PS-jono (Processor Sharing)

- Yhden palvelimen PS-jonossa palvelimen kapasiteetti C jaetaan tasan systeemissä olevien asiakkaiden kesken,
  - kun järjestelmässä on n asiakasta sisällä, kutakin palvellaan nopeudella C/n
  - asiakkaat eivät joudu odottamaan, vaan palvelu alkaa välittömästi
- PS-jono on tullut tärkeäksi välineeksi esim. Internetin vuotason mallinnuksessa
  - TCP-protokolla pyrkii jakamaan verkon resursseja karkeasti ottaen tasan käynnissä olevien voiden (esim. www-sivun tai muun dokumentin siirto) kesken
- PS-jono on idealisoitu malli, sillä yleensä palvelimen kapasiteettia ei voida jakaa ainakaan jatkuvasti (reaaliarvoisesti) osiin. PS-jonomalli on kuitenkin hyvä approksimaatio
  - kiertävälle palvelujärjestykselle (RR, round robin), jossa kutakin asiakasta vuorollaan palvellaan lyhyen ajan (esim. tietokoneiden aikajakoiset käyttöjärjestelmät)
  - dokumenttien ja tiedostojen siirrolle, kun nämä on jaettu pieniin paketteihin, joita palvellaan vuorotellen tai joiden lähetysnopeudet muutoin tasataan yhtä suuriksi
- PS-jono on teoreettisesti mielenkiintoinen mm. siksi, että sen keskiarvo-ominaisuudet ovat insensitiivejä asiakkaiden palveluvaadejakauman suhteen (toisin kuin FIFO-jonolla).

M/M/1-PS-jono

- Saapumisprosessi on poissoninen intensiteetillä λ, ja palveluvaateen (työn koko) jakauma on eksponentiaalinen siten, että asiakkaan saadessa koko palvelimen kapasiteetin käyttöönsä palveluaika noudattaa jakaumaa $\mathrm{Exp}(\mu)$.

- Tällöin asiakkaiden lukumäärä järjestelmässä N noudattaa samaa SK-prosessia kuin tavallisen M/M/1-FIFO-jonon tapauksessa:
  - todennäköisyys aikayksikköä kohden, että systeemiin tulee uusi asiakas on λ
  - kun systeemissä on n asiakasta, kunkin asiakkaan palvelun päättymisintensiteetti on μ/n; siten kokonaistodennäköisyys aikayksikköä kohden, että jonkun asiakkaan palvelu päättyy, on μ

- PS-jonon jononpituusjakauma on siten sama kuin tavallisella M/M/1-FIFO-jonolla,

$$\pi_n = (1-\rho)\rho^n, \qquad \rho = \lambda/\mu$$

- Vastaavasti keskimääräinen asiakkaiden lukumäärä $\mathrm{E}[N]$ ja Littlen lauseen mukaan keskiviipymä systeemissä $\mathrm{E}[T]$ ovat

$$\mathrm{E}[N] = \frac{\rho}{1-\rho}, \qquad \mathrm{E}[T] = \frac{1/\mu}{1-\rho}$$

M/G/1-PS-jono

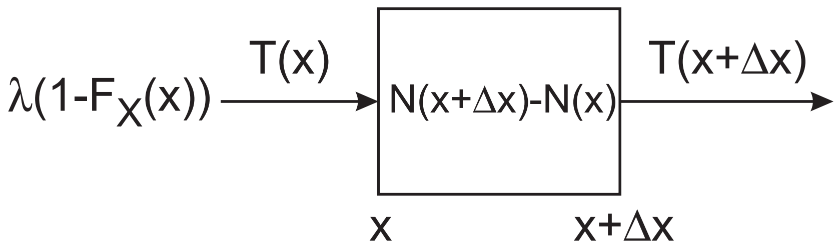
- Esitetään ensin aputulos, joka pätee yleiselle G/G/1-jonolle. Merkitään

$$\begin{cases} x & = \text{ asiakkaan saaman palvelun (työn) määrä} \\ F_X(x) & = \text{ asiakkaan palveluvaateen (työn) kertymäfunktio} \\ N(x) & = \text{ palvelua määrän } \leq x \text{ saaneiden asiakkaiden keskim. lukumäärä systeemissä} \\ T(x) & = \text{ palvelua määrän } x \text{ saaneiden asiakkaiden keskim. systeemissä viettämä aika} \\ n(x) & = \frac{dN(x)}{dx} = \text{ asiakkaiden keskim. lukumäärätiheys (saadun palvelun suhteen)} \end{cases}$$

- Sovelletaan Littlen lausetta mustaan laatikkoon, joka määritellään seuraavasti:
  - asiakas saapuu laatikkoon, kun saadun työn määrä ohittaa arvon x;
    asiakas on tällöin ollut systeemissä keskimäärin ajan $T(x)$
  - asiakas poistuu laatikosta, kun saadun työn määrä ohittaa arvon $x + \Delta x$;
    asiakas on tällöin ollut systeemissä keskimäärin ajan $T(x + \Delta x)$
- Ajatellaan palveluvaade diskreetiksi siten, että työn määrä on aina Δx:n monikerta
  - tällöin laatikosta ei tapahdu poistumisia palvelun päättymisen vuoksi
  - lopulta rajalla $\Delta x \to 0$ diskretoinnin merkitys häviää

M/G/1-PS-jono (jatkoa)

- Koska systeemiin saapuu asiakkaita nopeudella λ ja näistä osuus $1 - F_X(x)$ saavuttaa "palveluiän" x, saapumisnopeus laatikkoon on $\lambda(1 - F_X(x))$.
- Asiakkaan keskimääräinen viipymä laatikossa on $T(x + \Delta x) - T(x)$.



- Littlen lause antaa

$$N(x + \Delta x) - N(x) = \lambda(1 - F_X(x))(T(x + \Delta x) - T(x))$$

- Jakamalla Δx:llä saadaan rajalla $\Delta x \to 0$ haluttu aputulos

$$\boxed{n(x) \;=\; \lambda\,(1 - F_X(x))\,\frac{dT(x)}{dx}}$$

M/G/1-PS-jono (jatkoa)

- Toisaalta voidaan suoraan päätellä, että

$$\boxed{n(x) \;=\; n(0)\cdot(1-F_X(x))}$$

- Tämä perustuu siihen, että PS-jonossa kaikkia asiakkaita palvellaan samalla nopeudella
  - kullakin hetkellä kaikkien asiakkaiden "palveluikä" kasvaa siis samaa vauhtia
  - ainoa ero palveluiän mukaiseen asiakastiheyteen syntyy siitä, että asiakkaita poistuu systeemistä palvelun päättymisen vuoksi
  - ikään x mennessä osuus $F_X(x)$ asiakkaista on poistunut ja osuus $1-F_X(x)$ on jäljellä
- Merkitsemällä kehystettyjen kaavojen mukaiset $n(0)$:n lausekkeet yhtä suuriksi, saadaan

$$\frac{dT(x)}{dx}=\frac{n(0)}{\lambda} \qquad \text{eli} \qquad T(x)=\frac{n(0)}{\lambda}\,x$$

- $T(x)$ on paitsi x-ikäisten asiakkaiden keskimäärin systeemissä viettämä aika, samalla niiden asiakkaiden keskimääräinen kokonaisviipymä systeemissä, joiden palveluvaade on x, ts. palveluvaateeseen ehdollistettu keskiviipymä.

M/G/1-PS-jono (jatkoa)

- Voidaan edelleen päätellä, että

$$\lim_{x\to\infty} T(x) = \frac{x}{C(1-\rho)}$$

- Suuren työn saapuminen on harvinainen, yksittäinen tapahtuma. Työ viipyy systeemissä pitkään. PS-jonoon sinä aikana saapuneet muut (pienet) työt pääsevät "valumaan" ohi, ja iso työ näkee efektiivisenä palvelunopeutenaan muilta töiltä ylijääneen kaistan $C(1-\rho)$.

- Siten aikaisemmassa yhtälössä $T(x) = (n(0)/\lambda)\,x$ esiintyvä verrannollisuuskerroin $n(0)/\lambda$ on $1/C(1-\rho)$

$$\boxed{T(x) = \frac{x}{C(1-\rho)}}$$

- Keskiarvoistamalla tämä ehdollisen viipymäajan kaava työn jakauman suhteen ja soveltamalla Littlen tulosta, saadaan uudelleen

$$\boxed{\mathrm{E}[T] = \frac{1/\mu}{1-\rho}} \qquad 1/\mu = \mathrm{E}[X]/C, \qquad \boxed{\mathrm{E}[N] = \frac{\rho}{1-\rho}}$$

M/G/1-PS-jono (jatkoa)

- Uudelleenjohdetuissa keskiarvokaavoissa on merkittävää, että nyt ei tehty mitään oletusta palveluvaateen jakaumasta. PS-jonon keskiarvokaavat ovat siten insensitiivejä.
- Kaava $T(x) = x/C(1-\rho)$ kertoo, että asiakkaan keskimääräinen viipymä PS-jonossa on verrannollinen työn kokoon
  - jokaisen työn systeemissä keskimäärin viettämä aika on sen oma täyden kapasiteetin palveluaika x/C “venytettynä” tekijällä $1/(1-\rho)$
  - keskimäärin jokainen työ näkee efektiivisen palvelukapasiteetin $C(1-\rho)$.
- Näiden ominaisuuksiensa vuoksi PS-jonoa voidaan pitää tasapuolisimpana mahdollisena jonosysteeminä.

M/G/1-PS-jono (jatkoa)

- Pollaczek-Khinchinin kaavojen mukaan M/G/1-FIFO-jonossa keskimääräinen jononpituus ja viipymä ovat suurempia (pienempiä) kuin vastaavassa M/M/1-FIFO-jonossa, ja niin ollen M/G/1-PS-jonossa, mikäli palveluvaateen neliöllinen variaatiokerroin C_v^2 on suurempi (pienempi) kuin 1.
- Suuren neliöllisen variaatiokertoimen tapauksessa PS-jonokurin paremmuus keskiarvojen kannalta on helppo ymmärtää, koska
  - FIFO-jonossa suuri määrä pieniä töitä joutuu odottamaan suuren työn valmistumista
  - PS-jonossa ne pääsevät valumaan ohi
  - lukumääräisesti suuri määrä asiakkaita kokee siten PS-jonokurin parannuksena
- Pienen neliöllisen variaatiokertoimen tapauksessa (säännöllinen liikenne) kurinalaisempi FIFO on edullisempi.
- M/M/1-oletuksin, jolloin keskiarvot ovat samat, PS-jonon viivejakauman varianssi on suurempi kuin vastaavalla FIFO-jonolla. Niille voidaan johtaa lausekkeet:

$$\mathrm{V}[T]_{\mathrm{FIFO}} = \frac{1}{\mu^2(1-\rho)^2} \qquad \mathrm{V}[T]_{\mathrm{PS}} = \frac{1}{\mu^2(1-\rho)^2}\,\frac{2+\rho}{2-\rho}$$

jälkimmäinen tekijä on välillä $1 \ldots 3$

Esimerkki: alavirran dataliikenne solukkoverkossa

- 3G-solukkoverkkojen HSPDA-protokolla (High speed downlink packet access) perustuu aikajakoiseen multipleksointiin:
  - tukiasema lähettää täydellä teholla vain yhdelle käyttäjälle kussakin aikavälissä.
- Jos aikavälit jaetaan esim. kiertävässä järjestyksessä aktiivisten käyttäjien kesken, niin tukiasema muodostaa PS-jonon alavirran liikenteelle.
- Linkin mukautus (link adaptation): bittinopeus valitaan radiokanavan mukaisesti.
- Bittinopeus pienenee etäisyyden r kasvaessa signaalin heikentyessä, esim.,

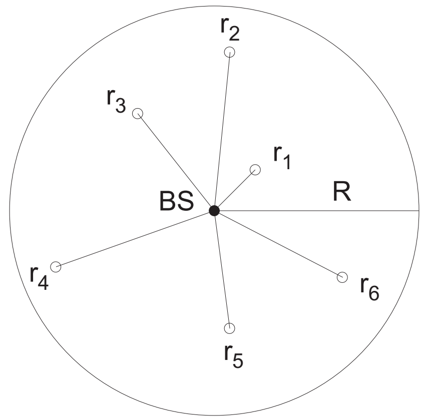
$$C(r) = \begin{cases} C_0 & r \le r_0 \\ C_0 \left(\dfrac{r_0}{r}\right)^{\alpha} & r > r_0 \end{cases}$$

  missä r_0 on kynnysetäisyys, jonka sisällä tietty maksiminopeus C_0 on saavutettavissa, ja eksponentti α on tyypillisesti välillä $2 \ldots 4$.
- $C(r)$ tarkoittaa etäisyydellä r olevan käyttäjän maksiminopeutta
  - jos solussa on n yhtäaikaista aktiivista käyttäjää, nopeus on $C(r)/n$.

Esimerkki (jatkoa)

Oletukset:

- Solua kuvataan R-säteisellä ympyrällä.
- Voita saapuu poissonisesti tukiasemalle nopeudella λ.
- Kunkin vuon koko X on arvottu riippumattomasti annetusta jakaumasta keskiarvolla $\bar{X}$.
- Kunkin vuon määränpään sijainti on satunnainen ympyrän sisällä (tasainen jakauma).



Palveluaika S (täydellä nopeudella, ilman jakoa voiden kesken)

$$S = \frac{X}{C(r)}$$

on satunnaismuuttuja, koska sekä X että r ovat satunnaismuuttujia. X ja r ovat kuitenkin toisistaan riippumattomia, joten

$$\bar{S} = \frac{\bar{X}}{\bar{C}} \quad \text{missä} \quad \frac{1}{\bar{C}} = \overline{\left(\frac{1}{C(r)}\right)} = \frac{1}{\pi R^2}\int_0^R \frac{2\pi r}{C(r)}\,dr$$

Esimerkki (jatkoa)

- Tukiaseman jono on PS-jono, jonka kuorma on $\rho = \lambda\bar{S} = \lambda\bar{X}/\bar{C}$.
- Stabiilissa jonossa pitää olla $\rho < 1$. Suurin ylläpidettävissä oleva liikennekuorma, $\lambda\bar{X}$, on siten $\bar{C}$ ja *solun kokonaikapasiteetti* [kbits/s] on

$$\bar{C} = \left(\frac{1}{\pi R^2} \int_0^R \frac{2\pi r}{C(r)}\, dr \right)^{-1}$$

- Koska systeemi muodostaa PS-jonon, tehollinen voiden näkemä bittinopeus (voiden läpäisy) etäisyydellä r on

$$C_{\mathit{eff}} = C(r)(1-\rho)$$

ja vuon, jonka koko on X, lähetysaika etäisyydellä r on keskimäärin

$$\bar{T}(r, X) = \frac{X}{C(r)(1-\rho)}$$