

# Efficiënte simulatie van wachtrijmodellen van telecommunicatiesystemen

Pieter-Tjerk de Boer

ptdeboer@cs.utwente.nl

maart 2001

## 1 Inleiding

Dit artikel gaat over een methode om efficiënt kansen op zeldzame gebeurtenissen in modellen van telecommunicatiesystemen te schatten, door middel van simulatie. Om duidelijk te maken wat het nut en de uiteindelijke praktische toepassing hiervan is, volgt eerst wat uitleg over deze wachtrijmodellen (hoofdstuk 2), en simulatie in het algemeen (hoofdstuk 3), gevolgd door een nadere beschouwing van de problematiek van het via simulatie schatten van zeer kleine kansen (hoofdstuk 4). In hoofdstuk 5 wordt dan de nieuw hiervoor ontwikkelde methode beschreven, gevolgd door wat slotopmerkingen in hoofdstuk 6. Hoewel dit onderwerp vrij wiskundig is, is het gebruik van wiskunde in de tekst zoveel mogelijk vermeden; de belangrijkste formules staan in een appendix. Zie [1] of [2] voor een veel gedetailleerdere beschrijving.

## 2 Telecommunicatie, prestatieanalyse en wachtrijmodellen

Moderne telecommunicatienetwerken transporteren informatie veelal in de vorm van pakketjes (“cellen” in ATM-netwerken) die van knooppunt naar knooppunt worden verstuurd. De verbindingen tussen de knooppunten worden gebruikt voor het transport van alle pakketten van alle connecties die over de betreffende knooppunten lopen. Dit is in principe efficiënt: als in een connectie op zeker moment relatief weinig pakketten gegenereerd worden, kan de daarbij vrij komende capaciteit voor pakketten van andere connecties benut worden. Vooral voor verkeer waarvan de intensiteit sterk wisselt (zoals WWW-dataverkeer, waarin afwisselend pagina’s worden opgehaald en gelezen) is dit voordelig. De totaal benodigde transmissiecapaciteit tussen twee knooppunten kan hierdoor kleiner worden gekozen dan indien er voor elke gebruiker een vaste capaciteit gereserveerd zou worden, zonder dat de gebruiker dit merkt in de kwaliteit van de verbinding.

Een probleem bij een pakket-geschakeld netwerk is echter dat het moeilijk is om garanties te geven over de geleverde prestaties van het netwerk. Bij circuit-geschakelde netwerken is dit relatief eenvoudig: op het moment dat de verbinding wordt opgebouwd, wordt er gekeken of er capaciteit beschikbaar is, en zo ja, dan wordt deze gereserveerd en is voor de rest van de duur van de verbinding exclusief voor deze verbinding beschikbaar. Bij een pakket-geschakeld netwerk moeten pakketten van verschillende connecties via één fysieke verbinding worden verzonden. Als nu twee pakketten (bijna) tegelijk aankomen, kan er in

eerste instantie maar één worden verzonden; het andere zal moeten wachten tot het eerste verzonden is, en daardoor enigszins vertraagd aankomen. Om zo'n pakket te laten wachten, is een vorm van geheugen nodig. Een knooppunt zal maar een eindige hoeveelheid geheugen hiervoor beschikbaar hebben, wat op een ander mogelijk probleem wijst: het kan gebeuren dat er onvoldoende geheugen beschikbaar is, en een pakket dus verworpen moet worden. Deze twee verschijnselen (vertraging en verlies) hebben uiteraard een negatieve invloed op de kwaliteit van de geleverde communicatiedienst ("Quality of Service"), en het is daarom wenselijk reeds in de ontwerpfase van een pakket-geschakeld telecommunicatiesysteem vertraging en verlieskansen te kunnen schatten.

Aspecten zoals vertraging en verlies in dergelijke systemen worden goed beschreven door de *wachtrijtheorie* (*queueing theory*). Deze tak der wiskunde beschrijft systemen waarin "klanten" aankomen bij een "loket" om "bediend" te worden, en waar zij eventueel moeten wachten op klanten die vóór hen zijn aangekomen. In een telecommunicatiecontext zijn de klanten datapakketten, is het loket een apparaat in een knooppunt van het netwerk, en is de bediening het verzenden over een fysieke verbinding. Dergelijke modellen zijn i.h.a. stochastisch van aard: de precieze tijdstippen waarop klanten aankomen zijn niet gegeven, doch slechts hun kansverdeling is bekend; evenzo kan ook de benodigde bedieningsduur onbekend zijn (in het geval van telecommunicatie: niet elk pakket is even lang, waardoor de benodigde verzendtijd verschilt).

In relatief eenvoudige gevallen laten dergelijke wachtrijmodellen zich wiskundig analyseren, en kunnen relevante grootheden zoals de gemiddelde wachttijd, of de kans dat de wachtrij groter wordt dan een gegeven buffergrootte, direct worden uitgerekend. In iets ingewikkeldere gevallen is dit niet meer mogelijk, maar kan nog wel een numerieke benadering gemaakt worden. Wanneer dat ook niet mogelijk is, zal men zijn toevlucht moeten nemen tot simulatie.

### 3 Simulatie

Simulatie houdt in, dat een model van het systeem nagebootst wordt in een computer. Het simuleren van een stochastisch model komt erop neer dat de computer mogelijke realisaties van de onderliggende stochastische processen genereert, door middel van een toevalsgetalgenerator. Bij een wachtrijmodel is bijvoorbeeld gegeven wat de kansverdeling van de tijd tussen twee aankomsten is; de computer kan dan getallen uit deze verdeling "trekken" en zodoende bepalen op welke tijdstippen een aankomst plaatsvindt. Op grond van deze trekkingen ontstaat dan een mogelijke realisatie van het gedrag van de wachtrij: een "pad" door de toestandsruimte van het systeem. Door dit vele malen te herhalen, waarbij t.g.v. het toeval telkens andere paden gerealiseerd worden, kan uiteindelijk een schatting van de waarde van een relevante prestatiemaat bepaald worden. Het systeem zou bijvoorbeeld 10000 keer gestart kunnen worden in de toestand waarin de wachtrij leeg is, en daarna telkens gesimuleerd kunnen worden tot de wachtrij weer leeg is, of overgestroomd is. Als van die 10000 pogingen in 5 gevallen overstroming optreedt, kunnen we stellen dat de kans op overstroming ongeveer  $5/10000 = 0,0005$  is.

Uit bovenstaande blijkt dat simulatie een zeer flexibele techniek is: in principe kan elk model gesimuleerd worden, en kan elke gewenste grootheid in het simulatiemodel geobserveerd, en daarmee geschat, worden. Wel zal natuurlijk de benodigde rekenkracht/-tijd toenemen met de complexiteit van het model.

Eén belangrijk aspect van simulatie is in het voorbeeld nog niet genoemd: de nauwkeurigheid. Immers, de simulatie is een toevalsproces, dus de gevonden uitkomst zal niet exact

zijn, maar wel met zekere (grote) kans in de buurt van de echte waarde liggen. Maten voor de te verwachten afwijking zijn de “variantie”, de “relatieve fout” of het “betrouwbaarheidsinterval”<sup>1</sup> In het algemeen geldt dat de schatting van zo’n overstromingskans een kleinere variantie heeft (d.w.z., beter wordt) als de overstroming vaker geobserveerd is. Vijf observaties, zoals in het gegeven voorbeeld, zijn eigenlijk te weinig voor een goede schatting; om de schatting te verbeteren, zou langer gesimuleerd moeten worden.

In praktische telecommunicatiesystemen zijn we vaak geïnteresseerd in zeer kleine kansen; althans, kansen die als ze niet klein zijn impliceren dat het systeem niet goed werkt. Voorbeelden hiervan zijn de kans dat een datapakket verloren gaat doordat er ergens onvoldoende bufferruimte aanwezig is, en de kans dat in een telefonienetwerk geen vrije lijn beschikbaar is. Als we proberen de kans op zo’n zeldzame gebeurtenis d.m.v. simulatie te bepalen, hebben we een probleem: ook in de simulatie zou de gebeurtenis maar zelden optreden, zodat we zeer lang zouden moeten simuleren om voldoende waarnemingen te krijgen om die kleine kans met een acceptabele nauwkeurigheid te schatten.

## 4 Simulatie van zeldzame gebeurtenissen: importance sampling

In de loop der jaren zijn er enkele technieken ontwikkeld om toch de kans op zeldzame gebeurtenissen efficiënt te kunnen bepalen door middel van simulatie. Eén daarvan is *importance sampling*. Bij deze techniek worden de aan de simulatie ten grondslag liggende kansverdelingen gewijzigd; dit heet een *kansmaatverandering*<sup>2</sup>. Concreet houdt dit bijvoorbeeld in, dat het gemiddelde aantal pakketten dat per seconde aan het gesimuleerde wachstelsysteem wordt toegevoerd, wordt verhoogd, en dat de gemiddelde bedieningsduur voor de pakketten wordt verlengd. Het zal duidelijk zijn dat dit een overstroming van de bufferruimte waarschijnlijker maakt, zodat de (normaliter zeldzame) overstromingsgebeurtenis minder zeldzaam wordt. (In de rest van dit artikel is “overstroming” de zeldzame gebeurtenis.)

Kunstmatig het aantal observaties van de (overstromings)gebeurtenis verhogen, zonder verder iets te doen, zou natuurlijk tot een te hoge schatting van de overstromingskans leiden. Om de kans toch juist te schatten, moeten we elke keer dat de overstroming optreedt deze niet volledig meetellen, doch slechts voor een fractie  $L$ . Deze  $L$ , de *likelihood ratio*, is gedefinieerd als de verhouding tussen de kans van wat er in de simulator gebeurd is onder de gewijzigde kansverdeling, en onder de echte kansverdeling. Het simulatieprogramma houdt  $L$  bij door telkens wanneer er een toevalsgetal is getrokken, de kans op dat getal onder de echte en onder de gewijzigde kansmaat te berekenen, op elkaar te delen, en de waarde van  $L$  tot dan toe ermee te vermenigvuldigen. Het kan wiskundig worden aangetoond dat deze procedure (kansmaat veranderen, en dan elke observatie met een factor  $L$  wegen) tot een correcte schatter leidt.

Het zal duidelijk zijn, dat de effectiviteit van importance sampling staat of valt met een goede keuze van de kansmaatverandering. Op het eerste gezicht lijkt dit simpel: wijzig het systeem zodanig dat de zeldzame gebeurtenis niet meer zeldzaam is, waardoor we deze gebeurtenis vaak observeren in de simulatie, en “dus” de kans erop nauwkeurig kunnen bepalen. Dit blijkt echter te simpel gedacht. Indien de kansmaatverandering ongunstig wordt gekozen, kan het gebeuren dat we weliswaar de zeldzame gebeurtenis vaak observeren, maar dat de schatter van diens kans een grote variantie heeft; theoretisch kan deze variantie zelfs

<sup>1</sup>Voorbeeld van het opgeven van een betrouwbaarheidsinterval: “met 95% zekerheid wijkt het simulatieresultaat niet meer dan 10% af van de werkelijke waarde.”

<sup>2</sup>Engels: *change of measure*; in veel gevallen ook *tilting*.

oneindig worden. Anderzijds kan theoretisch ook worden aangetoond dat er een kansmaatverandering bestaat die de variantie 0 maakt; helaas kan deze kansmaatverandering slechts berekend worden als de te schatten overstromingskans al bekend is, wat hem praktisch onbruikbaar maakt (een kip-en-ei-probleem).

Belangrijk voor praktische methoden om een goede kansmaatverandering te vinden, is de constatering (gesteund door theorie) dat de kansmaatverandering zó moet zijn, dat het systeem naar de zeldzame gebeurtenis “geduwd” wordt langs een pad dat het systeem ook zou volgen als het die zeldzame gebeurtenis zou bereiken zonder kansmaatverandering. Voor veel systemen is het mogelijk het “typische pad” naar overflow te vinden m.b.v. wiskunde, bijv. large-deviations-theory, en vervolgens een bijpassende kansmaatverandering te geven. Dit lukt bijvoorbeeld prima voor het schatten van de overstromingskans in een enkelvoudige wachtrij [3]. Helaas is de methode niet universeel: voor een niet eens zo veel ingewikkelder model, namelijk twee wachtrijen achter elkaar, blijkt de methode in sommige omstandigheden inderdaad nog steeds een prima resultaat te geven, maar in andere omstandigheden niet. Naast het feit dat ze in sommige omstandigheden geen goede simulatie opleveren, heeft dit soort methoden nog een ander nadeel: de benodigde wiskundige analyse is vaak niet eenvoudig. Daarmee vervalt een groot voordeel van simulatie, namelijk de mogelijkheid om willekeurige systemen te bestuderen, zonder beperkingen door wat wiskundig wel of niet analyseerbaar is.

Dit is een belangrijke reden voor de ontwikkeling van *adaptieve* methoden om de kansmaatverandering te vinden: methoden die op grond van één of meer initiële simulaties een goede kansmaatverandering voor de rest van de simulatie vinden. Deze methoden werken typischerwijs als volgt. De kansverdelingen in het model worden afhankelijk gemaakt van een beperkt aantal parameters. Vervolgens worden methoden uit de stochastische optimalisatie gebruikt om een minimum van de variantie (als functie van deze parameters) te vinden, door het uitvoeren van een aantal simulaties. Deze methoden geven vaak een kansmaatverandering die goed werkt, maar er zijn toch ook weer veel gevallen waarin geen goede kansmaatverandering gevonden wordt; bijvoorbeeld in het al eerder genoemde model met twee wachtrijen achter elkaar.

Niet alleen voor dat ene model met twee wachtrijen achter elkaar, maar ook voor diverse andere modellen van netwerkjes van wachtrijen, blijken de genoemde technieken geen goede simulatie op te leveren. Dit maakt toepassingen voor prestatieanalyse van telecommunicatienetwerken moeilijk; immers, een typisch telecommunicatiesysteem kan goed gemodelleerd worden als een netwerk van wachtsystemen, dus methoden die voor dat soort systemen vaak (en schijnbaar onvoorspelbaar) falen, zijn niet geschikt.

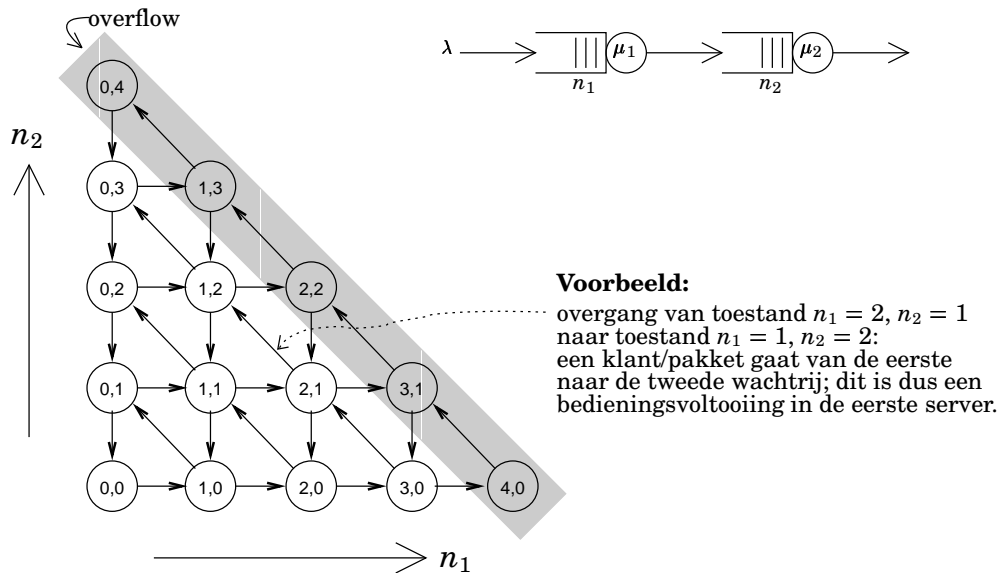
## **5 Adaptieve bepaling van een toestands-afhankelijke kansmaatverandering**

Het feit dat de in het vorige hoofdstuk genoemde adaptieve methoden er niet in slagen voor het model met twee wachtrijen achter elkaar een goede kansmaatverandering te vinden, terwijl theoretisch bekend is dat er een kansmaatverandering bestaat die de variantie 0 zou maken, wijst erop dat deze methoden een te kleine klasse van kansmaatveranderingen doorzoeken. Inderdaad, de genoemde adaptieve methoden bekijken alleen kansmaatveranderingen die één of meer parameters van de kansverdelingen aanpassen, waarbij zo'n aanpassing gedurende de hele simulatie onveranderd blijft, dus niet wordt aangepast aan de momentane toestand van het systeem. Ook voor veel met large-deviations-theory afgeleide kansmaatveranderingen geldt dit.

De reden voor de beperking tot dergelijke toestands-onafhankelijke kansmaatveranderingen is de eenvoud; met toestands-afhankelijkheid zou de benodigde wiskundige (large-deviations) analyse veel ingewikkelder worden, resp. zou een adaptieve methode in een veel grotere ruimte moeten zoeken naar een variantieminimum. In de literatuur zijn slechts weinig voorbeelden van toestands-afhankelijke kansmaatveranderingen gepubliceerd, en in alle gevallen lag daaraan een (deels heuristische) wiskundige analyse ten grondslag; daardoor zijn deze resultaten specifiek voor het betreffende probleem. Een adaptieve methode om een toestandsafhankelijke kansmaatverandering te bepalen zou niet specifiek zijn voor een model, en daarom beter toepasbaar op praktische problemen.

## 5.1 Markov-ketens

De rest van dit artikel beperkt zich tot zgn. Markovse wachtrijmodellen. Dit zijn modellen waarin de tijden tussen opeenvolgende aankomsten, en de duur van elke bediening, een exponentiële kansverdeling heeft. In veel (telecommunicatie)situaties is dat een realistisch model. Het grote voordeel van Markovse wachtrijmodellen is dat de toestand van een dergelijk systeem volledig wordt beschreven door het aantal klanten/pakketten dat zich in elk van de buffers bevindt; het is niet nodig andere grootheden, zoals de tijd sinds de laatste aankomst, of de duur van de momenteel gaande zijnde bediening, mee te nemen. Hierdoor wordt de toestandsruimte aanzienlijk verkleind, wat het bepalen van een toestandsafhankelijke kansmaatverandering gemakkelijker maakt.



Figuur 1: Twee wachtrijen achter elkaar, en de bijbehorende Markov-keten

In figuur 1 is schematisch een model van twee wachtrijen achter elkaar geschetst, en de bijbehorende “Markov-keten”: een diagram waarin alle mogelijke toestanden zijn weergegeven, alsmede (d.m.v. pijlen) de mogelijke overgangen tussen die toestanden. Een reeks toestanden is gemarkeerd als “overflow”: we zijn geïnteresseerd in de kans om één van deze toestanden bereiken<sup>3</sup>. Een toestandsafhankelijke kansmaatverandering voor dit model wordt

<sup>3</sup>nauwkeuriger geformuleerd: we laten het systeem in de toestand (0,0) beginnen (d.w.z., in de toestand waarin beide buffers leeg zijn), en willen de kans weten dat we eerder in een overflow-toestand terecht komen dan dat we in (0,0) terugkeren.

beschreven door voor elk van de pijltjes in de Markov-keten een nieuwe (transitie)kans te specificeren. Het is duidelijk dat, zeker als het overflowniveau wat hoger wordt gekozen, het aantal parameters (transitiekansen) dat bepaald moet worden zeer groot wordt.

## 5.2 De methode

Vanwege het grote aantal te bepalen transitiekansen valt een aantal adaptieve methoden al meteen af, zoals methoden die voor elke transitiekans een waarde proberen en zo zoeken naar het optimum. Een veel belovende methode is beschreven in [4]: bij deze methode wordt het minimaliseren van de variantie niet door middel van een zoektocht uitgevoerd, maar analytisch. Praktisch komt dat erop neer dat er een (initiële) kansmaatverandering wordt gekozen; daar wordt dan mee gesimuleerd, waarna op basis van deze ene set simulatiere-sultaten een nieuwe kansmaatverandering wordt berekend die in principe de variantie zou minimaliseren. Met de gevonden kansmaatverandering wordt weer gesimuleerd, waarna op basis van de resultaten daarvan weer een nieuwe (weer betere) kansmaatverandering wordt bepaald. Dit wordt enige malen herhaald, en convergeert zo naar een goede kansmaatverandering. Van deze methode bestaat een variant, de z.g. cross-entropy-methode, die bijna even goede resultaten geeft, maar numeriek gunstiger is.

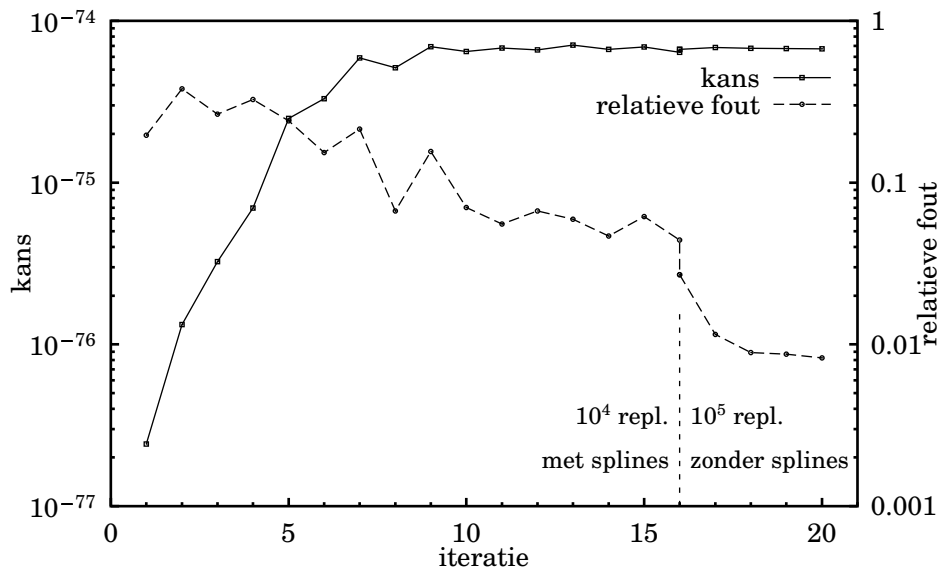
Uitwerken van de basisvergelijkingen van de cross-entropy-methode levert een formule op die de (in principe) optimale transitiekansen geeft, uitgedrukt in simulatiere-sultaten van de vorige iteratie. Deze formule ((6) in de appendix) is uiteindelijk redelijk intuïtief te begrijpen: hij stelt dat de optimale transitiekans van toestand  $X$  naar toestand  $Y$  gegeven wordt door te tellen hoe vaak de simulatie toestand  $X$  aandeed op paden die uiteindelijk naar de overflow leidden, en ook te tellen in hoeveel van die gevallen de volgende toestand  $Y$  was; de optimale kans is dan het quotiënt van deze twee tellingen<sup>4</sup>.

Voor sommige modellen is bovenstaande methode al goed bruikbaar, en levert goede resultaten op; voor veel andere modellen, met name de wat ingewikkeldere, echter niet. De oorzaak hiervan ligt in de beschreven manier om de nieuwe transitiekansen uit te rekenen: om een goede schatting van de transitiekans van  $X$  naar  $Y$  te bepalen, is het nodig dat die toestanden gedurende de (voorgaande) simulatie vaak zijn aangedaan. Echter, met een grote toestandsruimte is dat niet voor alle toestanden te garanderen. Om in die zelden aangedane toestanden toch nog een acceptabele schatting van de transitiekansen te vinden, maken we gebruik van de regelmatige structuur van de Markov-keten: het ligt voor de hand dat de kans op bijvoorbeeld een transitie corresponderend met de aankomst van een nieuwe klant in twee naburige toestanden niet al te veel verschilt. We zouden dus informatie van buurtoestanden kunnen gebruiken om de transitiekansen beter te schatten. Daartoe zijn drie technieken ontwikkeld:

- groeperen van naburige toestanden: verzamel resultaten van steeds meer omliggende toestanden tot de schatting van de transitiekans acceptabel is.
- grenslagen: experimenteel blijkt dat de transitiekansen vooral sterk van toestand tot toestand verschillen in de buurt van de grenzen van de toestandsruimte. Ver weg van die grenzen worden de transitiekansen praktisch toestandsonafhankelijk, zodat we al die toestanden als één toestand kunnen afhandelen. Naast een nauwkeurigere schatting van de optimale transitiekansen in die toestanden, levert dit bovendien een aanzienlijke verkleining van de toestandsruimte op, waardoor geheugen en rekenwerk bespaard wordt.

---

<sup>4</sup>Volledigheidshalve: bij deze tellingen wordt gewogen naar de likelihood-ratio  $L$ .



Figuur 2: Illustratie van het convergentieproces

- spline-fitting: omdat experimenteel blijkt dat de optimale transitiekansen vrij “glad” (bijv. monotoon dalend) van de toestand afhangen, is het zinvol de ruwe simulatieresultaten te verbeteren door er een gladde functie (zoals een spline) doorheen te fitten. Dit blijkt vooral de convergentie van het proces te versnellen.

Met deze drie methoden kunnen ook wat grotere netwerken prima worden gesimuleerd, zoals het voorbeeld hieronder aantoont.

### 5.3 Experimentele resultaten

Ter illustratie van de werking van de methode beschouwen we een systeem waar klanten (pakketten) vier wachtrijen na elkaar doorlopen. De aankomstintensiteit is 0,09 klanten per tijdseenheid, en de vier bedieningsintensiteiten zijn respectievelijk 0,23, 0,227, 0,227 en 0,226 bedieningen per tijdseenheid; dit alles met exponentiële kansverdeling, zodat het een Markovs systeem is. De zeldzame gebeurtenis waarop we ons richten is dat het totale aantal klanten in deze vier wachtrijen de 200 bereikt<sup>5</sup>.

Zoals reeds eerder gesteld, is de simulatieprocedure een iteratief proces: we simuleren een aantal malen (“replicaties”), berekenen uit de resultaten daarvan een schatting voor de optimale transitiekansen, gebruiken die als basis voor een volgende simulatie, waaruit weer transitiekansen volgen, enzovoort. Naast de schatting van de optimale transitiekansen levert elke iteratie ook een schatting voor de gezochte kans op, alsmede voor de onnauwkeurigheid (variantie) van die schatting. Door de schattingen van de gezochte kans en van de relatieve fout in een grafiek weer te geven, ontstaat een aardig beeld van hoe het proces convergeert; zie figuur 2.

<sup>5</sup>Deze kans is extreem klein,  $6,7 \cdot 10^{-75}$ , en daarmee veel kleiner dan voor praktische doeleinden interessant is. Echter, het feit dat de methode in deze extreme situatie nog goed werkt, toont z'n robuustheid aan. Bovendien is dit een model waarvan wiskundig bewezen is dat een op large-deviations-theory gebaseerde toestandsonafhankelijke kansmaatverandering geen goede resultaten geeft[5], zodat dit voorbeeld ook uit fundamenteel oogpunt interessant is.

In de grafiek is een aantal kenmerkende verschijnselen zichtbaar. In de eerste paar iteraties is de schatting van de kans veel te klein, zo rond de  $10^{-76}$  terwijl de juiste waarde  $6,7 \cdot 10^{-75}$  is. Zo'n te lage schatting is kenmerkend voor een importance-sampling-simulatie waarin een onjuiste kansmaatverandering is toegepast. Gaandeweg (ongeveer de eerste 7 iteraties) wordt de schatting van de gezochte kans steeds beter: de schatting convergeert. Ondertussen begint ook de geschatte fout te dalen, en na zo'n 12 iteraties is deze geconvergeerd op ca. 5 %. Daarna verandert er weinig meer, tot de 16e iteratie. Op dat punt is de simulator (handmatig) omgeschakeld van  $10^4$  replicaties per iteratie, naar  $10^5$  replicaties per iteratie; bovendien is de spline-fitting uitgeschakeld. Zoals te verwachten neemt bij een toenemend aantal replicaties de geschatte fout af; immers, we hebben nu meer observaties om de schatting op te baseren. De spline-fitting was aanvankelijk wel nodig om de convergentie te vergemakkelijken door de aanvankelijk slechte schattingen van de optimale transitiekansen "glad" en daarmee beter te maken; naarmate de ruwe schattingen van de transitiekansen beter worden in de loop van het convergentieproces, werkt dit juist averechts, en is het beter de spline-fitting uit te schakelen.

niveau	exact	simulatie-schatting	relatieve fout
25	$3,5283 \cdot 10^{-7}$	$3,504 \cdot 10^{-7}$	0,26%
50	–	$2,396 \cdot 10^{-16}$	0,42%
100	–	$1,422 \cdot 10^{-35}$	0,44%
200	–	$6,722 \cdot 10^{-75}$	0,82%

Tabel 1: Simulatieresultaten voor verschillende overflowniveaus (bij  $10^5$  replicaties)

Tabel 1 geeft de simulatieresultaten bij verschillende waarden van het overflowniveau. Uiteraard neemt de overflowkans sterk af bij toenemend overflow-niveau. We zien in de tabel echter dat de relatieve fout slechts langzaam toeneemt. Dit gedrag wordt ook wel *asymptotisch efficiënt* genoemd: de variantie neemt slechts polynomiaal toe, terwijl de geschatte kans exponentieel afneemt. Zonder gebruik van (correcte) importance-sampling zou de variantie ook exponentieel toenemen. Ter verificatie is ook het resultaat van een numerieke berekening van de overflowkans voor overflowniveau 25 gegeven; voor de hogere overflowniveaus is deze numerieke berekening niet uitvoerbaar vanwege de te grote toestandsruimte.

## 6 Conclusies

De ontwikkelde methode is een robuuste methode voor het efficiënt, door middel van simulatie, bepalen van kansen op zeldzame gebeurtenissen in Markovse wachtrijmodellen. De voorbeeldresultaten laten zien hoe de methode uitgaande van een erg slechte initiële kansmaatverandering na een beperkt aantal iteraties een goede kansmaatverandering heeft gevonden; ook blijkt daar dat de methode asymptotisch efficiënt is, wat praktisch betekent dat ook extreem kleine kansen nog efficiënt bepaald worden. De methode is verder met succes toegepast op andere modellen, zoals modellen waarin niet naar de overflow van de totale netwerkpopulatie werd gekeken, maar naar die van één buffer in het netwerk; en netwerken met feedback, d.w.z. dat een deel van de pakketten/klanten van de uitgang van het netwerk teruggevoerd werd naar een van de eerdere wachtrijen, waarmee bijvoorbeeld hertransmissies gemodelleerd kunnen worden.

De methode heeft echter ook beperkingen. Momenteel is de methode alleen toepasbaar op Markovse modellen; veel niet-Markovse modellen laten zich echter benaderen door Markovse modellen (bijv. door gebruik te maken van phase-type kansverdelingen), en er zijn al wat

ideeën om de methode algemener te maken. Ook ontstaan er nog problemen als het aantal wachtrijen in het netwerk (of meer in het algemeen: het aantal dimensies van de toestandsruimte) te groot wordt; de in hoofdstuk 5.2 genoemde technieken helpen weliswaar om met een grote toestandsruimte om te gaan, maar nog betere technieken zullen nodig zijn voor grotere systemen.

Hoewel de methode oorspronkelijk is ontwikkeld voor wachtrijmodellen van datapakketten die door telecommunicatienetwerken stromen, is hij ook op andere systemen toepasbaar die zich laten modelleren met een Markov-keten. Ook binnen de telecommunicatie zijn hiervan voorbeelden te vinden; te denken valt aan de belasting (aantal actieve gesprekken) in cellen in een GSM-netwerk, onder bepaalde aannamen over de gespreksduur en de kans op handovers.

## Appendix: wiskundige formulering

### A.1 Standaard simulatie

De basisformule om d.m.v. stochastische simulatie de kans op een bepaalde gebeurtenis te schatten uit  $N$  replicaties (pogingen) is

$$\hat{p} = \frac{\sum_{i=1}^N I_i}{N}. \quad (1)$$

Hierin is  $p$  de te schatten kans, en is  $I_i$  de zgn. indicatorfunctie: een functie die 1 is als in de  $i$ 'e replicatie de gezochte gebeurtenis optrad, en anders 0. Eigenlijk is de teller dus gewoon het aantal keer dat de gebeurtenis optrad, in  $N$  pogingen. Het dakje op  $\hat{p}$  geeft aan dat dit slechts een schatting van de kans  $p$  is.

De standaard-deviatie (= wortel uit de variantie) van de hierboven gegeven schatter wordt op zijn beurt geschat door

$$s.d. = \sqrt{\frac{\hat{p} - \hat{p}^2}{N - 1}}, \quad (2)$$

waarmee de relatieve fout (= standaard-deviatie gedeeld door gemiddelde) uitkomt op

$$RE = \sqrt{\frac{1/\hat{p} - 1}{N - 1}}. \quad (3)$$

Hieruit blijkt duidelijk dat om steeds kleinere kansen  $p$  te schatten met gelijkblijvende relatieve fout,  $N$  proportioneel moet toenemen.

### A.2 Importance sampling

Beschouwen we nu simulatie met importance sampling, waarbij de  $i$ 'e replicatie een likelihood-ratio  $L_i$  heeft. Dan geldt:

$$\hat{p} = \frac{\sum_{i=1}^N I_i L_i}{N} \quad (4)$$

$$RE = \sqrt{\frac{\frac{\sum_{i=1}^N I_i L_i^2}{\hat{p}^2 N} - 1}{N - 1}}. \quad (5)$$

Duidelijk is dat de relatieve fout afhangt van wat voor waarden  $L_i$  aanneemt.

### A.3 Adaptief bepaalde toestandsafhankelijke transitiekansen

Uitgaande van de cross-entropy-methode volgt de volgende uitdrukking voor de optimale transitiekansen, in termen van simulatieresultaten:

$$q_{lm} = \frac{\sum_{Z=Z_1}^{Z_N} I(Z)L(Z) \sum_{i:z_i=l} 1_{(z_{i+1}=m)}}{\sum_{Z=Z_1}^{Z_N} I(Z)L(Z) \sum_{i:z_i=l} 1}. \quad (6)$$

Hierin is  $\hat{q}_{lm}$  de schatting van de optimale transitiekans van toestand  $l$  naar toestand  $m$ . De symbolen  $Z_1$  t/m  $Z_N$  stellen de “sample-paden” voor die in de 1e t/m de  $N$ e replicatie van de simulatie zijn opgetreden. Zo’n  $Z$  is een rij getallen, die aangeven welke toestanden na elkaar zijn bereikt in die replicatie; de  $i$ e toestand van die replicatie noteren we als  $z_i$ . En  $I(Z)$  is de indicatorfunctie, dus een functie die 1 is dan en slechts dan als het pad  $Z$  de gezochte gebeurtenis bevat, en  $L(Z)$  is de likelihood-ratio van het pad  $Z$ . Uit bovenstaande volgt dat de uitdrukking  $\sum_{i:z_i=l} 1$  in de noemer telt hoe vaak het sample-pad  $Z$  door toestand  $l$  ging; en evenzo telt  $\sum_{i:z_i=l} 1_{(z_{i+1}=m)}$  hoe vaak het sample-pad  $Z$  door toestand  $l$  en direct daarna door toestand  $m$  ging.

## Literatuur

- [1] Pieter-Tjerk de Boer, Victor F. Nicola, and Reuven Y. Rubinstein. Adaptive importance sampling simulation of queueing networks. Winter Simulation Conference ’00, Orlando, FL, 2000. Online: <http://www.informs-cs.org/wsc00papers/086.PDF>
- [2] Pieter-Tjerk de Boer. *Analysis and efficient simulation of queueing models of telecommunication systems*. Proefschrift, Universiteit Twente, 2000. Online: <http://www.cs.utwente.nl/~ptdeboer/thesis/>
- [3] S. Parekh and J. Walrand. A quick simulation method for excessive backlogs in networks of queues. *IEEE Transactions on Automatic Control*, 34:54–66, 1989.
- [4] Reuven Y. Rubinstein. Optimization of computer simulation models with rare events. *European Journal of Operations Research*, 99:89–112, 1997.
- [5] Paul Glasserman and Shing-Gang Kou. Analysis of an importance sampling estimator for tandem queues. *ACM Transactions on Modeling and Computer Simulation*, 5(1):22–42, January 1995.

Voor een uitgebreidere literatuurlijst, zie [1] of [2].