

Norges teknisk-naturvitenskapelige universitet
Institutt for telematikk

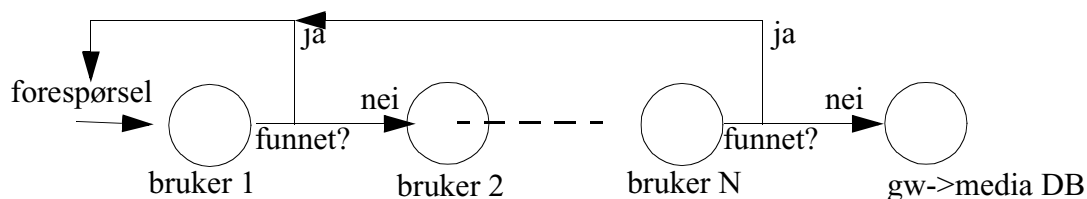


EKSAMEN I EMNE
SIE5015 PÅLITELIGHET OG YTELSE MED SIMULERING
2002

Løsningsforslag

Oppgave 1 Ytelsen til søkealgoritmen

- a) Anta at antall brukere N er uendelig. Ved bruk av søkealgoritmen som beskrevet over: Hva er fordelingen til antall søkesteg, X , i P2P nettet? (merk at hvis brukeren finner filmen hos seg selv så blir det $X = 0$ søkesteg) Hva kalles denne fordelingen? Beregn forventet søketid, $E(T)$, og varians, $Var(T)$, for den beskrevne algoritme. Sett $\tau_s = 2$ [ms], $p = 0.01$ og $N = \infty$



Figur 1 Søk etter film eller musikk

Antall søkesteg X er gitt av hvor mange påloggede brukere som besøkes før filmen finnes. Sannsynligheten for ikke å finne filmen hos en vilkårlig bruker er $1 - p$. Sannsynligheten for å gjøre x søkesteg er derved (merk at hvis filmen ligger på egen PC hvor søket starter så gjør vi $x = 0$ søkesteg med sannsynlighet $f_0 = p$):

$$f_x = P(X = x) = (1 - p)^x p, x = 0, 1, \dots \quad (1)$$

Fordelingen kalles *ordinær geometrisk fordeling*.

Forventet søketid T_s er

$$E(T_s) = E(X)\tau_s = \frac{0.99}{0.01} 2 = 198 \text{ [ms]} \quad (2)$$

Variansen til T_s er

$$Var(T_s) = Var(X)\tau_s^2 = \frac{0.99}{(0.01)^2} 2^2 = 39600 \quad (3)$$

- b) La nå antall brukere være endelig, $N < \infty$. Vis at sannsynligheten for å finne filmen blant de N brukerne er $P(X < N) = 1 - (1 - p)^N$.

Merk at hvis filmen ligger på egen PC så vil det bli $X = 0$ søkesteg, hvis den ligger på den siste av N (når N er endelig) så blir det $N - 1$ søkesteg. Dermed bli sannsynligheten for at filmen ligger i P2P nettet lik sannsynligheten for at vi har mindre enn N søkesteg.

$$\begin{aligned}
 P(X < N) &= \sum_{x=0}^{N-1} (1-p)^x p = 1 - \sum_{x=N}^{\infty} (1-p)^x p \\
 &= 1 - \sum_{x=0}^{\infty} (1-p)^{x+N} p = 1 - (1-p)^N \sum_{x=0}^{\infty} (1-p)^x p \\
 &= 1 - (1-p)^N
 \end{aligned} \tag{4}$$

- c) Beregn forventet søketid gitt at filmen finnes i P2P nettet med $N = 200$ brukere. Hva blir forventet søketid etter film som finnes i P2P eller i media-databasen? Tiden for søk etter film i media-databasene er $\tau_{db} = 150$ [ms]

Utgangspunkt av $E(X|(X < N))$ [denne ble oppgitt og skulle ikke regnes ut]

$$\begin{aligned}
 E(X|(X < N)) &= \sum_{x=0}^{\infty} x P(X=x|(X < N)) = \sum_{x=0}^{N-1} \frac{x(1-p)^x p}{1 - (1-p)^N} \\
 &= \sum_{x=0}^{\infty} \frac{x(1-p)^x p}{1 - (1-p)^N} - \sum_{x=N}^{\infty} \frac{x(1-p)^x p}{1 - (1-p)^N} \\
 &= \sum_{x=0}^{\infty} \frac{x(1-p)^x p}{1 - (1-p)^N} - \sum_{x=0}^{\infty} \frac{(x+N)(1-p)^{x+N} p}{1 - (1-p)^N} \\
 &= \sum_{x=0}^{\infty} \frac{x(1-p)^x p}{1 - (1-p)^N} - \frac{(1-p)^N}{1 - (1-p)^N} \left(\sum_{x=0}^{\infty} x(1-p)^x p + \sum_{x=0}^{\infty} N(1-p)^x p \right) \\
 &= \frac{E(X)}{1 - (1-p)^N} - \frac{(1-p)^N (E(X) + N)}{1 - (1-p)^N} \\
 &= \frac{E(X) - (1-p)^N (E(X) + N)}{1 - (1-p)^N} \\
 &= \frac{E(X)(1 - (1-p)^N) - N(1-p)^N}{1 - (1-p)^N} = E(X) - N \frac{1 - P(X < N)}{P(X < N)}
 \end{aligned} \tag{5}$$

Dermed blir søketid gitt at filmen finnes i P2P:

$$E(T) = \tau_s E(X) - \tau_s N \frac{1 - P(X < N)}{P(X < N)} = \tau_s E(X) - \tau_s N \frac{P(X \geq N)}{P(X < N)} \quad (6)$$

For søketid for filmer i P2P og database, $E(T)$, kan to alternative løsninger aksepteres, avhengig av hvordan søkealgoritmen fortolkes::

1. Søker alltid gjennom hele P2P nettet før søk i database
2. Vet hvilke filmer som ligger i P2P (men ikke hvor)

Løsning symbolsk for alternativ 1:

$$\begin{aligned} E(T) &= E(T|(X < N))P(X < N) + E(T|(X \geq N))P(X \geq N) \\ &= \tau_s E(X)P(X < N) - \tau_s N P(X \geq N) + (\tau_{db} + \tau_s)P(X \geq N) \\ &= \tau_s E(X)P(X < N) + \tau_{db} P(X \geq N) \\ &= \tau_s \frac{(1-p)}{p} (1 - (1-p)^N) + \tau_{db} (1-p)^N \end{aligned} \quad (7)$$

Løsning symbolsk for alternativ 2:

$$\begin{aligned} E(T) &= E(T|(X < N))P(X < N) + E(T|(X \geq N))P(X \geq N) \\ &= \tau_s E(X)P(X < N) - \tau_s N P(X \geq N) + \tau_{db} P(X \geq N) \\ &= \tau_s \frac{(1-p)}{p} (1 - (1-p)^N) - (\tau_s N - \tau_{db})(1-p)^N \end{aligned} \quad (8)$$

Nummerisk: Beregner først sannsynligheten for å finne filmen i P2P nettet

$$P(X < N) = 1 - (1-p)^N = 0.866 \quad (9)$$

deretter forventet antall søkesteg gitt at du finner filmen i P2P nettet:

$$E(X|X < N) = E(X) - N \frac{1 - P(X < N)}{P(X < N)} = \frac{1-p}{p} - \frac{(1-p)^N N}{1 - (1-p)^N} = 68.0585 \quad (10)$$

og så forventet søketid gitt at filmen er i P2P:

$$E(X|X < N) = \tau_s E(X|X < N) = 136.119 \text{ [ms]} \quad (11)$$

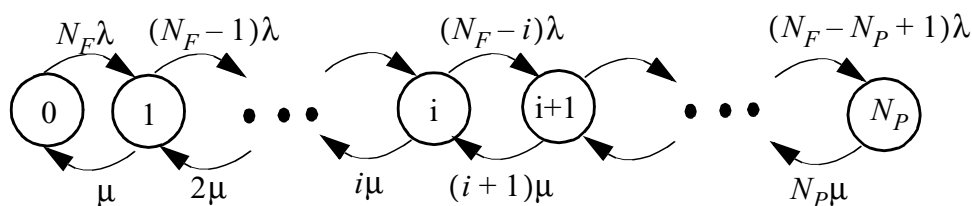
For $E(T)$,

1. Alternativ 1: $E(T) = 191.568$

2. Alternativ 2: $E(T) = 137.977$

- d) Beskriv en Markovmodell av systemet hvor tilstand i er lik antall filmer i P2P nettet. Marker tydelig ankomsts- og avgangsintensiteter og hvordan disse forandres som funksjon av tilstand i . Dette er en kjent modell. Hva kalles denne?

Modell 1



Figur 1.1 Tilstandsdiagram for Engsets systemmodell

Dette er en Engset-modell. Vi kan gjenkjenne at “ankomstintensiteten” (lik forespørselintensiteten) er lineært avtagende av hvor mange ressurser som er okkupert (filmer som er i nettet) og at avgangsintensiteten (intensiteten for at filmer blir utilgjengelig) er proporsjonal med antall filmer i systemer.

- e) Når $i = N_P$ vil nye forsøk på nedlasting av filmer fra mediedatabasene bli blokkert. Finn et uttrykk for sannsynligheten for at en film ikke kan lastes ned fordi det ikke er plass til flere. Finn et uttrykk for den andelen av tiden nettet er fullt.

Bruker π_i som betegnelse på stasjonærsannsynligheten for å være i tilstand i .

Nedlastningssperr, B , er den andelen nedlastningsforsøk som blir avvist delt på totalt antall nedlastingsforsøk. I vår modell blir dette:

$$B(N_P, N_F) = \frac{\lambda_{N_P} \pi_{N_P}}{\sum_{i=0}^{N_P} \lambda_i \pi_i} = \frac{(N_F - N_P) \lambda \pi_{N_P}}{\sum_{i=0}^{N_P} (N_F - i) \lambda \pi_i} \quad (12)$$

Tidssperr, E , er lik den andelen tid systemet har lastet ned filmer tilsvarende antall filmer tillatt i P2P nettet.

$$E(N_P, N_F) = \pi_{N_P} \quad (13)$$

Se forøvrig kompendiet hvor det er vist at $B(N_P, N_F) = E(N_P, N_F - 1)$.

Oppgave 2 Pålitelighet til en P2P-tjeneste

- a) Idet festen skal starte, ventes P2P nettet å være i en pålitelighetsmessig stasjonærtilstand mhp. antall arbeidende og feilte PCer. Forklar hva som menes dette. Hva er den pålitelighets-termen for sannsynligheten for at hun får levert film og musikk som ønsket idet festen starter?

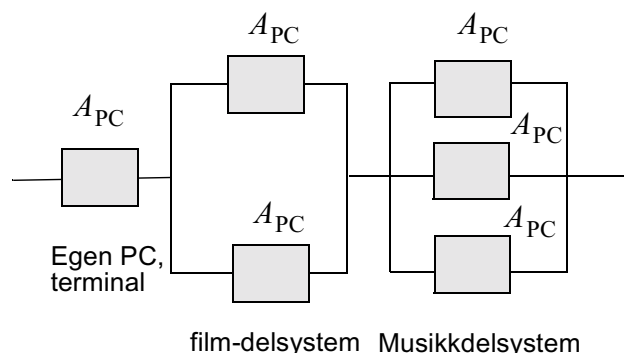
Systemet er stasjonært når sannsynligheten for et visst antall PCer er oppe og nede (alternativt: eller et spesifikt utvalg av PCer er oppe og nede) ikke endres med tiden.

Eksempel: $p_j(t) = P(j \text{ PCer er arbeidende ved tiden } t)$ så er systemet stasjonært ved tiden t_x dersom $(p_j(t_x) = p_j(t_y)), \forall (j, t_y > t_x)$, alternativt

$$\frac{d}{dt}p_j(t_x) = 0, \forall j.$$

Påliteligstermen er: (asymptotisk) tilgjengelighet.

- b) Hun ønsker å bestemme sannsynligheten, U_1 for at musikk eller film ikke leveres ved starten av festen. Tegn et pålitelighetsblokkskjema av systemet som leverer denne tjenesten og finn U_1 uttrykket ved A_{PC} .



Utilgjengeligheten (1- tilgjengeligheten) finnes ved ordinær blokkskjemareduksjon, her satt direkte opp:

$$\begin{aligned} U_1 &= 1 - (A_{PC}(1 - (1 - A_{PC})^2)(1 - (1 - A_{PC})^3)) \\ &= A_{PC}^6 - 5A_{PC}^5 + 9A_{PC}^4 - 6A_{PC}^3 + 1 \end{aligned} \quad (14)$$

- c) Angi hvilke elementer (PCer) i systemet som inngår i de minimale kuttsettene. Hvis vi antar at $(1 - A_{PC}) \ll 1$, forklar hvordan vi kan finne et rimelig anslag på svaret i punkt b ved å betrakte kuttsettene.

Det er tre minimale kuttsett med hhv. en, to og tre elementer,

- {Egen PC}
- {Film PC #1, Film PC #2}
- {Musikk PC #1, Musikk PC #2, Musikk PC #3}

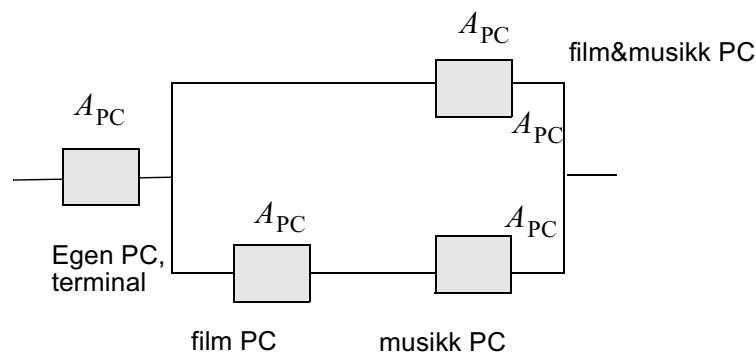
Siden alle elementene har samme sannsynlighet for ikke å virke vil det være det settet med færrest elementer som er størst. Siden denne sannsynligheten er vesentlig mindre enn en, vil dette settet dominere. I dette tilfelle er det utilgjeligheten av egen PC som vil dominere, dvs. $U_1 \approx 1 - A_{PC} = U_{PC}$.

Tilsvarende får vi fra foregående punkt:

$$\begin{aligned} U_1 &= 1 - ((1 - U_{PC})(1 - (U_{PC})^2)(1 - (U_{PC})^3)) \\ &= U_{PC}^6 - U_{PC}^4 + U_{PC} - U_{PC}^5 + U_{PC}^2 \approx U_{PC} \end{aligned} \quad (15)$$

- d) Finn et uttrykk for sannsynligheten for at festen kan gjennomføres uten at leveransen av film eller musikk feiler. Påpek hvilke av de ovenstående antakelsene du baserer fremgangsmåten på.

Med konfigurasjonen idet systemet tas i bruk, ved når festen starter, får vi følgende blokkdiagram



Siden alle PCer har konstant feilrate uavhengig av tidligere feil, har hver av de følgende funksjonssannsynlighet (Sannsynlighet for å feile etter tiden t):

$$R(t) = e^{-\lambda_{PC}t} \quad (16)$$

Siden vi har uavhengig feiling mellom PCene og disse antas å ikke bli reparert kan vi bruke blokkskjemametoden for å finne funksjonsannsynligheten til tjenesten:

$$R_S(t) = R(t)(1 - (1 - R(t))(1 - R(t)^2)) = R(t)^2 + R(t)^3 - R(t)^4 \quad (17)$$

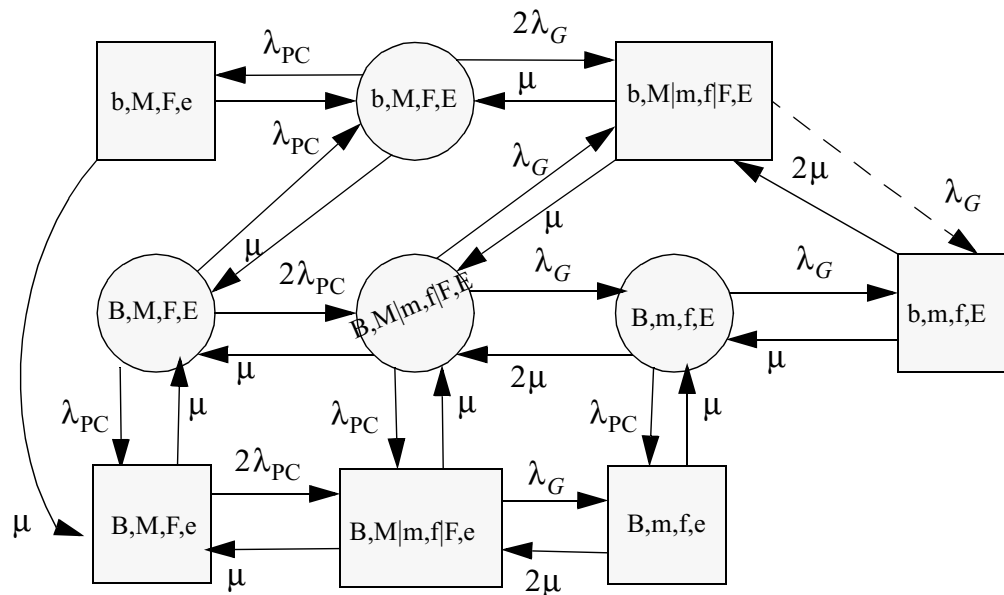
- e) Forklar kort hvorfor blokkskjema ikke kan benyttes til en pålitelighetsanalyse av systemet under disse forutsetningene. Hvilke(n) antakelse(r) må du gjøre ut over det som er gitt tidligere i oppgaven for at systemet skal kunne modelleres vha. et tilstandsdiagram (en Markov modell)?

Feilraten i de to “gjenlevende” PCene endres når ene går ned. Følgelig er ikke elementene i systemet uavhengig av hverandre, og blokkskjema kan ikke benyttes.

For å lage et tilstandsdiagram må vi anta et feil-avhjelpingstidene er negativt eksponensi-
alfordelt ([dette ikke nødvendig da det er gitt innledningsvis]og at tidene fra samme PC
også er uavhengige.)

- f) Tegn et tilstandsdiagram (en Markov modell) av systemet når tilleggsopplysningene tas i betraktning. (Det er ikke nødvendig å ekspandere modellen etter en systemfeiltilstand nås.) Angi tydelig hvilke PCer som er arbeidende og feilte i de ulike tilstandene i systemet samt feil og reparasjonsrater.

For å unngå unødige mange tilstander og et uoversiktlig diagram, utnyttes at feiling i PCen som inneholder kun musikk og PCen som inneholder kun film er symmetrisk.



Forkortelser:

X: PC X er arbeidende

x: PC X har feilet

B, b: PCen som "hoster" både filmen og musikken

M, m: PCen som "hoster" kun musikk.

F, e: PCen som "hoster" kun film

E, e: Studentens egen PC

M|m, f|F: Enten PC M har feilet eller PC f har feilet (men ikke begge)

μ : reparasjonsraten for en PC

Oppgave 3 Simulering av betalingsstrategier

- a) Beskriv en simuleringsmodell ved hjelp av aktivitetsdigram av en databasetjener hvor det er mulig å vurdere effekten av å endre antall brukere, N , køstørrelse, N_q , og antall tjenerprosesser (CPUer), N_{CPU} . (Tenk nøye gjennom hvilke detaljer som er tilstrekkelig og nødvendig for å vurdere tjenerens kvalitet under ulike konfigurasjoner). Hvordan kan du endre antall brukere, køstørrelsen, og antall CPUer i modellen din?

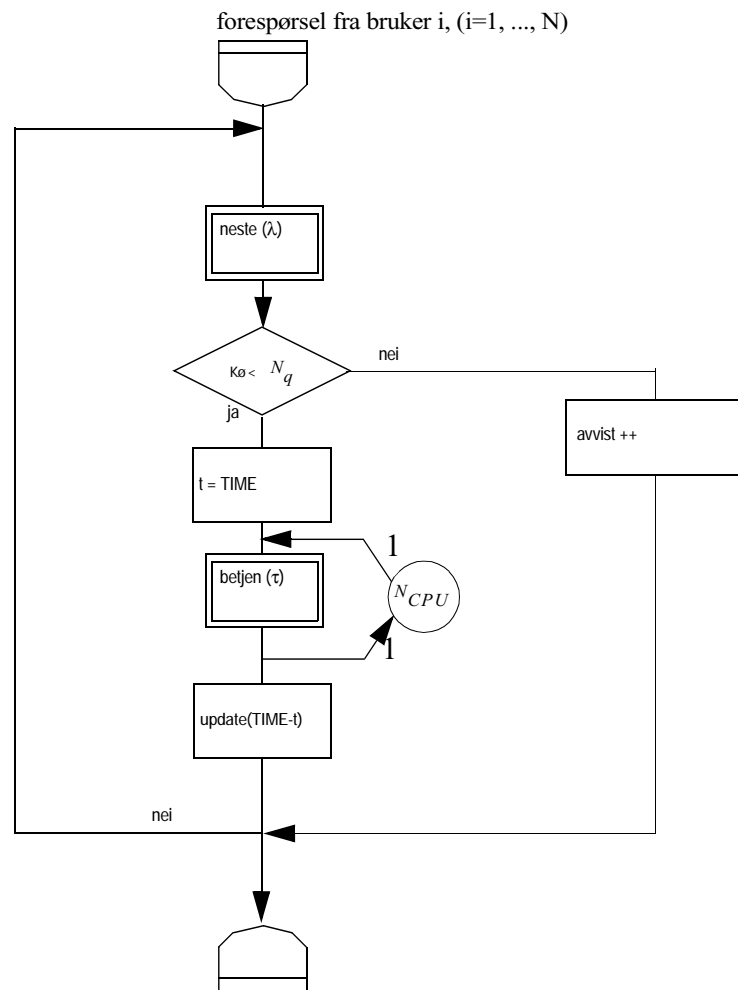
Velger å modellere “forespørsler” som entitet og tjenesteprosesser som ressurser. Felleskøen modelleres implisitt ved hjelp av køen som er relatert til ressurstypen, dvs. RES i DEMOS., Køgrensene sjekkes for en forespørsel legges i kø, dvs. før den gjør metodekall acquire til klassen RES. Databasen og query er ikke inkludert i modellen. Kun implisitt gjennom betjeningstiden τ .

Køgrensen blir en parameter som sjekkes i modellen, N_q .

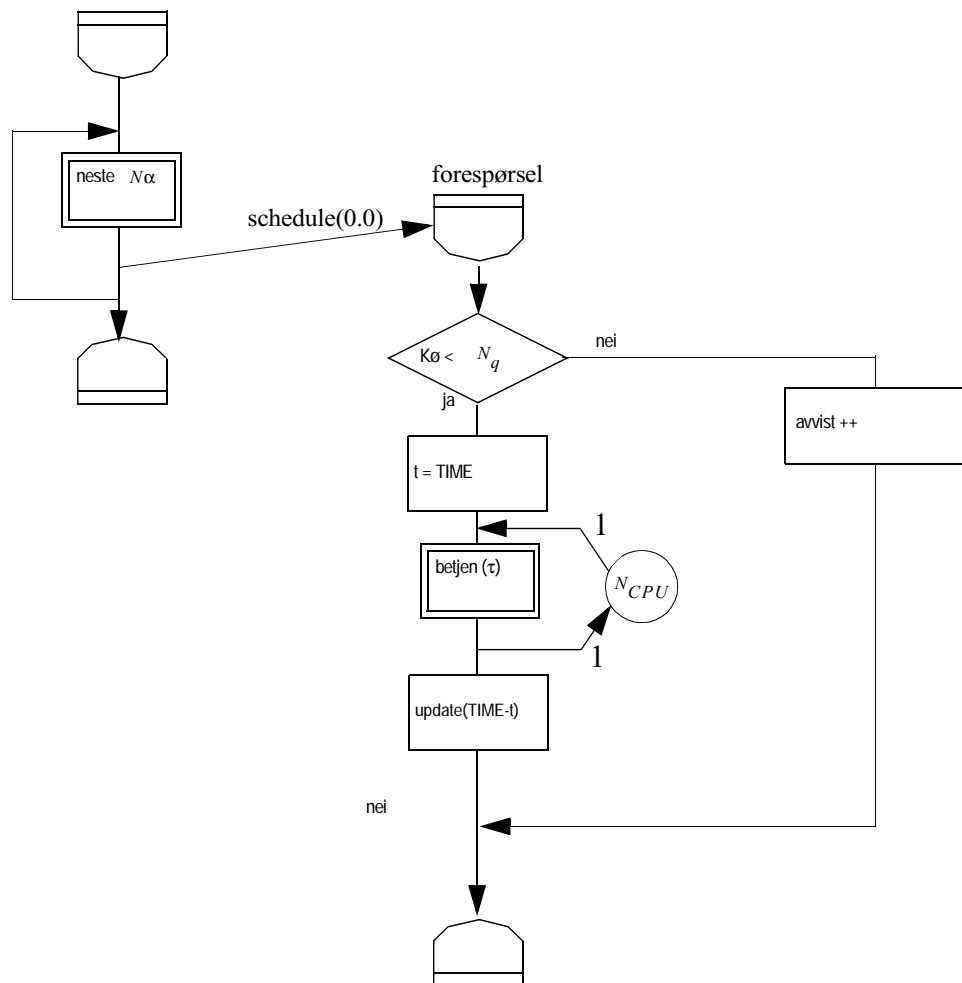
Antall CPUer settes ved oppstart som kapasitetsgrense for antall ressurser, N_{CPU} .

Generering av forespørsler kan modelleres enten ved hjelp av

1. Ekstern entitet eller bootstrap-metoden. Det startes opp en forespørsels-entiteter ved hjelp av en egen entitetsgenerator hvor tid mellom forespørsler er avhengig av antall brukere N , se figur 3.
2. Evig løkke i forespørselsentiteten (her vil nå tid mellom forespørsler påvirkes av av kø- og betjeningstid). Det startes opp en forespørsels-entitet per bruker N , se figur 2.



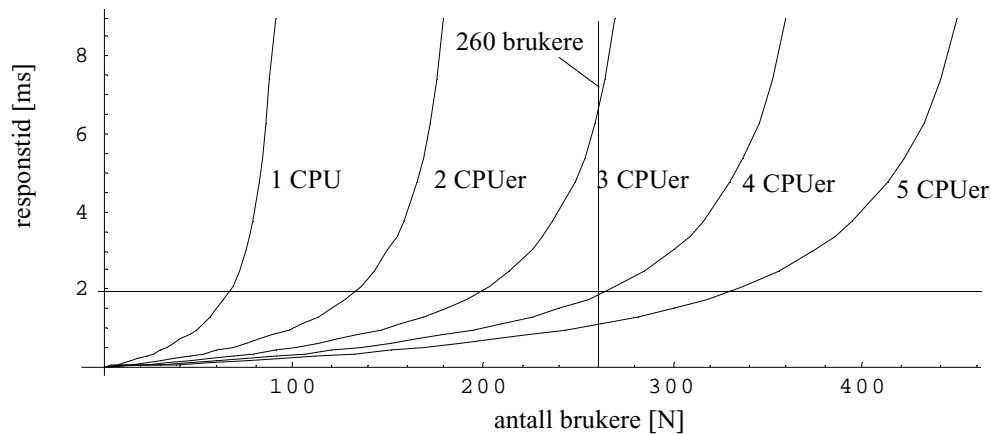
Figur 2 Nye forespørsler i evig løkke.



Figur 3 Nye forespørsler med ekstern entitetsgenerator.

- b) Anta at det er $N = 260$ brukere (markert som en strek i figuren). Spesifiser akseptablet responstidsnivå. Hvor mange tjenerprosesser er nødvendig for å oppfylle kravene? Gi en kvalitativ vurdering koplet til valg av responstidsnivå og antall brukere gitt en databasetjener som er konfigurert med alt fra 1 til 5 CPUer når det er $N = 260$ brukere. Forklar hvordan du kan innhente statistikken som

danner grunnlaget for kurvene i figur 4. Beskriv dette enten ved referanse til DEMOS primitiver og rapporter, eller markert i aktivitetsdiagrammene dine fra oppgave a).



Figur 4 Responsid som funksjon av antall brukere for ulike tjenerprosess-konfigurasjon

Antar et krav tilsvarende 2 millisekunder maksimal reponstid. (markert i figuren). Dette betyr at vi trenger 4 eller 5 CPUer.

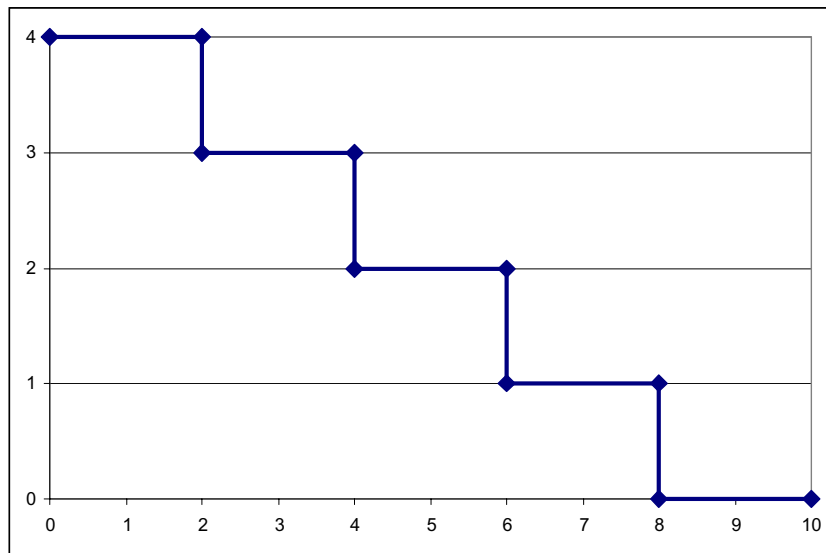
Kvalitativ vurdering:

- $N_{CPU} = 1, 2$: er helt uaktuelle uansett responstidskrav. Når antall brukere er så stort som 260 blir dette et tydelig et ustabilt køsystem hvor tid i kø går mot uendelig etter- som vi har uendelig kø (fra parameterene som er brukt til å generere kurvene så vil vi ha stabilt system når $N < 100N_{CPU}$, dvs. at for $N_{CPU} = 1, 2$ så er systemet stabilt når N er henholdsvis <100 og <200 .)
- $N_{CPU} = 3$: kan tilsynelatende brukes med responstidskrav større enn 7 ms, men bør unngås ettersom dette er i den bratte delen av kurven som gjør antakelsene våre svært følsomme for variasjoner i last og i antall brukere (f.eks. legg til 20 brukere og se hva som skjer)
- $N_{CPU} = 4$: kan brukes sålenge reponstidskravet er 2 ms og høyere, gjerne endel høyere
- $N_{CPU} = 5$: ser ut til å gi et godt dimensjonert system som oppfyller kravet vi har satt på 2 ms og kan tillate noe vekst i antall brukere.

Responstiden for ulike systemparametre kan vi ta ut av simuleringen ved å lese av slutt-rapporten til RES. Avg.wait + betjeningtiden som er konstant.

Alternativt er å legge inn lokal variabel som leser av TIME før RES tildelse, og etter at denne er frigitt. TALLY kan ta være på hver av responstidsobservasjonene og lage en sluttrapport.

- c) Lag et tidsplokk som viser antallet entiteter i systemet som funksjon av tiden i perioden $(0, T)$. Beregn de to verdiene i rapporten som mangler, % USAGE og AV. WAIT.



Figur 5 Plott av antall entiteter i systemet over tid.

Beregninger:

Utnyttelse: tiden hvor ressursen er i bruk/lengde av måleperiode = $8/10 = 80\%$

Gjennomsnittelig ventetid: $(\text{sum ventetid entiteter})/(\text{antall entiteter obs}) = (0+2+4+6)/4 = 3$

```

R E S O U R C E S
*****
TITLE      /   (RE)SET/   OBS/ LIM/ MIN/ NOW/   % USAGE/ AV. WAIT/QMAX
T           0.000      4   1   0   1   80.000   3.000   3

```

- d) Lag en simuleringsmodell hvor det er mulig å sammenlikne kostnadene ved valg av ulike strategier. Legg opp til at modellen studerer en spesifikk strategi ad gangen, og at modellen enkelt kan endres slik at andre strategier kan studeres (leie eller eie filmer? vente og “stream” filmer?). En mekanisme for å håndtere flyten elektroniske penger (betaling for film, inntekter fra vidredistribusjon, fylle opp konto) må eksplisitt modelleres. Beskriv modellen ved hjelp av aktivitetsdiagrammer. Mål for sammenlikning kan f.eks. være hva det kostet å laste ned et bestemt utvalg av filmer med bruk av ulike strategier.

Velger bruker som entitet. Penger modelleres ved hjelp av BIN som fylles opp med SK penger når brukeren logger på /registeres, og når brukeren ønsker å endre dette (når kontoen nærmer seg tom). Strategien for nedlasting kan gjøres hardkodet eller som en opsjon/argument ved instansiering.

