TAMPEREEN TEKNILLINEN YLIOPISTO

**JARI-PEKKA SULONEN**

# KESKITETTY LAAJAKAISTAKÄYTTÄJIEN TELETUNNISTETIETOJEN KERÄYS- JA HAKUJÄRJESTELMÄ

Diplomityö

Aihe ja tarkastaja hyväksytty
Tieto- ja sähkötekniikan tiedekuntaneuvoston
kokouksessa 8.12.2010

Tarkastaja:
Prof. Jarmo Harju (TTY)

# TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Signaalinkäsittelyn ja tietoliikennetekniikan koulutusohjelma
**SULONEN, JARI-PEKKA**: Keskitetty laajakaistakäyttäjien teletunnistetietojen keräys- ja hakujärjestelmä
Diplomityö, 44 sivua, 5 liitesivua
Joulukuu 2010
Pääaine: Tietoliikenneverkot ja -protokollat
Tarkastaja: prof. Jarmo Harju
Avainsanat: laajakaistaliittymä, keräysjärjestelmä, teletunnistetieto

Euroopan Unionin direktiivi ja Suomen lainsäädäntö määräävät tietoliikennepalvelujen tuottajat keräämään ja tallentamaan tunnistetietoja käyttäjistään. Perusteluna on tietojen tarpeellisuus rikosten tutkimuksissa sekä niiden ennaltaehkäisyssä. Tietojen keräämiseen ei kuulu viestinnän sisältö tai edes sen paljastava tieto, vaan ainoastaan viestinnän osapuolten tunnistamistiedot. Kerättyjen tietojen säilytysaika on Suomen lainsäädännössä rajattu yhteen vuoteen, jonka jälkeen kerätyt tiedot tulee tuhota tai muuttaa niin, ettei siitä enää voida tunnistaa yksilöä.

Tietojen keräämisen mahdollistavat tietoliikennetekniikat ovat olleet palveluntarjoajien käytössä jo pitkään. Direktiivin vaatimat kerättävät tiedot ovat kuitenkin tallentuneet tekstitiedostoihin, joista suuren datamäärän vuoksi on ollut aikaa vievää etsiä tiettyjä tietoja viranomaisten niitä pyytäessä. Lisäksi direktiivin määräämää tietojen käyttötilastointia ja tietojen käsittelyn läpinäkyvyyttä ei ole ollut helposti toteutettavissa.

Tässä työssä käydään läpi laajakaistaisen Internet-liittymän teletunnistetietojen keräämisessä mukana olevat tekniikat ja tietoliikenneprotokollat, kuten ADSL sekä IP- ja DHCP-protokollat. Näihin tekniikoihin perustuen luodaan tunnistetietojen keräysjärjestelmä, jonka tarkoituksena on nopeuttaa tietojen etsintää. Lisäksi järjestelmä mahdollistaa tunnistetietojen hakujen ja käytön seurannan, jotta järjestelmää käyttävä teleyhtiö voi antaa niistä raportin viranomaisille vuosittain. Järjestelmän pohjana käytetään tietokantaa, jonne tunnistetiedot tallennetaan, ja käyttöliittymänä on nettiselainpohjainen sovellus. Lisäksi tunnistetietojen tallennuksen tietokantaan hoitaa erillinen ohjelma.

# ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Signal Processing and Telecommunications
**SULONEN, JARI-PEKKA**: Centralized system for collecting and searching an identification information of broadband subscribers.
Master of Science Thesis, 44 pages, 5 Appendix pages
December 2010
Major: Communication networks and protocols
Examiner: Prof. Jarmo Harju
Keywords: broadband Internet access, collecting system, identification

Directive of the European Parliament and Finnish legislation orders that all telecommunication service providers should collect and save the subscriber identification data. This is justified by the necessity of such information in criminal investigation and crime prevention. The data to be collected must not contain any information about the content of communication. Only the data which is used to identify the user should be collected. The collected data should be retained for one year according to the Finnish legislation and after that period the data should be destroyed or modified so that no individual can be identified from it.

All the telecommunication techniques, which make this collecting possible, have been in telecommunication service providers' use already for a while. However, the data that Directive demands have been saved into text files. Because of the large amount of data, searching and fetching the needed information has been quite time consuming when legal authorities have demanded them. In addition, the reporting and logging the data usage has been complex or it has not been available at all.

This Master's Thesis will go through all the necessary telecommunication protocols and techniques which are part of the identification data collecting. Among these are ADSL technique and IP and DHCP protocols. The centralized system for collecting and searching the identification data will be created based on these protocols and techniques. The system will speed up the data search and it also makes the reporting possible. Telecommunication service providers can use the reporting ability to fulfill legal needs to give reports to the authorities yearly. The system is based on a database system where all the collected identification data will be stored. System's usage interface is web based, so search and reports can be made with the web browser. Besides these, the system has a separate application for collecting and saving the data to the database.

# ALKUSANAT

Diplomityöni on tehty Xetpoint Oy:n palveluksessa ja Pohjois-Hämeen Puhelinyhtiön (PHP Oy) toimeksiannosta. Työn aiheen idea oli lähtöisin puhelinyhtiöltä ja heidän tarpeistaan. Työnantajani kanssa aiheesta jalostettiin lopullinen muotonsa.

Esitän kiitokseni työn loppuun saattamisesta työn tarkastajalle professori Jarmo Harjulle, jonka avustavien kommenttien avulla työ on lopulta valmistunut. Erityiset kiitokset ansaitsee myös työnantajani, joka on antanut diplomityön tekemiseen loistavasti aikaa ja mahdollisuuksia muiden työtehtävien ohella. Kiitokset myös PHP Oy:lle, jonka kautta saatu tieto työn aihepiiristä on ollut erittäin olennainen.

Tampereella 9.12.2010

_____

Jari-Pekka Sulonen

# SISÄLTÖ

# TERMIT, KÄSITTEET JA LYHENTEET

| | |
|---|---|
| ADSL | Asynchronic Digital Subscriber Line. Asynkroninen digitaalinen tilaajalinja laajakaistaliittymien toteuttamiseen. |
| ARP | Address Resolution Protocol. Protokolla fyysisen osoitteen (MAC) selvittämiseen IP-osoitteen perusteella. |
| ATM | Asynchronous Transfer Mode. Fyysisen siirtotien tekniikka. |
| CSV | Comma-separated Values. Taulukkomuotoisen tiedon listausmuoto, jossa jokainen tietokenttä erotellaan toisistaan tietyllä erotinmerkillä. |
| DHCP | Dynamic Host Configuration Protocol. Protokolla automaattiseen IP-osoitteiden jakoon verkossa oleville käyttäjille. |
| DSLAM | Digital Subscriber Line Access Multiplexer. DSL-keskitin, joka kokoaa tietyn määrän asiakasyhteyksiä ja lähettää ne eteenpäin yhdistettyinä (multipleksattuna) yhteen runkolinjaan. |
| Ethernet | Fyysisen siirtotien protokolla, joka siirtää ylempien kerroksien dataa saman verkon laitteiden välillä. |
| HSPA | High Speed Packet Access. Matkapuhelinviestintäprotokollien kokoelma, joka parantaa tiedonsiirtonopeutta kolmannen sukupolven matkaviestinverkoissa. |
| HTML | Hypertext Markup Language. Verkkosivujen rakenteen ja sisällön kuvaava merkintäkieli. |
| HTTP | Hypertext Transfer Protocol. Verkkoselainten ja -palvelinten välinen tiedonsiirtoprotokolla. |
| IMEI | International Mobile Equipment Identity. Matkaviestimen laitetunnus, jolla laite voidaan tunnistaa matkaviestinverkosta. |
| IMSI | International Mobile Subscriber Indentity. Enintään 15-merkkinen numerosarja, jolla yksilöidään jokainen matkapuhelinverkon käyttäjä. |
| IP | Internet Protocol. Internet-liikenteessä käytössä oleva protokolla laitteiden väliseen viestintään. |
| IPv4 | IP-protokollan versio 4, joka määrittelee laitteille 32-bittiset osoitteet. |
| IPv6 | IP-protokollan uudempi versio 6, joka määrittelee laitteille 128-bittiset osoitteet |

| | |
|---|---|
| LTE | Long Term Evolution. Yhdistelmäkäsite neljännen sukupolven matkaviestinverkoille. |
| MAC | Media Access Control. Fyysisen verkkolaitteen yksikäsitteinen tunniste. Osoitteen avulla pystytään välittämään paketit oikeiden laitteiden välillä fyysisessä aliverkossa. |
| NAT | Network Address Translation. Osoitteenmuunnostekniikka, jolla sisäiset yksityiset IP-osoitteet piilotetaan yhden tai useamman julkisen osoitteen taakse. |
| oktetti | kahdeksan bitin muodostama lohko |
| PPP | Point-To-Point-Protocol. |
| PDF | Portable Document Format. Adoben kehittämä tiedostomuoto, jota käytetään sähköisten dokumenttien julkaisuun, tulostamiseen ja painamiseen samanlaisena käyttöjärjestelmästä riippumatta. |
| TCP | Transmission Control Protocol. TCP/IP-protokollapinon kuljetuskerroksen yhteydellinen protokolla, jonka avulla muodostetaan päästä päähän -yhteydet. |

# 1. JOHDANTO

Tietoverkkoja hyväksikäyttävän rikollisuuden määrä kasvaa sitä mukaa, kun tietoverkkojen käyttö laajenee globaalilla tasolla. Rikollisuuden ehkäisyn, torjunnan, selvittämisen ja tutkinnan tarpeisiin on Euroopan Parlamentti ja Euroopan Unionin Neuvosto säätänyt maaliskuussa 2006 direktiivin [1] sähköisten viestintäpalvelujen ja yleisten viestintäverkkojen yhteydessä tuotettavien tai käsiteltävien tietojen säilyttämisestä.

Direktiivin lähtöajatuksena on ollut jäsenmaiden toisistaan eroavat lainsäädännöt tietojen säilyttämistä koskien. Poikkeavuuksia on ollut tietojen säilyttämisen edellytyksissä, tietojen säilytysajoissa sekä itse säilytettävissä tiedoissa ja niiden kattavuudessa. Neuvosto on katsonut, että viestintäpalvelujen yhteydessä saatavien tietojen säilyttäminen on osoittautunut hyvin tarpeelliseksi sekä kansallisten että kansainvälisten rikoksien tutkinnassa sekä terrorismin ja järjestäytyneen rikollisuuden ehkäisyssä ja selvittelyssä. Muun muassa näillä perusteilla tulisi koko Euroopan Unionin talousalueella olla yhtenäiset lait ja säädökset viestintäpalvelujen tietojen säilyttämisestä, luovuttamisesta ja käytöstä.

Suomen lainsäädäntö toteuttaa Unionin direktiivin ja sen kautta velvoittaa tietoliikennepalveluja tuottavat yritykset keräämään tunnistetietoja käyttäjistään. Tunnistetietojen keräämisen mahdollistavat tekniikat ja protokollat erityisesti laajakaista- ja puhelinpalveluissa ovat olleet osa palveluiden toteutusta jo ennen lain ja direktiivin voimaan tuloa. Todellisuudessa nämä tiedot ovat kuitenkin pääosin hautautuneet tietojärjestelmien syövereihin ja niitä on käytetty vain soveltuvin osin liiketoiminnan tarpeisiin, kuten laskutukseen ja asiakashallintaan sekä asiakkaan tilaamien palvelujen määrittelyyn tietoliikenneverkon tasolla. Lain ja direktiivin velvoittama toiminta vaatii palveluntarjoajilta toimintatapoja, määrityksiä ja tietoja koostavia tekniikoita, jotta määrätyt tiedot saataisiin kustannustehokkaasti, nopeasti ja helposti toimitettua viranomaisille, silloin kun niitä tarvitaan. Nykyisistä järjestelmistä ei ole tarpeen vaatiessa saatu kaivettua tunnistetietoja ilman viivästyksiä ja kohtuutonta ihmistyöpanosta.

Työn tarkoituksena on luoda laajakaistaisen Internet-palvelun tuottajien tarpeisiin automatisoitu järjestelmä, joka hoitaa määrättyjen tunnistetietojen keräämisen ja tallentamisen sekä hyödyntää olemassa olevia tekniikoita ja järjestelmiä. Järjestelmän tulee ottaa huomioon kaikki palvelun tuottamisen yhteydessä kerättäviksi määritellyt tiedot sekä tarpeellinen tietoturva ja järjestelmän käytön läpinäkyvyys. Läpinäkyvyydellä tarkoitetaan mahdollisuutta koostaa myöhemmin raportteja ja tilastoja järjestelmän keräämien tunnistetietojen käytöstä. Keräysjärjestelmän edellyttämät tekniikat ovat operaattorien käytössä jo pääosin vakiintuneet, joten suhteellisen yleiskäyttöisen systeemin rakentaminen on mahdollista.

Tämän diplomityön luvussa 2 esitetään sekä Euroopan Unionin että Suomen eduskunnan lainsäädännön kautta tulevat määräykset ja ohjeistukset tunnistetietojen keräämisestä, käsittelystä ja tallentamisesta. Luvussa 3 kuvataan kuluttajatason laajakaistaliittymä ja siihen yleisimmin kuuluvat komponentit. Lisäksi esitellään tekniikat, joilla mahdollistetaan tunnistetietojen kerääminen lainsäädännön vaatimusten mukaisesti. Järjestelmän varsinainen toteutus ja toimintaperiaatteiden läpikäynti käsitellään luvussa 4. Luku 5 kokoaa yhteen lopullisen järjestelmän sekä sen vaatimukset laitteistoille. Samalla tuodaan esille myös erilaisia kokoonpanovaihtoehtoja järjestelmän toimintaan saattamiseksi. Luvussa 6 esitetään yhteenveto diplomityöstä ja toteutetusta järjestelmästä sekä mietitään mahdollisia jatkokehityskohteita.

# 2. LÄHTÖKOHTA

Euroopan Unionin direktiivi [1] tunnistetietojen säilyttämisestä kohdentaa määräykset ja säädökset vain sellaisiin tietoihin, joita palvelujen tarjoajat itse tuottavat tai käsittelevät, kuten esimerkiksi asiakkaan tunnistamiseen käytettävät tiedot. Direktiivin nojalla ei voida kerätä, tallentaa ja säilyttää viestinnän sisältöön liittyviä tietoja. Kerättävien tietojen säilyttämisteknologiaa direktiivi ei määrittele, vaan se tulee jokaisen jäsenmaan ratkaista omalla kansallisella tasollaan.

## 2.1. Direktiivin määrittelemät kerättävät tiedot

Direktiivin viidennessä artiklassa määritellään sen nojalla säilytettävät tiedot. Tiedot on jaoteltu kuuteen luokkaan sen mukaan mihin liittyvää tietoa kustakin luokasta on saatavilla. Luokilla ei ole mitään prioriteettia toistensa suhteen, vaan luokittelu on tehty vain kerättävien tietojen jaotteluun. Luokkien sisällä jaotellaan tiedot erikseen puhelinpalveluihin ja Internet-palveluihin liittyviksi. Puhelinpalveluilla tarkoitetaan kiinteitä lankapuhelinverkkoja sekä matkaviestinverkkoja ja kaikkia niihin liittyviä palveluita kuten äänipuheluja, neuvottelupuheluita, soitonsiirtoja sekä erilaisia viestipalveluja mukaan lukien tekstiviestit ja multimediaviestit). Internet-palveluihin kuuluvat datayhteys (esimerkiksi mobiilidata, laajakaista ja kaapelimodeemiyhteys), Internet-puhelut sekä sähköpostit.

Ensimmäinen luokka käsittää viestinnän lähteen tunnistetiedot, joihin kuuluu puhelinpalveluiden osalta puhelinnumero, josta on soitettu sekä kyseisen palvelun (puhelinliittymän) tilaajan tiedot. Internet-palveluiden osalta kerättäviin tietoihin kuuluu ensimmäisen luokan osalta yhteyteen tai lähettävään sähköpostiosoitteeseen liittyvä käyttäjätunnus tai -tunnukset sekä tiedot yhteyden tilaajasta, jonka nimiin viestinnän lähteenä käytetty yhteys oli rekisteröity yhteydenottohetkellä.

Toiseen luokkaan on koottu viestinnän kohteen tunnistamiseen ja jäljittämiseen tarkoitettuja tietoja. Puhelinpalveluiden tietoihin kuuluu puhelinnumero, johon on soitettu tai mikäli kohteen tapauksessa on käytetty jotain lisäpalvelua, kuten soitonsiirtoa, niin siirron kohteen numero tai numerot. Lisäksi kerättäväksi ja säilytettäväksi määritellään tämän kohteen liittymän tilaajan henkilötiedot (nimet ja osoitteet). Internet-palveluissa viestinnän kohde tulee esille pääasiassa vain sähköpostin ja Internet-puhelun tapauksessa ja näistä tuleekin tallentaa vastaanottajan puhelinnumero tai käyttäjätunnus sekä kohteen palvelun tilaajan tiedot.

Viestinnän ajankohdan määrittävät tiedot on luokiteltu direktiivin kolmanteen luokkaan ja ne käsittävät puhelun tai Internet-yhteyden alkamis- ja loppumisajat päivämäärä ja kellonaikatasolla. Internet-palvelun tapauksessa viestinnän ajankohtien määritystä tarkennetaan vielä paikallisella aikavyöhykkeellä sekä viestinnän tilaajalle käyttöön kyseisellä ajanhetkellä myönnetyllä Internet-osoitteella.

Neljännen luokan tallennettavat tiedot koskevat viestintätyyppiä. Puhelinpalveluissa kyseessä on käytetty puhelinpalvelu ja sähköpostien ja Internet-puhelun tapauksessa käytetty Internet-palvelu. Viidenteen luokkaan on jaoteltu käyttäjien viestintälaitteen tai oletetun viestintälaitteen tunnistamiseksi tarvittavat tiedot. Kiinteissä puhelinverkoissa tämä tarkoittaa lähettäjän ja vastaanottajan puhelinnumeroita. Matkaviestinverkoissa puhelintoiminnan osalta tiedot käsittävät puhelinnumeroiden lisäksi soittajan ja vastaanottajan matkaviestintilaajan tunnukset (IMSI-tunnus) sekä matkaviestinten tunnukset (IMEI-koodi). Lisäksi direktiivi huomioi, että mikäli kyseessä on anonyymina hankittu ennalta maksettu palvelu, esimerkiksi PrePaid-liittymä, niin tallennettavia tietoja ovat palvelun ensimmäisen aktivoinnin päivämäärä ja kellonaika sekä palvelun aktivoinnin sijaintitunniste. Matkapuhelinverkoissa sijaintitunniste on matkapuhelinverkon solun tunniste, minkä alueella viestintälaite on ollut palvelun aktivointihetkenä. Internet-palvelun laitteen tunnistamiseksi tulee tallentaa modeemiyhteyden tapauksessa puhelinnumero, josta yhteys on avattu ja digitaalisen laajakaistaliittymän tapauksessa liittymän loppupiste.

Viimeiseen kuudenteen luokkaan on eroteltu matkaviestintälaitteen sijainnin määrittämiseksi tarvittavat tiedot, joihin kuuluu matkaviestinverkon sijaintitunniste sekä tiedot kyseisen sijainnin (solun) maantieteellisestä sijainnista sillä hetkellä, kun yhteys on aloitettu.

## 2.2. Tietojen säilytys ja tilastointi

Direktiivissä annetaan kerättävien tietojen säilytysajalle aikahaarukka, jonka sisään jäsenvaltiot voivat itse määritellä säilytysajan. Direktiivin kuudennen artiklan mukaan tietoja tulee säilyttää vähintään kuuden kuukauden ja enintään kahden vuoden ajan viestinnän alkamisen päivämäärästä.

Säilytettäville tiedoille on direktiivissä myös määritelty vähimmäistietoturvaperiaatteet. Tämä tarkoittaa sitä, että säilytettävien tietojen pitää olla laadultaan samanlaisia ja samassa tietoturvatasossa, kuin samaiset tiedot ovat verkossa, josta niitä kerätään ja tallennetaan. Lisäksi tallennetuille tiedoille on tehtävä asianmukaiset tekniset ja hallinnolliset toimet, jotta ne voidaan suojata tahattomalta ja tahalliselta tuhoamiselta, muuttamiselta ja edelleen siirtämiseltä. Tähän liittyen tietoihin on teknisesti määritettävä pääsy vain erikoisvaltuudet omaaville henkilöille. Säilyttämisen lisäksi kerätyt tiedot tulee tuhota säilytysajan lopussa lukuun ottamatta tietoja, joita on käytetty ja tallennettu jonkin tutkinnan yhteydessä.

Tietojen varastoinnista säädetään lisäksi, että ne tulee tallentaa sellaisella tavalla, josta ne voidaan toimittaa pyynnöstä viranomaisille. Tämän toimituksen tulee tapahtua kohtuullisessa ajassa ilman tarpeetonta viivästystä.

Direktiivi määrää jäsenvaltioiden toimittamaan vuosittain tilastot tietojen säilyttämisestä Euroopan Komissiolle. Näiden tilastojen pitää sisältää

- tapaukset, joissa viranomaiset ovat tietoja pyytäneet,
- aika, joka on kussakin tapauksessa kulunut tietojen tallennuspäivästä siihen päivään, kun viranomainen kyseisiä tietoja pyysi sekä
- tapaukset, joissa viranomaisten tietopyyntöä ei voitu täyttää.

## 2.3. Suomen lainsäädäntö

Suomen eduskunta käsitteli Euroopan Unionin direktiivin (ks. luku 2.1.) vaatimukset vuoden 2007 lopussa ja 2008 alussa. Käsittelyn perusteella tuotettiin laki sähköisen viestinnän tietosuojalain muuttamisesta [2], joka tuli voimaan 1.6.2008. Lain määräämä tietojen keräämis- ja säilytysvelvollisuus alkoi palveluyritysten osalta kuitenkin vasta 15.3.2009. Muutoslaissa olennaisimmat muutokset sähköisen viestinnän tietosuojalakiin tehtiin pykälään 14.

Sähköisen viestinnän tietosuojalain [3] kahdeksannessa pykälässä otetaan huomioon direktiivin mukaisesti se, että tietojen kerääminen, käsittely ja säilyttäminen on sallittu ainoastaan asiaan kuuluvassa laajuudessa. Käsittely ei saa rajoittaa viestien sisällön luottamuksellisuutta tai henkilöiden yksityisyyden suojaa yhtään enempää kuin on tietojen keräämisen kannalta välttämätöntä. Lisäksi sama kahdeksas pykälä määrää, että kerättyjä tietoja saa luovuttaa vain viranomaisille heidän pyynnöstään ja että tunnistetiedot on määritellyn ajan kuluttua tuhottava tai tehtävä sellaisiksi, ettei niitä voida enää yksilöidä käyttäjätietoihin.

Tietosuojalaissa [3] määritellään direktiivin määräämien tietojen maksimisäilytysajaksi 12 kuukautta viestinnän päivämäärästä. Alkuperäisellä päivämäärällä tarkoitetaan esimerkiksi ajankohtaa, jolloin puhelu on aloitettu tai hetkeä, jolloin laajakaistayhteyden tilaajalle on osoitettu tietty Internet-osoite. Säilytysvelvollisuus rajataan koskemaan vain viestien tunnistetietoja eikä lain piiriin kuulu viestin sisällön tai verkkosivujen selaamisesta kertyneiden tunnistetietojen tallentaminen.

Säilyttämisen tekniikasta tietosuojalain 14b pykälässä todetaan, että "tietojen säilytyksen teknisen toteutuksen on oltava kustannustehokasta ja siinä on otettava huomioon palveluyrityksen liiketoiminnan tarpeet, järjestelmien tekniset ominaispiirteet ja säilyttämisestä aiheutuvat kustannukset maksavan viranomaisen tarpeet". [3] Tämän perusteella annetaan tietojen keräysvelvolliselle palveluyritykselle mahdollisuus itse toteuttaa tietojen kerääminen niillä resursseilla, joita sillä on olemassa. Säilytysvelvollisuus ei koske tietoja, joita yritys ei itse käsittele palvelun toteuttamisen yhteydessä, vaikka ne muuten olisivatkin tietosuojalain piiriin kuuluvia tietoja.

Palveluyrityksen vastuulla on tietojen keräämisen ja säilyttämisen lisäksi tietoturva. Tietoturvan vahvuutta ja sen teknisiä toteutustapoja ei laissa ole määritelty. Yrityksen on kuitenkin direktiivin pohjalta ylläpidettävä kerätyille tiedoille yhtä hyvää tietoturvaa, kuin tietoja käytettäessäkin on ollut. Lisäksi sen on määriteltävä organisaatiossaan henkilöt, joilla on oikeus käsitellä säilytettäviä tietoja.

## 2.4. Lainsäädännön ja direktiivin aiheuttamat järjestelmävaatimukset

Tietojen tallennusjärjestelmään tulee olla pääsy vain erikseen määrätyillä käyttäjillä. Tästä syystä järjestelmässä pitää olla käyttäjäautentikointi. Kaikki järjestelmään kirjautuneen käyttäjän tekemät toimet tulee tallentaa, jotta voidaan seurata kuka tietoja on milloinkin käsitellyt. Kerättävien tietojen määrä on arvioitava erikseen kullekin kerättävälle tietotyypille, jotta pystytään mitoittamaan tietojen tallennuskapasiteetin tarve. Lisäksi tietojen tallennustavassa on otettava huomioon tietojen mahdollisimman tehokas ja nopea haku. Myös haut, hakukriteerit ja niillä saatavat tulokset tulee tallentaa, jotta yritys voi vuosittain toimittaa tiedot järjestelmän käyttötapauksista valtiolle.

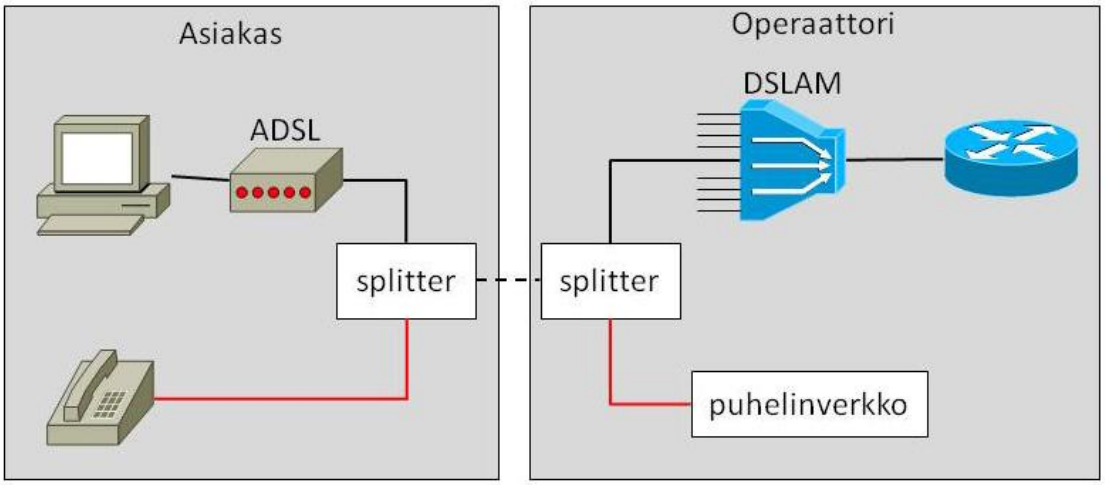
# 3. TEORIA

Teoriaosuudessa käydään läpi laajakaistaliittymän perustoteutus. Lisäksi käsitellään liittymän toteutukseen ja IP-osoitteistukseen liittyvät protokollat sekä käyttäjän identifiointiin liittyvät mekanismit laajakaistapalvelujen tapauksessa.

## 3.1. Kuluttajatason laajakaistaliittymän rakenne

Kuluttajatason laajakaistaliittymät toteutetaan joko langattomina käyttäen mobiiliverkkoja tai langallisina käyttäen perinteisiä puhelinverkkoja. Osa liittymistä voidaan toteuttaa talokohtaisina valokuituliittyminä, jolloin perinteistä puhelinverkon osaa ei ole mukana lainkaan. Mobiiliverkkoja käytettäessä verkkotekniikkana on tällä hetkellä 3G ja lisänä mahdollisesti myös sen laajennos HSPA (High Speed Packet Access). Neljännen sukupolven LTE (Long Term Evolution) -verkot ovat jo tulossa, mutta niitä ei vielä ole saatavilla ainakaan tuotantokäytössä.

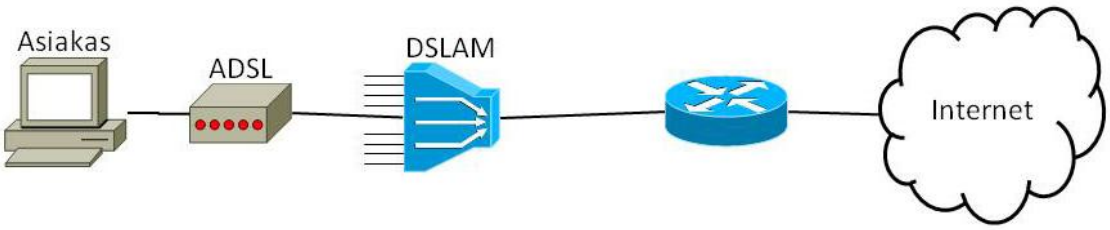
Perinteistä puhelinverkkoa käytettäessä verkon data siirretään perinteisiä kuparisia parikaapeleita pitkin. Tällöin tekniikkana on yleisimmin ADSL eli asynkroninen digitaalinen tilaajalinja. ADSL-tekniikalla toteutettujen laajakaistaliittymien teoreettiset maksiminopeudet vaihtelevat sen mukaan kuinka pitkä kuparilinja on asiakkaalta puhelinyhtiön keskukseen. Mitä pidempi välimatka ilman vahvistimia tai toistimia on kyseessä sitä hitaampi on teoreettinen maksiminopeus. Todellinen käytettävissä oleva nopeus on silti aina pienempi kuin teoreettinen nopeus johtuen useista eri seikoista, kuten esimerkiksi keskuksen ruuhkaisuudesta. ADSL-tekniikkaa käytettäessä asiakkaan (tilaajan) käytössä on ADSL-modeemi, joka moduloi dataliikenteen kuparikaapeliin. Tiedonsiirron kanssa samaan aikaan on mahdollista käyttää perinteistä puhelinyhteyttä (Kuva 1). Data ja puhe erotellaan ennen operaattorin DSL-keskitintä eli DSLAMia (Digital Subscriber Line Access Multiplexer). Puhetaajuudet ohjataan automaattisesti puhelukeskukseen ja siitä vastaanottajalle puhelinverkkoa pitkin ja data välitetään operaattorin dataverkkoon ja sitä kautta kohteeseen.




**Kuva 1: Laajakaista- ja lankaliittymä yhdessä.**

Laajakaistaliittymien toteuttajina määritellään kaksi eri operaattoria. Toinen, verkko-operaattori, omistaa tilaajan asuma-alueen puhelinparikaapelit ja puhelinkeskukset. Palveluoperaattori taas tuottaa, markkinoi ja myy tilaajille tietoliikenneyhteydet. Verkko- ja palveluoperaattori voivat olla joko yksi ja sama yritys tai kaksi kilpailevaa yritystä, jolloin palveluoperaattori vuokraa verkko-operaattorilta asiakkaiden siirtolinjoja ja asiakkaan yhteys kuljetetaan muuttumattomana verkko-operaattorin siirtoverkon kautta palveluoperaattorin verkkoon.

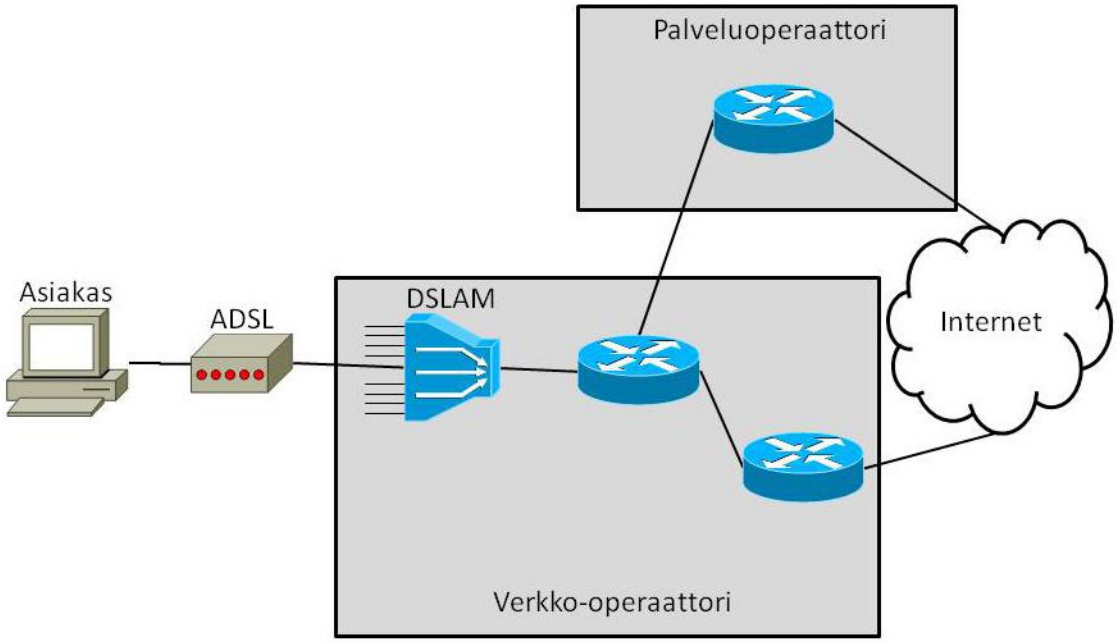
Yksinkertaisimmillaan asiakkaan laajakaistaliittymä on teknisesti toteutettu seuraavalla tavalla. Asiakkaan ADSL-modeemi kytkeytyy puhelinparikaapeleilla DSL-keskittimeen (Kuva 2).




**Kuva 2: Laajakaistaliittymän rakenne**

DSL-keskitin kanavoi usean tilaajan parikaapeliyhteydet yhdelle siirtojohdolle ja antaa kullekin asiakasyhteydelle yksikäsitteisen polku- ja kanava-arvon. Näillä arvoilla operaattorilla on aina tiedossa mistä puhelinliittymästä mikäkin asiakas liikennöi. DSL-keskittimen siirtotielle kanavoima data ohjautuu palveluoperaattorin verkkoon, josta operaattori reitittää liikenteen eteenpäin asiakkaan liittymän parametrien mukaisesti. Asiakkaan liittymän parametreihin kuuluu tyypillisesti ainakin liittymän maksiminopeus. Mikäli asiakkaan puhelinparikaapelit ja puhelinkeskukset omistava verkko-operaattori on eri kuin liittymän palvelun tarjoaja (Kuva 3), niin tällöin palvelu- ja verkko-operaattori ovat sopineet tietyt polku- ja kanava-arvot, joilla liikenne siirretään verkko-operaattorin verkosta palveluoperaattorin verkkoon. Edelleen palveluoperaattori

saa tietyillä arvoilla selville liikenteen tilaajayhteyden ja voi näin ollen ohjata datan eteenpäin tilaajan yhteysparametrien mukaisesti. Samanaikaisesti verkko-operaattori voi tarjota toki myös itse omaa palveluaan. Asiakkaan liikenne vain ohjataan DSLAMin jälkeen joko palveluoperaattorin verkkoon tai omaan verkkoon reititettäväksi Internetiin.




**Kuva 3: Laajakaistaliittymän rakenne, kun toteutuksessa on mukana kaksi eri operaattoria.**

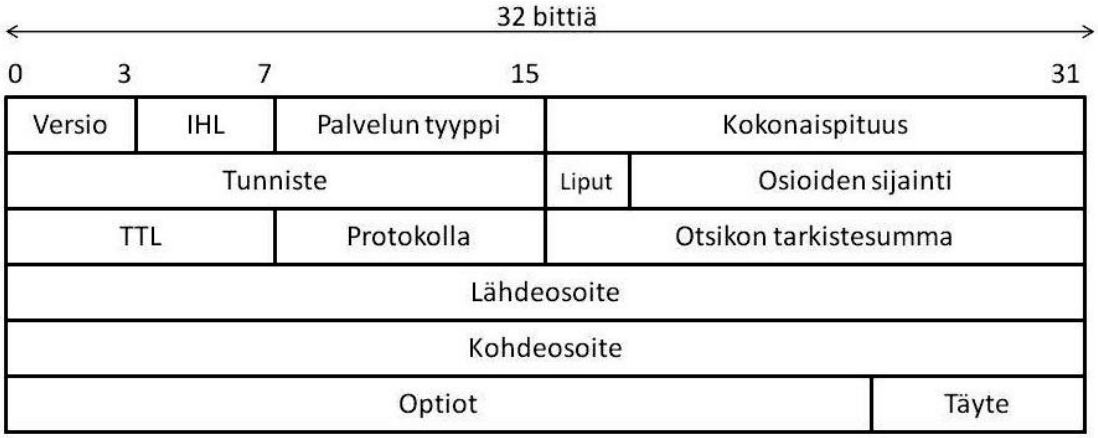
Yksityiselle laajakaistatilaajalle annetaan palveluntarjoajasta riippuen käyttöön yhdestä viiteen julkista Internet-osoitetta eli IP-osoitetta. Nämä osoitteet määritellään dynaamisesti jokaiselle käyttäjälle käyttäen DHCP-tekniikkaa. Kullekin asiakkaalle jaeltavat osoitteet otetaan tietystä operaattorin IP-osoiteavaruudesta. DHCP-palvelimeen voidaankin konfiguroida jaeltava osoiteavaruus sen mukaan, mitkä ovat asiakkaan yhteydelle määritellyt polku- ja kanava-arvot. Polku- ja kanava-arvot sekä asiakkaalle annettu osoite tallentuvat DHCP-palvelimen tietoihin. Näin voidaan myöhemmin yksilöidä kullakin ajanhetkellä kullakin asiakkaalla käytössä ollut IP-osoite.

## 3.2. Dynaaminen IP-osoitteiden jako

Kaikilla verkossa olevilla laitteilla on oltava yksikäsitteinen tunniste, jolla laite tunnistetaan ja jonka avulla liikennöinti verkon eri laitteiden välillä onnistuu. Yleisimmin käytössä on IP-osoitteistus ja siitä nykyisin vielä versio neljä (IPv4), vaikka tällä hetkellä sen sisältämien osoitemäärien rajat ovat tulossa vastaan ja siirtymä laajempaan osoiteavaruuteen ja versio kuutoseen (IPv6) on tulossa.

### 3.2.1. IP-protokolla

IP-protokolla on kehitetty välittämään datapaketteja verkosta toiseen. Ylemmältä protokollakerrokselta tuleva data ympäröidään IP-otsikkokentillä, jotka on esitetty kuvassa 4.




**Kuva 4: IP-protokollan otsikkokentät**

Versio-kentän neljä bittiä kertoo käytettävän IP-protokollan version. Tällä hetkellä kentän arvoksi tulee 4, mutta uutta IPv6-protokollaa käytettäessä kenttään asetetaan arvo 6.

IHL eli Internet Header Length kertoo otsikkokenttien kokonaispituuden mukaan lukien mahdolliset optiokentät ja täytteen. Arvo on nelibittinen ja kertoo 32-bittisten lohkojen kokonaismäärän eikä siis tavujen tai oktettien eli kahdeksan bitin lohkojen määrää. Ilman optioita eli minimissään kentän arvo on viisi 32-bitin lohkoa ja maksimissaan IP-otsikko voi siis käyttää 15 32-bitin lohkoa.

Palvelun tyypissä käytetään kahdeksaa bittiä. Sillä kerrotaan millaista palvelua kyseinen datapaketti verkolta haluaa. Tähän voi liittyä esimerkiksi minimiviive tai maksimikapasiteetti riippuen paketin käyttötarkoituksesta.

Kokonaispituus kertoo koko IP-paketin koon okteteissa. Paketin kokoon otetaan huomioon sekä otsikkokentät että data ja suurin paketin koko voi siis olla $2^{16}$-1 eli 65535 oktettia. Todellisuudessa tuo paketin koko on kuitenkin liian suuri laitteiden hallittavaksi ja standardi määrittelee, että IP-paketteja käsittelevän laitteen tulee kyetä vastaanottamaan paketteja, jotka ovat 576 oktettia pitkiä [4 s. 12]. Tuota suurempia paketteja suositellaan lähetettäväksi vain, kun ollaan varmoja vastaanottavan pään mahdollisuudesta vastaanottaa suurempia paketteja.

Tunnisteen 16 bittiä kertoo paketin yksilöllisen tunnisteen. Tätä tarvitaan siinä tapauksessa, että paketti on jossain vaiheessa matkan varrella osioitu. Vastaanottava laite pystyy tämän kentän avulla selvittämään mitkä osiot kuuluvat samaan pakettiin. Myös lippukentät ja osioiden sijainti -kenttä liittyvät IP-paketin osiointiin.

TTL-kenttä (Time-To-Live) eli paketin elinaika osoittaa kahdeksalla bitillä kuinka monen laitteen välillä pakettia maksimissaan kuljetetaan. Jokainen matkalla oleva reititin vähentää tämän kentän arvoa yhdellä ja kun arvo laskee nollaan, paketti tuhotaan.

Esimerkiksi Windows XP käyttää oletuksena arvoa 128 [5]. Eli sen mukaan mikä tahansa toinen piste Internetissä tulisi saavuttaa 128 aktiivilaitteen välisellä hypyllä.

Protokolla-kentän kahdeksan bittiä kertoo hyötykuorman eli IP-paketin dataosuuden protokollan numeron. Esimerkiksi, jos kuljetettavana on TCP-protokollan paketti, niin kenttään tulee arvo kuusi.

Tarkistussumman 16 bittiä käsittää koko IP-otsakkeen tarkistussumman. Sen avulla jokainen paketin vastaanottava laite voi tarkistaa, että otsikon sisältö on säilynyt muuttumattomana siirron ajan. Jokainen siirtotien varrella oleva laite tarkistaa ja laskee tämän arvon uudestaan. Näin pyritään IP-protokollan tasolla takaamaan paketin otsikoiden eheys.

32-bittiset lähde- ja kohdeosoitteet määrittävät tiedon siitä mistä paketti on tullut ja mihin paketti on menossa. Esimerkki: Asiakas lähettää datapaketin kohteeseen 74.125.39.105. Oletetaan että asiakkaalla on käytössään osoite 191.100.1.100. Paketti ei kuitenkaan siirry kohteeseen suoraan, koska päätepisteet ovat täysin eri IP-verkoissa ja useimmiten myös täysin eri fyysisessä sijainnissa. Paketti lähtee siis asiakkaalta hänen oletusyhdyskäytäväksi määritetylle reitittimelle, joka päättää oman reititystaulunsa perusteella mille reitittimelle paketti seuraavaksi tulee lähettää. Näin paketti etenee reitittimestä ja IP-aliverkosta toiseen, kunnes päästään kohteen sisältävän aliverkon laidalla olevalle reitittimelle. Siellä kyseinen reititin tietää, että paketin kohde onkin samassa aliverkossa, joten sitä ei ole enää tarpeellista lähettää eteenpäin muille reitittimille, vaan se voidaan toimittaa suoraan kohdekoneelle. Kohdekoneen mahdollinen vastauspaketti toimitetaan vastaavalla tavalla takaisin asiakkaalle. Tosin paketin kulkema reitti saattaa muuttua reitittimien ruuhkanhallinnan tai reittien katkeamisien takia, mutta sillä ei ole merkitystä kunhan reitittimellä on aina tiedossa seuraava laite, jolle paketin voi välittää.

IP-otsikon optiokentissä voidaan määritellä esimerkiksi turvallisuuteen ja pakettien reitittämiseen liittyviä juttuja. Mikäli optiokentät eivät ole kokonaisia 32 bitin lohkoja, niin otsikon perään lisätään täytebittejä, kunnes otsikkokenttien pituus on 32:lla jaollinen.

IP-osoitteistuksessa samaan aliverkkoon lasketaan kuuluvaksi laitteet, joiden verkko-osoite on sama. Verkko-osoitteen selvittämiseen käytetään 32-bittistä verkkomaskia, jonka avulla laitteen IP-osoitteesta pystytään laskemaan verkko-osoite. Laskenta tehdään bittitason operaatiolla AND IP-osoitteen ja verkkomaskin kesken. Bittitason AND antaa lopputulokseksi 1, kun molemmat sisääntulevat luvut ovat ykkösiä. Muuten lopputulos on 0. Koko operaation lopputulos kertoo aliverkon verkko-osoitteen.

Esimerkki: Laitteen IP-osoite on 192.168.17.125 ja sen verkkomaski on 255.255.240.0. Binäärisiksi muutettuina ja AND-operaatiolla laitteen verkko-osoitteeksi saadaan 192.168.16.0 (Kaava 1).

Kaava 1:

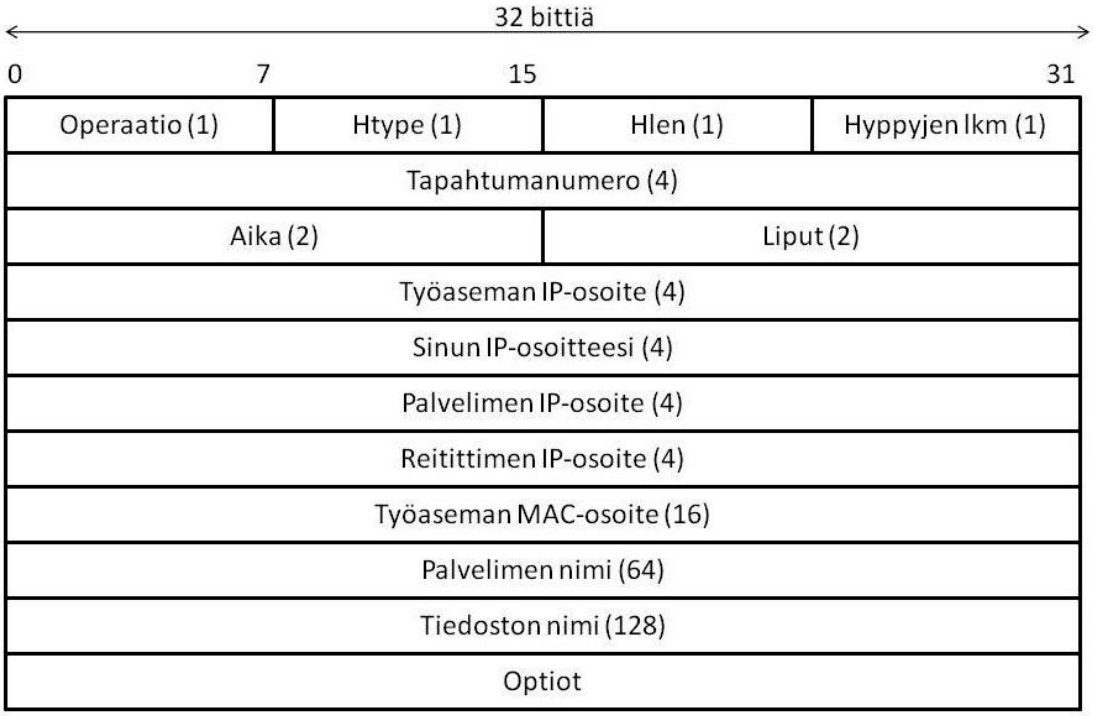
$$\begin{array}{r} 1100\ 0000\ 1010\ 1000\ 0001\ 0001\ 0111\ 1101 \\ \underline{\text{AND } 1111\ 1111\ 1111\ 1111\ 1111\ 0000\ 0000\ 0000} \\ 1100\ 0000\ 1010\ 1000\ 0001\ 0000\ 0000\ 0000 \end{array}$$

Tämän esimerkkilaitteen kanssa samaan aliverkkoon kuuluisivat siis laitteet, joiden osoitteet ovat välillä 192.168.16.1 - 192.168.31.254. Samaista osoitehaarukkaa kutsutaan myös nimellä osoiteavaruus ja se voidaan määritellä yksinkertaisesti verkko-osoitteen ja -maskin yhdistelmällä 192.168.16.0/20, jolloin maskia kuvaava numero kertoo bittitasolla olevien ykkösten määrän vasemmalta alkaen (Kaava 1).

### 3.2.2. **DHCP-protokolla**

Laitteiden IP-osoitteita voidaan hallita täysin manuaalisesti tai automaattisesti DHCP-protokollan avulla. DHCP-protokolla on kehitetty aiemmin käytössä olleen BOOTP-protokollan pohjalta. Automaattisen IP-osoitteiden konfiguroinnin idea on säilynyt samana eli laitteen kanssa samassa aliverkossa on palvelin, joka luovuttaa laitteille käyttöön tietyn IP-osoitteen tietyksi ajaksi. DHCP-palvelin antaa osoitteen ohella myös muita tarvittavia tietoja vastaanottavalle laitteelle. Näihin tietoihin kuuluvat normaalisti oletusyhdyskäytävän osoite (oletusreititin), verkkomaski, nimipalvelin tai nimipalvelimet sekä domain-nimi [6 s. 202].

Automaattinen DHCP-konfiguraatio alkaa laitteen lähettämällä joukkojakeluviestillä (broadcast), jonka kuulevat kaikki samassa IP-aliverkossa olevat DHCP-palvelimet. Kaikki kommunikaatio asiakkaan ja palvelimen välillä käyttää kuvan 5 mukaisia paketteja.




**Kuva 5: DHCP otsikkokentät (suluissa kentän pituus oktetteina)**

Operaatio-kentällä on kaksi määriteltyä arvoa. 1 tarkoittaa alustuspyyntöä ja 2 alustusvastausta. Työasema käyttää liikennöintiin operaatiota yksi ja palvelin vastauk-

sissaan tyyppiä 2. Erilliset protokollan toimintaan kuuluvat sanomat erotellaan sanomatyyppioption avulla.

Htype ja Hlen -kentät kertovat liikennöivän laitteen fyysisen osoitteen tyypin ja pituuden. Useimmiten fyysisenä siirtotienä on Ethernet, jolloin tyyppi saa arvon yksi ja pituus arvon kuusi (fyysisen Medium Access Control eli MAC-osoitteen pituus oktetteina). Hyppyjen lukumäärällä tarkoitetaan reitittimien tai muiden välittävien laitteiden määrää työaseman ja DHCP-palvelimen välillä. Hyppyjen lukumäärää kasvatetaan jokaisessa välillä olevassa verkon aktiivilaitteessa ja kenttä on lähinnä informatiivinen tieto palvelimelle.

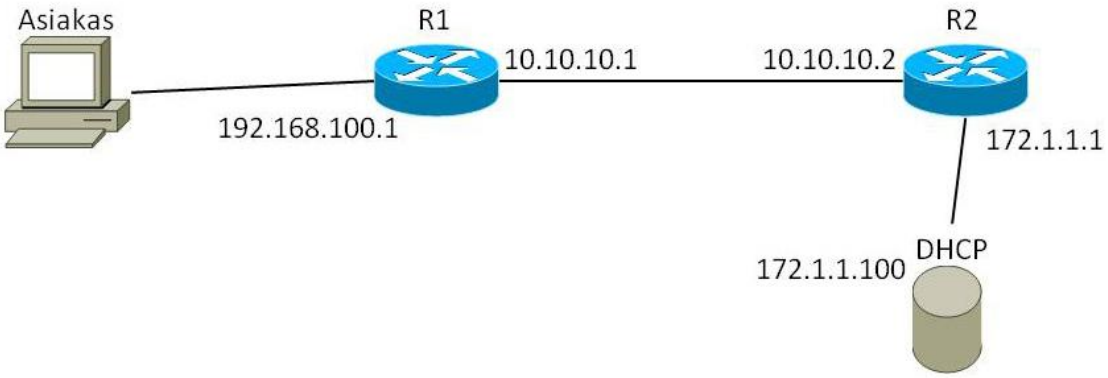
Tapahtumanumero identifioi jokaisen DHCP-transaktion. Tämän arvon avulla palvelin ja työasema tietävät mitkä paketit kuuluvat niiden väliseen DHCP-tapahtumaketjuun. Tapahtumanumero on oleellinen tieto, koska useimmat DHCP-viestit kulkevat verkossa joukkojakeluviesteinä ja samaan aikaan verkossa saattaa olla useita laitteita, jotka liikennöivät saman DHCP-palvelimen kanssa. Näin estetään se, että tietty osoite menisi väärälle laitteelle tai että kahdella tai useammalla laitteella olisi käytössä sama osoite samassa aliverkossa. Saman osoitteen käyttö samassa aliverkossa estäisi kaiken liikenteen, sekä saapuvan että lähtevän liikenteen, näiden kahden koneen osalta.

Aika-kenttään tallentuu aika IP-osoitteen hankinnan tai uusinnan aloittamisesta eli siitä kun työasema lähettää ensimmäisen DHCP-viestinsä. Aika tallennetaan sekunteina. Asiakkaan lähettämässä ensimmäisessä viestissä aikakentässä on arvona nolla.

Lipuilla työasema voi kertoa palvelimelle haluavansa vastaukset verkon joukkojakeluviestinä. Tämä saattaa olla ainoa mahdollinen tapa toimia silloin, kun laite ei voi vastaanottaa sen fyysiseen osoitteeseen kohdistettuja paketteja ennen kuin IP-konfiguraatio on suoritettu loppuun ja koska DHCP-viesteillä juuri tätä ollaan tekemässä, pitää paketit lähettää verkon levitysviesteinä.

IP-osoitekentillä määritellään palvelimen tarjoama osoite sekä myös mahdollinen työaseman haluama osoite. Työaseman IP-osoite -kentässä on arvo siinä tapauksessa että työasemalla on jo aiemmin ollut jokin IP-osoite, jonka vuokra-ajan se haluaa uudistaa. Mikäli työasemalla ei ole lainkaan IP-osoitetta, niin tämän kentän arvo on tietysti tyhjä eli toisin sanoen nolla. Sinun IP-osoitteesi -kentässä palvelin välittää työasemalle tarjoamansa osoitteen. Palvelimen osoite -kenttään palvelin tallentaa oman osoitteensa.

DHCP-pyynnön tekevää asiakasta lähinnä oleva reititin asettaa Reitittimen osoite -kentän. Siihen tulee reitittimen osoite siitä verkkoliitynnästä, jossa asiakas on kiinni. Kuvan 6 mukaisessa topologiassa reititin R1 asettaisi kentän arvoksi oman osoitteensa eli 192.168.100.1. Reititin R2 ei enää muuta kentän arvoa, vaan sen tehtävänä on ainoastaan välittää asiakkaan DHCP-paketit palvelimelle ja palvelimelta tulevat takaisin reitittimelle R1. Reitittimen osoitteen perusteella DHCP-palvelin voi tehdä päätöksiä työasemalle tarjottavasta IP-osoitteesta. Palvelimelle voi olla määriteltynä useita jaettavia IP-osoiteavaruuksia eri IP-aliverkoille ja sen pitää tietää mistä aliverkosta kukin pyyntö tulee.




**Kuva 6: Havainnekuva verkon rakenteesta DHCP-otsikkokenttiä täytettäessä.**

Työaseman MAC-osoite välitetään DHCP-viestissä sen takia, että useimmiten palvelin tallentaa omaan tauluunsa jaettujen IP-osoitteiden ja MAC-osoitteiden yhteyden ja tällöin palvelin voi helposti jakaa kullekin työasemalle sille aiemmin jaetun osoitteen, vaikka työasemalla ei mitään osoitetta olisikaan uusittavana. Työasemalle annetaan siis ensisijaisesti sama osoite, joka sillä on ollut aiemmin, mikäli sitä ei ole välillä luovutettu jonkun muun työaseman käyttöön.

Palvelimen nimi -kentässä voidaan välittää DHCP-palvelimen nimi. Tiedoston nimi -kentässä taas voidaan välittää BOOTP-tiedoston nimi tai polku tiedostoon. Nykyisin tuota kenttää ei kuitenkaan juurikaan käytetä, vaan IP-osoitteen lisänä tarvittavat parametrit, kuten esimerkiksi oletusreitittimen osoite ja käytettävät nimipalvelimet, välitetään DHCP -optioissa.

### *DHCP-protokollan sanomatyypit*

Kokonaisuudessaan DHCP-protokollassa on kahdeksan erilaista sanomatyyppiä, joita käytetään työaseman ja palvelimen välillä. Ensimmäinen viesti on DHCPDISCOVER ja siinä työasemalla ei ole mitään tietoa, missä on DHCP-palvelin. Tämä ensimmäinen viesti lähetetään koko verkkoon joukkojakeluviestinä, jotta verkossa olevat palvelimet osaisivat vastaanottaa ja reagoida pyyntöön. Toinen viesti sisältää palvelimen työasemalle tarjoaman IP-osoitteen ja se on tyyppiä DHCPOFFER. Mikäli työasema on otsikkokentän lipuilla ilmoittanut haluavansa viestit joukkojakeluviesteinä, niin tämäkin viesti lähtee verkkoon joukkojakeluviestinä. Muuten tämä viesti kohdistetaan MAC-osoitteella vain pyynnön lähettäneelle työasemalle.

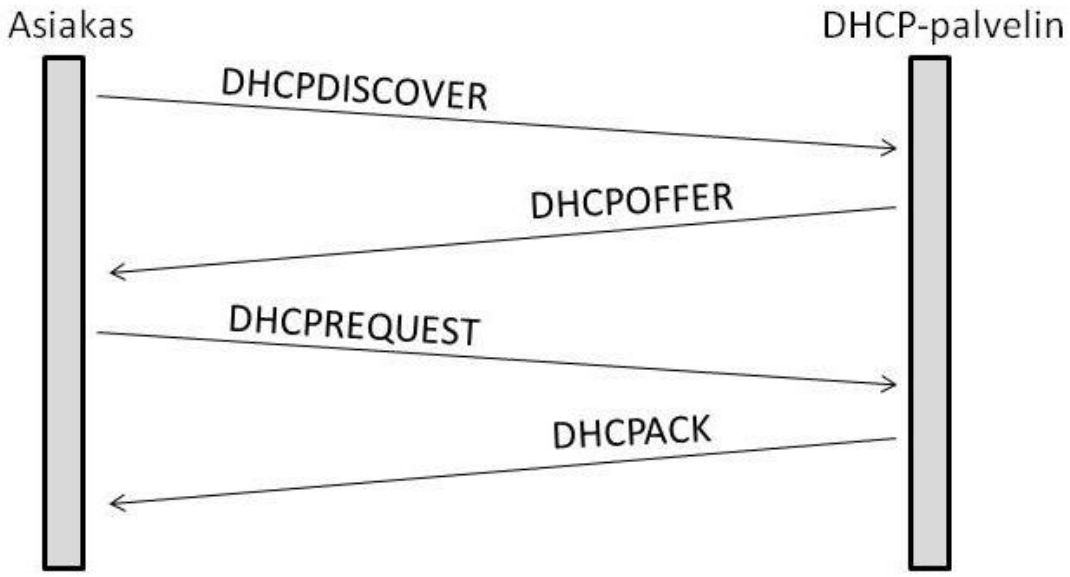
Kolmas viestityyppi on työaseman käyttöön ja siinä työasema pyytää palvelimelta tiettyä IP-osoitetta. Tätä viestiä käytetään sekä hyväksymään palvelimen tarjoama osoite että pyytämään jotain tiettyä IP-osoitetta palvelimelta esimerkiksi tilanteessa, jossa työasemalla on jo ollut jokin osoite ja se haluaisi jälleen saman osoitteen. Kolmas viesti on tyyppiä DHCPREQUEST. Neljäs ja viides viestityyppi ovat palvelimen käytössä ja niillä se voi joko kuitata halutun IP-osoitteen varauksen tai hylätä sen.

Lisäksi DHCP:ssä on RELEASE-, INFORM- ja DECLINE -viestit. RELEASE-viestillä työasema voi vapauttaa käyttämänsä osoitteen, jolloin palvelin voi jakaa ko.

osoitteen muille työasemille. INFORM-viestillä työasema pyytää IP-yhteyden lisäparametreja, kuten esimerkiksi nimipalvelimien osoitteita. DECLINE-viestillä työasema voi kertoa palvelimelle, ettei se halua käyttää sen tarjoamaa osoitetta. DECLINE-viestin jälkeen palvelin normaalitilanteessa tarjoaa työasemalle toista osoitetta.

### *DHCP-protokollan perustoiminta*

Automaattisen konfiguraation toiminnan aloittaa työasema lähettämällä DHCPDISCOVER-tyyppisen viestin levitysviestinä verkkoon (Kuva 7). Tähän verkossa sijaitsevat DHCP-palvelimet vastaavat DHCPOFFER-viesteillä, joissa siis palvelimet tarjoavat IP-osoitteita työasemalle. Työasema valitsee näistä yhden ja lähettää DHCPREQUEST-viestin, jolla se ilmaisee haluamansa osoitteen. Tässä samaisessa viestissä on myös osoitteen tarjonneen DHCP-palvelimen osoite, jotta muut verkossa mahdollisesti olevat palvelimet tietäisivät hylätä omat tarjouksensa ja vapauttaa ne tulevia pyyntöjä varten käyttöön. Tarjouksen tehnyt palvelin varaa halutun IP-osoitteen työaseman käyttöön, mikäli se edelleen oli vapaana ja kuittaa osoitteen varauksen DHCPACK-viestillä. Mikäli kuitenkin työaseman pyytämä osoite on jo varattu jollekin muulle laitteelle, niin palvelin vastaa DHCPNAK-viestillä, jolloin prosessia työasema joutuu valitsemaan toisen osoitteen saamistaan DHCPOFFER-viesteistä. DHCPACK-viestissä palvelin lähettää myös muut osoitteen konfigurointiin tarvittavat tiedot, kuten oletusreitittimen ja käytettävät nimipalvelimet. Lisäksi työasema voi osoitteen hyväksynnän jälkeen kysyä palvelimelta haluamiaan parametreja DHCPINFORM-viestillä.




**Kuva 7: DHCP-viestityyppien käyttö.**

Perustoiminta on kuvattu yksityiskohtaisina pakettitietoina pakettianalysaattorin kuvakaappauksina liitteessä 1. Ensin siis työasema lähettää levitysviestinä verkkoon DHCPDISCOVER-viestin. Tässä tapauksessa työasemalla on ollut aiemmin jokin IP-osoite ja se pyytää samaa osoitetta uudelleen käyttöönsä paketin optioissa (Liite 1, kuva

2, optio 50). Tämän jälkeen pakettianalysaattori näyttää palvelimen tekemän ARP-kyselyn (Address Resolution Protocol), jossa palvelin tarkistaa onko työaseman haluamaa IP-osoitetta verkossa käytössä (Liite 1, kuva 3). Tässä tapauksessa osoitetta ei ole käytössä, joten palvelin vastaa työasemalle tarjoavansa sille sen pyytämää osoitetta (Liite 1, kuva 4). Nyt palvelimen vastauksessa ei myöskään käytetä levitysviestiä, koska työasema ilmoitti ensimmäisessä viestissään pystyvänsä hyväksymään unicast-viestejä ilman IP-konfiguraatiota (Liite 1, kuva 2).

Kun työasema on vastaanottanut palvelimen tarjouksen, tekee se DHCPREQUEST-viestin, jossa pyydetään tarjottua IP-osoitetta käyttöön (Liite 1, kuva 5). Edelleen työasema lähettää viestin levitysviestinä, jotta kaikki verkossa olevat DHCP-palvelimet voisivat reagoida viestiin.

Tarjouksen tehnyt palvelin hyväksyy IP-osoitteen varauksen ja lähettää työasemalle loput IP-konfiguraatioon tarvittavat tiedot paketin optioissa (Liite 1, kuva 6). Tässä tapauksessa tietoihin kuului aliverkon maski, oletusreititin, nimipalvelimet, domain sekä tiedot osoitteen vuokra- ja uusinta-ajasta. Konfiguraation jälkeen työasema vielä mainostaa verkkoon ARP-viestillä oman IP-osoitteensa, jotta muut verkossa olevat laitteet voisivat sen lisätä ARP-tauluihinsa.

## 3.3. Käyttäjän identifiointi

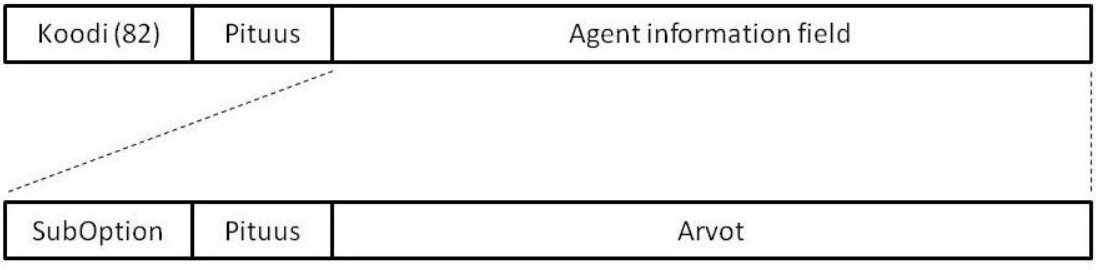
DHCP on perusominaisuuksiltaan kätevä mekanismi osoitteiden jakoon lähiverkossa oleville koneille. Yleisesti sitä käytetään esimerkiksi yrityksen laitteiden osoitteiden jakoon, jolloin kaikki osoitteen saavat koneet ovat yrityksen sisäverkossa ja sitä kautta myös luotettuja. Tällöin yrityksen ulospäin menevä liikenne kulkee yhden tai muutaman reitityspisteen kautta ja julkisia IP-osoitteita tarvitaan vain näille ulospäin näkyville liittymille. Tällöin kyseessä on yksittäisen julkisen osoitteen muuntaminen usean laitteen käyttöön eli NAT (Network Address Translation).

Nykyisin entistä useammat laitteet liikennöivät Internetissä. Näistä esimerkkeinä ovat kännykät ja muut kannettavat laitteet. Tämä johtaa entistä suurempaan paitsi julkisten IP-osoitteiden tarpeeseen myös joustavampaan IP-osoitteiden jakeluun verkko-operaattorien päästä katsottuna. Aiemmin tietokoneet pääsivät nettiin pääosin puhelinliittymään kytketyn modeemin avulla, jolloin IP-osoitteiden jakeluun käytettiin PPP-protokollaa. PPP-protokollan avulla voitiin toteuttaa käyttäjän autentikointi ja varmistaa ettei kukaan ulkopuolinen pääse verkon käyttäjäksi. Nykyisten laajakaista-, kaapeli- ja kiinteiden yhteyksien yleistyessä on kuitenkin jouduttu enemmän luopumaan PPP-protokollasta ja puhelinliittymien soittosarjoista. DHCP on vakiintunut PPP:n korvaajaksi IP-osoitteiden jakelussa laajakaistayhteyksille. DHCP-protokollan perusominaisuuksissa ei kuitenkaan ole asiakkaan tunnistusta muuten kuin MAC-osoitteen avulla, joka kuitenkin on melkoisen helppo vaihtaa toiseksi. Muun muassa tähän ongelmaan on DHCP-protokollaan kehitetty optio numero 82 [7].

DHCP-protokollan avulla viestiketju lähtee asiakkaan päästä lähetettävällä joukkojakeluviestillä. Joukkojakeluviesti näkyy kuitenkin vain samaan aliverkkoon kuulu-

ville laitteille eikä tähän normaalisti kuulu DHCP-palvelinta tai mitään muutakaan operaattorin hallitsemaa palvelinta. Reitittimeen voisi toki yhdistää DHCP-palvelimen toiminnallisuuden, mutta yleisimmin DHCP-palvelin on kuitenkin erillisenä laitteenaan operaattorin verkossa. Joukkojakeluviestiä ei näin ollen käsittelisi mikään verkon laite, koska lähiverkon reunareititinkään ei reititä joukkojakeluviestejä eteenpäin. Tästä syystä asiakasta lähinnä olevan reitittimen tulee osata relay agent -toiminta. Relay agent -toiminnassa reititin tunnistaa asiakkaalta tulleen DHCP-paketin ja reitittää sen kohti DHCP-palvelinta. Samalla se lisää pakettiin option numero 82. DHCP-palvelimelta vastauksena tulevista paketeista tämä samainen reititin poistaa optio 82 -optiokentän ennen paketin välittämistä asiakkaalle. Optio 82 sisältää informaatiota asiakkaan sijainnista ja se on konfiguroitavissa sisältämään esimerkiksi DSLAMin antaman virtuaalipiiri-arvon, reitittimen verkkoliitynnän numeron, kytkimen portin tai vaikka kaapelimodeemin tunnisteen.

Optio 82 lisätään DHCP-paketin optio-kenttään viimeiseksi, mutta kuitenkin ennen mahdollista optio 255 -kenttää, joka on DHCP-sanoman loppukenttä. Optio 82:n muoto on esitetty kuvassa 8. Option koodi sijoitetaan alkuun ja heti sen jälkeen tulee option datan pituus oktetteina eli kuvassa Agent information Field -kentän pituus. Agent information field taas sisältää vähintään yhden alioption, jolla on samalla tavalla optiokoodi, arvokentän pituus oktetteina sekä arvokenttä. Koska optio 82:lla tulee olla vähintään yksi alioptio tulee sen pituudeksi aina vähintään kaksi (alioption koodi- ja pituuskenttä). Alioptiokoodeja on kaksi, joista numero 1 tarkoittaa piirin tunnistetta (Circuit-ID) ja 2 yhteyslaitteen luotettua tunnistetta (Remote-ID). Piirin tunnisteella tarkoitetaan ATM-tekniikkaan (Asynchronous Transfer Mode) perustuvassa kytkennässä sen virtuaalipiirin tai virtuaalikanavan tunnistetta, mistä DHCP-sanoma on tullut. Ethernet-tekniikassa Circuit-ID:nä käytetään yleisimmin kytkimen porttia, johon asiakkaan linja on kytketty. Circuit-ID on aina yksikäsitteinen yhdelle relay agent -reitittimelle riippumatta fyysisen siirtotien tekniikasta. Yhteyslaitteen tunnisteena voi olla asiakkaan tunnistamistietoja, kuten esimerkiksi käyttäjänimi tai puhelinnumero, tai laitteen tietoja, kuten esimerkiksi kaapelimodeemin laitetunniste.

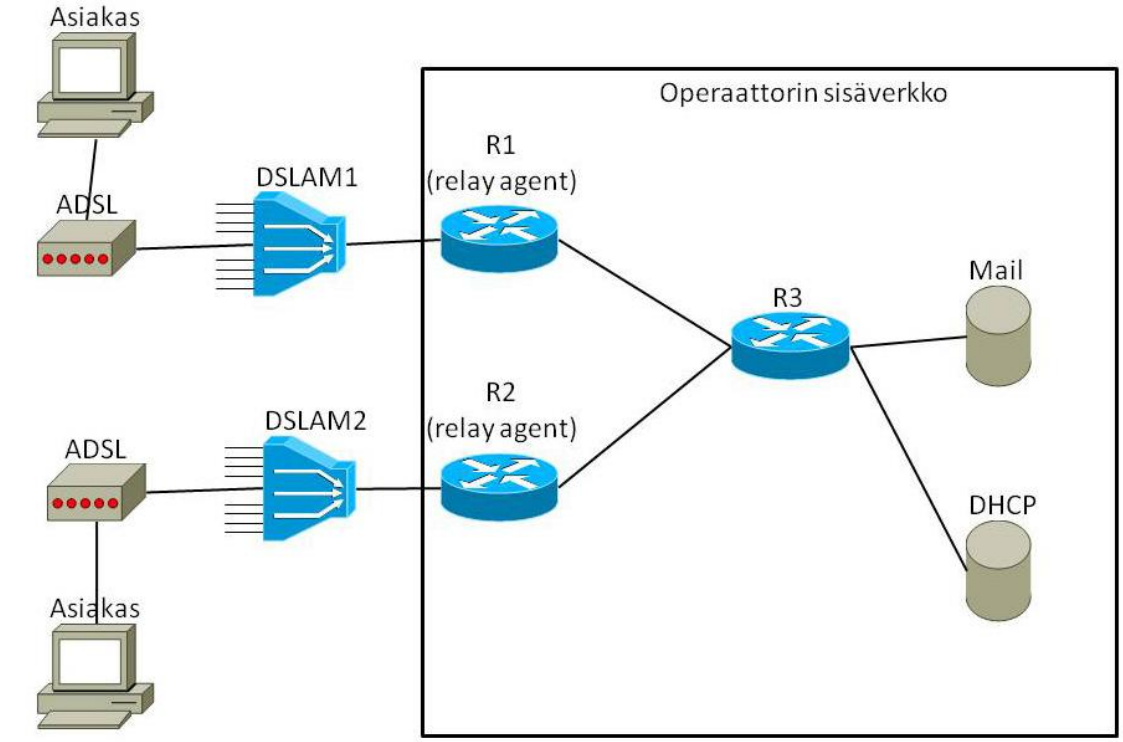



**Kuva 8: DHCP optio82:n kentät**

Reitittimessä, joka toteuttaa relay agent -toiminnan, tulee olla konfiguroituna tieto siitä milloin DHCP-pakettiin liitetään optio 82 ja mitkä ovat silloin alioptioihin tallennettavat tiedot. Relay agent -reititin tietää lisätä optio 82 -kentän, kun se vastaanottaa DHCP-paketin, jossa reitittimen IP-osoite -kenttä on asetettu nollaksi. Se tarkoittaa, että kyseinen reititin on ensimmäinen hyppy asiakkaan lähiverkosta eteenpäin. Lisäksi

vastaanotetussa DHCP-paketissa ei saa vielä olla optio 82 -kenttää. Mikäli tuo kenttä on kyseisessä paketissa, niin reititin poistaa sen. Tällä estetään se, että joku pyrkisi peittämään sijaintinsa lisäämällä lähtevään DHCP-pakettiin valmiiksi option 82. Mikäli reititin olettaisi, että optio on oikein ja välittäisi paketin eteenpäin, saisi asiakas mahdollisesti virheellisen IP-osoitteen virheellisten optio 82 -tietojen perusteella. Vaikka osoite olisikin toimiva ja oikea kyseiseen aliverkkoon, niin joka tapauksessa DHCP-palvelin ei enää tietäisi varmasti mille asiakkaalle kyseinen osoite on vuokrattu ja olisi mahdollista, että palvelimen lokitietoihin olisi tallentunut väärä asiakas.

Relay agent -toiminnassa on myös laillisesti mahdollista että DHCP-paketin reitittimen IP-osoite on nolla ja silti optio 82 on jo asetettu. Tällainen mahdollisuus on esimerkiksi ADSL-yhteyksissä, joissa DSLAMiin on määritetty optio 82:n lisääminen, mutta tällöin DSLAM ei kuitenkaan ole ensimmäinen reititin (Kuva 9). Tällöin reitittimen ja DSLAMin välillä on ns. luotettu verkko, jolloin reititin ei poista optio82-kenttää, vaan lisää ainoastaan DHCP-otsikkokenttään oman IP-osoitteensa reitittimen IP-osoite -kenttään ja välittää paketin eteenpäin.




**Kuva 9: Relay agent -verkkomalli.**

Relay agent -reitittimiä voi olla myös useamman relay agent -toimintaa toteuttavan tai toteuttamattoman reitittimen takana (esimerkiksi R3 kuvassa 9). Kun tällainen kauempana oleva reititin vastaanottaa DHCP-paketin, jossa jo on reitittimen IP-osoite ja mahdollisesti myös optio 82 -kenttä, se ei lisää pakettiin omia optio 82 -kenttiään eikä omaa IP-osoitettaan, vaan lähettää paketin eteenpäin muuttumattomana. Mikäli siis asiakasta lähinnä oleva reititin ei toteuta optio 82 -toimintaa, niin sitä ei toteuteta kauempanakaan asiakkaasta, vaikka siellä olevalla reitittimellä olisikin mahdollisuus kyseiseen toiminnallisuuteen. Relay agent -toiminnan poikkeustapaus tulee silloin, kun reitit-

timen vastaanottaman paketin DHCP-otsikossa on jo sen oma IP-osoite. Tällöin voidaan olettaa, että asiakas yrittää ohittaa tunnistetietojen lisäämisen ja piilottaa sijaintinsa, jolloin reitittimen tulee tuhota paketti välittämättä sitä eteenpäin. Kuvan 9 mukaisessa tapauksessa siis asiakas yrittäisi asettaa DHCP-otsikkokenttään reitittimen osoitteeksi reitittimen R1 IP-osoitteen, jotta kyseinen reititin ei lisäisi optio 82 -tietoa pakettiin (DSLAM ei tässä tapauksessa lisää optio 82 -kenttää). Tällöin reitittimen R1 tulee tuhota DHCP-paketti, koska asiakas yritti estää tunnistetietojen lisäämisen ja piilottaa sijaintinsa.

RFC3046 standardin [7] mukaan DHCP-palvelin voi olla konfiguroitu hyväksymään ja tallentamaan optio82 -kentän mukaiset tiedot. Tällöin sen pitää tallentaa lokitietoihinsa jokaisen luovutetun IP-osoitteen yhteyteen optio82 -kentän tiedot. Lisäksi DHCP-paketteihin, jotka palvelin lähettää asiakkaalle, tulee kopioida optio82 -kenttä muuttumattomana, jotta option lisännyt relay agent -reititin osaa poistaa sen DHCP-vastauksesta ennen sen välittämistä asiakkaalle. Mikäli DHCP-palvelin ei tue optio82 -kenttää, se jättää sen huomiotta eikä kopioi sitä omiin DHCP-vastauspaketteihin.

Optio 82:n alioptio Circuit-ID sisältää standardin mukaisesti informaatiota siitä verkon asiakkaasta, jolta DHCP-pyyntö on tullut. Circuit-ID alioptio onkin käytössä laajakaistayhteydessä pääosin puhelinverkon päätepisteessä eli DSLAMissa ja heti sitä seuraavassa reitittimessä, mikäli ne eivät ole samaa laitetta. Circuit-ID:n perusteella tiedetään tarkasti mistä reitittimestä ja mistä DSLAMin portista pyyntö on tehty. Remote-ID:n lisäävät laitteet, jotka pystyvät yksilöimään asiakkaan laitteen. Tällöin kyseessä on suoraan pakettikytkentäinen tai kiinteä tilaajaverkko, kuten esimerkiksi kaapelimodeemiyhteys. Remote-ID eroaa Circuit-ID:stä sillä, että sen tulee olla yksikäsitteinen kaikkialla, kun taas Circuit-ID on yksikäsitteinen vain relay agent -reitittimen yhteydessä. Sama Circuit-ID voi siis tulla DHCP-palvelimelle kahdesta eri relay agent -reitittimestä ja tällöin asiakkaan tunnistamiseksi olennaista on ID:n lisäksi reitittimen IP-osoite. Molemmat alioptiot voivat olla käytössä samanaikaisesti tai vain toinen niistä. DHCP-palvelimen tulee kuitenkin joka tapauksessa tallentaa molemmat läsnäolevat optiot IP-osoitteen vuokratapahtuman yhteyteen.

# 4. TOTEUTUS

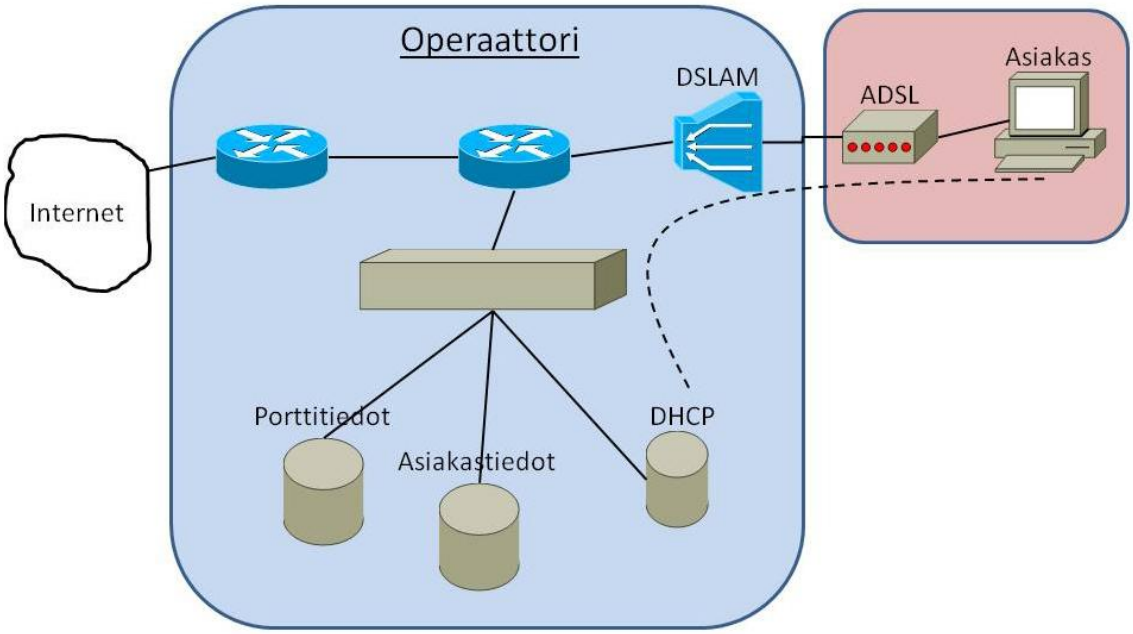
Toteutukseen otetaan huomioon sekä järjestelmän suunnittelu että mahdolliseen lopulliseen toteutukseen tulevat työn kannalta olennaiset yksityiskohdat. Järjestelmän ulkoasu ja käyttäjän lopulliset näkymät eivät kuulu tämän työn piiriin, vaan ne toteutetaan räätälöitäviksi. Järjestelmän käyttötapausten esittelyssä tosin käydään läpi havainnollistavia kuvia käyttöliittymän toiminnasta.

## 4.1. Operaattorin nykyinen verkkotopologia ja -komponentit

Operaattori hallitsee omia DSLAM-laitteitaan, joihin asiakkaiden ADSL-laitteet kytkeytyvät. Asiakkaan tiedot, kuten laskutustiedot ja liittymätyyppi, ovat omassa asiakastietojärjestelmässään. Lisäksi asiakkaan käytössä olevien porttien tiedot löytyvät toisesta järjestelmästä. Portti- ja asiakastietokanta jakavat tietoa keskenään ja niiden kautta voidaan jokainen asiakas yksilöidä tietyn DSLAMin tiettyyn porttiin.

DSLAMin (tai DSLAMista seuraavan reitittimen, mikäli DSLAM itse ei toimi reitittimenä) lisäämät asiakkaan tunnistamistiedot kulkeutuvat DHCP-kyselyn yhteydessä asiakkaalta DHCP-palvelimelle. Palvelimelle voi olla konfiguroitu useita asiakkaille jaeltavia IP-osoiteavaruuksia. Näiden jakoehtoihin voidaan liittää DSLAMista tulevia porttitietoja, jolloin kullekin asiakkaalle voidaan tarjota osoitetta oikeasta osoiteavaruudesta. Yleensä DHCP-palvelin on asetettu jakamaan osoitteita samasta avaruudesta suurin piirtein samalle maantieteelliselle alueelle ja normaalisti ainakin samaan DSLAMiin kytkeytyvät asiakkaat saavat osoitteensa samasta osoiteavaruudesta. Jokainen IP-osoitteen vuokratapahtuma tallennetaan DHCP-palvelimen lokitietoihin. Jokaisen tapahtuman yhteyteen tallennetaan myös tunnistetiedot [7]. Näin DHCP-palvelimen lokitiedoista selviää kunkin IP-osoitteen vuokraaja tietyllä ajanhetkellä.

Asiakkaan saatua IP-osoitteen, hän voi alkaa liikennöidä julkiseen Internetiin. Asiakastiedoista haetaan asiakkaan liittymän tyyppi, yhteyden nopeus ja mahdolliset muut liittymässä olevat yhteyteen liittyvät parametrit. Yhteyden nopeus rajoitetaan tilatulle tasolle DSLAMin portin perusteella viimeistään operaattorin reunareitittimessä Internetiin päin. Verkon aktiivilaitteiden konfigurointi voitaisiin jokaisen asiakkaan kohdalla tehdä manuaalisesti, mutta tehokkuuden ja virheiden välttämisen takia konfigurointi hoidetaan useimmiten erillisellä automaattisella järjestelmällä (Kuva 10). Järjestelmä tekee tarvittavat konfiguraatiomuutokset reitittimiin ja DSLAMiin asiakas- ja porttitietokannan perusteella. Sama automaattinen järjestelmä voi hallita myös DHCP-palvelimen IP-osoitteiden jakoa, jolloin voisi olla mahdollista yhdistää suoraan DHCP-palvelimen lokitiedot asiakastietoihin.




**Kuva 10: Operaattoriverkon topologia ja asiakasyhteyden hallinnassa mukana olevat laitteet (saattavat vaihdella eri operaattoreiden välillä).**

## 4.2. Tekniset vaatimukset

Järjestelmän pitää tallentaa tunnistetiedot vuoden ajalta. Tämän ajanjakson kuluttua sitä vanhemmat tiedot pitää tuhota. Järjestelmästä pitää pystyä hakemaan tietyn IP-osoitteen haltijan tiedot tietyllä ajanhetkellä tai -jaksolla. Järjestelmän tulee tallentaa kaikki asiakastietojen hakutoimenpiteet siten, että jokaisen toimenpiteen tekijä, tekoaika sekä hakukriteerit ja saadut vastaukset tallennetaan.

Yksinkertaisin käyttöliittymä järjestelmälle saadaan toteuttamalla palvelu selaimessa toimivana. Tällöin järjestelmään pääsy tulee rajoittaa paitsi tunnuksella ja salasanalla, myös mahdollisilla IP-osoiterajoitteilla. Näin voidaan varmistaa, ettei kukaan ulkopuolinen pääse järjestelmään ilman suurta vaivannäköä. Selainkäyttöliittymän valintaa tukee myös se, että palvelimelta ulospäin tarvitsee sallia vain HTTP-protokollan (Hypertext Transfer Protocol) käyttämä portti (80), jolloin esimerkiksi suorat tietokantayhteydet voidaan estää suoraan palomuurimäärityksissä. Tämänkaltaisilla rajoitteilla ja ominaisuuksilla toteutuu direktiivin ja Suomen lain vaatimukset tietoturvasta kohtuullisin kustannuksin.

Vaihtoehtona tietojen tallennuskohteelle voisi olla tekstitiedosto, jollaiseen DHCP-palvelinkin tallentaa lokitiedot. Tekstitiedostoon tulee jokaista vuokraustapahtumaa kohden vähintään neljä riviä. Yksi osoitteen pyynnölle, yksi osoitteen vuokrausvastaukselle sekä yksi kummallekin optio 82:n alioptiolle eli Circuit-ID:lle ja Remote-ID:lle. Jokaisella rivillä on tietosisällöstä riippumatta lisäksi aikaleima.

Esimerkiksi riviltä

```
Oct 11 04:02:27 ns2 dhcpd: DHCPREQUEST for 83.102.98.111 from
00:04:ed:60:6f:f0 via 83.102.96.1
```

voidaan lukea aikaleima, asiakkaan pyytämä IP-osoite, asiakkaan MAC-osoite sekä reititin, joka on optio 82 -kentän lisännyt. Yhdeltä riviltäkin pitää siis parsia olennaiset tiedot ohjelmallisesti. Tällaiseen tekstitiedostoon tulee rivejä huomattavia määriä (ote erään palvelimen lokitiedostosta liitteessä 2). Yhtä tiedostoa ei ole järkevää kasvattaa koko vuoden aikana kerätyllä tiedolla, vaan yleensä lokitiedosto tallennetaan kerran viikossa erilleen, jolloin yksi tekstitiedosto sisältää aina viikon aikana tehdyt IP-osoitteen vuokraustapahtumat. Tallennettuja tietoja vuoden ajalta on siis 52-53 tiedostossa. Tekstitiedoston läpikäynti tapahtuu ohjelmallisesti rivi kerrallaan joko alusta tai lopusta lähtien.

Kun halutaan hakea lokitiedoista yhtä tiettyä IP-osoitetta tietyllä ajanhetkellä, niin pahimmassa tapauksessa tämä ajanhetken ja osoitteen yhdistelmä löytyy vasta viimeisestä tiedostosta, jota käydään läpi. Jokaisella hakuoperaatiolla joudutaan siis keskimäärin käymään läpi 26 tiedostoa. Tähän on päästy sillä oletuksella, että osoitteiden ja ajankohtien hakukohteet ovat täysin satunnaisia ja voivat jakautua tasaisesti vuoden ajalle. Tätä keskiarvoa pystytään pienentämään, mikäli viikoittain tallennetut tiedostot tallennetaan sellaiselle nimelle, joka sisältää numeron siltä viikolta, jonka tiedot ko. tiedostossa ovat. Tällöin haun yhteydessä haettaisiin ensin halutun päivämäärän sisältävä viikko ja luettaisiin vain kyseisen viikon tiedosto.

Tuotantokäytössä olevan DHCP-palvelimen lokitiedostoja tutkiessa on tullut ilmi, että mikäli yksi tiedosto tallentaisi koko viikon tiedot kerralla, sen kooksi tulisi jopa yksi gigatavu. Rivejä tuollaisessa tiedostossa olisi jo reilut 9,5 miljoonaa. Tällaisen tiedoston läpikäymisen aika riippuu laitteiston suorituskyvystä ja lähinnä keskusmuistin määrästä. Mikäli tiedosto pystytään lukemaan kokonaan keskusmuistiin ja käydä läpi siellä, niin läpikäyminen on paljon nopeampaa, kuin jos välillä joudutaan lukemaan tiedoston osia levyltä. Oletetaan, että nykykoneilla tuollainen tiedosto mahtuu kerralla keskusmuistiin. Tällöinkin yhden tiedoston täydellinen läpikäynti kestää useita sekunteja. Mikäli hakuja jouduttaisiin tekemään jostain syystä useita lyhyen ajan sisällä, muodostuu puhdas tiedostojen läpikäyminen melkoisen hitaaksi ja turhaan resursseja kuluttavaksi operaatioksi.

Vaihtoehtona tiedostopohjaiseen hakutoimenpiteeseen on tietokantapohjaisuus. DHCP-palvelimet eivät tallenna lokitietojaan suoraan tietokantaan, joten tiedot on edelleen luettava tiedostoista ja tallennettava kantaan. Tietojen kerääminen on tietokantapohjaisella tavalla hitaampaa ylimääräisen toimenpiteen vuoksi, mutta hakujen tekeminen tietokantaan poistaa tiedostojen läpikäynnin tarpeen. Lisäksi tietokannat on kehitetty niin, että erityisesti hakujen tekeminen on nopeaa indeksoinnin ja avainten takia. Tietojen lisääminen tietokantaan aiheuttaa tiedoston ja sen rivien lukemisen ja tietojen parsimisen, mutta hakujen yhteydessä mitään tietoja ei tarvitse enää hakea

erikseen merkkijonoista, vaan jokainen olennainen tieto on omana kenttänään tietokannan taulussa. Tietokantapohjaisuuden valinnan tukena on myös ajatus järjestelmän nettiselainpohjaisuudesta. Yleisimmillä nettiohjelmointikielillä on yksinkertaisempaa ja nopeampaa lukea tietoa tietokannasta kuin jostakin palvelimen levyllä olevasta tiedostosta. Järjestelmän keräämien tietojen tallennuspaikaksi valitaan siis tietokanta.

Tietokanta ja erityisesti sen yksi taulu tulee muodostumaan melkoisen suureksi ja levytilaa vieväksi. Mikäli operaattorilla on asiakkaita 10000 ja jokainen asiakas vaihtaisi IP-osoitteen kerran vuorokaudessa uuteen (oletetaan siis, että osoitteiden vuokra-aikoja ei jatketa, vaan osoite vaihtuu jokaisella vuokrakerralla aina uuteen). Tällöin vuodessa tulisi

```
366 x 10000 = 3 660 000
```

merkintää tietokannan tauluun, johon tallennetaan tunnistetiedot. Kuitenkin tarkoituksena on mitoittaa kanta niin, että se varmasti kattaa Suomen tarpeet silläkin oletuksella, että yksi ja ainoa teleoperaattori hoitaisi kaikki yhteydet.

Tilastokeskuksen mukaan Suomessa on ollut vuonna 2007 308 917 yritystä [8]. Vuonna 2008 vakinaisesti asuttuja asuntokuntia eli kotitalouksia on ollut 2 499 000 [9]. Jos näiden lisäksi tehdään oletus, että jokaisella yrityksellä olisi käytössään viisi julkista IP-osoitetta sekä jokaisella kotitaloudella yksi IP-osoite. Tämä tarkoittaisi että käytössä olisi kaikkiaan

```
5 x 308 917 + 2 499 000 = 4 043 585
```

IP-osoitetta. Jos edelleen edellisen kappaleen mukaisesti ajatellaan, että jokainen näistä vaihtaisi IP-osoitetta kerran vuorokaudessa, niin tietokantaan tulisi 366 x 4 043 585 = 1 479 952 110 tietuetta. Tämä laskenta perustuu kuitenkin todella rajuihin olettamuksiin sekä yliarviointeihin ja siitä syystä tämä onkin niin sanottu ”pahimmassa tapauksessa” -tilanne. Järjestelmän mitoituksessa voidaan huoletta käyttää arvoina maksimissaan 50 prosenttia lasketusta määrästä, koska Suomessa on useampi teleoperaattori ja lisäksi käytettävien IP-osoitteiden määrä vuorokaudessa on huomattavasti tuota pienempi johtuen kiinteistä osoitteista ja siitä, että useat osoitteita tarvitsevat laitteet pyytävät vain osoitteen vuokra-ajan jatkoa eivätkä kokonaan uutta osoitetta. Vuokra-ajan jatkolla voidaan päivittää vain saman osoitteen vuokran yksityiskohtia tallentamatta tapahtumaa uutena vuokrauksena tietokantaan.

## 4.3. Toiminnallisuus

Luvussa 4.2. tehdyn valinnan mukaan DHCP-palvelimen lokitiedoista luetaan kaikki tunnistetiedot ohjelmallisesti tietokantaan. Työn kohdetapauksessa DHCP-palvelin on toteutettu Linux-pohjaisena järjestelmänä ja lokitiedot ovat selkokielisinä tekstitiedostossa. Ohjelmointikieleksi lokitiedostojen lukemiseen valitaan Python, koska sillä tekstimuotoisten tiedostojen läpikäynti on riittävän nopeaa sekä vähän resursseja kuluttavaa.

Lisäksi siihen on saatavilla valmiina lukuisa joukko moduuleja, joita voidaan käyttää hyödyksi järjestelmän toteutuksessa. Toinen hyvä vaihtoehto Linux-pohjaiseen järjestelmään olisi voinut olla Perl. Windows-arkkitehtuurin päällä toimiva DHCP-palvelin tallentaisi DHCP:m tunnistetiedot Windowsin tapahtumalokiin. Tosin Windows 2003-pohjainen DHCP-palvelin ei suoraan tue DHCP:n optio 82 -kenttää [10], joten Windows-pohjaista palvelinta käytetään pääosin vain yritysverkoissa jakamaan yrityksen sisäisiä yksityisiä osoitteita oman verkon työasemille. Windows-pohjaisen DHCP-palvelimen tietojen läpikäyntiä ei tämän työn osalta tulla käsittelemään.

DHCP:n lokitiedosto tulee käydä läpi säännöllisesti, jotta kaikki osoitteiden vuokratapahtumat olisivat aina ajan tasalla myös tietokannassa. Järjestelmän käynnistyksen yhteydessä sen hetkinen lokitiedosto luetaan ja käsitellään kokonaan, jotta saadaan aloitustilanne talteen. Tämän jälkeen jokaisella tiedoston lukukerralla käsitellään vain ne rivit, joita ei ole aiemmin käsitelty. Vanhat tiedot jätetään huomiotta, jotta samoja tietoja ei tallennettaisi tietokantaan useaan kertaan. Tiedoston läpikäynnin ajoitus voidaan määritellä suhteellisen vapaasti, mutta järkevintä on käydä tiedosto läpi vähintään kerran vuorokaudessa. Näin tietokannassa olisi maksimissaan vuorokauden viiveellä kaikki tiedot, joita DHCP-palvelimesta voidaan tunnistetietoina kerätä.

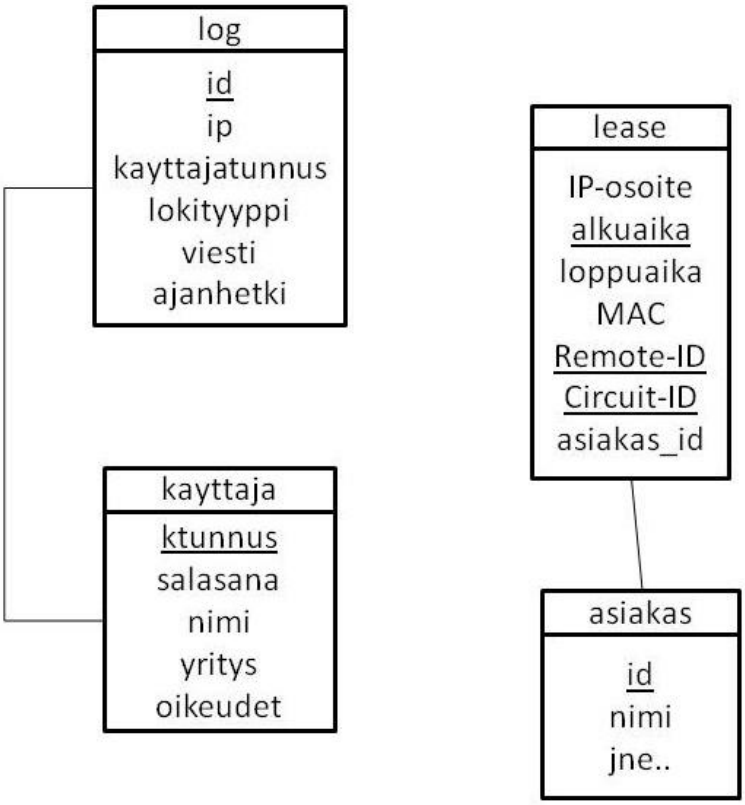
Käyttöliittymän kautta hallitaan käyttäjätunnuksia, joilla on pääsy järjestelmään, ja suoritetaan hakuja tietoihin. Käyttäjätunnuksien hallinnasta pois suljetaan fyysiset konsoliyhteydet, joilla pääsee käsiksi suoraan itse palvelimeen. Järjestelmien ylläpitoa varten niissä on pääkäyttäjätunnukset, joilla voi hallita järjestelmää koneen vierestä tai etäyhteyden päästä konsoliyhteydellä. Pääkäyttäjätunnuksia on perinteisesti vain yksi eikä niitä tarkkailemalla ole mahdollista selvittää henkilöä, joka on operaatiot tehnyt. Pääkäyttäjätunnuksilla on oikeus tehdä järjestelmässä sellaisia muutoksia, jotka vaikuttavat myös tunnistetietojen keräämiseen ja näin on mahdollista myös peittää jälkiä mahdollisesta direktiivin ja lain vastaisesta tietojen käsittelystä [11].

Luvussa 4.2. määriteltiin sovelluksen käyttöliittymä toteutettavaksi selainpohjaisena. Käyttöliittymä eli web-sovellus tullaan toteuttamaan PHP-ohjelmointikielellä, sillä sen avulla saadaan palvelun tietoturva halutulle tasolle helposti ja lisäksi kieli on riittävän kevyt, jotta hakuihin ei tule liian suuria viiveitä. Käyttöliittymän toteuttamiseen käytettävä ohjelmointikieli voisi olla mikä tahansa muukin, jolla pystytään tuottamaan dynaamisia verkkosivuja ja tekemään kyselyjä tietokantaan.

### 4.3.1. **Tietojen tallennus**

Tietokanta, johon DHCP:stä saatavat tiedot tallennetaan, tulee olemaan hyvin yksinkertainen. Tämä siitä syystä että kantaan tulevaa dataa on melkoisen paljon ja tällöin kantarakenteen yksinkertaisuus nopeuttaa huomattavasti toimintaa, kun ei tarvitse tehdä raskaita taulujen välisiä liitoksia tai ristikkäiskyselyjä. Suurin osa datasta tullaan tallentamaan lease-tauluun, joka nimensä mukaisesti pitää sisällään kaikki vuokratut IP-osoitteet alku- ja loppuaikoineen sekä yksilöintitietoineen (Kuva 11). Taulun avaimena on vuokran alkuaika, MAC-osoite sekä optio 82:n tiedot, Remote-ID ja Circuit-ID.

Lisäksi tauluun tehdään tietokantaindeksi kentälle IP-osoite. Indeksi parantaa taulusta hakemisen nopeutta ja koska yhtenä hakukriteerinä tulee aina olemaan IP-osoite, niin indeksiksi on luontevaa valita juuri se kenttä.

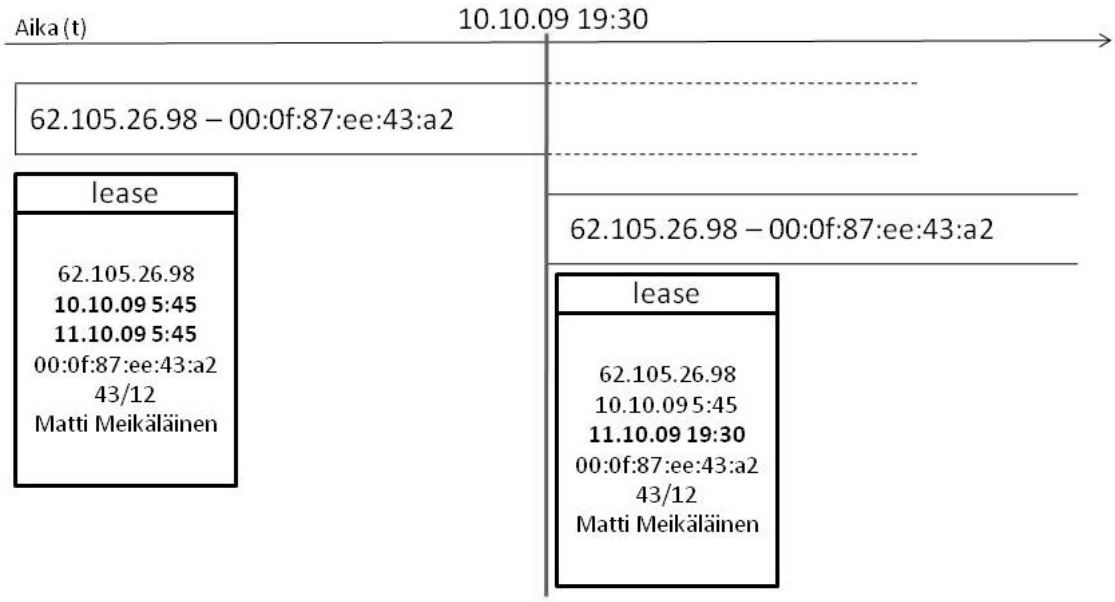



**Kuva 11: Järjestelmän tietokannan rakenne**

Tietojen tallennuksen osalta toiminta on seuraavanlainen. DHCP:n lokitiedostoa luetaan rivi kerrallaan ja eritellään rivistä kaikki tiedot. Päähuomio keskittyy kuitenkin IP-osoitteeseen, aikaan sekä optio 82:n tietoihin. Kaikkien yhteen tapahtumaan liittyvien tunnistetietojen lukemisen jälkeen tehdään haku tietokantaan saman IP-osoitteen ja ajan perusteella. Aika on tässä tapauksessa osoitteen vuokran alkuaika. Mikäli kyselystä saadaan tulos, on kyseinen rivi tallennettu jo aiemmin kantaan ja voidaan siirtyä lokitiedoston seuraavaan tapahtumaan. Jos kysely ei kuitenkaan tuota tulosta, niin kyseessä on uusi osoitteen vuokratapahtuma ja se pitää lisätä kantaan.

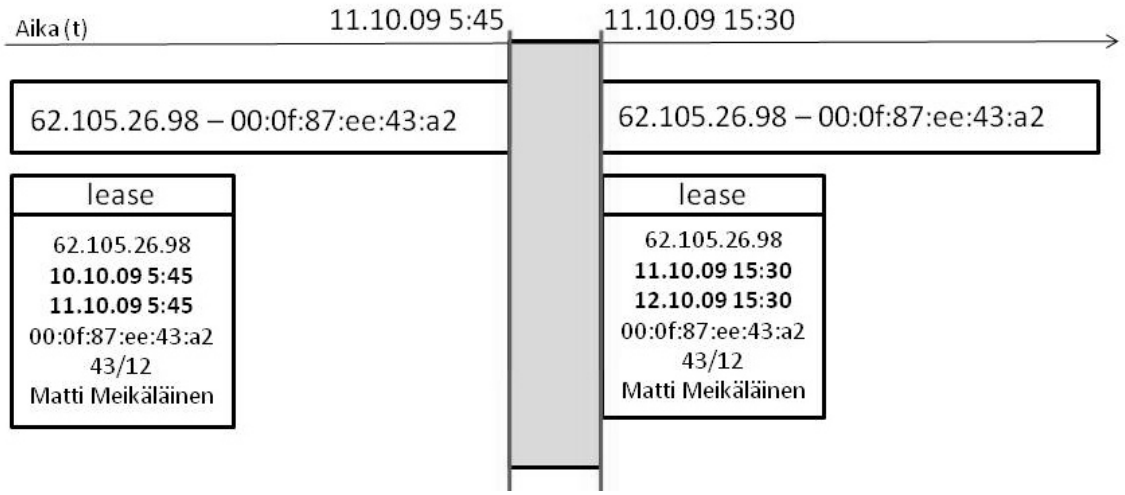
Lisäyksen aluksi kannasta haetaan edellinen kyseisen IP-osoitteen haltija. Tämä tapahtuu hakemalla samaa IP-osoitetta, jolla on korkein alkuaika. Mikäli edellinen ja uusi vuokraaja ovat samat (vertailu tehdään optio 82:n tiedoista) ja edellisen vuokran loppuaika on merkitty myöhemmäksi kuin tämän käsiteltävän tapahtuman alkuaika, niin kasvatetaan vain tietokannassa kyseisen tapahtuman sisältävässä tietueessa vuokran loppuaikaa (Kuva 12), sillä kyseessä on tällöin saman osoitteen vuokra-ajan jatkopyyntö. Kuvassa 12 asiakas Matti Meikäläinen on vuokrannut osoitteen käyttöönsä 10.10.09 ja aluksi vuokran kestoksi on talletettu 24 tuntia. Osoitteen vuokra-aika loppuisi 11.10.10 kello 5.45. Matin laite kuitenkin pyytää saman osoitteen uusintaa myöhemmin samana päivänä kello 19.30. Palvelin on hyväksynyt pyynnön ja tietokannassa päivite-

tään saman osoitteen vuokra-ajan loppuajaksi 11.10.10 kello 19.30 eli 24 tuntia eteenpäin tämän uuden pyynnön hetkestä. DHCP-palvelin voidaan konfiguroida käyttämään mitä tahansa aikaa vuokran kestona, mutta yleisesti kuluttajatason laajakaistaliittymissä aikana on yksi vuorokausi. Tätä arvoa voidaan muuttaa järjestelmän käyttöliittymässä sen mukaiseksi, miksi se on palvelimessa asetettu.




**Kuva 12: Vuokratapahtuman tallennus IP-osoitteen vuokra-ajan uusinnan tapauksessa.**

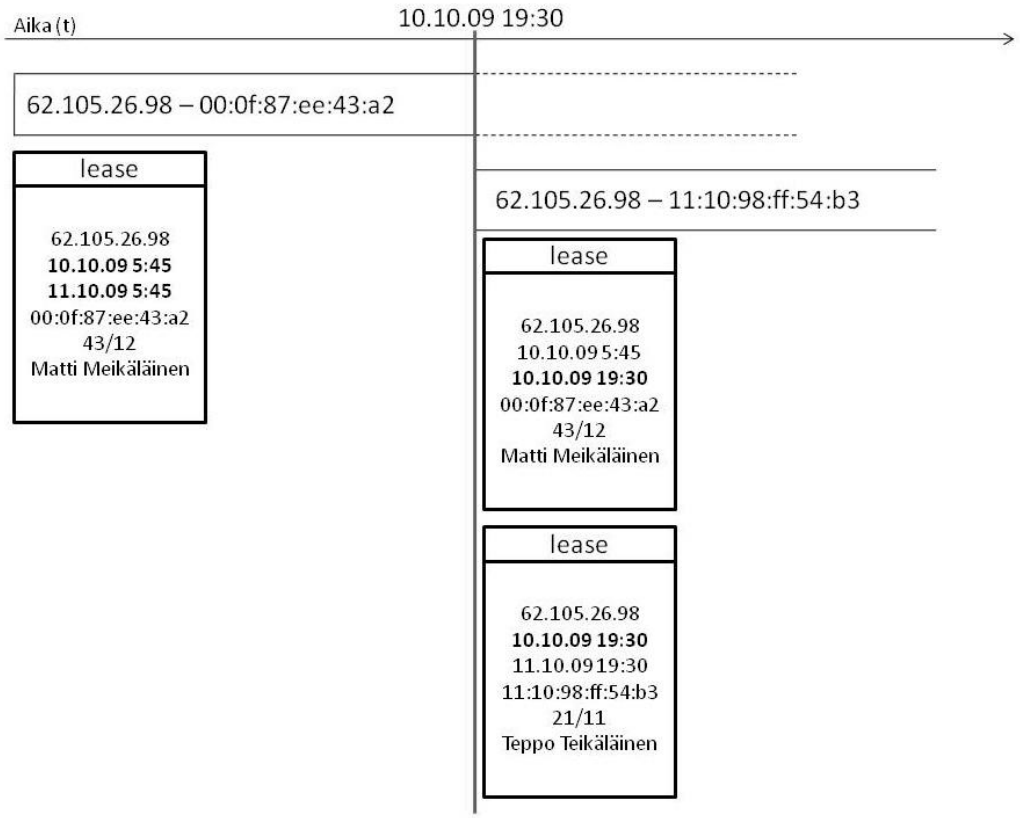
Osoitteen vuokraajat voivat olla samat myös silloin, kun uuden tapahtuman alkuaika on myöhemmin kuin aiemman saman osoitteen tapahtuman loppuaika. Tällöin kyseessä on uusi vuokratapahtuma, jossa asiakaslaite pyytää palvelimelta samaa osoitetta, joka sillä on aiemmin ollut käytössään. Tapahtumasta tulee kuitenkin tallentaa uusi tietue tietokantaan, koska aiemman tapahtuman loppuajan ja uuden tapahtuman alkuajan välillä osoite ei ole ollut käytössä kenelläkään. Kuvassa 13 Matti Meikäläisen vuokraaman osoitteen vuokra-aika päättyi 11.10.09 kello 5.45. Tämän jälkeen osoite ei ole ollut kenelläkään käytössä, kunnes Matti pyytää palvelimelta samaa osoitetta käyttöönsä 11.10.09 kello 15.30. Tästä luodaan tietokantaan uusi tietue, johon aloitusajaksi tallentuu pyynnön aika ja loppuajaksi 24 tuntia myöhäisempi aika.




**Kuva 13: Vuokratapahtuman tallennus, kun asiakas on sama, mutta ajat poikkeavat.**

Kuvan 13 mukainen vuokratapahtuma on kyseessä myös silloin, kun vuokraajat ovat erilaiset, mutta tapahtuman alkuaika on myöhempi kuin edellisen vuokran loppuaika. Tapahtuma tallennetaan uutena tapahtumana tietokantaan uuden vuokraajan tiedoilla eli kuvassa 13 oikean puoleinen lease -taulun tietue sisältäisikin Matti Meikäläisen tietojen sijasta tämän uuden vuokraajan tiedot.

Monimutkaisin tilanne tallennuksen yhteydessä tulee silloin kun vuokratapahtuman alkuaika on aiemmin kuin saman osoitteen edellisen vuokran loppuaika ja vuokraajat ovat erilaiset. Tällöin aiempi vuokraaja on jostain syystä vapauttanut vuokraamansa osoitteen ennen vuokran loppumista ja DHCP-palvelin on merkannut osoitteen vapaaksi. Sitten on tullut uusi osoitepyyntö, johon DHCP-palvelin on vastannut luovuttavansa vuokralle tämän aiemmin vapautetun osoitteen. Tietokannassa aiemman vuokratapahtuman sisältämän tietueen loppuajaksi pitää päivittää tämän uuden tapahtuman alkuaika, koska silloin osoitteen käyttäjä vaihtuu. Uuden tapahtuman tallennus suoritetaan normaalisti asettaen tapahtuman loppuajaksi 24 tuntia alkuajasta. Tämän tilanteen tallennuksen proseduuri on esitetty kuvassa 14. Siinä IP-osoite on aluksi vuokrattu kuvitteellisella ajanhetkellä 10.10.09 kello 5:45 asiakkaalle Matti Meikäläinen, jonka optio82:n tiedot ovat 43/12. Vuokran loppumisajaksi on vuokran alkuvaiheessa määritelty yksi vuorokausi. Kuitenkin asiakas on jostain syystä vapauttanut vuokraamansa IP-osoitteen ja DHCP-palvelin on vuokrannut sen uudelle henkilölle 10.10.09 kello 19:30. Entisen käyttäjän eli Matti Meikäläisen vuokrauksen loppuajaksi päivitetään tämän uuden vuokran alkuaika, koska osoite on tällöin vuokrattu toiselle käyttäjälle. Samalla tallennetaan tämän uuden käyttäjän vuokratapahtuma uutena tietokantaan. Uudeksi alkamisajaksi asetetaan tämä vaihtoaika ja loppuajaksi jälleen oletuksena oleva yksi vuorokausi eteenpäin.




**Kuva 14: Järjestelmän toiminta IP-osoitteen tallennuksessa, kun osoitteen vuokraaja vaihtuu kesken oletetun vuokra-ajan.**

### 4.3.2. **Tietojen haku**

Tunnistetietojen keräys- ja tallennusjärjestelmän näkökulmasta tallennettuihin tietoihin on pääsy vain selainkäyttöliittymän kautta. Tietokannan ja itse palvelimen tietoihin on toki pääsy myös järjestelmän pääkäyttäjällä ja konsoliyhteyksillä (luku 4.3). Keräys- ja tallennusjärjestelmän hallinta hoidetaan kuitenkin täysin selaimella.

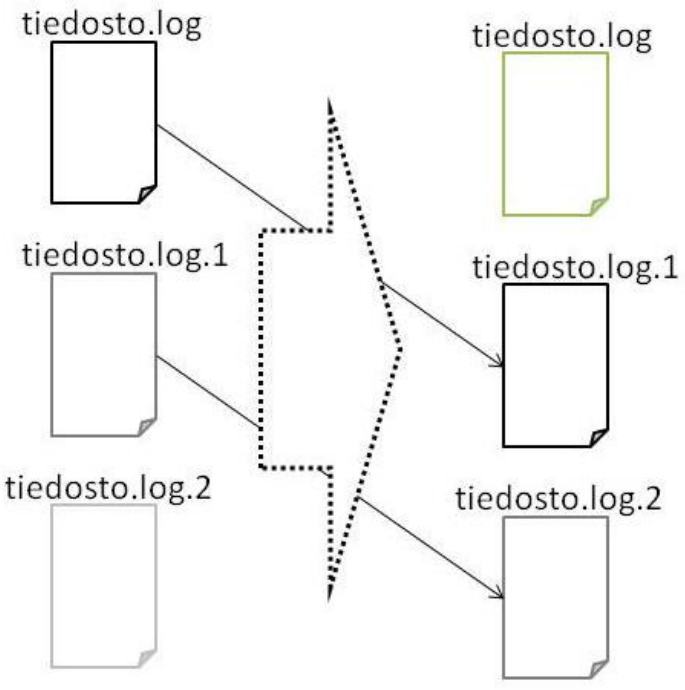
Tunnistetietojen hakuprosessi lähtee siitä, että viranomaiset haluavat tietää tietyllä ajanhetkellä tietyn IP-osoitteen käyttäjän (vuokraajan). He tekevät siitä pyynnön tele-operaattorille, joka sitten halutuilla kriteereillä suorittaa haun tähän keräysjärjestelmään. Haku suoritetaan aina järjestelmään ennalta määritellyillä tunnuksilla, jotta voidaan varmistua haun tekijän henkilöllisyydestä. Ennen haun tulosten näyttämistä hausta tallennetaan järjestelmään tekoaika, haun tekijä (henkilön käyttäjätunnus ja nimi) sekä hakukriteerit eli IP-osoite ja aika. Myös haun tulokset tallennetaan.

Haku tehdään tietokannan lease-tauluun. Lease-taulu sisältää jokaisen DHCP-palvelimen lokitiedostosta luetun tunnistetiedon. Taulusta haetaan kaikki halutun IP-osoitteen sisältävät rivit. Sitten tulokset rajataan koskemaan vain rivejä, joissa haluttu aika on alkuajan ja loppuajan välissä. Tämän toimenpiteen tuloksena pitäisi tulla vain yksi tietokannan rivi, koska vain yhdellä asiakkaalla yhdellä ajanhetkellä on voinut olla kyseinen osoite käytössään. Riviltä luetaan kaikki tallennetut tunnistetiedot (Circuit-ID, Remote-ID, MAC-osoite, IP-osoite sekä vuokran alku- ja loppuajat) ja tehdään vielä haku asiakastauluun, jotta saadaan selville asiakkaan tiedot (erityisesti nimi ja yhteys-

tiedot). Tuloksen IP-osoite, MAC-osoite, optio 82:n tiedot, ajat sekä asiakkaan tiedot tallennetaan haun yhteyteen ja näytetään taulukoidussa muodossa haun tekijälle. Lisäksi järjestelmä mahdollistaa tulosten viennin esimerkiksi Microsoft Excel -taulukkoon CSV-muodossa (Comma-separated Values), jolloin tietojen jatkomuotoilu raportointia ja viranomaistoimitusta varten helpottuu.

### 4.3.3. Järjestelmän siivous

Koska järjestelmään pitää tallentaa tiedot vuoden ajalta, mutta sitä vanhemmat on poistettava tai tehtävä sellaisiksi, ettei niistä ole enää mahdollista yhdistää tunnistetietoja käyttäjiin, on järjestelmän tehtävä myös tyhjennys- ja siivousoperaatioita. Linux-ympäristössä erilaisten lokitiedostojen kierrättämiseen on olemassa valmiita ohjelmia (esim. logrotate [12]), jotka käyvät läpi systeemin lokitiedostoja kopioiden aina vanhimmat tiedot uuteen tiedostoon ja poistaen kaikkein vanhimmat tiedot. Samanlaista operaatiota käytetään myös DHCP-palvelimen tuottamaan lokitiedostoon. Yksinkertaisimmillaan siirretään määrätyin väliajoin lokitiedosto ja nimetään siirretyt juoksevalla numeroinnilla. Lisäksi säilytysajan tai tietyn järjestysnumeron saavuttua tuhotaan vanhin tiedosto. Numeroinnissa uusin tiedosto on aina 1 ja vanhimmalla on suurin numeroarvo. Esimerkki kuvassa 15 tiedosto.log on se lokitiedosto, johon kirjoitetaan. Tiedostojen kierrättämisen yhteydessä se siirretään nimelle tiedosto.log.1 ja entinen tiedosto.log.1 siirretään nimelle tiedosto.log.2 ja niin edelleen. Lopulta vastaan tulee tiedosto, jonka järjestysnumero kasvaisi yli asetetun rajan, jolloin kyseinen tiedosto poistetaan. Kuvassa 15 tiedosto.log.2 siis poistetaan, koska tiedosto.log.1 siirretään tiedosto.log.2 -nimelle ja tilanteessa oletetaan, että suurin säilytettävän tiedoston järjestysnumero on kaksi.




**Kuva 15: Tiedostojen kierrättämisen yksinkertainen toimintatapa.**

Tiedostojen kierrättämisen avulla DHCP-palvelimen lokitiedostot on mahdollista pitää edelleen tallessa halutun ajan, jonka jälkeen ne voi turvallisesti tuhota. Kyseisiä tiedostoja voisi kierrättää esimerkiksi kerran päivässä, jolloin järjestysnumeroltaan 366 ylittävät tiedostot poistettaisiin automaattisesti.

Tietojen keräys- ja hakujärjestelmä tallentaa tunnistetiedot näistä DCHP-lokitiedostoista kuitenkin tietokantaan, josta pitää myös tuhota vuotta vanhemmat tiedot. Tietojen poisto suoritetaan vastaavalla ajastuksella kuin fyysisten tiedostojenkin tapauksessa. Kerran vuorokaudessa verrataan kannassa olevia osoitteiden vuokra-aikojen loppuja kuluvaan ajanhetkeen. Kaikki kannan tunnistetietorivit, joissa loppuaika on 366 vuorokautta vanhempi kuin nykyinen aika, poistetaan. 366 vuorokautta on valittu järjestelmän tietojen säilytysajaksi sen takia, että laki määritteli tietojen maksimisäilytysajaksi vuoden, mutta ei määritellyt vuoden pituutta. 366 vuorokauden säilytysajalla varmistetaan, että myös karkausvuosien ajalta säilytetään varmasti kalenterivuoden mittainen jakso tunnistetietoja.

Tuhoamalla riittävän vanhat palvelimen tekemät tiedostot sekä tietokannan rivit, voidaan varmistua siitä, ettei tunnistetietoja jää lain määräämällä tavalla mihinkään sillä tavalla, että niistä voisi vielä yhdistellä käyttäjien tunnistetietoja. Tietokannan tapauksessa siis riittäisi myös tuhota riittävän vanhoista tiedoista optio 82:n tiedot, jolloin tiedoista ei enää voitaisi yksilöidä käyttäjää. Tämä on yhtä kauan kestävä operaatio kuin rivien poistaminen, mutta järjestelmän kannalta se kasvattaa turhan tiedon määrää ja tallennustarvetta, joten tietokannassa rivien poistaminen on perusteltua.

### 4.3.4. **Järjestelmän toiminnan lokitus**

Suomen sähköisen viestinnän tietosuojalain [3] mukaan järjestelmästä tulee saada tieto siitä, kuka ja milloin järjestelmän tietoja on hakenut, mitä tietoja on haettu ja millä kriteereillä sekä millaisia tuloksia haku on tuottanut. Tietokantaan luodaan tätä varten log-taulu (kuva 11), jonne tallennetaan kaikki käyttäjän tekemät toimenpiteet järjestelmässä. Tämä tuo mahdollisuuden seurata kuinka paljon järjestelmää käytetään ja minkälaisiin tapauksiin. Lisäksi tietojen avulla on helppo koostaa vuosittainen viranomaisraportti tunnistetietojen käytöstä.

Järjestelmään luodaan tietokantapohjainen käyttäjäautentikointi. Tietokantaan tulee siis jokaiselle järjestelmään pääsevälle oma käyttäjätunnuksensa ja salasanansa, joilla käyttäjä kirjautuu järjestelmään. Käyttäjäautentikoinnin avulla järjestelmässä nähdään kuka käyttäjä on milloinkin järjestelmässä jotain tehnyt. Käyttäjätunnus tallennetaan tietokantaan selväkielisenä, mutta salasana säilytetään kannassakin tiivistemuodossa, jotta sitä ei pystytä lukemaan sellaisenaan edes suoraan tietokannasta. Tiivistemuoto on merkkijonosta tehty "sormenjälki". Tiivistemuoto on aina sama täsmälleen samalle merkkijonolle, kun käytetään samaa tiivistefunktiota suorittamaan operaatio. Tiivistemuodon etuna on se, että siitä ei ole mahdollista tehdä vastakkaista operaatiota eli selvittää selkokielistä merkkijonoa tiivistemuodosta [13]. Tästä syystä kyseistä muotoa on hyvä käyttää salasanan tallentamiseen tietokantaan. Vaikka ei-toivottu hyökkääjä

pääsisikin lukemaan tietokantaa, hän ei silti näkisi järjestelmän käyttäjien salasanoja, eikä näin ollen voisi suorittaa operaatioita toisena käyttäjänä. Tiivistemuotoa voidaan kuitenkin käyttää autentikoinnin yhteydessä niin, että verrataan kirjautumisen yhteydessä annetun salasanan tiivistettä tietokannassa olevaan ja mikäli ne ovat samat, niin salasana on ollut oikea. Missään vaiheessa järjestelmän ei siis tarvitse verrata selkokielisiä salasanoja.

Järjestelmän käyttäjän on siis annettava käyttäjätunnus ja salasana ennen kuin hän pääsee suorittamaan muita operaatioita järjestelmässä. Kaikista operaatioista, joita järjestelmässä suoritetaan käyttöliittymän kautta, tallennetaan merkintä tietokantaan. Näihin operaatioihin lukeutuvat sisäänkirjautumisen lisäksi uloskirjautuminen, sisäänkirjautumisyritys, haun tekeminen sekä järjestelmän yleisten asetusten muuttaminen. Sisäänkirjautumisyritykset tallennetaan mahdollisten hyökkäysten varalta, sillä mikäli niitä alkaa tulla useita ilman syytä, niin voidaan alkaa selvittää yritysten lähdettä. Yritys-merkinnän toki tuottaa myös laillinen käyttäjä unohdettuaan salasanan tai kirjoitettuaan sen väärin. Jokainen tallennettu lokimerkintä sisältää käytetyn koneen IP-osoitteen, operaation suorituksen aikaleiman, käyttäjätunnuksen sekä operaation tyypin.

Sisään- ja uloskirjautumisten tallennuksien avulla saadaan tieto kunkin käyttäjän järjestelmässä käytetystä ajasta, joka tosin ei aina ole välttämättä tosiaikainen, sillä mikäli käyttäjä unohtaa uloskirjautumisen, niin hänelle ei tietenkään tule siitä merkintää. Käyttäjän uloskirjaus tapahtuu kuitenkin aina, kun selainikkuna suljetaan eli käyttäjän kirjautuminen on sidottu yhteen selainistuntoon.

Käyttäjän tekemästä hausta tallennetaan operaation yhteyteen viesti-kenttään (kuva 11) sekä annetut hakuehdot että haulla saadut tulokset. Huomion arvoista on, että vaikka järjestelmän tallentamia asiakkaiden tunnistetietoja siivotaan (luku 4.3.3), niin hakujen yhteydessä tallennettuja tulostietoja ei poisteta. Tämä siitä syystä, että tunnistetietojen pyynnön yhteydessä viranomaisilla on epäily tai tutkinta rikoksesta, johon liittyviä tunnistetietoja ei poisteta vuoden vanhenemisajan jälkeen. Järjestelmän siivous on kuitenkin yksinkertaisempaa, kun poistetaan tunnistetiedot normaalisti ja rikostutkinnan kannalta etsityt tärkeät tiedot säilyvät log-taulussa.

## 4.4. Järjestelmän automaattinen toiminta skriptitasolla

Järjestelmän sydän on DHCP-palvelimen lokitiedostoa läpikäyvä Python-skripti. Sitä ajetaan ajastetusti Linux-käyttöjärjestelmän cron-ohjelmaa käyttäen halutuin väliajoin. Luvussa 4.3 määriteltiin, että lokitiedosto tulisi lukea vähintään kerran vuorokaudessa. Oletuksena ajastus laitetaan juuri tähän, mutta samalla tulee varmistaa, että lokitiedostoja kierrättävä ohjelma on ajastettu suorittamaan kierrätyksen vasta tämän järjestelmän lukuskriptin jälkeen. Muuten on mahdollista, että jokin tunnistetieto jää lukematta ja tallentamatta, koska kierrättävä ohjelma on ehtinyt siirtää kyseisen tiedon talteen toiseen tiedostoon.

Järjestelmän siivoavaa skriptiä ajetaan myös kerran vuorokaudessa. Sen ajastus ei ole niin kriittinen eikä riippuvainen muista ajettavista ohjelmista, mutta silti se on hyvä

ajaa riittävän usein, ettei sorruta säilyttämään tunnistetietoja lain määräämää ajanjaksoa pidempään.

## 4.5. Palvelun rakenne

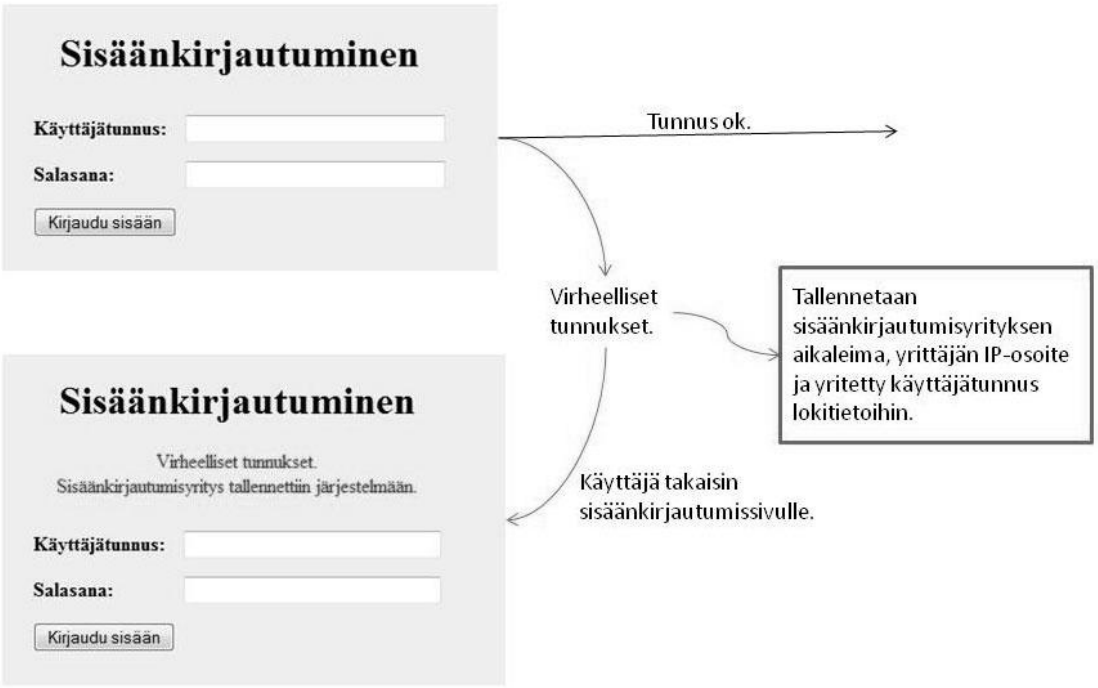
Palvelusta tuotetaan tämän työn tiimoilta prototyyppitasoinen järjestelmä, jossa kuitenkin perustoiminnallisuus on yhteneväinen lopullisen tuotantokäyttöön tarkoitetun palvelun kanssa. Tässä luvussa käydään läpi palvelun käyttöä sen käyttäjien näkökulmasta.

### 4.5.1. Sisäänkirjautuminen

Käyttäjän tulee kirjautua sisään omilla, järjestelmään luoduilla, tunnuksillaan ennen kuin hän pääsee tekemään hakuja järjestelmään. Käyttäjätunnuksilla on kaksi eri oikeustasoa. Korkeammalla tasolla olevilla tunnuksilla pääsee hallitsemaan järjestelmää, luomaan uusia käyttäjätunnuksia sekä tuottamaan järjestelmän käyttöraportteja ja tarkastelemaan käyttötilastoja. Alemman tason tunnuksilla pääsee vain suorittamaan hakuja sekä lataamaan haun tulokset. Käyttäjän on joka tapauksessa kirjauduttava sisään ennen kuin pääsee suorittamaan oikeustasonsa mukaisia operaatioita.

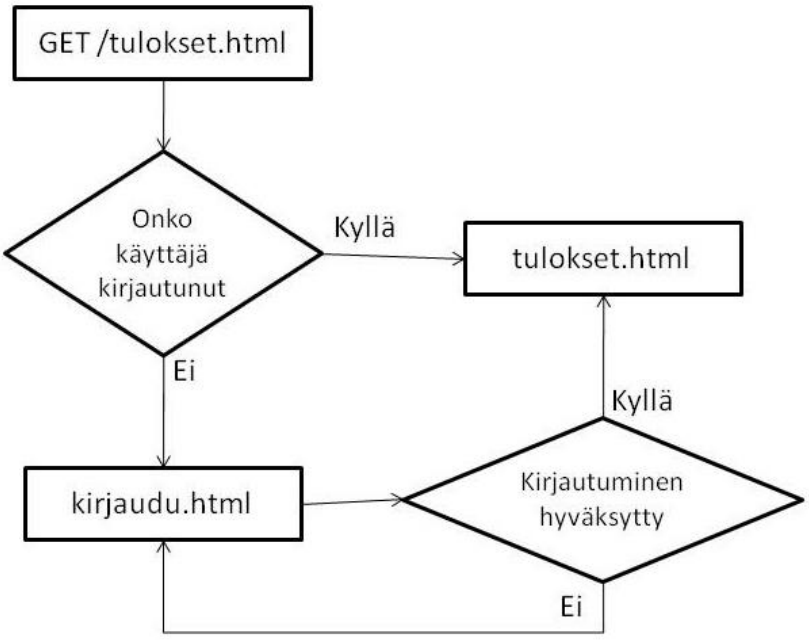
Sisäänkirjautumislomake näytetään mille tahansa järjestelmän sivuille yritettäessä, mikäli käyttäjä ei ole kirjautunut. Käyttäjällä on mahdollisuus kirjoittaa suoraan selaimen osoiteriville haluamansa sivun osoitteen. Tämä sivu saattaa olla järjestelmässä sellainen, jonne ei ole mahdollista päästä ilman sisäänkirjautumista (haku, hakutulokset) tai riittäviä oikeuksia (järjestelmän asetukset, käyttäjähallinta). Jokaisella järjestelmän käyttöliittymän tuottamalla verkkosivulla tulee siis tarkistaa, että käyttäjä on kirjautunut sekä että käyttäjällä on riittävät oikeudet päästä pyytämälleen sivulle.

Kuvassa 16 havainnollistetaan sisäänkirjautumisprosessi. Aluksi käyttäjä näkee tyhjän lomakkeen, johon hän antaa tunnuksensa ja salasanansa. Tämän jälkeen järjestelmä tarkistaa ovatko tunnukset oikein vertaamalla käyttäjätunnus-salasana -paria tietokannassa oleviin (salasanasta vertaillaan tiivisteitä, kuten luvussa 4.3.4 on määritelty) ja mikäli eivät, niin käyttäjä palautetaan samalle lomakkeelle virheilmoituksen kera. Virheellisestä sisäänkirjautumisyrityksestä tallennetaan lisäksi tietokantaan merkintä, jossa on yrityksen tarkka aika, yritetty tunnus sekä IP-osoite koneesta, josta tämä yritys tehtiin.




**Kuva 16: Palvelun sisäänkirjautumisen toimintakaavio.**

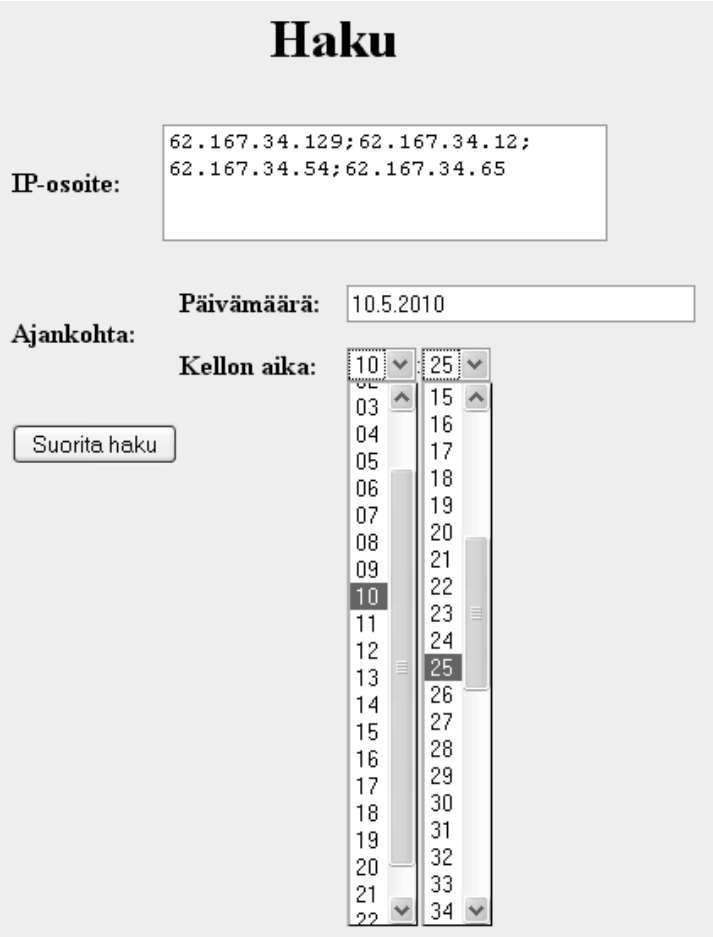
Onnistuneesta sisäänkirjautumisesta tallennetaan tietokantaan vastaavat tiedot kuin epäonnistuneesta. Lisäksi onnistunut kirjautuminen tallennetaan selainistunnon tietoihin. Istunnon tietoja tarkastelemalla voidaan jokaisella sivulla todeta käyttäjän voimassa oleva kirjautuminen. Onnistuneen sisäänkirjautumisen jälkeen käyttäjä ohjataan käyttäjätason mukaan joko hakulomakkeelle, hallinnan etusivulle tai sivulle, jota käyttäjä pyysi ennen kirjautumistaan (Kuva 17).




**Kuva 17: Käyttäjän ohjaus kirjautumisen jälkeen.**

### 4.5.2. **Haku**

Esiehto hakulomakkeelle pääsyyn on käyttäjän onnistunut sisäänkirjautuminen. Hakulomakkeella hakuehtoina ovat IP-osoite ja ajankohta. Minkään muiden tietojen perusteella viranomaiset eivät tietopyyntöä voi tehdä, joten nämä ovat riittävät kriteerit perushaun tekemiseen (Kuva 18).




**Kuva 18: Järjestelmän hakulomake.**

Päivämäärän esitysmuoto voi olla erilainen eri maiden ja järjestelmien välillä. Suomessa käytetään pistenotaatiota järjestyksessä päivä, kuukausi, vuosi. Päivämäärän esitystapoja on kuitenkin lukuisia, joista tietokannat ja muut järjestelmät nykyisin käyttävät ISO 8601 -standardia. Sen mukaisesti päivämäärä esitetään muodossa vuosi-kuukausi-päivä etunollilla varustettuna; esimerkiksi 2010-07-02 [14]. Kuitenkin käytön helpottamiseksi päivämäärän valinta toteutetaan mahdollisuudella valita päivä suoraan kalenterinäkymästä (Kuva 19).


**Kuva 19: Kalenterinäkymä haun päivämäärän valinnassa.**

Järjestelmä lisää päivämäärän käyttäjälle tutuimmassa muodossa hakulomakkeen kenttään (kuva 18) ja päivämäärän käsittelyn yhteydessä päivämäärä muutetaan ISO 8601 -standardin mukaiseksi, jotta sitä on helpompi vertailla kannassa oleviin kerättyihin tunnistetietoihin.

Kellonaika toteutetaan pudotusvalikoilla, joissa tunnit ja minuutit ovat erikseen. Tällä estetään käyttäjää syöttämästä virheellisiä arvoja. Samalla säästetään aikaa myös haun suorittavassa osuudessa, koska aika-kentän arvojen oikeellisuutta ei tarvitse tarkistaa niin täydellisesti, vaan voidaan luottaa esimerkiksi siihen, ettei lomakkeelta tule numeroista poikkeavia arvoja. Normaalistihan kaikki käyttäjän lomakkeelle syöttämät arvot tulisi tarkistaa, ettei järjestelmään päästä vahingossa tai tarkoituksella syöttämään sellaista informaatiota, jolla voitaisiin vaikuttaa järjestelmän toimintaan.

Hakukentistä vain joko IP-osoite tai ajankohta on pakollinen. Tosin, mikäli käyttäjä hakee vain osoitetta rajaamatta sitä mitenkään ajallisesti, niin tuloksena tulee vuoden ajalta kaikki kyseisen osoitteen vuokratapahtumat ja käyttäjät eikä tämä yleensä ole haettu vastaus. Vastaavalla tavalla mikäli haetaan ilman IP-osoitetta vain ajankohdalla, niin järjestelmä palauttaa kaikki IP-osoitteet, jotka kyseisellä ajankohdalla ovat olleet jonkun asiakkaan käytössä. Ajankohdan täyttämisessä sekä päivämäärä että kellonaika ovat pakollisia tietoja. Kellonajan oletusajaksi tosin annetaan 00:00, joten mikäli käyttäjä ei sitä tarkemmin määrittele, niin haku kohdistuu tuohon kellon aikaan.

IP-osoitteeksi on mahdollista syöttää myös useampi osoite samaan aikaan. Tällöin haun tuloksena saadaan tiedot kaikkien annettujen osoitteiden osalta kyseiseltä ajanhetkeltä tai mikäli ajankohtaa ei annettu, niin koko vuodelta. IP-osoitteet tulee syöttää puolipisteellä eroteltuina.

### 4.5.3. **Hakutulokset**

Ennen haun suorittamista ja tulosten näyttämistä tarkistetaan edelleen, että haun tekijä on todellakin kirjautunut sisään. Myös hakutulos-sivulle on mahdollista päästä suoraan tekemällä oikeanlaisen HTTP POST-pyynnön ja mikäli kirjautumista ei jälleen tarkistettaisi, pääsisi tuloksia näkemään täysin ilman tietoa haun tehneestä käyttäjästä. Websivuston lähdekoodista pystytään näkemään lomakkeen kenttien nimet, joiden avulla kenttien tietoon päästään käsiksi lomakkeen käsittelijässä. Lisäksi sieltä näkee mille

palvelinpään sivulle ja mitä HTTP-pyyntöä käyttäen lomakkeen tiedot ohjataan. Näiden tietojen avulla on mahdollista tuottaa HTTP-protokollan mukainen pyyntö suoraan lomakkeen käsittelijälle ilman, että käydään hakusivulla täyttämässä lomake. Tämä onnistuu esimerkiksi telnet-yhteydellä suoraan palvelimen HTTP-protokollan käyttämään porttiin 80. Yhteyden avulla pyydettäisiin lomakkeen käsittelijän mukaista sivua, jolle annettaisiin parametrina lomakkeen kentän nimisiä kenttiä, mutta joiden arvoina voisi olla mitä tahansa tekstidataa.

Esimerkki: Lomake on sivulla lomake.html ja käsittelijä sivulla lomakekasittelija.html. Telnet-yhteys palvelimen porttiin 80 ja pyyntö

```
GET /lomake.html HTTP/1.0
```

tuottaisi vastauksena lomakkeen HTML-muotoillun (Hypertext Markup Language) sivun eli toisin sanoen lomakkeen näkymän lähdekoodin. Olettamalla, että lomakkeella olisi kenttä nimeltä osoite, voitaisiin lomakkeen käsittelijää kutsua vastaavalla tavalla suoraan antamalla riittävät HTTP-pyynnön otsikkotiedot [15]:

```
POST /lomakekasittelija.html HTTP/1.0
Content-Type: application/x-www-form-urlencoded
Content-Length: 18
osoite=10.10.10.10
```

Tällöin palvelin käsittelisi osoite-nimisen kentän, kuten normaalisti lomakkeelta tultaessa, mutta ei voi varmistua siitä, että pyynnön tekijä on todella tehnyt pyynnön lomake.html -sivulta käsin.

Esimerkki ei ole täydellinen kuvaus pakollisista HTTP-protokollan otsikkotiedoista, vaan sen tarkoitus on kuvata mahdollisuutta pyytää tietoja (HTML-sivuja) palvelimelta ilman nettiselaimen kautta tapahtuvaa lomakkeen lähetystä. Vastaavanlaisen pyynnön voi toki tehdä myös silloin, kun käyttäjä on ensin kirjautunut, mutta tämä on yleensä huomattavasti vaikeampaa, koska kirjautuminen on toteutettu selainistuntoon pohjautuen eikä tällöin ole enää oleellista millä tavalla käyttäjä lomakkeen käsittelijäsivulle saapuu, koska hän on joka tapauksessa tunnistettu käyttäjä.

Kun käyttäjän kirjautuminen on varmistettu selaimen istuntotiedoista, näytetään käyttäjälle haun mukaiset tulokset. Käyttäjälle kerrotaan joka tapauksessa, että hänen tekemänsä haun parametrit sekä haun tuottamat tulokset on talletettu järjestelmän lokitietoihin. Mikäli hakuun soveltuvia tuloksia ei löydy, niin käyttäjälle luonnollisesti kerrotaan asiasta (kuva 20).


**Kuva 20: Hakutulos-sivu, kun hakuehtoihin täsmääviä tuloksia ei löytynyt.**

Hakutulokset muotoillaan taulukoksi, jossa on omilla riveillään esitetty jokainen haettu IP-osoite, osoitteen vuokran alku- ja loppuajat sekä asiakas, jolla osoite on tuona ajanhetkenä ollut käytössä. Hakutuloksista käyttäjä voi tuottaa CSV-tiedoston, joka on helppo tulostaa arkistointia tai eteenpäin lähettämistä varten. Esimerkki kuvan 18 hakuehtojen tuottamista kuvitteellisista tuloksista on kuvassa 21. Tuloksiin tulee siis osoitteen vuokran oikea alku- ja loppuaika, joiden väliin hakulomakkeella annettu ajankohta osuu.


**Kuva 21: Esimerkkihakutulokset kuvan 18 mukaisista hakuehdoista.**

## 4.6. Järjestelmän pääkäyttäjän toiminnallisuus

Järjestelmän pääkäyttäjän oikeudet omaavilla henkilöillä on mahdollisuuksia laajempaan toiminnallisuuteen käyttöliittymässä. Heidän tunnuksillaan on mahdollista tehdä täysin samaan tapaan hakuja kuin alemman käyttäjätason tunnuksillakin. Hakujen lisäksi korkeammilla oikeuksilla nähdään järjestelmästä erilaisia tilastoja sekä voidaan koostaa raportteja järjestelmän käytöstä tietyllä ajanjaksolla. Myös uusien käyttäjätunnusten luominen ja olemassa olevien tunnusten muokkaaminen on mahdollista.

Tunnukset on haluttu jakaa kahteen eri luokkaan sen takia, että yleensä järjestelmän ylläpitoa hoitaa vain muutama henkilö ja toisaalta hakutoimenpiteitä saattaa tehdä useampi henkilö, joilla ei ole mitään tarvetta järjestelmän ylläpitoon tai sen asetuksiin. Lisäksi järjestely mahdollistaa hakutunnusten tekemisen esimerkiksi suoraan viran-

omaisille, joiden ei ole tarkoitus päästä tekemään mitään palvelimen tai järjestelmän hallintaan liittyviä toimenpiteitä. Lisäksi järjestelmän käyttäjätietokannasta nähdään näin helposti ja nopeasti kaikki henkilöt, joilla on mahdollisuus tehdä oleellisia muutoksia itse järjestelmään.

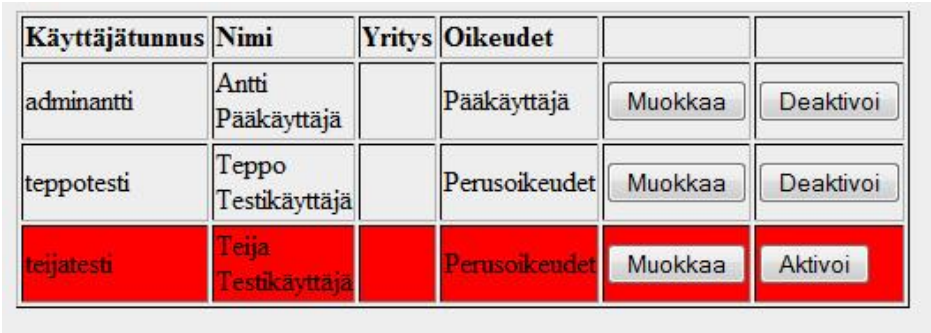
### 4.6.1. **Käyttäjähallinta**

Yksi pääkäyttäjän olennaisimmista tehtävistä on hallita muita järjestelmään pääsijöitä. Pääkäyttäjän selainkäyttöliittymässä tunnusten hallinta on suoraviivaista ja lisäksi kaikkien käyttäjien listaaminen raportointia varten on helppoa. Uuden käyttäjän lisäämisen yhteydessä pitää antaa henkilön nimi, käyttäjätunnus sekä salasana. Lisäksi käyttäjälle pitää määrittää oikeustaso (Kuva 22). Pääkäyttäjä voi siis luoda uusia pääkäyttäjäoikeuksin varustettuja käyttäjiä normaalien käyttäjien lisäksi. Tunnuksen yhteyteen on mahdollisuus lisätä myös käyttäjän yritys. Tämä mahdollisuus on tehty sitä varten, jos halutaan luoda hakutunnuksia myös operaattorin yritysrakenteen ulkopuolisille henkilöille (esimerkiksi viranomaisille kuten edellä mainittiin), eikä yritystiedon syöttäminen ole pakollista tunnuksen luonnin yhteydessä. Salasanan pitää olla minimissään kahdeksan merkkiä pitkä ja sen tulee sisältää sekä isoja että pieniä kirjaimia ja numeroita.


**Kuva 22: Uuden käyttäjän luontilomake.**

Pääkäyttäjä voi myös listata kaikki järjestelmässä käytössä olevat tunnukset, muokata niiden tietoja ja deaktivoida tunnuksia (Kuva 23). Tunnuksia ei voida kokonaan poistaa, koska silloin olisi vaarana, että hukataan tieto siitä millä tunnuksella tietty haku tai muu toimenpide on tehty. Järjestelmän lokitiedot linkittyvät toimenpiteen tehneeseen tunnukseen ja tunnuksen poisto rikkoisi tämän linkin. Deaktivoidut tunnukset voitaisiin poistaa vasta, kun oltaisiin varmoja että kaikki kyseisellä tunnuksella tehdyt toimenpiteet ovat niin vanhoja, ettei niitä enää tultaisi vaatimaan raportteihin tai tutkimuksiin. Tätä aikaa ei kuitenkaan selkeästi ole määritelty, joten tässä vaiheessa

järjestelmästä ei tunnuksia ole mahdollista poistaa käyttöliittymän kautta. Poisto on toki mahdollista suoraan tietokannasta, mikäli päästään käsiksi fyysisesti koneen konsoliin, mutta sitä operaatiota on mahdotonta järjestelmän kannalta tarkkailla [11] ja voidaan olettaa, että järjestelmän konsoliin on pääsy vain harvoilla huoltotoimenpiteitä suorittavilla henkilöillä. Deaktivoiduilla tunnuksilla ei ole mahdollista kirjautua sisään järjestelmän käyttöliittymään. Deaktivoidut tunnukset on mahdollista aktivoida uudelleen, jolloin ne toimivat kuten ennenkin. Pääkäyttäjän ei ole kuitenkaan mahdollista deaktivoida omaa tunnustaan, koska näin voisi olla mahdollista, että järjestelmän ainoa pääkäyttäjä lukitsee tunnuksensa eikä pääkäyttäjän hallintapuolelle olisi enää mahdollista päästä.

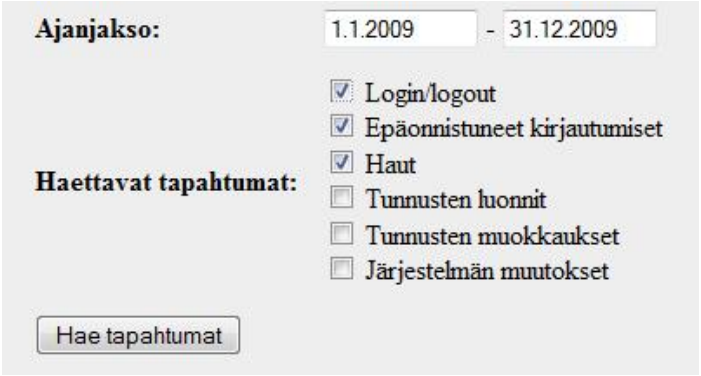



**Kuva 23: Käyttäjätunnusten listaus. Deaktivoidut tunnukset näytetään punaisella pohjavärillä.**

Yleisimmät tunnusten muokkaustarpeet kohdistuvat käyttäjän salasanan resetointiin sekä oikeustason muuttamiseen. Käyttäjätunnusta ei voi myöskään muuttaa, koska se on oleellinen lokitietojen ja käyttäjän linkityksessä (Kuva 11). Salasanan resetointia tarvitaan vain silloin, kun käyttäjä on unohtanut salasanansa. Muuten käyttäjä pystyy itse halutessaan vaihtamaan salasanansa.

### 4.6.2. **Raportointi**

Pääkäyttäjän on mahdollista normaalien tunnistetietojen hakujen tuloksien lisäksi raportoida järjestelmän käytön lokitietoja. Raporttiin lasketaan halutulta ajanjaksolta haluttujen toimenpiteiden määrät sekä listataan toimenpiteet ja niiden tekijät. Tuotettavaa raporttia hallitaan käyttöliittymän lomakkeella (Kuva 24), josta valitaan ajanjakso sekä raporttiin tuotavat toimenpiteet järjestelmässä.

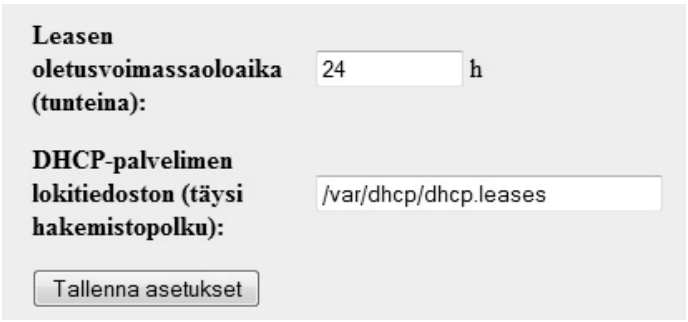



**Kuva 24: Raportin koostamislomake.**

Oletuksena lomakkeella on hakujen raportointi, mikä on viranomaisille luovutettavaa tietoa. Muiden operaatioiden raportoinnilla voidaan seurata esimerkiksi käyttäjien kirjautumistiheyttä tai epäonnistuneita kirjautumisyrityksiä ja tehdä näistä johtopäätöksiä organisaation sisällä. Ajanjakso voidaan valita jälleen samanlaisesta kalenterinäkymästä, kuin haun puolella (Kuva 19), ja aikaväli voidaan valita vapaasti, kunhan alkuaika on aiempi kuin loppuaika.

### 4.6.3. Järjestelmän asetukset

Järjestelmän käyttöliittymän kautta voidaan muuttaa IP-osoitteen vuokran oletusvoimassaoloaikaa DHCP-palvelimen asetusten mukaiseksi (Kuva 25). Järjestelmä siis olettaa että osoitteen vuokra on tietyn mittainen. Mikäli järjestelmän olettama aika on kuitenkin lyhyempi kuin DHCP-palvelimen konfiguraatiossa asetettu vuokra-aika, on mahdollista että järjestelmä luulee osoitteen vuokran loppuneen ennen kuin se oikeasti on loppunut. Tästä syystä järjestelmään tulee päivittää sama osoitteen vuokra-aika kuin DHCP-palvelimen konfiguraatiossa. Näin mahdolliset palvelimen konfiguraatioon tehtävät muutokset eivät aiheuta muutoksia suoraan järjestelmän lähdekoodiin, jotta toiminta säilyisi ennallaan, vaan muutos on mahdollista tehdä dynaamisesti. Lisäksi DHCP-palvelimen lokitiedostojen sijainti voidaan määrittää käyttöliittymän kautta.




**Kuva 25: Järjestelmän asetusten muuttamisnäkymä.**

Järjestelmä ei kuitenkaan tue DHCP-palvelimen fyysisen konfiguraation muuttamista suoraan käyttöliittymän kautta. Konfiguraatiotiedostot sisältävät suhteellisen paljon sellaista tietoa, jota ei missään nimessä tultaisi muuttamaan järjestelmän kautta, joten suoraa muuttamista ei ole nähty tarpeellisena toteuttaa keräysjärjestelmään.

# 5. TULOKSET

Toteutettu järjestelmä on mahdollista asettaa ajoon suoraan DHCP-palvelimeen, jolloin kyseisellä palvelimella pitää olla yhteys tietokantapalvelimeen tai palvelimessa itsessään pitää olla asennettuna ja ajossa jokin tietokantahallintajärjestelmä. Lisäksi palvelimessa tulee olla web-palvelin PHP-kielen tuella, jotta voidaan tarjota järjestelmän käyttöliittymä.

Järjestelmästä voidaan myös erottaa eri koneille keräysosuus ja käyttöliittymäosuus, jolloin DHCP-palvelimelle tulee vain tunnistetietoja keräävä skripti, joka tallentaa tiedot muualla olevaan tietokantaan. Käyttöliittymän tarjoavalla koneella tulee olla pääsy tietokantaan, jonne kerätyt tiedot tallennetaan sekä web-palvelinsovellus. Myös tietokanta on mahdollista hajauttaa, niin että tunnistetiedot tallennetaan omaan kantaansa ja järjestelmän käyttöön liittyvät loki- ja käyttäjätiedot tallennetaan omaansa.

Mikä tahansa onkin lopullinen asennustapa, niin joka tapauksessa käyttöliittymän tulee saada keräyskannasta tunnistetietojen lisäksi asiakastiedot, jotka tunnistetietoihin voidaan liittää. Useimmissa tapauksissa tämä on helpointa toteuttaa yhdellä liittymällä yrityksen asiakastietokantaan, joka sisältää kullekin asiakkaalle määritellyt Circuit- ja Remote-ID:t.

Kuten luvussa 4.3 todettiin, järjestelmä tukee vain Linux/Unix-pohjaista DHCP-palvelinta. Järjestelmän toiminnan hajauttamisen tapauksessa siis keräysjärjestelmän tulee olla Linux-palvelimessa. Käyttöliittymä ja tietokantapalvelimet voivat olla millä tahansa käyttöjärjestelmällä pyöriviä, kunhan käyttöliittymää palvelevassa koneessa on web-palvelin sekä siinä tuki PHP-komentokielelle. Tuki on saatavissa useimpiin yleisimmin [16] käytössä oleviin palvelinsovelluksiin mukaan lukien mm. Apache, Microsoft IIS ja Nginx. Järjestelmän tietokannan toteutus pohjautuu vapaan lähdekoodin MySQL-tietokannan hallintajärjestelmään, mutta kaikki käytetyt tietokantakomennot ovat perus SQL-kieltä, joten tietokannan hallintajärjestelmäkin voidaan valita suhteellisen vapaasti.

# 6. LOPPUPÄÄTELMÄT

Työn tuloksena luodun järjestelmän avulla laajakaistapalveluja tarjoavat operaattorit saavat laajakaistaliittymien tunnistetiedot kerättyä helposti ja automaattisesti. Järjestelmä säästää aikaa ja vaivaa erityisesti silloin, kun viranomaiset pyytävät tunnistetietoja rikostutkinnan tai -epäilyn yhteydessä. Lisäksi järjestelmästä saadaan tärkeitä tietoja esimerkiksi käytössä olleiden IP-osoitteiden määrästä.

Järjestelmän heikkoutena on mahdolliset muutokset tai viallisuudet suoraan DHCP-palvelimen toiminnassa. Suurin heikkous tulee ilmi DHCP-palvelimen tallentamien lokitiedostojen sijainnin muutoksessa, jolloin keräysjärjestelmä ei automaattisesti osaa hakea oikeaa tiedostoa palvelimelta. Tämä on kuitenkin huomioitu järjestelmän käyttöliittymässä, jossa voidaan määritellä palvelimen lokitiedoston täydellinen sijaintipolku. Toinen ongelmia aiheuttava muutos on DHCP-palvelimen vuokraaman osoitteen vuokra-aika. Järjestelmä ei lue vuokra-aikaa suoraan DHCP:n konfiguraatioista, vaan se pitää manuaalisesti vaihtaa oikeaksi käyttöliittymässä.

Järjestelmän jatkokehitysajatuksena on juuri edellä mainittujen haasteiden voittaminen. Tähän voitaisiin päästä tekemällä muutoksia keräysjärjestelmän kautta suoraan DHCP-palvelimen konfiguraatiotiedostoihin. Toinen vaihtoehto olisi, että DHCP-palvelimen toimintaan lisättäisiin tunnistetietojen tallentaminen lokitiedoston sijaan tai lisäksi suoraan tietokantaan. Tällöin keräysjärjestelmästä poistuisi lokitiedostoa lukeva osuus. Tämän etuna olisi myös se, ettei fyysiseen DHCP-palvelimeen tarvitsisi enää asentaa mitään keräysjärjestelmän osia, vaan käyttöliittymät ja tietokannat voitaisiin viedä toisiin koneisiin. Mitään suurta tehoetua näillä toimenpiteillä tuskin saataisiin, koska lokitiedoston lukuosuus on maltillisesti resursseja käyttävä ja toimenpiteiden jälkeen myös DHCP-palvelinsovelluksen toiminnan resurssienkulutus lisääntyisi tietokantayhteyden takia. Tämä saattaisi jopa hidastaa palvelimen toimintaa skriptikäyttöiseen keräysjärjestelmään verrattuna.

# LÄHTEET

1. **Euroopan Parlamentti.** Euroopan parlamentin ja neuvoston direktiivi 2006/24/EY, yleisesti saatavilla olevien sähköisten viestintäpalvelujen tai yleisten viestintäverkkojen yhteydessä tuotettavien tai käsiteltävien tietojen säilyttämisestä ja direktiivin 2002/58/EY muuttamisesta. *Euroopan Unionin virallinen lehti nro L 105.* 13. 4. 2006.

2. **Suomen eduskunta.** Laki sähköisen viestinnän tietosuojalain muuttamisesta. Helsinki, 13. 3. 2009.

3. **Suomen eduskunta.** Sähköisen viestinnän tietosuojalaki. Helsinki, 16. 5. 2004.

4. **Information Sciences Institute, University of Southern California.** RFC 791 - Internet Protocol. *Darpa Internet Program Protocol Specification.* Syyskuu 1981.

5. Microsoft Windows XP - Ping. [Online] Microsoft. [Viitattu: 10. 9. 2009.] http://www.microsoft.com/resources/documentation/windows/xp/all/proddocs/en-us/ping.mspx?mfr=true.

6. **Anttila, Aki.** *TCP/IP-tekniikka.* 1. painos. Juva : WSOY Kirjapainoyksikkö, 2000.

7. **Patrick, M.** DHCP Relay Agent Information Option, RFC 3046, The Internet Society, Tammikuu 2001.

8. Yritysten toimialarakenne TOL 2008 -luokituksen mukaan vuonna 2007. [Online] Tilastokeskus, 10. 2. 2009. [Viitattu: 31. 10. 2009.] http://www.stat.fi/til/syr/2007/02/syr_2007_02_2009-02-10_tie_001.html.

9. Asunnot ja asuinolot. [Online] Tilastokeskus, 26. 5. 2009. [Viitattu: 31. 10. 2009.] http://www.stat.fi/til/asas/index.html.

10. DHCP Server Callout API usage - Microsoft Windows DHCP Team Blog - Site Home - TechNet Blogs. [Online] 6. 7. 2009. [Viitattu: 18. 10. 2010.] http://blogs.technet.com/b/teamdhcp/archive/2009/07/06/dhcp-server-callout-api-usage.aspx.

11. **Viestintävirasto.** Viestintäviraston suositus tunnistamistietojen käsittelyä koskevien tietojen tallentamisesta. *Viestintäviraston julkaisuja.* Viestintävirasto, 24. 11. 2004. s. 6.

12. **Erik, Troan ja Brown, Preston.** Logrotate - System Administrator's Manual. [Online] 5. 11. 2002. http://linuxcommand.org/man_pages/logrotate8.html.

13. **Viestintävirasto.** Tiivistefunktiot. [Online] 27. 9. 2007. [Viitattu: 25. 10. 2010.] http://www.ficora.fi/index/palvelut/palvelutaiheittain/tietoturva/salausmenetelmat/tiivistefunktiot.html.

14. **International Organization for Standardization.** ISO 8601:2004. *Data elements and interchange formats -- Information interchange -- Representation of dates and times.* 2004.

15. **Fielding, R. et. al.** Hypertext Transfer Protocol - HTTP/1.1. *RFC 2616,* The Internet Society, 1999.

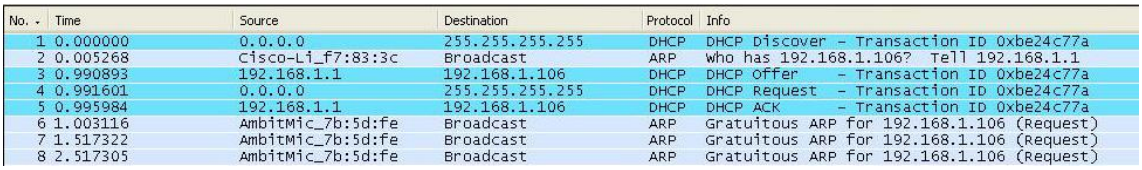
16. **Netcraft Ltd.** January 2010 Web Server Survey. [Online] 7. 1. 2010. [Viitattu: 27. 10. 2010.] http://news.netcraft.com/archives/2010/01/07/january_2010_web_server_survey.html.

# LIITTEET

## *Liite 1*

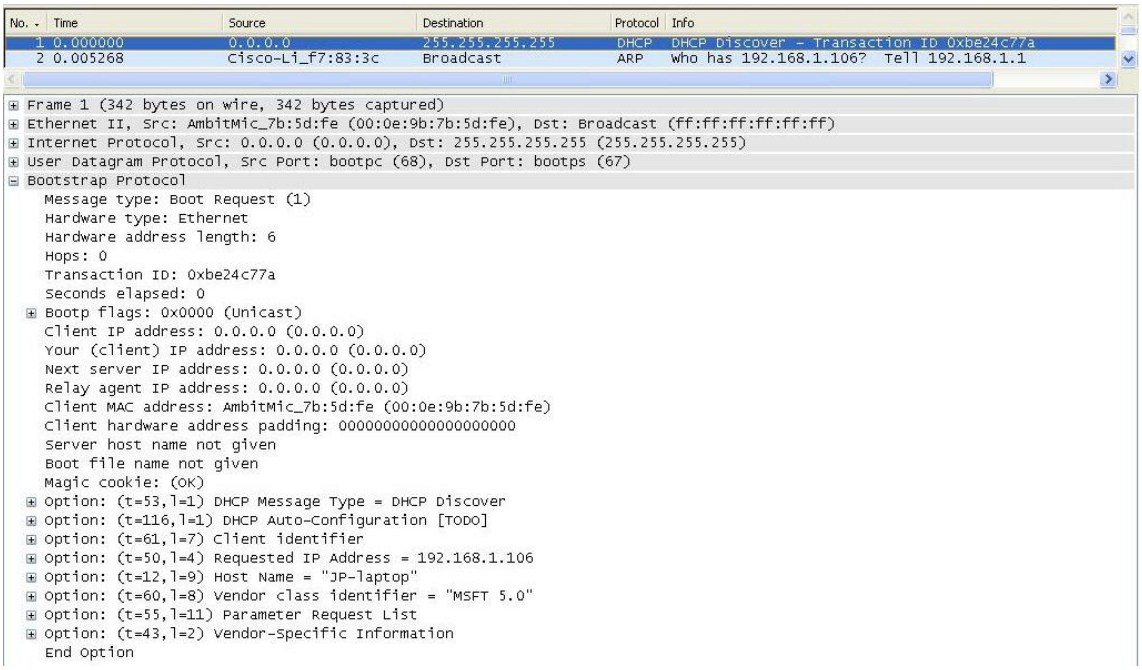
DHCP-protokollan toiminta pakettianalysaattorilla.

**Kuva 1**



| No. | Time | Source | Destination | Protocol | Info |
|---|---|---|---|---|---|
| 1 | 0.000000 | 0.0.0.0 | 255.255.255.255 | DHCP | DHCP Discover - Transaction ID 0xbe24c77a |
| 2 | 0.005268 | Cisco-Li_f7:83:3c | Broadcast | ARP | Who has 192.168.1.106? Tell 192.168.1.1 |
| 3 | 0.990893 | 192.168.1.1 | 192.168.1.106 | DHCP | DHCP Offer - Transaction ID 0xbe24c77a |
| 4 | 0.991601 | 0.0.0.0 | 255.255.255.255 | DHCP | DHCP Request - Transaction ID 0xbe24c77a |
| 5 | 0.995984 | 192.168.1.1 | 192.168.1.106 | DHCP | DHCP ACK - Transaction ID 0xbe24c77a |
| 6 | 1.003116 | AmbitMic_7b:5d:fe | Broadcast | ARP | Gratuitous ARP for 192.168.1.106 (Request) |
| 7 | 1.517322 | AmbitMic_7b:5d:fe | Broadcast | ARP | Gratuitous ARP for 192.168.1.106 (Request) |
| 8 | 2.517305 | AmbitMic_7b:5d:fe | Broadcast | ARP | Gratuitous ARP for 192.168.1.106 (Request) |

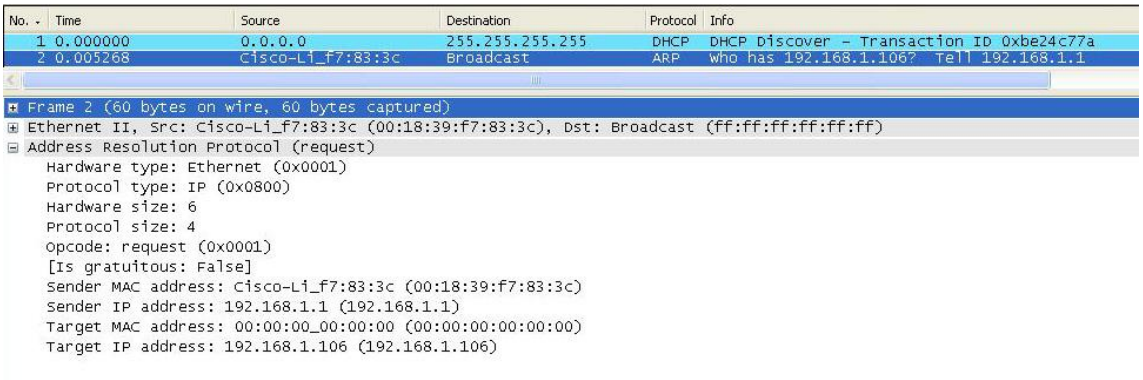
**Kuva 2**



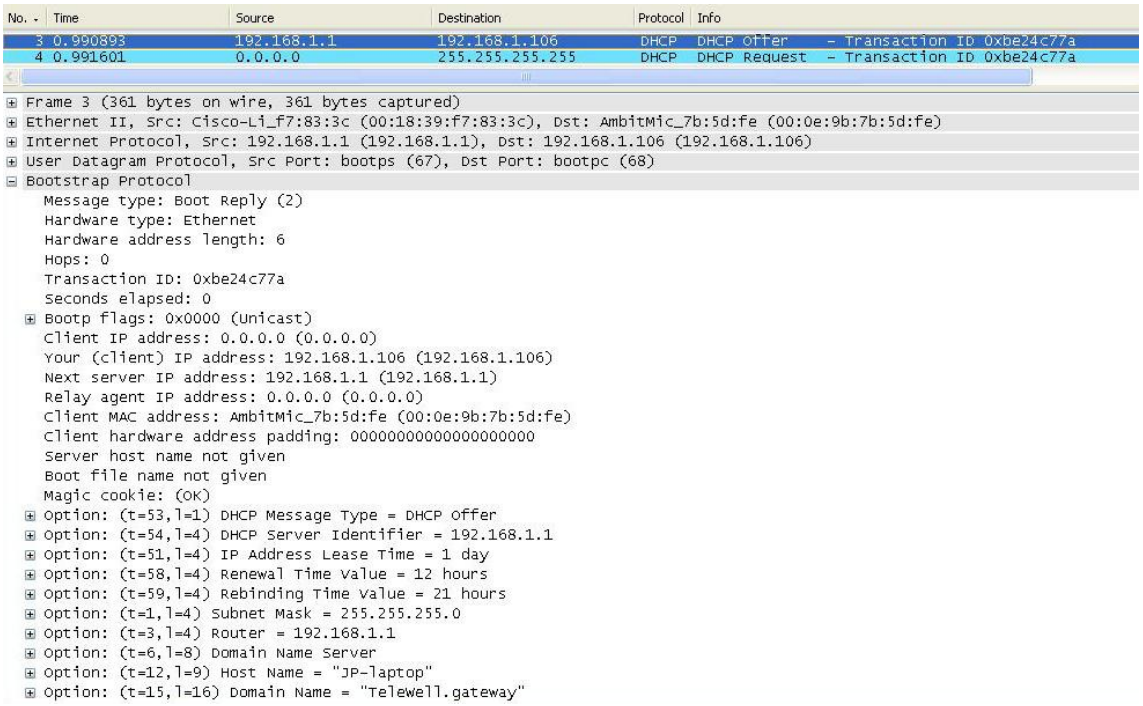
| No. | Time | Source | Destination | Protocol | Info |
|---|---|---|---|---|---|
| 1 | 0.000000 | 0.0.0.0 | 255.255.255.255 | DHCP | DHCP Discover - Transaction ID 0xbe24c77a |
| 2 | 0.005268 | Cisco-Li_f7:83:3c | Broadcast | ARP | Who has 192.168.1.106? Tell 192.168.1.1 |

```
⊞ Frame 1 (342 bytes on wire, 342 bytes captured)
⊞ Ethernet II, Src: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
⊞ Internet Protocol, Src: 0.0.0.0 (0.0.0.0), Dst: 255.255.255.255 (255.255.255.255)
⊞ User Datagram Protocol, Src Port: bootpc (68), Dst Port: bootps (67)
⊟ Bootstrap Protocol
    Message type: Boot Request (1)
    Hardware type: Ethernet
    Hardware address length: 6
    Hops: 0
    Transaction ID: 0xbe24c77a
    Seconds elapsed: 0
  ⊞ Bootp flags: 0x0000 (Unicast)
    Client IP address: 0.0.0.0 (0.0.0.0)
    Your (client) IP address: 0.0.0.0 (0.0.0.0)
    Next server IP address: 0.0.0.0 (0.0.0.0)
    Relay agent IP address: 0.0.0.0 (0.0.0.0)
    Client MAC address: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe)
    Client hardware address padding: 00000000000000000000
    Server host name not given
    Boot file name not given
    Magic cookie: (OK)
  ⊞ Option: (t=53,l=1) DHCP Message Type = DHCP Discover
  ⊞ Option: (t=116,l=1) DHCP Auto-Configuration [TODO]
  ⊞ Option: (t=61,l=7) Client identifier
  ⊞ Option: (t=50,l=4) Requested IP Address = 192.168.1.106
  ⊞ Option: (t=12,l=9) Host Name = "JP-laptop"
  ⊞ Option: (t=60,l=8) Vendor class identifier = "MSFT 5.0"
  ⊞ Option: (t=55,l=11) Parameter Request List
  ⊞ Option: (t=43,l=2) Vendor-Specific Information
    End Option
```

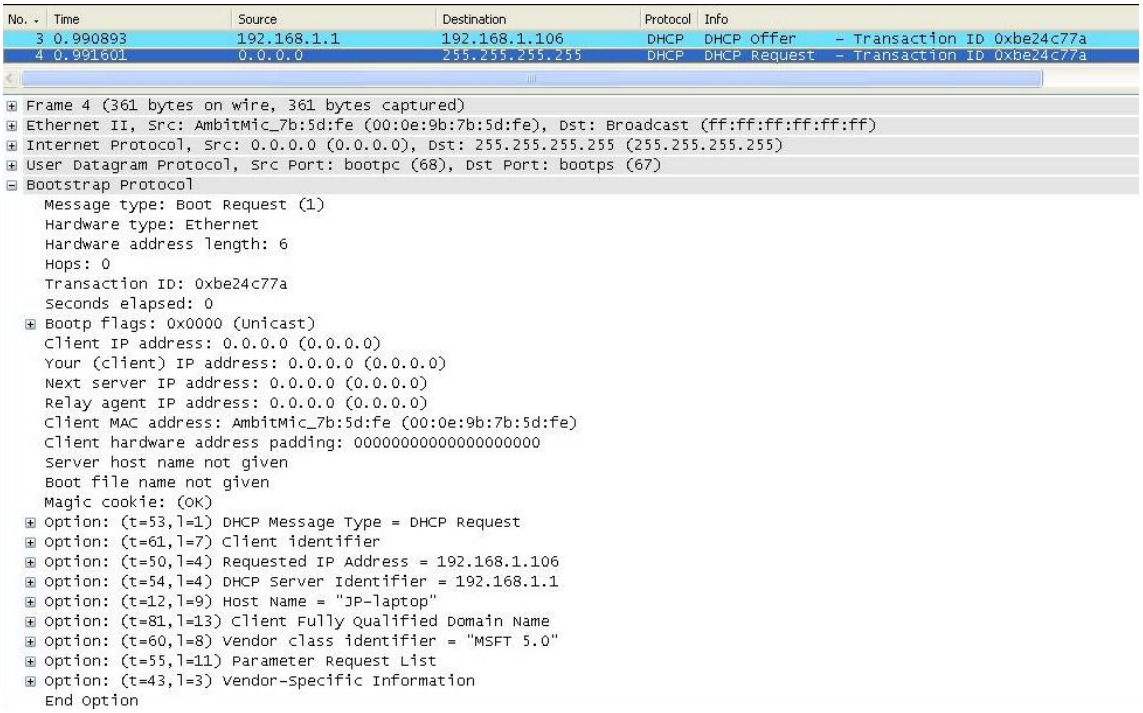
**Kuva 3**


No. Time Source Destination Protocol Info
1 0.000000 0.0.0.0 255.255.255.255 DHCP DHCP Discover - Transaction ID 0xbe24c77a
2 0.005268 Cisco-Li_f7:83:3c Broadcast ARP Who has 192.168.1.106? Tell 192.168.1.1
Frame 2 (60 bytes on wire, 60 bytes captured)
Ethernet II, Src: Cisco-Li_f7:83:3c (00:18:39:f7:83:3c), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
Address Resolution Protocol (request)
Hardware type: Ethernet (0x0001)
Protocol type: IP (0x0800)
Hardware size: 6
Protocol size: 4
Opcode: request (0x0001)
[Is gratuitous: False]
Sender MAC address: Cisco-Li_f7:83:3c (00:18:39:f7:83:3c)
Sender IP address: 192.168.1.1 (192.168.1.1)
Target MAC address: 00:00:00_00:00:00 (00:00:00:00:00:00)
Target IP address: 192.168.1.106 (192.168.1.106)

**Kuva 4**


No. Time Source Destination Protocol Info
3 0.990893 192.168.1.1 192.168.1.106 DHCP DHCP Offer - Transaction ID 0xbe24c77a
4 0.991601 0.0.0.0 255.255.255.255 DHCP DHCP Request - Transaction ID 0xbe24c77a
Frame 3 (361 bytes on wire, 361 bytes captured)
Ethernet II, Src: Cisco-Li_f7:83:3c (00:18:39:f7:83:3c), Dst: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe)
Internet Protocol, Src: 192.168.1.1 (192.168.1.1), Dst: 192.168.1.106 (192.168.1.106)
User Datagram Protocol, Src Port: bootps (67), Dst Port: bootpc (68)
Bootstrap Protocol
Message type: Boot Reply (2)
Hardware type: Ethernet
Hardware address length: 6
Hops: 0
Transaction ID: 0xbe24c77a
Seconds elapsed: 0
Bootp flags: 0x0000 (Unicast)
Client IP address: 0.0.0.0 (0.0.0.0)
Your (client) IP address: 192.168.1.106 (192.168.1.106)
Next server IP address: 192.168.1.1 (192.168.1.1)
Relay agent IP address: 0.0.0.0 (0.0.0.0)
Client MAC address: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe)
Client hardware address padding: 00000000000000000000
Server host name not given
Boot file name not given
Magic cookie: (OK)
Option: (t=53,l=1) DHCP Message Type = DHCP Offer
Option: (t=54,l=4) DHCP Server Identifier = 192.168.1.1
Option: (t=51,l=4) IP Address Lease Time = 1 day
Option: (t=58,l=4) Renewal Time Value = 12 hours
Option: (t=59,l=4) Rebinding Time Value = 21 hours
Option: (t=1,l=4) Subnet Mask = 255.255.255.0
Option: (t=3,l=4) Router = 192.168.1.1
Option: (t=6,l=8) Domain Name Server
Option: (t=12,l=9) Host Name = "JP-laptop"
Option: (t=15,l=16) Domain Name = "TeleWell.gateway"

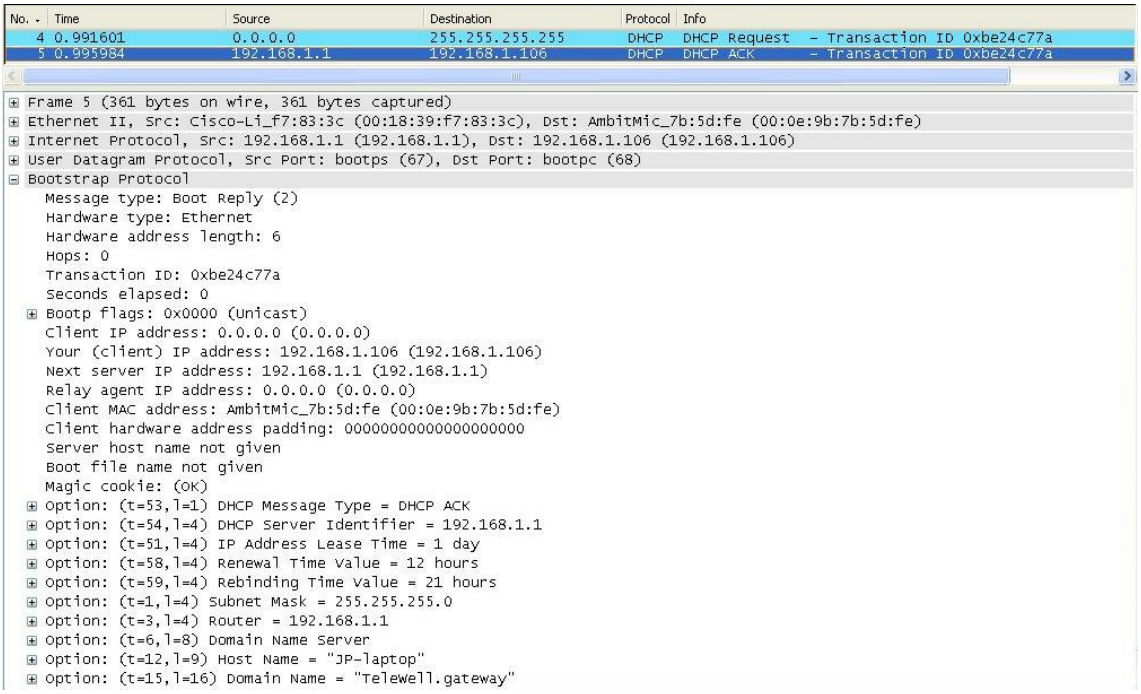
**Kuva 5**



```
No.  Time          Source          Destination     Protocol  Info
  3  0.990893      192.168.1.1     192.168.1.106   DHCP      DHCP Offer    - Transaction ID 0xbe24c77a
  4  0.991601      0.0.0.0         255.255.255.255 DHCP      DHCP Request  - Transaction ID 0xbe24c77a

⊞ Frame 4 (361 bytes on wire, 361 bytes captured)
⊞ Ethernet II, Src: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
⊞ Internet Protocol, Src: 0.0.0.0 (0.0.0.0), Dst: 255.255.255.255 (255.255.255.255)
⊞ User Datagram Protocol, Src Port: bootpc (68), Dst Port: bootps (67)
⊟ Bootstrap Protocol
    Message type: Boot Request (1)
    Hardware type: Ethernet
    Hardware address length: 6
    Hops: 0
    Transaction ID: 0xbe24c77a
    Seconds elapsed: 0
  ⊞ Bootp flags: 0x0000 (Unicast)
    Client IP address: 0.0.0.0 (0.0.0.0)
    Your (client) IP address: 0.0.0.0 (0.0.0.0)
    Next server IP address: 0.0.0.0 (0.0.0.0)
    Relay agent IP address: 0.0.0.0 (0.0.0.0)
    Client MAC address: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe)
    Client hardware address padding: 00000000000000000000
    Server host name not given
    Boot file name not given
    Magic cookie: (OK)
  ⊞ Option: (t=53,l=1) DHCP Message Type = DHCP Request
  ⊞ Option: (t=61,l=7) Client identifier
  ⊞ Option: (t=50,l=4) Requested IP Address = 192.168.1.106
  ⊞ Option: (t=54,l=4) DHCP Server Identifier = 192.168.1.1
  ⊞ Option: (t=12,l=9) Host Name = "JP-laptop"
  ⊞ Option: (t=81,l=13) Client Fully Qualified Domain Name
  ⊞ Option: (t=60,l=8) Vendor class identifier = "MSFT 5.0"
  ⊞ Option: (t=55,l=11) Parameter Request List
  ⊞ Option: (t=43,l=3) Vendor-Specific Information
    End Option
```

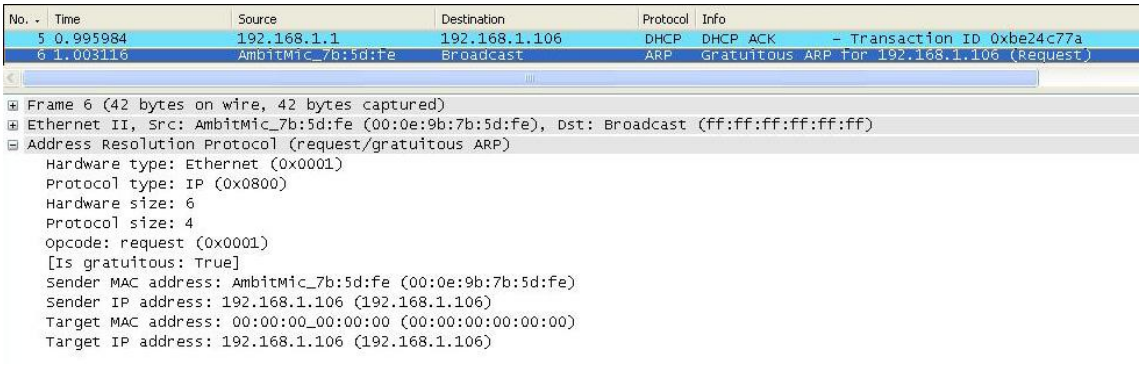
**Kuva 6**



```
No.  Time          Source          Destination     Protocol  Info
  4  0.991601      0.0.0.0         255.255.255.255 DHCP      DHCP Request  - Transaction ID 0xbe24c77a
  5  0.995984      192.168.1.1     192.168.1.106   DHCP      DHCP ACK      - Transaction ID 0xbe24c77a

⊞ Frame 5 (361 bytes on wire, 361 bytes captured)
⊞ Ethernet II, Src: Cisco-Li_f7:83:3c (00:18:39:f7:83:3c), Dst: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe)
⊞ Internet Protocol, Src: 192.168.1.1 (192.168.1.1), Dst: 192.168.1.106 (192.168.1.106)
⊞ User Datagram Protocol, Src Port: bootps (67), Dst Port: bootpc (68)
⊟ Bootstrap Protocol
    Message type: Boot Reply (2)
    Hardware type: Ethernet
    Hardware address length: 6
    Hops: 0
    Transaction ID: 0xbe24c77a
    Seconds elapsed: 0
  ⊞ Bootp flags: 0x0000 (Unicast)
    Client IP address: 0.0.0.0 (0.0.0.0)
    Your (client) IP address: 192.168.1.106 (192.168.1.106)
    Next server IP address: 192.168.1.1 (192.168.1.1)
    Relay agent IP address: 0.0.0.0 (0.0.0.0)
    Client MAC address: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe)
    Client hardware address padding: 00000000000000000000
    Server host name not given
    Boot file name not given
    Magic cookie: (OK)
  ⊞ Option: (t=53,l=1) DHCP Message Type = DHCP ACK
  ⊞ Option: (t=54,l=4) DHCP Server Identifier = 192.168.1.1
  ⊞ Option: (t=51,l=4) IP Address Lease Time = 1 day
  ⊞ Option: (t=58,l=4) Renewal Time Value = 12 hours
  ⊞ Option: (t=59,l=4) Rebinding Time Value = 21 hours
  ⊞ Option: (t=1,l=4) Subnet Mask = 255.255.255.0
  ⊞ Option: (t=3,l=4) Router = 192.168.1.1
  ⊞ Option: (t=6,l=8) Domain Name Server
  ⊞ Option: (t=12,l=9) Host Name = "JP-laptop"
  ⊞ Option: (t=15,l=16) Domain Name = "Telewell.gateway"
```

**Kuva 7**


No. Time Source Destination Protocol Info
5 0.995984 192.168.1.1 192.168.1.106 DHCP DHCP ACK - Transaction ID 0xbe24c77a
6 1.003116 AmbitMic_7b:5d:fe Broadcast ARP Gratuitous ARP for 192.168.1.106 (Request)
Frame 6 (42 bytes on wire, 42 bytes captured)
Ethernet II, Src: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
Address Resolution Protocol (request/gratuitous ARP)
Hardware type: Ethernet (0x0001)
Protocol type: IP (0x0800)
Hardware size: 6
Protocol size: 4
Opcode: request (0x0001)
[Is gratuitous: True]
Sender MAC address: AmbitMic_7b:5d:fe (00:0e:9b:7b:5d:fe)
Sender IP address: 192.168.1.106 (192.168.1.106)
Target MAC address: 00:00:00_00:00:00 (00:00:00:00:00:00)
Target IP address: 192.168.1.106 (192.168.1.106)

## *Liite 2*

### Ote DHCP-palvelimen lokitiedostosta

Oct 11 20:03:26 ns2 dhcpd: Remote-ID for MAC [32:0:21:1d:f8:15] is [teiskontie22-24/5]. IP is [87.94.135.155]
Oct 11 20:03:26 ns2 dhcpd: Circuit-ID MISSING for MAC [32:0:21:1d:f8:15]
Oct 11 20:03:26 ns2 dhcpd: DHCPREQUEST for 87.94.135.155 from 32:00:21:1d:f8:15 via 87.94.128.1
Oct 11 20:03:26 ns2 dhcpd: DHCPACK on 87.94.135.155 to 32:00:21:1d:f8:15 via 87.94.128.1
Oct 11 20:03:27 ns2 dhcpd: Circuit-ID for MAC [0:4:ed:63:b3:64] is [Elisa/173 901289]. IP is [83.102.99.168]
Oct 11 20:03:27 ns2 dhcpd: Remote-ID for MAC [0:4:ed:63:b3:64] is [
Oct 11 20:03:27 ns2 dhcpd: DHCPREQUEST for 83.102.99.168 from 00:04:ed:63:b3:64 via 83.102.96.1
Oct 11 20:03:27 ns2 dhcpd: DHCPACK on 83.102.99.168 to 00:04:ed:63:b3:64 via 83.102.96.1
Oct 11 20:03:27 ns2 dhcpd: Circuit-ID MISSING for MAC [0:80:37:86:ed:fe]
Oct 11 20:03:27 ns2 dhcpd: Remote-ID MISSING for MAC [0:80:37:86:ed:fe]
Oct 11 20:03:27 ns2 dhcpd: DHCPDISCOVER from 00:80:37:86:ed:fe via 87.94.80.1
Oct 11 20:03:27 ns2 dhcpd: DHCPOFFER on 87.94.81.60 to 00:80:37:86:ed:fe via 87.94.80.1
Oct 11 20:03:27 ns2 dhcpd: Remote-ID for MAC [fe:0:19:8:20:3b] is [MAKI28-30/1/4]. IP is [83.102.27.157]
Oct 11 20:03:27 ns2 dhcpd: Circuit-ID MISSING for MAC [fe:0:19:8:20:3b]
Oct 11 20:03:27 ns2 dhcpd: DHCPREQUEST for 83.102.27.157 from fe:00:19:08:20:3b via 83.102.16.1
Oct 11 20:03:27 ns2 dhcpd: DHCPACK on 83.102.27.157 to fe:00:19:08:20:3b via 83.102.16.1
Oct 11 20:03:27 ns2 dhcpd: Circuit-ID for MAC [0:4:ed:67:4e:c6] is [Elisa/1168/901786]. IP is [83.102.96.92]
Oct 11 20:03:27 ns2 dhcpd: Remote-ID for MAC [0:4:ed:67:4e:c6] is [
Oct 11 20:03:27 ns2 dhcpd: DHCPREQUEST for 83.102.96.92 from 00:04:ed:67:4e:c6 via 83.102.96.1
Oct 11 20:03:27 ns2 dhcpd: DHCPACK on 83.102.96.92 to 00:04:ed:67:4e:c6 via 83.102.96.1
Oct 11 20:03:27 ns2 dhcpd: Remote-ID for MAC [fe:0:39:18:0:4a] is [HER2/1/A01068]. IP is [87.94.134.215]
Oct 11 20:03:27 ns2 dhcpd: Circuit-ID MISSING for MAC [fe:0:39:18:0:4a]
Oct 11 20:03:27 ns2 dhcpd: DHCPREQUEST for 87.94.134.215 from fe:00:39:18:00:4a (RAS) via 87.94.128.1
Oct 11 20:03:27 ns2 dhcpd: DHCPACK on 87.94.134.215 to fe:00:39:18:00:4a (RAS) via 87.94.128.1
Oct 11 20:03:28 ns2 dhcpd: Remote-ID for MAC [fe:0:31:4c:0:27] is [HER/2/035383]. IP is [87.94.152.248]
Oct 11 20:03:28 ns2 dhcpd: Circuit-ID MISSING for MAC [fe:0:31:4c:0:27]
Oct 11 20:03:28 ns2 dhcpd: DHCPREQUEST for 87.94.152.248 from fe:00:31:4c:00:27 via 87.94.128.1
Oct 11 20:03:28 ns2 dhcpd: DHCPACK on 87.94.152.248 to fe:00:31:4c:00:27 via 87.94.128.1