

Rob van der Mei

KPN Research, afdeling Network Planning
2260 AK Leidschendam
R.D.vanderMei@kpn.com

Overzichtsartikel

Het modelleren van netwerkverkeer

De afgelopen jaren is Informatie- en Communicatietechnologie (ICT) sterk in opkomst, en anno 2000 is telecommunicatie niet meer uit het dagelijks leven weg te denken. Telecommunicatienetwerken moeten daardoor steeds grotere hoeveelheden informatie transporteren, hetgeen kan leiden tot opstoppingen in het netwerk en vermindering van de kwaliteit die de eindgebruiker ervaart. Om verkeersproblemen te voorspellen, te voorkomen en op te lossen is het belangrijk modellen te ontwikkelen die het stochastische gedrag van het verkeer in het netwerk beschrijven. In dit verhaal wordt een kort overzicht gegeven van de ontwikkelingen op het gebied van verkeersmodelling in telecommunicatienetwerken.

Telecommunicatie is een vak met een rijke geschiedenis. De eerste vormen van communicatie over afstand hebben een sterk militaire component. De toneelschrijver Aeschylus (vijfde eeuw voor Christus) beschrijft al dat de val van Troje op snelle wijze door de overwinnaars aan Zeus wordt doorgegeven. Via grote vuren op enkel bergtoppen wordt de overwinning aan de oppergod gemeld. Op deze wijze kan snel een afstand van 500 kilometer worden overbrugd. Er is hier sprake van een zeer vroege vorm van digitale optische communicatie. Een verbeterd optisch communicatiesysteem vinden we in Europa aan het eind van de 18e eeuw. Het systeem is gebaseerd op een vinding van de Fransman Claude Chappe. Op torens worden beweegbare armen geplaatst. Iedere stand van de armen heeft een bepaalde betekenis. Op deze wijze kan teken voor teken een bericht worden doorgegeven. Dit systeem staat bekend als de luchttelegraaf. Een bericht van vijftwintig woorden doet er zes uur over om van Straatsburg naar Parijs te worden getransporteerd. Met de uitvinding van het Morse-toestel rond 1836 ontstaat er een nieuwe en veel snellere communicatiestructuur, waarmee ook transatlantische communicatie

mogelijk wordt. Hoewel deze communicatiemogelijkheid voornamelijk door overheden wordt gebruikt, ontstaat er ook een publieke dienst die voor commerciële doeleinden kan worden ingezet. Snelle communicatie blijkt een lucratieve faciliteit. Rond 1877 vindt Graham Bell de telefoon uit, die onder de naam klanken-telegraaf in Nederland wordt geïntroduceerd. In de jaren twintig doet de radio zijn intrede, rond 1930 komt de televisie op de markt en in de jaren zestig begint mobiele telefonie op te komen.

In de jaren zeventig zien we een nieuwe ontwikkeling: de computer doet zijn intrede in het communicatieveld. Via op afstand gekoppelde terminals kan de computer zonder tussenkomst van een computercentrum rechtstreeks worden gebruikt. Bovendien komen er minicomputers, en later zelfs microcomputers. Uiteindelijk resulteren de technische ontwikkelingen in de PC zoals we die nu kennen. Er ontstaat vervolgens behoefte aan telecommunicatie tussen deze systemen: de datacommunicatie doet zijn intrede. Door het groeiend gebruik van computers neemt het belang van datacommunicatie verder toe. Het gebruik van modems maakt het mogelijk telefonienetten te gebruiken voor datacommunicatie.

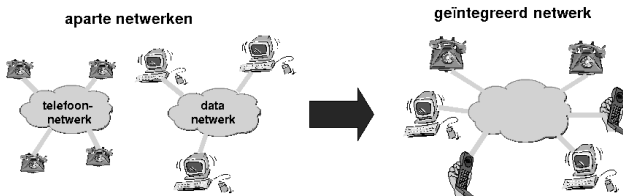
Door de opkomst van het Internet aan het eind van de jaren tachtig zijn de ontwikkelingen op het gebied van datacommunicatie nog eens versneld. Uiteindelijk is het de universitaire wereld die de mogelijkheden van de computer maximaal benut om nieuwe communicatievormen en netstructuren te ontwikkelen. Het resultaat is de opkomst van multimedia-diensten, een communicatievorm waarin spraak, tekst en beeld zijn geïntegreerd. Het Internet wordt echter pas echt *booming* in de jaren negentig met de doorbraak van het World-Wide Web (WWW) en Electronic mail. Anno 2000 is het Internet niet meer uit het dagelijks leven weg te denken.

1 Technologische ontwikkelingen

De opkomst van ICT heeft grote gevolgen voor de maatschappij: voor het bedrijfsleven, de overheid en voor particulieren. Aan deze ontwikkeling ligt een aantal technologische ontwikkelingen ten grondslag, zoals de sterke toename van de rekenkracht van computers, de beschikbare hoeveelheid opslagruimte (geheugen, disk-ruimte) en de transmissiecapaciteit (bandbreedte), en de digitalisering van audio en video. Daarnaast zien we nog een aantal trends op het gebied van netwerken, zoals hun onderlinge integratie, het ontstaan van nieuwe netwerkprotocollen en de overgang van circuit- naar pakketgeschakelde netwerken.

1.1 Integratie van netwerken

Een belangrijke ontwikkeling die aan de basis heeft gestaan van de opkomst van multimedia-diensten is de integratie van netwerken. Tot die tijd werden voor het leveren van de meeste diensten *aparte* netwerken aangelegd. Zo bestond de infrastructuur uit een verzameling van netwerken, voor verschillende toepassingen als telefoon-, data-, telex-, huurlijnen- en autotelefoonverkeer. Integratie van deze aparte netwerken heeft geleid tot de ontwikkeling van *Broadband Integrated Services Digital Networks (B-ISDNs)*, breedbandige geïntegreerde digitale netwerken. B-ISDNs maken het mogelijk allerlei nieuwe diensten te leveren, zowel voor de consumentenmarkt (betaal-TV, telewerken, beeldtelefoon, videovergaderen) als de zakelijke markt (digitale huurlijnen, koppeling van Local Area Networks, bandwidth-on-demand).

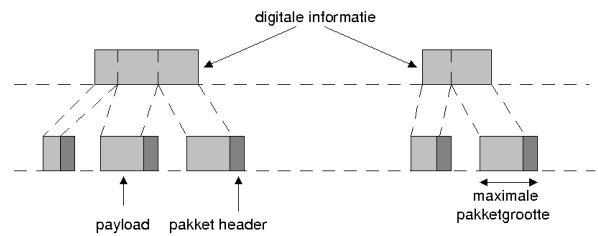


Figuur 1 Integratie van netwerken

1.2 Netwerkprotocollen

De integratie van netwerken leidt tot de ontwikkeling van nieuwe *netwerkprotocollen*, noodzakelijk om informatie over een netwerk te kunnen transporteren. Een protocol is een verzameling afspraken, bijvoorbeeld over het formaat waarin een bericht verstuurd wordt en de betekenis van die berichten. Om communicatie over een netwerken voor iedereen mogelijk te maken, zijn protocollen gedefinieerd op verschillende abstractie-niveaus, en gestandaardiseerd in het zogenaamde Open System Interconnect (OSI) referentiemodel. Voorbeelden van protocollen zijn HyperText Transfer Protocol (HTTP), File Transfer Protocol (FTP), Transfer Control Protocol (TCP), Internet Protocol (IP) en Asynchronous Transfer Mode (ATM). Figuur 2 geeft een vereenvoudigde weergave van het gebruik van een netwerkprotocol: een hoeveelheid digitale informatie wordt opgesplitst in een aantal pakketten (de maximale lengte van een pakket hangt af van het netwerkprotocol), die vervolgens over het fysieke netwerk (in nullen en enen) naar hun bestemming worden gestuurd. Een pakket bestaat meestal uit twee delen: de *pakket-header* (met informatie over het protocol, de bestemming van het pakket, enzovoorts) en de *payload* (de daadwer-

kelijke digitale informatie). Figuur 2 illustreert het gebruik van netwerkprotocollen.

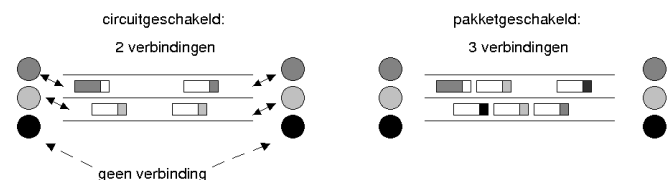


Figuur 2 Protocollen: digitale informatie wordt in pakketten opgedeeld en over het netwerk verstuurd

1.3 Overgang van circuit- naar pakketgeschakelde netwerken

Het traditionele telefoonnetwerk is een zogenaamd *circuitgeschakeld* netwerk: om een telefoonverbinding tussen twee eindpunten mogelijk te maken wordt een bi-directionele verbinding (circuit) opgezet die actief blijft totdat de telefoonverbinding wordt afgebroken. Daarbij wordt voor iedere verbinding een *vaste hoeveelheid bandbreedte* in het netwerk gereserveerd. Voor standaard telefonie is dat 64 kilobits per seconde. In pakketgeschakelde netwerken daarentegen wordt niet vantevoren een vaste hoeveelheid bandbreedte gereserveerd, maar wordt bandbreedte alleen gebruikt wanneer er daadwerkelijk pakketten verstuurd worden. Het verschil tussen circuitgeschakelde en pakketgeschakelde netwerken vindt een aardige analogie met het wegnetwerk. In circuitgeschakelde netwerken heeft iedere weggebruiker een gereserveerde rijstrook ter beschikking, terwijl in pakketgeschakelde netwerken meerdere weggebruikers op dezelfde rijstrook kunnen rijden.

Deze analogie met het wegnetwerk maakt ook meteen de wederzijdse voor- en nadelen duidelijk. Het grote voordeel van circuitgeschakelde netwerken is dat er feitelijk geen verkeersproblemen (congestie) in het netwerk voorkomen, terwijl er in pakketgeschakelde netwerken de kans op verkeersopstoppingen aanwezig is, zeker wanneer het netwerk zwaar belast is. De prijs die circuitgeschakelde netwerken hiervoor moeten betalen is verlies aan efficiëntie van het gebruik van transportcapaciteit. Zo kunnen in een pakketgeschakeld netwerk in principe meer verbindingen worden toegelaten. Figuur 3 illustreert een situatie waarin drie verbinden zijn aangevraagd. In een circuitgeschakelde implementatie kunnen slechts twee verbindingen tegelijk actief zijn, terwijl in een pakketgeschakelde implementatie alle drie de verbindingen tegelijk actief kunnen zijn.



Figuur 3 Circuit- versus pakketgeschakelde netwerken

Het verschil in efficiëntie tussen circuit- en pakketgeschakelde netwerken manifesteert zich vooral wanneer het verkeer dat door een applicatie gegenereerd wordt *grillig* is, dat wil zeggen dat de applicatie perioden van hoge en lage activiteit afwisselt. De meeste moderne applicaties, zoals World-Wide Web, Electronic mail en Video-on-Demand, genereren grillig verkeer. Pakketgeschakelde netwerken zijn dan ook geschikter voor het ondersteunen van de meeste moderne applicaties dan circuitgeschakelde netwerken. Dat is dan ook een van de redenen dat pakketgeschakelde netwerken de laatste jaren steeds populairder zijn geworden.

2 Tijdsschalen

De verkeerskarakteristieken in telecommunicatienetwerken hangen af van de tijdsschaal waarop het verkeer beschouwd wordt. Voor het gemak onderscheiden we drie tijdsschalen:

- call-level (minuten)
- burst-level (seconden)
- pakket-level (milli-seconden).

Op *call*-level worden verbindingen of sessies opgestart en afgebroken. De tijdsschaal is in de orde van minuten. Voorbeelden van calls zijn 'gewone' telefoongesprekken, fax-sessies, of verbindingen met het Internet. Op kleinere tijdsschaal fluctueert de hoeveelheid verkeer die gegenereerd wordt tijdens een call: bij veel applicaties worden bursts van verkeer gegenereerd. Op *burst*-level fluctueert het verkeersaanbod daardoor tussen zogenaamde ON-perioden waarin een burst verkeer gegenereerd wordt, en OFF-perioden waarin geen (of nauwelijks) verkeer gegenereerd wordt. Tijdens een WWW-sessie, bijvoorbeeld, haalt de gebruiker van tijd tot tijd informatie via het WWW op, leest de informatie, haalt nieuwe informatie op, enzovoorts. Zo fluctueert het verkeer op een tijdsschaal van seconden tussen ON- en OFF-perioden. Op de nog kleinere tijdsschaal van milli-seconden is een burst van verkeer weer op te delen in een aantal pakketten die over het netwerk getransporteerd worden. De tijdsschaal waarop het verkeer wordt beschouwd als een serie pakketten wordt *pakket*-level genoemd.

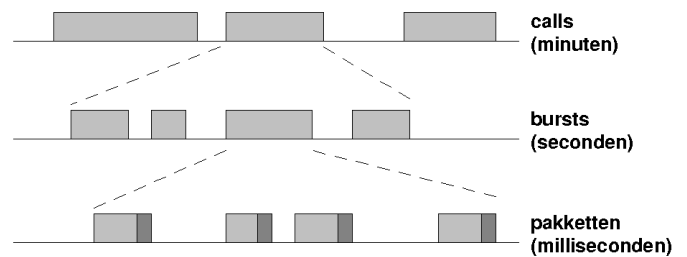
Een eerste stap in een wiskundige analyse van telecommunicatienetwerken is het opstellen van een model dat het stochastische gedrag van het verkeer beschrijft. In dit artikel beperken we ons tot het bespreken van de gangbare modellen. Het zou te ver voeren ook in te gaan op de analyse van deze modellen.

3 Verkeersmodellering van telefoonnetwerken

Het call-level verkeer in telefoonnetwerken wordt gekarakteriseerd door twee entiteiten: het aankomstproces van inkomende calls, en de verdeling van het lengte van een call. Merk op dat verkeersmodellering of burst- en pakket-level hier niet relevant is, omdat het telefoonnetwerk circuitgeschakeld is.

3.1 Aankomstproces van calls

Verkeersmetingen in telefoonnetwerken hebben experimenteel aangetoond dat calls over een tijdperiode t ($t = 1, 2, \dots$) bij benadering worden gegenereerd volgens een homogeen Poisson-



Figuur 4 Verschillende tijdsschalen

proces met aankomstintensiteit $\lambda(t)$. Dat wil zeggen dat de tijden tussen opeenvolgende calls onderling onafhankelijk en exponentieel verdeeld zijn met gemiddelde $1/\lambda(t)$: als X de tijd tussen twee opeenvolgende calls is, dan geldt voor $t = 1, 2, \dots$,

$$\text{Prob}\{X > s\} = e^{-s\lambda(t)} \quad (s > 0).$$

Een interessante eigenschap van de exponentiële verdeling is dat deze *geheugenloos* is, dat wil zeggen

$$\text{Prob}\{X > s + t \mid X > s\} = \text{Prob}\{X > t\} \quad (s, t > 0).$$

De geheugenloosheidseigenschap van de exponentiële verdeling, die aan het Poisson-proces ten grondslag ligt, is in overeenstemming met de intuïtie dat telefoongesprekken 'volledig random' gegenereerd worden. Stel bijvoorbeeld dat het gemiddeld aantal telefoongesprekken in een uur 60 is, maar dat gedurende twee minuten geen calls zijn aangekomen. Dan zegt deze informatie niets over hoe lang we nog moeten wachten totdat de volgende call aankomt.

Merk op dat het gemiddeld aantal calls per tijdseenheid over het algemeen varieert van uur tot uur, van dag tot dag, van week tot week, van maand tot maand, etcetera. Het tijdsinterval t waarover $\lambda(t)$ constant verondersteld kan worden is in de orde van grootte van een kwartier tot een uur. In de praktijk is $\lambda(t)$, het gemiddeld aantal calls ter tijdseenheid voor tijdsinterval t ($t = 1, 2, \dots$) vrij nauwkeurig te voorspellen op grond van historische gegevens.

3.2 Lengte van een call

Traditioneel wordt aangenomen dat de gespreksduren in telefoonnetwerken ook exponentieel verdeeld zijn. Recentelijke verkeersmetingen in telefoonnetwerken hebben echter (experimenteel) aangetoond dat de gespreksduur een kansverdeling met een *dikke staart* heeft.

Definitie. Een niet-negatieve stochastische variabele X heeft een *dikke staart* als

$$\text{Prob}\{X > t\} = c(t)t^{-\alpha}, \quad t \rightarrow \infty, \quad 1 < \alpha < 2,$$

waarbij $c(\cdot)$ de eigenschap heeft dat $\lim_{t \rightarrow \infty} c(tx)/c(t) = 1$ voor alle $x > 0$. De functie $c(\cdot)$ wordt ook wel *langzaam variërend* bij oneindig genoemd. Voorbeelden van langzaam variërende functies zijn de constante functie $c(t) = c_0$ en de logaritmische functie $c(t) = \log(t)$. De bovengenoemde stochastische variabele X wordt ook wel *regulier variërend met index α* genoemd.

De intuïtie achter stochastische variabelen met dikke staarten is dat de kans op uitschieters relatief groot is ten opzichte van kansvariabelen met dunne staarten (dat wil zeggen met een exponentieel dalende staartkansen voor $t \rightarrow \infty$). Merk op dat onder de bovengenoemde definitie een stochastische variabele met een dikke staart een eindige verwachting, maar een oneindige variantie, heeft.

De vraag rijst *waarom* de duur van een telefoongesprek een dikke staart heeft. Een interessante verklaring wordt gegeven door Bolotin [4], die stelt dat aan de lengte van telefoongesprekken een *logaritmisch principe* ten grondslag ligt. Het idee daarachter is dat tijdens conversaties de menselijke perceptie van tijd de Wet van Weber volgt, die zegt dat de 'gelijkwaardige' perceptie van verschillen wordt gemeten op een logaritmische schaal: *'just noticeable difference' in a stimulus bears a constant ratio to the stimulus*. In de context van telefoonconversaties: een uitbreiding van een gesprek van 30 naar 40 seconden 'voelt hetzelfde' als een uitbreiding van een gesprek van 3 naar 4 minuten.

4 Verkeersmodellering in B-ISDN netwerken

Verkeerskarakterisering in B-ISDN netwerken is veel complexer dan in het standaard circuitgeschakelde telefonienetwerk. De complexiteit is het gevolg van een groot aantal factoren die invloed uitoefenen op de verkeerskarakteristieken. B-ISDN netwerken ondersteunen een verscheidenheid aan applicaties, zoals World-Wide Web (WWW), remote login, TELNET, File Transfer Protocol (FTP), en het Simple Mail Transfer Protocol (SMTP). Deze applicaties maken op hun beurt weer gebruik van allerlei verschillende netwerkprotocollen, zoals Transfer Control Protocol (TCP), Internet Protocol (IP), Asynchronous Transfer Mode (ATM) en Ethernet, die elk hun specifieke eigenschappen hebben. Tenslotte veranderen de eigenschappen van het verkeer ook nog door allerlei Traffic Management en flow-control mechanismen die in het netwerk geïmplementeerd zijn om het verkeer te stroomlijnen. Als gevolg van al deze invloeden is het verkeer in B-ISDN netwerken vaak onoverzichtelijk en buitengewoon complex. De verkeersmodellering in B-ISDN netwerken moet dan ook in deze context beschouwd worden.

4.1 Call-level karakterisering

Op grond van gedetailleerde verkeersmetingen komen Paxson en Floyd [9] tot de intuïtief duidelijke conclusie dat het aankomstproces van *door de gebruiker geïnitieerde* applicaties nauwkeurig kan worden beschreven door middel van een homogeen Poissonproces, waarbij de intensiteit over verschillende tijdsperioden kan variëren. Deze observatie is onafhankelijk van het soort call. Het maakt dus bijvoorbeeld niet uit of het WWW-, remote login-, TELNET- of FTP-sessies betreft.

In talloze empirische studies op grond van verkeersmetingen is aangetoond dat de lengte van de 'calls' dikke staarten hebben, waarbij de specifieke parameters verschillen per applicatie en gebruikersgroep. Merk op dat Bolotins verklaring gebaseerd op een logaritmisch principe hier ook toepasbaar is (zie sectie 3.2). Interessant is het daarbij op te merken dat de resultaten in [4] het voorkomen van dikke staarten aantonen voor bijvoorbeeld de

lengte van een voice-mail bericht, de transmissietijd voor het versturen van een fax, de tijd dat een Internet-gebruiker ingelogd is, en de lengte van een E-mail bericht. De details van de genoemde empirische resultaten worden besproken in [4-5].

4.2 Burst-level karakterisering

De burst-level karakterisering hangt af van de specifieke applicatie, maar bestaat doorgaans uit een vernieuwingsproces van alternerende ON en OFF-perioden. De verdeling van de lengte van ON- en OFF-perioden, en de transmissiesnelheden tijdens de ON- en OFF-perioden verschilt per applicatie, en hangt ook af van bijvoorbeeld de beschikbare hoeveelheid bandbreedte in het access netwerk (bijvoorbeeld de modemsnelheid) en de hoeveelheid informatie die tijdens ON-perioden wordt getransporteerd.

Voor de WWW applicatie alterneert de gebruiker typisch tussen ON-perioden waarin een informatie van het WWW wordt opgehaald, en OFF-perioden waarin de informatie verwerkt wordt. Op grond van meerdere empirische studies is aangetoond dat de hoeveelheid informatie opgehaald tijdens een ON-periode (zeg de file-grootte) een dikke staart heeft (zie bijvoorbeeld [2]). Vergelijkbare resultaten gelden voor andere applicaties, zoals FTP en E-mail.

4.3 Pakket-level karakterisering

Het opstellen van een gedetailleerde karakterisering van het pakket-level verkeer van een applicatie in een B-ISDN netwerk is zeer gecompliceerd ten gevolge van een verscheidenheid aan factoren, zoals Traffic Management mechanismen (traffic shaping, traffic policing, Random Early Discard), flow control mechanismen (TCP slow start, retransmissions, congestion avoidance [11]), prioriteits- en schedulingmechanismen (Weighted Fair Queueing, Generalized Processor Sharing), enzovoorts. Al deze factoren zorgen ervoor dat de karakteristieken van een verkeersstroom gegeneerd door een bepaalde applicatie sterk kunnen veranderen wanneer het verkeer door het netwerk stroomt. Deze complicerende factoren maken een gedetailleerde per-applicatie karakterisering van het verkeer op pakket-level praktisch onmogelijk.

5 Long-Range Dependence en schaalinvariantie

Op grond van uitgebreide verkeersmetingen is empirisch aangetoond dat B-ISDN netwerken een aantal intrigerende eigenschappen bezitten die in de traditionele verkeersmodellen niet voorkomen: long-range dependence en schaalinvariantie. De definities en interpretaties van deze eigenschappen worden in deze sectie besproken.

5.1 Long-range dependence

Definitie. Zij $X = \{X_t, t > 0\}$ een stationair stochastisch proces met autocovariantie-functie

$$r_X(s) := \text{Cov}\{X_t, X_{t+s}\} \quad (s > 0),$$

onafhankelijk van t . Dan heet X *long-range dependent* (LRD) als

$$\int_{s=0}^{\infty} r_X(s) ds = \infty. \quad (1)$$

Anders heet X *short-range dependent* (SRD).

De niet-integreerbaarheid van de autocovariantie-functie weerspiegelt de intuïtie achter LRD, dat ondanks het feit dat de autocovarianties $r_X(s)$ klein zijn voor grote waarden van s , het cumulatieve effect van deze autocovarianties significant is. Het fenomeen LRD wordt ook wel het *Jozef-effect* genoemd, refererend naar de persoon in het Oude Testament die de droom van de farao over 'de zeven magere en zeven vette koeien' uitlegde als 'zeven magere en zeven vette jaren' dat het oude Egypte in het vooruitzicht had. Intuïtief zegt long-range dependence iets over tijdsafhankelijkheden die in veel natuurlijke processen voorkomen, waarbij clusters van grote en kleine waarden voorkomen.

Een alternatieve karakterisering van LRD die in de literatuur op het gebied van telecommunicatienetwerken vaak wordt gebruikt is dat de autocovariantie-functie $r_X(s)$ hyperbolisch naar 0 convergeert voor grote waarden van s :

$$r_X(s) = c(s)s^{-\alpha} \quad (s \rightarrow \infty), \tag{2}$$

met $0 < \alpha < 1$, en waarbij $c(\cdot)$ een langzaam variërende functie bij oneindig is. SRD processen daarentegen worden in deze context ook wel gekarakteriseerd door

$$r_X(s) = c(s)e^{-\eta s} \quad (s \rightarrow \infty),$$

met $\eta > 0$, ofwel door een autocovariantie-functie die exponentieel snel naar 0 convergeert voor grote waarden van s .

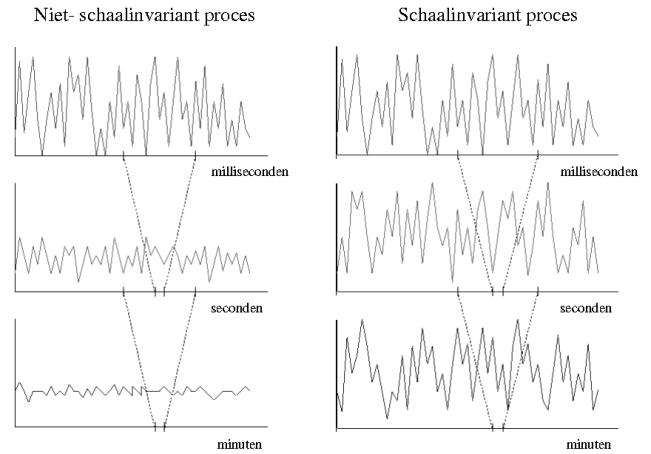
Vanuit puur mathematisch oogpunt is zijn definities van LRD in (1) en (2) dubieus. In het algemeen hoeft een autocovariantiefunctie, noch de integraal in (1), te convergeren — neem bijvoorbeeld een periodiek proces. Binnen het toepassingsgebied van telecommunicatienetwerken is de praktische relevantie van de verschillen tussen bovengenoemde definities van LRD echter beperkt. In het vervolg zullen de alternatieve definities dan ook door elkaar worden gebruikt.

Het fenomeen LRD manifesteert zich op verschillende manieren. Daarom zijn er naast de bovengenoemde definities van LRD equivalente karakteriseringen van LRD. Bijvoorbeeld, in het frequentiedomein manifesteert LRD zich in een singulariteit van de spectraal-dichtheid in $\lambda = 0$:

$$f_X(\lambda) := \int_{s=0}^{\infty} r_X(s)e^{-is\lambda} ds = L(\lambda)\lambda^{-\gamma} \quad (\lambda \rightarrow 0),$$

waarbij $0 < \gamma < 1$, en waarbij $L(\cdot)$ langzaam variërend bij 0 is. Omdat $f_X(0) = \int_0^{\infty} r_X(s)ds$, wordt LRD gekarakteriseerd als $f_X(0) = \infty$. In dat geval spreekt men wel van $1/f$ -noise. Voor SRD processen daarentegen geldt dat de spectraal-dichtheid naar een eindige limiet convergeert als $\lambda \rightarrow 0$.

Een eenvoudig voorbeeld van een LRD proces is een simpel ON-OFF proces met ON-perioden van oneindige variantie. Preciezer, zij $X = \{X_t, t > 0\}$ een alternerend ON-OFF proces met onderling onafhankelijke ON- en OFF-perioden ter lengte L_{ON} en L_{OFF} , met $E[L_{ON}] < \infty$ en $E[L_{OFF}] < \infty$. Dan is het eenvoudig na te gaan dat als $\text{Var}[L_{ON}] = \infty$, dan geldt $\int_0^{\infty} \text{Cov}\{X_0, X_s\}ds = \infty$, en dus is X LRD. Merk op dat omgekeerd als $\text{Var}[L_{ON}] < \infty$, dan geldt $\int_0^{\infty} \text{Cov}\{X_0, X_s\}ds < \infty$, en dus is X SRD.



Figuur 5 Schaalinvariantie

5.2 Schaalinvariantie

Recente metingen van het verkeer in B-ISDN netwerken hebben nog een andere eigenschap van het verkeer aan het licht gebracht: *schaalinvariantie*. De verkeerskarakteristieken zijn hetzelfde over meerdere tijdsschalen: de hoeveelheid gemeten verkeer over een tijdsinterval ter lengte T laat een grillig verloop zien, zowel voor kleine T (milli-seconden) als grote T (uren). Figuur 5 illustreert het verschil tussen schaalinvariante en niet-schaalvariante processen door de gemiddelde hoeveelheid verkeer over verschillende tijdsschalen uit te zetten als functie van de tijd.

Het is interessant op te merken dat traditionele verkeersmodellen in de literatuur in het algemeen niet schaalinvariant zijn. Neem bijvoorbeeld het Poisson-proces, het traditionele verkeersmodel voor telefoonverkeer. Zij X_T het aantal aankomsten in een tijdsinterval ter lengte T , dan is X_T Poisson verdeeld gemiddelde λT , waarbij λ de aankomstintensiteit is. Dan is eenvoudig in te zien dat $X_T/T \xrightarrow{d} \lambda$ voor $T \rightarrow \infty$, dat wil zeggen dat de hoeveelheid verkeer 'uittendelt' over langere tijdsintervallen. Het Poisson-proces is dus niet schaalinvariant.

Definitie. Zij $X = \{X_t, t > 0\}$ een continu-tijds-proces met drift 0 en stationaire aangroeiingen. Dan heet X (strict) *schaalinvariant* wanneer voor alle $a > 0$ geldt

$$X_{at} \stackrel{d}{=} a^H X_t \quad (t > 0),$$

waarbij de notatie " $\stackrel{d}{=}$ " staat voor gelijkheid voor alle eindig-dimensionale verdelingen. X heet *asymptotisch schaalinvariant* als X_{at} (genormaliseerd) zwak convergeert naar een strict schaalinvariant proces voor $a \rightarrow \infty$.

In de context van telecommunicatienetwerken wordt vaak gekeken naar de iets zwakkere eigenschap van tweede-orde schaalinvariantie. X heet *tweede-orde schaalinvariant* als voor elke $a > 0$ X_{at} en $a^H X_t$ dezelfde tweede orde karakteristieken hebben. Tenslotte heet X *asymptotisch tweede orde schaalinvariant* als de tweede orde karakteristieken X_{at} (genormaliseerd) convergeren naar de tweede orde karakteristieken van een tweede orde schaalinvariant proces als $a \rightarrow \infty$.

Schaalinvariantie (ook wel *self-similarity* genoemd) is een fenomeen dat voorkomt in verschillende fysische en mathematische systemen. In de theorie van de *fractals* wordt getracht het voorkomen van schaalinvariante vormen in de natuur (zoals bijvoorbeeld de vorm van een slakkenhuis, de vorm van een blad een boom, de vorm van een bloemkool, of een kustlijn) te verklaren als gevolg van een basaal onderliggende mechanisme dat iteratief herhaald wordt. Een stuk prachtige wiskundige theorie over schaalinvariantie wordt beschreven in Peitgen et al. [10].

De parameter H heet de *Hurst*-parameter. Het is eenvoudig na te gaan dat voor een niet-triviaal tweede orde schaalinvariant proces met stationaire aangroeiingen geldt dat $H \in (0, 1)$. Strict schaalinvariante processen met stationaire aangroeiingen en $H \geq 1$ bestaan weliswaar, maar zijn praktisch niet relevant in de context van telecommunicatienetwerken. De Hurst-parameter H is een indicatie voor de graad van schaalinvariantie. $H \approx 1$ betekent dat bursts meer prominent aanwezig zijn op langere tijdschalen. Het is eenvoudig in te zien dat $H \in (1/2, 1)$ impliceert dat het proces X LRD is, terwijl X SRD is als $H \in (0, 1/2]$.

LRD en schaalinvariantie zijn nauw aan elkaar gerelateerd. Aangevoerd kan worden dat de autocovariantie-functie van een (tweede-orde) schaalinvariant proces X met Hurst parameter $H \in (1/2, 1)$ aan eigenschap (2) voldoet, met $\alpha = 2 - 2H \in (0, 1)$, en dus LRD is.

6 Verkeersmodellen

In de literatuur zijn allerlei verkeersmodellen voorgesteld en bestudeerd om het gedrag van stochastische systemen te beschrijven. Voorbeelden van traditionele verkeersmodellen zijn Poisson-processen, vernieuwingsprocessen, verschillende varianten van Markov-modellen, fluid-flow modellen, Autogressive Integrated Moving Average modellen en Transform-Expand-Sample (TES)-modellen. Het is interessant op te merken dat deze modellen niet direct kunnen worden gebruikt om schaalinvariante processen te modelleren. Bovendien zijn de traditionele verkeersmodellen vrijwel in alle gevallen SRD, en dus niet geschikt zijn voor het modelleren van LRD processen zoals die zijn geobserveerd in B-ISDN netwerken.

Om de verkeerskarakteristieken van schaalinvariante processen te beschrijven zijn andere modellen, of aanpassingen van bestaande modellen, nodig. Een belangrijke randvoorwaarde voor de toepasbaarheid van dit soort modellen is dat het aantal modelparameters beperkt is — in de literatuur wordt dit ook wel *parsimonious modeling* genoemd. Hieronder worden twee veelgebruikte verkeersmodellen kort besproken: de Fractional Brownian Motion en Fractional ARIMA.

6.1 Fractional Brownian Motion

Een voorbeeld van een (exact) self-similar verkeersmodel is het Fractional Brownian Motion (FBM) model, een generalisatie van de standaard Brownse beweging die vaak in heavy-traffic analyses van stochastische systemen voorkomt. De FBM wordt gedefinieerd door het cumulatieve aankomstproces

$$A(t) = mt + \sqrt{am}Z(t),$$

waarbij de genormaliseerde Fractional Brownian motion $Z(t)$ wordt gekarakteriseerd door de volgende eigenschappen:

- $Z(t)$ heeft stationaire aangroeiingen,
- $Z(0) = 0$, en $E[Z(t)] = 0$ voor alle t ,
- $\text{var}[Z(t + \tau) - Z(t)] = \tau^{2H}$ voor alle t en τ ,
- $Z(t)$ heeft continue paden,
- de eindig-dimensionale verdelingen van $Z(t)$ zijn multivariaat normale verdelingen,
- $Z(at) =_d a^H Z(t)$ voor $a > 0$.

Merk op dat in het speciale geval $H = 1/2$, $Z(t)$ overeenkomt met de standaard Brownse beweging. Het FBM model komt veel voor in de modellering van hydrologische systemen.

6.2 Fractional ARIMA model

Een verkeersmodel met (tweede orde) schaalinvariantie is het Fractional Auto Regressive Integrated Moving Average (F-ARIMA) proces. Een F-ARIMA model met parameters p, q (integers) en d (reëel) is een stochastisch proces $X = \{X_k, k = 1, 2, \dots\}$ van de vorm

$$\Phi[B] \Delta^d(X_k) = \Theta[B] \epsilon_k,$$

waarbij $\Phi[B] = 1 - \sum_{i=1}^p \varphi_i B^i$ en $\Theta[B] = 1 - \sum_{i=1}^q \theta_i B^i$ polynomen zijn in de shift-operator $B(X_k) = X_{k-1}$, $\Delta = 1 - B$, en waarbij $\Delta^d = (1 - B)^d \sum_{k=0}^{\infty} f(d, k) (-B)^k$, met $f(d, k) (-1)^k = \Gamma(-d + k) / \Gamma(-d) \Gamma(k + 1)$, en waarbij $\{\epsilon_k, k = 0, 1, \dots\}$ witte ruis is. Dan is vrij eenvoudig in te zien dat voor $0 < d < 1/2$ het proces X asymptotisch (tweede orde) schaalinvariant is.

7 Hoe ontstaat schaalinvariantie?

De ontdekking van het fenomeen schaalinvariantie heeft geleid tot een aantal intrigerende vragen zoals ‘Hoe ontstaat schaalinvariantie?’, en ‘Welk effect heeft schaalinvariantie op netwerk-engineering?’ Het antwoord op deze vragen is niet eenduidig en al jaren onderwerp van discussie tussen experts.

7.1 Het ontstaan van schaalinvariantie

Een interessante verklaring voor het ontstaan van het fenomeen schaalinvariantie is gelegen in het feit dat individuele verkeersstromen feitelijk bestaan uit bursts van verkeer, waarbij de lengte van een burst een verdeling met een dikke staart heeft. Om precies te zijn, neem de superpositie van een N onderling onafhankelijke ON/OFF-bronnen, met waarden 1 (voor ON-perioden) en 0 (voor OFF-perioden), waarbij de lengte L_{ON} regulier variërend is met index α (voor de definitie, zie sectie 3.2). Willinger et al. [13] bewijzen dat deze superpositie voor $N \rightarrow \infty$ convergeert (in de zin van de eindig-dimensionale verdelingen) naar een Fractional Brownian Motion (FBM), een strict schaalinvariant proces, met Hurst parameter $H = (3 - \alpha)/2$.

Met andere woorden, de totale hoeveelheid verkeer gegenereerd door een groot aantal ON-OFF bronnen met dikstaartige ON-perioden (die het burst-level verkeer gegenereerd door applicaties zoals WWW, FTP en E-mail modelleren, zie sectie 4.2) is bij benadering schaalinvariant.

Een andere verklaring voor het voorkomen van schaalinvariantie is gebaseerd op de hiërarchische structuur van het verkeer in B-ISDN netwerken. Een pakket maakt vaak deel uit van een burst, die op zijn beurt deel uitmaakt van een sessie (of call). Op een grotere tijdschaal maakt een call weer deel uit van een serie van calls door dezelfde gebruikers, enzovoorts. Op deze manier maken acties steeds weer deel uit van een uitgebreidere verzameling acties. Deze redenering verklaart (tot op zekere hoogte) het ontstaan van het fenomeen schaalinvariantie.

7.2 Effect van schaalinvariantie op netwerk engineering

Ook de vraag welke implicaties schaalinvariantie (en LRD) hebben op netwerk engineering is al jaren onderwerp van discussie. Een belangrijk discussiepunt is netwerk-dimensionering: hoeveel capaciteit moet aan de netwerkverbindingen worden toegekend. Als een verkeersstroom niet direct op een gegeven link kan worden toegelaten treedt buffering op. De vraag is: gegeven de karakteristieken van de verkeersstroom, hoeveel capaciteit moet de uitgaande link hebben, zodanig dat

$$\text{Prob}\{Q > x\} < \alpha,$$

voor een of andere α , en waarbij Q de buffervulling is. Vanuit wachtrijtheoretisch oogpunt is dit een interessante vraag. Zo is gebleken dat voor de traditionele SRD verkeersmodellen de buffervulling, Q , geldt (zie [1]):

$$\text{Prob}\{Q > x\} \approx e^{-\eta x} \quad (x \rightarrow \infty), \quad (3)$$

waarbij $\eta > 0$, en waarbij $f(x) \approx g(x)$ als en alleen als $\lim_{x \rightarrow \infty} f(x)/g(x) = c$ voor een of andere constante c ($0 < c < \infty$). Voor LRD modellen (in het bijzonder voor het FBM-model met parameter H) geldt onder bepaalde aannamen dat (zie [7])

$$\text{Prob}\{Q > x\} \approx e^{-\gamma x^\beta} \quad (x \rightarrow \infty), \quad (4)$$

waarbij $\gamma > 0$ en $\beta = 2 - 2H \in (0, 1)$. Relaties (3) en (4) impliceren dat de staartkansen van de buffervulling voor LRD verkeersstromen significant langzamer naar 0 gaan als $x \rightarrow \infty$ dan in SRD verkeersstromen. Deze resultaten impliceren dat het verwaarlozen van LRD kan resulteren in een significante onderdimensionering van netwerk-buffers.

Arvidsson en Karlsson [3], daarentegen, laten op grond van experimentele resultaten zien dat de theoretische resultaten in (4) helemaal niet verschijnen op gebruikersniveau.

8 Verder onderzoek

Vandaag de dag is het belang en de implicaties van LRD en schaalinvariantie in B-ISDN netwerken nog steeds een punt van discussie. Een overzicht van de state-of-the-art is op dit gebied wordt gegeven door een recent verschenen boek van Park en Willinger [8]. Ook in Nederland wordt onderzoek gedaan naar de implicaties van eigenschappen zoals LRD en schaalinvariantie op de performance in telecommunicatie. Zo houden de Stochastic Operations Research-groep van de Technische Universiteit Eindhoven (onder leiding van Onno Boxma), de Communication and Computer Networks-groep op de Centrum voor Wiskunde en Informatica (onder leiding van Sem Borst) en de Stochastic Operations Research-groep van de Universiteit Twente (onder leiding van Henk Zijm) zich onder andere bezig met het bestuderen van wachtrijmodellen met dikstaartige kansverdelingen. ◀

Dankwoord

De auteur wil graag Ger Koole en André Ran van de Vrije Universiteit Amsterdam bedanken voor hun suggesties die hebben geleid tot een verbetering van dit artikel en Sem Borst van het CWI voor het leveren van de input voor de plaatjes in figuur 5. De historische gegevens in de introductie zijn deels gebaseerd op [6].

Referenties

- J. Abate, G.L. Choudhury and W. Whitt (1994). Asymptotics for steady-state tail probabilities in structures Markov chains. *Stochastic Models* **10**, 99–143.
- M.F. Arlitt en C.L. Williamson (1997). Internet Web servers: workload characterization and performance implications. *IEEE/ACM Transactions on Networking* **5**, 631–645.
- A. Arvidsson and P. Karlsson (1999). On traffic models for TCP/IP. In: *Teletraffic Engineering in a Competitive World*, (eds. P. Key and D. Smith), Elsevier Science, 457–466.
- V.A. Bolotin (1994). Telephone circuit holding time distributions. In: *The Fundamental Role of Teletraffic in the Evolution of Telecommunication Networks*, (eds. J. Labetoulle and J.W. Roberts), Elsevier Science, 125–134.
- V.A. Bolotin, Y. Levy en D. Liu (1999). Characterizing data connection and messages by mixtures of distributions on logarithmic scale. In: *Teletraffic Engineering in a Competitive World*, (eds. P. Key and D. Smith), Elsevier Science, 887–894.
- L.A.A.M. Coolen (1998). *Er was telefoon*. Inaugurale rede Universiteit Maastricht.
- I. Norros (1994). A storage model with self-similar input. *Queueing Systems* **16**, 387–396.
- K. Park en W. Willinger (1998). *Self-Similar Network Traffic and Performance Evaluation*, eds. K. Park en W. Willinger (Wiley). ISBN 0-471-31974-0.
- V.A. Paxson en S. Floyd (1995). Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking* **3**, 226–244.
- H.O. Peitgen, H. Jurgens en D. Saupe (1992). *Chaos and Fractals - New Frontiers of Science* (Springer-Verlag, New York).
- W.R. Stevens (1996). *TCP/IP Illustrated*, Volume 1 (Addison-Wesley).
- G. Fechner (1996). *Elements of Psychophysics*, Volume 1.
- W. Willinger, M.S. Taqqu, R. Sherman en D.V. Wilson (1995). Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. In: *Proc. ACM Sigcomm '95* (Cambridge, MA), 100–113.