlebot
canon-webrecd	eit-link-verifier-robot	googlebot
cartographer	eleetftp	go!zilla
castanet tuner	emacs-w3	grab-a-site
charlotte	emailwolf	gulliver
checkbot	emc web hopper	hamahakki
checkurl	enigma	happybot

harvest  
harvest cache  
hcat  
headdump  
heraspider  
hgrepurl  
hi (html index)  
hku www octopus  
hotjava  
hp openview/network node  
manager  
hp web printsmart  
htdig  
html\_analyzer  
htmlgobble  
httpaccessor  
httpdown  
http-library  
hyper-decontextualizer  
ibeam  
ibm\_network\_dispatcher  
ibm\_planetwide  
ibm-spidey  
ibot  
ifox98  
incywincy  
infogist  
infolink  
inforavirobot  
Inforia Quest  
Inforia+Quest+97  
Inforian Quest  
Inforian+Quest  
inforover  
infospiders  
infostack  
infotray  
insite  
intelligent  
internet angel  
internet ninja  
investor ticker  
inwebstigator  
ipbot  
ispi  
israeli-search  
iti spider  
jangoweb libgreg  
javelink  
jazzsoft  
jcrawler  
jobot  
joebot  
justpop3  
justview  
kilroy  
kit-fireball  
klink  
libertech-rover  
likse  
linecker  
linkalarm  
linklint  
linklooker  
linkscan  
linksweeper  
linkwalker  
lmcospider  
surfsaver  
surfwatchspider  
symantec notify internet session  
tarantula  
tarspider  
tcl w3

lmtasspider  
localeyes sputnik  
mac wwwworm  
mag-net  
magnifi  
magpie  
major submit  
mapuccino  
mastercard bytewatcher  
mata hari  
matchlogicaautosubmitter  
mcbot  
memoweb  
mercator  
metacrawler  
microsoft url control  
microsoft\_site\_analyst  
microsoft\_webmapper  
minebot  
mirago  
mister pix  
mobot  
monitor  
monster  
morning paper  
motor  
moviefinder-crawler  
moxilla  
Mozilla 2.0  
mozilla/ispi  
ms frontpage  
msie 4.0 crawler  
msiecrawler  
muscat  
mybot  
mycnnspider  
myownrobot  
naughtyrobot  
navroad  
ncn-sweeper  
nearsite  
net\_vampire  
netattach  
metbrief-light  
netcarta\_cyberpilot  
netcarta\_Webmapper  
netcaster  
netdelivery  
netjet  
netmechanic  
netmind-minder  
netpim  
netscape-catalog  
newsspace  
nhse web forager  
nicerspro  
nomad  
northstar  
nutscrape  
nwi robot  
nzexplorer  
offline explorer  
ogspider  
okware\_robot  
ontop  
openfind  
openurl  
orbitalwebindexer  
teleport  
templeton  
the informant  
t-h-u-n-d-e-r-s-t-o-n-e  
titan  
tkwww

pack rat  
pagesentry  
PageWatch  
parasite  
patric  
peregrinator  
perf-trak  
perman surfer  
pfsubmitbot  
pgp key agent  
phantom  
philips-websearch  
pioneer  
pita  
plumtreewebaccessor  
poacher  
pointcast internet fetch agent  
powerloader  
powermapper  
powermarks  
powernet  
processor  
python  
raleighrobot  
raveler  
redalert  
resume robot  
robot francoroute  
robotagent  
roverbot  
safetynet  
samslapper  
sandbot  
scanner  
scour.net crawler  
secret agent  
senrigan  
sg-scout  
shai'hulud  
simmany  
simonespider  
simulbrowse  
site server  
sitemap  
sitesnagger  
sitesweeper  
sitetech rover  
sitetech-rover  
skycache  
slowbot  
snoop\_spidey\_spide  
spanner  
spider  
spike  
sproxy  
spry wizard  
spyder3  
squid cache  
squirrel  
starsweeper  
statfetcher  
stress  
summit site validator  
summycrawler  
superbot  
supersnooper  
surfbot  
surfgopher  
tooter  
top10contactbot  
transceive  
travel-finder  
trawler  
turboexplorer

ucsd crawl  
ultraseek  
unmozify  
useit  
vci webviewer  
vdkwebi  
ventura  
viking  
vision-search  
vizion  
vnet  
vwbot  
w3ccommandline  
w3clinemode  
w3crobot  
w3index  
w3m2  
w3mir  
w3mirror  
w3new  
wakano\_robot  
wanderer  
web browser intelligence  
web core  
web downloader  
web moose  
web sucker  
web wombat  
webauto  
webbandit  
webbanger  
webclip  
webcompass

webcopier session  
webcopy  
webex authenticator  
webferret  
webfetch  
webfetcher  
webfoot  
webhound  
webinator  
webjammer  
webjuicer  
webking  
weblayers  
weblicator  
weblinker  
webmapper  
webmechanic  
webmirror  
webnerd  
webpest  
webpluck  
webreaper  
webretriever  
websaver  
webscanner  
webseer  
websentry  
webshift searchserver  
webshuttle  
websnake  
websnarf  
webspector  
webspider

webtamer  
webtrends  
webtrends link analyzer  
webtwain  
webvac  
webvcr  
webwalk  
webwalker  
webwalker4indexing  
webwatch  
webwhacker  
webwilly  
webzip  
west wind web monitor  
whostalking  
whowhere  
whowhere\_robot  
wild ferret web hoppe  
winddance webchallenger  
windows ce crawler  
wire news robot  
wisebot  
wobot  
wping  
wurlwind  
www collector  
www.pl  
wwwexplorer  
wwwoffle  
xenu's link sleuth  
yahoo-fetch  
yeehaw robot