O.J.G. van der Velde

Optimaal stroomverbruik in een server farm

Bachelorscriptie, 23 juli 2014

Scriptiebegeleider: H. Blok, MSc.

Mathematisch Instituut, Universiteit Leiden

Inhoudsopgave

1 Inleiding

Een server farm bestaat uit een grote verzameling servers die samen een of meerdere taken uitvoeren die te groot zijn om op een machine uitgevoerd te kunnen worden. Server farms worden door steeds meer bedrijven gebruikt en voor bedrijven is het belangrijk om zo laag mogelijke kosten te hebben. Omdat server farms uit heel veel servers bestaan die allemaal stroom gebruiken is het voor een bedrijf heel interessant om een optimale strategie te hebben wat betreft het uitzetten of op standby zetten van servers. Hier kan namelijk veel geld mee bespaard worden. Er moet echter ook geld betaald worden om een server aan of uit te zetten. Hier moet dus ook rekening mee gehouden worden bij het bepalen van een strategie om de verwachte kosten zo laag mogelijk te houden.

In deze bachelorscriptie gaan we een model met oneindig veel servers beschouwen waarbij we ook rekening houden met het feit dat je geld op de bank kunt zetten waar je rente over kunt krijgen. Hierdoor hoef je dus minder geld op de rekening te zetten dan je uiteindelijk moet betalen, wat dan dus ook van invloed kan zijn op de optimale strategie.

We gaan met dit uitgebreide model, zoals ook beschreven in [2], een optimale strategie zoeken zodanig dat de verwachte verdisconteerde kosten zo laag mogelijk zijn en daarna gaan we op zoek gaan naar een bovengrens op het aantal klanten waarboven je alle servers altijd uitzet. Dit gaan we doen door globaal dezelfde technieken en ideeën die R. Kappetein in [1] geeft te gebruiken, maar dan voor het model waar we met verdisconteerde kosten rekenen in plaats van met gemiddelde verwachte kosten. In deze scriptie zal dan ook meestal het model met de verdisconteringsfactor beschouwd worden. Een enkele keer maken we echter gebruik van het model met gemiddelde verwachte kosten.

2 Het model

Beschouw een server farm met oneindig veel servers. Iedere server bevindt zich in een van de volgende drie toestanden: aan, op standby of uit. Servers die aan staan worden gebruikt door klanten, waarbij iedere klant één server gebruikt. Dus er geldt dat het aantal klanten in het systeem is gelijk aan het aantal servers dat aan staat. De klanten komen aan volgens een Poissonproces met parameter $\lambda > 0$. Alle klanten worden toegelaten tot het systeem. De tijd dat een klant een server gebruikt wordt gegeven door de exponentiële verdeling met parameter $\mu > 0$. Als de klant klaar is vertrekt hij weer uit het systeem. Op dat moment moet er beslist worden of de server op standby gezet wordt of dat de server uitgezet wordt.

Stel er zijn j klanten in het systeem. Dan is de tijd totdat de eerste klant vertrekt exponentieel verdeeld met parameter $j\mu$. Dit geldt wegens het feit dat de exponentiële verdeling geheugenloos is.

Het doel van het model is om een strategie te vinden die de kosten minimaliseert. Dit is als volgt te modelleren:

Beschouw de toestandsruimte $S = \mathbb{N}_0 \times \mathbb{N}_0$. Voor toestand $(i, j) \in S$ geldt dat $i =$ *het aantal servers dat op standby staat* en $j =$ *het aantal servers dat aan staat* $=$ *het aantal klanten*.

Er geldt dat een server die aan staat meer stroom verbruikt dan een server die op standby staat. Servers die uit staan verbruiken geen stroom. De stroomkosten van een server die op standby staat zijn per tijdseenheid $c_{standby} > 0$. Als er i servers op standby staan geeft dit per tijdseenheid dus de stroomkosten $i \cdot c_{standby}$. Een server die aan staat geeft per tijdseenheid stroomkosten $c_{aan} > 0$, maar omdat alle klanten geholpen worden, hebben deze kosten geen invloed op de strategie voor het uitzetten of op standby zetten van servers als er een klant vertrekt. Daarom mogen we de kosten c_{aan} negeren bij het bepalen van de strategie. Definieer daarom vanaf nu $c = c_{standby}$.

Als er geen servers meer op standby staan als er een klant arriveert, dan moet er een server aangezet worden. Dit opstarten van een server geeft eenmalig extra kosten $K^{aan} \geq 0$. Dit opstarten van de server kost geen tijd. Als er nog wel (een of meer) servers op standby staan, wordt er een server die op standby staat aangezet. Het uitzetten van een server die op standby staat geeft eenmalige extra kosten $K^{uit} \geq 0$. De kosten K^{aan} en K^{uit} worden niet op hetzelfde moment betaald. Dit kan uitmaken voor de optimale strategie wanneer we de verdisconteringsfactor geïntroduceerd hebben en daarom beschouwen we zowel de kosten K^{aan} als de kosten K^{uit}.

3 Verdisconteringsfactor en verdisconteerde kosten

Voordat we op zoek kunnen naar de optimale strategie, moet er eerst uitgelegd worden wat de verdisconteringsfactor is en hoe de verdisconteerde kosten in de toestanden $(0, j)$ en (i, j) met $i \in \mathbb{N}_{>0}$ en $j \in \mathbb{N}_{\geq 0}$ berekend worden. Dit zal gebeuren in dit hoofdstuk.

3.1 Verdisconteringsfactor

Het model beschreven in hoofdstuk 2 is een model met gemiddelde kosten en bevat dus geen verdisconteringsfactor. Er wordt dus geen rekening gehouden met het feit dat er geld op de bank gezet kan worden en dat daar rente over ontvangen kan worden. Hierdoor hoeft er voor kosten die in de toekomst gemaakt gaan worden minder geld op de bank gezet te worden dan er betaald moet worden. Met de verdisconteringsfactor α wordt dit toegevoegd aan het model.

In een discreet Markovproces met perioden $t = 0, 1, 2, ...$ geldt voor een beginkapitaal y en met rente r per tijdseenheid dat op tijdstip t een kapitaal van $(1 + r)^t \cdot y = x$ verkregen is. Dus om met een eindkapitaal x te eindigen, moet er begonnen worden met een beginkapitaal $y = (1 + r)^{-t} \cdot x = \left(\frac{1}{1+r}\right)^t \cdot x$. Hier is de verdisconteringsfactor gegeven door $\beta = \frac{1}{1+r} > 0$.

In een continu Markovproces is de verdisconteringsfactor gegeven door $\alpha > 0$. Dus om op tijdstip $t \geq 0$, ofwel $t \in \mathbb{R}_{\geq 0}$, een kapitaal x te hebben moet gelden dat het beginkapitaal $y = e^{-\alpha t} \cdot x$ is.

In het model in deze scriptie wordt een continu Markovproces beschouwd. Daarom wordt in de komende hoofdstukken $e^{-\alpha}$ met de verdisconteringsfactor $\alpha > 0$ beschouwd.

Voor ons model betekent de verdisconteringsfactor α dus dat kosten die je in de toekomst maakt minder zwaar meewegen in het bepalen van de strategie wat betreft het uitzetten of op standby zetten van servers dan kosten die je meteen moet maken.

Met behulp van dit model gaan we op zoek naar een optimale strategie die de verwachte totale verdisconteerde kosten minimaliseert.

3.2 Verdisconteerde kosten

Met behulp van de verdisconteringsfactor $\alpha > 0$ kunnen we vervolgens voor de toestanden $(0, j)$ en (i, j), met $i \in \mathbb{N}_{>0}$ en $j \in \mathbb{N}_{\geq 0}$ de verdisconteerde kosten bepalen.

3.2.1 Verdisconteerde kosten in toestand $(0, j)$ met $j \in \mathbb{N}_{\geq 0}$

Voor alle toestanden $(0, j)$ met $j \in \mathbb{N}_{\geq 0}$ geldt dat er een server aangezet moet worden als er een klant arriveert. In alle overige toestanden hoeft er nooit een server aangezet te worden, want dan staan er nog servers op standby en die worden eerst gebruikt. Om de verdisconteringsfactor en de verdisconteerde kosten uit te leggen beschouwen we het geval dat de kosten om een server aan te zetten $K^{aan} \geq 0$ zijn en om een server uit te zetten $K^{uit} = 0$ zijn.

Stel je bent in toestand $(0, j)$. Dan willen we weten wat de verwachte verdisconteerde kosten in toestand $(0, j)$ zijn totdat we wegspringen naar een andere toestand. Je springt weg uit toestand $(0, j)$ als er een klant aankomt of als er een klant vertrekt. De kans dat een klant arriveert is $\frac{\lambda}{\lambda+\mu j}$ en de kans dat er een klant vertrekt is dus $\frac{\mu j}{\lambda+\mu j}$.
Stel op tijdstip τ is de tijd tot de eerste sprong naar een andere toestand. Op tijdstip τ kan er dus zowel een klant arriveren als vertrekken. Dan zijn de verdisconteerde kosten voor het aanzetten van de server $K^{aan} e^{-\alpha\tau}$. Er geldt in toestand $(0, j)$ dat τ exponentieel verdeeld is met parameter $\lambda + \mu j$. Dan volgt dat de verwachte verdisconteerde kosten in toestand $(0, j)$ gegeven worden door

$$\begin{aligned} \mathbb{E}\left[K^{aan} e^{-\alpha\tau}\right] &= \int_{t=0}^{\infty} K^{aan} e^{-\alpha t} (\lambda + \mu j) e^{-(\lambda+\mu j)t} dt \\ &= K^{aan} (\lambda + \mu j) \int_{t=0}^{\infty} e^{-(\lambda+\mu j+\alpha)t} dt \\ &= K^{aan} \cdot \frac{\lambda + \mu j}{\lambda + \mu j + \alpha}. \end{aligned}$$

Omdat de kosten om een server aan te zetten K^{aan} zijn en om uit te zetten 0 zijn, volgt dat de verwachte verdisconteerde kosten over de tijd die in $(0, j)$ doorgebracht wordt gegeven zijn door

$$\frac{\lambda}{\lambda + \mu j} \cdot \mathbb{E}\left[K^{aan} e^{-\alpha\tau}\right] + \frac{\mu j}{\lambda + \mu j} \cdot 0 = \frac{\lambda}{\lambda + \mu j} \cdot K^{aan} \cdot \frac{\lambda + \mu j}{\lambda + \mu j + \alpha} + 0 = \frac{K^{aan} \lambda}{\lambda + \mu j + \alpha},$$

want als we toestand $(0, j)$ bezoeken moeten we een server aanzetten met kans $\frac{\lambda}{\lambda+\mu j}$ en we betalen niets als we een server uitzetten.

3.2.2 Verdisconteerde kosten in toestand (i, j) met $i \in \mathbb{N}_{>0}$ en $j \in \mathbb{N}_{\geq 0}$

Beschouw toestand (i, j) met $i \in \mathbb{N}_{>0}$ en $j \in \mathbb{N}_{\geq 0}$. Voor deze toestanden geldt dat er geen servers aangezet moeten worden, want er geldt dat $i \neq 0$. Voor deze toestanden beschouwen we dus alleen de stroomkosten van de servers die op standby staan. In toestand (i, j) met $i \neq 0$ zijn de stroomkosten per tijdseenheid gegeven door ci. Dan zijn de verdisconteerde kosten per tijdseenheid op tijdstip u gegeven door $cie^{-\alpha u}$, waarbij u de tijd is die we in toestand (i, j) doorbrengen.

Stel op tijdstip τ springen we naar een andere toestand. Dan geldt dat de verdisconteerde kosten op het tijdsinterval tussen tijdstip 0 en tijdstip τ gegeven worden door

$$\begin{aligned}\int_{u=0}^{\tau} ci \cdot e^{-\alpha u} du &= ci \cdot \int_{u=0}^{\tau} e^{-\alpha u} du \\ &= ci \cdot \left[-\frac{1}{\alpha} e^{-\alpha u}\right]_{u=0}^{\tau} \\ &= ci \left(-\frac{1}{\alpha} e^{-\alpha \tau} + \frac{1}{\alpha}\right) \\ &= \frac{ci}{\alpha}\left(1 - e^{-\alpha \tau}\right).\end{aligned}$$

Er geldt in toestand (i, j) met $i \neq 0$ dat τ exponentieel verdeeld is met parameter $\lambda + \mu j$. Dan volgt dat de verwachte verdisconteerde kosten in toestand (i, j), met $i \neq 0$, gegeven worden door

$$\begin{aligned}\mathbb{E}\left[\frac{ci}{\alpha}\left(1 - e^{-\alpha \tau}\right)\right] &= \int_{t=0}^{\infty} \frac{ci}{\alpha}\left(1 - e^{-\alpha t}\right)(\lambda + \mu j) e^{-(\lambda+\mu j)t} dt \\ &= \frac{ci}{\alpha}(\lambda + \mu j) \int_{t=0}^{\infty} \left(1 - e^{-\alpha t}\right) e^{-(\lambda+\mu j)t} dt \\ &= \frac{ci}{\alpha}(\lambda + \mu j)\left(\int_{t=0}^{\infty} e^{-(\lambda+\mu j)t} dt - \int_{t=0}^{\infty} e^{-(\lambda+\mu j+\alpha)t} dt\right) \\ &= \frac{ci}{\alpha}(\lambda + \mu j)\left(\frac{1}{\lambda + \mu j} - \frac{1}{\lambda + \mu j + \alpha}\right) \\ &= \frac{ci}{\alpha}\left(1 - \frac{\lambda + \mu j}{\lambda + \mu j + \alpha}\right) \\ &= \frac{ci}{\alpha} \cdot \frac{\alpha}{\lambda + \mu j + \alpha} \\ &= \frac{ci}{\lambda + \mu j + \alpha}.\end{aligned}$$

4 Eigenschappen van het model

In dit hoofdstuk zullen we een aantal definities geven en met behulp van deze definities zullen we een aantal eigenschappen van het model gaan bewijzen die heel handig blijken te zijn in de komende hoofdstukken.

Definitie 4.1. Voor $k \in \mathbb{N}_{\geq 0}$ wordt de diagonaal $D(k)$ gegeven door de volgende verzameling toestanden:

$$D(k) = \{(i, j) | i + j = k\}.$$

Definitie 4.2. Voor $k, l \in \mathbb{N}_{\geq 0}$ is diagonaal $D(l)$ een hogere diagonaal dan diagonaal $D(k)$ als $l > k$.

Definitie 4.3. Voor twee toestanden (i, j) en (k, l), met $i + j = m = k + l$, op diagonaal $D(m)$ geldt dat toestand (k, l) lager op diagonaal $D(m)$ ligt dan toestand (i, j) als geldt dat $i < k$ èn $j > l$.

Definitie 4.4. Voor alle $k \in \mathbb{N}_{\geq 0}$ wordt de driehoek $T(k)$ gegeven door de volgende verzameling toestanden:

$$T(k) = \{(i, j) | 0 \leq i + j \leq k\}.$$

Lemma 4.5. *Je kan alleen naar een hogere diagonaal in de toestanden* $(0, j)$*, met* $j \in \mathbb{N}_{\geq 0}$.

Bewijs. Sprongen naar een hogere diagonaal kunnen alleen gebeuren als er geen klanten in het systeem zijn en er arriveert een klant, want dan moet je een server aanzetten waardoor je naar een hogere diagonaal gaat. Dus alleen in toestand $(0, j)$, met $j \in \mathbb{N}_{\geq 0}$, kun je naar een hogere diagonaal springen. □

Lemma 4.6. (Lemma 2, Hoofdstuk 4 van [1]) *Je kan driehoek* $T(k)$ *alleen verlaten in het punt* $(0, k)$ *en de totale tijd om de driehoek te verlaten is onafhankelijk van de gekozen strategie.*

Een gevolg van zowel lemma 4.5 als lemma 4.6 is het volgende.
Gevolg 4.7. *Voor het model met verwachte gemiddelde kosten volgt dat het voldoende is om de gemiddelde verwachte kosten over eindige horizon om driehoek* $T(k)$ *te verlaten te beschouwen.*

Bewijs. Stel je bent in toestand (i, j) met $i + j = k$. Dan ben je dus in driehoek $T(k)$. Lemma 4.6 geeft dat het moment waarop je de driehoek verlaat onafhankelijk is van de strategie die je kiest en dat je de driehoek altijd in toestand $(0, k)$ verlaat. De strategie die je kiest heeft dus alleen invloed op de kosten die je maakt terwijl je in driehoek $T(k)$ bent. Als je naar een grotere driehoek gaat, dan zoek je vanaf dat tijdstip opnieuw naar een optimale strategie die de kosten minimaliseert.

Je kiest dus de strategie die de totale verwachte gemiddelde kosten om de driehoek te verlaten minimaliseert. Als je in toestand (i, j) bent en er vertrekt een klant, dan kun je kiezen om de server uit te zetten of op standby te zetten. De optimale keuze is dan om naar het punt te gaan die de laagste verwachte gemiddelde kosten heeft om driehoek $T(k)$ te verlaten. Omdat het maar een eindige tijd duurt voordat driehoek $T(k)$ verlaten wordt, volgt dat het voldoende is om de gemiddelde verwachte kosten over eindige horizon te beschouwen. □

Analoog aan het bewijs van gevolg 4.7, maar dan met verwachte verdisconteerde kosten in plaats van verwachte gemiddelde kosten, volgt gevolg 4.8.
Gevolg 4.8. *Voor het model met de verdisconteringsfactor* α *volgt dat het voldoende is om de totale verwachte verdisconteerde kosten over eindige horizon om driehoek* $T(k)$ *te verlaten te beschouwen.*

Er geldt dus dat het voldoende is om de verwachte (verdisconteerde) kosten te minimaliseren tot je driehoek $T(k)$ verlaat. Vanwege de geheugenloosheid van het systeem is het zelfs al voldoende om de kosten te minimaliseren tot je in toestand $(0, k)$ terecht komt.

5 Strategieën op een diagonaal

Met behulp van gevolg 4.7 en gevolg 4.8 volgt dus dat het niet nodig is om bij het zoeken naar een optimale strategie de hele toestandsruimte te beschouwen. Het is namelijk voldoende om de kosten te minimaliseren in driehoek $T(k)$. In dit hoofdstuk zullen we zowel voor het model met verwachte gemiddelde kosten als voor het model met verdisconteerde kosten de optimale actie bepalen in toestand $(i + 1, j - 1)$, met $i \in \mathbb{N}_{\geq 0}$ en $j \in \mathbb{N}_{>1}$, als geldt dat de optimale actie in toestand (i, j) het uitzetten van de server is als er een klant vertrekt.

5.1 Het model met verwachte gemiddelde kosten

Beschouw nu eerst het model met verwachte gemiddelde kosten. We beschouwen de situatie dat de kosten K^{aan} betaald moeten worden bij het aanzetten van een server en waarbij de kosten om een server uit te zetten buiten beschouwing laten. De situatie dat de kosten K^{uit} betaald moeten worden bij het uitzetten van een server is gegeven in stelling 1 in hoofdstuk 6 van [1].

Definitie 5.1. Zij $U^k(i,j)$ de verwachte kosten om driehoek $T(k)$ te verlaten vanuit toestand (i,j) als de optimale strategie gebruikt wordt.

Definitie 5.2. Zij $b^k_{(i,j)}(S)$ de verwachte kosten om diehoek $T(k)$ te verlaten vanuit toestand (i,j) als strategie S gebruikt wordt.

Definitie 5.3. Zij $\pi(i,j)$, met $i,j \in \mathbb{N}_{>0}$, de gekozen actie in toestand (i,j) wat betreft het uitzetten of op standby zetten van de server als er in toestand (i,j) een klant vertrekt. Definieer $\pi^*(i,j)$ als de optimale actie in toestand (i,j).

Er geldt dat als het optimaal is om een server uit te zetten in toestand (i,j) met $j > 1$, dan is het optimaal om de server uit te zetten in toestand $(i+1, j-1)$. Dit kunnen we verwoorden in de volgende stelling.

Stelling 5.4. *Beschouw het model met verwachte gemiddelde kosten. Zij K^{aan} de kosten om een server aan te zetten. Als voor $j > 1$ geldt dat $\pi^*(i,j) =$ uitzetten, dan geldt dat $\pi^*(i+1, j-1) =$ uitzetten.*

Bewijs. Dit bewijs volgt het idee van het bewijs van stelling 1 in hoofdstuk 6 van [1]. We gebruiken hier echter andere notatie en meer uitleg bij sommige stappen.

Stel $\pi^*(i,j) =$ uitzetten. Dan geldt dat de verwachte kosten om driehoek $T(i+j)$ te verlaten lager zijn als je naar toestand $(i, j-1)$ gaat, dan wanneer je naar toestand $(i+1, j-1)$ zou gaan, ofwel

$$U^{i+j}(i, j-1) \leq U^{i+j}(i+1, j-1). \tag{1}$$

Stel nu dat je in toestand $(i+1, j-1)$ begint. Deze toestand ligt op diagonaal $D(i+j)$. Je wilt weten of je in toestand $(i+1, j-1)$ de server wel of niet uit moet zetten als er een klant vertrekt.

Als geldt dat

$$U^{i+j}(i+1, j-2) \leq U^{i+j}(i+2, j-2), \tag{2}$$

dan volgt dat $\pi^*(i+1, j-1) =$ uitzetten.

De verwachte kosten voor de strategie die de verwachte kosten om driehoek $T(i+j)$ te verlaten als je in toestand $(i+1, j-2)$ bent minimaliseert, zijn van boven begrensd door de verwachte kosten van iedere andere strategie die je kiest in toestand $(i+1, j-2)$. Dus als je één strategie voor toestand $(i+1, j-2)$ hebt die goedkoper is dan iedere strategie in toestand $(i+2, j-2)$, dan is het optimaal om wel naar toestand $(i+1, j-2)$ te gaan en om niet naar toestand $(i+2, j-2)$ te gaan.

Voor iedere strategie vanaf toestand $(i+2, j-2)$ moet er een moment zijn dat je naar een lagere diagonaal gaat of dat je naar toestand $(i+1, j-1)$ gaat. Noem het tijdstip dat dit gebeurt tijdstip t. Zo'n tijdstip t bestaat, want wegens lemma 4.6 is de totale tijd om driehoek $T(i+j)$ te verlaten onafhankelijk van de gekozen strategie. Verder geldt dat de verwachte tijd die het kost om $T(i+j)$ te verlaten eindig is. Ook geldt er dat om $T(i+j)$ te verlaten je òf via toestand $(i+1, j-1)$ òf via een lagere diagonaal moet gaan. Beschouw dus de strategie

$$\begin{aligned} S_2 = \{ & \text{Zet op tijdstip 0 de server op standby in toestand } (i+1, j-1), \\ & \text{voer daarna de optimale strategie uit tot je uit } T(i+j) \text{ bent}\}. \end{aligned}$$

Dan is er voor strategie S_2 zo'n tijdstip t, want als je een server op standby zet, dan kom je in toestand $(i+2, j-2)$. Er geldt dat

$$b^{i+j}_{(i+1,j-1)}(S_2) = U^{i+j}(i+2, j-2), \tag{3}$$

want S_2 is de optimale strategie vanaf toestand $(i+2, j-2)$.

Kies dan nu in toestand $(i+1, j-2)$ de volgende strategie: als een klant vertrekt voor tijdstip t, zet dan de server op standby. Dit betekent dat je tot tijdstip t op de diagonaal $D(i+j-1)$ blijft. We beschouwen dus de strategie

$$\begin{aligned} S_1 \quad = \quad & \{\text{Zet op tijdstip 0 de server uit in toestand } (i+1, j-1), \\ & \text{zet daarna tot tijdstip } t \text{ de servers op standby als er een klant vertrekt} \\ & \text{en voer daarna de optimale strategie uit tot je uit } T(i+j) \text{ bent}\}\,. \end{aligned}$$

Er geldt dat

$$b^{i+j}_{(i+1,j-1)}(S_1) \geq U^{i+j}(i+1, j-2), \tag{4}$$

want S_1 is een strategie vanaf toestand $(i+1, j-2)$ en niet noodzakelijk de optimale strategie.

We zullen laten zien dat

$$b^{i+j}_{(i+1,j-1)}(S_1) \leq b^{i+j}_{(i+1,j-1)}(S_2)$$

geldt, want dan volgt dat vergelijking 2 geldt en dat is wat we moeten bewijzen.

We gaan nu strategie S_1 vergelijken met strategie S_2 tot tijdstip t. Dan zijn er 2 mogelijkheden:

1) òf beide strategieën bevinden zich op tijdstip t in dezelfde toestand,
2) òf strategie S_1 bevindt zich in toestand $(i, j-1)$ en strategie S_2 in toestand $(i+1, j-1)$.

Waarom hebben we deze twee mogelijkheden?
Voor de toestanden $(i+1, j-2)$ en $(i+2, j-2)$ geldt dat klanten met dezelfde snelheid vertrekken, want er zijn in beide toestanden evenveel klanten in het systeem. Klanten vertrekken in deze toestanden met snelheid $(j-2)\mu$. Dus voor deze twee toestanden geldt dat als je in de ene strategie een aankomst of vertrek hebt, je tijdens het vergelijken van de strategieën S_1 en S_2 ook een aankomst of vertrek hebt in de andere strategie. Er geldt namelijk dat het aantal klanten in het systeem onafhankelijk is van de gekozen strategie. Voor strategie S_1 geldt dat je bij het vertrekken van een klant de server tot tijdstip t altijd op standby zet. En voor strategie S_2 geldt dat je op tijdstip t naar een lagere diagonaal gaat of dat je in toestand $(i+1, j-1)$ bent.

1) Stel je beschouwt strategie S_2 en je gaat op tijdstip t naar een lagere diagonaal. Omdat voor beide strategieën geldt dat aankomsten en vertrekken op hetzelfde moment gebeuren, volgt dus dat je op tijdstip t met strategie S_1 je de server uitzet en met strategie S_2 zet je de server op standby. Omdat de strategieën voor tijdstip t iedere keer dat er een aankomst of vertrek evenveel klanten in het systeem hebben, maar strategie S_1 altijd één server minder op standby heeft staan dan strategie S_2, volgt dus dat op tijdstip t beide strategieën in dezelfde toestand eindigen.

2) Verder kan strategie S_1 in toestand $(i, j-1)$ zijn, want strategie S_2 kan voor de sprong op tijdstip t in toestand $(i+1, j-1)$ zijn, want voor iedere strategie S is er een moment dat je naar een lagere diagonaal springt of dat je naar toestand $(i+1, j-1)$ gaat. Dan volgt weer dat de strategieën voor tijdstip t iedere keer dat er een aankomst of vertrek is evenveel klanten in het systeem hebben, maar strategie S_1 heeft altijd één server minder op standby

staan dan strategie S_2. Voor strategie S_2 geldt dat je op een gegeven moment in toestand $(i+1, j-1)$ kunt zijn en dan volgt dus voor strategie S_1 dat je op dat moment in toestand $(i, j-1)$ bent, want met beide strategieën zijn er evenveel klanten in het systeem.

Deze twee mogelijkheden waar je je op tijdstip t kunt bevinden zijn de enige twee mogelijkheden om je te bevinden, want voor de strategie van toestand $(i+1, j-2)$ blijf je tot tijdstip t op diagonaal $D(i+j)$ en daarna kun je pas wegspringen naar een lagere diagonaal. Verder is tijdstip t ook zo gedefinieerd dat tijdstip t het moment is dat je naar een lagere diagonaal springt of dat je in toestand $(i+1, j-1)$ bent.

Beschouw dan nu beide mogelijkheden waar je je op tijdstip t kunt bevinden:

1) Beschouw de toestand (i', j'), met $i' + j' = i + j - 1$, zodanig dat op tijdstip t de strategieën S_1 en S_2 beide in toestand (i', j') zijn. Als beide strategieën op tijdstip t in toestand (i', j') zijn beland, dan heb je meer betaald als je strategie S_2 volgde dan als je strategie S_1 volgde, want je hebt tot tijdstip t één server meer op standby staan in strategie S_2, wat extra stroomkosten oplevert. Verder zijn vanaf tijdstip t de verwachte gemiddelde kosten gelijk, want beide strategieën starten in hetzelfde punt en dan kun je vanaf dit punt dezelfde optimale strategie kiezen. Hieruit volgt dat

$$b^{i+j}_{(i+1,j-1)}(S_1) \leq b^{i+j}_{(i+1,j-1)}(S_2),$$

want tot tijdstip t heb je met strategie S_2 meer betaald dan met strategie S_1.

2) Als op tijdstip t geldt dat strategie S_1 in toestand $(i, j-1)$ beland is, dan is strategie S_2 in toestand $(i+1, j-1)$ beland. Tot tijdstip t geldt dat strategie S_1 goedkoper is dan strategie S_2, want tot tijdstip t geldt dat er voor strategie S_2 één server meer op standby staat dan voor strategie S_1. Verder geldt er dat $b^{i+j}_{(i+1,j-1)}(S_1) = f(S_1) + U^{i+j}(i, j-1)$ en dat $b^{i+j}_{(i+1,j-1)}(S_2) = f(S_2) + U^{i+j}(i+1, j-1)$, waarbij $f(S)$ de gemiddelde verwachte stroomkosten zijn van de servers die tussen tijdstip 0 en t op standby staan via strategie S. Omdat er met strategie S_1 één server minder op standby staat dan met strategie S_2 volgt dat $f(S_1) \leq f(S_2)$. Met behulp van vergelijking 1 en met behulp van de ongelijkheid $f(S_1) \leq f(S_2)$ volgt er dat geldt dat

$$b^{i+j}_{(i+1,j-1)}(S_1) = f(S_1) + U^{i+j}(i, j-1) \leq f(S_2) + U^{i+j}(i+1, j-1) = b^{i+j}_{(i+1,j-1)}(S_2).$$

Hieruit volgt dan dat

$$b^{i+j}_{(i+1,j-1)}(S_1) \leq b^{i+j}_{(i+1,j-1)}(S_2).$$

Dus in beide situaties geldt dat

$$b^{i+j}_{(i+1,j-1)}(S_1) \leq b^{i+j}_{(i+1,j-1)}(S_2).$$

Hieruit volgt met behulp van vergelijking 3 en vergelijking 4 dat

$$U^{i+j}(i+1, j-2) \leq b^{i+j}_{i+1,j-1}(S_1) \leq b^{i+j}_{i+1,j-1}(S_2) = U^{i+j}(i+2, j-2),$$

waaruit volgt dat $\pi^*(i+1, j-1) =$ uitzetten. □

Door stelling 5.4 herhaaldelijk toe te passen volgt het volgende.

Gevolg 5.5. *Beschouw het model met verwachte gemiddelde kosten. Zij K^{aan} de kosten om een server aan te zetten en zij $K^{uit} = 0$. Als er een toestand $(i, j) \in D(k)$ is waarvoor geldt dat $\pi^*(i, j) =$ uitzetten, dan volgt voor alle toestanden op diagonaal $D(k)$ lager dan toestand (i, j) ook dat het optimaal is om de server uit te zetten als er een klant vertrekt.*

Voor het model met verwachte gemiddelde kosten geldt er dus dat als het in een toestand optimaal is om de server uit te zetten, dat het dan ook voor alle toestanden die lager op dezelfde diagonaal liggen optimaal is om de server uit te zetten.

5.2 Het model met verdisconteerde kosten waarbij K^{aan} betaald moet worden bij het aanzetten van een server

We beschouwen nu het model met verdisconteerde kosten. Hierbij beschouwen we de situatie dat K^{aan} betaald moet worden bij het aanzetten van een server en waarbij we de kosten om een server uit te zetten buiten beschouwing laten.
Definitie 5.6. Zij $Q^k(i,j)$ de verwachte verdisconteerde kosten om driehoek $T(k)$ te verlaten vanuit toestand $(i,j) \in T(k)$ als de optimale strategie gebruikt wordt.
Definitie 5.7. Zij $d^k_{(i,j)}(S)$ de verwachte verdisconteerde kosten om driehoek $T(k)$ te verlaten vanuit toestand $(i,j) \in T(k)$ als de strategie S gebruikt wordt.

Verder wordt definitie 5.3 weer gebruikt.
Stelling 5.8. *Beschouw het model met de verdisconteringsfactor $\alpha > 0$. Zij K^{aan} de kosten die betaald moeten worden bij het aanzetten van een server. Als voor $j > 1$ geldt dat $\pi^*(i,j)$ = uitzetten, dan geldt dat $\pi^*(i+1,j-1)$ = uitzetten.*

Bewijs. Dit bewijs heeft eenzelfde structuur als het bewijs van stelling 5.4.

Stel $\pi^*(i,j)$ = uitzetten. Dan geldt dat de verwachte verdisconteerde kosten om driehoek $T(i+j)$ te verlaten lager zijn als je naar toestand $(i,j-1)$ gaat, dan wanneer je naar toestand $(i+1,j-1)$ zou gaan, ofwel
$$Q^{i+j}(i,j-1) \leq Q^{i+j}(i+1,j-1). \tag{5}$$

Stel nu dat je in toestand $(i+1,j-1)$ begint. Deze toestand ligt op diagonaal $D(i+j)$. Je wilt weten of je in toestand $(i+1,j-1)$ de server wel of niet uit moet zetten als er een klant vertrekt. Als geldt dat
$$Q^{i+j}(i+1,j-2) \leq Q^{i+j}(i+2,j-2), \tag{6}$$
dan weten we dat $\pi^*(i+1,j-1)$ = uitzetten.

De verwachte verdisconteerde kosten voor de strategie die de verwachte verdisconteerde kosten om driehoek $T(i+j)$ te verlaten als je in toestand $(i+1,j-2)$ bent minimaliseert, zijn van boven begrensd door de verwachte verdisconteerde kosten van iedere andere strategie die je kiest in toestand $(i+1,j-2)$. Dus als je één strategie voor toestand $(i+1,j-2)$ hebt die goedkoper is dan iedere strategie in toestand $(i+2,j-2)$, dan is het optimaal om wel naar toestand $(i+1,j-2)$ te gaan en om niet naar toestand $(i+2,j-2)$ te gaan.

Voor iedere strategie in toestand $(i+2,j-2)$ moet er een moment zijn dat je naar een lagere diagonaal gaat of dat je naar toestand $(i+1,j-1)$ gaat. Noem het tijdstip dat dit gebeurt tijdstip t. Zo'n tijdstip t bestaat, want wegens lemma 4.6 is de totale tijd om driehoek $T(i+j)$ te verlaten onafhankelijk van de gekozen strategie. Verder geldt dat de verwachte tijd die het kost om $T(i+j)$ te verlaten eindig is. Ook geldt er dat om $T(i+j)$ te verlaten je òf via toestand $(i+1,j-1)$ òf via een lagere diagonaal moet gaan. Beschouw dus de strategie

$$\begin{aligned} S_2 \quad = \quad & \{\text{Zet op tijdstip 0 de server op standby in toestand } (i+1,j-1), \\ & \text{voer daarna de optimale strategie uit tot je uit } T(i+j) \text{ bent}\}. \end{aligned}$$

Dan is er voor strategie S_2 zo'n tijdstip t, want als je een server op standby zet, dan kom je in toestand $(i+2, j-2)$. Er geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_2) = Q^{i+j}(i+2, j-2), \tag{7}$$

want S_2 is de optimale strategie vanaf toestand $(i+2, j-2)$.

Kies dan nu in toestand $(i+1, j-2)$ de volgende strategie: als een klant vertrekt voor tijdstip t, zet dan de server op standby. Dit betekent dat je tot tijdstip t op de diagonaal $D(i+j-1)$ blijft. We beschouwen dus de strategie

$$\begin{aligned} S_1 \quad = \quad & \{\text{Zet op tijdstip 0 de server uit in toestand } (i+1, j-1), \\ & \text{zet daarna tot tijdstip } t \text{ de server op standby als er een klant vertrekt} \\ & \text{en voer daarna de optimale strategie uit tot je uit } T(i+j) \text{ bent}\}. \end{aligned}$$

Er geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \geq Q^{i+j}(i+1, j-2), \tag{8}$$

want S_1 is een strategie vanaf toestand $(i+1, j-2)$ en niet noodzakelijk de optimale strategie.

We zullen laten zien dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2)$$

geldt, want dan volgt dat vergelijking 6 geldt en dat is wat we moeten bewijzen.

We gaan nu strategie S_1 vergelijken met strategie S_2 tot tijdstip t. Dan zijn er 2 mogelijkheden:

1) òf beide strategieën bevinden zich op tijdstip t in dezelfde toestand,
2) òf strategie S_1 bevindt zich in toestand $(i, j-1)$ en strategie S_2 in toestand $(i+1, j-1)$.

Analoog aan het bewijs van stelling 5.4 voor de situatie zonder verdisconteerde kosten is te bewijzen dat dit inderdaad de 2 mogelijkheden zijn.

Beschouw dan nu beide mogelijkheden waar je je op tijdstip t kunt bevinden:

1) Beschouw de toestand (i', j'), met $i' + j' = i + j - 1$, zodanig dat op tijdstip t de strategieën S_1 en S_2 beide in toestand (i', j') zijn. Als beide strategieën op tijdstip t in toestand (i', j') zijn beland, dan heb meer betaald als je strategie S_2 volgde dan als je strategie S_1 volgde, want je hebt tot tijdstip t met strategie S_2 één server meer op standby staan met strategie S_1. Verder zijn vanaf tijdstip t de verwachte verdisconteerde kosten gelijk, want vanaf tijdstip t kun je dezelfde optimale strategie kiezen. Hieruit volgt dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2),$$

want tot tijdstip t heb je met strategie S_2 meer moeten betalen dan met strategie S_1.

2) Als op tijdstip t geldt dat strategie S_1 in toestand $(i, j-1)$ beland is, dan is strategie S_2 in toestand $(i+1, j-1)$ beland. Tot tijdstip t geldt dat strategie S_1 goedkoper is dan strategie S_2, want tot tijdstip t geldt dat er voor strategie S_2 één server meer op standby staat dan voor strategie S_1. Verder geldt er dat $d^{i+j}_{(i+1,j-1)}(S_1) = f(S_1) + Q^{i+j}(i, j-1)e^{-\alpha t}$ en dat $d^{i+j}_{(i+1,j-1)}(S_2) = f(S_2) + Q^{i+j}(i+1, j-1)e^{-\alpha t}$, waarbij $f(S)$ de verwachte verdisconteerde stroomkosten zijn van de servers die tussen tijdstip 0 en t op standby staan via strategie S. Omdat er met strategie S_1 één server minder op standby staat dan met strategie S_2

volgt dat $f(S_1) \leq f(S_2)$. Met behulp van vergelijking 5 en met behulp van de ongelijkheid $f(S_1) \leq f(S_2)$ volgt er dat geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_1) = f(S_1)+Q^{i+j}(i,j-1)e^{-\alpha t} \leq f(S_2)+Q^{i+j}(i+1,j-1)e^{-\alpha t} = d^{i+j}_{(i+1,j-1)}(S_2).$$

Hieruit volgt dan dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2).$$

Als op tijdstip t geldt dat strategie S_1 in toestand $(i, j-1)$ beland is, dan is strategie S_2 in toestand $(i+1, j-1)$ beland. Dan geldt er dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \cdot e^{-\alpha t} \leq d^{i+j}_{(i+1,j-1)}(S_2) \cdot e^{-\alpha t},$$

ofwel,

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2),$$

wegens de aanname dat $\pi^*(i,j) =$ uitzetten en omdat beide kanten met dezelfde verdisconteringsfactor $e^{-\alpha t}$ vermenigvuldigd moeten worden omdat het tot tijdstip t duurt voordat je in de toestanden $(i, j-1)$ en $(i+1, j-1)$ bent. Tot tijdstip t geldt dat strategie S_1 goedkoper is dan strategie S_2, want tot tijdstip t geldt dat er voor strategie S_2 één server meer op standby staat dan voor strategie S_1. Hieruit volgt dan dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2).$$

Dus in beide situaties geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2).$$

Hieruit volgt dan met behulp van vergelijking 7 en vergelijking 8 dat

$$Q^{i+j}(i+1, j-2) \leq d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2) = Q^{i+j}(i+2, j-2),$$

waaruit volgt dat $\pi^*(i+1, j-1) =$ uitzetten. □

Door stelling 5.8 herhaaldelijk toe te passen krijgen we het volgende.

Gevolg 5.9. *Beschouw het model met de verdisconteringsfactor $\alpha > 0$. Zij K^{aan} de kosten om een server aan te zetten. Als er een toestand $(i,j) \in D(k)$ is waarvoor geldt dat $\pi^*(i,j) =$ uitzetten, dan volgt voor alle toestanden op $D(k)$ lager dan toestand (i,j) ook dat het optimaal is om de server uit te zetten als er een klant vertrekt.*

Voor het model met de verdisconteringsfactor $\alpha > 0$ waarbij K^{aan} de kosten zijn om er server aan te zetten en waarbij we de kosten om een server uit te zetten buiten beschouwing laten geldt er dus dat als het in een toestand optimaal is om de server uit te zetten, dat dan ook voor alle toestanden lager op dezelfde diagonaal dat het optimaal is om de server uit te zetten.

5.3 Het model met verdisconteerde kosten waarbij K^{uit} betaald moet worden bij het uitzetten van een server

Beschouw de situatie dat K^{uit} betaald moet worden bij het uitzetten van een server en waarbij we de kosten om een server aan te zetten buiten beschouwing laten. Voor deze situatie moeten we twee situaties beschouwen, namelijk $\frac{c}{\alpha} \leq K^{uit}$ en $\frac{c}{\alpha} \geq K^{uit}$. We beginnen met de situatie dat $\frac{c}{\alpha} \leq K^{uit}$.

Stelling 5.10. *Beschouw het model met de verdisconteringsfactor $\alpha > 0$. Zij K^{uit} de kosten die betaald moeten worden bij het uitzetten van een server waarbij geldt dat $\frac{c}{\alpha} \leq K^{uit}$. Dan geldt voor alle toestanden (i,j), met $i \in \mathbb{N}_{\geq 0}$ en $j \in \mathbb{N}_{>0}$, dat $\pi^*(i,j) =$ op standby zetten.*

Bewijs. Zij $t \geq 0$ de tijd om vanaf een toestand (i,j), met $i \in \mathbb{N}_{\geq 0}$ en $j \in \mathbb{N}_{>0}$, in toestand $(0, i+j)$ te belanden. Voor alle $t \geq 0$ geldt dat

$$K^{uit} \geq \frac{c}{\alpha} \geq \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right).$$

Omdat dit voor alle $t \geq 0$ en iedere willekeurige begintoestand (i,j) geldt, volgt dat het voor alle toestanden goedkoper is om de server op standby te zetten dan om de server uit te zetten.

Zij S_1 de strategie waarbij, startend op tijdstip 0 in toestand (i,j), tot tijdstip t in elke toestand de server op standby gezet wordt als er een klant vertrekt. Dan geldt voor alle strategieën S, ongelijk aan S_1, dat het goedkoper is om strategie S_1 te volgen dan strategie S, want $K^{uit} \geq \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right)$. Dus voor alle toestanden (i,j), met $i \in \mathbb{N}_{\geq 0}$ en $j \in \mathbb{N}_{>0}$, geldt dat $\pi^*(i,j)$ op standby zetten. ☐

Nu we weten wat de strategie is als geldt dat $\frac{c}{\alpha} \leq K^{uit}$ kunnen we de situatie beschouwen waarbij $\frac{c}{\alpha} \geq K^{uit}$. Voor het bewijzen van stelling 5.11 maken we weer gebruik van definitie 5.6 en definitie 5.7. Ook definitie 5.3 gebruiken we weer.

Dan kunnen we nu de stelling formuleren.

Stelling 5.11. *Beschouw het model met de verdisconteringsfactor $\alpha > 0$. Zij K^{uit} de kosten die betaald moeten worden bij het uitzetten van een server waarbij geldt dat $\frac{c}{\alpha} \geq K^{uit}$. Als voor $j > 1$ geldt dat $\pi^*(i,j) =$ uitzetten, dan geldt dat $\pi^*(i+1, j-1) =$ uitzetten.*

Bewijs. Dit bewijs heeft eenzelfde structuur als het bewijs van stelling 5.4 en stelling 5.8.

Stel $\pi^*(i,j) =$ uitzetten. Dan geldt dat de verwachte verdisconteerde kosten om driehoek $T(i+j)$ te verlaten minstens K^{uit} lager zijn als je naar toestand $(i, j-1)$ gaat, dan wanneer je naar toestand $(i+1, j-1)$ zou gaan. Ofwel,

$$Q^{i+j}(i, j-1) + K^{uit} \leq Q^{i+j}(i+1, j-1). \tag{9}$$

Stel nu dat je in toestand $(i+1, j-1)$ begint. Deze toestand ligt op diagonaal $D(i+j)$. Je wilt weten of je in toestand $(i+1, j-1)$ de server wel of niet uit moet zetten als er een klant vertrekt. Als geldt dat

$$Q^{i+j}(i+1, j-2) + K^{uit} \leq Q^{i+j}(i+2, j-2), \tag{10}$$

dan weten we dat $\pi^*(i+1, j-1) =$ uitzetten.

De verwachte verdisconteerde kosten voor de strategie die de verwachte verdisconteerde kosten om driehoek $T(i+j)$ te verlaten als je in toestand $(i+1, j-2)$ bent minimaliseert, zijn van boven begrensd door de verwachte verdisconteerde kosten van iedere andere strategie die je kiest in toestand $(i+1, j-2)$. Dus als je één strategie voor toestand $(i+1, j-2)$ hebt die minstens K^{uit} goedkoper is dan iedere strategie in toestand $(i+2, j-2)$, dan is het optimaal om wel naar toestand $(i+1, j-2)$ te gaan en niet naar toestand $(i+2, j-2)$ te gaan.

Voor iedere strategie in toestand $(i+2, j-2)$ moet er een moment zijn dat je naar een lagere diagonaal gaat of dat je naar toestand $(i+1, j-1)$ gaat. Noem het tijdstip dat dit gebeurt tijdstip t. Zo'n tijdstip t bestaat, want wegens lemma 4.6 is de totale tijd om driehoek $T(i+j)$ te verlaten onafhankelijk van de gekozen strategie. Verder geldt er dat de verwachte tijd om $T(i+j)$ te verlaten eindig is. Ook geldt er dat om $T(i+j)$ te verlaten je òf via toestand $(i+1, j-1)$ òf via een lagere diagonaal moet gaan. Beschouw dus de strategie

$$\begin{aligned} S_2 \quad = \quad & \{\text{Zet op tijdstip 0 de server op standby in toestand } (i+1, j-1), \\ & \text{voer daarna de optimale strategie uit tot je uit } T(i+j) \text{ bent}\}. \end{aligned}$$

Dan is er voor strategie S_2 zo'n tijdstip t. Er geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_2) = Q^{i+j}(i+2, j-2), \tag{11}$$

want S_2 is de optimale strategie vanaf toestand $(i+2, j-2)$.

Kies dan nu in toestand $(i+1, j-2)$ de volgende strategie: als een klant vertrekt voor tijdstip t, zet dan de server op standby. Dit betekent dat je tot tijdstip t op de diagonaal $D(i+j-1)$ blijft. We beschouwen dus de strategie

$$\begin{aligned} S_1 \quad = \quad & \{\text{Zet op tijdstip 0 de server uit in toestand } (i+1, j-1), \\ & \text{zet daarna tot tijdstip } t \text{ de server op standby als er een klant vertrekt} \\ & \text{en voer daarna de optimale strategie uit tot je uit } T(i+j) \text{ bent}\}. \end{aligned}$$

Er geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \geq Q^{i+j}(i+1, j-2) + K^{uit}, \tag{12}$$

want S_1 is een strategie vanaf toestand $(i+1, j-2)$ en niet noodzakelijk de optimale strategie.

We zullen laten zien dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2)$$

geldt, want dan volgt namelijk dat vergelijking 10 geldt en dat is wat we moeten bewijzen.

We gaan nu strategie S_1 vergelijken met strategie S_2 tot tijdstip t. Dan zijn er twee mogelijkheden:

1) òf beide strategieën bevinden zich op tijdstip t in dezelfde toestand,

2) òf strategie S_1 bevindt zich in toestand $(i, j-1)$ en strategie S_2 in toestand $(i+1, j-1)$.

Analoog aan het bewijs van stelling 5.4 voor de situatie zonder verdisconteerde kosten is te bewijzen dat dit inderdaad de twee mogelijkheden zijn.

Beschouw dan nu beide mogelijkheden waarop je je op tijdstip t kunt bevinden:

1) Beschouw de toestand (i', j'), met $i' + j' = i + j - 1$, zodanig dat op tijdstip t de strategieën S_1 en S_2 beide in toestand (i', j') zijn. Definieer dan $f(S)$ = verdisconteerde stroomkosten van de servers die tussen tijdstip 0 en t op standby staan via strategie S. Omdat er voor strategie S_2 geldt dat er tot tijdstip t één server meer op standby heeft gestaan, volgt met behulp van hoofdstuk 3.2.2 met $i = 1$ dat $f(S_2) = f(S_1) + \frac{c}{\alpha}(1 - e^{-\alpha t})$. Er geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_1) = K^{uit} + f(S_1) + Q^{i+j}(i', j')\, e^{-\alpha t}$$

en er geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_2) = K^{uit} e^{-\alpha t} + f(S_1) + \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) + Q^{i+j}(i', j')\, e^{-\alpha t}.$$

Om te bepalen of geldt dat $d^{i+j}_{(i+1,j-1)}(S_2) \geq d^{i+j}_{(i+1,j-1)}(S_1)$ beschouwen we het verschil $d^{i+j}_{(i+1,j-1)}(S_2) - d^{i+j}_{(i+1,j-1)}(S_1)$. Dan volgt dat

$$\begin{aligned}
d^{i+j}_{(i+1,j-1)}(S_2) - d^{i+j}_{(i+1,j-1)}(S_1) &= K^{uit}e^{-\alpha t} + f\left(S_1\right) + \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) + Q^{i+j}\left(i', j'\right)e^{-\alpha t} \\
&\quad - \left(K^{uit} + f\left(S_1\right) + Q^{i+j}\left(i', j'\right)e^{-\alpha t}\right) \\
&= K^{uit}e^{-\alpha t} + f\left(S_1\right) + \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) + Q^{i+j}\left(i', j'\right)e^{-\alpha t} \\
&\quad -K^{uit} - f\left(S_1\right) - Q^{i+j}\left(i', j'\right)e^{-\alpha t} \\
&= K^{uit}e^{-\alpha t} + \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) - K^{uit} \\
&= \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) - K^{uit}\left(1 - e^{-\alpha t}\right) \\
&= \left(\frac{c}{\alpha} - K^{uit}\right)\left(1 - e^{-\alpha t}\right) \\
&\overset{(*)}{\geq} 0.
\end{aligned}$$

(*) Want $\frac{c}{\alpha} \geq K^{uit}$.

Er geldt dus dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2).$$

2) Als op tijdstip t geldt dat strategie S_1 in toestand $(i, j-1)$ beland is, dan is strategie S_2 in toestand $(i+1, j-1)$ beland. Definieer dan weer $f(S)$ = verdisconteerde stroomkosten van de servers die tussen tijdstip 0 en t op standby staan via strategie S. Omdat er voor strategie S_2 geldt dat er tot tijdstip t één server meer op standby heeft gestaan, volgt met behulp van hoofdstuk 3.2.2 met $i = 1$ dat $f(S_2) = f(S_1) + \frac{c}{\alpha}(1 - e^{-\alpha t})$. Er geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_1) = K^{uit} + f\left(S_1\right) + Q^{i+j}\left(i, j-1\right)e^{-\alpha t}$$

en er geldt dat

$$d^{i+j}_{(i+1,j-1)}(S_2) = f\left(S_1\right) + \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) + Q^{i+j}\left(i+1, j-1\right)e^{-\alpha t}.$$

Om te bepalen of geldt dat $d^{i+j}_{(i+1,j-1)}(S_2) \geq d^{i+j}_{(i+1,j-1)}(S_1)$ beschouwen we het verschil $d^{i+j}_{(i+1,j-1)}(S_2) - d^{i+j}_{(i+1,j-1)}(S_1)$. Dan volgt dat

$$\begin{aligned}
d^{i+j}_{(i+1,j-1)}(S_2) - d^{i+j}_{(i+1,j-1)}(S_1) &= f\left(S_1\right) + \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) + Q^{i+j}\left(i+1, j-1\right)e^{-\alpha t} \\
&\quad - \left(K^{uit} + f\left(S_1\right) + Q^{i+j}\left(i, j-1\right)e^{-\alpha t}\right) \\
&= \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) + Q^{i+j}\left(i+1, j-1\right)e^{-\alpha t} \\
&\quad -K^{uit} - Q^{i+j}\left(i, j-1\right)e^{-\alpha t} \\
&= \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) + Q^{i+j}\left(i+1, j-1\right)e^{-\alpha t} \\
&\quad -K^{uit} + K^{uit}e^{-\alpha t} - K^{uit}e^{-\alpha t} - Q^{i+j}\left(i, j-1\right)e^{-\alpha t} \\
&= \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) + Q^{i+j}\left(i+1, j-1\right)e^{-\alpha t} \\
&\quad -K^{uit}\left(1 - e^{-\alpha t}\right) - \left(K^{uit} + Q^{i+j}\left(i, j-1\right)\right)e^{-\alpha t}.
\end{aligned}$$

Dan volgt met behulp van $\frac{c}{\alpha} \geq K^{uit}$ en vergelijking 9 dat

$$d^{i+j}_{(i+1,j-1)}(S_2) - d^{i+j}_{(i+1,j-1)}(S_1) \quad \geq \quad \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right) - K^{uit}\left(1 - e^{-\alpha t}\right) \geq 0.$$

Er geldt dus dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2).$$

In beide situaties volgt dus, omdat $\frac{c}{\alpha} \geq K^{uit}$, dat

$$d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2).$$

Hieruit volgt dan met behulp van vergelijking 11 en vergelijking 12 dat

$$Q^{i+j}(i+1, j-2) + K^{uit} \leq d^{i+j}_{(i+1,j-1)}(S_1) \leq d^{i+j}_{(i+1,j-1)}(S_2) = Q^{i+j}(i+2, j-2),$$

waaruit volgt dat $\pi^*(i+1, j-1) =$ uitzetten voor $\frac{c}{\alpha} \geq K^{uit}$. □

Door stelling 5.11 herhaaldelijk toe te passen krijgen we het volgende.

Gevolg 5.12. *Beschouw het model met de verdisconteringsfactor $\alpha > 0$. Zij K^{uit} de kosten om een server uit te zetten. Zij $\frac{c}{\alpha} \geq K^{uit} \geq 0$. Als er een toestand $(i, j) \in D(k)$ is waarvoor geldt dat $\pi^*(i, j) =$ uitzetten, dan volgt voor alle toestanden op $D(k)$ lager dan toestand (i, j) ook dat het optimaal is om de server uit te zetten als er een klant vertrekt.*

Door stelling 5.8 en stelling 5.11 tegelijk toe te passen volgt het volgende.

Gevolg 5.13. *Beschouw het model met de verdisconteringsfactor $\alpha > 0$. Zij K^{uit} de kosten om een server uit te zetten en K^{aan} de kosten om een server aan te zetten. Zij $\frac{c}{\alpha} \geq K^{uit} \geq 0$ en $K^{aan} \geq 0$. Als voor $j > 0$ geldt dat $\pi^*(i, j) =$ uitzetten, dan geldt dat $\pi^*(i+1, j-1) =$ uitzetten.*

De volgende stelling volgt uit gevolg 5.13.

Stelling 5.14. *Beschouw het model met de verdisconteringsfactor $\alpha > 0$. Zij K^{uit} de kosten om een server uit te zetten en K^{aan} de kosten om een server aan te zetten. Zij $\frac{c}{\alpha} \geq K^{uit} \geq 0$. Als er een toestand $(i, j) \in D(k)$ is waarvoor geldt dat $\pi^*(i, j) =$ uitzetten, dan volgt voor alle toestanden op $D(k)$ lager dan toestand (i, j) ook dat het optimaal is om de server uit te zetten als er een klant vertrekt.*

Bewijs. Dit volgt door gevolg 5.13 herhaaldelijk toe te passen. □

Een ander belangrijk gevolg is de volgende stelling.

Stelling 5.15. *Beschouw het model met de verdisconteringsfactor $\alpha > 0$. Zij K^{aan} de kosten die betaald moeten worden bij het aanzetten van een server en zij K^{uit} de kosten die betaald moeten worden bij het uitzetten van een server. Zij $\frac{c}{\alpha} \leq K^{uit}$. Dan geldt voor alle toestanden (i, j) dat $\pi^*(i, j) =$ op standby zetten.*

Bewijs. Zij $t \geq 0$ de tijd om vanaf een toestand (i, j), met $i \in \mathbb{N}_{\geq 0}$ en $j \in \mathbb{N}_{>0}$, in toestand $(0, i+j)$ te belanden. Voor alle $t \geq 0$ geldt dat

$$\begin{aligned} K^{aan}e^{-\alpha t} + K^{uit} &\geq K^{aan}e^{-\alpha t} + \frac{c}{\alpha} \\ &\geq \frac{c}{\alpha} \\ &\geq \frac{c}{\alpha}\left(1 - e^{-\alpha t}\right). \end{aligned}$$

Omdat dit voor alle $t \geq 0$ geldt en het niet uitmaakt in welke toestand je bent, volgt dat het altijd optimaal is om de server op standby te zetten als er een klant vertrekt. Dus er geldt voor alle toestanden (i,j) in het systeem dat $\pi^*(i,j)$ op standby zetten. $\square$

Nu we weten wat er op de diagonalen gebeurt, gaan we kijken of er een bovengrens te vinden is op het gebied waarbinnen een server op standby zetten optimaal kan zijn. Als we dit vinden, dan volgt namelijk dat we een begrensd gebied hebben waarbinnen een server op standby zetten optimaal kan zijn.

6 De strategie in de toestanden $(0,j)$, met $j \in \mathbb{N}_{\geq 0}$

Door in de toestanden $(0,j)$, met $j \in \mathbb{N}_{\geq 0}$, de optimale strategie te bepalen kunnen we deze strategie daarna gebruiken om in andere toestanden de optimale strategie bepalen. Voordat we echter de optimale strategie in de toestanden $(0,j)$, met $j \in \mathbb{N}_{\geq 0}$, kunnen geven moeten we eerst nog twee lemma's geven waaruit de optimale strategie in toestand $(0,1)$ volgt.

Lemma 6.1. *Zij X een stochastische variabele die exponentieel verdeeld is met parameter $\lambda+n\mu$, waarbij $n \in \mathbb{N}_{\geq 0}$. Er geldt dat $\mathbb{E}\left[e^{-\alpha X}\right] = \frac{\lambda+n\mu}{\alpha+\lambda+n\mu}$.*

Bewijs. Zij X een stochastische variabele die exponentieel verdeeld is met parameter $\lambda + n\mu$, waarbij $n \in \mathbb{N}_{\geq 0}$. Dan geldt er dat

$$\mathbb{E}[X] = \frac{1}{\lambda + n\mu},$$

dus er geldt dat

$$e^{\mathbb{E}[X]} = e^{\frac{1}{\lambda+n\mu}}.$$

Zij $f(x)$ de exponentiële verdelingsfunctie van X. Dan geldt dat

$$f(x) = \begin{cases} (\lambda + n\mu)e^{-(\lambda+n\mu)x} & \text{, als } x \geq 0 \\ 0 & \text{, als } x < 0 \end{cases}$$

Zij nu $Y = g(X)$, met $g(x) = e^{-\alpha x}$ voor alle $x \in X$. Dan volgt dat

$$\begin{aligned} \mathbb{E}[e^{-\alpha X}] = \mathbb{E}[Y] &= \int_0^\infty g(x)f(x)dx \\ &= \int_0^\infty e^{-\alpha x} \cdot (\lambda + n\mu)e^{-(\lambda+n\mu)x}dx \\ &= (\lambda + n\mu)\int_0^\infty e^{-(\alpha+\lambda+n\mu)x}dx \\ &= (\lambda + n\mu)\left[-\frac{1}{\alpha+\lambda+n\mu}e^{-(\alpha+\lambda+n\mu)x}\right]_{x=0}^\infty \\ &= \begin{cases} \infty & \text{, als } \alpha+\lambda+n\mu \leq 0 \\ \frac{\lambda+n\mu}{\alpha+\lambda+n\mu} & \text{, als } \alpha+\lambda+n\mu > 0 \end{cases}. \end{aligned}$$

Omdat in ons geval geldt dat $\alpha, \lambda, \mu > 0$ en $n \in \mathbb{N}_{\geq 0}$, volgt dat $\mathbb{E}[e^{-\alpha X}] = \frac{\lambda+n\mu}{\alpha+\lambda+n\mu}$. $\square$

Lemma 6.2. *Beschouw het model met verdisconteerde kosten waarbij K^{aan} betaald moet worden bij het aanzetten van een server en K^{uit} bij het uitzetten van een server. Het is optimaal om de server in toestand $(0,1)$ op standby te zetten als geldt dat $c \leq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$. En het is optimaal om de server in toestand $(0,1)$ uit te zetten als geldt dat $c \geq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$.*

Bewijs. Stel je bent in toestand $(0,1)$. Dan ben je in driehoek $T(1)$. Als er een klant vertrekt kun je kiezen om de server uit te zetten, dan ga je naar toestand $(0,0)$, of om de server op standby te zetten, dan ga je naar toestand $(1,0)$. We beschouwen daarom de volgende twee strategieën S_1 en S_2:

$$\begin{aligned} S_1 &= \{\text{Zet op tijdstip 0 in toestand } (0,1) \text{ de server op standby} \\ &\quad\ \text{en volg daarna de optimale strategie}\} \end{aligned}$$

en

$$\begin{aligned} S_2 &= \{\text{Zet op tijdstip 0 in toestand } (0,1) \text{ de server uit} \\ &\quad\ \text{en volg daarna de optimale strategie}\} \end{aligned}$$

Met kans $\frac{\lambda}{\lambda+\mu}$ komt er in toestand $(0,1)$ een klant aan voordat er een klant vertrekt, en met kans $\frac{\mu}{\lambda+\mu}$ vertrekt er in toestand $(0,1)$ een klant voordat er een klant arriveert. Als er eerst een klant arriveert voordat er een klant vertrekt maakt het niet uit of je strategie S_1 of S_2 gekozen had, want dan zijn S_1 en S_2 even duur totdat je voor het eerst een toestand op diagonaal $D(2)$ bereikt.

Beschouw daarom de situatie waarbij er eerst een klant vertrekt. Als er een klant vertrekt ga je òf naar toestand $(0,0)$ òf naar toestand $(1,0)$. Voor beide toestanden geldt dat de tijd die het systeem in deze toestanden verblijft exponentieel verdeeld is met parameter λ. Zij τ de verblijftijd in $(0,0)$ en $(1,0)$. Dan geldt dat τ exponentieel verdeeld is met parameter λ. Er geldt dat de verwachte verdisconteerde kosten tot je weer in toestand $(0,1)$ terug bent (je komt namelijk met kans 1 weer terug in $(0,1)$) voor strategie S_1 gegeven door

$$\begin{aligned} \mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)\right] &= \frac{c}{\alpha}\mathbb{E}\left[1-e^{-\alpha\tau}\right] \\ &= \frac{c}{\alpha}\left(1-\frac{\lambda}{\alpha+\lambda}\right) \\ &= \frac{c}{\alpha}\cdot\frac{\alpha}{\alpha+\lambda} \\ &= \frac{c}{\alpha+\lambda} \end{aligned}$$

en voor strategie S_2 gegeven door

$$\begin{aligned} \mathbb{E}\left[K^{aan}e^{-\alpha\tau}+K^{uit}\right] &= K^{aan}\mathbb{E}\left[e^{-\alpha\tau}\right]+K^{uit} \\ &= \frac{K^{aan}\lambda}{\alpha+\lambda}+K^{uit} \\ &= \frac{K^{aan}\lambda+K^{uit}(\alpha+\lambda)}{\alpha+\lambda}. \end{aligned}$$

Hieruit volgt dan dat het voor $c \leq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$ optimaal is om in toestand $(0,1)$ de server op standby te zetten en dat het voor $c > K^{aan}\lambda + K^{uit}(\alpha+\lambda)$ optimaal is om in toestand $(0,1)$ de server uit te zetten. □

Nu bekend is wat voor de verschillende waarden van de parameters de optimale strategie in toestand $(0,1)$ is, kunnen we de optimale strategie in de toestanden $(0,j)$ gaan bepalen.

6.1 Op standby zetten in toestand $(0,1)$ is optimaal

Beschouw nu eerst de situatie dat het in toestand $(0,1)$ optimaal is om de server op standby te zetten. We gaan nu bepalen wat dan de strategie in de toestanden $(0,j)$ is.

Stelling 6.3. *Beschouw het model met verdisconteerde kosten waarbij K^{aan} betaald moet worden bij het aanzetten van een server en K^{uit} bij het uitzetten van een server. Als geldt dat $c \leq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$, dan geldt dat het optimaal is om in de toestanden $(0,j)$, met $j \in \mathbb{N}_{\geq 0}$, de servers op standby te zetten als er een klant vertrekt.*

Bewijs. Zij $c \leq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$. Met behulp van volledige inductie naar j zal bewezen worden dat het in alle toestanden $(0,j)$, met $j \in \mathbb{N}_{\geq 0}$, optimaal is om de server op standby te zetten als er een klant vertrekt.

Met behulp van lemma 6.2 volgt dat het optimaal is om in toestand $(0,1)$ de server op standby te zetten. Neem nu aan dat je in de optimale strategie minstens tot en met toestand $(0,j)$ de servers op standby zet als er een klant vertrekt. Beschouw dan toestand $(0,j+1)$.

In toestand $(0,j+1)$ geldt dat de kans op een aankomst $\frac{\lambda}{\lambda+(j+1)\mu}$ is en de kans op een vertrek $\frac{(j+1)\mu}{\lambda+(j+1)\mu}$ is. We beschouwen weer de situatie dat er eerst een klant vertrekt, want als er eerst een klant arriveert zijn alle strategieën weer even duur. Dan beschouwen we de strategieën

$$\begin{aligned} S_1 &= \{\text{Zet op tijdstip 0 in toestand } (0,j+1) \text{ de server op standby,} \\ &\quad\ \text{zet in toestand } (1,j) \text{ de server uit als er een klant vertrekt,} \\ &\quad\ \text{daarna optimaal tot uit driehoek } T(j+1)\} \end{aligned}$$

en

$$\begin{aligned} S_2 &= \{\text{ Zet op tijdstip 0 in toestand } (0,j+1) \text{ de server uit,} \\ &\quad\ \text{daarna optimaal tot uit driehoek } T(j+1)\} \\ &= \{\text{ Zet op tijdstip 0 in toestand } (0,j+1) \text{ de server uit,} \\ &\quad\ \text{zet in toestand } (0,j) \text{ de server op standby als er een klant vertrekt,} \\ &\quad\ \text{daarna optimaal tot uit driehoek } T(j+1)\}. \end{aligned}$$

Dan gaan we vervolgens deze twee strategieën vergelijken. Er zijn twee situaties die we moeten beschouwen:

(1) Eerst vertrekt er een klant en vervolgens arriveert er een klant,

(2) Er vertrekken twee klanten na elkaar.

Vanwege de geheugenloosheid van het systeem is het voldoende om de verdisconteerde kosten in toestand $(0,j+1)$, met $j \in \mathbb{N}_{\geq 1}$, te beschouwen tot je daar terug komt of totdat twee strategieën in dezelfde toestand belanden.

Zij τ de verblijfstijd in de toestanden $(0,j)$ en $(1,j)$. Dan geldt dat τ exponentieel verdeeld is met parameter $\lambda + j\mu$.

Beschouw nu eerst de verwachte verdisconteerde kosten van de strategieën S_1 en S_2 in situatie (1). Voor strategie S_1 geldt dat de verwachte verdisconteerde kosten tot terugkeer in toestand $(0, j+1)$ gegeven worden door:

$$\begin{aligned}\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)\right] &= \frac{c}{\alpha}\left(1-\mathbb{E}\left[e^{-\alpha\tau}\right]\right)\\ &= \frac{c}{\alpha}\left(1-\frac{\lambda+j\mu}{\alpha+\lambda+j\mu}\right)\\ &= \frac{c}{\alpha}\cdot\frac{\alpha}{\alpha+\lambda+j\mu}\\ &= \frac{c}{\alpha+\lambda+j\mu}.\end{aligned}$$

Voor strategie S_2 geldt dat de verwachte verdisconteerde kosten tot terugkeer in toestand $(0, j+1)$ gegeven worden door:

$$\begin{aligned}\mathbb{E}\left[K^{aan}e^{-\alpha\tau}+K^{uit}\right] &= K^{aan}\mathbb{E}\left[e^{-\alpha\tau}\right]+K^{uit}\\ &= \frac{K^{aan}\left(\lambda+j\mu\right)}{\alpha+\lambda+j\mu}+K^{uit}\\ &= \frac{K^{aan}\left(\lambda+j\mu\right)+K^{uit}\left(\alpha+\lambda+j\mu\right)}{\alpha+\lambda+j\mu}.\end{aligned}$$

Beschouw vervolgens de verwachte verdisconteerde kosten van de strategieën S_1 en S_2 in situatie (2). Voor strategie S_1 geldt dat de verwachte verdisconteerde kosten totdat je in toestand $(1, j-1)$ beland gegeven worden door:

$$\begin{aligned}\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)+K^{uit}e^{-\alpha\tau}\right] &= \mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)\right]+K^{uit}\cdot\mathbb{E}\left[e^{-\alpha\tau}\right]\\ &= \frac{c}{\alpha+\lambda+j\mu}+\frac{K^{uit}\left(\lambda+j\mu\right)}{\alpha+\lambda+j\mu}.\end{aligned}$$

Voor de strategie S_2 geldt dat de verwachte verdisconteerde kosten totdat je in toestand $(1, j-1)$ beland gegeven worden door:

$$\mathbb{E}\left[K^{uit}\right] = K^{uit}.$$

Omdat de kansen op een aankomst of vertrek in de toestanden $(0, j)$ en $(1, j)$ bekend zijn volgt dat de verwachte verdisconteerde kosten van beide strategieën nu berekend kan worden. Voor strategie S_1 geldt dat de verwachte verdisconteerde kosten tot tijdstip τ gegeven worden door:

$$\begin{aligned}&\frac{\lambda}{\lambda+j\mu}\cdot\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)\right]\\ &+\frac{j\mu}{\lambda+j\mu}\cdot\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)+K^{uit}e^{-\alpha\tau}\right] = \mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)\right]+\frac{j\mu}{\lambda+j\mu}\cdot K^{uit}\cdot\mathbb{E}\left[e^{-\alpha\tau}\right]\\ &\qquad = \frac{c}{\alpha+\lambda+j\mu}+K^{uit}\cdot\frac{j\mu}{\lambda+j\mu}\cdot\frac{\lambda+j\mu}{\alpha+\lambda+j\mu}\\ &\qquad = \frac{c}{\alpha+\lambda+j\mu}+\frac{K^{uit}j\mu}{\alpha+\lambda+j\mu}.\end{aligned}$$

Voor strategie S_2 geldt dat de verwachte verdisconteerde kosten tot tijdstip τ gegeven worden door:

$$\begin{aligned}\frac{\lambda}{\lambda+j\mu}\cdot\mathbb{E}\left[K^{aan}e^{-\alpha\tau}+K^{uit}\right]+\frac{j\mu}{\lambda+j\mu}\cdot\mathbb{E}\left[K^{uit}\right] &= \frac{K^{aan}\lambda}{\lambda+j\mu}\cdot\mathbb{E}\left[e^{-\alpha\tau}\right]+\mathbb{E}\left[K^{uit}\right] \\ &= \frac{K^{aan}\lambda}{\lambda+j\mu}\frac{\lambda+j\mu}{\alpha+\lambda+j\mu}+K^{uit} \\ &= \frac{K^{aan}\lambda}{\alpha+\lambda+j\mu}+K^{uit}.\end{aligned}$$

We hebben aangenomen dat $c \leq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$. Hieruit volgt dat het in toestand $(0, j+1)$ altijd optimaal is om strategie S_1 te volgen, want er geldt dat

$$\begin{aligned}\frac{c}{\alpha+\lambda+j\mu}+\frac{K^{uit}j\mu}{\alpha+\lambda+j\mu}-\left(\frac{K^{aan}\lambda}{\alpha+\lambda+j\mu}+K^{uit}\right) &= \frac{c}{\alpha+\lambda+j\mu}+\frac{K^{uit}j\mu}{\alpha+\lambda+j\mu} \\ &\quad -\frac{K^{aan}\lambda}{\alpha+\lambda+j\mu}-\frac{K^{uit}(\alpha+\lambda+j\mu)}{\alpha+\lambda+j\mu} \\ &= \frac{c-K^{aan}\lambda-K^{uit}(\alpha+\lambda)}{\alpha+\lambda+j\mu} \\ &\overset{(*)}{\leq} 0.\end{aligned}$$

(*) Want er geldt dat $c \leq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$.

Dus als geldt dat $c \leq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$, dan is het in toestand $(0, j+1)$ optimaal om de server op standby te zetten. □

6.2 Uitzetten in toestand $(0, 1)$ is optimaal

Voordat we de stelling geven die zegt wat de optimale strategie in de toestanden $(0, j)$ is zullen we eerst nog een lemma geven wat we vervolgens in het bewijs van de stelling gaan gebruiken.

Lemma 6.4. *Laat $n \in \mathbb{N}_{>0}$. Zij $X(n)$ een stochastische variabele. Zij $Y(n)$ een stochastische variabele die de Erlang verdeling heeft met parameters n, het aantal aankomsten, en λ, de aankomstsnelheid. Als geldt dat $X(n) \overset{d}{\geq} Y(n)$, dan geldt voor alle n dat*

$$\mathbb{E}\left[e^{-\alpha X(n)}\right] \leq \mathbb{E}\left[e^{-\alpha Y(n)}\right].$$

Bewijs. Omdat geldt dat $X(n) \overset{d}{\geq} Y(n)$ volgt dat $-\alpha X(n) \overset{d}{\leq} -\alpha Y(n)$. Dan volgt hieruit dat $e^{-\alpha X(n)} \overset{d}{\leq} e^{-\alpha Y(n)}$. En dan volgt hieruit dat geldt dat $\mathbb{E}\left[e^{-\alpha X(n)}\right] \leq \mathbb{E}\left[e^{-\alpha Y(n)}\right]$. □

Beschouw dan nu de situatie dat het in toestand $(0, 1)$ optimaal is om de server uit te zetten.

Stelling 6.5. *Beschouw het model met verdisconteerde kosten waarbij K^{aan} betaald moet worden bij het aanzetten van een server en K^{uit} bij het uitzetten van een server. Als geldt dat $c \geq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$, dan geldt dat het optimaal is om in alle toestanden $(0, j)$, met $j \in \mathbb{N}_{>0}$, de servers uit te zetten als er een klant vertrekt.*

Bewijs. Zij $c \geq K^{aan}\lambda + K^{uit}(\alpha+\lambda)$. Met behulp van volledige inductie naar j zal bewezen worden dat het in alle toestanden $(0, j)$, met $j \in \mathbb{N}_{\geq 0}$, optimaal is om de server op standby te zetten als er een klant vertrekt.

Dan volgt met behulp van lemma 6.2 dat het optimaal is om in toestand $(0,1)$ de server uit te zetten. Neem nu aan dat je in de optimale strategie minstens tot en met toestand $(0, j)$ de servers uitzet als er een klant vertrekt.

Stel nu dat je in toestand $(0, j+1)$ bent en er vertrekt een klant. We gaan de strategie waarbij je de server uitzet in toestand $(0, j+1)$ vergelijken met de strategie waarbij je de server op standby zet in deze toestand. Als uitzetten goedkoper is, dan is het genoeg om een strategie te geven die goedkoper is dan iedere strategie waar je in toestand $(0, j+1)$ de server op standby zet.

De strategieën die we beschouwen zijn:

$$\begin{aligned} S_2 \;&=\; \{\text{Zet op tijdstip 0 in toestand } (0, j+1) \text{ de server op standby,} \\ &\qquad \text{daarna optimaal tot uit driehoek } T(j+1)\} \end{aligned}$$

en

$$\begin{aligned} S_1 \;&=\; \{\text{Zet op tijdstip 0 in toestand } (0, j+1) \text{ de server uit,} \\ &\qquad \text{imiteer daarna strategie } S_2 \text{ tot uit driehoek } T(j+1)\}. \end{aligned}$$

Als strategie S_2 in toestand $(1, j)$ is en strategie S_1 in toestand $(0, j)$ is, met $j \in \mathbb{N}_{\geq 0}$, en er arriveert een klant, dan belanden beiden strategieën in dezelfde toestand, namelijk toestand $(0, j+1)$. Noem dit moment tijdstip τ. Er geldt dat

$$\tau \overset{d}{\geq} X,$$

met X een stochastische variabele die exponentieel verdeeld is met parameter λ, ofwel X is Erlang verdeeld met parameters $n = 1$ en λ.

Voor strategie S_2 geldt dan dat je $K^{uit} + K^{aan}e^{-\alpha\tau}$ extra moet betalen die je bij S_1 niet hoeft te betalen. Maar bij S_1 heb je altijd één server meer op standby hebt staan dan bij S_2 en deze stroomkosten zijn $\frac{c}{\alpha}(1 - e^{-\alpha\tau})$. Als geldt dat $K^{uit} + K^{aan}e^{-\alpha\tau} \leq \frac{c}{\alpha}(1 - e^{-\alpha\tau})$, dan geldt dat het goedkoper is om strategie S_1 te volgen dan S_2. Er geldt dat

$$\begin{aligned} \mathbb{E}\left[\frac{c}{\alpha}\left(1 - e^{-\alpha\tau}\right) - \left(K^{uit} + K^{aan}e^{-\alpha\tau}\right)\right] &= \mathbb{E}\left[\frac{c}{\alpha}\left(1 - e^{-\alpha\tau}\right) - K^{uit} - K^{aan}e^{-\alpha\tau}\right] \\ &= \frac{c}{\alpha}\left(1 - \mathbb{E}\left[e^{-\alpha\tau}\right]\right) - K^{uit} - K^{aan}\mathbb{E}\left[e^{-\alpha\tau}\right] \\ &\overset{(*)}{\geq} \frac{c}{\alpha}\left(1 - \mathbb{E}\left[e^{-\alpha X}\right]\right) - K^{uit} - K^{aan}\mathbb{E}\left[e^{-\alpha X}\right] \end{aligned}$$

(*) Wegens lemma 6.4 met $n = 1$.

Hieruit volgt dat

$$\begin{aligned}
\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)-\left(K^{uit}+K^{aan}e^{-\alpha\tau}\right)\right] &= \frac{c}{\alpha}\left(1-\frac{\lambda}{\alpha+\lambda}\right)-K^{uit}-K^{aan}\frac{\lambda}{\alpha+\lambda} \\
&\overset{(**)}{\geq} \frac{K^{aan}\lambda+K^{uit}\left(\alpha+\lambda\right)}{\alpha}\left(1-\frac{\lambda}{\alpha+\lambda}\right) \\
&\quad -K^{uit}-K^{aan}\frac{\lambda}{\alpha+\lambda} \\
&= \frac{K^{aan}\lambda+K^{uit}\left(\alpha+\lambda\right)}{\alpha+\lambda}-K^{uit}-K^{aan}\frac{\lambda}{\alpha+\lambda} \\
&= 0.
\end{aligned}$$

(**) Want er geldt dat $c \geq K^{aan}\lambda + K^{uit}\left(\alpha+\lambda\right)$.

Omdat geldt dat $\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha\tau}\right)-\left(K^{uit}+K^{aan}e^{-\alpha\tau}\right)\right] \geq 0$, volgt dat het goedkoper is om de server uit te zetten in toestand $(0, j+1)$. Hieruit volgt dat als geldt dat $c \geq K^{aan}\lambda + K^{uit}\left(\alpha+\lambda\right)$, dat het optimaal is in alle toestanden $(0, j)$ om de servers uit te zetten als er een klant vertrekt. □

Door stelling 5.14 op lemma 6.5 toe te passen volgt voor $c \geq K^{aan}\lambda + K^{uit}\left(\alpha+\lambda\right)$ dat het optimaal is om in het hele systeem de servers uit te zetten als er een klant vertrekt.

Gevolg 6.6. *Beschouw het model met verdisconteerde kosten waarbij K^{aan} betaald moet worden bij het aanzetten van een server en K^{uit} bij het uitzetten van een server. Als geldt dat $c \geq K^{aan}\lambda + K^{uit}\left(\alpha+\lambda\right)$, dan geldt dat voor alle toestanden (i, j), met $i \in \mathbb{N}_{\geq 0}$ en $j \in \mathbb{N}_{>0}$, dat $\pi^*(i, j) =$ uitzetten.*

Bewijs. Omdat geldt dat $c \geq K^{aan}\lambda + K^{uit}\left(\alpha+\lambda\right)$ volgt met behulp van lemma 6.5 dat het voor alle toestanden $(0, k)$, met $k \in \mathbb{N}_{>0}$, optimaal is om de server uit te zetten als er een klant vertrekt. Dan volgt met behulp van stelling 5.14 dat voor alle toestanden (i, j) op diagonaal $D(k)$ geldt dat $\pi^*(i, j) =$ uitzetten. Dus voor alle toestanden in het systeem is het optimaal om de server uit te zetten als er een klant vertrekt. □

7 De bovengrens op het uitzetten van servers

Voor $c \geq K^{aan}\lambda + K^{uit}\left(\alpha+\lambda\right)$ geldt dat het in het hele systeem optimaal is om de servers uit te zetten als er een klant vertrekt. Neem daarom in dit hoofdstuk aan dat $c \leq K^{aan}\lambda + K^{uit}\left(\alpha+\lambda\right)$.

In dit hoofdstuk gaan we op zoek naar een diagonaal $D(j+1)$ waar het in toestand $(1, j)$ optimaal is om de server uit te zetten als er een klant vertrekt. Als we deze diagonaal gevonden hebben, hebben we een bovengrens gevonden op het gebied waarbinnen op standby zetten optimaal kan zijn. Buiten dit gebied zet je dan dus alle servers uit als er een klant vertrekt.

7.1 De strategie in toestand $(i, 1)$, met $i \in \mathbb{N}_{\geq 0}$

Voordat we de strategie in de toestanden $(1, j)$, met $j \in \mathbb{N}_{\geq 0}$, kunnen geven moeten we eerst te weten komen wat de strategie in de toestanden $(i, 1)$ is, met $i \in \mathbb{N}_{\geq 0}$. Voordat we de stelling kunnen geven die zegt wat de optimale strategie voor de toestanden $(1, j)$ is, met $j \in \mathbb{N}_{\geq 0}$, hebben we eerst nog een definitie en een aantal lemma's nodig.

Definitie 7.1. Zij $X(n)$, met $n \in \mathbb{N}_{\geq 0}$, de tijd om vanaf een toestand $(i, 0)$, met $i \in \mathbb{N}_{\geq 0}$ waarbij $i \leq n$, in toestand $(0, n)$ te belanden.

Lemma 7.2. *Zij $Y(n)$ een stochastische variabele die de Erlang verdeling heeft met parameters $n \in \mathbb{N}_{>0}$, het aantal aankomsten, en λ, de aankomstsnelheid. Zij $X(n)$ als in definitie 7.1. Dan geldt voor alle $n \in \mathbb{N}_{>0}$ dat*

(1) $X(n) \overset{d}{\geq} Y(n)$,

(2) $\mathbb{E}\left[e^{-\alpha X(n)}\right] \leq \mathbb{E}\left[e^{-\alpha Y(n)}\right]$.

Bewijs.

(1) Voor de stochastische variabele $Y(n)$ geldt dat het de tijd aangeeft tot er precies n aankomsten zijn. Voor $X(n)$ geldt dat er minstens n aankomsten nodig zijn tot je in toestand $(0, n)$ belandt. Deze aankomsten zijn allemaal via een exponentiële verdeling met parameter λ. Hieruit volgt dat $X(n) \overset{d}{\geq} Y(n)$.

(2) Uit lemma 7.2.1 volgt dat $X(n) \overset{d}{\geq} Y(n)$. Dan volgt met behulp van lemma 6.4 dat $\mathbb{E}\left[e^{-\alpha X(n)}\right] \leq \mathbb{E}\left[e^{-\alpha Y(n)}\right]$.

□

Het volgende lemma hebben we nodig om in het lemma erna in het bewijs een stelling te kunnen toepassen.

Lemma 7.3. *Er geldt dat*

(1) $\mathbb{E}\left[e^{-\alpha Y(n)}\right] = \frac{\lambda^n}{(\alpha+\lambda)^n}$ *voor alle* $n \in \mathbb{N}_{>0}$,

(2) $\lim_{n\to\infty} \mathbb{E}\left[e^{-\alpha Y(n)}\right] = 0$.

Bewijs. (1) Voor alle $n \in \mathbb{N}_{>0}$ geldt dat

$$\mathbb{E}\left[e^{-\alpha Y(n)}\right] = \int_{-\infty}^{\infty} e^{-\alpha y} f(y) dy,$$

waarbij $f(y)$ gegeven wordt door

$$f(y) = \begin{cases} \frac{(\lambda y)^{n-1}}{(n-1)!}\lambda e^{-\lambda y} & , \text{als } y \geq 0 \\ 0 & , \text{als } y < 0 \end{cases}$$

want $Y(n)$ is Erlang verdeeld met parameters n en λ.
Dan volgt er dat

$$\begin{aligned} \mathbb{E}\left[e^{-\alpha Y(n)}\right] &= \int_{-\infty}^{0} e^{-\alpha y} \cdot 0 \; dy + \int_{0}^{\infty} e^{-\alpha y} \cdot \frac{(\lambda y)^{n-1}}{(n-1)!}\lambda e^{-\lambda y} \; dy \\ &= 0 + \frac{\lambda^n}{(n-1)!}\int_0^\infty y^{n-1} e^{-(\alpha+\lambda)y} \; dy \\ &= \frac{\lambda^n}{(n-1)!}\int_0^\infty y^{n-1} e^{-(\alpha+\lambda)y} \; dy \end{aligned}$$

Voor $n = 1$ volgt dan dat

$$\begin{aligned}
\mathbb{E}\left[e^{-\alpha Y(n)}\right] &= \mathbb{E}\left[e^{-\alpha Y(1)}\right] \\
&= \frac{\lambda^1}{(1-1)!}\int_0^\infty y^{1-1}e^{-(\alpha+\lambda)y}\ dy \\
&= \lambda\int_0^\infty e^{-(\alpha+\lambda)y}\ dy \\
&= \frac{\lambda}{\alpha+\lambda}.
\end{aligned}$$

Voor $n = 2$ volgt met behulp van partiële integratie dat

$$\begin{aligned}
\mathbb{E}\left[e^{-\alpha Y(n)}\right] &= \mathbb{E}\left[e^{-\alpha Y(2)}\right] \\
&= \frac{\lambda^2}{(2-1)!}\left(\left[-\frac{1}{\alpha+\lambda}y^{2-1}e^{-(\alpha+\lambda)y}\right]_{y=0}^\infty - \int_0^\infty (2-1)y^{(2-1)-1}\cdot -\frac{1}{\alpha+\lambda}e^{-(\alpha+\lambda)y}\ dy\right) \\
&= \lambda^2\left((0+0) + \frac{1}{\alpha+\lambda}\int_0^\infty e^{-(\alpha+\lambda)y}\ dy\right) \\
&= \frac{\lambda^2}{\alpha+\lambda}\int_0^\infty e^{-(\alpha+\lambda)y}\ dy \\
&= \frac{\lambda^2}{(\alpha+\lambda)^2}
\end{aligned}$$

Voor $n \in \mathbb{N}_{\geq 3}$ volgt door herhaaldelijk partiële integratie toe te passen dat

$$\begin{aligned}
\mathbb{E}\left[e^{-\alpha Y(n)}\right] &= \frac{\lambda^n}{(n-1)!}\left(\left[-\frac{1}{\alpha+\lambda}y^{n-1}e^{-(\alpha+\lambda)y}\right]_{y=0}^\infty - \int_0^\infty (n-1)y^{(n-1)-1}\cdot -\frac{1}{\alpha+\lambda}e^{-(\alpha+\lambda)y}\ dy\right) \\
&= \frac{\lambda^n}{(n-1)!}\left((0+0) + \frac{n-1}{\alpha+\lambda}\int_0^\infty y^{n-2}e^{-(\alpha+\lambda)y}\ dy\right) \\
&= \frac{\lambda^n}{(n-1)!}\frac{n-1}{\alpha+\lambda}\int_0^\infty y^{n-2}e^{-(\alpha+\lambda)y}\ dy \\
&= \frac{\lambda^n}{(n-1)!}\frac{n-1}{\alpha+\lambda}\left(\left[-\frac{1}{\alpha+\lambda}y^{n-2}e^{-(\alpha+\lambda)y}\right]_{y=0}^\infty + \frac{n-2}{\alpha+\lambda}\int_0^\infty y^{n-3}e^{-(\alpha+\lambda)y}\ dy\right) \\
&= \frac{\lambda^n}{(n-1)!}\frac{(n-1)(n-2)}{(\alpha+\lambda)^2}\left(0 + \int_0^\infty y^{n-3}e^{-(\alpha+\lambda)y}\ dy\right) \\
&= \frac{\lambda^n}{(n-1)!}\frac{(n-1)(n-2)}{(\alpha+\lambda)^2}\int_0^\infty y^{n-3}e^{-(\alpha+\lambda)y}\ dy \\
&= \frac{\lambda^n}{(n-1)!}\frac{(n-1)!}{(\alpha+\lambda)^n} \\
&= \frac{\lambda^n}{(\alpha+\lambda)^n}.
\end{aligned}$$

Dus voor alle $n \in \mathbb{N}_{>0}$ geldt dat $\mathbb{E}\left[e^{-\alpha Y(n)}\right] = \frac{\lambda^n}{(\alpha+\lambda)^n}$.

(2) Voor de limiet van n naar oneindig geldt dan dat

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\left[e^{-\alpha Y(n)}\right] &= \lim_{n\to\infty} \frac{\lambda^n}{(\alpha+\lambda)^n} \\ &= 0.\end{aligned}$$

□

Een gevolg van lemma's 7.2 en 7.3 is het volgende.
Gevolg 7.4. *Er geldt dat* $\lim_{n\to\infty} \mathbb{E}\left[e^{-\alpha X(n)}\right] = 0$.

Bewijs. Wegens lemma 7.2 volgt dat $\mathbb{E}\left[e^{-\alpha X(n)}\right] \leq \mathbb{E}\left[e^{-\alpha Y(n)}\right]$, met $Y(n)$ een stochastische variabele die Erlang verdeeld is met parameters $n \in \mathbb{N}_{>0}$ en λ. Verder geldt dan ook dat

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-\alpha X(n)}\right] \leq \lim_{n\to\infty} \mathbb{E}\left[e^{-\alpha Y(n)}\right].$$

Dan volgt met behulp van lemma 7.3 dat

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-\alpha X(n)}\right] \leq 0.$$

Omdat voor alle n geldt dat $\mathbb{E}\left[e^{-\alpha X(n)}\right] \geq 0$ volgt dan dus dat geldt dat

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-\alpha X(n)}\right] = 0.$$

□

Nu we deze lemma's bewezen hebben kunnen we de volgende stelling gaan bewijzen.
Stelling 7.5. *Beschouw het model met verdisconteerde kosten waarbij K^{aan} betaald moet worden bij het aanzetten van een server en K^{uit} bij het uitzetten van een server. Neem aan dat $K^{uit} < \frac{c}{\alpha}$. Dan geldt voor alle $n \in \mathbb{N}_{>0}$ met $n \geq n_0$, waarbij*

$$n_0 = \frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda}{\alpha+\lambda}\right)},$$

dat het op diagonaal $D(n)$ optimaal is om minstens één server uit te zetten.

Bewijs. Zij $n \in \mathbb{N}_{>0}$. Om te bewijzen dat het voor alle $n \geq n_0$, met

$$n_0 = \frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda}{\alpha+\lambda}\right)},$$

optimaal is om op diagonaal $D(n)$ minstens één server uit te zetten moet er bewezen worden dat het voor $n \geq n_0$, met

$$n_0 = \frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda}{\alpha+\lambda}\right)},$$

optimaal is om in toestand $(n-1,1)$ de server uit te zetten als er een klant vertrekt.

Om dit te bewijzen beschouwen we de strategieën

$$\begin{aligned} S_2 &= \{\text{Zet op tijdstip 0 in toestand } (n-1,1) \text{ de server op standby,} \\ &\quad \text{daarna optimaal tot uit driehoek } T(n)\} \end{aligned}$$

en

$$\begin{aligned} S_1 &= \{\text{Zet op tijdstip 0 in toestand } (n-1,1) \text{ de server uit,} \\ &\quad \text{imiteer daarna strategie } S_2 \text{ tot uit driehoek } T(n)\}. \end{aligned}$$

Als geldt dat $\mathbb{E}\left[K^{aan}e^{-\alpha X(n)} + K^{uit}\right] \leq \mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha X(n)}\right)\right]$, dan volgt dat strategie S_1 goedkoper is dan strategie S_2 en dan is het dus optimaal om in toestand $(n-1,1)$ de server uit te zetten. Om te bepalen of deze ongelijkheid geldt beschouwen we het verschil tussen beide verwachtingswaarden. Dan volgt dat

$$\begin{aligned} &\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha X(n)}\right)\right] \\ &-\mathbb{E}\left[K^{aan}e^{-\alpha X(n)} + K^{uit}\right] &=& \frac{c}{\alpha} - \frac{c}{\alpha}\mathbb{E}\left[e^{-\alpha X(n)}\right] - K^{aan}\mathbb{E}\left[e^{-\alpha X(n)}\right] - K^{uit} \\ && =& \frac{c}{\alpha} - K^{uit} - \left(\frac{c}{\alpha}+K^{aan}\right)\mathbb{E}\left[e^{-\alpha X(n)}\right] \\ && \overset{(*)}{\geq}& \frac{c}{\alpha} - K^{uit} - \left(\frac{c}{\alpha}+K^{aan}\right)\mathbb{E}\left[e^{-\alpha Y(n)}\right] \\ && \overset{(**)}{=}& \frac{c}{\alpha} - K^{uit} - \left(\frac{c}{\alpha}+K^{aan}\right)\frac{\lambda^n}{(\alpha+\lambda)^n} \\ && \overset{(***)}{\geq}& \frac{c}{\alpha} - K^{uit} - \left(\frac{c}{\alpha}+K^{aan}\right)\left(\frac{\lambda}{\alpha+\lambda}\right)^{n_0} \\ && =& 0. \end{aligned}$$

(*) Wegens lemma 7.2.
(**) Wegens lemma 7.3.1.
(***) Waarbij n_0 de kleinste waarde is waarvoor geldt dat $\frac{c}{\alpha} - K^{uit} - \left(\frac{c}{\alpha}+K^{aan}\right)\left(\frac{\lambda}{\alpha+\lambda}\right)^{n_0} = 0$.

Uit de vergelijking $\frac{c}{\alpha} - K^{uit} - \left(\frac{c}{\alpha}+K^{aan}\right)\left(\frac{\lambda}{\alpha+\lambda}\right)^{n_0} = 0$ is n_0 te bepalen. Er volgt dat

$$\left(\frac{\lambda}{\alpha+\lambda}\right)^{n_0} = \frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}},$$

ofwel

$$n_0 = \tfrac{\lambda}{\alpha+\lambda}\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right).$$

Hieruit volgt dus dat

$$n_0 = \frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda}{\alpha+\lambda}\right)}.$$

Dus voor $n_0 = \frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda}{\alpha+\lambda}\right)}$ geldt dat $\frac{c}{\alpha} - K^{uit} - \left(\frac{c}{\alpha} + K^{aan}\right)\frac{\lambda^{n_0}}{(\alpha+\lambda)^{n_0}} = 0$.

Voor alle $n \in \mathbb{N}_{>0}$, waarvoor geldt dat $n \geq n_0$, geldt dat

$$\frac{\lambda^{n_0}}{(\alpha+\lambda)^{n_0}} \geq \frac{\lambda^n}{(\alpha+\lambda)^n}.$$

Dus er volgt voor $n \in \mathbb{N}_{>0}$ met $n \geq n_0$ dat

$$\frac{c}{\alpha} - K^{uit} - \left(\frac{c}{\alpha} + K^{aan}\right)\frac{\lambda^n}{(\alpha+\lambda)^n} \geq 0.$$

Voor alle $n \in \mathbb{N}_{>0}$ met $n \geq n_0 = \frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda}{\alpha+\lambda}\right)}$ geldt er dus dat

$$\mathbb{E}\left[\frac{c}{\alpha}\left(1 - e^{-\alpha X(n)}\right)\right] - \mathbb{E}\left[K^{aan}e^{-\alpha X(n)} + K^{uit}\right] \geq 0$$

en is het dus optimaal om in toestand $(n-1,1)$ de server uit te zetten. Hieruit volgt dan dat het voor alle $n \in \mathbb{N}_{>0}$, met $n \geq n_0 = \frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda}{\alpha+\lambda}\right)}$, op diagonaal $D(n)$ optimaal is om minstens één server uit te zetten. □

7.2 De strategie in toestand $(1, j)$, met $j \in \mathbb{N}_{\geq 0}$

Nu bekend is wat de optimale strategie in de toestanden $(i, 1)$ is, met $i \in \mathbb{N}_{\geq 0}$, kunnen we gaan kijken naar de optimale strategie in de toestanden $(1, j)$ met $j \in \mathbb{N}_{\geq 0}$.

Stelling 7.6. *Beschouw het model met verdisconteerde kosten waarbij K^{aan} betaald moet worden bij het aanzetten van een server en K^{uit} bij het uitzetten van een server. Neem aan dat $K^{uit} < \frac{c}{\alpha}$. Dan geldt voor alle $n \in \mathbb{N}_{>0}$ met $n \geq m_0$, waarbij*

$$m_0 = \frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda(\lambda+\mu)}{\alpha\mu+\lambda(\lambda+\mu)}\right)},$$

Bewijs. Zij $n \in \mathbb{N}_{>0}$. Stel je bent in toestand $(1, n-1)$ en er vertrekt een klant. We gaan bepalen wat de optimale strategie in deze toestand is. We gaan dit doen door de strategie waarbij de server uitgezet wordt in toestand $(1, n-1)$ te vergelijken met de optimale strategie nadat je de server in deze toestand op standby zet. Als geldt dat we een strategie kunnen vinden waarbij uitzetten goedkoper is, dan volgt dat het in toestand $(1, n-1)$ optimaal is om de server uit te zetten. We beschouwen daarom de volgende strategieën:

$$\begin{aligned} S_2 \;=\; &\{\text{Zet op tijdstip 0 in toestand } (1, n-1) \text{ de server op standby,}\\ &\text{daarna optimaal tot uit driehoek } T(n)\} \end{aligned}$$

en

$$\begin{aligned} S_1 \;=\; &\{\text{Zet op tijdstip 0 in toestand } (1, n-1) \text{ de server uit,}\\ &\text{zet daarna de servers op standby tot strategie } S_2 \text{ ontmoet wordt}\\ &\text{daarna optimaal tot uit driehoek } T(n)\}. \end{aligned}$$

We gaan de kosten beschouwen totdat de strategiën S_1 en S_2 elkaar ontmoeten. Er geldt dat de strategiën S_1 en S_2 op twee manieren elkaar kunnen ontmoeten, namelijk

(1) het moment dat strategie S_2 in toestand $(0, n)$ belandt,

(2) het moment dat strategie S_2 naar diagonaal $D(n-1)$ gaat.

Omdat strategie S_1 zich altijd op dezelfde hoogte, wat betreft het aantal klanten, als strategie S_2 bevindt, en altijd één server minder op standby heeft staan, zijn deze twee manieren de enige twee manieren dat de twee strategieën in dezelfde toestand kunnen belanden.

We gaan nu de verwachte verdisconteerde kosten voor beide strategieën vergelijken.

(1) Beschouw nu eerst het moment dat beide strategieën elkaar ontmoeten in toestand $(0, n)$. Het moment dat beide strategieën elkaar ontmoeten in toestand $(0, n)$ noemen we $Z(n)$.

Voor de kosten tot tijdstip $Z(n)$ geldt dat je bij strategie S_1 een keer meer K^{aan} en K^{uit} moet betalen dan bij strategie S_2, maar bij strategie S_2 heb je tot tijdstip $Z(n)$ altijd één server meer op standby staan. We beschouwen daarom het volgende:

$$\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha Z(n)}\right)-\left(K^{uit}+K^{aan}e^{-\alpha Z(n)}\right)\right] \quad = \quad \frac{c}{\alpha}-K^{uit}-\left(\frac{c}{\alpha}+K^{aan}\right)\mathbb{E}\left[e^{-\alpha Z(n)}\right].$$

Zij $a(n)$ de kans dat er $n+1$ klanten in het systeem zijn voordat het systeem leegraakt, gegeven dat je met n klanten start. Zij $Y(n)$ een stochastische variabele die Erlang verdeeld is met parameters $n \in \mathbb{N}_{>0}$ en λ. Zij verder $U(n)$ een stochastische variabele die exponentieel verdeeld met parameter $\lambda + (n-1)\mu$. Dan geldt dat

$$\begin{aligned} Z(n) &\overset{d}{\geq} a(n)U(n) + (1-a(n))Y(n) \\ &\overset{d}{\geq} (1-a(n))Y(n). \end{aligned}$$

De laatste ongelijkheid geldt wegens het feit dat zowel $a(n)$ als $U(n)$ groter dan 0 zijn voor alle n.
Er geldt dat $a(n) \leq a(1)$ en hieruit volgt dan dat

$$\begin{aligned} Z(n) &\overset{d}{\geq} (1-a(1))Y(n) \\ &\overset{d}{=} \left(1-\frac{\lambda}{\lambda+\mu}\right)Y(n) \\ &\overset{d}{=} \frac{\mu}{\lambda+\mu}Y(n). \end{aligned}$$

Er geldt dus dat

$$Z(n) \overset{d}{\geq} \frac{\mu}{\lambda+\mu}Y(n).$$

Analoog aan het bewijs van lemma 7.2, dat $\mathbb{E}\left[e^{-\alpha Z(n)}\right] \leq \mathbb{E}\left[e^{-\alpha\frac{\mu}{\lambda+\mu}Y(n)}\right]$. Verder volgt analoog aan het bewijs van lemma 7.3.1 dat $\mathbb{E}\left[e^{-\alpha\frac{\mu}{\lambda+\mu}Y(n)}\right] = \left(\frac{\lambda(\lambda+\mu)}{\alpha\mu+\lambda(\lambda+\mu)}\right)^n$.

Er volgt nu dat

$$\begin{aligned}
\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha Z(n)}\right)-\left(K^{uit}+K^{aan}e^{-\alpha Z(n)}\right)\right] &= \frac{c}{\alpha}-K^{uit}-\left(\frac{c}{\alpha}+K^{aan}\right)\mathbb{E}\left[e^{-\alpha Z(n)}\right] \\
&\geq \frac{c}{\alpha}-K^{uit}-\left(\frac{c}{\alpha}+K^{aan}\right)\mathbb{E}\left[e^{-\alpha\frac{\mu}{\lambda+\mu}Y(n)}\right] \\
&= \frac{c}{\alpha}-K^{uit}-\left(\frac{c}{\alpha}+K^{aan}\right)\left(\frac{\lambda\left(\lambda+\mu\right)}{\alpha\mu+\lambda\left(\lambda+\mu\right)}\right)^{n} \\
&\overset{(*)}{\geq} \frac{c}{\alpha}-K^{uit}-\left(\frac{c}{\alpha}+K^{aan}\right)\left(\frac{\lambda\left(\lambda+\mu\right)}{\alpha\mu+\lambda\left(\lambda+\mu\right)}\right)^{m_0} \\
&= 0.
\end{aligned}$$

(*) Waarbij m_0 de kleinste waarde is waarvoor geldt dat

$$\frac{c}{\alpha}-K^{uit}-\left(\frac{c}{\alpha}+K^{aan}\right)\left(\frac{\lambda\left(\lambda+\mu\right)}{\alpha\mu+\lambda\left(\lambda+\mu\right)}\right)^{m_0}=0.$$

Uit de vergelijking $\frac{c}{\alpha}-K^{uit}-\left(\frac{c}{\alpha}+K^{aan}\right)\left(\frac{\lambda(\lambda+\mu)}{\alpha\mu+\lambda(\lambda+\mu)}\right)^{m_0}=0$ is m_0 te bepalen. Er volgt dat

$$\left(\frac{\lambda\left(\lambda+\mu\right)}{\alpha\mu+\lambda\left(\lambda+\mu\right)}\right)^{m_0} = \frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}},$$

ofwel

$$m_0 = {}^{\frac{\lambda(\lambda+\mu)}{\alpha\mu+\lambda(\lambda+\mu)}}\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right).$$

Hieruit volgt dus dat

$$m_0=\frac{\log\left(\frac{\frac{c}{\alpha}-K^{uit}}{\frac{c}{\alpha}+K^{aan}}\right)}{\log\left(\frac{\lambda(\lambda+\mu)}{\alpha\mu+\lambda(\lambda+\mu)}\right)}.$$

Dus voor alle $n\in\mathbb{N}_{>0}$, met $n\geq m_0$, geldt dat het in toestand $(1,n-1)$ optimaal is om de server uit te zetten als er een klant vertrekt.

(2) Beschouw nu het moment dat beide strategieën elkaar ontmoeten wanneer strategie S_2 naar diagonaal $D(n-1)$ gaat. Noem dit moment $X(n)$. Voor de kosten tot tijdstip $X(n)$ geldt dat je bij strategie S_1 een keer meer K^{uit} moet betalen dan bij strategie S_2, maar bij strategie S_2 heb je tot tijdstip $Z(n)$ altijd één server meer op standby staan en op tijdstip $Z(n)$ moet je K^{uit} extra betalen. Er geldt dat

$$\begin{aligned}
\mathbb{E}\left[\frac{c}{\alpha}\left(1-e^{-\alpha X(n)}\right)+K^{uit}e^{-\alpha X(n)}-K^{uit}\right] &= \frac{c}{\alpha}-K^{uit}-\left(\frac{c}{\alpha}-K^{uit}\right)\mathbb{E}\left[e^{-\alpha X(n)}\right] \\
&\overset{(**)}{\geq} 0.
\end{aligned}$$

(**) Want er geldt dat $\frac{c}{\alpha}>K^{uit}$.

In deze situatie geldt dus voor alle $n\in\mathbb{N}_{>0}$ dat het in toestand $(1,n-1)$ optimaal is om de server uit te zetten als er een klant vertrekt.

In situatie (1) geldt dat het voor alle $n \in \mathbb{N}_{>0}$ met $n \geq m_0$, waarbij

$$m_0 = \frac{\log\left(\frac{\frac{c}{\alpha} - K^{uit}}{\frac{c}{\alpha} + K^{aan}}\right)}{\log\left(\frac{\lambda(\lambda+\mu)}{\alpha\mu + \lambda(\lambda+\mu)}\right)},$$

optimaal is om in toestand $(1, n-1)$ de server uit te zetten als er een klant vertrekt. Verder geldt in situatie (2) dat het voor alle $n \in \mathbb{N}_{>0}$ in toestand $(1, n-1)$ optimaal is om de server uit te zetten als er een klant vertrekt.

Hieruit volgt dat dat het voor alle $n \in \mathbb{N}_{>0}$ met $n \geq m_0$, waarbij

$$m_0 = \frac{\log\left(\frac{\frac{c}{\alpha} - K^{uit}}{\frac{c}{\alpha} + K^{aan}}\right)}{\log\left(\frac{\lambda(\lambda+\mu)}{\alpha\mu + \lambda(\lambda+\mu)}\right)},$$

optimaal is om in toestand $(1, n-1)$ de server uit te zetten als er een klant vertrekt. $\square$

8 Conclusie

We hebben laten zien dat er onder bepaalde voorwaarden van de parameters in iedere toestand (i, j) een optimale strategie te geven is wat betreft het op standby zetten of uitzetten van de servers als er een klant vertrekt. Deze optimale strategie geeft dan de minimale verwachte verdisconteerde kosten, wat is waar we naar op zoek waren. Er zijn drie verschillende situaties wat betreft de voorwaarden van de parameters en in alle drie deze situaties is de optimale strategie te geven.

Voor de situatie dat geldt dat $c \geq K^{aan}\lambda + K^{uit}(\alpha + \lambda)$ geldt dat het in alle toestanden optimaal is om de server uit te zetten als er een klant vertrekt.

Als geldt dat $c \leq K^{aan}\lambda + K^{uit}(\alpha + \lambda)$ en $\frac{c}{\alpha} \leq K^{uit}$, dan geldt ook dat het in alle toestanden optimaal is om de server uit te zetten als er een klant vertrekt.

Echter als geldt dat $c \leq K^{aan}\lambda + K^{uit}(\alpha + \lambda)$ en $\frac{c}{\alpha} > K^{uit}$, dan is de optimale strategie gegeven door:

- Voor alle toestanden $(0, n)$, met $n \in \mathbb{N}_{>0}$, is het optimaal om de server op standby te zetten als er een klant vertrekt.
- Voor alle $n \in \mathbb{N}_{>0}$, met $n \geq n_0$, waarbij

  $$n_0 = \frac{\log\left(\frac{\frac{c}{\alpha} - K^{uit}}{\frac{c}{\alpha} + K^{aan}}\right)}{\log\left(\frac{\lambda}{\alpha+\lambda}\right)},$$

  is het in ieder geval in toestand $(n-1, 1)$ optimaal om de server uit te zetten als er een klant vertrekt.
- Voor alle $n \in \mathbb{N}_{>0}$, met $n \geq m_0$, waarbij

  $$m_0 = \frac{\log\left(\frac{\frac{c}{\alpha} - K^{uit}}{\frac{c}{\alpha} + K^{aan}}\right)}{\log\left(\frac{\lambda(\lambda+\mu)}{\alpha\mu + \lambda(\lambda+\mu)}\right)},$$

is het optimaal om in toestand $(1, n-1)$ de server uit te zetten als er een klant vertrekt.

- Voor alle toestanden (i, j) waarin geldt dat het optimaal is om de server uit te zetten, geldt voor alle toestanden die lager liggen op diagonaal $D(i+j)$ ook dat het optimaal is om de server uit te zetten als er een klant vertrekt.

Referenties

[1] R.B. Kappetein, *Optimal control of a server farm. Universiteit Leiden*, Masterscriptie, Leiden, januari 2014.
[2] A.C.C. van Wijk, *Pooling and Polling: Creation of pooling in inventory and queueing, Technische Universiteit Eindhoven*, Proefschrift, pp. 123-136, Eindhoven, 2012.