

# VU Research Portal

## Optimal Quality of Service Control in Communication Systems

Bosman, J.W.

2014

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Bosman, J. W. (2014). *Optimal Quality of Service Control in Communication Systems*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

### **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
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

---

# Samenvatting (Dutch Summary)

---

## Optimale besturing van serviceniveaus in ICT-systemen

De hedendaagse, complexe, samengestelde ICT-systemen worden vaak op een ongecentraliseerde wijze bestuurd zonder dat er een goed inzicht is in de gevolgen van storingen in *ketens* van ITC-diensten. Hoewel deze aanpak goed kan werken voor kleine ketens, wordt dit onhaalbaar voor complexe wereldwijde dienstenketens. Daarom zijn er mechanismen nodig die op een efficiënte wijze beschikbare (netwerk)systeemcapaciteit kunnen benutten zonder aan het gewenste serviceniveau voor de eindgebruikers in te boeten.

Het belangrijkste vraagstuk dat in deze dissertatie wordt behandeld is: *hoe kunnen serviceniveaumechanismen effectief worden toegepast op complexe grootschalige ICT-systemen met gedeelde systeemcapaciteit?*

Om deze vraag te beantwoorden worden er in deze dissertatie kwantitatieve modellen ontwikkeld, geanalyseerd, geoptimaliseerd en geëvalueerd die de essentiële dynamiek beschrijven van op service gerichte besturingsmechanismen en wordt het gevolg bestudeerd van het gebruik van deze modellen op het (door de gebruikers ervaren) serviceniveau. Het achterliggende doel van deze aanpak is om schaalbare, robuuste algoritmen, beslistabellen en vuistregels te ontwikkelen die het mogelijk maken om service gerichte besturingsmechanismen optimaal toe te passen. Bij de toepassing hiervan worden drie complicerende factoren onderscheiden:

*Veranderlijkheid van de vraag naar systeemcapaciteit.* Een groot deel van de vraag naar systeemcapaciteit wordt bepaald door voorspelbaar gedrag van gebruikers. Naast voorspelbaar gedrag zijn er ook moeilijk te voorspellen fenomenen die een grote invloed hebben op de beschikbaarheid van systeembronnen, zoals aanvallen via het internet of onverwachte drukte door bijvoorbeeld het bekend worden van een grote gebeurtenis in de media.

*Onzekerheid over de beschikbaarheid van systemen.* Verschillende zaken dragen bij aan de variatie in beschikbaarheid van systemen. Voorbeelden hiervan zijn het delen van systeemcapaciteit, uitval van netwerken en systemen, chaotisch gedrag van systemen en tijdelijke overbelasting. In de meeste ICT-systemen is het delen van systeemcapaciteit de belangrijkste bron van variabiliteit in de beschikbaarheid. Een andere belangrijke factor is chaotisch gedrag van systemen als gevolg van een onverwachte wisselwerking tussen verschillende systemen. Vaak wordt dergelijk gedrag veroorzaakt door configuratiefouten. In het uiterste geval kunnen netwerken en systemen vastlopen. Dit is in het bijzonder het geval voor globale systemen waar

de vraagvolumes dusdanig groot zijn dat losse systemen niet in staat zijn om alle vraag individueel af te handelen.

*Beperkte informatie over wat zich afspeelt in externe systemen.* Veel gebruikte modellen veronderstellen dat het stochastische gedrag van vraag en systeemcomponenten of capaciteit bekend is. In de praktijk is dit zelden het geval. Meestal hebben gebruikers slechts beperkt zicht op wat er zich afspeelt in de systemen van externe partijen die zij gebruiken. Bovendien is het mogelijk dat de systemen draaien in een omgeving die onderhevig is aan onzekere en onvoorspelbare factoren. Om met die beperkte informatie om te kunnen gaan, zijn mechanismen nodig die zich kunnen aanpassen aan deze onzekere en veranderende factoren.

Hoofdstukken 2 en 3 beschouwen een vloeistofmodel dat het gedrag van video over het internet, bijvoorbeeld YouTube, beschrijft. Een storende factor in video's over het internet is dat deze kunnen gaan haperen tijdens het afspelen. Uit het vloeistofmodel volgt een aanpak waarmee de laadtijd en andere parameters zoals videokwaliteit en bandbreedte zo kunnen worden gekozen dat een video met een grote waarschijnlijkheid onafgebroken afspeelt.

In hoofdstuk 4 wordt de situatie beschouwd waarin bestanden kunnen worden verstuurd over meerdere draadloze netwerken. Voor deze situatie wordt een eenvoudig toepasbare beslisregel geformuleerd, genaamd convexe combinatie (CC), die bepaalt over welk netwerk een bestand moet worden verstuurd, gebruikmakend van de geobserveerde drukte in de netwerken. De beslisregel is gebaseerd op een combinatie van twee regels die goed werken in verschillende situaties. Om de effectiviteit van onze beslisregel te evalueren is de CC regel geïmplementeerd in een simulatieomgeving die het gedrag van netwerken realistisch nabootst. Uit de resultaten blijkt dat de CC regel goed presteert bij een breed scala aan belastings- en capaciteitsparameters van draadloze netwerken.

In hoofdstuk 5 wordt een toewijzingsprobleem behandeld. Er zijn twee diensten aanwezig: een interne (goedkopere) dienst en een (duurdere) dienst van een externe partij. Verder is er een variërend vraagproces dat zowel lange termijn patronen vertoont als korte termijn schommelingen. De uitdaging is om de interne capaciteit goed te kiezen, zodat de schommelingen kunnen worden opgevangen. Indien er meer vraag is dan toegewezen interne capaciteit gaat er vraag verloren. In het geval dat teveel capaciteit is toegewezen, wordt er betaald voor ongebruikte capaciteit. Echter, aan het aanpassen van de interne capaciteit zijn ook kosten verbonden. Het is van belang om een goede afweging te maken tussen aanpassingskosten van de interne capaciteit en de door schommelingen in het vraagproces veroorzaakte onder- of overcapaciteit. In dit hoofdstuk wordt een eenvoudig te implementeren mechanisme geformuleerd dat goed met deze schommelingen om kan gaan. Om dit mechanisme te demonstreren is er een simulatieomgeving opgezet. Uit de experimenten blijkt dat het beschreven besturingsmechanisme zeer goed functioneert.

---

Hoofdstuk 6 beschouwt dynamische compositie van samengestelde webdiensten. Daarbij wordt de samengestelde webdienst als een keten van taken gerepresenteerd die sequentiëel moeten worden uitgevoerd. Voor elke taak in de keten zijn implementaties beschikbaar van externe partijen met elk hun eigen prijs-kwaliteitsverhouding. De samengestelde webdienst is onderdeel van een serviceovereenkomst waarin staat dat de respons op elke vraag binnen een vastgestelde termijn plaats moet vinden. Met deze overeenkomst voor ogen is in dit hoofdstuk een algoritme ontwikkeld dat een dynamische beslisstrategie berekent voor de gegeven serviceovereenkomst en prijs-kwaliteitsverhouding van de gebruikte diensten van externe partijen. Het algoritme neemt beslissingen op basis van de resterende responstijd voordat het in de serviceovereenkomst gestelde tijdsdoel wordt overschreden. Uit experimenten blijkt dat er enorme winst valt te behalen door de compositie dynamisch te laten aanpassen aan de resterende responstijd.

In hoofdstuk 7 wordt uitgegaan van de dynamische beslisstructuur van hoofdstuk 6. Echter, dit maal wordt er verondersteld dat het gedrag in termen van responstijd van derde partijen niet bekend is en geleerd moet worden uit geobserveerde responstijden. Dit hoofdstuk een ontwikkelt aanpak waarbij een dynamische programmeertechniek wordt toegepast die is gebaseerd op de *empirische responstijdverdelingen*. De empirische responstijdverdelingen worden actueel gehouden door middel van twee mogelijke principes het vensterprincipe en het uitdoofprincipe. Op de empirische verdelingen worden statistische toetsen toegepast om te kijken of er significante veranderingen zijn geweest in de responstijdverdelingen. Op deze manier hoeft het dynamisch programmeeralgoritme niet voor elke waarneming een nieuwe beslistabel te berekenen. Om te voorkomen dat bepaalde diensten nooit bezocht worden, omdat deze diensten niet in de dynamische beslistabel zitten. Mogelijk vormen deze onbezochte diensten toch een aantrekkelijk alternatief, omdat ze beter zijn gaan presteren. Daarom worden er testaanvragen verstuurd. Dit zijn aanvragen die informatie inwinnen over de onbezochte diensten. Om de beschreven aanpak goed te laten werken moeten verschillende parameters worden afgewogen. Uit de simulatie-experimenten blijkt dat de beschreven aanpak in veranderende omgevingen veel winst kan opleveren ten opzichte van statische aanpakken die worden berekend over een langere termijn.