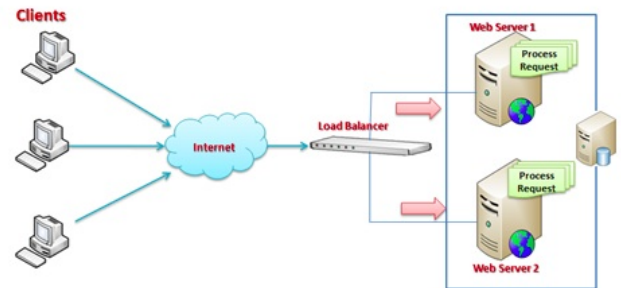


Optimaal beheer van een webfarm

Een webserver trekt enorme klantenaantallen voor de ondersteunde websites. Eén server is dan niet voldoende om de klanten adequaat te helpen. Meerdere servers zijn noodzakelijk om aan de vraag te voldoen. Het aan laten staan van servers is duur. Uit laten staan is goedkoop, maar het kost geld om een server aan te zetten.

Als een klant een website aanklikt, wordt hij toegewezen aan een standby server, als die er is. Zo niet, dan wordt er een server aangezet. Als een klant klaar is, kan de desbetreffende server op standby gezet worden of hij kan uitgezet worden. Welke strategie minimaliseert de verwachte kosten?



Modellering Dit probleem kan als een Markov beslissingproces worden gemodelleerd. Omdat het probleem de vraag betreft hoeveel webserver er optimaal aan moeten staan, veronderstellen we dat we de beschikking hebben over een oneindig aantal webserver. Als het aankomstproces van klanten een Poisson proces is en klanten een exponentieel verdeelde tijd een website bezoeken, hoeven we slechts bij te houden hoeveel server bezig zijn en hoeveel er standby aan staan.

Tijd is continu, maar voor het berekenen van een optimale strategie moet je de tijd discretiseren. Dan kan het zogenaamd *waarde-iteratie algoritme* toegepast worden, waarmee de optimale strategie berekend kan worden.

Omdat het aantal klanten dat de websites bezoekt, onbegrensd kan zijn, is het aantal klanten per tijdseenheid dat uitgesurft is, ook onbegrensd. Discretisatie is in zo'n geval niet mogelijk. Eén mogelijkheid om hiermee om te gaan, is door aan te nemen dat het aantal klanten per tijdseenheid dat uitgesurft is, te begrenzen. Het resultaat hiervan is nog niet gepubliceerd, maar wel beschreven in de presentatie [1] van Adan, Kulkarni en van Wijk. De optimale strategie blijkt met behulp van een zogenaamde 'switching curve' te karakteriseren: als een webserver uitgezet wordt, wanneer er x klanten aan het surfen zijn, dan wordt deze ook uitgezet wanneer meer klanten bezig zijn. Als een web server aan blijft wanneer er x klanten aan het surfen zijn, dan blijft hij zeker aan staan wanneer er minder zijn.

Een open probleem is wat het effect van deze begrenzing is op de structuur van de optimale strategie van het oorspronkelijke probleem. Is de optimale strategie simpeler van vorm, namelijk een *drempelstrategie* (zet een vrijgekomen server alleen uit als er meer dan k standby server aan staan)?

In de theorie zit wat dat betreft een 'gat'. Bhulai, Brooms en Spieksma [2] hebben een alternatieve aanpak ontwikkeld om problemen met onbegrensd overgangssnelheden te bestuderen. Deze aanpak lijkt de structurele eigenschappen van optimale strategieën goed te conserveren. Het is interessant om te weten wat dit oplevert voor het webfarmmodel.

Project Het project kan de volgende onderdelen omvatten:

1. Het bestuderen van de basistheorie van Markov processes en Markov Beslissingsketens, met name het waarde-iteratie algoritme. Voorts de aanpak in artikel [2].
2. Aan de hand van de scriptie van Herman Blok [3] (geënt op Koole [4]), en [1] analyse van het web-farmmodel.
3. Numerieke analyse en vergelijk met de resultaten in [1].

Referenties

- [1] I. Adan, V. Kulkarni, and S. van Wijk (2011). Optimization of Web Farms. Presentation at Annual INFORMS Conference, Charlotte, U.S.A.
- [2] S. Bhulai, A. Brooms and F.M. Spieksma (2010). A processor sharing retrial queue: structural properties of the value function for an unbounded Markov jump process. Submitted to Qesta.
- [3] H. Blok (2010). Markov decision processes with unbounded transition rates: structural properties of the relative value function. Master Thesis UU.
- [4] G.M. Koole (2006). Monotonicity in Markov Reward and Decision Chains: Theory and Applications. Foundations and Trends in Stochastic Systems **1**, 1 –76.

Floske Spieksma
spieksma@math.leidenuniv.nl
Bachelorproject voor het AS&B seminarium
voorjaar 2012