

Karen Aardal

*Delft Institute of Applied Mathematics
TU Delft
k.i.aardal@tudelft.nl*

Thije van Barneveld

*Centrum Wiskunde & Informatica
Amsterdam
t.c.van.barneveld@cw.nl*

Pieter van den Berg

*Delft Institute of Applied Mathematics
TU Delft
p.l.vandenberg@tudelft.nl*

Sandjai Bhulai

*Afdeling Wiskunde
Vrije Universiteit Amsterdam
s.bhulai@vu.nl*

Martin van Buuren

*Centrum Wiskunde & Informatica
Amsterdam
m.van.buuren@cw.nl*

Theresia van Essen

*Centrum Wiskunde & Informatica, Amsterdam, en
Delft Institute of Applied Mathematics, TU Delft
j.t.vanessen@tudelft.nl*

Caroline Jagtenberg

*Centrum Wiskunde & Informatica
Amsterdam
c.j.jagtenberg@cw.nl*

Geert Jan Kommer

*Rijksinstituut voor Volksgezondheid en Milieu
Bilthoven
geertjan.kommer@rivm.nl*

Guido Legemaate

*Brandweer Regionale Eenheid Amsterdam, en
Centrum Wiskunde & Informatica, Amsterdam
g.a.g.legemaate@cw.nl*

Rob van der Mei

*Centrum Wiskunde & Informatica
Amsterdam
r.d.van.der.mei@cw.nl*

Van reactieve naar proactieve planning van ambulancediensten

In dit artikel berichten Karen Aardal e.a. over het REPRO-project, een onderzoeksproject dat zich richt op het ontwikkelen van slimme voorspellings- en planningsmethoden voor ambulancediensten. In Nederland is de norm dat een ambulance in 95 procent van de spoedeisende gevallen binnen vijftien minuten ter plaatse moet zijn. Het nieuw ontwikkelde systeem kan hiervoor uitrekenen wat de optimale locaties van de standplaatsen zijn en de spreiding van ambulances over de standplaatsen.

In levensbedreigende situaties waarin elke seconde telt, kan het al dan niet op tijd ter plaatse zijn van een ambulance het verschil maken tussen leven en dood. In deze gevallen is het streven om zo snel mogelijk en bij voorkeur binnen vijftien minuten ter plaatse van de patiënt te zijn. Om zulke korte responstijden te realiseren is een uitgekiende planning van ambulanceritten noodzakelijk. Bij het realiseren van een efficiënte aansturing van ambulanceritten spelen allerlei vragen een rol, zowel op strategisch, tactisch als operationeel niveau. Typische vragen zijn: "Hoe kunnen we op de juiste wijze anticiperen en reageren op pieken in de vraag naar ambulance-

ritten?", "Hoeveel standplaatsen hebben we nodig en wat zijn de goede locaties voor die standplaatsen?", "Hoeveel teams moeten we minimaal inzetten om de gewenste servicegraad te realiseren, en waar en wanneer?" en "Hoe kunnen we op elk moment van de dag een goede dekking van ambulances over een regio realiseren door slimme dynamische en proactieve real-time herpositionering van ambulances?"

In het STW-project 'Van Reactieve naar Proactieve Planning van Ambulancediensten', kortweg REPRO, staan bovengenoemde vragen centraal. Binnen het project worden nieuwe modellen en oplossingsmethoden

niet alleen ontwikkeld en geëvalueerd, maar ook daadwerkelijk toegepast in de praktijk. Het project is een samenwerking tussen het Centrum Wiskunde & Informatica (CWI), de Technische Universiteit Delft, de Vrije Universiteit Amsterdam, het Rijksinstituut voor Volksgezondheid en Milieu (RIVM), Ambulance Amsterdam, Witte Kruis Ambulancezorg en de regionale ambulancevoorzieningen Flevoland, Gooi- en Vechtstreek en Utrecht. Daarnaast zijn ook Brandweer Regionale Eenheid Amsterdam, hulpdiensten-software specialist CityGIS, Politie Regionale Eenheid Amsterdam, Ambulancezorg Nederland, het Universitair Medisch Centrum Groningen, het Oranje Kruis en de ANWB bij het project betrokken. Het is belangrijk om op te merken dat de vraagstukken die voorkomen in de ambulancesector vergelijkbaar zijn met vraagstukken die voorkomen bij de planning van dienstverlening in andere sectoren, zoals de brandweer, de politie, de wegwacht, maar ook

bij het managen van *spare parts* waarbij (dure) reserveonderdelen over een regio moeten worden verspreid zodanig dat ze snel kunnen worden ingezet als onderdelen van apparatuur en machines uitvallen. De wiskundige modellen die wij ontwikkelen en gebruiken hebben daardoor een breder nut dan alleen voor ambulanceplanning.

Wij modelleren de planningsproblemen rondom de ambulancezorg als *optimaliseringsproblemen in netwerken*. De netwerkstructuur is meer of minder expliciet aanwezig afhankelijk van het onderliggende vraagstuk. Om de algoritmes en oplossingen voor de vraagstukken te analyseren, testen en evalueren, hebben we een simulatietool ontwikkeld.

Een sterk complicerende, maar wetenschappelijk zeer uitdagende, factor bij het oplossen van dit soort planningsvraagstukken is de grote mate van *onzekerheid* die inherent is aan het ambulance-serviceproces. De bestaande planningsmethodieken gaan er doorgaans van uit dat de vraag naar en beschikbaarheid van ambulances van tevoren bekend zijn, terwijl veel van deze methoden zeer gevoelig zijn voor veranderingen van de invoerparameters.

In dit artikel zullen verschillende aspecten van ambulanceplanning worden belicht. In de volgende paragraaf worden verschillende modellen besproken voor het bepalen van de goede locaties voor standplaatsen, en de bijbehorende bezetting van ambulances. Daarna wordt ingegaan op locatiemodellen specifiek voor de dienstverlening van de brandweer. Vervolgens richten wij ons op modellen en algoritmen door dynamisch ambulance management, waarbij de dekking in real-time wordt verbeterd door middel van het uitvoeren van zogenaamde *proactieve relocaties* op basis van de actuele en toekomstige locaties van de ambulancevoertuigen en de incidenten. Dan volgt nog een paragraaf waarin wij verder ingaan op het optimaal plannen van besteld ambulancevervoer en een paragraaf waarin zal worden ingegaan op de simulatietool *Testing Interface for Ambulance Research* (TIFAR) waarin de ontwikkelde plannings- en relocatiealgoritmen zijn geïmplementeerd. Tot slot wordt in de laatste paragraaf een simulatiemodel voor ambulancemeldkamers gepresenteerd.

Voordat we op de verscheidene aspecten van de ambulanceplanning ingaan, geven we een korte beschrijving van hoe de ambulancesector in Nederland is georganiseerd. De organisatie van de ambulancezorg is verdeeld over 25 regio's. In elke regio is er een Regiona-

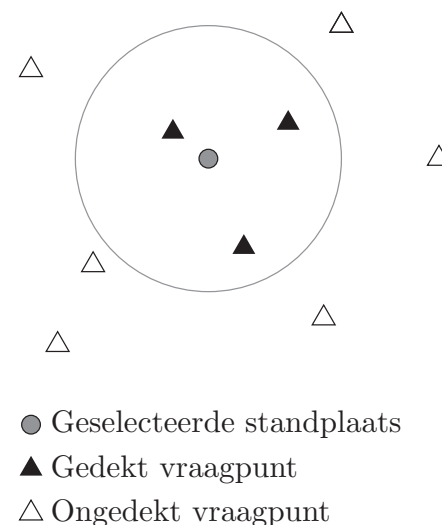
le Ambulancevoorziening (RAV) verantwoordelijk voor het verlenen van de ambulancezorg, zowel voor het meldkamer- als het ritten-domein. De ambulancesector in Nederland is de laatste jaren sterk in beweging. Door de invoering van de Tijdelijke Wet op Ambulancezorg worden de prestaties van de verschillende RAV's gemeten en beoordeeld aan de hand van een aantal prestatie-indicatoren. De door een RAV geleverde prestatie kan gevolgen hebben voor de financiering, waardoor het leveren van topkwaliteit ambulancezorg voor RAV's cruciaal is.

Binnenkomende meldingen worden via een standaardvraagprotocol door de meldkamercentralist geëvalueerd in drie urgentieclassen: A1, A2 en B. Een melding van een ernstige levensbedreigende situatie wordt geclassificeerd als een A1-urgentie, en in dat geval is de richtlijn dat er binnen vijftien minuten hulp ter plaatse moet zijn. Is de situatie minder ernstig, dan wordt de melding als A2 aangemerkt en geldt een responstijd van dertig minuten. Tot slot zijn er B-ritten. Dat zijn planbare ritten waarbij een patiënt bijvoorbeeld vanaf een ziekenhuis naar huis of naar een andere zorginstelling wordt vervoerd. In Nederland is het aantal binnenkomende meldingen ongeveer 1,1 miljoen per jaar, waarvan ongeveer 49 procent als A1 wordt aangemerkt, 24 procent A2 en 27 procent B.

In het vervolg wordt deze Nederlandse situatie als uitgangspunt gebruikt.

Locatieproblemen voor ambulancevervoer

Een van de eerste problemen in de ambulancecelogistiek waarop de wiskunde is toegepast is het bepalen van locaties voor standplaat-



Figuur 1

sen. Het doel van de hiervoor gebruikte modellen is het verlagen van de aanrijtijden naar patiënten. Als prestatie maat wordt hier vaak de dekking binnen een gegeven tijdslicet gebruikt. Deze dekking is de fractie van het aantal oproepen dat binnen de dekkingsnorm wordt bereikt. In Nederland is deze norm vijftien minuten voor inzetten met A1-urgentie.

Bij het definiëren van deze locatiemodellen wordt gebruik gemaakt van een verzameling met vraagpunten V en een verzameling met potentiële standplaatslocaties W . De rijtijd van een standplaats $i \in W$ naar een vraagpunt $j \in V$ wordt bekend verondersteld en wordt genoteerd als t_{ij} . In de rijtijd zit een zogenaamd 'pre-trip delay' ingebakken. Dat is de tijd tussen het binnenkomen van een melding en het moment dat de ambulance het basisstation verlaat. Op basis van deze rijtijden kan voor elk vraagpunt j een verzameling $W_j \subset W$ worden afgeleid die alle potentiële standplaatslocaties bevat die dekking kunnen verlenen aan vraagpunt j . Dus $W_j = \{i \in W \mid t_{ij} \leq r\}$, waarbij r de gestelde dekkingsnorm is. Om de dekking van de verschillende vraagpunten te kunnen wegen, wordt aan elk vraagpunt j een weging d_j toegekend. Voor deze weging zou bijvoorbeeld het aantal inwoners of het verwachte aantal oproepen in een gebied gebruikt kunnen worden. Zie Figuur 1.

De eerste en eenvoudigste modellen die gebruikt worden in de literatuur proberen de enkelvoudige dekking te maximaliseren. Hierbij wordt een gebied als gedekt beschouwd als tenminste één ambulance gestationeerd is binnen de gestelde norm. Dit model staat bekend als het *Maximum Coverage Location Problem* (MCLP) [7]. In dit model hebben we een binaire variabele x_i die waarde 1 aanneemt als er een ambulance is gestationeerd op locatie $i \in W$. Verder geeft de binaire variabele y_j aan of vraagpunt $j \in V$ is gedekt door tenminste één ambulance. Het model maximaliseert de dekking die behaald kan worden met maximaal p ambulances. Het model is als volgt:

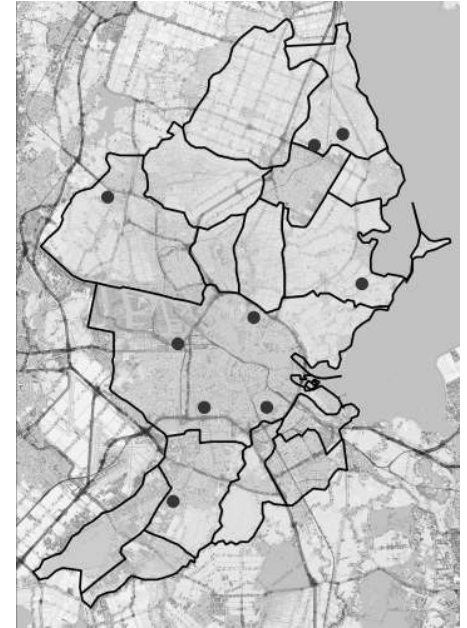
$$\begin{aligned} \max \quad & \sum_{j \in V} d_j y_j \\ \text{s.t.} \quad & \sum_{i \in W_j} x_i \geq y_j \quad (\forall j \in V), \\ & \sum_{i \in W} x_i = p, \\ & x_i \in \{0, 1\} \quad (\forall i \in W), \\ & y_j \in \{0, 1\} \quad (\forall j \in V). \end{aligned}$$

De eerste voorwaarde dwingt af dat de variabele y_j enkel waarde 1 krijgt als er daadwerkelijk een ambulance in staat is dit gebied te dekken. De tweede voorwaarde beperkt het aantal gebruikte standplaatsen. Dit relatief simpele model is NP-moeilijk en is een speciaal geval van het maximaliseren van een submodulaire functie. Ondanks de complexiteitsstatus blijkt dit model voor grote instanties nog goed oplosbaar. Het model is voor meer dan 10.000 gebieden oplosbaar binnen redelijke rekentijd. Dit terwijl grote RAV's in Nederland meestal niet meer dan 500 gebieden bevatten. Een gebied is typisch een postcodegebied als we de vier cijfers van de postcode beschouwen en is dan zowel een vraagpunt als een potentiële standplaatslocatie.

Hoewel MCLP vaak gebruikt is voor het bepalen van standplaatslocaties, is de toepasbaarheid beperkt als naast de standplaatsen ook de ambulanceverdeling bepaald dient te worden. In dat geval gaat MCLP ervan uit dat een ambulance altijd op zijn post aanwezig is, zodat een enkele ambulance voldoende is voor volledige dekking. In de praktijk is dit niet het geval en, met name in stedelijk gebied, is het van groot belang back-updekking te garanderen. We presenteren twee modellen die als maat voor de verwachte beschikbaarheid van de ambulances gebruik maken van het concept van een bezettingsgraad. Hier is de bezettingsgraad gedefinieerd als de verwachte totale werklast gedeeld door de totale beschikbare capaciteit.

Het eerste model dat deze bezettingsgraad q gebruikt is het *Maximum Availability Location Problem* (MALP) [21], waarbij met behulp van q wordt berekend hoeveel ambulances nodig zijn om met betrouwbaarheid α een ambulance beschikbaar te hebben. Door aan te nemen dat ambulances onafhankelijk van elkaar beschikbaar zijn, geeft dit dat $b = \lceil \frac{\log(1-\alpha)}{\log q} \rceil$ ambulances nodig zijn. Vervolgens wordt de dekking met tenminste b ambulances gemaximaliseerd.

Een andere aanpak is om de verwachte dekking, op basis van de bezettingsgraad q , mee te nemen in de doelfunctie. Dit wordt gedaan in het *Maximum Expected Coverage Location Problem* (MEXCLP) [8], waarbij de kans dat tenminste één ambulance beschikbaar is binnen de norm wordt gemaximaliseerd. Gegeven een dekking met k ambulances is de verwachte dekking $E_k = 1 - q^k$. Voor modelleerdoeleinden wordt vaak gebruik gemaakt van de marginale dekking van de k -de ambulance, gegeven door $E_k - E_{k-1} = q^{k-1}(1 - q)$. Dit geeft de volgende doelstellingsfunctie:



Figuur 2 In de linkerplaattegrond worden de huidige standplaatsen weergegeven. Rechts zien we de standplaatsen die het resultaat zijn van het oplossen van MEXCLP met tijdsafhankelijke input [3] met de extra eis dat elk gebied door tenminste één ambulance gedekt moet zijn.

$$\max \sum_{j \in V} \sum_{k=1}^p d_j (1 - q)^{k-1} y_{jk}.$$

Hierbij is y_{jk} een binaire variabele die aangeeft of vraagpunt j met tenminste k ambulances is gedekt, en x_i representeert nu het aantal ambulances op standplaats i . Deze variabele is niet langer binair, omdat meer dan één ambulance op een standplaats gestationeerd kan worden. Om de variabelen y_{jk} de juiste waarde te geven wordt de volgende voorwaarde toegevoegd:

$$\sum_{i \in W_j} x_i \geq \sum_{k=1}^p y_{jk} (\forall j \in V).$$

Er zijn vele uitbreidingen mogelijk van de genoemde modellen. Een mogelijke uitbreiding is om rekening te houden met gedifferentieerd vervoer. Veel ambulancediensten werken niet alleen met auto's die patiënten kunnen vervoeren, maar ook met zogenaamde 'rapid responders' die snel naar de patiënt kunnen komen om hem te stabiliseren in afwachting op vervoer naar het ziekenhuis. Het is makkelijk om het model uit te breiden met verscheidene voertuigtypen met erbij behorende responstijden. Voor het bepalen van ambulancstandplaatsen heeft dit tot nu toe geen hoge prioriteit gehad, maar deze uitbreiding wordt wel besproken in de volgende paragraaf over uitbreidingen naar de brandweeren en verderop in de paragraaf over besteld

vervoer. Een belangrijke uitbreiding die wel is onderzocht binnen REPRO is om rekening te houden met tijdsafhankelijkheid in de input. In [3] zijn tijdsafhankelijke aspecten toegevoegd aan het model MEXCLP. Wij geven in Figuur 2 resultaten van het oplossen van het model ontwikkeld in [3] met input van de regio Amsterdam.

Het bepalen van standplaatsen is een strategische vraag binnen de ambulanceplanning, en rekentijden mogen dan lang zijn zonder dat het gebruik van de modellen wordt belemmerd. Alle besproken modellen zijn NP-moeilijk en we kunnen daardoor verwachten dat rekentijden behoorlijk oplopen als de instanties groter worden. Onze ervaring is echter dat dit meevalt, in ieder geval als we problemen voor een regio oplossen en niet voor heel Nederland. Om een voorbeeld te noemen: als we het model MEXCLP nemen met input van de grootste regio in Nederland, regio Groningen, krijgen we een instantie met 456 geheeltallige variabelen, 8208 binaire variabelen en 457 constraints. Op een laptop kan je die instantie oplossen in ongeveer één minuut rekentijd met behulp van het pakket IBM CPLEX, dat gebruik maakt van een combinatie van *branch-and-bound* en *cutting planes*.

Voor meer referenties verwijzen we naar het overzichtsartikel [16].

Locatieproblemen voor brandweerkazernes

De problematiek van het bepalen van de optimale locaties van de standplaatsen speelt ook bij de brandweer. Typische vragen zijn:

“Wat zijn de optimale locaties van de brandweerkazernes?” en “Hoeveel brandweerwagens van de verschillende types moeten op elk van de kazernes gestationeerd zijn, zodanig dat de dekking zo groot mogelijk is?” Er zijn echter een paar belangrijke verschillen tussen de ambulance en de brandweer. Ten eerste zijn er veel minder incidenten waarbij de brandweer moet uitrukken, waardoor het zelden voorkomt dat meer dan één brandweerincident tegelijkertijd plaatsvindt. Ten tweede heeft de brandweer te maken met verschillende typen voertuigen, die afhankelijk van het type incident ingezet kunnen worden. Ten derde zijn de maximale aanrijtijden bij de brandweer veel kleiner dan ingeval van ambulances.

Niet als in de vorige sectie is V de verzameling vraagpunten van mogelijke incidentlocaties en W de verzameling van potentiële basislocaties. De verzameling beschikbare voertuigtypen noemen we K . In totaal zijn er c_k voertuigen van type $k \in K$ beschikbaar. Aan elk vraagpunt $j \in V$ en voor elk voertuigtype $k \in K$ wordt een gewicht d_{jk} toegekend dat het relatieve belang aangeeft van het dekken van vraagpunt j door een voertuig van type k .

Voor gegeven vraagpunt $j \in V$ en voertuigtype $k \in K$ definiëren we $W_{jk} \subset W$ als de verzameling van alle locaties die vraagpunt j dekken door voertuigtype k . De vraag of vraagpunt j wordt gedekt door basislocatie $i \in W$ en voertuigtype $k \in K$ hangt af van twee aspecten: (1) de dekkingsnorm r_{jk} , de maximaal toelaatbare aanrijtijd voor vraagpunt $j \in V$ en voertuigtype $k \in K$, en (2) t_{ij} , de rijtijd van i naar j inclusief de pre-trip delay. We nemen aan dat de rijtijden niet afhangen van het voertuigtype en dat de pre-trip delay vast is. Dus geldt voor $j \in V$ en $k \in K$ dat $i \in W_{jk}$ als en alleen als $t_{ij} \leq r_{jk}$.

In het model onderscheiden we drie typen variabelen. De variabele x_{ik} is het aantal voertuigen van type k dat gesitueerd is op potentiële basislocatie i ; omdat we in dit model geen meervoudige dekking meenemen geldt dat $x_{ik} \in \{0, 1\}$. De variabele $y_{jk} \in \{0, 1\}$ neemt waarde 1 aan als vraagpunt j gedekt is door voertuigtype k . Tot slot definiëren we $z_i := 1$ als en alleen als basislocatie i wordt gebruikt door tenminste één voertuigtype, en $z_i := 0$ anders. Als een basislocatie i wordt gebruikt moeten we kosten β betalen. Het gebruik van hetzelfde basisstation voor verschillende voertuigtypen wordt hiermee gestimuleerd.

Het planingsprobleem kan als volgt worden geformuleerd als een geheel-tallig lineair optimaliseringsprobleem:

$$\begin{aligned} \max \quad & \sum_{j \in V} \sum_{k \in K} d_{jk} y_{jk} - \beta \sum_{i \in W} z_i \\ \text{s.t.} \quad & \sum_{i \in W_{jk}} x_{ik} \geq y_{jk}, \quad j \in V, k \in K, \\ & \sum_{i \in W} x_{ik} \leq c_k, \quad k \in K, \\ & x_{ik} \leq z_i, \quad i \in W, k \in K, \\ & y_{jk}, z_i, x_{ik} \in \{0, 1\}. \end{aligned}$$

De doelfunctie bestaat uit twee delen: het gewogen aantal meldingen die ‘gedekt’ zijn door het juiste voertuigtype, en kosten voor het aantal gebruikte basisstations. De eerste restrictie geeft aan dat vraagpunt j alleen gedekt is door voertuigtype k als tenminste één voertuig van type k voldoet aan de dekkingsnorm. De tweede restrictie beperkt het aantal voertuigen van elk type. Als gevolg van de derde restrictie kunnen we alleen maar een voertuig hebben op een basisstation dat geopend is.

In de gegeven formulering modelleert β de kosten van het in gebruik nemen van een basisstation. Deze parameter definieert de trade-off tussen de extra dekking als gevolg van het openen van een nieuwe basislocatie enerzijds en de daarmee gepaard gaande onderhoudskosten anderzijds. De waarde van β kan grofweg worden geïnterpreteerd als het aantal voorheen niet-gedekte incidentmeldingen die wel gedekt zouden zijn als een extra basislocatie zou worden geopend. In de

praktijk is het soms lastig om een geschikte waarde voor β te bepalen. In dat geval kunnen we dit deel van de doelfunctie weglaten en in plaats daarvan een restrictie toevoegen op het aan basislocaties dat kan worden geopend.

Het boven gegeven model is NP-moeilijk. Echter, commerciële solvers zoals CPLEX kunnen voor realistische modelinstanties tot een optimale oplossing komen binnen redelijke tijd. Het model zoals beschreven is gebruikt voor het optimaliseren van de locaties en bezetting van de standplaatsen voor de Brandweer Amsterdam-Amstelland. In deze dataset worden vier voertuigtypen onderscheiden: tankautosputten (TS), redvoertuigen (RV), hulpverleningsvoertuigen (HV) en waterongevallenwagens (WO). Het aantal beschikbare voertuigen van de vier typen bedragen respectievelijk 21, 9, 3 en 2. We nemen het aantal inzetten van deze voertuigen in het jaar 2011 als gegeven. In deze periode zijn deze 39.516 keer ingezet, per voertuigtype respectievelijk 29.016, 9.182, 615 en 703. We onderscheiden 2643 vraagpunten, die overeenkomen met de secties die in de praktijk gebruikt worden. Daarnaast onderscheiden we 2223 potentiële basislocaties. De rijtijden t_{ij} tussen de basislocaties en de vraagpunten zijn verkregen van de brandweer en zijn gebaseerd op geschatte reistijden op het wegennet.

Voertuigtype	TS	RV	HV	WO
Aantal beschikbaar	21	9	3	2
Aantal incidenten	29.016	9.182	615	703
Minimum dekkingsnorm	6	6	15	15
Gemiddelde dekkingsnorm	7,98	14,68	15	15
Maximum dekkingsnorm	10	15	15	15

Tabel 1 Overzicht van de gebruikte dataset.

# wijzigingen	Bedeckingsgraad per voertuigtype				
	TS	RV	HV	WO	Totaal
0	87,68	98,23	96,84	88,64	90,83
1	89,99	98,23	96,84	88,64	92,29
2	91,76	99,64	96,84	88,64	93,74
3	93,20	99,64	97,27	89,78	94,76
4	94,38	99,64	96,84	90,68	95,53
Ongelimiteerd	98,62	99,86	98,10	93,37	98,53

Tabel 2 Dekkingsgraad bij verschillend aantal wijzigingen (in procenten).



Figuur 3 Illustratie van fictieve verplaatsingen in de Regionale Eenheid Amsterdam.

Voor voertuigtypes TS en RV gelden wettelijke dekkingsnormen, die variëren tussen 5, 6, 8 en 10 minuten, afhankelijk van het type en de functie van de gebouwen bij de verschillende vraagpunten. Voor type HV geldt een maximale dekkingsnorm van 15 minuten, voor type WO is dit gesteld op 18 minuten. In onze berekeningen zetten we de dekkingsnorm op 15 minuten voor elk vraagpunt voor beide voertuigen. De pre-trip delay is 150 seconden. Tabel 1 geeft een overzicht over de gebruikte data.

Het hierboven beschreven model stelt ons in staat goede locaties van de standplaatsen te berekenen voor een willekeurig gegeven aantal potentiële standplaatsen. Daarnaast bestaat de mogelijkheid, gegeven de negentien kazernes die op dit moment in gebruik zijn, om te bepalen wat het zou opleveren om een gelimiteerd aantal kazernes te wijzigen. Als we één wijziging toestaan bestaat de oplossing uit achttien van de huidige kazernes en één kazerne is mogelijk anders. Tabel 2 laat per voertuigtype de dekkingsgraad zien bij een verschillend aantal wijzigingen. Een wijziging levert al een winst op van 1,46 procentpunt, de twee daaropvolgende eveneens ook telkens ruim één procentpunt. De situatie bij vier wijzigingen is weergegeven in Figuur 3.

De resultaten in Tabel 2 laten zien dat grote winsten in de dekkingsgraad te behalen zijn door het wijzigen van een klein aantal kazernelocaties en het optimaal verdelen van de voertuigen over de kazernes.

Dynamisch Ambulance Management

In de traditionele ambulancedienstverlening heeft elk ambulancevoertuig een vaste standplaats: elke wagen rijdt, wanneer deze vrij-

komt, terug naar zijn *eigen* thuisbasis. Tegenwoordig wordt dit echter vaak gezien als een verouderde aanpak, en beseft men dat het terugsturen van de ambulance naar zijn thuisbasis niet altijd de beste keuze is. Afhankelijk van de situatie op dat moment, bijvoorbeeld de locaties en status van de huidige incidenten en van de andere voertuigen, kan het verplaatsen naar een andere basis beter zijn. Dit is de kern van het *Dynamisch Ambulance Management*, kortweg DAM.

In Nederland is een verandering in paradigma merkbaar: steeds vaker wordt de statische planning voor een, al dan niet door de ambulancevoorzieningen zelf ontworpen, dynamische planning ingeruild. Meestal wordt daarbij gebruik gemaakt van een eenvoudige *schuifregeltabel*, die voor elk aantal wagens aangeeft op welke basislocaties zij zich dienen te bevinden. De schuifregeltabel kan leiden tot *voorwaardeschepende ritten*, kortweg VWS-ritten, waarbij één of meer ambulancevoertuigen naar een andere locatie worden gedirigeerd.

Het grote voordeel van DAM is dat door de proactieve verplaatsingen van voertuigen veel flexibeler kan worden geanticipeerd op toekomstige incidenten, waardoor de responstijden, en daarmee ook de mortaliteit en morbiditeit, kunnen worden gereduceerd. Een veelgenoemd nadeel van DAM is dat ambulances vaker moeten verplaatsen, en dat de werkdruk daardoor wordt verhoogt. Bovendien geldt voor gebieden met veel incidenten dat het aantal vrije ambulances zeer snel kan wijzigen, vaak zo snel dat de situatie al wijzigt voordat de ambulances de nieuwe gewenste configuratie hebben bereikt.

De schuifregeltabelen zijn gebaseerd op *enkelvoudige dekking* van de regio, waarbij ambulances zodanig zijn verspreid over de regio dat de kans dat het *eerstvolgende* incident binnen de vijftienminutenstraal is geoptimaliseerd. Het probleem is echter dat in de praktijk in veel gevallen meerdere incidenten tegelijkertijd moeten worden afgehandeld, waardoor de ‘statische’ schuifregels houden geen rekening met de actuele locaties van de voertuigen, en voorzien niet in een meervoudige dekking.

Om dat probleem op te lossen wordt binnen het REPRO-project onderzoek verricht naar *online optimaliseringsalgoritmen*, waarbij optimale relocaties worden bepaald, bijvoorbeeld voor iedere mogelijke ‘toestand’ van het systeem (zoals incident- en voertuiglocaties). Het probleem hierbij is dat het aantal mogelijke toestanden van het systeem

exponentieel snel stijgt in het aantal voertuigen en incidentlocaties, waardoor de reken tijden voor het bepalen van ‘optimale’ relocaties van voertuigen veel te lang worden voor praktische toepassingen. Uit onderzoek binnen REPRO blijkt dat het mogelijk is om online optimaliseringsalgoritmen te ontwikkelen, die in real-time bijna-optimale relocatieadviezen geven. Een voorbeeld is het *Dynamic Maximum Expected Coverage Location Problem* (D-MEXCLP)-algoritme, geïnspireerd door het MEXCLP-model [8], waarin de bezettingsgraad q wordt meegenomen voor optimaliseren van meervoudige dekking.

Het D-MEXCLP-algoritme [14] beperkt zich in eerste instantie tot het verplaatsen van een ambulance wanneer deze zich vrij meldt. Hierbij wordt slechts een beperkt deel van de real-time data meegenomen, namelijk enkel de *bestemmingen* van andere vrije ambulances. Dit wordt afgezet tegen de verwachte behoefte aan ambulances.

We bepalen de gewenste locatie voor de zojuist vrijgekomen ambulance door te berekenen op welke standplaats de ambulance de *grootste toename in bedekking* zou geven volgens het MEXCLP-model. Voor de notatie, zie de eerste paragraaf. Het model definieert de waarde van het toevoegen van een k -de ambulance die vraagpunt $j \in V$ op tijd kan bereiken als $d_j(1-q)q^{k-1}$. De zo gedefinieerde strategie $\pi(\underline{n})$, met $\underline{n} = (n_1, \dots, n_{|W|})$, kan genoteerd worden aan de hand van het aantal vrije ambulances n_x dat bestemming x heeft. Merk op dat vrije ambulances altijd stilstaan op een basis, of daar naar onderweg zijn, dus het volstaat om $x \in W$ te beschouwen. Voor gegeven voertuigconfiguratie \underline{n} stuurt het D-MEXCLP-algoritme een vrijkomende ambulance naar de bestemming $w \in W$ waarvoor de verwachte toename in de totale dekkingsgraad

$$\sum_{j \in V} d_j(1-q)q^{k(j,w,\underline{n})-1}$$

maximaal is, waarbij

$$k(j, w, \underline{n}) := \sum_{i=1}^{|W|} n_i \cdot \mathbf{1}(t_{ij} \leq r) + \mathbf{1}(t_{wj} \leq r),$$

het aantal voertuigen dat knoop j kan bereiken binnen r minuten, voor gegeven voertuigconfiguratie \underline{n} , aannemende dat de vrijkomende ambulance instantaan op de knoop van bestemming $w \in W$ is.

Doordat de keuzemogelijkheden beperkt zijn (in de praktijk is $|W|$ meestal niet groter dan 30 of 40), kunnen we de optimale

bestemming 'brute force' uitrekenen. De D-MEXCLP-oplossing is uitgetest voor een realistische dataset van de provincie Utrecht. De resultaten laten zien dat door het gebruik van D-MEXCLP het aantal tijdsoverschrijdingen met ongeveer 15–20 procent kan worden gereduceerd ten opzichte van de optimale statische oplossing [14].

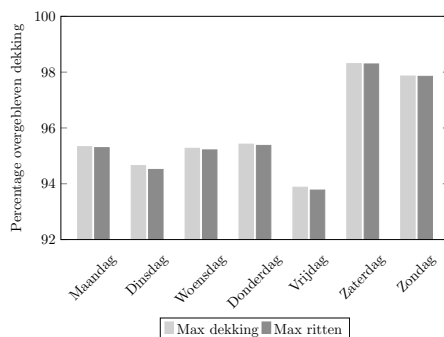
De modellen voor DAM worden momenteel uitgetest in een pilot die wordt uitgevoerd in samenwerking met GGD Flevoland, CityGIS en het CWI.

Besteld vervoer

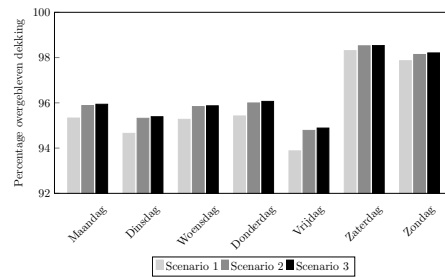
Ambulances worden naast spoedvervoer ook gebruikt voor besteld vervoer. Bij besteld vervoer wordt er een afspraak gemaakt voor vervoer tussen het woon- of verblijfadres van een patiënt of zorginstellingen. In het geval van een behandeling of onderzoek in een ziekenhuis wordt deze afspraak vaak ruim van tevoren gemaakt. Bij ontslag uit of opname in het ziekenhuis wordt deze afspraak pas kort van tevoren gemaakt. Dit betekent dat een deel van de afspraken ruim op tijd kan worden ingepland, maar dat deze planning verstoord wordt door afspraken die later binnenkomen. Hierdoor moet deze planning gedurende de dag worden aangepast zodat alle afspraken kunnen worden nagekomen.

Als er een afspraak voor besteld vervoer gemaakt wordt, wordt ook een gewenste ophaaltijd van de patiënt doorgegeven. Bij het maken van de planning wordt zoveel mogelijk rekening gehouden met deze gewenste tijd, maar hier kan niet altijd aan worden voldaan. Vaak proberen ambulancediensten in het tijdsinterval van een uur voor en een uur na de gewenste tijd de patiënt op te halen.

De meeste ambulancediensten in Nederland maken gebruik van verschillende soorten ambulances, waaronder *Advanced Life Support* (ALS)-ambulances en *Basic Life Support* (BLS)-ambulances. Voor spoedvervoer kunnen alleen ALS-ambulances worden inge-



Figuur 4 Overgebleven verwachte dekking voor spoedvervoer.



Figuur 5 Overgebleven verwachte dekking voor spoedvervoer.

zet welke bemand worden door een ambulancechauffeur en een ambulanceverpleegkundige. Voor besteld vervoer kunnen zowel ALS- en BLS-ambulances worden ingezet waarbij een BLS-ambulance eenvoudiger uitgerust is dan een ALS-ambulance en wordt bemand door een chauffeur en een basisverpleegkundige. Voor sommige patiënten is de medische indicatie zodanig dat een ALS-ambulance nodig is bij besteld vervoer. Als de patiënt geen ALS-ambulance nodig heeft probeert men zoveel mogelijk besteld vervoer uit te voeren met BLS-ambulances, maar soms is deze capaciteit niet toereikend. Omdat voldoende ALS-ambulances nodig zijn om een korte responstijd voor spoedvervoer te garanderen, is het noodzakelijk om het gebruik van ALS-ambulances voor besteld vervoer goed te plannen. Dit betekent dat het bepalen van de routes voor de BLS-ambulances en de inzet van ALS-ambulances voor besteld vervoer op zo'n manier moet gebeuren dat de verwachte dekking van de ALS-ambulances voor spoedvervoer zo hoog mogelijk blijft.

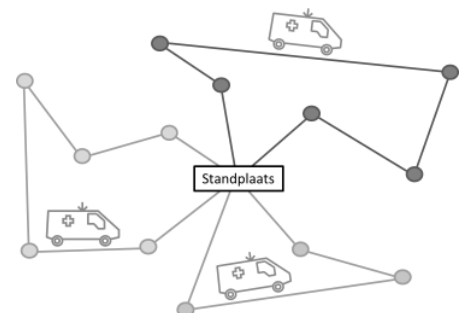
Aangezien een groot deel van de afspraken pas gedurende de dag wordt gemaakt, hebben we een *geheeltallig lineair optimaliseringsprobleem* geformuleerd [4] dat wordt opgelost zodra er een nieuwe afspraak binnenkomt. Dit model neemt alle tot dan toe beschikbare informatie mee en bepaalt de planning voor het besteld vervoer. De planning van de ritten die afgerond of bezig zijn staat vast, maar de planning van toekomstige afspraken mag worden gewijzigd. De doel functie van het model is het maximaliseren van de overgebleven dekking voor spoedvervoer. Het model is ingewikkeld en we geven daarom hier een beschrijving van de beslissingen die volgen uit het oplossen van het model:

- Toewijzing van besteld vervoer aan een ALS- of BLS-ambulance.
- Bij gebruik van ALS-ambulance voor besteld vervoer: wijs de afspraak toe aan een bepaalde standplaats en een bepaalde tijd.

- Bij gebruik van BLS-ambulance voor besteld vervoer: bepaal de routes voor de verschillende BLS-ambulances waarbij rekening wordt gehouden met de gewenste ophaaltijd van de patiënt.

Een voor de hand liggende aanpak is om het aantal ritten uitgevoerd met een BLS-ambulance te maximaliseren om zo de overgebleven dekking voor spoedvervoer te maximaliseren. Een andere aanpak is om de dekking expliciet mee te nemen, bijvoorbeeld met behulp van de doelfunctie van MEXCLP [8]. Om deze twee aanpakken met elkaar te vergelijken, kijken we naar de overgebleven dekking voor spoedvervoer als percentage van de dekking wanneer er geen besteld vervoer wordt uitgevoerd met ALS-ambulances. De berekeningen zijn gemaakt voor RAV Utrecht. Figuur 4 laat zien dat het expliciet meenemen van de dekking resulteert in een hogere overgebleven dekking, terwijl het aantal ritten uitgevoerd met een BLS-ambulance afneemt van 91,7 procent naar 90,2 procent vergeleken met het maximaliseren van het aantal met BLS uitgevoerde ritten.

Figuur 5 laat zien wat de invloed op de overgebleven dekking is van het feit dat niet alle afspraken 's ochtends bekend zijn. Hiertoe bekijken we drie verschillende scenario's. Het eerste scenario is de reële setting waarbij afspraken gedurende de dag bekend worden en daarna pas kunnen worden ingepland. Bij dit scenario kan 90,2 procent van de afspraken met een BLS-ambulance worden uitgevoerd. Het tweede scenario is de situatie waarin we veronderstellen dat alle afspraken 's morgens al bekend zijn en dus al ingepland kunnen worden. In dit scenario zit nog wel de restrictie dat afspraken pas na het tijdstip waarop ze aangevraagd worden mogen worden uitgevoerd. Dit betekent dat we alsnog flexibiliteit verliezen omdat het voor kan komen dat een afspraak niet een uur voor de gewenste ophaaltijd kan worden ingepland omdat de rit toen nog niet was aangevraagd.



Figuur 6

Doordat er wordt verondersteld dat alle afspraken 's ochtends al bekend zijn, neemt het percentage ritten uitgevoerd met een BLS-ambulance toe tot 95,5 procent. In het derde scenario hebben we de meeste flexibiliteit omdat in dit scenario niet alleen wordt verondersteld dat alle afspraken bekend zijn in de morgen, maar dat we ook voor alle afspraken onbeperkte flexibiliteit hebben. Dit betekent dat alle afspraken in het interval van een uur voor tot een uur na de gewenste ophaaltijd kunnen worden ingepland. Dit verhoogt het aantal ritten dat uitgevoerd kan worden met een BLS-ambulance tot 97,1 procent.

Uit Figuur 5 blijkt dat het eerder maken van afspraken een grote invloed heeft op de overgebleven dekking van ALS-ambulances voor spoedvervoer. Deze resultaten kunnen door de ambulancediensten gebruikt worden om de ziekenhuizen te overtuigen om eerder te bellen als er een ontslag of opname gepland staat. Een volgende stap is om de gebruikte oplossingsmethode te implementeren in de praktijk om zo de efficiëntie van de BLS-ambulances en daarmee de overgebleven dekking voor spoedvervoer te verhogen. Ons model is tot stand gekomen na intensieve samenwerking met RAV Utrecht.

Simulatie

Aanbieders van ambulancediensten moeten veel strategische beslissingen nemen. "Hoeveel beter presteert de ambulancedienst in een gebied als we daar één extra ambulance inzetten?", "Wat is de invloed door het toevoegen van de beste plek voor een nieuwe standplaats die uit locatieproblemen volgt?" of "Wat is nu echt de toegevoegde waarde van het introduceren van DAM-algoritmen op de meldkamer?" Dit soort vragen kunnen vaak beantwoord worden met behulp van simulaties.

In het kader van REPRO is een simulatiemodel ontwikkeld onder de naam *Testing Interface For Ambulance Research*, kortweg TIFAR [5]. In deze sectie wordt de werking van TIFAR toegelicht. Eerst kijken we naar de gegevens die we als in- en uitvoer gebruiken, en daarna beschrijven we de diverse objecten van het simulatiemodel en hoe die samenwerken.

Vanwege onze intensieve samenwerking met verscheidene ambulancediensten in Nederland zijn er veel bronnen beschikbaar gesteld die gebruikt kunnen worden als *invoer* van het simulatiemodel. Een van de belangrijkste bronnen voor de nulmeting zijn de coördinaten van de huidige standplaatsen en de huidige ambulance- en meldkamerroos-

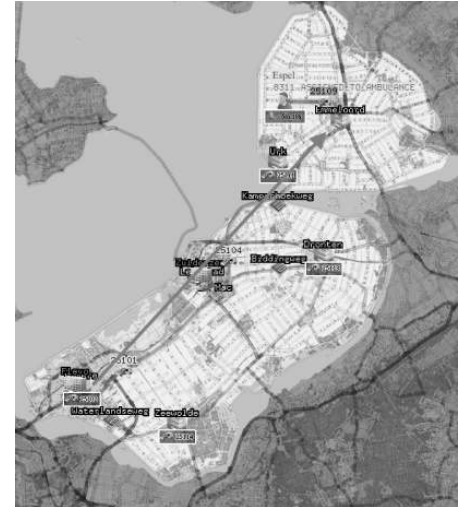
ters. Door het meenemen van de roosters wordt op elk tijdstip de juiste hoeveelheden ambulances in de berekening meegenomen.

De *zorgvraag* wordt geaggregeerd op vier positie postcodeniveau (4pp); dit zijn de cijfers van het postcodegebied waarin het incident plaatsheeft. Voor elk postcodegebied weten we de coördinaten van de centroïde. Het RIVM heeft een rijtijdentabel beschikbaar gesteld die aangeeft hoe lang de reistijd voor ambulances is tussen elk koppel van 4pp-centroïdes bij spoedvervoer in de spits. Deze tabel is nuttig bij het berekenen van locatieproblemen, en DAM-algoritmes die onderdeel uitmaken van relocatievoorstellen. City-GIS levert een andere *routeplanner* die speciaal gemaakt is voor ambulancediensten. Met deze routeplanner valt de reistijd, de afstand en het traject tussen twee coördinaten op de kaart op te vragen. Dat is handig, want in de regel bevinden ambulances zich niet op de centroïde van een postcodegebied. Deze twee routeplanners zijn speciaal afgestemd op ambulancezorg, en houden rekening met onder meer aangepaste snelheid, busbanen, fietspaden en speciaal voor hulpdiensten aangelegde afritten.

Naast deze geografische eigenschappen en roosters hebben de ambulancediensten de logistieke keuzes van de meldkamercentralisten en ambulances in beslisseregels gevat en beschikbaar gesteld. Hierover wordt later in de tekst nog teruggekomen.

De laatste en waarschijnlijk belangrijkste dataset zijn de *historische ritgegevens*. Vanaf 2008 weten we van elke rit de urgentie die toegekend is, het ambulancenummer, de vertrekplaats van de ambulance, de incidentlocatie, de bezorglocatie en de standplaats waarnaar de ambulance teruggekeerd is. Deze dataset bevat ook de tijdstippen, rijtijden, en behandel- en bezorgduur voor elke rit. Vrijwel alle parameters die nodig zijn om modellen te vullen vallen hieruit te halen, in het bijzonder de zorgvraag per 4pp. Merk op dat een alternatief hiervoor dat beter kan zijn bij groeiende steden, zijn demografische voorspellingen door het CBS. Uit de ritgegevens worden ook de prestatie-maten berekend, waaronder de fractie A1-ritten met een responstijd van onder de vijftien minuten.

De *uitvoer* van de simulatiesoftware is een nieuwe ritdatabase die precies dezelfde structuur heeft als de historische ritdatabase. Hierdoor kunnen de simulatieresultaten goed vergeleken worden met historische prestaties. Het is zelfs mogelijk om naderhand andere prestatie-maten te introduceren en vergelijken, wat vooral voordelen heeft in de commu-



Figuur 7 Illustratie van het gebruikersinterface van simulatie-tool TIFAR.

nicatie met managers die soms aanvullende informatie vragen.

De *werking* van het simulatiemodel is als volgt. Initieel worden alle ambulances leeg op hun standplaats gezet en zijn er geen incidenten in het systeem. Vanaf hier wordt een *discrete event simulation* (DES) uitgevoerd. De computer houdt een geordende lijst bij met alle tijdstippen waarop een beslissing genomen moet worden. Vervolgens wordt van een beslismoment naar het eerstvolgende beslismoment vooruit in de tijd gesprongen, en alleen op deze discrete tijdstippen wordt de toestand van het systeem herberekend. Een herberekening bestaat uit de volgende opeenvolgende stappen: incidentgeneratie, ritvernieuwing, meldkamervernieuwing, ambulancevernieuwing, en uiteindelijk de relocatieprocedure. Direct na het updaten van een parameter begint het hele herberekeningsproces op hetzelfde tijdstip van voor af aan. Uiteraard zijn er in de implementatie wel optimalisatieslagen gedaan om de rekentijd kort te houden. Figuur 7 geeft een illustratie van de gebruikersinterface van TIFAR, waarbij een proactieve relocatie wordt voorgesteld.

Een *incidentengenerator* maakt incidenten aan op 4pp's, wanneer er geen toekomstige incidenten in de wachtrij staan en voor zolang de simulatieduur nog niet verstreken is. Zowel de incidentlocatie, het tijdstip van binnenkomst, als de diverse andere tijdsduren zijn stochastisch bepaald. Hierbij is de historische ritdatabase een goede bron voor de parameterbepaling van de verdelingen.

Bij de *ritvernieuwing* worden ritten die op dat moment binnenkomen als actief aange-merkt, en afgesloten ritten in de outputdatabase weggeschreven en het bijbehorende werkgeheugen schoongemaakt.

Indien ervoor gekozen is om de *meldkamerprocessen* in de simulatie mee te nemen, worden telefoongesprekken in dit proces gekoppeld aan of ontkoppeld van de centralisten. Vaak wordt ervoor gekozen om de meldkamer niet in dit detail mee te nemen, aangezien de focus dan ligt op het wegdomein. In dat geval beperkt de meldkamer zich tot het uitgeven van een rit, en het ogenblikkelijk beantwoorden van vragen door ambulances: normaliter wordt bij een uitgifte de dichtstbijzijnde beschikbare ambulance gestuurd, gaat de ambulance naar het dichtstbijzijnde ziekenhuis, en keert ze terug naar de standplaats waar de dienst begonnen is. Later in deze tekst zullen we een meldkamermodel uitvoerig behandelen.

In de *ambulancevernieuwing* wordt de locatie van elke ambulance vernieuwd. De ambulance is verantwoordelijk voor het statussen van een rit, net als in de werkelijkheid. Dit betekent dat hij aangeeft wanneer hij aankomt op zijn bestemming, wanneer een patiënt behandeld is, of wanneer de overdracht bij de bestemming van de patiënt voltooid is. De meldkamer blijft verantwoordelijk voor het aansturen van alle logistieke processen, dus voor het nemen van dit type beslissingen neemt een ambulance altijd contact op met de meldkamer. Elke vernieuwingsslag wordt beëindigd met een *relocatie-*

berekening. Hiervoor kunnen onder andere DAM-methodes gebruikt worden.

In alle stappen heeft degene die de simulatie uitvoert veel vrijheid. Hierdoor kunnen dus veel vragen nauwkeurig beantwoord worden, zonder dat de ambulancedienst eerst geheel op een andere werkwijze hoeft over te schakelen. Een keuze die gemaakt kan worden is om een simulatie *trace-driven* uit te voeren; in dat geval gebeuren de incidenten op dezelfde tijdstippen als in de historie, waarbij ook de behandelduur, overdrachtstijd en bestemming uit de historische ritdatabase gehaald wordt. Vergelijken met een historische nulmeting is goed voor een duidelijke beeldvorming omtrent de implicaties van het wijzigen van een beleidsregel.

Simulatie is een beproefde methode die een belangrijke ondersteunende rol in een zorgvuldig beslisproces vervult. Door in detail te simuleren wordt het effect van diverse beleidsaanpassingen inzichtelijk gemaakt.

Meldkamermodel

De *meldkamer ambulancezorg* (MKA) is de plek waar de aanvragen voor ambulancezorg binnenkomen, waarvandaan de voertuigen worden aangestuurd. De MKA vervult de rol als communicatieknoppunt tussen ambulanceverpleegkundigen, ziekenhuizen en andere professionals die betrokkenen zijn

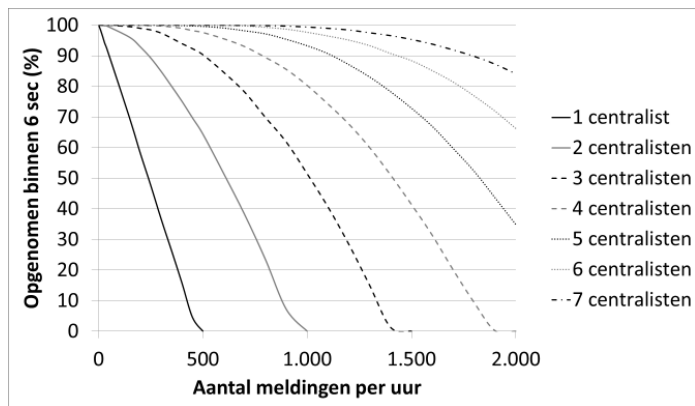
bij de operationele ambulancezorg. Op de meldkamer zijn er meerdere inkomende lijnen via welke ritten worden aangevraagd. Ongeveer 40 tot 50 procent van de aanvragen voor spoedeisende ambulancezorg betreft een melding via het landelijke telefoonnummer 1-1-2. Daarnaast zijn er veel aanvragen die via een huisarts lopen. Het besteld vervoer wordt door ziekenhuizen of andere zorginstellingen aangevraagd.

In de Nederlandse MKA's vindt sinds 2010 schaalvergroting plaats waarbij meerdere kleinere meldkamers samengaan tot een grotere organisatie. Ook hebben nieuwe triage-systemen hun intrede gedaan waardoor de uitvraag en beoordeling van spoedeisende ambulancezorg meer geprotocolleerd plaatsvindt. Door deze ontwikkelingen is er een behoefte gekomen aan een simulatiemodel dat de processen op de meldkamer nabootst en waarmee het presteren van de MKA kan worden onderzocht. Ons simulatiemodel TIFAR kan gebruikt worden om de processen op de meldkamer te simuleren. Het model omvat de processen rondom de aanvragen en uitgiftes voor zowel spoedeisende ambulancezorg als besteld vervoer.

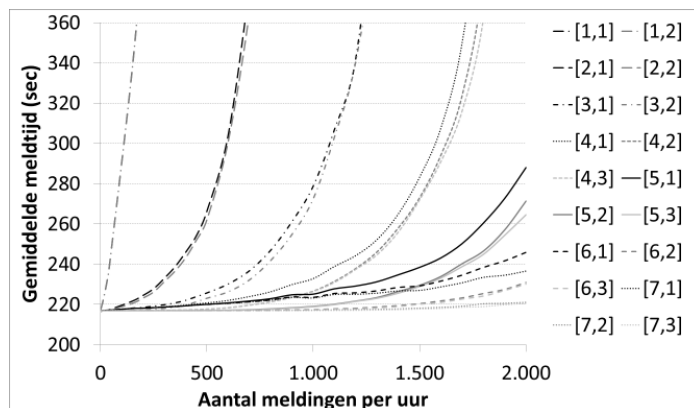
Het voert te ver om het model in detail te bespreken. Daarom geven we in deze paragraaf een kort overzicht van de belangrijkste elementen in het binnen REPRO ontwikkelde



Foto: Till Krech



Figuur 8 Aantal meldingen binnen zes seconden bij toenemend aantal meldingen per dag.



Figuur 9 Doorlooptijd als functie van het gemiddelde aantal meldingen per dag ([centralisten, uitgifte-centralisten]).

MKA-model. De geïnteresseerde lezer wordt verwezen naar [6] voor meer details.

Op een MKA werken *aanname- en uitgifte-centralisten*. Een aanvraag voor ambulancezorg wordt door een aanname-centralist beoordeeld op urgentie, en bij een indicatie voor het verlenen van ambulancezorg wordt een ritopdracht uitgegeven door een uitgifte-centralist. Deze geeft het meest geschikte ambulanceteam de opdracht om naar het incident te rijden, ter plaatse zorg te verlenen en, indien nodig, vervoer naar een ziekenhuis te bieden. Veelal is dit het ambulanceteam dat het snelste ter plaatse kan zijn.

In het geval van een spoedeisende inzet gaat een ambulance met optische en geluidsignalen (OGS) naar de locatie van het incident en verleent daar levensreddende zorg. Beide centralisten kunnen op meerdere momenten tijdens de inzet communiceren: de aanname-centralist met de melder, bijvoorbeeld om instructie voor eerste hulp te geven of om de patiënt in veiligheid te brengen. De uitgifte-centralist kan een contact hebben om vragen van het ambulanceteam te beantwoorden of de komst van de patiënt met het ziekenhuis te communiceren. We noemen deze vervolcontacten van de uitgifte-centralist 'terugkoppelmomenten'.

Indien het druk wordt op de meldkamer, wordt er geprioriteerd. Voor de aanname-centralist hebben de 1-1-2-aanvragen en spoedeisende aanvragen van huisartsen de hoogste prioriteit aangezien het dan om een levensbedreigende situatie kan gaan. Het besteld vervoer heeft de laagste urgentie omdat deze aanvragen vaak niet direct respons nodig hebben. We gaan er in ons model van uit dat alle taken 'geduldig' zijn en het systeem niet zullen verlaten tot ze afgehandeld zijn. In werkelijkheid kan een brandweer- of politie-centralist een aanvraag voor ambulancezorg oppakken als alle ambulance-centralisten

bezet zijn. Voor de uitgifte-centralist heeft het uitgeven van spoedritten een hogere prioriteit dan de terugkoppelmomenten. Bij gelijke prioriteit hebben langer wachtenden altijd voorrang.

Het MKA-model onderscheidt zeven verschillende inkomende lijnen voor ambulancezorg. Afhankelijk van de lijn heeft de aanvraag een hoge, middelmatige of lage prioriteit. Gedurende de loop van het gesprek bepaalt de aanname-centralist of ambulancezorg gewenst is en kent aan de melding een urgentie toe. In sommige gevallen is de urgentie a priori bekend doordat de inkomende lijn alleen voor een bepaald soort meldingen gebruikt wordt. We veronderstellen dat nieuwe aanvragen middels een Poisson-proces optreden. De frequentie $\lambda_{i,u}$ en gespreksduur $\mu_{i,u}$ van de aanvragen zijn afhankelijk van de binnenkomende lijn i en urgentie u , en kunnen ook in de tijd variëren. Tijdens het uitvoeren van de inzet door het ambulanceteam zijn er verschillende contactmomenten met de meldkamer. We gaan ervan uit dat er maximaal één contactmoment is bij iedere status van de ambulance, waaronder aanrijdend, behandelend, bezorgend en overdragend. Het contactmoment laten we altijd plaatsvinden aan het einde van de status. Door deze keuzes wordt de kans op een contactmoment tijdens elke status een Bernoulli-proces, waarin we uitgaan van aannames van meldkamerexperts voor de parameterwaarden. De verdeling voor de duur van elke status en bijbehorende parameters worden verkregen uit de ritdatabase.

De bedieningsduurverdelingen zijn geschat aan de hand van een dataset die is opgebouwd uit een koppeling tussen gegevens uit de telefooncentrale van de meldkamer, de arbi-data, en de ritgegevens van de ambulance-inzetter. De arbi-data bevatten geen kenmerken van de inzetter, zoals

de aanvrager van de inzet en of bij een aanvraag daadwerkelijk een inzet heeft plaatsgevonden. Gegevens over de periode april-juni 2012 waren beschikbaar voor ons onderzoek; dit betrof gegevens van 109.000 telefonische contacten waaruit 41.000 ambulanceinzetten hebben plaatsgevonden.

Met de uitkomsten van het simulatiemodel kan het functioneren van de meldkamer worden geanalyseerd onder verschillende werklasten en met verschillende centralisten-aantallen. Belangrijke prestatie-indicatoren hierbij zijn de wachttijd voordat de telefoon wordt opgenomen, de totale doorlooptijd van de meldkamer voor elk van de verschillende soorten telefoongesprekken, en de tijd die nodig is voor de uitgifte van een inzet.

Ter illustratie is het simulatiemodel TIFAR gebruikt om het presteren van een meldkamer te analyseren bij een veranderende werkdruk. Het aantal aanvragen is stapsgewijs verhoogd van $\lambda = 100$ naar 2000 aanvragen per dag, waarbij de verhouding tussen het aantal aanvragen dat elke inkomende lijn levert gelijk blijft. Hierbij is per λ in een aantal stappen het aantal centralisten verhoogd. Bij een vast aantal centralisten en een toenemend aantal aanvragen neemt de wachttijd voor de aanvrager toe: deze zal vaker in een wachtrij komen en de melding zal niet direct kunnen worden beantwoord. Als vuistregel geldt een prestatienorm dat een binnenkomende 1-1-2-aanvraag binnen zes seconden moet worden opgenomen. Wij hanteren dat tenminste 95 procent van de hulpverzoeken aan deze aanvragen moet voldoen.

Door de scheiding in functie ondervinden de aanname-centralisten geen nadeel door eventuele vertragingen bij uitgifte-centralisten. In Figuur 8 zien we dat tussen de 50 en 700 meldingen per dag twee aanname-centralisten voldoende zijn, terwijl je tot 1600

meldingen per dag met drie aankomsten. Hierin zijn schaalvoordelen al zichtbaar.

De meldkamer-doorlooptijd van een aanvraag is de tijdsduur vanaf het opnemen van een telefoon tot het moment van een opdracht aan een ambulance. De doorlooptijd is sterk afhankelijk van het aantal centralisten en de verhouding tussen het aantal aanname- en het aantal uitgifte-centralisten. Bij een gegeven λ kan gezocht worden naar het optimale aantal centralisten waarbij de doorlooptijd nog acceptabel is. In Figuur 9 zien we dat tot ongeveer 1200 meldingen per dag één

uitgifte-centralist voldoende zal zijn. Dit valt te zien doordat de uitgifte-centralist hier niet de beperkende factor is op de doorlooptijd. Tot 2000 meldingen per dag zien we dat een derde niet nodig is.

Door enkele keuzes in het modelleren wordt er een minimale doorlooptijd van 220 seconden weergegeven, in plaats van ongeveer twee minuten die in de praktijk geldt. Dit komt doordat enkele parallelle processen in ons model sequentieel uitgevoerd worden. Deze keuze heeft echter geen invloed op de werklust van de centralisten en daarmee de

toename in de doorlooptijd, waardoor conclusies over de toename in werklust en doorloop veilig getrokken kunnen worden.

Het simulatiemodel is een krachtig middel om het presteren van een MKA te analyseren zonder dat er grootschalige organisatorische veranderingen nodig zijn. Door simulatie kunnen fictieve situaties worden doorgerekend, bijvoorbeeld de situatie dat meldkamers worden samengevoegd en het volume van aanvragen toeneemt. De relatie tussen het aantal aanvragen en het aantal benodigde centralisten kan inzichtelijk gemaakt worden. ◀

Referenties

- 1 L. Aboueljinane, E. Sahin en Z. Jemaï, A review on simulation models applied to emergency medical service operations, *Computers & Industrial Engineering* 66(4) (2013), 734–750.
- 2 R. Alanis, A. Ingolfsson en B. Kolfal, A Markov chain model for an EMS system with repositioning, *Production and Operations Management* 22(1) (2013), 216–231.
- 3 P.L. van den Berg en K.I. Aardal, Time-dependent MEXCLP with start-up and relocation cost, *European Journal of Operational Research* 242(2) (2015), 383–389.
- 4 P.L. van den Berg en J.T. van Essen, Scheduling non-urgent patient transport while maximizing emergency coverage, submitted, 2015.
- 5 M. van Buuren, R.D. van der Mei, K.I. Aardal en H. Post, Evaluating Dynamic Dispatch Strategies for Emergency Medical Services: TIFAR Simulation Tool, in *Proceedings of the Winter Simulation Conference*, 2012.
- 6 M. van Buuren, G.J. Kommer, R.D. van der Mei en S. Bhulai, A simulation model for Emergency Medical Service call centers, te verschijnen in *Proceedings Winter Simulation Congress, Los Angeles*, 2015.
- 7 R. Church en C. S. ReVelle, The maximal coverage location problem, *Papers in Regional Science* 32(1) (1974), 101–118.
- 8 M.S. Daskin, A maximum expected covering location model: formulation, properties and heuristic solution, *Transportation Science* 17(1) (1983), 48–70.
- 9 M. Dzator en J. Dzator, An effective heuristic for the p -median problem with application to ambulance location, *OPSEARCH* 50(1) (2013), 60–74.
- 10 E. Erkut, A. Ingolfsson en G. Erdoğan, Ambulance location for maximum survival, *Naval Research Logistics* 55(1) (2008), 42–58.
- 11 M. Gendreau, G. Laporte en F. Semet, Solving an ambulance location model by tabu search, *Location Science* 5(2) (1997), 75–88.
- 12 K. Hogan en C.S. ReVelle, Concepts and applications of backup coverage, *Management Science* 32(11) (1986), 1434–1444.
- 13 A. Ingolfsson, S. Budge en E. Erkut, Optimal ambulance location with random delays and travel times, *Health Care Management Science* 11(3) (2008), 262–274.
- 14 C.J. Jagtenberg, S. Bhulai en R.D. van der Mei, An efficient heuristic for real-time ambulance redeployment, *Operations Research for Health Care* 4 (2015), 27–35.
- 15 V.A. Knight, P.R. Harper en L. Smith, Ambulance allocation for maximal survival with heterogeneous outcome measures, *Omega* 40(6) (2012), 918–926.
- 16 X. Li, Z. Zhao, X. Zhu en T. Wyatt, Covering models and optimization techniques for emergency response facility location and planning: a review, *Mathematical Methods of Operations Research* 74(3) (2011), 281–310.
- 17 M.S. Maxwell, S.G. Henderson en H. Topaloglu, Tuning approximate dynamic programming policies for ambulance redeployment via direct search, *Stochastic Systems* 3(2) (2013), 322–361.
- 18 M.S. Maxwell, M. Restrepo, S.G. Henderson en H. Topaloglu, Covering models and optimization techniques for emergency response facility location and planning: a review, *Mathematical Methods of Operations Research* 74(3) (2011), 281–310.
- 19 W.B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, John Wiley & Sons, 2011.
- 20 J.F. Repede en J.J. Bernardo, Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky, *European Journal of Operational Research* 75(3) (1994), 567–581.
- 21 C.S. ReVelle en K. Hogan, The maximum availability location problem, *Transportation Science* 23(3) (1989), 192.
- 22 V. Schmid en K.F. Doerner, Ambulance location and relocation problems with time-dependent travel times, *European Journal of Operational Research* 207(3) (2010), 1293–1303.
- 23 O. Zhang, A.J. Mason en A.B. Philpott, Simulation and optimisation for ambulance logistics and relocation, presentatie op *INFORMS 2008 Conference*.