



Aalto-yliopisto
Perustieteiden
korkeakoulu

Verkkokerros 2: Reititys

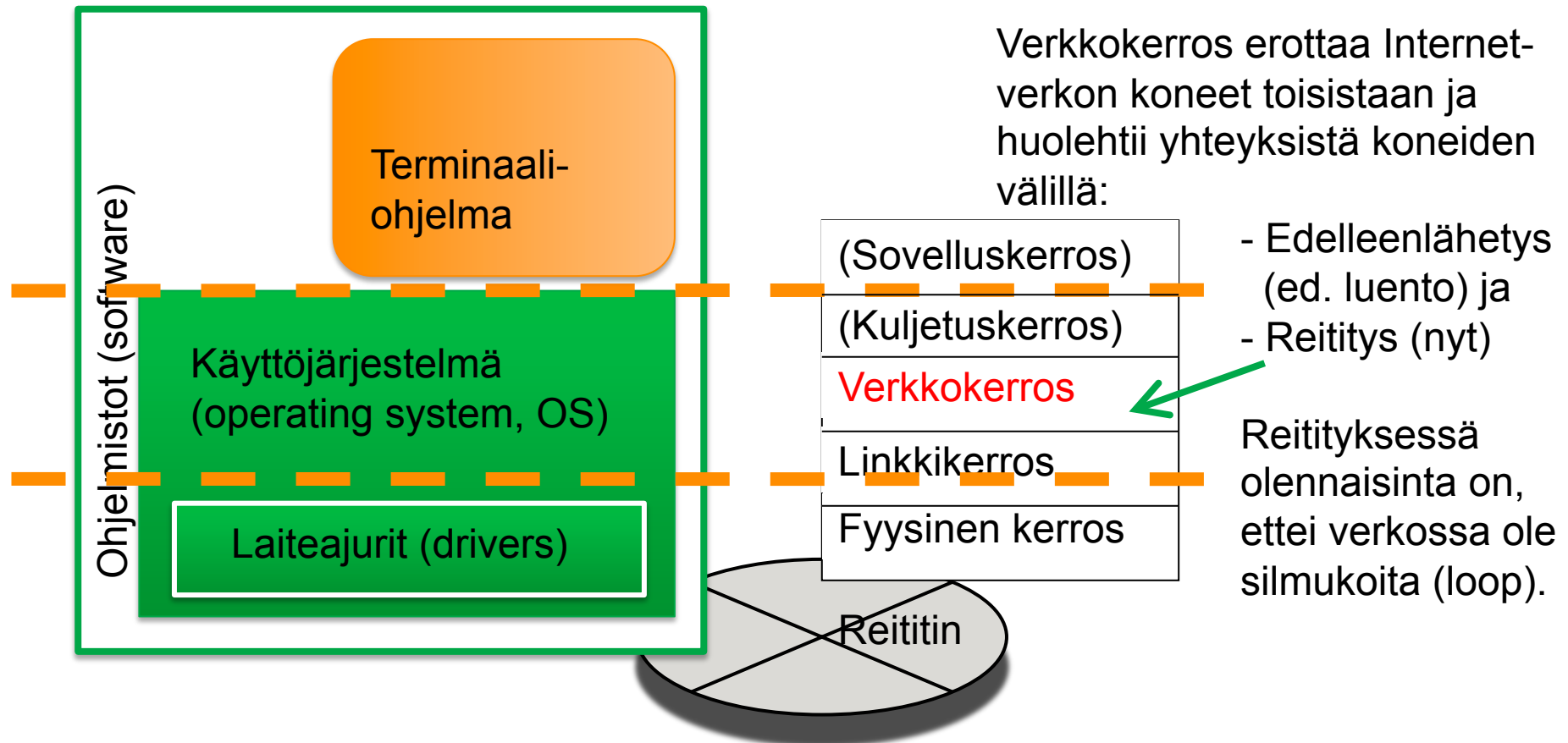
CSE-C2400 Tietokoneverkot
Kirjasta 4.2-4.3, 4.5-4.8

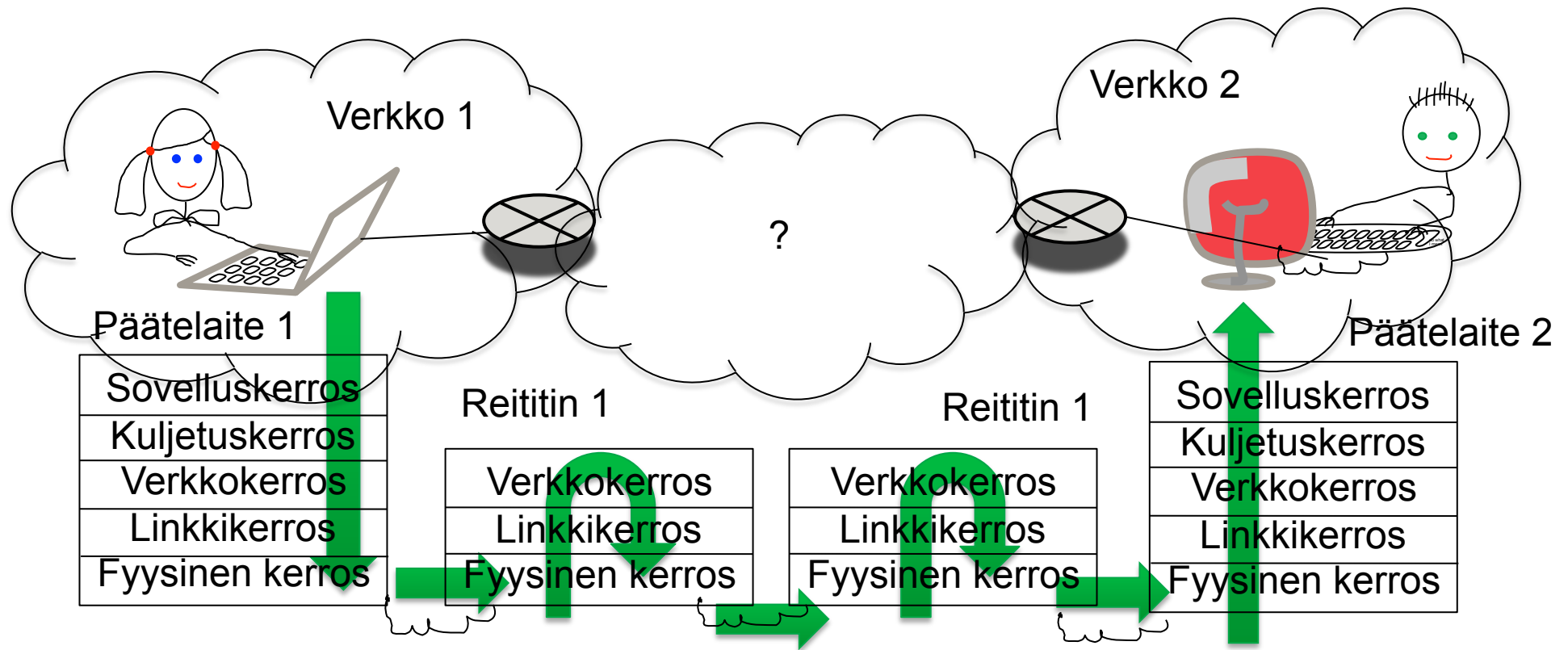
Sanna Suoranta

Verkkokerros

- Tällä luennolla
 - Reititys Internet-verkossa ja internet-verkoissa – Internetin rakenne
 - Reititysprotokollat ja algoritmit
 - Reitittimen toiminta
 - Yleislähetys (broadcast) ja monilähetyksen (multicast) reititys

Internet-protokollapino ja verkkokerroksen tehtävä





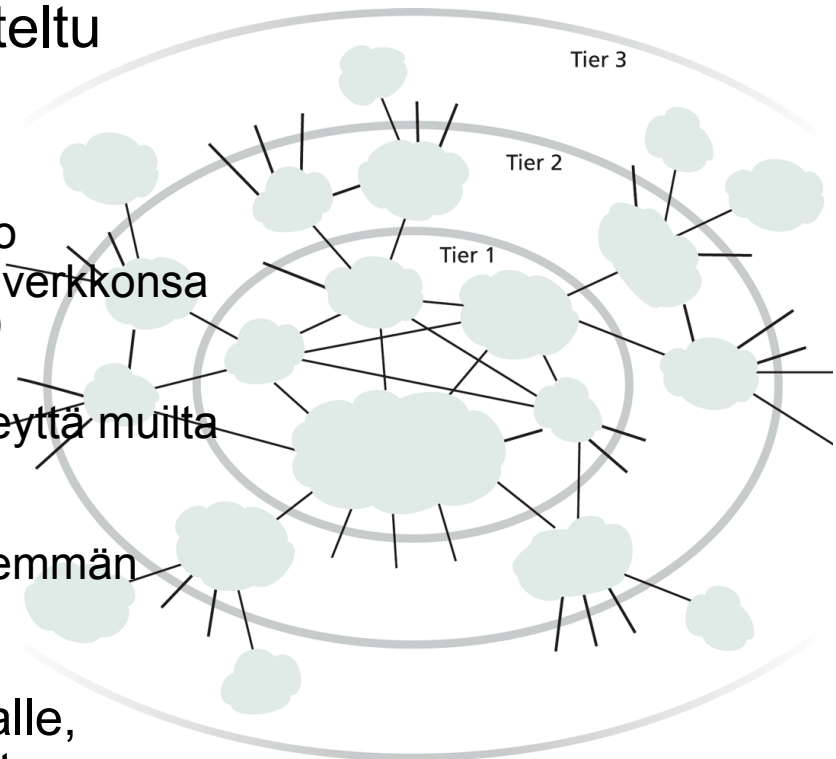
- Sovelluskerroksen data pakataan kuljetuskerroksen tietosähkeeseen, joka pakataan verkkokerroksen pakettiin
- Reitittimet katsovat edelleenlähetystaulustaan, mihin rajapintaan paketti lähetetään seuraavaksi
 - Reitittimet ovat siis verkkokerroksen laitteita

Reititys – termejä

- Reititys: polun löytäminen osoitteesta toiseen
- Reititysprotokolla: reitittimien tapa vaihtaa tietoa verkon tilasta ja verkoista (osoitteista), joihin reitittimellä on yhteys
- Reititysalgoritmi: Algoritmin avulla lasketaan parhaat reitit verkon läpi perustuen reititysprotokollan avulla saatuun tietoon
- Reititystaulu: Tiedot, mille reitittimelle saapunut paketti lähtetään seuraavaksi ja kuinka pitkä reitti on
- Reititin: laite, joka huolehtii reitityksestä ja edelleenlähetyksestä

Internetin rakenne

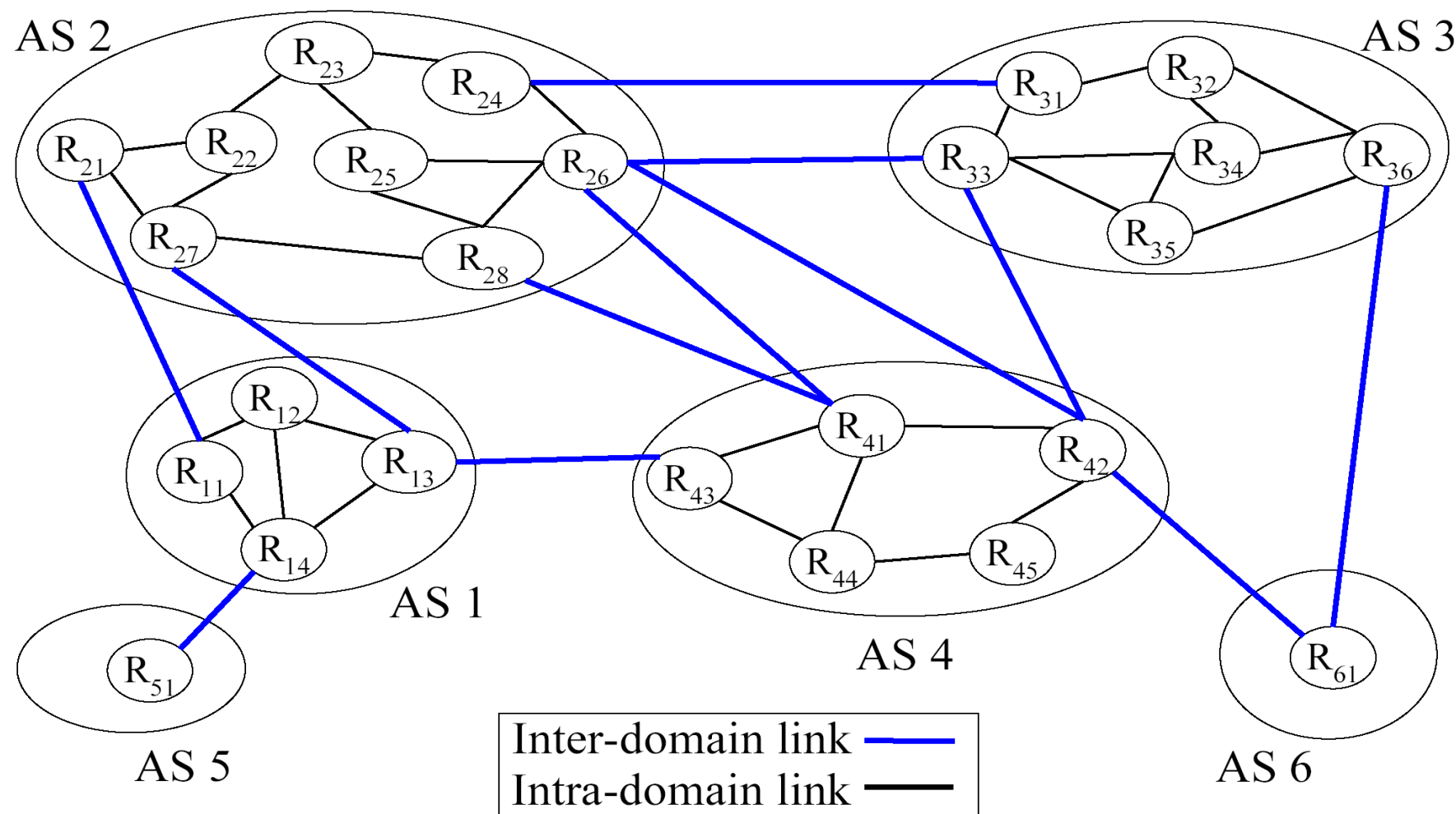
- Internet-yhteydentarjoajat (ISP) jaoteltu kolmeen luokkaan
 - tier 1: globaali
 - Vähän yli kymmenen kappaletta
 - Internetin “selkäranka” eli runkoverkko
 - Sallivat toistensa liikenteen maksutta verkkonsa kautta (a.k.a. settlement free peering)
 - tier 2: alueellinen
 - Myös peering, lisäksi myös ostaa yhteyttä muilta (transit palvelu)
 - tier 3: lokaali
 - Yksinomaan ostaa yhteyttä muilta (ylemmän kategorian) ISP:lta
- Toisenlainen jaottelu:
 - Reunaverkot (stub), yksi yhteys muualle,
 - monikotiset (multihomed) reunaverkot, useampi yhteys muualle, mutta ei tarjota muille
 - Transit-verkot tarjoavat läpikulkua toisiin verkkoihin



Internet koostuu autonomisista järjestelmistä (autonomous system, AS)

- AS on joukko verkko-osoitteita, joita hallinnoi yksi organisaatio, joka ylläpitää yhtenäistä reitityspolitiikkaa
- Jokaisella on tunnisteena 32-bittinen AS-numero (ASN)
- ASien välillä käytetään EGP-reititystä (exterior gateway protocol), Internetissä Border Gateway Protocol (BGP)
 - BGP:llä välitetään saavutettavuustietoa verkkojen välillä
 - Saavutettavuus riippuu politiikasta eli sopimuksista AS:ien välillä
 - reitti ei välttämättä ole lyhin tai nopein!
- Runkoverkon reitittimillä ei ole oletusreittejä, vaan kaikki reitittimet tietävät, mille reitittimelle paketti seuraavaksi ohjataan

Internetin rakenne



Border Gateway Protocol v4 (BGP4)

- BGP-reunareitittimet pitävät tietoa muiden verkkojen saavutettavuudesta external BGP:llä (eBGP)
- Jos AS:llä on useampi yhteys, sen BGP-reitittimet vaihtavat tietoa toistensa kanssa interior BGP:llä (iBGP)
 - Ulos mainostettava yhteinen käsitys kokonaisuudesta säilyy
 - Reunareititin toimii oletusreitittimenä tai, jos reunareitittimiä on useita, reunareitittimet mainostavat reittejä ulosverkosta oman verkon muille reitittimille sisäverkon reititysprotokollien avulla
- Verkkojen tunnisteenä niiden prefix
 - Useampia verkkoja voidaan mainostaa kerrallaan
- Polku Internetin läpi ilmaistaan AS-numeroiden avulla
 - Seuraavan hypyn tunnisteenä naapurin AS-numero, sitten muut AS:t, joiden läpi polu kulkee

Reitin valinta BGP:ssä

- Edelleenlähetyksessä verkkoprefixien perusteella
 1. Paikallinen priorisointi verkonhaltijan asetusten mukaan
 2. Lyhin AS-polku
 3. Jos yhtä pitkät polut, lähin reunareititin AS:n sisällä
- Reititystaulun päivityksessä:
 - Sopimuspohjaista, minkä toisen liikennettä välitetään ja millä ehdoilla (engl. Policy)
 - Reunaverkot (stub) eivät mainosta reittejä kuin omaan verkkoon vaikka olisivatkin monikotisia – läpikulkuliikenteestä olisi kuluja

BGP:n toiminta

- BGP-reitittimet avaavat TCP-yhteyden toisiinsa (portti 179) ja pitävät sen auki koko ajan keep-alive-viesteillä
 - Reitti on käytettävissä, kun viestit kulkevat
 - Jos keep-alive-viestiä ei saada, reitit muualle poistetaan
- Muutoksista vaihdetaan update-viestejä, joissa tietoa uusista ja poistuneista reiteistä
 - Reittejä voidaan yhdistää yhdeksi mainostukseksi ja siten yhdeksi reititystaulun riviksi (route aggregation)

AS:n sisäinen reititys voidaan järjestää monin tavoin

- Staattinen reititys: ylläpitäjä lisää reitin käsin reititystauluun
 - Oletusreitti lisätään käsin esim. työasemiin
- Dynaaminen reititys reititysprotokollien avulla
 - Tieto paikallisista verkoista eli suoraan reitittimeen kytketyistä lähiverkoista lisätään käsin reitittimen reititystauluun
 - Reitittimet mainostavat toisilleen omia verkkoja reititysprotokollan avulla eli muutokset huomioidaan automaattisesti
 - Reittimainostuksista rakennetaan reititystaulu, jonka perusteella sitten verkossa kulkeville paketeille löydetään paras reitti (edelleenlähetyksessä)
- IGP-reititykseen useita protokollia tarjolla
 - RIP, OSPF, EIGRP (Ciscon oma), IS-IS,...

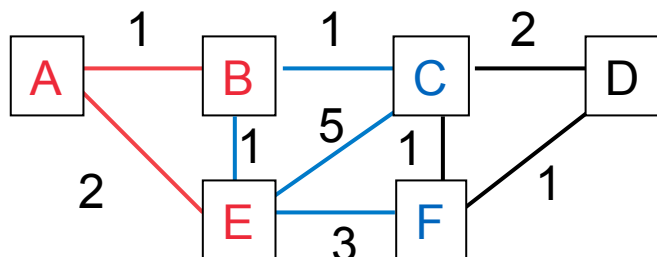
Routing Information Protocol (RIP) – Etäisyysvektori (distance vector) -reititys

- Yksinkertainen, sopii pieniin AS-verkkoihin
 - Etäisyyden metriikkana käytetään hyppyjen lukumäärää, max 15 hyppyä
 - Ei kokonaiskuvaa verkosta, reititystaulu lasketaan naapureiden kertomien etäisyysvektorien perusteella iteratiivisesti
- RIPin toiminta (yksinkertaistettuna)
 - Naapureiden (=RIP-reitittimet) selvittäminen monilähetyksellä 244.0.0.9 (tai yleislähetyksenä) UDP-porttiin 520
 - Reitittimet vaihtavat reititystaulunsa naapurien kanssa joka 30 sekunti RIP Advertisement -viesteillä
 - Lasketaan parhaat reitit perustuen naapurien ilmoituksiin
- Reititystaulussa: <kohde, seuraava hyppy, etäisyys>
- RIPv1 RFC 1058 ja RIPv2 RFC 2453

Etäisyysvektoreititysalgoritmi (Bellman-Ford-algoritmi)

0. Merkitse etäisyydet kaikille noodin suorille naapureille, muille ∞
1. Lähetä naapureille kaikki tiedot reititystaulun sisällöstä
 - <kohde, seuraava hyppy, etäisyys>-vektoreina
2. Laske naapureilta saatujen etäisyysvektorien perusteella, muuttuiko mikään
 - jos löytyi uusi lyhyempi reitti, päivitä se reititystauluun
 - reitti voi olla parempi kuin aiemmin tiedetty tai uuteen, ennen tuntemattomaan verkon noodiin
3. jatka kohdasta 1
 - Jossain vaiheessa mikään ei vaan enää muutu, jollei jokin yhteyksistä katkea tai tule lisää, joten algoritmi lopettaa itse (reititysprotokolla toki jatkaa kokoajan tietojen vaihtoa naapureiden kanssa)
- Algoritmi on asynkroninen, iteratiivinen, itselopettava ja hajautettu

Etäisyysvektori - Esimerkki



A:n taulu	B	C	D	E	F
0. kierros	1, -	∞	∞	2, -	∞
1. kierros	1, -	2, B	∞	2, -	5, E
2. kierros	1, -	2, B	4, B	2, -	3, B

Alkutila: suorat naapurit

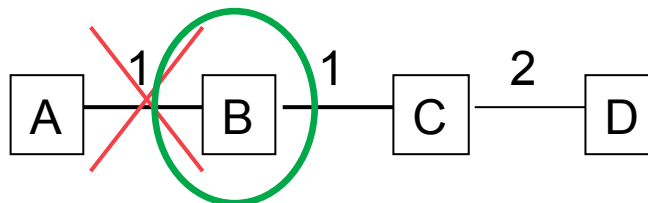
B:n taulu		E:n taulu	
A	1	A	2
C	1	B	1
E	1	C	5
		F	3

1. kierroksen jälkeen:

B:n taulu		E:n taulu	
A	1	A	2
C	1	B	1
D	3, C	C	2, B
E	1	D	4, F
F	2, C	F	3

- Hyvät uutiset eli uudet reitit leviävät nopeasti

Etäisyysvektori - Ongelmat



- Reitin katkeamistieto leviää hitaasti
 - B kuvittelee, että C:llä on vaihtoehtoinen reitti A:han
- Äärettömyyteen laskemisen ongelma (count- to- infinity problem)
- Ratkaisut:
 - Jaettu horisontti (split horizon): reittiä ei mainosteta takaisinpäin sinne, mistä tieto saatiin
 - Lähteen myrkytys (poisoned reverse): ilmoitetaan metriikaksi ääretön sille kohteelle, josta reitti saatiin
 - Tietojen nopea välitys (triggered updates): välitetään muuttunut tieto heti naapureille

	B	C	D
A	1	2	4
B	∞	2	4
C	3	2	4
D	3	4	4
E	5	4	6
F	...	...	...
G	∞	∞	∞

Open Shortest Path First (OSFP) – Linkkitila (link state) -reititysprotokolla

- Sopii laajoihin (esim. ISP) verkkoihin
 - Verkon voi jakaa hierarkisesti alueisiin (area), joista huolehtii jokin edustajareititin (designated router)
- Jokainen reititin rakentaa itselleen kuvan koko AS:n verkon topologiasta
- Viestit lähetetään suoraan IP:n päällä, OSPF:n protokollanumero on 89, monilähetysosoitteeseen 224.0.0.5
- OSPF:n toiminta (yksinkertaistettuna) – kolme osaa
 - Hello-protokolla naapureiden tavoitettavuuden selvittämiseksi säännöllisin väliajoin
 - Exchange-protokolla yhteyden aluksi naapurien reititystaulujen saamiseksi, käytetään oman reititystaulun laskemiseen
 - Flooding-protokolla tilatietojen päivittämiseen säännöllisin väliajoin
- OSPFv2, RFC 2328

Linkkitila - Dijkstran lyhimmän polun algoritmi

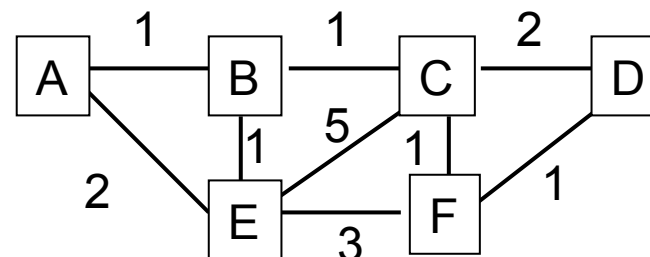
0. Polku itseen = 0 ja muihin ∞ , merkitse noodit vierailemattomiksi, aloita tarkastelu nykynoodista (A itse)

1. Laske etäisyys nykynoodista naapureihin,

- jos lyhyempi kuin tiedossa (huom: laske koko polku), vaihda
- merkitse noodi vierailluksi

2. Jos jäljellä on vierailemattomia noodeja, siirry niistä lähimpään ja jatka kohdasta 1

- jos kaikissa noodeissa on jo vierailtu, lopeta



kier-ros	noo-di	B	C	D	E	F
1.	A	1	∞	∞	2	∞
2	B	1	2	∞	2	∞
3.	E	1	2	∞	2	5
4.	C	1	2	4	2	3
5.	F	1	2	4	2	3
6.	D	1	2	4	2	3

Linkkitila - Ongelmat

- Sarjanumeroiden ylivuodot
 - sarjanumeroilla pyritään estämään vanhan tiedon käyttöä
- Reitittimen kaatuminen
 - raskasta aloittaa alusta
- Melkoisen monimutkainen, mutta sopii sitten isonkin AS:n tarpeisiin

Reititysprotokollien turvallisuus

- AS:ien välinen reititys ja BGP: millä perusteella on oikeus mainostaa reittiä?
 - RFC 6483 route origin authorization, allekirjoituksin toteutettu
- AS:ien sisäinen reititys ja OSPF:
 - Reitittimet voidaan tunnistaa jaetun salaisuuden ja erikseen jaettujen avainten avulla, tai selväkielisen salasanan avulla
- AS:ien sisäinen reititys ja RIP:
 - Alunperin ei tunnistettu mitenkään
 - Myöhemmin lisätty mahdollisuus käyttää muuta kanavaa pitkin jaettua salaista avainta

Reititys perustuu verkkojen IP-osoitteisiin

- Kun paketti saapuu reitittimelle, sen osoitetta verrataan reititystauluun tallennettuihin verkkojen osoitteisiin (network prefix)
 - Jotta reititystaulut eivät paisu jättimäisiksi, saman verkkorajapinnan reittejä voidaan yhdistellä (route aggregation)

Destination	Next Hop
30.1.0.0/24	R3
30.1.2.0/24	direct
30.2.1.0/24	direct
30.3.1.0/24	R3
20.2.0.0/16	R2
20.1.1.0/28	R2

Destination	Next Hop
30.1.0.0/24	R3
30.1.2.0/24	direct
30.2.1.0/24	direct
30.3.1.0/24	R3
20.0.0.0/8	R2

Pisin soveltuva sääntö valitaan (Longest Prefix Matching Rule)

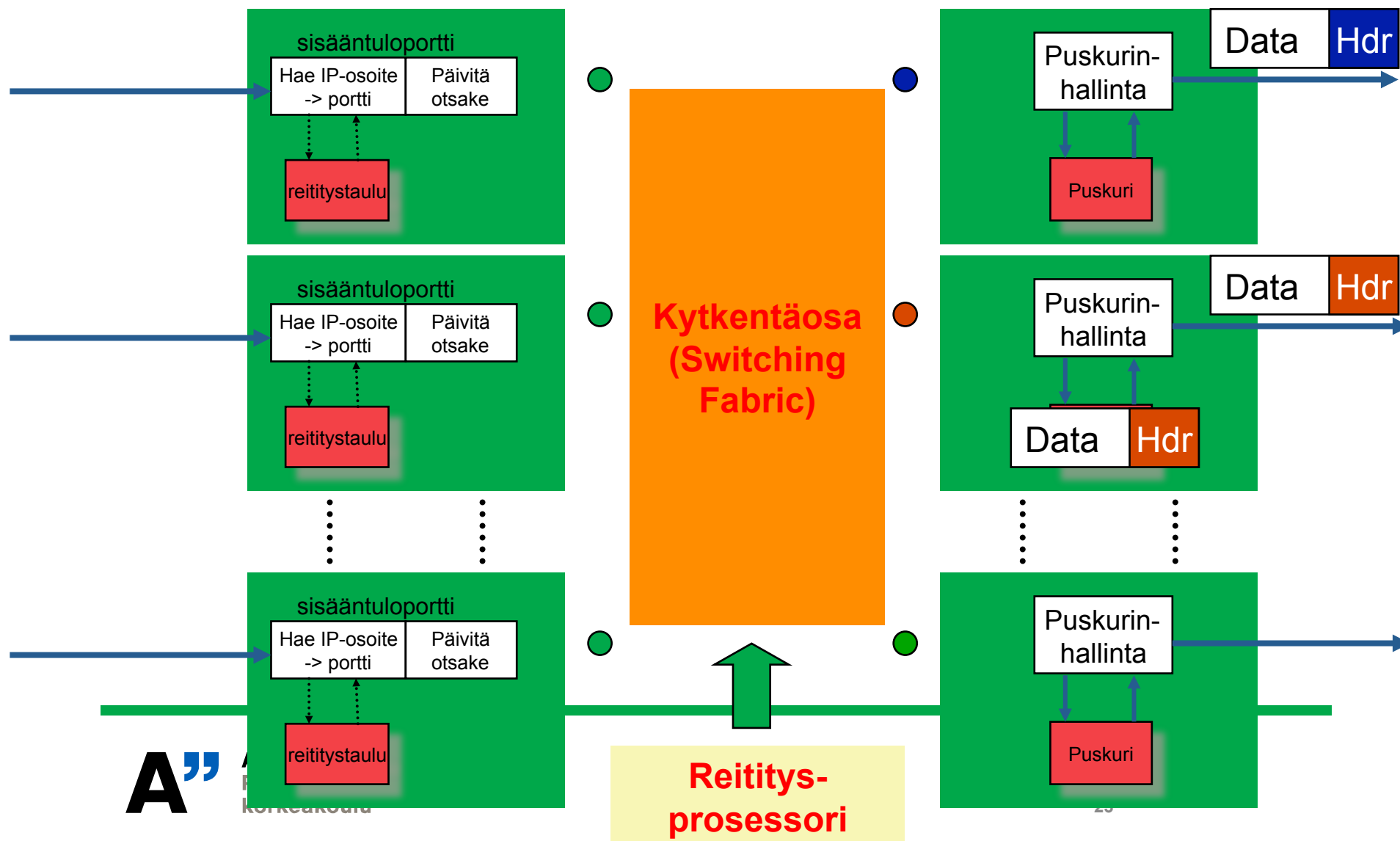
Destination	Next Hop
200.223.0.0/16	R2
200.16.0.0/13	R4
200.22.0.0/15	R1

- Mihin tämä paketti lähetetään?

Kohdeosoite on: 200.23.146.51

11001000 00010111 10010010 00110011

Reitittimen arkkitehtuuri



Useampi vastaanottaja kerralla?

- Usein olisi tehokkaampaa, jos verkko monistaisi paketit mahdollisimman lähellä vastaanottajia (eikä lähettäjä tekisi jokaiselle vastaanottajalle omaa pakettia erikseen)
 - Pohjautuvat yksilähetyksen (unicast) reititykseen (luento tähän asti)
- Yleislähetys (broadcast): “kaikille” verkossa
 - Paitsi sille, jolta viesti saatiin
 - Tulviminen (flooding): lähetetään kaikille -> yleislähetysmyrsky (broadcast storm)
 - Reverse path forwarding (RPF): lähetetään eteenpäin, vain jos itse on lyhimmällä polulla lähettäjään nähden
 - Aika vähän käytetty verkkokerroksella (linkkikerroksella lisää), IPv6 ei tue lainkaan
- Monilähetys/ryhmälähetys (multicast): tietylle joukolle vastaanottajia, verkko monistaa viestin

Monilähetysosoitteet

- Monilähetysosoite toimii monilähetysryhmän tunnisteena
 - Ei kerro, missä ryhmän jäsenet sijaitsevat
 - IPv4: 224.0.0.0/4 ja IPv6: alkaa kahdeksalla 1-bitillä
- Kuka tahansa voi lähettää viestin, jonka vastaanottaja on monilähetysryhmä
 - Mutta myös tapa rajoittaa ryhmäkohtaisia lähettäjiä
- Sekä pysyviä että dynaamisesti jaettavia osoitteita
 - IANA allokoit pysyvät, esim 224.0.0/24-osoitteet varattu paikallisverkon kontrolliin
- Sekä vain paikalliseen verkkoon että Internetin laajuisesti välitettäviä osoitteita
 - IPv6:ssa osoitteen lisähierarkiana organisaation verkko
 - IPv4:ssä TTL tai RFC 2365:n avulla IPv6:n kaltainen

Monilähetys lähettäjän näkökulmasta

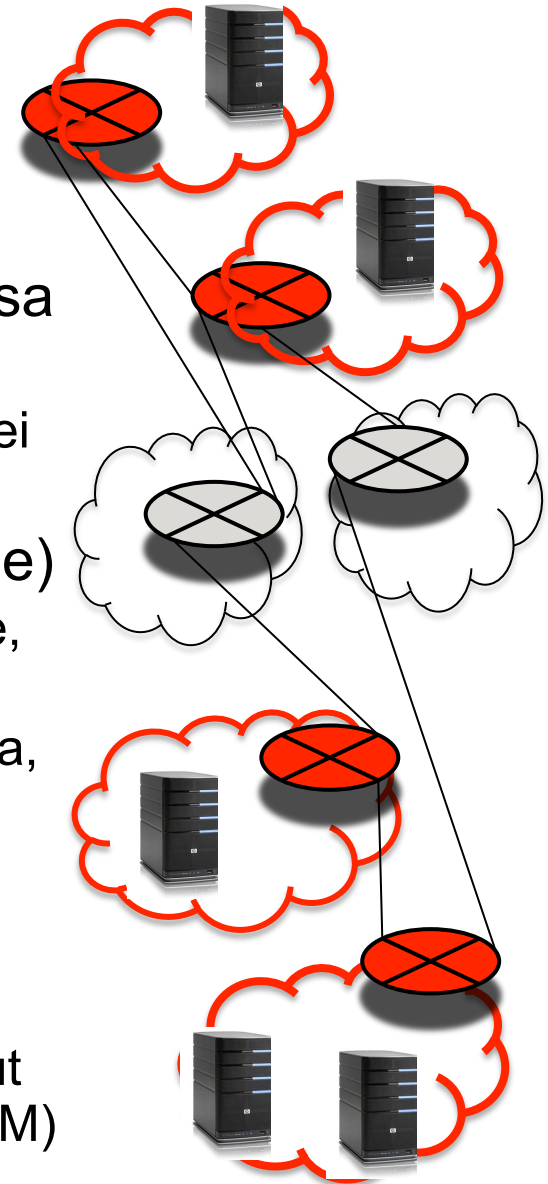
- IP-kerroksen luoma monilähetysosoitteeseen lähetetty viesti kapseloidaan linkkikerroksen kehykseen ja lähetetään linkkikerroksen monilähetysosoitteeseen, jos se tukee monilähetystä, muuten yleislähetystenä
 - Mäppäys linkki- ja verkkokerroksen monilähetysosoitteiden välillä
 - Ei siis jokaiselle vastaanottajalle erikseen koskaan
- Reititin kuuntelee monilähetysryhmiä, ja välittää viestin eteenpäin muihin verkkoihin perustuen reititystaulussaan olevaan tietoon
 - jos siis kyse osoitteesta, joka välitetään ulos

Dynaamiseen monilähetysryhmän kuuntelijaksi pitää liittyä erikseen

- IPv4-verkossa käytetään Internet Group Management Protocol IGMPv3 (RFC 3376)
 - ICMPn kaltainen, toimii suoraan IP:n päällä (protokollanro=2)
 - Yksittäinen kone käyttää IGMP:tä ryhmään liittymiseen
 - IGMP-viesti lähetetään halutun ryhmän osoitteeseen
 - Paikallinen reititin vastaanottaa viestin
 - Reititin pitää tietoa ryhmistä, joissa on jäseniä
 - tarkistus säännöllisesti, onko reitittimen hallitsemassa verkossa yhä kuulijoita (yksikin vastaus riittää, kaikkien ei tarvi vastata, eikä reititin pidä kirjaa vastaanottajien lukumäärästä tai IP-osoitteista)
 - jos ei ole, kerrotaan muille reitittimille ja päivitetään reititystaulua
 - IPv6-verkossa: Multicast Listener Discovery v2 (RFC 3810)
-

Monilähetyksen reititys

- Liikenne välitetään vain niihin verkkoihin, joissa on kuuntelijoita
 - Myös sellaisten reitittimien läpi, joiden verkossa ei ole kuuntelijoita, jos ne on matkalla
- Kaksi tapaa rakentaa virityspuu (spanning tree)
 - Datalähtöinen (data driven): lähetetään kaikkialle, kunnes joku kertoo, että tänne ei tarvi lähettää
 - Tarvelähtöinen (demand driven): lähetetään vasta, kun joku pyytää
- Lähettäjä merkitsee myös
 - Alunperin kuka vaan saattoi lähettää, virityspuun rakentaminen hankalaa
 - Nyt yleensä vain yksi lähettäjä (jonka kautta muut voi myös lähettää) Source-specific multicast (SSM)



Monilähetyksen reititysprotokollat

- Distance-Vector Multicast Routing Protocol (DVMRP)
 - Pohjautuu RIPiin, datalähtöinen, ei skaalaudu kovin hyvin
- Protocol-Independent Multicast (PIM) routing protocol on kaksiosainen
 - Dense mode: (RFC 3973): kuuntelijoita on paljon, datalähtöinen tapa
 - Sparse mode (RFC 4601): vähän kuuntelijoita verkon reunoilla, tarvelähtöinen tapa, jossa yksi reititin toimii rendez-vous pointina
- BGP:hen on monilähetyksreitityslaajennus (RFC 6513)
- Monilähetyksreititysprotokollia on paljon muitakin...

Yhteenvedo reitityksestä

- Edelleenlähetyksessä verrataan vastaanottajan IP-osoitetta reititystaulusta löytyviin verkko-osoitteisiin ja niistä valitaan pisin sopiva
- Internet-verkko on jaettu autonomisiin järjestelmiin
 - AS:n välillä reititys on sopimuspohjaista
 - Reitityksen skaalautumiseksi tarvitaan hierarkiatasoja
 - Tehokkuus tärkeintä yhden toimijan verkossa
- AS:n välillä käytetään BGP4:ää
- AS:n sisällä voi käyttää mitä haluaa
 - Staattinen reititys
 - Linkkitilareititys esim OSPF
 - Etäisyysvektoreititys esim RIP
- Monilähetyksessä verkko monistaa viestin kaikille tiettyyn ryhmään kuuluville tai siihen liittyneille mahdollisimman lähellä vastaanottajaa
 - Tehostaa verkon käyttöä

Lyhenteet

- LS link-state
- DV distance vector
- AS autonomous system
- ASN autonomous system number
- EGP exterior gateway protocol
- IGP interior gateway protocol
- RIP routing information protocol
- OSPF open shortest path first
- BGP border gateway protocol
- RPF reverse path forwarding
- IGMP internet group management protocol
- DVMRP distance vector multicast routing protocol
- PIM protocol-independent multicast routing protocol

Termit ja käsitteet

- Etuliite (prefix), longest prefix matching rule, verkkorajapinta (network/link interface), reitittimen portti, reititystaulu (routing table), edelleenlähetystaulu (forwarding table),
- Edelleenlähetys (forwarding), kontrollitaso (control plane), datataso (data plane), reititys (routing),
- Jonotus, puskuri, switching
- Pakettihäviö (packet loss), pakettiskeduleri (packet scheduler), palvelunlaatu (quality of service), hännän tiputus (drop tail) random early detection (RED), head-of-the-line (HOL) blocking,
- Reititysalgoritmi, reititysprotokolla, oletusreititin (default router), reititin (router, gateway), ensimmäisen hypyn reititin, lähettäjäreititin, vastaanottajareititin, graafi (graph), naapuri (neighbor), polku (path), lyhin polku, halvin polku, globaali reititysalgoritmi, linkkitila-algoritmi, etäisyysvektorialgoritmi, hajautettu reititysalgoritmi, staattinen reititys, dynaaminen reititys, hajautettu, iteratiivinen, topologia, reititystaulu, topologiataulu, reitityssilmukka (routing loop), autonominen järjestelmä (autonomous system), sisäinen reititys, ulkoinen reititys, alueen reunareititin (area border router), runkoverkko (backbone), painoarvo, kustannus (cost), BGP-pari (BGP peer), ulkoinen BGP (external BGP), sisäinen BGP (internal BGP), reitti (route), politiikka (huono suomennos sanalle policy, ehkä sopimusperusteinen olisi parempi), reunaverkko (stub net), monikotinen (multihomed), läpikulkuverkko (transit network), asiakasverkko (customer network), operaattoriverkko (provider network), skaalautuvuus (scalability), tehokkuus (performance),
- Yleislähetys (broadcast), yksilähetys (unicast), monilähetys/ryhmälähetys (multicast), tulviminen (flooding), yleislähetysmyrsky (broadcast storm), reverse path forwarding (RPF), virityspuu (spanning tree), rendezvous point, core, monilähetysryhmä, pehmeä tila (soft state), monilähetysreititys, pruning, lähettäjäriippuva monilähetys (source-specific multicast), datadriven, demand driven
- Ilmoitusviesti (advertisement),

Lähteet

- BGPv4 RFC 4271, RFC 4274
- RIPv1 RFC 1058
- RIPv2 RFC 2453
- OSPFv2 RFC 2328
- IGMPv3 RFC 3376
- MLD RFC 3810 (IGMP:n vastine IPv6:ssa)
- DVMRP RFC 1075
- PIM RFC 3973, RFC 5601
- Source specific multicast (SSM) RFC 3569, RFC 4607
- BGP:n monilähetys RFC 6513