

BACHELOR

Call center mathematics

van de Laar, B.J.J.

Award date:
2007

[Link to publication](#)

Disclaimer

This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

Call Center Mathematics

Versienummer: 2.0

Datum: 18-09-2007

Auteur: Bart van de Laar

Opleiding: Technische Wiskunde

Universiteit: Technische Universiteit Eindhoven

Begeleider: Ivo Adan

Inhoudsopgave

1	Call Center model	4
1.1	Terminologie	4
1.2	Karakteristieken	5
2	Performance van een Call Center	7
3	<i>M/M/s</i>-model	9
4	<i>M/M/1</i>-model	15
5	Voorspellingen	17
6	Omgaan met onzekerheden	21
7	Multiple Skills	22
8	Skilled based Routing	23
9	TUI	25
9.1	Algemeen	25
9.2	Call Center	25
9.3	Onderzoeksopdracht	26
10	Prioriteitsmodellen	27
10.1	Overall prioriteit type 1	28
10.1.1	Karakteristieken	28
10.1.2	Wachttijden met overall prioriteit	29
10.1.3	Simulatie in Arena	31
10.2	Prioriteit type 1 zonder onderbreken	35
10.2.1	Karakteristieken	35
10.2.2	Wachttijden met prioriteit zonder onderbreken	36
11	Slow-server queue	38
12	Model met generalisten	45
12.1	Simulatie	45
12.2	Conclusie	48
12.3	Discussiepunten	48
13	Bijlage	49
13.1	<i>Arena</i>	49
13.1.1	Model met prioriteit	49
13.1.2	Model met generalisten	51

13.2	Grafieken van tabel	52
13.3	Slow-server queue	53
13.3.1	Afleiding van tweedegraadsvergelijking	53
13.3.2	Bestaan van grenswaardes	54
13.3.3	ρ_+ -analyse	54
13.3.4	ρ_- -analyse	54
13.4	R -taal	55

Inleiding

In een Call Center speelt wiskunde een belangrijke rol. Zo kan er met behulp van wiskundige modellen een goede verhouding verkregen worden tussen servicekwaliteit en de kosten. Om een Call Center goed te besturen zijn verschillende vaardigheden nodig. Bij deze vaardigheden heb je van de ene kant te maken met algemene vaardigheden, vaardigheden die bij ieder bedrijf horen zoals motivatie en het trainen van werknemers, aan de andere kant heb je die vaardigheden die specifiek zijn voor je product. Voor de laatste vaardigheden kan de wiskunde erg behulpzaam zijn.

Voor een werkgever is het van belang om de wiskunde achter een Call Center te weten, aangezien niet alles door een computer geregeld kan worden. Zo is het niet mogelijk om in een model de 'mens' te modelleren. Altijd blijf je afhankelijk van het feit dat je met mensen te maken hebt. Om goed op bepaalde situaties te kunnen reageren is het dan ook van belang voor een Call Center manager om de wiskunde achter een Call Center te begrijpen.

Daarnaast kan wiskunde gebruikt worden om te testen of een verandering een positief effect zal hebben. Door het wiskundige model te simuleren en te implementeren kan voordat de verandering ingezet wordt al duidelijk worden wat het effect zal zijn. Dit zorgt vanzelfsprekend voor kostenbesparing en voor inzicht voor de manager of deze wel dan niet de verandering in wil voeren.

In dit verslag zal eerst worden verteld wat er allemaal ontdekt is uit de literatuurstudie. Daarna zal het onderzoek starten. Allereerst zal onderzocht gaan worden wat de invloed is van verschillende prioriteitsregels in een Call Center. Verder komt het *slow-server queue*-probleem besproken. Daarna zullen met het programma *Arena* verschillende modellen gesimuleerd worden en onderzocht zal worden hoeveel generalisten en specialisten er nodig zijn om de performance van het Call Center te vergroten.

Literatuuronderzoek

De bronnen die voor de literatuur gebruikt zijn, staan in de sectie Referenties, aan het einde van dit hoofdstuk. In de tekst zal verwezen worden naar de bronnen wanneer dit nodig is, waardoor duidelijk zal worden waar de informatie vandaan gekomen is

1 Call Center model

1.1 Terminologie

Wanneer een klant belt naar het Call Center, moet het Call Center eigenlijk twee taken verrichten:

- Het identificeren van het binnenkomende gesprek door een agent of door de ACD (*Automatic Call Distributor*) waarna het gesprek doorgestuurd wordt naar de juiste agent.
- Het behandelen van het gesprek door de gevraagde informatie te verschaffen of door de gevraagde acties uit te voeren.

Met de term 'agent' worden zowel generalisten als specialisten bedoeld. Generalisten behandelen de standaardgesprekken waarbij simpele handelingen of algemene informatie verschaft dient te worden, acties waarvoor dus algemene kennis nodig is. Specialisten behandelen gesprekken die meer diepgaand zijn, waarvoor meer kennis nodig is. Generalisten zijn dan ook vaak in te zetten op meerdere typen gesprekken, echter heeft dit een langzamere bediening tot gevolg.

De parameters die het belangrijkste zijn in een Call Center zijn de volgende: de aankomstintensiteit (gemiddelde aantal aankomsten per tijdseenheid) λ , de gemiddelde behandeltijd door een agent μ , en het aantal agenten vaak aangegeven door de letter s . Deze parameters bepalen de bezetting van de agenten en de snelheid van het proces.

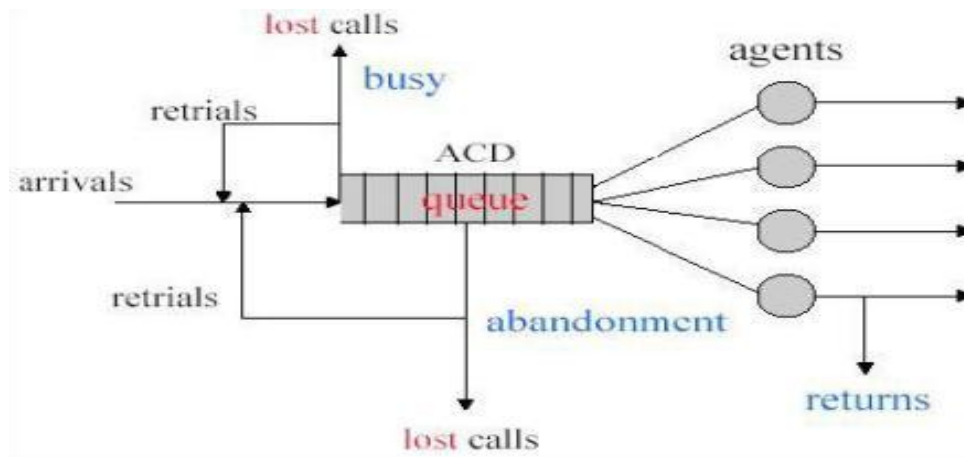
1.2 Karakteristieken

Een Call Center kan op verschillende manieren gekarakteriseerd worden, wat vaak afhangt van de doelstellingen van het Call Center. Welke vaardigheden hebben de agenten nodig, welk type gesprekken neemt het Call Center aan, hoe groot is het Call Center, wat voor soort klanten heeft het Call Center, wat zijn de parameters (λ , μ en s), wat is de ligging, behandelt het Call Center ook meer dan alleen telefoongesprekken, etc. Voor ieder Call Center kunnen deze vragen beantwoord worden maar hét Call Center bestaat niet. Deze karakteristieken zullen dan ook bepaald moeten worden voor het Call Center waar op het moment mee gewerkt wordt.

In de ideale situatie krijgen klanten die bellen meteen een agent aan de lijn die kan helpen. Helaas is dit zo goed als nooit het geval. Het kan voorkomen dat alle agenten bezig zijn wanneer een klant belt, waardoor deze klant niet direct geholpen kan worden.

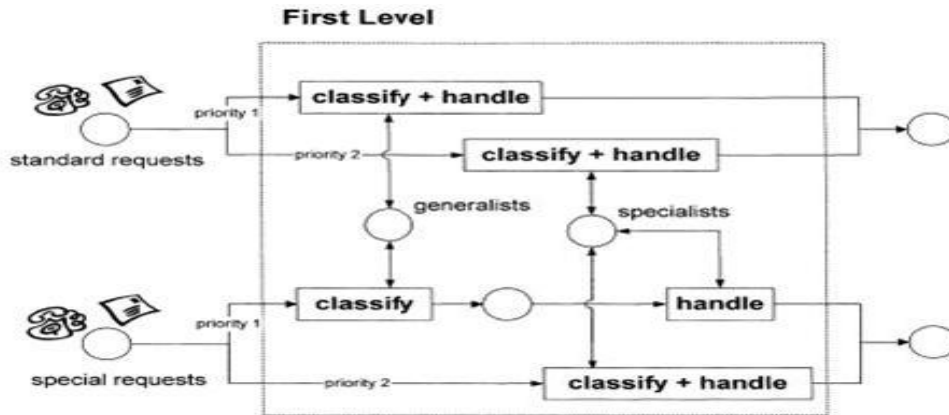
Deze klant kan dan in een wachtrij komen te staan of, als de server het niet aan kan, dan wordt de klant geblokkeerd en wordt toegang tot de server onthouden.

In figuur 1 staat een model van een Call Center. Van de binnenkomende gesprekken



Figuur 1: Mogelijke representatie van een Call Center

zijn er een paar klanten die niet in de server (ACD=Automatic Call Distributor) komen. Een deel van deze klanten probeert het opnieuw (retrials), andere klanten gaan verloren (lost calls). Van de klanten die in de queue komen zijn er een paar die eerder ophangen alvorens ze geholpen zijn (abandonment), waarvan er vervolgens een deel is dat het opnieuw probeert (retrials) en een deel dat het niet meer probeert (lost calls). Klanten die uiteindelijk bij de agenten komen worden geholpen.



Figuur 2: Mogelijke representatie van een Call Center

Een andere representatie wordt gegeven in figuur 2. Dit Call Center maakt onderscheid in gesprekken met algemene vragen en met speciale vragen. Daartussen maakt het onderscheid in prioriteit van sommige van deze gesprekken. Het kan zijn dat mensen die bijvoorbeeld vaker bellen prioriteit krijgen boven mensen die voor de eerste keer bellen. Dit onderscheid wordt door het management gemaakt en uitgevoerd door de ACD.

Sommige generalisten identificeren de binnenkomende gesprekken en sturen ze dan door naar specialisten of naar een andere generalist, al naar gelang het onderwerp van het gesprek of wat het probleem is van de klant. Zo worden speciale vragen met prioriteit 1 geïdentificeerd door generalisten en daarna behandeld door specialisten of generalisten. Standaardvragen met prioriteit 1 worden behandeld door generalisten, etc.

Hierboven zijn twee voorbeelden gegeven van mogelijke modellen van een Call Center. Er zijn meer modellen te geven, echter is deze configuratie per Call Center verschillend. Voordat aan een probleem van een Call Center gewerkt kan worden, zal eerst een model gemaakt dienen te worden.

2 Performance van een Call Center

Ieder Call Center wil er voor zorgen dat ze goed presteert. Of een Call Center goed presteert hangt af van verschillende aspecten, aspecten die gerelateerd zijn aan het wachtproces en aspecten die gerelateerd zijn aan het omgaan met de klanten zelf. Het moge duidelijk zijn dat het moeilijker wordt om beter te presteren wanneer gesprekken binnenkomen die verschillende vaardigheden vereisen. Om deze te kunnen behandelen zullen werknemers moeten beschikken over meerdere vaardigheden.

Het Service-Level (SL) kan gedefinieerd worden als de mate van tevredenheid van de klant. Het kan daarom gezien worden als een maat voor de prestatie van het Call Center. Hoe hoger het SL uitvalt, hoe hoger de prestatie is. Het SL wordt vaak aangegeven met twee cijfers, bijvoorbeeld 80/20, wat wil zeggen dat 80% van de gesprekken binnen 20 seconden opgenomen is. Hoe hoog het Call Center het Service-Level wil hebben, hangt af van het type Call Center en de doelstellingen van dat Call Center.

Er zijn verschillende niveaus waarop beslissingen genomen worden om het Service-Level te verbeteren:

- Strategische beslissingen. Beslissingen over het budget dat beschikbaar is en wat voor type service er verleend wordt. Deze worden genomen door hoofdbestuur en worden een paar keer in het jaar genomen.
- Tactische beslissingen. Beslissingen over het trainen van werknemers en hoe de middelen te gebruiken die beschikbaar zijn. Vaak zijn deze het gevolg van strategische beslissingen die genomen zijn.
- Planning beslissingen. Hoe de werknemers in het rooster te plannen. Deze worden vaak wekelijks, soms ook dagelijks genomen.
- Dagelijkse controle. Hoe te reageren op de dagelijkse situaties en hoe de binnenkomende gesprekken te verdelen over de werknemers. Zo kan het zijn dat een beller voorrang krijgt boven een andere, etc.

De belangrijkste prestatie-indicatoren voor effectiviteit zijn First-Time-Resolution (FTR) en Average-Holding-Time (AHT). FTR is de mate waarin klanten na één keer bellen voldoende hebben. Deze klanten zijn dus voldoende geholpen na hun eerste telefoontje. Als de FTR hoog is, is het aantal follow-up gesprekken laag. AHT staat voor de gemiddelde tijd dat een klant moet wachten.

Een belangrijk doel van elke Call Center is het verhogen van de FTR en het verlagen van de AHT. Het rendement is een belangrijk aspect hiervan, die aangeeft wat de fractie van de beschikbare tijd is dat de werknemer een klant aan het helpen is. Het rendement is te berekenen met formule (1).

$$Rendement = \frac{w_t}{t_b} * 100\% \quad (1)$$

Hierin stelt w_t de totale tijd voor die de werknemer aan het werk is en t_b de totale tijd die een werknemer beschikbaar is. Dit is dus de tijd dat een agent werkt plus de tijd dat deze

niet aan het werk is, maar wel aanwezig is op de werkvloer.

In veel gevallen wordt het Service-Level bepaald door te kijken naar de fractie van gesprekken waarvan de wachttijd een bepaalde vaste tijd ('Acceptable-Waiting-Time' (AWT)) niet overschrijdt. Dit percentage wordt ook wel 'Telephone-Service-Factor' (TSF) genoemd. Een andere veelgebruikte wachttijd metriek is 'Average-Speed-of-Answer' (ASA). Of een bepaald TSF-percentage goed is voor het Service-Level hangt vaak af van het gedrag van de klant. Het is namelijk onbekend of een klant een tweede keer probeert te bellen, als het de eerste keer niet gelukt is om iemand aan de lijn te krijgen. Het kan ook dat de klant niet wil wachten en al ophangt voordat hij de 'Acceptable-Waiting-Time' gewacht heeft. Een ander gevolg, wat ontstaat wanneer een Call Center streeft naar een zo hoog mogelijk TSF, is dat het langst-wachtende gesprek nooit beantwoord zal worden. Dit vanwege het feit dat men er voor wil zorgen dat het aantal gesprekken dat voor de AWT beantwoord wordt zo hoog mogelijk is. Dit leidt vanzelfsprekend tot ontevredenheid van de klanten die al langer dan AWT aan het wachten zijn.

Een ander manier om Service-Level te berekenen is aan de hand van een model met testklanten. Hierbij maak je in je model gebruik van klanten die oneindig lang willen wachten. Het grote voordeel van deze manier van interpreteren is dat je onafhankelijk bent van het gedrag van de klant. Deze test-klanten hangen namelijk niet eerder op en verlaten dus de wachtrij niet, voordat ze behandeld zijn. Echter is deze benadering niet representatief voor de werkelijkheid. Wel kunnen bepaalde scenario's getest worden door met een model met testklanten te werken. Hierdoor kan de manager inzichten in hoe een Call Center werkt verkrijgen.

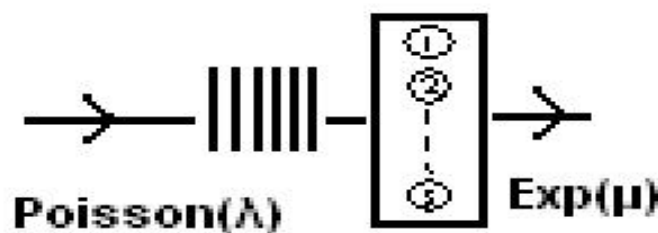
Het doel van een Call Center is het zo hoog mogelijk krijgen van het Service-Level. Om dit voor elkaar te krijgen, zijn meerdere benaderingen bekend. Welke de beste is hangt af van het type Call Center en welke doelstellingen het Call Center voor ogen heeft.

3 $M/M/s$ -model

Voordat er moeilijkere modellen gemaakt worden, zal er eerst een eenvoudig model worden beschreven en geanalyseerd. Hierbij worden de volgende aannames gemaakt:

- Er werken s agenten in het Call Center.
- Gesprekken komen aan via een Poisson-Proces met intensiteit λ . Dat betekent dat de tussenaankomsttijden onafhankelijk en exponentieel verdeeld zijn met parameter λ .
- De servicetijd van iedere agent is exponentieel verdeeld met parameter μ en onafhankelijk van andere bedieningstijden en tussenaankomsttijden.
- Als een klant belt en deze ziet dat de agenten aan het werk zijn, dan zal deze klant achteraan in de wachtrij komen.
- Er is geen voorkeursbehandeling van klanten.
- Klanten hangen niet eerder op. Als een klant in de wachtrij zit, zal deze de server niet meer verlaten alvorens hij geholpen is.

Dit model wordt ook wel het $M/M/s$ -model genoemd. Hierbij duidt de eerste M er op dat het aankomstproces een Poisson Proces is, de tweede M dat de servicetijden exponentieel verdeeld zijn en de s duidt erop dat er s agenten werkzaam zijn. Dit is schematisch weergegeven in figuur 3.



Figuur 3: Schematische weergave $M/M/s$ -model

Stel dat gesprekken arriveren volgens dit model. Dit houdt in dat de tussenaankomsttijden onafhankelijk exponentiële stochasten zijn met gemiddelde $\frac{1}{\lambda}$. Als een van de agenten klaar is met een klant, zal deze klant de server verlaten en zal de klant die voor in de wachtrij staat geholpen worden. De servicetijden zijn onafhankelijk exponentiële stochasten met gemiddelde $\frac{1}{\mu}$.

Veronderstel dat het proces zich in toestand i bevindt. Dit betekent dat er i mensen in

de server zitten. Definieer nu $N(t)$ als het aantal klanten in het systeem op tijdstip t . Dit zijn dus zowel de klanten die in de wachtrij staan, als de klanten die op dat moment geholpen worden. $\{N(t), t \geq 0\}$ is nu een geboorte-sterfteproces aangezien er alleen maar een sprong gemaakt kan worden van ten hoogste grootte één. Wanneer er namelijk i mensen in de server zitten, zal het proces naar toestand $i-1$ gaan wanneer een klant geholpen is, óf naar toestand $i+1$ gaan wanneer er een nieuw gesprek binnenkomt. Omdat zowel de aankomsttijden als de servicetijden exponentieel verdeeld zijn, begint het proces als het ware opnieuw nadat er een gebeurtenis plaatsgevonden heeft. Dit houdt in dat als er een gebeurtenis plaatsgevonden heeft, dat vanuit de volgende toestand het aankomstproces en het serviceproces opnieuw beginnen. Hierbij is het niet van belang dat klanten al geholpen werden. De gesprekken van deze klanten beginnen als het ware opnieuw (dit wordt ook wel de geheugenloosheidseigenschap genoemd van de exponentiële verdeling). Een gebeurtenis zoals hierboven beschreven stelt dus zowel een aankomst als een vertrek van een klant voor.

Definieer nu λ_i als de overgangsintensiteit van toestand i naar toestand $i+1$ en definieer μ_i als de overgangsintensiteit van toestand i naar toestand $i-1$. Nu zullen de λ_i en μ_i bepaald worden.

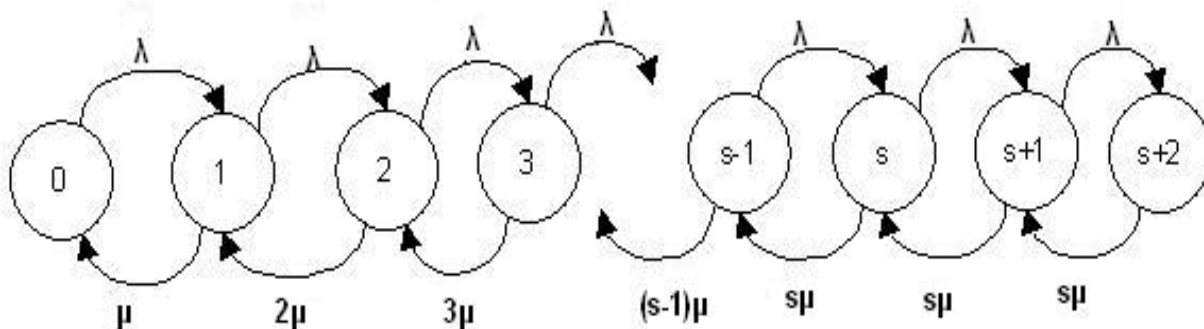
Wanneer het systeem zich in toestand 0 bevindt, is de enige sprong kan die gemaakt kan worden die naar toestand 1. Er bevindt zich namelijk niemand in het systeem en er wordt dus op dat moment ook niemand geholpen. Deze overgang gebeurt met intensiteit λ en dus is $\lambda_0 = \lambda$. Aangezien er per aankomst maar één gesprek binnenkomt geldt voor iedere overgang van i naar toestand $i+1$ dat deze gebeurt met intensiteit λ en daarom geldt dat $\lambda_i = \lambda$ voor $i \geq 0$.

Wanneer het proces zich in toestand 1 bevindt, wordt er op dat moment een klant geholpen. Wanneer deze klant geholpen is alvorens er een nieuw gesprek binnengekomen is, zal deze klant het systeem verlaten en zal er dus een sprong gemaakt worden van toestand 1 naar toestand 0. Deze overgang gebeurt met intensiteit μ en dus geldt dat $\mu_1 = \mu$. Als het proces zich nu in toestand 2 bevindt, zullen er twee agenten zijn die klanten aan het helpen zijn. Aangezien iedere agent met snelheid μ werkt, wordt de intensiteit waarmee van toestand 2 naar toestand 1 gesprongen wordt 2μ . Dus geldt dat $\mu_2 = 2\mu$. Hetzelfde geldt tot en met toestand s , dus $\mu_i = i\mu$ voor $1 \leq i \leq s$.

Nu is het nog de vraag wat μ_i is voor $i > s$. Aangezien er maar s agenten zijn, zullen er ook maar s klanten tegelijk geholpen kunnen worden. De totale vertrekintensiteit is dan ook $s\mu$. Samengevat zijn de intensiteiten dus:

$$\begin{cases} \lambda_i = \lambda & \text{voor } i \geq 0 \\ \mu_i = i\mu & \text{voor } 1 \leq i \leq s \\ \mu_i = s\mu & \text{voor } i > s \end{cases}$$

In figuur 4 wordt schematisch weergegeven wat de toestanden in dit proces zijn. Het aantal klanten in het systeem loopt op tot ∞ . Met pijlen is aangegeven wat de sprongen zijn die het proces kan maken. Iedere sprong is ter grootte één, hetgeen komt doordat het een geboorte-sterfteproces is. Bij elke pijl is aangegeven wat de overgangsintensiteit is waarmee de betreffende sprong gemaakt wordt.



Figuur 4: Schematische weergave van toestanden van proces

Om iets te kunnen zeggen over de wachttijden van binnenkomende klanten zal eerst bekend moeten zijn wat de fractie van de tijd is dat het proces zich in de verschillende toestanden bevindt. Definieer daarom $P_j \equiv \lim_{t \rightarrow \infty} P_{ij}(t)$. Dit is de limietwaarde van de kans dat het proces zich in toestand j op tijdstip t bevindt. Dit is gelijk aan de fractie van de tijd dat het proces zich in toestand j bevindt. Om P_j te mogen definiëren, moet de limietkans voor elke toestand bestaan. Twee voldoende condities hiervoor zijn:

1. Vanuit elke toestand moet er een positieve kans zijn om naar iedere andere toestand te komen (toestanden moeten met elkaar communiceren).
2. De gemiddelde tijd om van toestand i terug te komen in toestand i moet eindig zijn. De kans dat je terugkomt in de toestand is 1 en de verwachte terugkeertijd moet eindig zijn, gegeven dat het proces start in toestand i .

De conditie voor het bestaan van de limietkansen in het $M/M/s$ -model is $\lambda < s\mu$. In woorden betekent dit dat er gemiddeld per tijdseenheid minder gesprekken binnen moeten komen dan dat s agenten gemiddeld per tijdseenheid kunnen verwerken. Daardoor zal er nooit een oneindige lange wachtrij ontstaan.

Voor elk interval $(0, t)$ geldt dat het aantal sprongen naar toestand j gelijk is aan het aantal sprongen uit toestand j (op een verschil van 1). Om vanuit toestand j een sprong te kunnen maken, zal het proces namelijk eerst een sprong naar toestand j moeten maken en andersom. Naarmate het aantal gesprekken groter wordt, zal het verschil van één niet meer op te merken zijn en geldt dat het aantal sprongen naar toestand j per tijdseenheid net zo groot is als het aantal sprongen vanuit toestand j (deze zijn dus in balans). Als dit nu per toestand bekeken wordt, krijg je de zogenaamde 'balansvergelijkingen'.

De balansvergelijkingen van het $M/M/s$ -model kunnen nu bepaald worden (door gebruik te maken van figuur 4):

$$\begin{aligned}
\lambda P_0 &= \mu P_1 \\
(\lambda + \mu)P_1 &= 2\mu P_2 + \lambda P_0 \\
(\lambda + 2\mu)P_2 &= 3\mu P_3 + \lambda P_1 \\
&\vdots \\
(\lambda + s\mu)P_s &= s\mu P_{s+1} + \lambda P_{s-1} \\
&\vdots
\end{aligned}$$

De bovenstaande vergelijkingen zijn als volgt samen te vatten:

$$\begin{aligned}
\lambda P_0 &= \mu P_1 \\
(\lambda + i\mu)P_i &= (i+1)\mu P_{i+1} + \lambda P_{i-1} \quad \text{als } 1 \leq i \leq s-1 \\
(\lambda + s\mu)P_i &= s\mu P_{i+1} + \lambda P_{i-1} \quad \text{als } i \geq s
\end{aligned} \tag{2}$$

Uit de vorige vergelijkingen is nu voor elke P_j een formule af te leiden in termen van P_0 :

$$\begin{aligned}
\lambda P_0 = \mu P_1 &\Rightarrow P_1 = (\lambda/\mu)P_0 \\
(\lambda + \mu)P_1 = 2\mu P_2 + \lambda P_0 &\Rightarrow P_2 = (\lambda^2/2\mu^2)P_0 \\
(\lambda + 2\mu)P_2 = 3\mu P_3 + \lambda P_1 &\Rightarrow P_3 = (\lambda^3/(6\mu^3))P_0 \\
&\vdots \\
(\lambda + (s-1)\mu)P_{s-1} = s\mu P_s + \lambda P_{s-2} &\Rightarrow P_s = (\lambda^s/(s!\mu^s))P_0 \\
(\lambda + (s)\mu)P_s = s\mu P_{s+1} + \lambda P_{s-1} &\Rightarrow P_{s+1} = (\lambda^{s+1}/(s!\mu^{s+1}))\frac{1}{s}P_0 \\
&\vdots
\end{aligned}$$

Samengevat worden dit de volgende vergelijkingen (met $\rho = \frac{\lambda}{s\mu}$):

$$P_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} P_0 = (s\rho)^i \frac{1}{i!} P_0 \quad \text{als } 0 \leq i \leq s \tag{3}$$

$$P_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{s^{i-s} s!} P_0 = \rho^{i-s} P_s \quad \text{als } i \geq s+1 \tag{4}$$

Aangezien P_j de fractie van de tijd is dat het proces zich in toestand j bevindt en er geldt dat het proces zich altijd in één van de toestanden bevindt, geldt:

$$\sum_{j=0}^{\infty} P_j = 1 \tag{5}$$

Wanneer nu de vergelijkingen (3), (4) en (5) samen gebruikt worden, volgt er een uitdrukking voor P_0 :

$$\begin{aligned} \sum_{j=0}^{\infty} P_j &= 1 = P_0 \left\{ \sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \sum_{j=s}^{\infty} \rho^{j-s} P_s \right\} = \dots \\ &= P_0 \left\{ \sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \sum_{j=s}^{\infty} \rho^{j-s} \right\} = P_0 \left\{ \sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \frac{1}{1-\rho} \right\} \Rightarrow \\ P_0 &= \left\{ \sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \frac{1}{1-\rho} \right\}^{-1} \end{aligned}$$

Door deze vergelijking te substitueren in formules (3) en (4) krijg je vergelijkingen voor P_i :

$$\begin{aligned} P_i &= \frac{(s\rho)^i \frac{1}{i!}}{\sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \frac{1}{1-\rho}} \quad \text{als } 0 \leq i \leq s \\ P_i &= \frac{(s\rho)^i \frac{1}{s^{i-s} s!}}{\sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \frac{1}{1-\rho}} \quad \text{als } i \geq s+1 \end{aligned}$$

De kans dat een klant moet wachten, P_W , is gelijk aan de fractie van de tijd dat de server bezet is. Dat is in termen van P_i :

$$\begin{aligned} P_W &= \sum_{j=s}^{\infty} P_j = P_s + P_{s+1} + P_{s+2} + \dots = P_s \sum_{i=0}^{\infty} \rho^i = \dots \\ &= \frac{P_s}{1-\rho} = \frac{1}{1-\rho} \frac{\frac{(s\rho)^s}{s!}}{\sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \frac{1}{1-\rho}} = \frac{(s\rho)^s}{s!(1-\rho) \sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + (s\rho)^s} \end{aligned}$$

Directe berekeningen doen met deze formule kan voor numerieke problemen zorgen. Als λ , μ en s groot zijn, zal de som die uitgerekend dient te worden een deling zijn van twee grote getallen. Dit veroorzaakt numerieke problemen en om dit te voorkomen zal de formule herschreven worden. Hiervoor zal een iteratieformule gemaakt worden.

$$B[s] := \frac{(s\rho)^s}{s!} \frac{1}{\sum_{j=0}^s \frac{(s\rho)^j}{j!}} \quad (6)$$

$$\begin{aligned} B[s] &= \frac{(s\rho)(s\rho)^{s-1}}{s(s-1)!} \frac{1}{\frac{(s\rho)(s\rho)^{s-1}}{s(s-1)!} + \sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!}} \\ &= \frac{\frac{s\rho}{s} B[s-1]}{\frac{s\rho}{s} B[s-1] + 1} = \frac{\frac{\lambda}{\mu} B[s-1]}{\frac{\lambda}{\mu} B[s-1] + s} \end{aligned}$$

In vergelijking (6) is de beginvoorwaarde gegeven door $B[0]=1$. Deze recursie geeft numeriek geen problemen. Nu is P_W :

$$P_W[s] = \frac{\frac{1}{1-\rho} \frac{(s\rho)^s}{s!}}{\frac{(s\rho)^s}{s!} \frac{1}{1-\rho} + \sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!}} = \frac{\frac{s\rho}{s} \frac{1}{1-\rho} B[s-1]}{1 + \frac{s\rho}{s} \frac{1}{1-\rho} B[s-1]} = \frac{\lambda B[s-1]}{s\mu(1-\rho) + \lambda B[s-1]} \quad (7)$$

Nu de kans bekend is dat een klant moet wachten, is het mogelijk om de ASA (Average-Speed-of-Answer) uit te drukken in deze kans. Om dit te bewerkstelligen zal gebruik gemaakt worden van de *Wet van Little*:

$$L_q = \lambda_a W_q \quad (8)$$

In vergelijking (8) is:

L_q het gemiddeld aantal klanten in de wachtrij

W_q de gemiddelde tijd dat een klant in de wachtrij zit

λ_a de gemiddelde aankomstsnelheid van klanten.

Omdat ASA gelijk is aan de gemiddelde tijd dat een klant in de wachtrij zit, dus aan W_q , geldt:

$$L_q = P_{s+1} + 2P_{s+2} + \dots = \sum_{i=s}^{\infty} (i-s)P_i \quad (9)$$

L_q kan nu als volgt worden uitgerekend:

$$\begin{aligned} L_q &= 0 \cdot P_s + P_{s+1} + 2P_{s+2} + \dots = P_s \{0 + 1\rho + 2\rho^2 + \dots\} \\ &= P_s \left(\sum_{i=0}^{\infty} i\rho^i \right) = P_s \frac{\rho}{(1-\rho)^2} = \frac{P_s}{1-\rho} \cdot \frac{\rho}{1-\rho} = P_W \frac{\rho}{1-\rho} \end{aligned}$$

Met dit resultaat en vergelijking (8) volgt voor de gemiddelde tijd dat een klant in de wachtrij zit:

$$W_q = \frac{L_q}{\lambda_a} = \frac{P_W \rho}{1-\rho} \frac{1}{\lambda} = \frac{P_W}{s\mu - \lambda}$$

4 $M/M/1$ -model

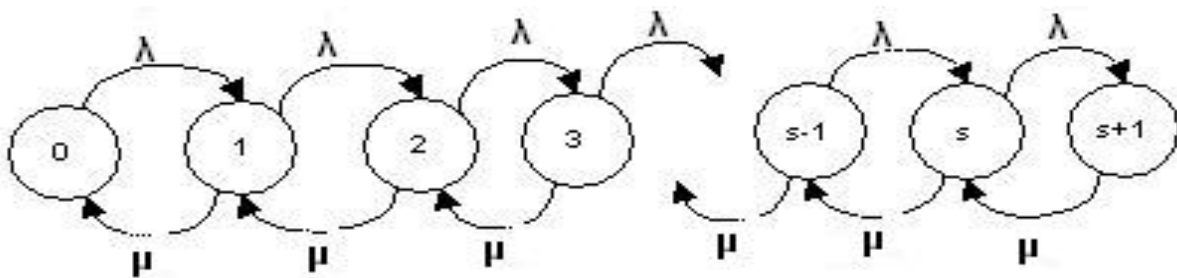
Het $M/M/1$ -model is een model waarin de volgende aannames gemaakt worden:

- Er werkt één agent in het Call Center.
- Gesprekken komen aan volgens een Poisson-Proces met intensiteit λ . Dat betekent dat de tussenaankomsttijden exponentieel verdeeld zijn met parameter λ .
- De servicetijd van de agent is exponentieel verdeeld met parameter μ
- Als een klant belt en deze ziet de agent aan het werk, dan zal deze klant achteraan in de wachtrij komen.
- Er is geen voorkeursbehandeling van klanten.
- Klanten hangen niet eerder op. Als een klant in de wachtrij zit, zal deze de server niet meer verlaten alvorens de klant geholpen is.

Definieer net zoals in het $M/M/s$ -model dat het systeem zich in toestand i bevindt, wanneer er i mensen in het systeem zijn. Voor het $M/M/1$ -model gelden de volgende overgangsintensiteiten:

$$\begin{cases} \lambda_i = \lambda & \text{voor } i \geq 0 \\ \mu_i = \mu & \text{voor } i \geq 1 \end{cases}$$

In figuur 5 is overzichtelijk weergegeven welke sprongen met welke intensiteiten genomen worden.



Figuur 5: Schematische weergave van toestanden van proces van $M/M/1$

De balansvergelijkingen voor dit model worden:

$$\begin{aligned} \lambda P_0 &= \mu P_1 \\ \lambda P_1 &= \mu P_2 \\ \lambda P_2 &= \mu P_3 \\ &\vdots \\ \lambda P_i &= \mu P_{i+1} \end{aligned}$$

In termen van P_0 is nu voor de P_i te schrijven:

$$P_i = \left(\frac{\lambda}{\mu}\right)^i P_0$$

Samen met vergelijking (5) volgt nu voor P_0 en daaruit voor P_i :

$$\begin{aligned} P_0 &= 1 - \frac{\lambda}{\mu} \\ P_i &= \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right) \end{aligned} \tag{10}$$

Een andere variant op de wet van Little is de volgende:

$$L = \lambda_a W \tag{11}$$

In vergelijking (11) is:

L het gemiddeld aantal klanten in het systeem

λ_a de gemiddelde aankomstsnelheid van binnenkomende gesprekken

W de gemiddelde tijd dat een klant in het systeem is.

Voor het gemiddeld aantal klanten in het systeem geldt dat deze gelijk is aan:

$$\begin{aligned} L &= \sum_{i=0}^{\infty} iP_i = \sum_{i=0}^{\infty} i \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right) = \dots \\ &\Rightarrow L = \frac{\lambda}{\mu - \lambda} \end{aligned}$$

Hier volgt nu uit dat de gemiddelde tijd dat een klant in het systeem zit, gelijk is aan:

$$W = \frac{L}{\lambda_a} = \frac{\frac{\lambda}{\mu - \lambda}}{\lambda} = \frac{1}{\mu - \lambda}$$

5 Voorspellingen

Data

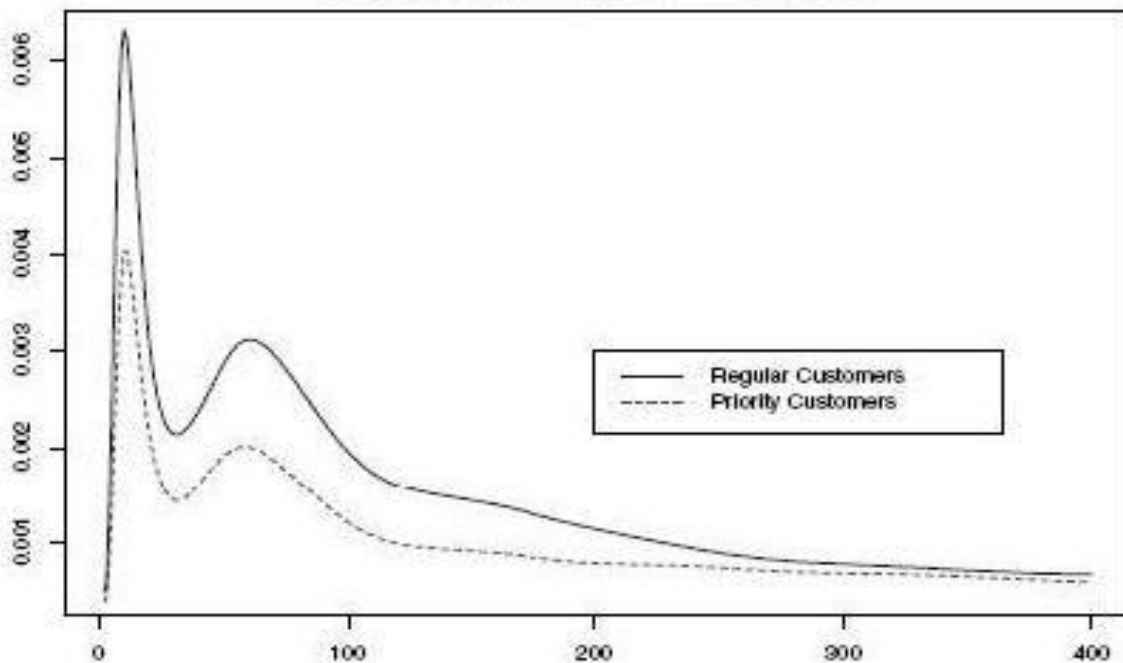
Het schatten van de parameters λ en μ is een van de belangrijkste dingen wat een Call Center moet doen, vanwege het feit dat op basis van de voorspelde parameters bijvoorbeeld de werknemers ingepland worden. Daarom moeten de parameters zo nauwkeurig mogelijk bepaald worden. Echter kan niemand in de toekomst kijken en daardoor is het moeilijk om exacte voorspellingen te kunnen doen.

Om goede schattingen te kunnen doen is het belangrijk dat de juiste data beschikbaar is. Voor de analyse is data nodig van verschillende typen:

- Data over klanten.
 - Hierbij kan gedacht worden aan aankomstpatronen, wachttijden van klanten, behandelingstijden en data die specificeert wat de onderwerpen van de klanten waren. Hierdoor kan bijvoorbeeld duidelijk worden wat voor onderwerp het meest gevraagd wordt en hierdoor kan het management er voor kiezen welke vaardigheden de agenten moeten hebben.
 - Daarnaast is het verkrijgen van data over het gedrag van klanten noodzakelijk (hierop wordt verder ingegaan in de volgende paragraaf)
- Data over agenten. Hierbij kan gedacht worden aan beschikbaarheid, vaardigheden en behandelings-snelheden. Hierdoor kan men bijvoorbeeld roosters maken waarbij verschillende vaardigheden op hetzelfde tijdstip beschikbaar zijn voor de klant.
- Data over interne gebeurtenissen. Dit zijn gebeurtenissen die door het Call Center veroorzaakt zijn, zoals reorganisatie en reclamecampagnes. Wanneer er bijvoorbeeld een reclamecampagne opgestart wordt, zal het aantal binnenkomende gesprekken toenemen (dat is in ieder geval het doel in de meeste gevallen). Met de juiste data kan de impact van een reclamecampagne voorspeld worden.
- Data over externe gebeurtenissen. Hierbij kan gedacht worden aan feestdagen, weersomstandigheden en bijvoorbeeld griepvirus. Wanneer bekend is dat er een feestdag aankomt, kan het aantal gesprekken fors dalen dan wel fors stijgen (afhankelijk van de feestdag en type Call Center). Hiermee kan met het inroosteren van agenten rekening gehouden worden. Daarnaast kunnen, wanneer bekend is dat er een griepvirus heerst in het land, op voorhand extra agenten ingeroosterd worden. Hierdoor kan voorkomen worden dat er teveel klanten te lang op een agent moeten wachten, wanneer blijkt dat veel agenten griep hebben.

Gedrag van klanten

Een van de moeilijkste factoren om rekening mee te houden bij het schatten van de parameters is het feit dat klanten eerder kunnen ophangen. Een vraag die meestal onbeantwoord blijft, is of deze klanten wel of niet proberen terug te bellen. Wanneer hierover data beschikbaar is, kan er mogelijk een patroon in ontdekt worden. Op basis van dit patroon is het dan mogelijk het gedrag van de klanten te voorspellen, waardoor er iets gezegd kan worden over de kans dat een persoon opnieuw probeert een agent aan de lijn te krijgen. In figuur 6 is een grafiek te zien van een functie van ongeduldigheid voor een bepaald Call Center.



Figuur 6: De intensiteit waarmee klanten ophangen uitgezet tegen de tijd in seconden van twee verschillende klanten

In figuur 6 (bron [2]) blz 36) staat de intensiteit waarmee klanten ophangen uitgezet tegen de verstreken wachttijd in seconden. In het figuur is verschil gemaakt tussen normale klanten (Regular Customers) en klanten met hoge prioriteit (Priority Customers). Klanten met hoge prioriteit zijn klanten die de hulp hard nodig hebben op dat moment.

In de grafiek vallen twee zaken op:

1. Aangezien klanten met hoge prioriteit de hulp harder nodig hebben, zullen deze ook geduldiger zijn en zullen ze dus bereid zijn langer te wachten totdat ze een agent aan de lijn krijgen

2. In de grafieken zitten twee pieken, eentje dicht tegen de oorsprong en eentje na ongeveer 60 seconden. De eerste piek wordt veroorzaakt doordat sommige klanten niet willen wachten en daardoor meteen ophangen. De tweede piek is veroorzaakt doordat na 60 seconden een bericht kwam, waarin klanten geïnformeerd werden wat de relatieve positie was in de wachtrij. Voor sommige klanten is dat het teken dat ze het een andere keer moeten proberen en besluiten daarop dan ook op te hangen. De piek bij 60 seconden is dan ook optioneel. Het is alleen om aan te geven dat wanneer klanten verteld wordt wat de relatieve positie is, klanten in veel gevallen beslissen om het een andere keer te proberen.

Helaas is iedere klant anders en is er dus geen 'model'-klant te maken. Echter hoe meer data beschikbaar is over het ophangen van klanten en het terugbellen van de klanten, hoe beter een 'model'-klant gemaakt kan worden voor het Call Center. Aan de hand daarvan kunnen de tijden berekend worden die klanten bereid zijn om te wachten. Hierop kan met het maken van het rooster dan ook ingespeeld worden.

Termijn

Nadat de data bekend is, zal het Call Center er voor moeten kiezen in wat voor termijnen (maanden, weken, dagen, etc) ze de parameters willen schatten. Naarmate de termijnen waarover schattingen gedaan worden kleiner zijn, hoe beter Call Centers kunnen anticiperen op gebeurtenissen. Echter heeft dit het nadeel dat er heel flexibel gewerkt dient te worden. Daarnaast moet het termijn niet al te klein gekozen worden. Wanneer dit wel het geval is, kan het zijn dat agenten er niet voor kunnen zorgen dat ze er zijn. Als een week van tevoren bekend is wie wanneer moet werken, kunnen de agenten hier rekening mee houden. Dit wordt moeilijker als agenten een uur van tevoren opgeroepen worden.

Als bijvoorbeeld de maandelijkse fluctuatie erg hoog is, kan het beter zijn om te kijken naar fluctuaties in voorgaande maanden. Dit kan worden gedaan met bijvoorbeeld de methodes 'moving average' of 'exponential smoothing'. Bij de eerste methode neem je steeds het gemiddelde van een n aantal maanden terug. Bij de tweede methode maak je gebruik van de geconstateerde fluctuaties met voorgaande maanden en je houdt het meeste rekening met de meest recente fluctuaties. Deze geef je als het ware de grootste invloed (vermenigvuldigen met de grootste factor). Dit blijkt een betrouwbare manier te zijn, wanneer de fluctuaties per maand niet te drastisch zijn.

Intervallen

Aangezien de parameters (λ en μ) kunnen fluctueren in intervallen, wordt er vaak gekozen voor het berekenen van zogenaamde betrouwbaarheidsintervallen, waardoor je bijvoorbeeld met 90% zekerheid kan zeggen dat een specifiek interval de parameter bevat. Deze intervallen worden vaak berekend met data van voorgaande tijden en op basis van ervaringen met de verschillende omstandigheden. Zo kan het voorkomen dat er iedere week op maandag meer gesprekken binnenkomen dan op de vrijdag. Met deze kennis kan dan rekening gehouden worden.

6 Omgaan met onzekerheden

In een Call Center zijn bepaalde factoren die het Service Level doen dalen. Een van deze factoren is de beschikbaarheid van de agenten. Onzekerheden hierin ontstaan door bijvoorbeeld ziekten, bijeenkomsten waar agenten naartoe moeten en andere verplichtingen. Manieren om met deze omstandigheden om te gaan:

- **Overbezetting.** Er kan voor gekozen worden om meer werknemers in te zetten dan dat er minimaal nodig zijn om het Service Level te bereiken.
Hierdoor kunnen ziektes direct opgevangen worden en kan het Call Center een hoger Service-Level halen. Nadeel is wel dat in geval van geen ziektes er teveel werknemers aan het werken zijn waardoor de kosten hoger zijn dan nodig.
- **Flexibele werknemers.** Er kan er voor gekozen worden om sommige werknemers zo te contracteren, dat er bijvoorbeeld per dag gekeken wordt of deze werknemer wel of niet moet komen. Dit is voordeliger dan overbezetting, aangezien bij overbezetting de mensen er altijd zijn, ook als de hoeveelheid binnenkomende gesprekken niet hoger uitvalt dan voorspeld was. Hierbij gezegd dat de hoeveelheid gesprekken in het begin van de dag een goede indicatie blijkt voor de hoeveelheid gesprekken in het midden en op het einde van de dag.
Het grote voordeel is dat er niet teveel werknemers ingepland worden. Er wordt per dag gekeken of dit wel dan niet nodig is. Verder kunnen zo de piekdagen opgevangen worden.
- **Mobiele werknemers.** Werknemers die mobiel bereikbaar zijn in het geval het Call Center in problemen komt.
Het voordeel is dat deze werknemers altijd beschikbaar zijn. Echter zullen deze werknemers wel extra getraind moeten worden en de nodige apparatuur zal hiervoor aangeschaft dienen te worden, wat de kosten weer hoger maakt.
- **Werknemers laten overwerken.** Dit is een relatief simpele manier om te reageren op hoge werklasten. Echter geldt dit alleen voor werk dat achteraf gedaan kan worden. Het aannemen van telefoontjes moet vanzelfsprekend op het moment zelf gebeuren, niet achteraf.
- **'Blending'.** Werknemers met andere taken waarvoor de druk niet zo hoog is, zoals faxen en uitgaande gesprekken, binnenkomende gesprekken aan laten nemen. De andere taken worden zo als het ware opvulling van de tijd waarop geen binnenkomende gesprekken zijn.
Voordeel hiervan is dat er geen extra mensen ingeroosterd dienen te worden en dat onvoorspelbare piektijden opgevangen kunnen worden door werknemers met normaal andere taken.

Een andere onzekerheid waar Call Centers tegen aan lopen, is het feit dat het aantal binnenkomende gesprekken van tevoren niet exact te voorspellen is. Het kan voorkomen dat er opeens een piek ontstaat in het aantal gesprekken. Daardoor moeten sommige klanten veel te lang wachten. Om dat laatste te voorkomen, kan het Call Center tegen de mensen zeggen dat ze later teruggebeld worden. Op deze manier kan het Call Center op momenten waarop het niet druk is de mensen wel te woord staan. Hierdoor kan er een hoger Service Level gehaald met minder werknemers. Daarnaast zijn de klanten ook tevreden aangezien ze weten dat ze wel geholpen gaan worden.

Verder kan er voor gekozen worden om mensen eerst met email contact te laten maken met het Call Center waarin ze aangeven welke tijden ze bereikbaar zijn, zodat wanneer de werknemer tijd heeft, deze zelf kan bepalen wanneer hij de klant terugbelt. Dit is een variant op de eerste manier en heeft dan ook nagenoeg hetzelfde effect. Klanten zijn hierdoor tevreden en het ontstaan van pieken wordt zo in de hand gehouden.

7 Multiple Skills

Call Centers kunnen niet alleen groeien door uitbreiding van gebouwen en van gesprekken, maar ook door meerdere taken te vervullen. Dat wil zeggen dat het Call Center ook gesprekken aanneemt met een ander onderwerp dan voorheen of een gesprek die een extra vaardigheid vereist, zoals een extra taal. Aangezien er verschillende vaardigheden nodig zijn, kan het management er voor kiezen om mensen te specialiseren. Hier zijn verschillende redenen voor:

- Gespecialiseerde agenten werken efficiënter.
- Gespecialiseerde agenten hebben minder training nodig.

Echter zijn er ook nadelen. Gespecialiseerde agenten zijn vaak minder flexibel. Wanneer er grote fluctuaties zijn in binnenkomende gesprekken waarbij verschillende vaardigheden nodig zijn, is het nadelig om alleen maar gespecialiseerde agenten te hebben. Hier zou het beter zijn om multi-skilled agenten (*Generalisten*) te hebben, die daar kunnen helpen waar het druk is. Echter zijn generalisten in het algemeen weer minder efficiënt en het blijkt dan ook dat een combinatie van generalisten en specialisten het beste is voor de performance van een groot Call Center.

De verdeling van generalisten en specialisten kan bepaald worden door simulaties uit te voeren. Bij deze simulaties zijn verschillende parameters nodig. Deze parameters zijn niet identiek voor iedere agent. Zoals al eerder aangegeven is het belangrijk dat deze parameters goed geschat worden. Met deze parameters is het dan mogelijk calculators te gebruiken. Sommige van deze calculators zijn te vinden op internet. Hiermee kan op een eenvoudige manier het aantal specialisten en generalisten gevarieerd worden wanneer de overige parameters bekend zijn ([5] is hier een voorbeeld van).

8 Skilled based Routing

Het verdelen van binnenkomende gesprekken op gesprekniveau noemt men **Skilled-Based Routing** (SBR). In *SBR* zijn er twee verschillende beslissingen die genomen moeten worden.

- Wanneer een agent beschikbaar komt, welke type gesprek deze als eerste moet behandelen.
- Wanneer er een gesprek binnenkomt en er zijn meerdere verschillende agenten beschikbaar, welke agent het gesprek dan toegewezen moet krijgen.

De complexiteit van *SBR* kan heel erg verschillen tussen verschillende Call Centers. Hoe meer verschillende vaardigheden een Call Center behandelt, hoe complexer het probleem wordt. Aangezien de meeste Call Centers gesprekken binnen krijgen die verschillende vaardigheden nodig hebben, is het noodzakelijk om de gesprekken te kunnen identificeren. Dit kan het Call Center bijvoorbeeld doen door verschillende telefoonnummers aan verschillende vaardigheden toe te kennen.

Als de identificatie gebeurd is, zal het Call Center moeten bepalen welke werknemers de gesprekken gaan beantwoorden. Er moet besloten worden naar welke **Skill-groep** (team van werknemers die beschikken over dezelfde vaardigheden) het gesprek doorgestuurd dient te worden, zodat de klant juist behandeld kan worden. Het kiezen van deze skill-groepen en het verdelen van de werknemers over deze skill-groepen hoort allemaal tot *SBR*. Vragen waar *SBR* mee te maken heeft zijn bijvoorbeeld: Hoeveel werknemers heeft de ene skill-groep nodig om een bepaald Service Level te halen, hoeveel specialisten moeten er getraind worden, etc.

Daarnaast kan het in een Call Center voorkomen dat sommige klanten voorrang krijgen boven andere klanten. Het kan voorkomen dat sommige gesprekken over een bepaald type onderwerp voorrang dienen te krijgen boven andere onderwerpen. Het omgaan met deze prioriteiten kan op verschillende manieren:

1. Er kan voor gekozen worden om wanneer er een klant belt met hoge prioriteit, deze klant meteen te helpen door daarvoor een andere klant in wacht te zetten of door zelfs de verbinding te verbreken met deze klant (overall-priority)
2. Vermenigvuldig met verschillende factoren (al naar gelang een type gesprek belangrijker is een hogere factor) de types gesprekken. Kijk daarna naar het product van de wachttijd en de factor om te bepalen welk gesprek als eerst geholpen dient te worden, wanneer men streeft naar de wachttijden zo klein mogelijk hebben. De gesprekken hebben zo als het ware een gewogen prioriteit.
3. Het Call Center kan er voor kiezen een aantal werknemers te reserveren voor een hoge prioriteit. Deze wachten dan totdat er zo een gesprek binnenkomt, ongeacht het binnenkomen van gesprekken met lagere prioriteit.

De vraag is nu hoe veel efficiënter specialisten werken boven generalisten. Maakt het relatief niets uit als het Call Center uit specialisten bestaat, dan is het reserveren van werknemers voor bepaalde taken niet nodig.

Wanneer generalisten en specialisten even efficiënt werken, is het voordeliger om zoveel mogelijk werknemers generalisten te maken. Bij binnenkomende gesprekken is het dan het slimste om eerst deze gesprekken toe te wijzen aan specialisten, om ze daarna, wanneer er geen specialisten meer over zijn voor dat type gesprek, toe te kennen aan de generalisten. Generalisten kunnen dan als het ware de pieken van een bepaald type gesprek opvangen. In het onderzoek zal gekeken worden of deze verwachtingen ook daadwerkelijk waar zijn.

Referenties

- [1] Ger Koole, *Call Center Mathematics*, 26 januari 2007
- [2] M.J.Stegeman, M.H.Jansen-Vullers, *Determining Optimal Staffing Levels In Multiple Skill Inbound Call Centers*,
http://fp.tm.tue.nl/beta/publications/working%20papers/beta_wp160.pdf
- [3] http://en.wikipedia.org/wiki/Erlang_unit
- [4] <http://www.student.tue.nl/Q/l.d.jonge/modelleren3/index.htm>
- [5] <http://www.math.vu.nl/~koole/ccmath/sim2skill/>

Onderzoek

9 TUI

9.1 Algemeen

TUI Nederland maakt sinds december 2001 onderdeel uit van World of TUI, het overkoepelende begrip voor een verscheidenheid aan reisorganisaties van de toerisme-organisatie TUI AG. Een wereld waarin meer dan 600 vooraanstaande merken en ondernemingen uit zestien landen samenwerken onder één hoofd. Dit zorgt ervoor dat er veel deskundigheid en ervaring is, waardoor er meer mogelijkheden zijn om service te verlenen. Dit betekent een bundeling van deskundigheid en ervaring, meer mogelijkheden en service op maat.

TUI Nederland is op de volgende terreinen actief:

- Touroperating. Naast de merken Holland International, Arke en KRAS SterVakanties behoren er ook nog een groot aantal kleinere merken tot TUI Nederland.
- Toeristische retail. De retailgroep bestaat uit ruim tweehonderd reisbureaus: reisbureaus van Arke en vestigingen van Holland International Reisbureau.
- Internet. Veel reizen worden aangeboden via het internet.
- Luchtvaart. Sinds april 2005 heeft TUI Nederland haar eigen airline, Arkefly.

9.2 Call Center

Reizen die niet via internet geboekt worden, worden in de meeste gevallen geboekt met de telefoon. Daarom heeft TUI ook een Call-Center waarin alle gesprekken binnenkomen. Aangezien het marktaandeel van TUI groot is en daarmee het aantal telefoongesprekken, is het belangrijk dat gesprekken die bij het Call Center binnenkomen, zo vlug en goed mogelijk behandeld worden. Er zijn twee verschillende gesprekken die binnenkomen, te weten 'Business-to-Business' (BTB) en 'Business-to-Customer' (BTC). De eerste zijn gesprekken die ontstaan wanneer een klant naar een reisbureau gaat om een reis te boeken. Dit bureau belt dan zelf naar TUI om de betreffende reis te boeken. Ongeveer 85% van de gesprekken zijn BTB-gesprekken. De overige 15% van de gesprekken zijn de gesprekken waarbij de klant rechtstreeks belt naar TUI.

Vakanties zijn in verschillende groepen in te delen. Voorbeelden van groepen zijn:

1. Zon (vakanties naar de Middellandse Zee).
2. Overland (vakanties naar plaatsen waar met de auto naar toe gereisd wordt).
3. City (vakanties naar steden).
4. Verre vakanties - Caribisch gebied.

5. Verre vakanties - Azië/Afrika.

6. Verre vakanties - USA/Canada.

In het Call Center zijn verschillende skill-groepen. Met de bovengenoemde nummering zijn een paar voorbeelden hiervan: 1, 2, 3, 1-2, 1-4-5-6, etc. Iedere agent die behoort tot één van deze groepen behoort te beschikken over de vaardigheden die de skill-groep verlangt. Enkele vaardigheden zijn:

- Kennis van hotels. Welke hotels heeft het betreffende land allemaal? Hierbij moeten de agenten bijvoorbeeld weten welke hotels voor jongeren zijn en welke voor gezinnen.
- Kennis van vervoersmogelijkheden. Hierbij moeten de agenten weten wat er allemaal voor mogelijkheden zijn wat betreft het vervoer.
- Bezienswaardigheden.

9.3 Onderzoeksopdracht

Op dit moment is een veelgebruikte formule in het Call Center de 'Erlang C'-formule. Deze gaat uit van het $M/M/s$ -model die beschreven is in sectie 3. De formule is afzonderlijk op de verschillende skill-groepen toe te passen. Echter zijn er bij TUI ook generalisten die over meerdere vaardigheden beschikken. Het nadeel van dit model is dat het geen rekening houdt met deze generalisten. Generalisten zijn flexibeler dan specialisten en kunnen dan ook op plaatsen ingezet worden waar het op dat moment druk is. Dit kan tot een betere performance van het Call Center leiden.

Als contactpersoon binnen TUI is de heer Ronald Kupsch benaderd. Deze gaf aan dat een van de zaken waar hij in geïnteresseerd is, het ontwikkelen van een calculator is. Vanwege de verschillende skill-groepen is het van belang dat er een goede balans is in de vraag naar een bepaalde skill door de klanten en het aanbod van deze skill door het Call Center. Er moet namelijk voorkomen worden dat er te veel gesprekken afhaken. Om te weten te komen hoeveel agenten er nodig zijn voor de verschillende skills is het ontwikkelen van een calculator belangrijk. Met deze calculator is het makkelijk om te kijken wat voor gevolgen het veranderen van de parameters (λ , μ en s) heeft op bijvoorbeeld het Service-Level. Deze calculator zal dan ook rekening moeten houden met de verschillende agenten die TUI in het bedrijf heeft.

Het ontwikkelen van de calculator is niet binnen de gegeven tijd te halen. Daarom is besloten dat het ontwikkelen van deze calculator niet meer tot het onderzoek behoort.

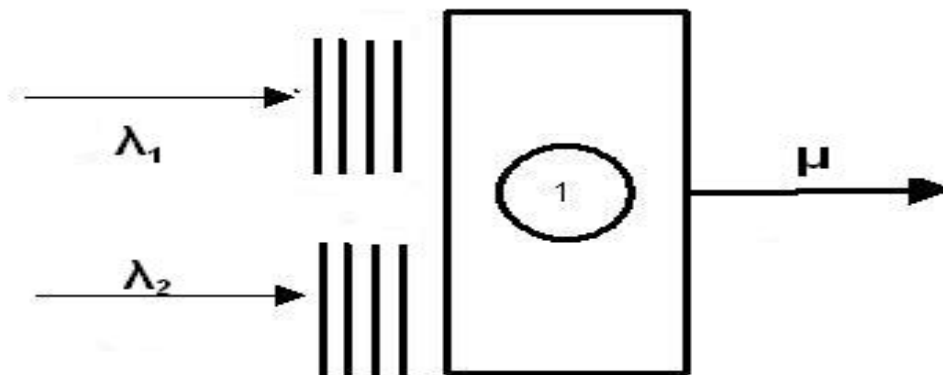
Allereerst zal gekeken worden naar wat voor gevolgen het heeft om een model te gebruiken waarin een prioriteitsregel wordt gehanteerd. Hierbij zal er eerst analytisch naar gekeken worden, vervolgens zullen er resultaten volgen van simulaties die gemaakt zijn. Vervolgens zal het *slow-server*-probleem behandeld worden. Hierbij is onderzocht onder welke voorwaarde het verstandiger is om een 'langzame' agent in te zetten.

Ten slotte is er gekeken naar een Call Center met 10 agenten. De vraag die hierbij gesteld wordt is wat de ideale samenstelling van specialisten en generalisten is bij verschillende bedieningstijden van de generalist. Er zijn drie gevallen onderzocht met behulp van simulaties in het programma *Arena*. De resultaten van deze simulatie zijn gegeven en er zijn conclusies aan verbonden.

10 Prioriteitsmodellen

Bij een Call Center kan het voorkomen dat sommige klanten belangrijker zijn dan andere klanten. Het management van het Call Center zou er dan voor kunnen kiezen om deze klanten voorrang te geven boven andere klanten. Wat voor gevolgen dit heeft voor het Call Center is voor het management belangrijk om te weten. Vragen die dan gesteld kunnen worden zijn: Wanneer heeft het zin om klanten voorrang te geven, wat voor gevolgen heeft dit op de wachttijd van de minder belangrijke klanten, etcetera. Om hier wat over te kunnen zeggen, zal er eerst een model gemaakt moeten worden van het probleem.

Allereerst zal het model besproken worden met daarin twee verschillende typen gesprekken en één agent. Dit model is een variant op het $M/M/1$ -model en in figuur 7 is te zien hoe dit model er schematisch uit ziet. Dit model kan op verschillende manieren ingevuld



Figuur 7: Schematische weergave van model met prioriteit

worden. Allereerst zal in sectie 10.1 gekeken worden naar een model waarin type 1-klanten overall-prioriteit hebben op type 2-klanten. Dit houdt in dat, wanneer er een type 1-klant het systeem binnenkomt en er is geen wachtrij voor dit type klant, de type 1-klant meteen

geholpen zal worden, ongeacht of er een type 2-klant behandeld wordt of niet. In sectie 10.2 zal het model beschreven worden dat type 1-klanten moeten wachten totdat het systeem vrij is. Het verschil met het andere model is dat als er een type 2-klant geholpen wordt, deze eerst helemaal behandeld zal worden. Daarna zullen de type 1-klanten weer behandeld worden, totdat deze er niet meer zijn.

10.1 Overall prioriteit type 1

10.1.1 Karakteristieken

De volgende aannames worden gemaakt:

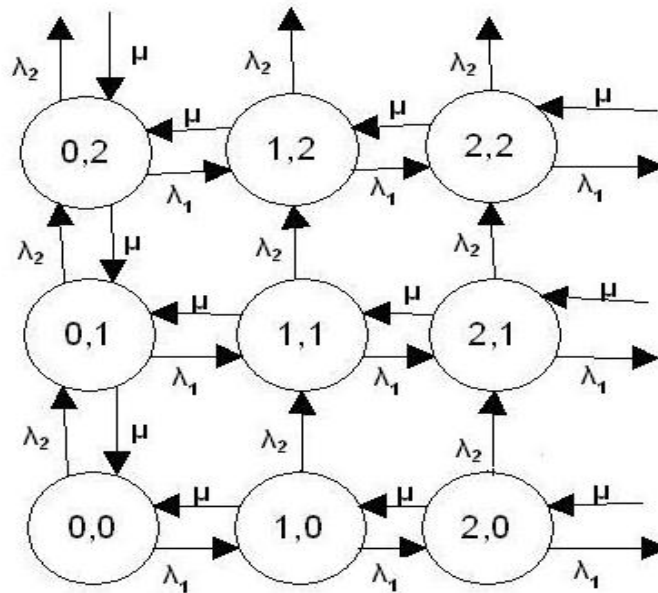
- Aankomsten van de gesprekken volgen een Poisson-Proces met intensiteit λ_1 en λ_2 voor respectievelijk type 1 en type 2 gesprekken.
- Het systeem heeft één agent die beide typen gesprekken behandelt.
- Type 1 gesprekken hebben een hogere prioriteit dan type 2 gesprekken. Dit houdt in dat zodra er een type 1 gesprek binnenkomt, het type 2 gesprek wordt onderbroken en in de wachtrij gezet wordt en er wordt overgegaan op behandeling van het type 1 gesprek.
- De behandelingstijd van de agent is exponentieel verdeeld met intensiteit μ voor beide typen gesprekken.
- Klanten hangen niet eerder op. Als een klant in de wachtrij komt, wacht deze totdat deze behandeld wordt.
- Als een klant behandeld is, verlaat deze klant de server en daarmee het systeem.
- Voor ieder type gesprek is er een aparte wachtrij. Klanten die binnenkomen en die zien dat de server bezet is, sluiten in de juiste rij aan.
- $\lambda_1 + \lambda_2 < \mu$.

De volgende notaties worden gebruikt:

- $L_{q,i}$ verwachte aantal type i-klanten in de wachtrij, L_q het verwachte aantal klanten in de totale wachtrij.
- L_i verwachte aantal type i-klanten in het systeem.
- $W_{q,i}$ verwachte tijd dat arriverende type i-klant in de wachtrij is.
- $\rho = \rho_1 + \rho_2 = \frac{\lambda_1}{\mu} + \frac{\lambda_2}{\mu}$ fractie van de tijd dat er een willekeurige klant geholpen wordt.

Net zoals voor het $M/M/s$ -model geldt, is hier ook sprake van een geboorte-sterfteproces. Er worden alleen maar sprongen gemaakt grootte één. In figuur 8 is schematisch weergegeven hoe de sprongen genomen worden en wat de intensiteit is waarmee dit gebeurt.

Met toestand (i, j) wordt hier bedoeld dat er i mensen van type 1 en j mensen van type



Figuur 8: Toestanden bij model van prioriteit

2 in het systeem zitten. Wanneer het systeem zich in toestand $(0,0)$ bevindt, is de enige gebeurtenis die plaats kan vinden een aankomst. Deze aankomst kan zowel van type 1 met intensiteit λ_1 of een aankomst van type 2 met intensiteit λ_2 zijn.

Figuur 8 geeft een overzicht van de sprongen en vergroot hierdoor het inzicht in het probleem. Voor het berekenen van de wachttijden zijn deze limietkansen echter niet nodig en het wordt dan ook niet uitgewerkt, zoals dit bijvoorbeeld wel bij het $M/M/s$ -model gedaan is.

10.1.2 Wachttijden met overall prioriteit

Type 1-klienten

Voor de prioriteit 1-klienten (klienten met een type 1 gesprek) is de gemiddelde wachttijd makkelijk te bepalen. Er moet hierbij in acht genomen worden dat type 1-klienten niet merken of er nu type 2-klienten in het systeem zijn of niet. Aangezien type 1-klienten geholpen worden als er geen type 1-klienten meer in de wachtrij voor deze zijn, ongeacht of er nu type 2-klienten zijn of niet, is in dit geval het model gelijk aan het $M/M/1$ -model. De type 2-klienten kunnen als het ware verwaarloosd worden wanneer er gekeken wordt naar de wachttijd van de type 1-klienten.

De wachttijd is te bepalen door het resultaat uit sectie 4 te gebruiken. Daar is een formule voor het gemiddeld aantal klanten in het systeem afgeleid. Aangezien hier de wachttijd in de wachtrij onderzocht wordt, is $L_{q,1}$ nodig:

$$L_{q,1} = L_1 - \frac{\lambda_1}{\mu} = \frac{\lambda_1}{\mu - \lambda_1} - \frac{\lambda_1}{\mu} = \frac{\lambda_1^2}{\mu(\mu - \lambda_1)} = \frac{\rho_1^2}{1 - \rho_1}$$

$W_{q,1}$ wordt dan:

$$W_{q,1} = \frac{L_{q,1}}{\lambda_1} = \frac{\frac{\rho_1^2}{1 - \rho_1}}{\lambda_1} = \frac{\rho_1^2}{\lambda_1(1 - \rho_1)}$$

Type 2-klanten

Om de wachttijd van een arriverende type 2-klant te bepalen, zal gebruik gemaakt worden van het volgende inzicht. Wanneer er gekeken wordt naar het totale systeem en dan naar de agent die de gesprekken moet behandelen, valt het volgende op. Aangezien de behandelingstijden van beide type gesprekken hetzelfde zijn, maakt het voor de agent niet uit welk type gesprek deze behandelt. Een ander gevolg hiervan is dat de agent eigenlijk maar één wachtrij voor zich ziet. Alleen maar naar de agent gekeken kan het model dus ook gezien worden als een $M/M/1$ -model.

De aankomstintensiteit is hierin gelijk aan $\lambda = \lambda_1 + \lambda_2$ en de vertrekintensiteit is gelijk aan μ . L_q , het verwachte totaal aantal klanten in de wachtrij, is gelijk aan het verwachte aantal klanten in de wachtrij voor type 1-klanten opgeteld bij het verwachte aantal klanten in de wachtrij voor type 2-klanten. Dit ziet er in formulevorm als volgt uit:

$$L_q = L_{q,1} + L_{q,2} \quad (12)$$

In sectie 4 is de formule al afgeleid voor het verwachte totaal aantal klanten in de wachtrij van het systeem. Door het resultaat uit die sectie te gebruiken volgt voor $L_{q,2}$:

$$L_{q,2} = \frac{\rho^2}{1 - \rho} - L_{q,1} \quad (13)$$

Door nu de 'Wet van Little' toe te passen op de type 2-klanten en door vergelijking (13) te gebruiken, wordt het volgende resultaat verkregen (met $L_{q,1} = \frac{\rho_1^2}{1 - \rho_1}$):

$$W_{q,2} = \frac{L_{q,2}}{\lambda_2} = \frac{\rho^2 - (1 - \rho)L_{q,1}}{\lambda_2(1 - \rho)} \quad (14)$$

10.1.3 Simulatie in Arena

Om te kijken of de analyses gemaakt in de vorige paragraaf kloppen, is in het programma *Arena* het model geïmplementeerd. In bijlage 13.1.1 is een plaatje van dit model in het programma te zien.

Met dit model is gekeken naar de gemiddelde wachttijd van de twee verschillende klanten. Hierbij is de aankomstintensiteit van beide klanten gelijk genomen en constant gehouden. Verder heeft de agent voor beide gesprekken eenzelfde bedieningstijd.

In het programma zijn de volgende twee modellen tegenover elkaar gezet: het model met de prioriteitsregel met overall-prioriteit en het model waarin klanten worden geholpen met de regel 'First-Come-First-Served' (FCFS). Dit betekent dat klanten worden behandeld in de volgorde waarin deze aankomen.

Variëren van bedieningstijd agent

λ_1	λ_2	μ		W_{gem}	Half-width	Analyse-resultaat
5	5	10	klant 1 prior	0,1000	0,001188	-
			klant 2 prior	71,9971	Correlated	-
			klant 1 FCFS	37,7159	Correlated	-
			klant 2 FCFS	37,6604	Correlated	-
5	5	12	klant 1 prior	0,0598	0,000697	0,0595
			klant 2 prior	0,6982	0,008749	0,7738
			klant 1 FCFS	0,4190	0,002532	0,4167
			klant 2 FCFS	0,4182	0,002556	0,4167
5	5	14	klant 1 prior	0,0395	0,000555	0,0397
			klant 2 prior	0,2641	0,005405	0,3175
			klant 1 FCFS	0,1793	0,002996	0,1786
			klant 2 FCFS	0,1791	0,003388	0,1786
5	5	17	klant 1 prior	0,0242	0,000266	0,0245
			klant 2 prior	0,1063	0,001260	0,14356
			klant 1 FCFS	0,0846	0,000915	0,0840
			klant 2 FCFS	0,0843	0,000986	0,0840
5	5	20	klant 1 prior	0,0168	0,000258	0,0167
			klant 2 prior	0,0695	0,000974	0,0833
			klant 1 FCFS	0,0499	0,000580	0,05
			klant 2 FCFS	0,0501	0,000596	0,05

Tabel 1. Gemiddelde wachttijden bij model met prioriteit en FCFS, variëren van μ

De volgende zaken vallen op wanneer gekeken wordt naar tabel 1:

- Wanneer er meer mensen binnenkomen dan dat de agenten aan kunnen, explodeert de wachttijd van type-2 klanten. Dit is te zien in de simulatie waarbij $\mu=10$. Daarom is er voor gekozen om μ groter te kiezen dan de som van de twee aankomstintensiteiten bij de volgende simulaties (één van de aannames aan het begin van deze sectie). Een Call Center zal dan ook altijd moeten zorgen dat de behandelingsnelheid van alle agenten groter is dan de aankomstintensiteit (volgt ook uit sectie 10.2.2)
- Type 1-klanten die de prioriteit hebben boven de type 2-klanten hebben bij iedere simulatie een veel kleinere gemiddelde wachttijd. Bij $\mu=12$ is dit verschil ongeveer het twaalfvoudige, bij $\mu=20$ is dit nog ongeveer het vijfvoudige.
- De gemiddelde wachttijd van type 1-klanten, die volgens FCFS-regel behandeld zijn, verschilt aanzienlijk met de gemiddelde wachttijd van de andere type 1-klanten. Bij $\mu=12$ is dit ongeveer het achtvoudige, bij $\mu=20$ nog ongeveer het drievoudige.
- Het effect op de gemiddelde wachttijd van type 2-klanten is voor alle waarden van μ groot. Bij $\mu=12$ is het ongeveer anderhalf keer zoveel ten opzichte van het systeem waarin de FCFS-regel toegepast wordt, bij $\mu=20$ is de wachttijd ongeveer met $2/5$ van de andere wachttijd vergroot.
- Het effect op de gemiddelde wachttijd van de type 1-klanten wordt kleiner naarmate μ groter wordt. Bij $\mu=20$ is dit effect een reductie van ongeveer 70%, bij $\mu=12$ levert het een reductie van 85% op.
Een van de oorzaken hiervan is het feit dat beide klanten veel vlugger behandeld worden dan dat ze aankomen. Het gevolg voor het systeem is dan dat de agent in de meeste gevallen geen gesprek behandelt en daardoor ontstaat er geen wachtrij, waardoor de prioriteit ook niet veel effect heeft.
- De gemiddelde wachttijd van zowel type 1-klanten als type 2-klanten bij het model met FCFS-regel, zijn aan elkaar gelijk. Dit volgt ook uit de definitie van de regel en wordt ook versterkt door dit resultaat.
- De analyseresultaten geven een hogere gemiddelde wachttijd bij de type 2-klanten in het model met de prioriteitsregel. Dit komt door de implementatie in *Arena*. Wanneer een type 2-klant in het proces-module is, kan deze niet eruit gehaald worden op het moment dat er een type 1-klant bij de agent komt. Op het moment dat de proces-module klaar is met de type 2-klant en deze ziet dat er nog een type 1-klant in de module zit, gaat deze type 2-klant naar een aparte wachtrij. Deze wachtrij verlaat hij weer op het moment dat de type 1-klant behandeld is door de agent. De tijd dat de type 2-klant teveel in de proces-module zit, wordt niet meegeteld als wachttijd.

De conclusie die volgt uit de simulatie, waarbij zowel λ als μ voor beide klanten even groot is, is dat het effect op de gemiddelde wachttijd aanzienlijk is. Klanten die een hoge urgentie hebben kunnen als het systeem gebruik maakt van de prioriteitsregel veel sneller behandeld worden en zo een veel lagere wachttijd hebben. Echter is het effect op de klanten die een lage prioriteit hebben ook groot. Naarmate μ groter wordt, zal het effect van de prioriteitsregel afnemen en zal het rendement kleiner worden.

Variëren in aankomstintensiteiten

In de volgende simulatie worden de aankomstintensiteiten gevarieerd. De bedieningstijd van de agent is constant genomen en de aankomstintensiteiten zijn zo gekozen dat $\lambda_1 + \lambda_2 = \lambda$ constant is. Dit heeft tot gevolg dat de bezettingsgraad bij deze simulatie constant gehouden wordt.

Verder is ervoor gekozen om de som van de aankomstintensiteiten kleiner te laten zijn dan de vertrekintensiteit. Dit omdat er anders een oneindig lange wachtrij zou kunnen ontstaan en aangenomen is dat een Call Center dat probeert te voorkomen door genoeg agenten in te zetten.

Overigens wordt er in de meeste gevallen prioriteit gegeven aan een kleine groep mensen. Er zullen in het algemeen meer 'gewone' klanten zijn dan dat er klanten zijn die hoge prioriteit moeten krijgen. Het resultaat is hieronder te zien in Tabel 2.

λ_1	λ_2	μ		W_{gem}	Half-width	Analyse-resultaat
1	6	10	klant 1 prior	0,0118	0,000419	0,0111
			klant 2 prior	0,2653	0,005150	0,2704
			klant 1 FCFS	0,2332	0,004648	0,2333
			klant 2 FCFS	0,2326	0,004936	0,2333
2	5	10	klant 1 prior	0,0250	0,000367	0,0250
			klant 2 prior	0,2991	0,005504	0,3167
			klant 1 FCFS	0,2333	0,002562	0,2333
			klant 2 FCFS	0,2328	0,002781	0,2333
3	4	10	klant 1 prior	0,0428	0,000552	0,0429
			klant 2 prior	0,3482	0,006008	0,37619
			klant 1 FCFS	0,2304	0,003371	0,2333
			klant 2 FCFS	0,2305	0,002949	0,2333
4	3	10	klant 1 prior	0,0672	0,00112	0,0667
			klant 2 prior	0,4172	0,010791	0,4556
			klant 1 FCFS	0,2345	0,005558	0,2333
			klant 2 FCFS	0,2336	0,004056	0,2333

Tabel 2. Gemiddelde wachttijden bij model met prioriteit en FCFS, variëren van λ_1 en λ_2 bij constante ρ

De volgende zaken vallen op wanneer gekeken wordt naar tabel 2.

- Het effect op de gemiddelde wachttijd van type 1-klienten is groot. Bij $\lambda_1=1$ is het een vermindering van 90%, bij $\lambda_1=4$ is dit 75% ten opzicht van de gemiddelde wachttijd die het zou zijn bij het *FCFS*-model.
- Het effect op de gemiddelde wachttijd van type 2-klienten is bij kleine λ_1 klein, maar wordt groter naarmate λ_1 groter wordt. Voor $\lambda_1=4$ wordt de gemiddelde wachttijd het dubbele van de gemiddelde wachttijd bij het *FCFS*-model.
- De gemiddelde wachttijden van de klienten bij het *FCFS*-model zijn nagenoeg gelijk bij iedere simulatie. Dat volgt uit de definitie van dit model en uit het feit dat de totale aankomstintensiteit bij de simulaties constant gehouden is.
- De analyseresultaten geven een hogere gemiddelde wachttijd bij de type 2-klienten in het model met de prioriteitsregel dan dat uit de analyse blijkt. Dit komt door de implementatie in *Arena*. Wanneer een type 2-klient in het proces-module is, kan deze niet eruit gehaald worden op het moment dat er een type 1-klient bij de agent komt. Anders zou de agent op dat moment met twee klienten bezig zijn. Op het moment dat de proces-module klaar is met de type 2-klient en deze ziet dat er nog een type 1-klient in de module zit, gaat de type 2-klient naar een aparte wachtrij in het submodel van de proces-module. Deze wachtrij verlaat hij weer op het moment dat de type 1-klient behandeld is door de agent. De tijd dat de type 2-klient teveel in de proces-module zit doordat hij behandeld wordt op het moment dat de agent in werkelijkheid bezig zou zijn met een type 1-klient, wordt niet meegeteld als wachttijd.

De conclusie die volgt uit de simulatie is dat voor kleine groepen mensen het goed kan zijn om prioriteit te geven. Deze 'belangrijke' mensen hebben dan een veel kleinere gemiddelde wachttijd en de toename van de gemiddelde wachttijd van de grote groep is bijna te verwaarlozen. Als er teveel 'belangrijke' mensen komen, is het effect op de gemiddelde wachttijd groot, positief voor de grote groep maar negatief voor de andere groep. Het Call Center zal dus moeten bekijken hoe groot de groep is die prioriteit krijgt.

10.2 Prioriteit type 1 zonder onderbreken

10.2.1 Karakteristieken

De volgende aannames worden gemaakt:

- Aankomsten van de type 1-gesprekken volgen een Poisson-Proces met intensiteit λ_1 en de type 2-gesprekken volgen een Poission-Proces met intensiteit λ_2 .
- Het systeem heeft één agent die beide typen gesprekken behandelt.
- Type 1-gesprekken hebben een hogere prioriteit dan type 2-gesprekken. Het verschil met de overall-prioriteit is dat gesprekken van type 2-klanten in dit geval niet onderbroken worden. De agent behandelt dus eerst het type 2-gesprek af voordat deze aan het behandelen van type 1-gesprekken overgaat.
- De behandelingstijd van de agent is exponentieel verdeeld met intensiteit μ voor beide typen gesprekken.
- Klanten hangen niet eerder op. Als een klant in de wachtrij komt, wacht deze totdat hij behandeld wordt.
- Als een klant behandeld is, verlaat deze klant de server en daarmee het systeem.
- Voor ieder type gesprek is er een aparte wachtrij. Klanten die binnenkomen en die zien dat de server bezet is, sluiten in de juiste rij aan.
- $\rho < 1$.

De volgende notaties worden gebruikt:

- $L_{q,i}$ verwachte aantal type i-klanten in wachtrij, L_q het verwachte aantal klanten in de wachtrij van het totale systeem.
- $W_{q,i}$ verwachte wachttijd van arriverende type i-klant.
- $\rho = \rho_1 + \rho_2 = \frac{\lambda_1}{\mu} + \frac{\lambda_2}{\mu}$ fractie van de tijd dat er een willekeurige klant geholpen wordt.

Het maken van een overzicht met daarin de sprongen en daarbij de intensiteiten van deze sprongen, is in dit model moeilijker. Het is namelijk afhankelijk van het type gesprek dat behandeld wordt op dat moment. Het berekenen van de limietkansen is bij dit model niet nodig. Met inzichten die duidelijk zullen worden in de volgende secties is het bepalen van de gemiddelde wachttijd veel vlugger.

10.2.2 Wachttijden met prioriteit zonder onderbreken

Type 1-klanten

De wachttijd van een type 1-klant, die arriveert bij het systeem, bestaat uit de tijd dat het duurt om de wachtrij van type 1-klanten tot nul te reduceren opgeteld bij de servicetijd van de klant die op dat moment behandeld wordt. Dit ziet er in formulevorm als volgt uit:

$$W_{q,1} = E[x] + E[z] \quad (15)$$

In vergelijking (15) is:

$W_{q,1}$ de verwachte wachttijd in de wachtrij van arriverende type 1-klant
 x bedieningstijd van totaal aantal type 1-klanten in wachtrij
 z wachttijd opgelopen doordat er een klant in behandeling is

Het verwachte aantal klanten van type 1 in de wachtrij is gedefinieerd als $L_{q,1}$. Dewerkingstijden van deze klanten zijn onafhankelijk en identiek verdeeld met gemiddelde $\frac{1}{\mu}$. Daarom geldt:

$$E[x] = L_{q,1} \frac{1}{\mu}$$

Aangenomen is dat de behandelingstijden van beide typen gesprekken exponentieel verdeeld zijn met gemiddelde $\frac{1}{\mu}$. De verwachte wachttijd, die opgelopen wordt doordat er een klant in behandeling is, is gelijk aan het product van de kans dat er een klant behandeld wordt met de verwachte behandelingstijd van die klant. Dit ziet er in formulevorm als volgt uit:

$$E[z] = \frac{\lambda_1 + \lambda_2}{\mu} \frac{1}{\mu} = \frac{\rho}{\mu}$$

Vergelijking (15) wordt dan:

$$W_{q,1} = L_{q,1} \frac{1}{\mu} + \frac{\rho}{\mu} \quad (16)$$

Verder geldt in dit model de 'Wet van Little'. In sectie 4 is deze al besproken. Echter zal deze formule nu in de volgende vorm gebruikt worden:

$$L_{q,i} = \lambda_i W_{q,i} \quad (17)$$

Aangezien er onderzocht wordt wat de wachttijd is, zal vergelijking (17) gesubstitueerd worden in vergelijking (16). Deze formule is dan te schrijven als:

$$W_{q,1} = \frac{\lambda_1 + \lambda_2}{\mu(\mu - \lambda_1)} = \frac{\rho}{\mu - \lambda_1} \quad (18)$$

Type-2 klanten

Om de wachttijd van een arriverende type 2-klant te bepalen zal gebruik gemaakt worden van hetzelfde inzicht als dat in sectie 10.1.2 gebruikt is. Alleen maar naar de agent gekeken kan het model gezien worden als een $M/M/1$ -model met aankomstintensiteit λ ($= \lambda_1 + \lambda_2$) en vertrekintensiteit μ . L_q , het verwachte aantal klanten in de wachtrij, is gelijk aan het verwachte aantal klanten in de wachtrij voor type 1-klanten opgeteld bij het verwachte aantal klanten in de wachtrij voor type 2-klanten. Dit ziet er in formulevorm als volgt uit:

$$L_q = L_{q,1} + L_{q,2} \quad (19)$$

In sectie 4 is de formule afgeleid voor het gemiddeld aantal klanten in het systeem. Aangezien hier L_q nodig is, zal van dat resultaat eerst het gemiddeld aantal klanten in bediening afgehaald moeten worden:

$$L_q = \frac{\rho}{1 - \rho} - \rho = \frac{\rho^2}{1 - \rho}$$

$L_{q,2}$ wordt nu:

$$L_{q,2} = \frac{\rho^2}{1 - \rho} - L_{q,1} \quad (20)$$

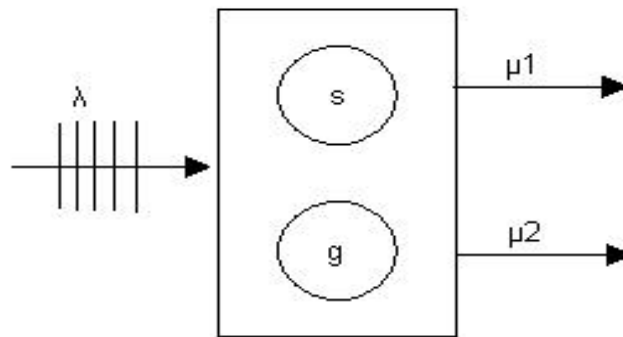
Door nu de 'Wet van Little' toe te passen op de type 2-klanten en door vergelijking (20) te gebruiken, wordt het volgende resultaat verkregen (met $L_{q,1} = \frac{\rho\lambda_1}{\mu - \lambda_1}$):

$$W_{q,2} = \frac{L_{q,2}}{\lambda_2} = \frac{\rho^2 - (1 - \rho)L_{q,1}}{(1 - \rho)\lambda_2} \quad (21)$$

11 Slow-server queue

Generalisten in een Call Center zijn agenten die op meerdere typen gesprekken inzetbaar zijn. Het inzetten van veel generalisten kan slim zijn in een Call Center. Flexibele werknemers zijn makkelijk omdat ze op veel gebieden inzetbaar zijn. Er zijn minder agenten nodig door deze generalisten. Echter zijn generalisten vaak langzamer dan specialisten en of de prestatie van het Call Center verbeterd is de vraag. Daarom zal eerst onderzocht worden wanneer het zinvol is om een agent tot generalist op te leiden, waar het anders een specialist wordt.

Eerst zal gekeken worden naar het model waarbij er één type klant is. Er zijn twee verschillende agenten: een snelle (de specialist) en een langzame agent (de generalist). Dit model wordt ook wel de *Slow-server queue* genoemd. Het model is in figuur 9 te zien:



Figuur 9: Schematische weergave van de slow-server queue

De volgende aannames worden gemaakt:

- Er werken twee agenten in het Call Center.
- Gesprekken komen aan via aan Poisson-Proces met intensiteit λ . Dat betekent dat de tussenaankomsttijden onafhankelijk en exponentieel verdeeld zijn met parameter λ .
- Er is maar één wachtrij voor de klanten.
- De specialist heeft een servicetijd die exponentieel verdeeld is met parameter μ_1 .
- De generalist heeft een servicetijd die exponentieel verdeeld is met parameter μ_2 .
- Als beide agenten vrij zijn gaat de klant naar de specialist.

- Klanten hangen niet eerder op. Als een klant in de wachtrij zit, zal deze de server niet meer verlaten alvorens hij geholpen is.
- $\mu_1 \geq \mu_2$
- $\lambda < \mu_1 + \mu_2$

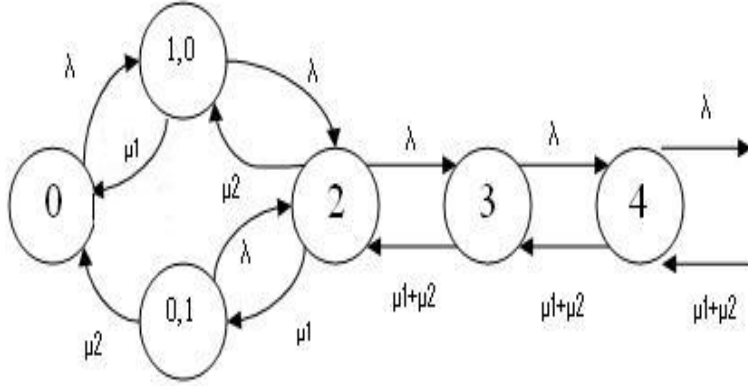
De volgende notaties worden gebruikt:

- $\mu = \mu_1 + \mu_2$
- $\rho = \lambda / (\mu_1 + \mu_2) = \lambda / \mu$
- L_{gen} het verwachte aantal klanten in het systeem.
- P_i fractie van de tijd dat er i klanten in het systeem zitten.
- $P_1 = P_{0,1} + P_{1,0}$, waarbij $P_{1,0}$ wil zeggen dat er een klant bij de specialist zit en niemand bij de generalist ($P_{0,1}$ vice versa).

De vraag is wat de klant, die als eerste in de wachtrij staat, moet doen wanneer de specialist al bezig is met een klant. De klant kan wachten totdat de specialist klaar is of hij kan naar de generalist gaan. De verwachting is dat als de verwachte wachttijd korter is dan het verschil in behandelen tussen de generalist en specialist, het verstandiger voor de klant is om te wachten. De onderzoeksvraag wordt: 'Onder welke condities moet de generalist weggelaten worden uit het systeem?'. Een constatering hierbij is dat wanneer de generalist weggelaten wordt, het systeem de klanten behandelt volgens het $M/M/1$ -model.

Als criterium zal gekeken worden naar het verwachte aantal klanten in het systeem. Is het namelijk zo dat de verwachte wachttijd van een klant in het beschreven model kleiner is dan bij het $M/M/1$ -model, dan is het verstandiger om de generalist uit het systeem te laten. De *Wet van Little* geeft dan, aangezien de aankomstintensiteit gelijk is bij beide modellen, dat het verwachte aantal klanten in het systeem kleiner is.

Om L_{gen} te bepalen zullen eerst de toestanden gedefinieerd worden. In figuur 10 zijn de toestanden en de overgangsintensiteiten weergegeven.



Figuur 10: Toestanden bij slow-server queue

De balansvergelijkingen bij het model zijn:

$$\lambda P_0 = \mu_1 P_{1,0} + \mu_2 P_{0,1} \quad (22)$$

$$(\lambda + \mu_1) P_{1,0} = \mu_2 P_2 + \lambda P_0 \quad (23)$$

$$(\lambda + \mu_2) P_{0,1} = \mu_1 P_2 \quad (24)$$

$$\sum_{i=0}^{\infty} P_i = 1 \quad (25)$$

$$P_n = \rho^{n-1} P_1 \text{ voor } n \geq 1 \quad (26)$$

Vergelijking (25) werkt het beste wanneer alle termen in termen van één toestand geschreven zijn. Gekozen is om P_i te schrijven in termen van P_0 voor alle i . Hiervoor moet P_1 in termen van P_0 geschreven worden, aangezien dan met vergelijking (26) alle termen te schrijven zijn in P_0 . Door gebruik te maken van het feit dat $P_1 = P_{0,1} + P_{1,0}$ volgt:

$$P_2 = \rho P_1 = \rho(P_{0,1} + P_{1,0}) \quad (27)$$

Door (27) te substitueren in vergelijking (25) kan $P_{0,1}$ geschreven worden in termen van $P_{1,0}$. Dat levert het volgende resultaat:

$$\begin{aligned} (\lambda + \mu_2) P_{0,1} &= \mu_1 P_2 = \mu_1 \rho (P_{0,1} + P_{1,0}) \\ P_{0,1} &= \frac{\mu_1 \rho P_{1,0}}{\lambda + \mu_2 - \mu_1 \rho} = \frac{\mu_1 \lambda P_{1,0}}{\mu \lambda + \mu \mu_2 - \mu_1 \lambda} = \frac{\mu_1 \lambda P_{1,0}}{\mu_2 \lambda + \mu \mu_2} \end{aligned} \quad (28)$$

Met behulp van de vergelijkingen (28) en (24) kan $P_{1,0}$ in termen van P_0 geschreven worden:

$$\begin{aligned}
(\lambda + \mu_1)P_{1,0} &= \lambda P_0 + \mu_2 \rho \left(P_{1,0} + \frac{\mu_1 \lambda P_{1,0}}{\mu_2 \lambda + \mu \mu_2} \right) = \lambda P_0 + \mu_2 \lambda P_{1,0} \left(\frac{\lambda + \mu_2}{\mu_2 \lambda + \mu \mu_2} \right) = \dots \\
&= \lambda P_0 + \lambda P_{1,0} \left(\frac{\lambda + \mu_2}{\lambda + \mu} \right) \\
\lambda P_0 &= P_{1,0} \left(\lambda + \mu_1 - \frac{\lambda^2 + \lambda \mu_2}{\lambda + \mu} \right) \\
&= P_{1,0} \left((\lambda^2 + 2\lambda \mu_1 + \lambda \mu_2 + \mu_1 \mu - \lambda^2 - \lambda \mu_2) / (\lambda + \mu) \right) \\
&= P_{1,0} \left(\frac{\mu_1 (2\lambda + \mu)}{\lambda + \mu} \right) \\
P_{1,0} &= \frac{\lambda (\lambda + \mu) P_0}{\mu_1 (2\lambda + \mu)} \tag{29}
\end{aligned}$$

Door vergelijking (29) te substitueren in vergelijking (28):

$$\begin{aligned}
P_{0,1} &= \frac{\mu_1 \lambda P_{1,0}}{\mu_2 \lambda + \mu \mu_2} = \frac{\mu_1 \lambda^2 (\lambda + \mu) P_0}{(\mu_2 \lambda + \mu \mu_2) (2\lambda \mu_1 + \mu_1 \mu)} \\
&= \frac{\mu_1 \lambda^2 (\lambda + \mu) P_0}{(\mu_1 \lambda + \mu \mu_1) (2\lambda \mu_2 + \mu_2 \mu)} \\
&= \frac{\lambda^2 P_0}{\mu_2 (2\lambda + \mu)}
\end{aligned}$$

De formule van P_1 is nu te verkrijgen door de laatste vergelijking op te tellen bij vergelijking (29):

$$\begin{aligned}
P_1 &= \frac{\lambda^2 P_0}{\mu_2 (2\lambda + \mu)} + \frac{(\lambda^2 + \lambda \mu) P_0}{\mu_1 (2\lambda + \mu)} = \dots = \frac{\mu \lambda (\lambda + \mu_2)}{\mu_1 \mu_2 (2\lambda + \mu)} P_0 = \frac{\lambda}{a} P_0 \tag{30} \\
a &= \frac{\mu_1 \mu_2 (2\lambda + \mu)}{\mu (\lambda + \mu_2)}
\end{aligned}$$

Door vergelijking (30) te substitueren in (25) is elke P_i geschreven in termen van P_0 :

$$P_n = \left(\frac{\lambda}{a} \right) \rho^{n-1} P_0 \quad \text{voor } n \geq 1 \tag{31}$$

P_0 is nu te bepalen door vergelijking (25) te gebruiken als volgt:

$$\begin{aligned}
\sum_{i=0}^{\infty} P_i &= 1 = P_0 \left(1 + \frac{\lambda}{a} + \frac{\lambda}{a} \rho + \frac{\lambda}{a} \rho^2 + \dots \right) = P_0 \left(1 + \frac{\lambda}{a} (1 + \rho + \rho^2 + \dots) \right) \\
P_0 &= \frac{1 - \rho}{1 - \rho + \frac{\lambda}{a}} \tag{32}
\end{aligned}$$

Nu voor alle toestanden bekend is wat de fractie van de tijd is dat het systeem zich in deze toestand bevindt, is het mogelijk om te bepalen wat het gemiddeld aantal klanten in het systeem is:

$$L_{gen} = \sum_{i=0}^{\infty} iP_i = \frac{\lambda}{a}P_0 \sum_{i=0}^{\infty} i\rho^{i-1} = \left(\frac{\lambda}{a}\right)P_0 \frac{1}{(1-\rho)^2} = \frac{\lambda}{(a(1-\rho) + \lambda)(1-\rho)} \quad (33)$$

Van zowel de *slow-server queue* als van het $M/M/1$ -model zijn vergelijkingen gevonden voor het gemiddeld aantal klanten in het systeem. Hieruit kan de voorwaarde bepaald worden wanneer de generalist wel dan niet ingezet moet worden.

De vergelijking die opgelost moet worden is in formulevorm:

$$L_{gen} = \frac{\lambda}{(a(1-\rho) + \lambda)(1-\rho)} > \frac{\lambda}{\mu_1 - \lambda} = L_{M/M/1} \quad (34)$$

Om de drempelwaarde uit te rekenen zal vergelijking (34) met een gelijkteken opgelost worden.

$$\begin{aligned} \frac{\lambda}{(a(1-\rho) + \lambda)(1-\rho)} &= \frac{\lambda}{\mu_1 - \lambda} \\ \mu_1 - \lambda &= \lambda - \lambda\rho + a - 2a\rho + a\rho^2 \\ -a\rho^2 + \rho(2a + \lambda) + \mu_1 - 2\lambda + a &= 0 \\ \rho^2(\mu - a) + 2\rho(a - \mu) + (\mu_1 - a) &= 0 \end{aligned}$$

Aangezien a afhankelijk is van λ en μ , is a ook afhankelijk van ρ . Om deze reden is het nodig om de vergelijking te herschrijven. Om dit te bewerkstelligen zal a ingevuld moeten worden in bovenstaande vergelijking en zal gebruik gemaakt worden van de ratio μ_2/μ_1 ($=r$).

Dit levert:

$$\rho^2(1 + r^2) - 2\rho(2 + r^2) - (2r - 1)(r + 1) = 0 \quad (35)$$

In sectie 13.3.1 in de bijlage is te zien hoe de afleiding van vergelijking (35) is.

De oplossing van deze kwadratische vergelijking is te verkrijgen door de ABC -formule toe te passen. De gevonden ρ_{\pm} zijn drempelwaarden voor het probleem. Als de bezettingsgraad van het Call Center gelijk is deze drempelwaarde, dan maakt het voor het Call Center niet uit of de langzame server weggehaald wordt (gekeken naar de gemiddelde wachttijd).

$$\rho_{\pm} = \frac{2 + r^2 \pm \sqrt{(2 + r^2)^2 + 4(1 + r^2)(2r - 1)(1 + r)}}{2(1 + r^2)} \quad (36)$$

Laat ρ_- de kleinste waarde en ρ_+ de grootste waarde. Dat de grenswaardes voor $0 \leq r \leq 1$ aangenomen worden is bewezen in sectie 13.3.2 van de bijlage. Dat ρ_+ niet voldoet aan de stabiliteitsvoorwaarde van ρ wordt bewezen in sectie 13.3.3 van de bijlage. ρ_+ zal buiten beschouwing gelaten worden.

De voorwaarden die bij deze formule van belang zijn:

1. $0 \leq r \leq 1$
2. $0 \leq \rho < 1$
3. $\frac{\lambda}{\mu_1} < 1 \Rightarrow \rho < \frac{1}{1+r}$

Voorwaarde 1 volgt uit het feit dat de behandelingsintensiteiten van de agenten niet negatief kunnen zijn en omdat aangenomen is dat $\mu_1 \geq \mu_2$. Een andere aanname die gemaakt is in het begin van deze sectie, is dat $\rho < 1$. Daarnaast kan λ ook niet negatief zijn. Hieruit volgt de tweede voorwaarde. Voorwaarde 3 volgt uit het feit dat er vergeleken wordt met het $M/M/1$ -model. Daar geldt als voorwaarde dat $\lambda < \mu_1$.

Nu de formules voor de drempelwaardes bekend zijn, moet gekeken worden voor welke waardes van ρ $L_{gen} > L_{M/M/1}$ geldt. Deze vergelijking is te herschrijven tot (in de afleiding in de bijlage in plaats van ' $=$ ' ' $>$ ' laten staan):

$$\rho^2(1 + r^2) - 2\rho(2 + r^2) - (2r - 1)(r + 1) > 0 \quad (37)$$

Voor verschillende waardes van r is de grenswaarde ρ_- berekend en deze is weergegeven in tabel 3:

r	ρ_-	$\frac{1}{1+r}$
0.0	1	0.909091
0.1	0.6503	0.90909
0.2	0.4615	0.83333
0.3	0.2938	0.76923
0.4	0.1402	0.71429
0.5	0	0.66667
0.6	-0.1264	0.625
0.7	-0.2389	0.58824
0.8	-0.3381	0.55556
0.9	-0.4247	0.52632
1.0	-0.5	0.5

Tabel 3. Voor verschillende waardes van r de bijbehorende grenswaarde ρ_- .

Voor $0 < r \leq 0.5$ geldt dat $0 \leq \rho_- < 1$ (zie bijlage sectie 13.3.4). De drempelwaarde ρ_+ valt buiten het toegestane gebied van ρ . Nu moet nog bepaald gaan worden voor welke waarde van ρ (' $<$ ' of ' $>$ ') de langzame server weggehaald moet worden in het geval van $0 < r \leq 0.5$. Een simpele calculatie in vergelijking (37) geeft dat voor $\rho < \rho_-$ er aan de vergelijking voldaan is en dus moet voor deze waardes de langzame server weggehaald worden.

Voor $0.5 < r \leq 1$ geldt $\rho_- < 0$ en $\rho_+ > 1$. Dit betekent dat de langzame server niet weggehaald moet worden voor $0 < \rho < 1$ als $0.5 < r \leq 1$.

Conclusie

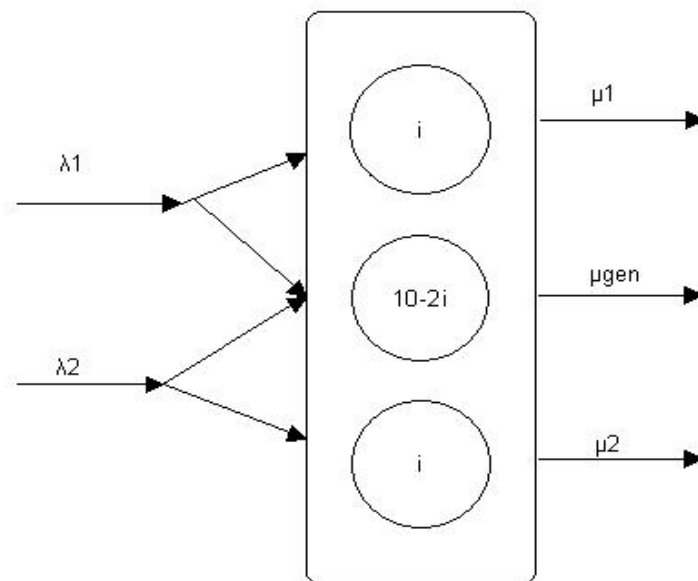
Als $r > 0.5$ moet de langzame server nooit weggehaald worden. Dit impliceert dat de langzame server niet weggehaald moet worden in het geval dat $\mu_1 < 2 \mu_2$. Als $0 < r \leq 0.5$ ($\mu_1 > 2 \mu_2$) en $\rho \leq \rho_-$, waarbij ρ_- gegeven is in vergelijking (36), kan het Call Center het beste de langzame server weghalen (wanneer er gekeken wordt naar de gemiddelde wachttijd). Het gemiddelde aantal klanten in de wachtrij, en daarmee de gemiddelde wachttijd, is kleiner als de langzame server weggehaald wordt.

12 Model met generalisten

12.1 Simulatie

In deze sectie zal gekeken worden naar een Call Center waarin 10 agenten zijn. Deze agenten kunnen zowel generalist als specialist zijn en er zal gevarieerd worden in het aantal generalisten. Hiermee verandert vanzelfsprekend het aantal specialisten. Daarnaast zal gevarieerd worden in de behandelingssnelheid van de generalisten.

Het model dat gebruikt wordt is in figuur 11 te zien.



Figuur 11: Toestanden bij slow-server queue

Er zijn twee verschillende typen gesprekken die binnenkomen. De aankomstintensiteiten zijn respectievelijk λ_1 en λ_2 . De parameters μ_1 en μ_2 zijn tijdens de simulaties 5 gekozen. De parameter μ_{gen} en het aantal generalisten zal gevarieerd worden bij de verschillende simulaties. De simulaties zullen met het programma *Arena* gemaakt worden. Het model dat gebruikt is, is in sectie 13.1.2 te zien.

In de eerste simulatie is de behandelingssnelheid van de generalist gelijk gekozen aan die van een specialist (Tabel 4). Het ideale geval (gedefinieerd als de situatie waarin de gemiddelde wachttijd het kleinste is) is naar verwachting dat het beste 10 generalisten ingeroosterd moeten worden. Hoe de wachttijd oploopt wanneer het aantal generalisten afneemt is nu de onderzoeksvraag.

In de tweede simulatie wordt de behandelingssnelheid van de generalist 60% van de behandelingssnelheid van de specialist (Tabel 5) en in de derde simulatie 40% (Tabel 6). Of in

deze gevallen de ideale situatie nog steeds bereikt wordt met 10 generalisten is een van de zaken die bekeken gaat worden.

$W_{q,1}$ is de gemiddelde wachttijd (in uren) van type 1-klanten

$W_{q,2}$ is de gemiddelde wachttijd (in uren) van type 2-klanten

W_q is de gemiddelde wachttijd (in uren) van beide type klanten

# gen.	# 1-agenten	# 2-agenten	$W_{q,1}$	$W_{q,2}$	W_q	Analyse
10	0	0	0,01336765	0,06736978	0,040368715	0,040918
8	1	1	0,01599484	0,07162363	0,043809235	-
6	2	2	0,01990437	0,07757585	0,04874011	-
4	3	3	0,02704215	0,08382626	0,055434205	-
2	4	4	0,04536086	0,1014	0,07338043	-
0	5	5	0,1099	0,109	0,10945	0,110823

Tabel 4. Gemiddelde wachttijd bij model met generalist, $\lambda=20$ en $\mu_{gen}=5$

# gen.	# 1-agenten	# 2-agenten	$W_{q,1}$	$W_{q,2}$	W_q	Analyse
10	0	0	0,00856493	0,02468227	0,0166236	0,01746
8	1	1	0,00484743	0,00982772	0,007337575	-
6	2	2	0,00290688	0,00465085	0,003778865	-
4	3	3	0,00232045	0,00275389	0,00253717	-
2	4	4	0,00252974	0,00255979	0,002544765	-
0	5	5	0,0039103	0,00421472	0,00406251	0,00400

Tabel 5. Gemiddelde wachttijd bij model met generalist, $\lambda=10$ en $\mu_{gen}=3$

# gen.	# 1-agenten	# 2-agenten	$W_{q,1}$	$W_{q,2}$	W_q	Analyse
10	0	0	0,03349862	0,1683	0,10089931	0,10230
8	1	1	0,00930404	0,01933691	0,014320475	-
6	2	2	0,00318428	0,00418427	0,003684275	-
4	3	3	0,00140225	0,00159424	0,001498245	-
2	4	4	0,00110456	0,00108734	0,00109595	-
0	5	5	0,00155614	0,00151205	0,001534095	0,001523

Tabel 6. Gemiddelde wachttijd bij model met generalist, $\lambda=8$ en $\mu_{gen}=2$

De grafieken die horen bij de tabellen staan in sectie 13.2. Deze grafieken zijn getekend met behulp van het programma R . De punten zijn door een lijn met elkaar verbonden en de punten worden aangegeven met de rondjes. In deze grafieken zijn de gemiddelde wachttijden van beide klanten berekend en uitgezet tegen het aantal generalisten. De gebruikte R -code voor een bedieningstijd van de generalist die gelijk is aan die van de specialist staat in sectie 13.4. Voor de andere twee gevallen is eenzelfde code gebruikt met de daarbijhorende waarden.

Een van de zaken dat uit de tabellen blijkt, is dat de gemiddelde wachttijd van type 2-klienten hoger is dan van type 1-klienten.

Om een realistische simulatie te bewerkstelligen is het noodzakelijk om een van beide typen klienten een hogere prioriteit te geven. Deze prioriteit houdt in dat wanneer zowel type 1-klienten als type 2-klienten aan het wachten zijn en de generalist komt vrij, de type 1-klient voorrang heeft op de type 2-klient. *Arena* werkt met tijd en op het moment dat een generalist beschikbaar komt zien beide typen klienten dat. Op dat moment zou in het programma van beide wachtrijen een klient naar de generalist gaan. Vanzelfsprekend kan één generalist niet twee klienten tegelijk behandelen. Daarom is gekozen om de type 2-klienten bij het zien van het vrijkomen van de generalist een kleine vertraging te geven. Deze vertraging is verwaarloosbaar ten opzichte van de wachttijd/behandelingstijd, maar is wel lang genoeg om te zorgen dat de generalist maar één klient toegevoerd krijgt. De type 2-klient komt in dit geval weer in de wachtrij terug. Je zou kunnen zeggen dat de generalist prioriteit geeft aan type 1-klienten.

Om deze reden is gekozen om het gemiddelde van de twee klienten te nemen als gemiddelde wachttijd in de grafieken. Normaal zou namelijk de wachttijd van beide klienten nagenoeg gelijk moeten zijn. De gemiddelde wachttijd van de type 1-klient kan dan gezien worden als een ondergrens en de gemiddelde wachttijd van de type 2-klient kan gezien worden als een bovengrens. Door hier tussenin te gaan zitten, wordt de werkelijke gemiddelde wachttijd beter benaderd.

De volgende zaken vallen op in tabel 4 en in figuur 14:

- Zoals verwacht is de ideale situatie 10 generalisten inzetten in het geval dat de generalist net zo snel werkt als de specialist.
- Bij een afname van twee generalisten is het effect op de gemiddelde wachttijd een stijging van 10%. Naarmate er minder generalisten ingeroosterd worden, hoe groter het effect zal zijn. Als er bijvoorbeeld maar twee generalisten ingeroosterd worden, heeft dat een verdubbeling van de gemiddelde wachttijd ten opzichte van de ideale situatie tot gevolg.

De volgende zaken vallen op in tabel 5 en in figuur 15:

- De ideale situatie ligt bij een snelheid van de generalist van 60% van die van de specialist niet meer bij 10 generalisten. De ideale situatie ligt hier bij het inzetten van drie generalisten.
- Het inzetten van geen generalisten zorgt voor een hogere gemiddelde wachttijd dan de ideale situatie. Blijkbaar is het gebruik van generalisten nog wel effectief op het reduceren van de wachttijd, echter wel als het aantal generalisten klein is.
- Naarmate het aantal generalisten groter wordt, stijgt de gemiddelde wachttijd sterk. Wanneer van een situatie met 8 generalisten naar een situatie met 10 generalisten gegaan wordt, levert dit een verdubbeling op van de gemiddelde wachttijd.

De volgende zaken vallen op in tabel 6 en in figuur 16:

- De ideale situatie ligt bij een snelheid van de generalist van 40% van die van de specialist niet meer bij 10 generalisten. De ideale situatie is is hier het inzetten van twee generalisten. Blijkbaar is het voordeel (flexibiliteit) wat generalisten wat zich meebrengen niet groot genoeg om de snelheid van de specialisten te achterhalen.
- De grootste gemiddelde wachttijd is bij het inzetten van 10 generalisten.
- Het maakt voor de gemiddelde wachttijd niet veel uit of er geen generalisten of vijf generalisten ingezet worden. Dit zal te maken hebben met de bezettingsgraad van het model. In dit geval zullen veel meer klanten door de specialisten geholpen worden. Deze specialisten zorgen ervoor dat de wachttijd klein blijft. Voor de klanten die naar de generalisten gaan, geldt dat deze in dat geval niet hoeven te wachten en de generalisten vormen dan een soort buffer voor de wachtrij.

12.2 Conclusie

De ideale situatie in een Call Center die volgens het beschreven model opereert, is zeker niet altijd het inzetten van allemaal generalisten. De bedieningstijd van deze generalist is van belang. Uit de simulatie blijkt dat in het geval van een bedieningsintensiteit van 100% de ideale situatie bereikt wordt als er 10 generalisten ingezet worden. Wat opvalt is dat de ideale situatie wél generalisten gebruikt. In het geval van een bedieningsintensiteit van 60% is het namelijk het beste om drie generalisten in te zetten, voor 40% geldt dat er twee generalisten ingezet worden. Blijkbaar helpen generalisten, ook al zijn ze veel langzamer dan de specialisten, de gemiddelde wachttijd te reduceren.

Een Call Center zal moeten onderzoeken wat de snelheid van de generalist is ten opzichte van die van een specialist. Door middel van een simulatie zal dan per Call Center duidelijk worden wat het ideale geval is.

12.3 Discussiepunten

Natuurlijk zijn er nog vele andere samenstellingen van agenten te bedenken en zijn ook veel meer simulaties mogelijk. Wat precies het ideale geval is wanneer de aankomstintensiteiten bijvoorbeeld verschillend zijn, is niet onderzocht. Dit zou als vervolgonderzoek gedaan kunnen worden.

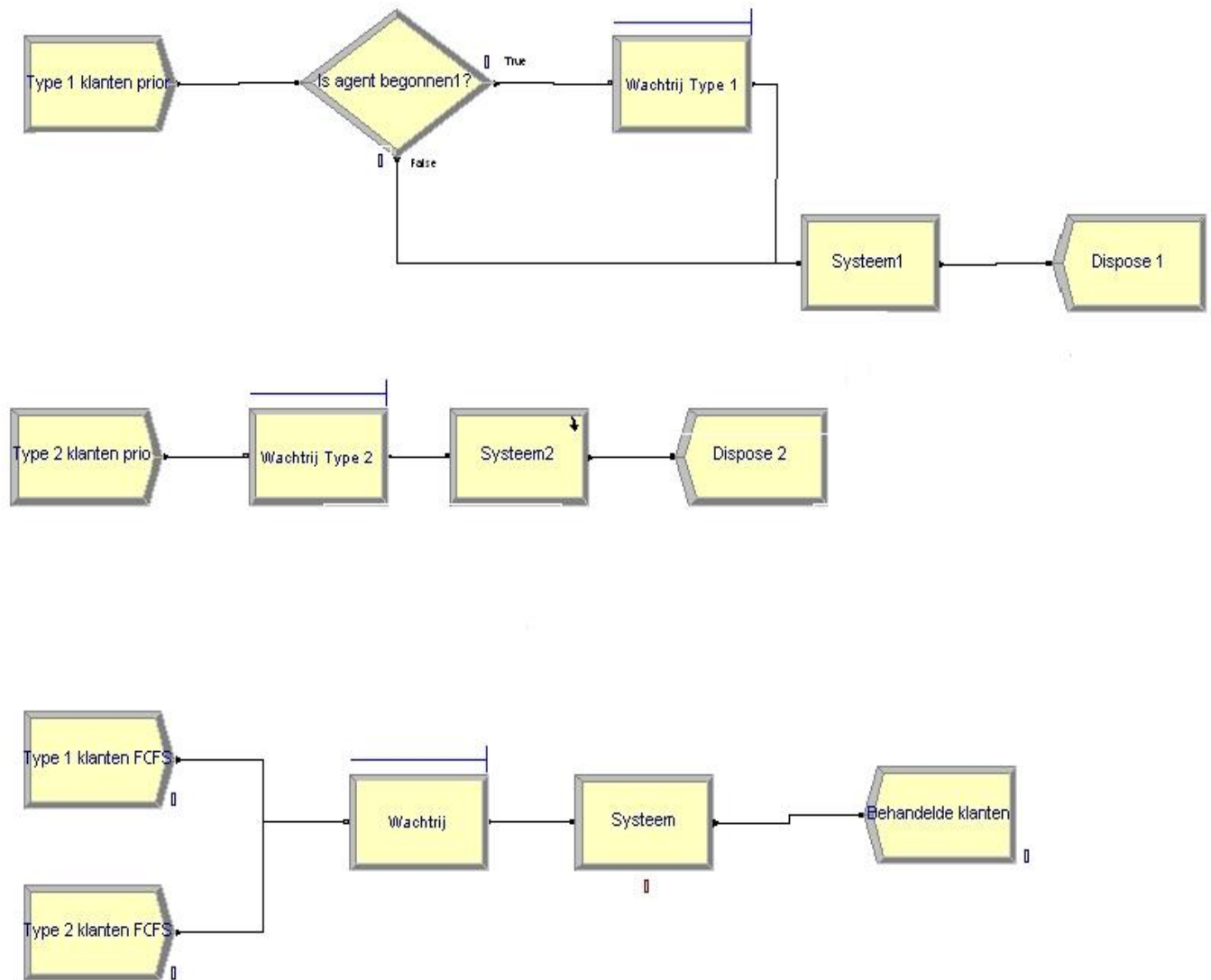
Verder zou gekeken kunnen worden naar een situatie waarin er maar één generalist is voor type 1-klanten en de rest voor de type 2-klanten.

Gekozen is er in dit onderzoek om alleen te kijken naar de gevolgen van een variatie in het aantal generalisten en een variatie in de bedieningssnelheid van de generalisten bij gelijkblijvende aankomstintensiteiten en bij gelijke bedieningssnelheid van beide type agenten.

13 Bijlage

13.1 *Arena*

13.1.1 Model met prioriteit



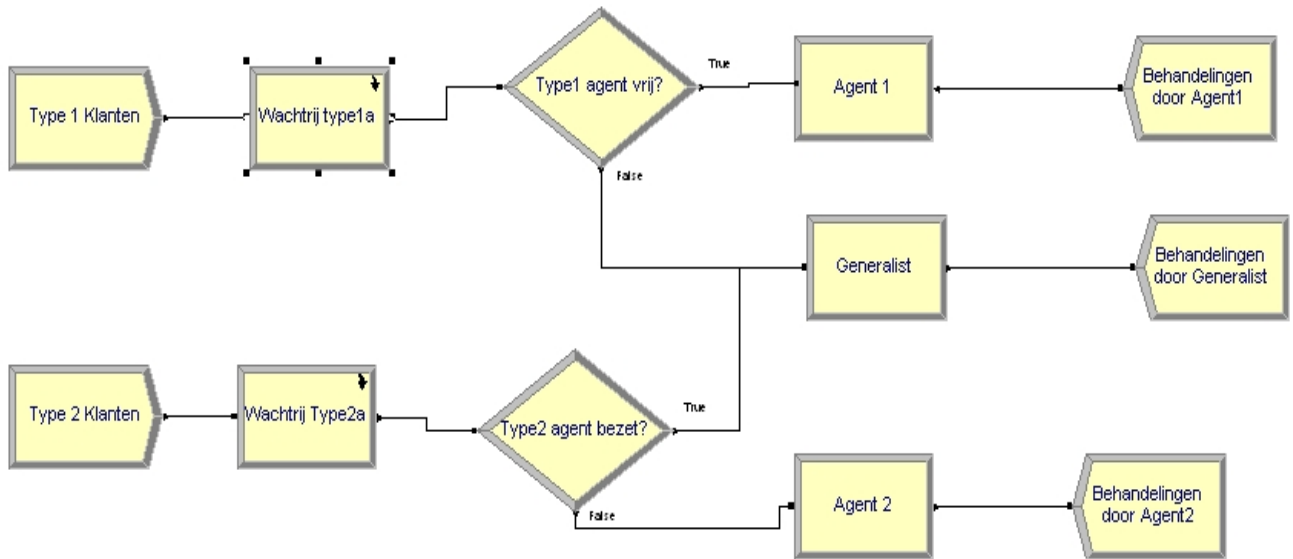
Figuur 12: Implementatie in Arena van model met prioriteit

- *Type 1 klanten prior*: Hier worden de type 1-klanten aangemaakt.
- *Type 2 klanten prior*: Hier worden de type 2-klanten aangemaakt.

- *Is agent begonnen?*: Er wordt gekeken of de agent al begonnen is, is dit niet het geval dan gaat de klant meteen naar de agent anders komt deze in de wachtrij.
- *Wachtrij Type 1*: Klanten worden hier vastgehouden totdat de agent klaar is.
- *Systeem1*: Dit is de agent en deze behandelt de klanten.
- *Dispose 1*: Hier gaan de type 1-klanten die behandeld zijn naartoe.
- *Wachtrij Type 2*: Deze houdt een type 2-klant net zolang vast totdat de agent klaar is met de type 2-klant.
- *Systeem2*: Als een klant hier arriveert, dan is de agent niet bezig met een type 2-klant. In dit submodel wordt gekeken of de agent bezig is met een type 1-klant. Is dit niet meer het geval en is de wachtrij van type 1-klanten leeg, dan wordt deze klant behandeld.
- *Dispose 2*: Hier gaan de type 2-klanten die behandeld zijn naartoe.
- *Type 1 klanten FCFS*: Hier worden type 1-klanten aangemaakt.
- *Type 2 klanten FCFS*: Hier worden type 2-klanten aangemaakt.
- *Wachtrij*: Hier komen beide type klanten bij elkaar en komen ze in de wachtrij volgens het FCFS-principe. Ze worden vastgehouden totdat de agent klaar is met een klant.
- *Systeem*: De agent in dit systeem.
- *Behandelde klanten*: Hier gaan de behandelde klanten naartoe.

Systeem 1 en *Systeem 2* zijn een en dezelfde agent.

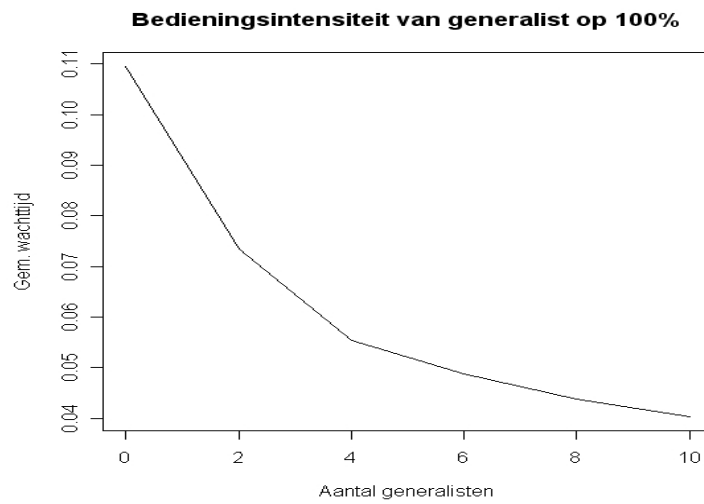
13.1.2 Model met generalisten



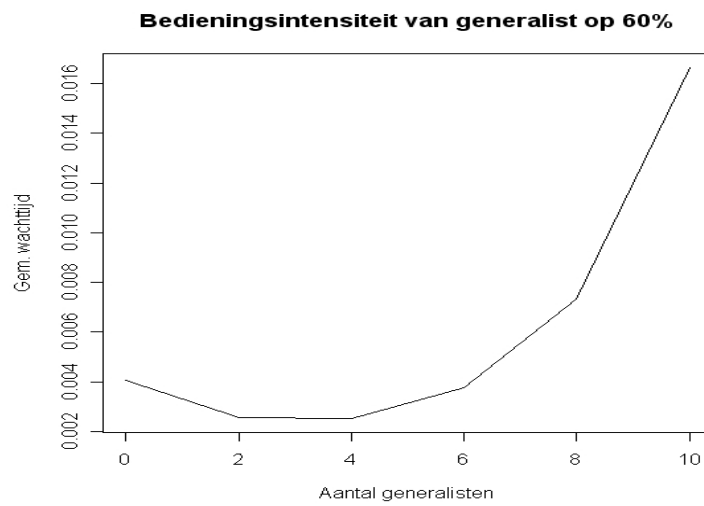
Figuur 13: Implementatie in Arena van model met generalisten

- *Type 1 klanten, Type 2 klanten*: Hier worden de klanten aangemaakt.
- *Wachtrij type1a*: Submodel waarin type 1-klanten blijven zitten totdat er een type 1-agent of generalist vrijkomt.
- *Type 1 agent vrij?*: Controleert of klant naar generalist of naar de type 1-agent moet.
- *Agent 1, Agent 2, Generalist*: Alle type 1-agenten, type 2-agenten en generalisten respectievelijk. Het aantal hangt af van de simulatie die uitgevoerd wordt.
- *Wachtrij Type2a*: Submodel waarin type 2-klanten blijven zitten totdat er een type 2-agent of generalist vrijkomt. Als er ook een type 1-klant aan het wachten was, zal in dit submodel de type 2-klant een vertraging krijgen, zodat deze in de wachtrij blijft en er niet meer klanten komen dan dat er generalisten zijn.
- *Type2 agent bezet?*: Controleert of er een type 2-agent vrij is.
- *Behandelingen door Agent1, Behandelingen door Generalist, Behandelingen door Agent2*: Hier gaan de behandelde klanten naartoe.

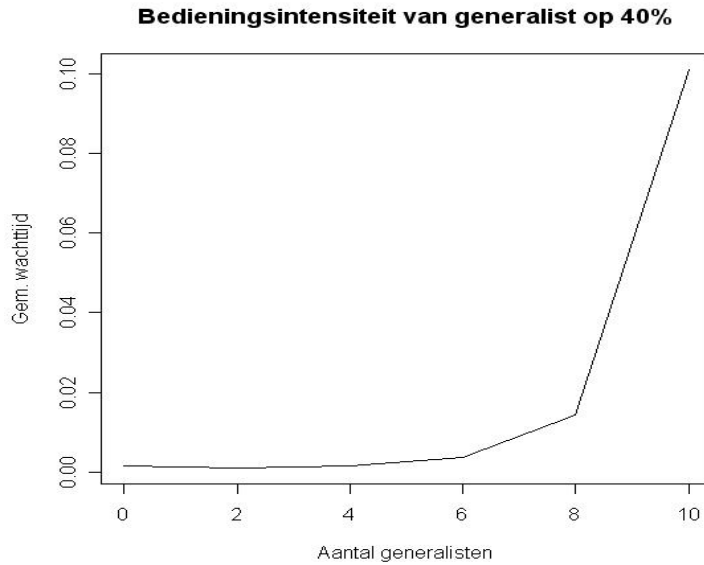
13.2 Grafieken van tabel



Figuur 14:



Figuur 15:



Figuur 16:

13.3 Slow-server queue

13.3.1 Afleiding van tweedegraadsvergelijking

In deze sectie is de afleiding van de tweedegraadsvergelijking te vinden in termen van de ratio $r = \mu_2 / \mu_1$ en ρ . Allereerst zullen de verschillende variabelen geschreven worden in termen van r en μ_1 wordt 1 veronderstelt. Zo wordt alles ten opzichte van $\mu_1 = 1$ gesteld, iets wat voor gemiddeldes geen verschil maakt.

$$\begin{aligned}
 \lambda &= \rho(\mu_1 + \mu_2) &= \mu_1 \rho(1 + r) &= \rho(1 + r) \\
 \mu &= \mu_1 + \mu_2 &= \mu_1(1 + r) &= 1 + r \\
 a &= \frac{\mu_1 \mu_2 (2\lambda + \mu)}{\mu(\lambda + \mu_2)} &= \frac{\mu_2 (2\rho(1+r) + 1 + r)}{(1+r)(\rho(1+r) + r)} &= \frac{r(2\rho + 1)}{\rho(1+r) + r} \\
 \frac{\mu_1 - a}{\mu - a} &= \frac{1 - \frac{r(2\rho + 1)}{\rho(1+r) + r}}{1 + r - \frac{r(2\rho + 1)}{\rho(1+r) + r}} &= \frac{\rho(1+r) + r - 2r\rho - r}{\rho(1+r)^2 + r(1+r) - r(2\rho + 1)} &= \dots \\
 &= \frac{\rho + \rho r - 2r\rho}{\rho + 2r\rho + \rho r^2 + r + r^2 - 2r\rho - r} &= \frac{\rho(1-r)}{\rho r^2 + \rho + r^2} &= \frac{\rho(1-r)}{\rho(1+r^2) + r^2}
 \end{aligned}$$

$$\begin{aligned}
 \rho^2(\mu - a) - 2\rho(\mu - a) + (\mu_1 - a) &= 0 \\
 \rho^2 - 2\rho + \frac{\mu_1 - a}{\mu - a} &= 0 \\
 0 &= \rho^2 - 2\rho + \frac{\rho(1-r)}{\rho(1+r^2) + r^2} \\
 &= \rho^2(\rho(r^2 + 1) + r^2) - 2\rho(\rho(r^2 + 1) + r^2) + \rho(1 - r) \\
 &= \rho(\rho(r^2 + 1) + r^2) - 2(\rho(r^2 + 1) + r^2) + (1 - r) \\
 &= \rho^2(r^2 + 1) + \rho r^2 - 2\rho r^2 - 2r^2 - 2\rho + 1 - r \\
 &= \rho^2(r^2 + 1) - \rho(2 + r^2) - (2r - 1)(r + 1)
 \end{aligned}$$

13.3.2 Bestaan van grenswaardes

De grenswaardes bestaan wanneer geldt dat de wortel in vergelijking (36) ≥ 0 is voor $0 \leq r \leq 1$.

$$\begin{aligned}
 (2+r^2)^2 + 4(1+r^2)(2r-1)(1+r) &= 4 + 4r^2 + r^4 + (4+4r^2)(2r^2+r-1) \\
 &= 4 + 4r^2 + r^4 + 8r^2 + 4r - 4 + 8r^4 + 4r^3 - 4r^2 \\
 &= 9r^4 + 4r^3 + 8r^2 + 4r
 \end{aligned}$$

De laatste vergelijking is voor alle waarden $0 \leq r \leq 1$ groter of gelijk aan 0.

13.3.3 ρ_+ -analyse

In deze sectie zal bewezen worden dat ρ_+ niet aan de stabiliteitsconditie ($0 \leq \rho \leq 1$) van ρ voldoet. Er geldt:

$$\begin{aligned}
 \rho_+ &= \frac{2+r^2+\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)}}{2(1+r^2)} && \geq 1 \\
 &2+r^2+\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)} && \geq 2+2r^2 \\
 &\frac{\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)}}{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)} && \geq r^2 \\
 &\frac{9r^4+4r^3+8r^2+4r}{8r^4+4r^3+8r^2+4r} && \geq r^4 \\
 &\frac{8r^4+4r^3+8r^2+4r}{8r^4+4r^3+8r^2+4r} && \geq 0
 \end{aligned}$$

De laatste vergelijking is voor alle waarden $0 \leq r \leq 1$ groter of gelijk aan 0. Dus $\rho_+ \geq 1$ en dus voldoet deze niet aan de condities voor ρ .

13.3.4 ρ_- -analyse

In deze sectie zal bewezen worden dat $0 \leq \rho_- < 1$ geldt voor $0 < r \leq 0.5$.

$$\begin{aligned}
 \frac{2+r^2-\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)}}{2(1+r^2)} &< 1 \\
 2+r^2-\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)} &< 2+2r^2 \\
 \frac{\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)}}{9r^4+4r^3+8r^2+4r} &> -r^2 \\
 &> r^4 \\
 \frac{8r^4+4r^3+8r^2+4r}{8r^4+4r^3+8r^2+4r} &> 0
 \end{aligned}$$

Dit geldt voor $r > 0$.

$$\begin{aligned}
 \frac{2+r^2-\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)}}{2(1+r^2)} &\geq 0 \\
 2+r^2-\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)} &\geq 0 \\
 \frac{\sqrt{(2+r^2)^2+4(1+r^2)(2r-1)(1+r)}}{9r^4+4r^3+8r^2+4r} &\leq 2+r^2 \\
 &\leq r^4+4r^2+4 \\
 \frac{8r^4+4r^3+4r^2+4r-4}{8r^4+4r^3+8r^2+4r} &\leq 0
 \end{aligned}$$

Dit geldt voor $0 \leq r \leq 0.5$.

Door de twee voorwaardes te combineren krijg je $0 \leq \rho_- < 1$ voor $0 < r \leq 0.5$.

13.4 *R*-taal

```
xas<- c(0,2,4,6,8,10)
yasgem <- c(0.10945, 0.07338043, 0.055434205, 0.04874011, 0.043809235, 0.040368715)
plot(xas,yasgem,xlab="Aantal generalisten", ylab="Gem. wachttijd", type="l",
main="Bedieningsintensiteit van generalist op 100%")
```