



## 4. Liikenteen mallinnus ja mittaus

## Sisältö

- Liikenteen mittaus
- Liikenteen vaihtelu
- Puhelinliikenteen mallinnus
- Dataliikenteen mallinnus
- Uudet dataliikenteen mallit

## Liikenteen mittaus (1)

- Liikennemittauksia tarvitaan
  - verkonsuunnittelua varten
    - mitoituksen pohjaksi
    - liikennemallien luomiseksi
    - liikenne-ennusteiden tekemiseksi
  - dynaamisen liikenteenhallinnan toteuttamiseksi
    - mittauksiin perustuva yhteyksien hyväksyntämenettely
    - dynaaminen reititys
    - ruuhkan havaitseminen ja hallinta
  - verkon ja sen eri komponenttien suorituskyvyn ja niiden tarjoaman palvelun laadun selvittämiseksi
  - mutta myös laskutustiedon keräämiseksi
- Liikennemittausten tarve kasvaa, sillä ...
  - tulee uusia palveluita ja uudenlaisia verkkoja
  - liikenteen luonne muuttuu, käyttötottumukset muuttuvat, tariffit muuttuvat
  - liikenne tietoverkoissa on erilaista kuin perinteisessä puhelinverkossa

## Liikenteen mittaus (2)

- Puhelinverkon mittaukset
  - liikenne eri linkeillä
    - liikenneprosessi (liikenteen voimakkuus varattuina kanavina)
    - kutsujen saapumisprosessi (saapumisväliajat)
    - kutsujen pitoajat
  - liikenne runkoverkon solmuissa
    - eri suuntiin lähtevän liikenteen jakauma
    - eri suunnista tulevan liikenteen jakauma
  - liikenne liityntäverkon solmuissa
    - liikenteen jakauma lähteen tyyppin mukaan
    - eri palveluiden käyttö

## Liikenteen mittaus (3)

- Dataverkkojen mittaukset
  - liikenne eri linkeillä
    - liikenneprosessi (linkin käyttöaste, läpäisy)
    - pakettitasolla (esim. IP)
      - pakettien saapumisprosessi (saapumisväliajat)
      - pakettien koot
      - pakettien menetyssuhde (loss ratio)
    - yhteys/vuotasolla (esim. TCP)
      - yhteyksien/voiden saapumisprosessi (saapumisväliajat)
      - yhteyksien/voiden kestot
      - yhteyksien/voiden aikana siirrettyjen bittien määrät
      - em. tiedot sovellutuksittain (http / ftp / email / telnet etc.)
  - päästä-päähän liikennemittaukset
    - pakettien kokema viive (delay) ja viiveen vaihtelu (jitter)
    - yhteyden saama keskimääräinen kaista eli läpäisy (throughput)

## Liikennemittausten analysointi

- Perinteiset tilastolliset menetelmät:
  - liikennemallien parametrien **estimointi**, esim.
    - liikenneintensiteetti
    - liikenteen vaihtelu (lyhyen aikavälin varianssi)
    - liikenteen huipukkuus
  - liikenteen autokorrelaation estimointi
- Uusi lähetymistapa:
  - liikenteen **skaalautuvuuden** analyysi
    - itsesimilaarisuus
    - multifraktaalisuus

## Sisältö

- Liikenteen mittaus
- Liikenteen vaihtelu
- Puhelinliikenteen mallinnus
- Dataliikenteen mallinnus
- Uudet dataliikenteen mallit

## Liikenteen vaihtelu eri aikaskaaloissa (1)

- **Ennustettavat vaihtelut**
  - **Trendikehitys** (vuosia)
    - liikenteen määrän kasvu:
      - olemassaolevat palvelut: käyttäjien määrän kasvu, käyttötottumuksen muutokset, tariffien muutokset
      - uudet palvelut
  - **Vuosittainen kausivaihtelu**, vuosiprofiili (kuukausia)
  - **Viikkorytmiin liittyvät vaihtelut**, viikkoprofiili (päiviä)
  - **Päivärytmiin liittyvät vaihtelut**, päiväprofiili (tunteja)
    - sisältäen “kiiretunnin”
    - huom. eri käyttäjäryhmillä erilaiset (päivä-/viikko-/vuosi-) profiilit
  - **Ulkoisista tapahtumista** johtuvat ennustettavat vaihtelut
    - säännölliset: esim. joulukuusi
    - epäsäännölliset: esim. puhelinäänestykset



## Liikenteen vaihtelu eri aikaskaaloissa (2)

- **Satunnaiset vaihtelut**
  - **Lyhyen aikavälin satunnaisvaihtelut** (sekunteja - minuutteja)
    - toisistaan riippumattomien käyttäjien käyttäytymiseen liittyvät vaihtelut
      - satunnaiset kutsujen saapumiset
      - satunnaiset kutsujen pitoajat
  - **Pitkän aikavälin satunnaisvaihtelut** (tunteja - ...)
    - satunnaiset vaihtelut päivä-, viikko-, vuosiprofiilin ympärillä
    - jokainen päivä, viikko, ... on erilainen
  - **Ulkoisten tapahtumien** aiheuttamat satunnaisvaihtelut
    - esim. maanjäristykset ja muut luonnonmullistukset

## Kiiretunti (1)

- Verkon mitoittaminen suurimman milloinkaan esiintyvän liikennehuipun varalle ei ole tarkoituksenmukaista
- Puhelinverkoissa mitoitusta varten kehitetty (laskennallinen) liikenteen huippua kuvaava suure on ns. **kiiretunnin** (busy hour) **liikenne**

**Kiiretunti**  $\approx$  se yhden tunnin pituinen jakso, jona liikenteen määrä on suurin

- tämä määritelmä on yksikäsitteinen vain yksittäisille päiville
- kyseessä on siis (tarkkaan ottaen) päivittäinen huipputunti (daily peak hour)
- Mitoitukseen sopivien (useampaan päivään liittyvien) keskimääräisten arvojen määrittämiseksi suosituksissa on määritelty kiiretunnin liikenteelle erilaisia mittaustapoja
  - ADPH (Average Daily Peak Hour)
  - TCBH (Time Consistent Busy Hour)
  - FDMH (Fixed Daily Measurement Hour)

## Kiiretunti (2)

- Merkitään
  - $N$  = mitattauspäivien lkm (esim.  $N = 10$ )
  - $a_n(\Delta)$  = liikenteen keskim. voimakkuus mittauspäivän  $n$  aikavälillä  $\Delta$
  - $\max_{\Delta} a_n(\Delta)$  = huipputunnin liikenne mittauspäivänä  $n$
- Kiiretunnin liikenne  $a$  eri menetelmillä laskettuna:

$$a_{\text{ADPH}} = \frac{1}{N} \sum_{n=1}^N \max_{\Delta} a_n(\Delta)$$

$$a_{\text{TCBH}} = \max_{\Delta} \frac{1}{N} \sum_{n=1}^N a_n(\Delta)$$

$$a_{\text{FDMH}} = \frac{1}{N} \sum_{n=1}^N a_n(\Delta_{\text{fixed}})$$

- Selvästikin

$$a_{\text{FDMH}} \leq a_{\text{TCBH}} \leq a_{\text{ADPH}}$$

## Sisältö

- Liikenteen mittaus
- Liikenteen vaihtelu
- Puhelinliikenteen mallinnus
- Dataliikenteen mallinnus
- Uudet dataliikenteen mallit

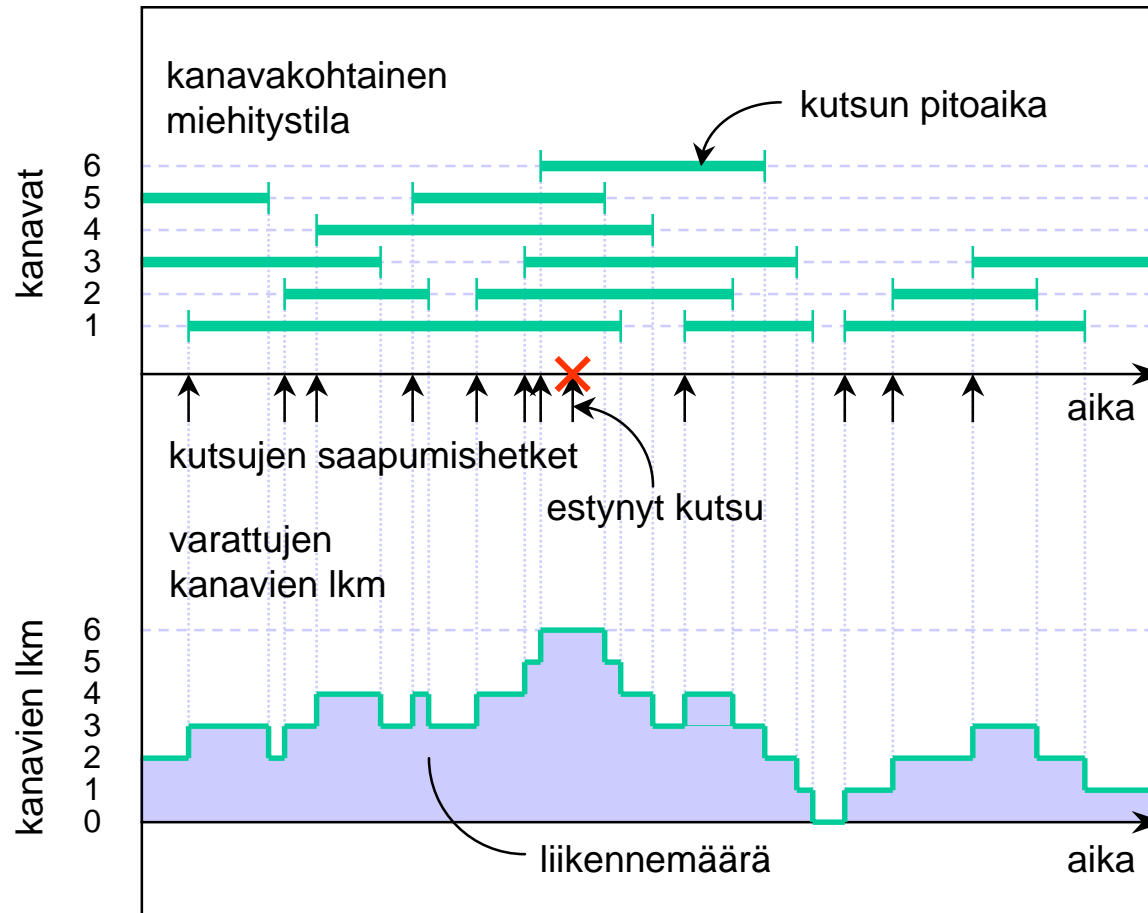
## Puhelinliikenteen mallinnus

- Puhelinverkossa liikenne koostuu **kutsuista** (puheluista)

Liikenne ↔ Kutsut

- Verkon kannalta on tärkeää kuvata
  - **kutsujen saapumisprosessi** (millä hetkillä uusia puheluita yritetään käynnistää)
  - **kutsujen pitoaikajakauma** (miten kauan yksittäiset puhelut kestävät)
- Nämä yhdessä muodostavat
  - **liikenneprosessin**, joka kertoo montako puhelua on yhtäaikaan käynnissä  
= montako kanavaa on yhtäaikaan varattuna  
= kuljetetun liikenteen (hetkellisen voimakkuuden) erlangeina
- Huom. Tiettynä aikavälinä kuljetettua kokonaisliikennettä sanotaan
  - **liikennemääräksi** eli liikenteen volyymiksi  
= liikenteen voimakkuuden integraali yli ko. aikavälin

# Liikenneprosessi

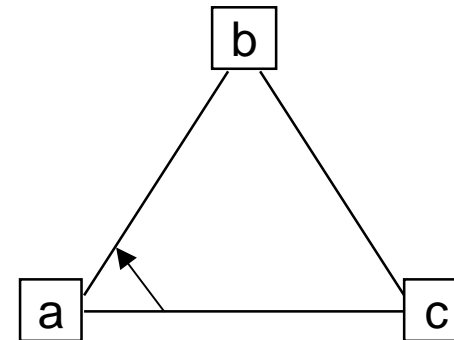


## Kutsujen saapumisprosessin mallinnus (1)

- **Aggregoitu liikenne** (runkoverkossa)
  - perinteinen malli: Poisson-prosessi (intensiteetillä  $\lambda$ )
    - hyvin lyhyellä aikavälillä  $\Delta$  tulee joko uusi kutsu (tn:llä  $\lambda\Delta$ ) tai sitten ei (tn:llä  $1 - \lambda\Delta$ )
    - eri aikavälit ovat toisistaan riippumattomia
    - tuloksena riippumattomat ja eksponentiaaliset saapumisten väliajat keskiarvolla  $1/\lambda$
  - malli toimii, jos hyvin suuri käyttäjäpopulaatio, joka tekee päätöksiä kutsujen käynnistämisestä toisistaan (ja aiemmista päätöksistään) riippumatta
  - vastaavat järjestelmämallit:
    - **Poisson-malli** (ääretön kapasiteetti)
    - **Erlang-malli** (äärellinen kapasiteetti)

## Kutsujen saapumisprosessin mallinnus (2)

- Hierarkkiseen vaihtoehtoisreititykseen liittyvä **ylivuoto-liikenne** (runkoverkossa)
  - mallinnettu esim. keskeytyvänä Poisson-prosessina (Interrupted Poisson Process, IPP) tai yleisemmin Markov-moduloituna Poisson-prosessina (Markov modulated Poisson process, MMPP)
  - ideana on, että Poisson-prosessin mukaiset saapumiset sallitaan tai estetään riippuen ns. moduloivan prosessin tilasta



suora reitti: a - c  
vaihtoehtoinen reitti: a - b - c



## Kutsujen saapumisprosessin mallinnus (3)

- **Yksittäinen käyttäjä** (tilaajan käyttäytymisen mallinnus)
  - perinteinen malli: eksponentiaalinen on-off-prosessi
    - käyttäjä voi olla kahdessa eri tilassa: 'on' ja 'off'
      - kun puhelu on käynnissä, tila on 'on'
      - muutoin tila on 'off'
    - sekä kutsun pitoajat että eri kutsujen väliset ajat oletetaan eksponentiaalisesti jakautuneiksi (eri parametreilla)
- Yksittäisten käyttäjien **superpositio** (tilaajaverkossa)
  - äärellinen määrä yksittäisiä käyttäjiä, jotka
    - mallinnetaan kuten yllä ja
    - toimivat toisistaan riippumatta
  - vastaavat järjestelmämallit:
    - **Binomi-malli** (riittävä kapasiteetti)
    - **Engset-malli** (riittämätön kapasiteetti)

## Kutsujen pitoajan mallinnus (1)

- Perusoletus:
  - pitoajat ovat **riippumattomia** ja **samoin jakautuneita** (independent and identically distributed, **IID**)
- Pitoajan jakauma
  - perinteinen malli: eksponentiaaliset pitoajat
    - 1 parametri  $\Rightarrow$  **yksinkertaisuus!**
    - muistittomuus: kestätyään ajan  $t$ , kutsu päättyy lyhyessä aikavälissä  $(t, t+\Delta)$  ajasta  $t$  riippumattomalla tn:llä  $\mu\Delta$
    - häntäjakauma pienenee eksponentiaalista vauhtia
  - monimutkaisempia malleja:
    - normaalijakauma (2 parametria: keskiarvo ja varianssi)
    - log-normaalijakauma (2 parametria)
    - Weibull-jakauma (2 parametria)
    - hyperekspontiaalinen (3 parametria)

## Kutsujen pitoajan mallinnus (2)

- Pitoaika riippuu myös esim. siitä,
  - onko kyseessä liike- vai yksityispuhelu (päivä vs. ilta)
  - onko kyseessä tavallinen puheyhteys vai kenties puhelinyhteyden käyttö muihin tarkoituksiin (kuten Internet-surffailu, faxin lähetys yms.)

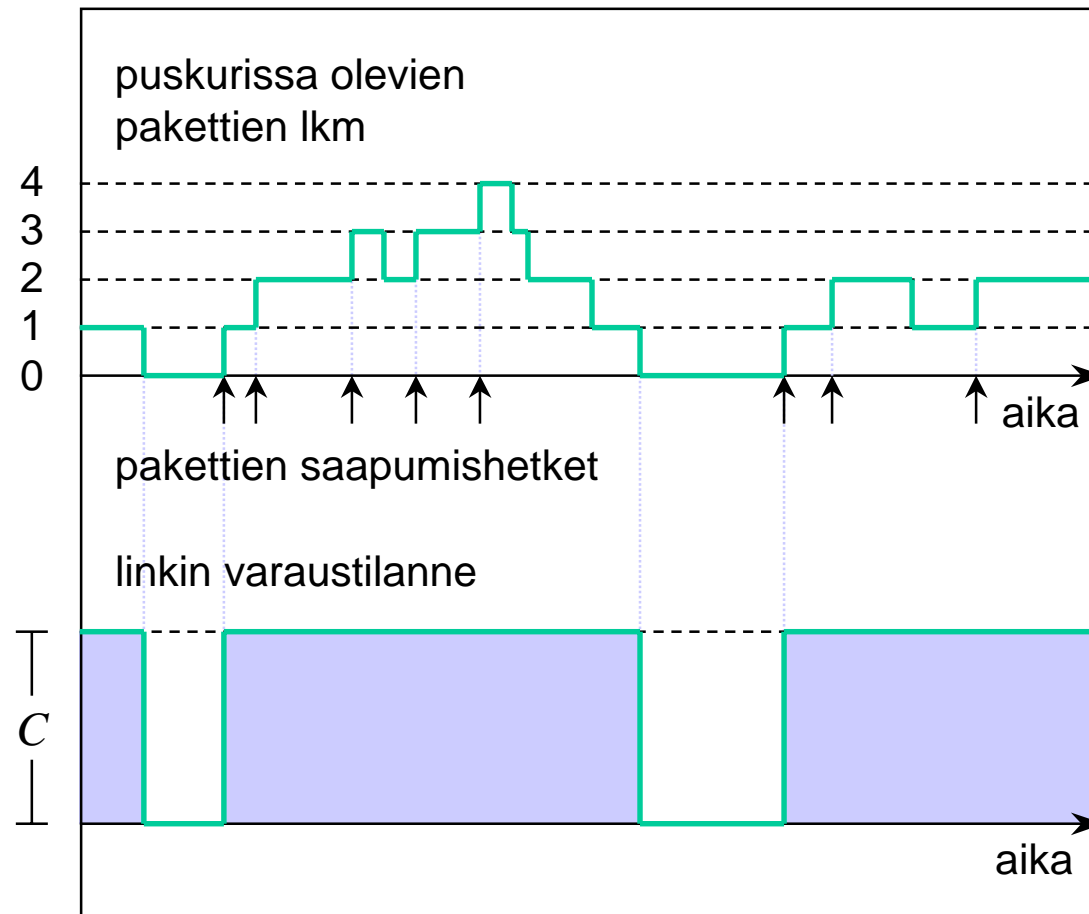
## Sisältö

- Liikenteen mittaus
- Liikenteen vaihtelu
- Puhelinliikenteen mallinnus
- **Dataliikenteen mallinnus**
- Uudet dataliikenteen mallit

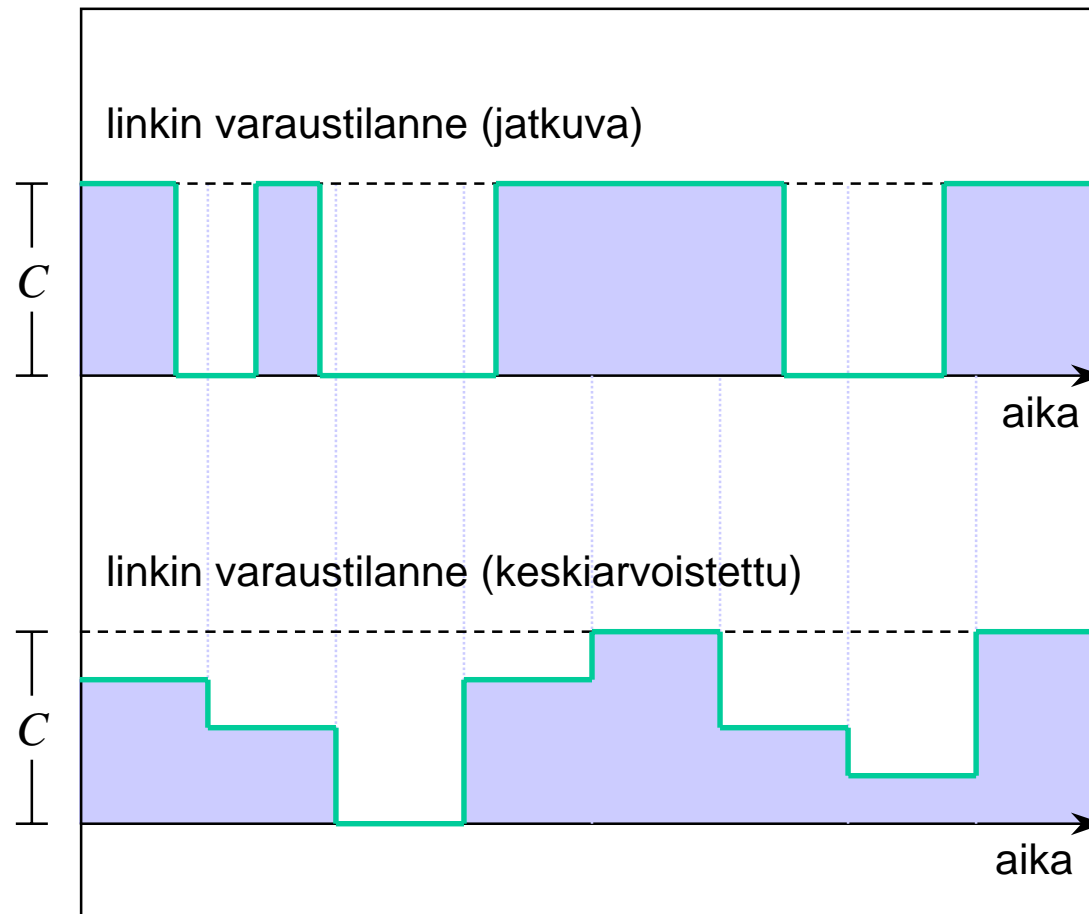
## Dataliikenteen mallinnus pakettitasolla

- Perinteinen malli **pakettitasolla** (packet level)
  - uusia paketteja saapuu Poisson-prosessin mukaisesti  
⇒ pakettien väliajat riippumattomia ja samoin jakautuneita noudattaen eksponenttijakaumaa
  - pakettien pituudet riippumattomia ja samoin jakautuneita noudattaen eksponenttijakaumaa  
⇒ pakettien lähetysajat riippumattomia ja samoin jakautuneita noudattaen eksponenttijakaumaa
  - järjestelmämallina yhden palvelijan **jonotusjärjestelmä** (M/M/1-FIFO jonomalli)

## Pakettitason liikenneprosessi (1)



## Pakettitason liikenneprosessi (2)



## Dataliikenteen mallinnus vuotasolla (1)

- Vuo on eräälainen pakettitasolta havaittu approksimaatio yhteystason liikenteestä
- **Vuo** (flow) = sarja peräkkäisiä ja toisiinsa kuuluvia paketteja, joilla sama lähde ja määränpää
  - Karkeassa luokittelussa huomioidaan vain lähde- ja määränpääosoite, hienommassa luokittelussa voidaan erotella esim. eri ylemmät protokollat (esim. TCP, UDP, HTTP, FTP, ...)
  - Peräkkäiset vuot erotetaan toisistaan ajastimella: peräkkäiset paketit kuuluvat samaan vuohon vain, jos niiden ajallinen etäisyys on tarpeeksi pieni
  - Huom. Vuon määritelmä on hyvin joustava. Se, miten vuot käytännössä pitäisi luokitella, on oma taiteenlajinsa:
    - karkeuden (granularity) valinta
    - ajastuksen (timeout) valinta



## Dataliikenteen mallinnus vuotasolla (2)

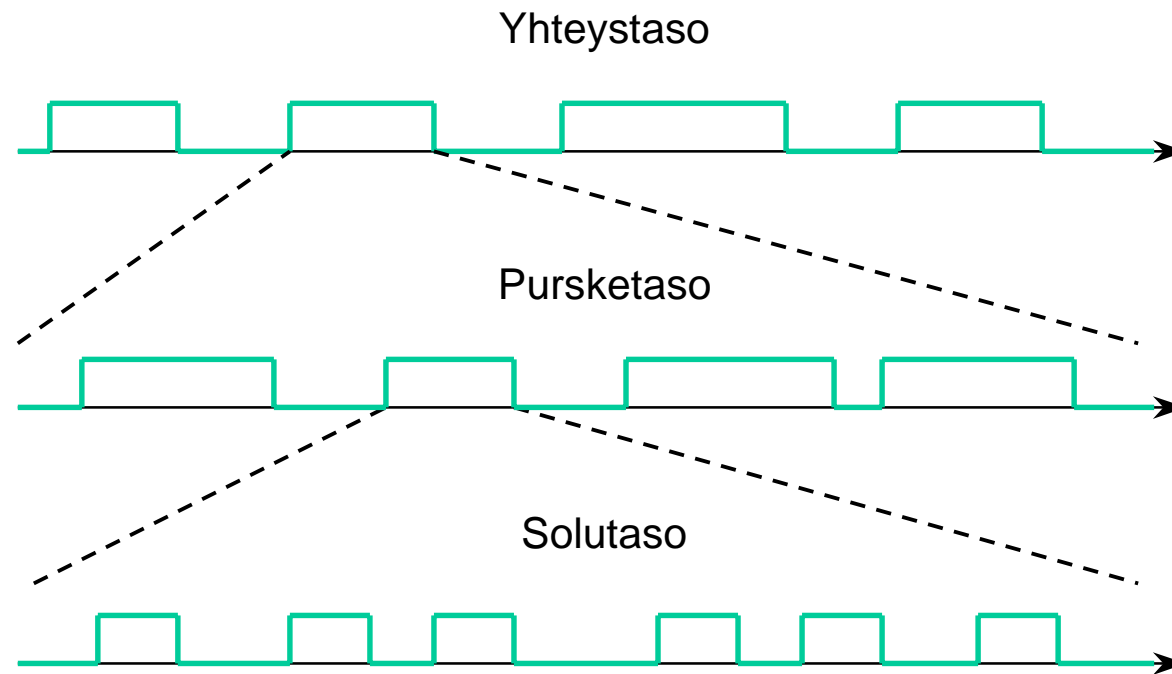
- Voiden jaottelu liikenteen luonteen mukaisesti:
  - **Virtaavat** (streaming) vuot, joita karakterisoi vaadittu kaista ja vuon kesto
    - esim. audio- ja videoyhteydet UDP:llä siirrettynä
  - **Elastiset** (elastic) vuot, joita karakterisoi siirrettävä tietomäärä eli vuon koko
    - esim. digitaalisten dokumenttien siirto (HTTP,FTP,...) TCP:llä
    - huom. elastisen vuon vaatima kaista ja vuon kesto määräytyvät dynaamisesti verkon tilasta riippuen

## Dataliikenteen mallinnus vuotasolla (3)

- TCP:tä käyttävän **elastisen** liikenteen malli vuotasolla
  - uusia voita syntyy Poisson-prosessin mukaisesti
  - voiden kokojakauma vapaasti valittavissa (esim. mittauksiin sovittaen)
  - jos tarkastelukohteena yksittäinen linkki, järjestelmämallina yhden palvelijan **jonotusjärjestelmä**, jossa kaikkia asiakkaita palvellaan yhtäaikaan samalla, systeemissä olevien asiakkaiden lkm:stä riippuvalla nopeudella (M/G/1-PS jonomalli)
  - PS jonokuri idealisoi ajatuksen, että TCP-vuot reagoivat ruuhkatilanteeseen vähentämällä nopeuttaan (ts. sopeuttavat nopeuttaan tilanteen mukaan)
- UDP:tä käyttävän **virtaavan** liikenteen malli vuotasolla
  - uusia voita syntyy Poisson-prosessin mukaisesti
  - voiden kestojakauma vapaasti valittavissa (esim. mittauksiin sovittaen)
  - voiden varaama kaista kiinteä tai muuttuu satunnaisesti (verkon tilasta riippumatta)
  - jos pelkästään UDP-liikennettä, yksinkertaiseksi malliksi käyvät erilaiset estomallit (vrt. ATM-liikennemallit yhteystasolla)

## ATM-liikenteen mallinnus (1)

- Kolme eri aikaskaalaa:



## ATM-liikenteen mallinnus (2)

- **Yhteystaso** (connection level)
  - “liikenteen yksikkö” = yhteys
  - järjestelmämallina **menetysjärjestelmät** (CBR- ja VBR-yhteyksille)
- **Pursketaso** (burst level)
  - “liikenteen yksikkö” = vaihtelevanmittainen (ja mahdollisesti vaihtelevan-nopeuksinen) purske
  - järjestelmämallina ns. **nestejonomallit** (fluid queue):
    - Anick-Mitra-Sondhi (A-M-S): eksponentiaaliset on-off-lähteet ( $l_{km} < \infty$ )
    - Kosten: eksponeniaaliset on-off-lähteet ( $l_{km} = \infty$ )
- **Solutaso** (cell level)
  - “liikenteen yksikkö” = kiinteänpituinen paketti = solu
  - järjestelmämallina **jonotusjärjestelmät**:
    - periodisten lähteiden superpositio ( $N^*D/D/1$ )
    - solujen saapumiset Poisson-prosessin mukaisia ( $M/D/1$ )

## Sisältö

- Liikenteen mittaus
- Liikenteen vaihtelu
- Puhelinliikenteen mallinnus
- Dataliikenteen mallinnus
- Uudet dataliikenteen mallit

## Bellcorella tehdyt Ethernet-mittaukset

- Bellcorella tehdyt Ethernet (LAN) mittaukset ('89-'92)
  - pitkäaikainen (sis. satoja miljoonia paketteja) ja erittäin tarkka mittaus
    - jokaisen paketin koko ja lähetysaika rekisteröitiin
  - kts. IEEE/ACM Trans. Networking, vol. 2, nr. 1, pp. 1-15, February 1994
- Johtopäätökset:
  - Ethernet-liikenne näyttää olevan **äärimmäisen vaihtelevaa**
    - purkeisuutta kaikissa aikaskaaloissa mikrosekunneista millisekunneihin, sekunteihin, tunteihin, kuukausiin, ...
  - Ethernet-liikenne näyttää olevan **itsesimilaarista** (self-similar) ja sisältävän **pitkän aikavälin riippuvuutta** (long-range dependence, LRD)
    - jos sitä tarkastellaan eri aikaskaaloissa (zoomataan), se näyttää samanlaiselta, ainoastaan vaihteluiden amplitudi muuttuu
    - ns. Hurstin parametri (siis yksi ainoa) kuvaa riippuvuuden astetta
  - Perinteiset dataliikenteen mallit (Poisson-prosessi, eksponenttijakauma) eivät tuo esiin näitä havaittuja, uusia Ethernet-liikenteen ominaisuuksia!

## Paxsonin ja Floydin Internet-mittaukset

- Paxsonin ja Floydin Internet (WAN) mittaukset ('93-'95)
  - sekä yhteys- että pakettitason tarkastelu
  - kts. IEEE/ACM Trans. Networking vol. 3, nr. 3, pp. 226-244, June 1995
- Yhteystason johtopäätökset:
  - yhteydet, joihin liittyy todellinen (elävä) käyttäjä ja joissa ko. käyttäjä ei käynnistä uusia yhteyksiä istuntonsa aikana (TELNET, FTP sessions), voidaan mallintaa Poisson-saapumisprosessilla (kuitenkin tuntikohtaisin keskiarvoin)
  - sen sijaan muunlaiset yhteydet, ts. aidosti tietokoneiden väliset (SMTP, NNTP) sekä sellaiset, joissa käyttäjä luo useita yhteyksiä saman istunnon aikana (HTTP, FTP data), ovat **purskeisempia** kuin mitä Poisson-malli antaisi olettaa; lisäksi saapumisväliaikojen **korreloituneisuus**
- Pakettitason johtopäätökset
  - TELNET pakettien väliaikojen jakauma on **paksuhäntäinen** (heavy-tailed) eikä siis eksponentiaalinen, kuten perinteiset mallit olettavat

## Dataliikenteen uudet mallit (1)

- “Uusia” jakaumia:
  - **alieksponentiaaliset** jakaumat (subexponential distributions)
    - “worse than exponential tail”
    - esim. log-normaalinen, Weibull- ja Pareto-jakauma
  - **paksuhäntäiset** jakaumat (heavy tailed distributions)
    - “power-law tail”
    - esim. Pareto-jakauma (sijaintiparametrina  $a$  ja muotoparametrina  $\beta$ )

$$P\{X > x\} = \left(\frac{a}{x}\right)^\beta, \quad x \geq a > 0, \quad \beta > 0$$

- Uusissa dataliikenteen malleissa näillä jakaumilla kuvataan
  - pakettien pituuksia ja saapumisväliaikoja
  - yhteyksien kestoja ja saapumisväliaikoja

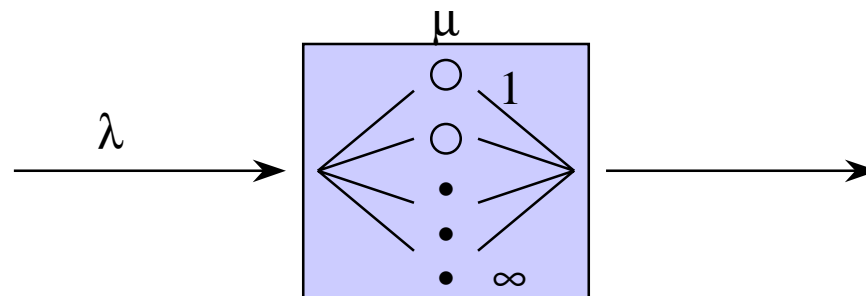


## Dataliikenteen uudet mallit (2)

- “Uusia” prosessimalleja:
  - **pitkän aikavälin riippuvuutta** sisältävät prosessit (long range dependence, LRD)
  - **itsesimilaariset** prosessit (self-similarity)
- Jos stokastinen prosessi on (asymptoottisesti) itsesimilaarinen ja positiivisesti korreloitunut, niin se sisältää pitkän aikavälin riippuvuutta
  - esim. **fraktionaalinen Brownin liike** (fractional Brownian motion, FBM)
    - soveltuu kuvaamaan aggregoitua dataliikennettä (runkoverkossa)
    - itsesimilaarinen ja pos. korreloitunut prosessi  $\Rightarrow$  LRD
    - vain kolme parametria (säästeliästä!)
    - näistä yksi, **Hurstin parametri**  $H \in (1/2, 1)$ , kuvaa riippuvuuden astetta (mitä suurempi  $H$  sitä pidemmälle riippuvuus kantaa)
- Itsesimilaarisuus ja pitkän aikavälin riippuvuus voi aiheutua paksuhäntäisistä jakaumista (kts. seuraava esimerkki)

## Esimerkki

- Tarkastellaan ääretöntä systeemiä (M/G/∞)
  - asiakkaita saapuu Poisson-prosessin mukaisesti
  - palveluajat riippumattomia ja samoin jakautuneita
    - jakauma paksuhäntäinen
    - jakaumalla ääretön varianssi
    - esim. Pareto-jakauma muotoparametrinaan  $\beta < 2$
- Tällöin vastaava liikennemääräprosessi on asymptoottisesti itsesimilaarinen sisältäen pitkän aikavälin riippuvuutta



## Sanastoa

- liikenne = traffic
- liikenteen voimakkuus = traffic intensity
- liikenneprosessi = traffic process
- liikennemäärä = traffic volume
- saapumisprosessi = arrival process
- ylivuotoliikenne = overflow traffic
- kiiretunti = busy hour
- estimaatti = estimate
- harhaton = unbiased
- aikaskaala = time scale
- yhteystaso = connection level
- pursketaso = burst level
- solutaso = cell level
- pakettitaso = packet level
- vuotaso = flow level
- virtaava = streaming
- elastinen = elastic
- nestejono = fluid queue
- itsesimilaarisuus = self-similarity
- pitkän aikavälin riippuvuus = long range dependence = LRD
- paksuhäntäinen jakauma = heavy-tailed distribution

**THE END**

