

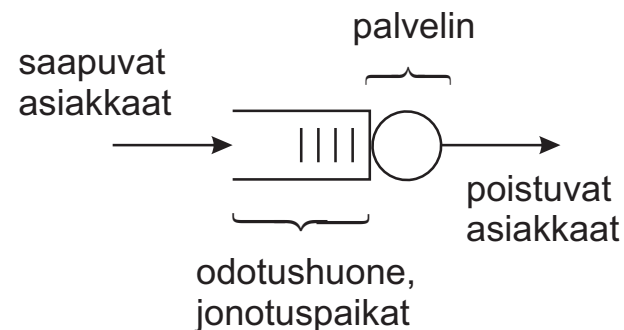
JONOJÄRJESTELMÄT

Yleistä

- Jonojärjestelmät muodostavat keskeisen mallinnuksen välineen mm. tietoliikenne- ja tietokonejärjestelmien suorituskyvyn analysoinnissa.
- Kuvaa kilpailua resursseista
 - jonokielellä resursseja kutsutaan palvelimiksi
 - sovelluksissa tämä voi tarkoittaa johtoja, kapasiteettia ...
- Jonoon saapuvat asiakkaat vastaavat sanomia, paketteja, puheluita, töitä, taskeja ...
- Useimmiten kuvattava järjestelmä on monimutkainen (esim. tietoliikenneverkko, käyttöjärjestelmä) ja sisältää useita jonoja, joista muodostuu jonoverkko.
 - alussa keskitytään tarkastelemaan yksittäisten jonojen muodostamia systeemejä
 - näitäkin on montaa eri tyyppiä, ja näille on kehitetty rikas, monipuolinen teoria

Esimerkki.

Yhden palvelimen jono



Jonojärjestelmiä erottelevia tekijöitä

- Saapumisprosessi
 - väliajat
 - ryhmäsaapumiset
- Palveluprosessi
 - palveluvaateet
- Palvelinten lukumäärä
- Jonojen lukumäärä
- Odotuspaikkojen lukumäärä
 - odotustilan jako eri jonojen kesken
- Asiakkaiden palvelujärjestys jonossa
 - FIFO, LIFO
 - nopeimmat työt ensin
 - rahakkaimmat työt ensin
- Jonojen vuoronjako (skedulointi), jonokuri
 - kiertävä järjestys
 - prosessointikapasiteetin jako
 - prioriteetit
- Informaation määrä
 - kun valitaan jono, tiedetäänkö jonojen pituudet, tarvittavat palveluajat ...
- Diskreettiaikaisuus / jatkuva-aikaisuus
- Muita tekijöitä
 - asiakkaiden karsinta
 - etuilu
 - lahjonta ...

Jonojärjestelmien merkintä (Kendall)

Jonojärjestelmien määrittelemiseksi käytetään usein seuraavaa Kendallin ehdottamaa merkintätapaa: $A/S/m/c/p$, missä

\underbrace{A}	/	\underbrace{S}	/	\underbrace{m}	/	\underbrace{c}	/	\underbrace{p}
saapumis- prosessi		palvelu- prosessi		palvelinten lukumäärä		systeemi- paikkojen lukumäärä		asiakas- populaation koko

- A ja S korvataan tapauksen mukaan jollakin yleisessä käytössä olevista symboleista.
- Yleisen käytännön mukaan jononpituudella tarkoitetaan järjestelmässä sisällä olevien asiakkaiden kokonaismäärää (sisältäen sekä palveltaviana olevat että odottavat asiakkaat).
- Parametri c sisältää sekä palvelupaikat että odotuspaikat
 - voidaan jättää merkitsemättä, jolloin sen arvo oletusarvoisesti on ääretön
- Asiakaspopulaation koko on niin ikään optionaalinen parametri
 - voidaan jättää merkitsemättä, jolloin sen arvo oletusarvoisesti on ääretön

A (saapumisprosessi)

- Määrittelee saapumisprosessin
- Usein ajatellaan, että saapumisväliajat ovat riippumattomia (ns. uusiutumispösessi), jolloin riittää määritellä väliaikajakauman tyyppi.

Yleisesti käytettyjä merkintöjä ovat

M eksponentiaalinen väliaikajakauma ($M =$ Markovian, memoryless); Poisson-pösessi

D deterministinen, vakioväliajat

G yleinen ($G =$ general)

E_k Erlang- k -jakauma

PH vaihejakauma (phase distribution)

Cox Coxin jakauma

- Lyhenteitä keksitään lisää aina tarpeen mukaan.

S (palveluprosessi)

- Määrittelee asiakkaan palveluajan jakauman
- Palvelu aikaan vaikuttaa kaksi eri tekijää
 - asiakkaan palveluvaade eli työn määrä (esim. lähetettävän datapaketin koko, kB)
 - palvelimen nopeus (esim. kB/s)
 - palveluaika on näiden osamäärä
- Kendallin merkinnässä palveluaikajakauman tyyppi ilmaistaan merkitsemällä S :n paikalle sopiva symboli; yleisesti käytössä ovat samat symbolit (M , D , G jne) kuin saapumisväliaikojen kuvauksessa

Esim. 1. $M/M/1$ -jono

- Poisson-saapumisprosessi
- eksponentiaalinen palveluaikajakauma
- yksi palvelin
- odotuspaikkojen määrä rajaton

Esim. 2. $M/M/m/m$ -jono

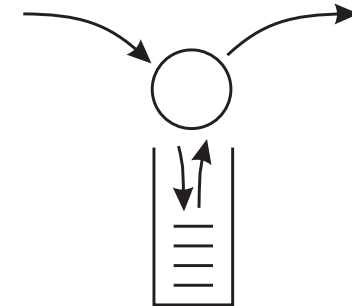
- Poisson-saapumisprosessi
- eksponentiaalinen palveluaikajakauma
- m palvelinta ja m systeemipaikkaa \Rightarrow ei lainkaan odotuspaikkoja, ns. menetysjärjestelmä

Jonokuri / vuoronjako (queueing discipline / scheduling)

- Tavallinen saapumisjärjestyksessä palveltu jono

$$\left\{ \begin{array}{l} \underline{\text{FCFS}} \quad \underline{\text{First Come First Served}} \\ \underline{\text{FIFO}} \quad \underline{\text{First In First Out}} \end{array} \right.$$


- Pino, viimeksi tullut asiakas pääsee ensiksi palveluun

$$\left\{ \begin{array}{l} \underline{\text{LIFS}} \quad \underline{\text{Last Come First Served}} \\ \underline{\text{LIFO}} \quad \underline{\text{Last In First Out}} \end{array} \right.$$


- Pinosta on vielä kolme alalajia

- pre-emptive resume

saapuva asiakas keskeyttää meneillään olevan palvelun, joka sitten aikanaan keskeytetyn asiakkaan päästessä uudelleen palveluun, jatkuu samasta pisteestä, mihin jäätiin

- pre-emptive restart

saapuva asiakas keskeyttää meneillään olevan palvelun; keskeytynyt palvelu aloitetaan alusta keskeytetyn asiakkaan päästessä uudelleen palveluun

- non-pre-emptive

saapuva asiakas odottaa meneillään olevan palvelun päättymistä ennen kuin pääsee palveluun

Jonokuri / vuoronjako (jatkoa)

- Kiertävä palvelu

RR Round robin

- jokainen asiakas saa vuorollaan pienen “aikasiivun” palvelua
- pollaus

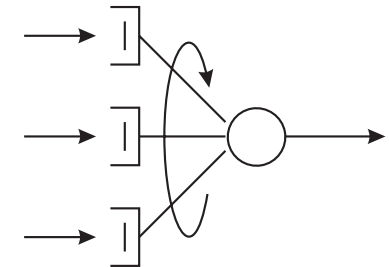
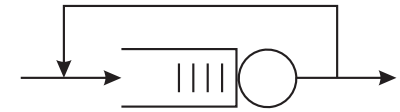
- Prosessorikapasiteetin jako asiakkaiden kesken

PS Processor sharing

- kaikkia jonossa olevia asiakkaita palvellaan samanaikaisesti
- kapasiteetti jaetaan tasan asiakkaiden kesken (kunkin palvelu hidastuu kääntäen verrannollisena jonossa olevien asiakkaiden lukumäärään)
- idealisoitu muoto RR:stä (aikasiivut äärettömän lyhyitä)

Muita vuoronjakomenetelmiä ovat mm.

- SIRO (Service In Random Order): palvelu tapahtuu satunnaisessa järjestyksessä
- SSF (Shortest Jobs First): lyhimmän palveluajan työt suoritetaan ensiksi; palveluaika tulee tietää edeltäkäsien; minimoi keskimääräisen odotusajan



Jonokuri / vuoronjako (jatkoa)

- Jonokuria sanotaan työnsäilyttäväksi, jos palvelimen / palvelimien kapasiteettia ei hukata eli yksikään palvelin ei ole jouten, kun järjestelmässä on odottavia asiakkaita.
- Esimerkkejä jonokurista, jotka eivät ole työnsäilyttäviä
 - LCFS / pre-emptive restart
 - järjestelmät, joissa palvelin voi välillä “lähteä vapaalle”