

Hoofdstuk 1

Inleiding

Ondanks de recente economische terugval na het uiteenspatten van de technologie-zeepbel, stelt men ook vandaag nog steeds een aanhoudende groei vast van het telecommunicatieverkeer [40]. Het is al lang geen geheim meer dat het oorspronkelijk dominerende telefoonverkeer in volume overtroffen wordt door Internet-verkeer [41]. Geïnspireerd op de technologie van dat Internet, wat gekarakteriseerd wordt door een pakketgebaseerde overdracht van heterogene informatie, is het idee gegroeid om ook in de ondersteunende transportnetwerken pakketschakelen toe te passen, en dit om de aanwezige bandbreedte zo efficiënt mogelijk te benutten.

1.1 Optisch PakketSchakelen

De media die benut worden voor de transmissie van informatie zijn erg gediversifiëerd: men kan koper gebruiken (zoals in de klassieke telefoonlijnen, nu ook uitgebuit door xDSL technologieën, of coaxiale kabel) of zelfs lucht voor draadloze verbindingen, maar in de kern van het netwerk is optische vezel ongetwijfeld het sterkst vertegenwoordigd. Transmissie van licht in de vrije ruimte was al in 1880 door Bell gebruikt voor z'n zogenaamde fotofoon, maar het gebruik van tastbare optische dragers werd pas in de jaren vijftig van de volgende eeuw gepostuleerd. Praktisch gebruik van optische vezels bleef evenwel nog geruime tijd beperkt tot het vervaardigen van gebruiksvoorwerpen en kledij. Zowat de enige communicatie-gerelateerde toepassing was de vezelscoop (Eng. fibrescope), onder meer populair in medische kringen voor bv. gastroscopie. Medio jaren zestig, na de komst van de laser, begon men het ware potentieel van optische vezel als transmissiemedium te ontsluiten, maar was geremd door de te hoge attenuatie in toenmalige vezels: Kao geloofde in optische vezelcommunicatie en schatte in 1966 dat men erin moest slagen de attenuatie onder de 20dB/km te krijgen [42]. Het was wachten tot 1970, toen onderzoekers bij Corning erin slaagden uit met titanium gedopeerd kwartsglas een vezel te fabriceren met voldoende lage attenuatie, zodat die geschikt was voor communicatie. Die vinding gaf een nieuwe impuls aan het onderzoekswerk, wat resulteerde in het ontdekken van enerzijds geschikte golflengtes waarbij de attenuatie minimaal was (eerst nog rond 800nm, maar nadien de superieure 1310nm en later 1550nm) en anderzijds vezels met verbeterde karakteristieken (in het

bijzonder de zogenaamde “single mode fibre”, SMF [43]). In de vroege jaren tachtig waren de eerste lange afstandsnetwerken gebaseerd op optische vezelcommunicatie dan ook een feit (cf. de netwerken van MCI en Sprint in de V.S.).

Een ontwikkeling die het gebruik van optische transmissietechnologie een extra ruggensteun gaf, was golflengtemultiplexering (Eng. Wavelength Division Multiplexing, WDM). Door de komst van versterkers met een relatief breed golflengtebereik (de zogenaamde EDFA's), die simultaan verschillende golflengten kunnen versterken, was het in de jaren negentig mogelijk geworden kost-efficiënt meerdere golflengten gelijktijdig over een zelfde vezel te versturen. Intussen is men erin geslaagd het aantal beschikbare golflengten nog op te drijven door nauwere kanaal-spatieëring toe te passen en zijn bv. al bronnen met een bereik van meer dan 150 golflengten gedemonstreerd [44], terwijl multiplexers gekend zijn die zelfs tot 400 golflengten kunnen onderscheiden [45].

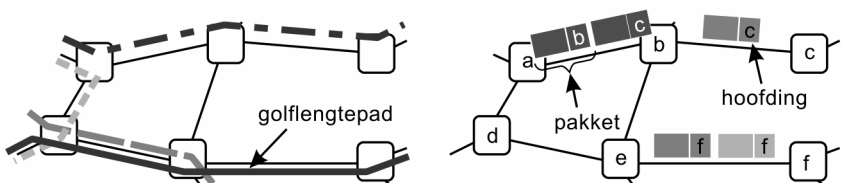
De eerste netwerken en zelfs het gros van hun hedendaagse opvolgers, gebruik(t)en optische communicatie voor punt-tot-punt verbindingen. Aan de uiteinden van de vezels wordt het licht telkens terug naar een elektrisch signaal omgezet en elektronisch verwerkt. Het correct routeren van de informatie doorheen het netwerk gebeurt dan puur elektronisch: van netwerk-functionaliteit in het optische domein is nog geen sprake. Recentere ontwikkelingen maken het evenwel mogelijk ook in de schakelpunten waar verschillende vezels samenkomen, de signalen optisch te behandelen. Een beroemd voorbeeld zijn de Micro-Elektro-Mechanische Systemen (MEMS, [46]), waarmee minuscule spiegeltjes gestuurd kunnen worden zodat ze licht van een ingangs- naar een uitgangsvazel overbrengen.

De functionaliteit van optische schakelaars die vandaag verkrijgbaar zijn laat toe om optische netwerken uit te bouwen die het concept van circuitschakelen toepassen. Dit betekent dat, om informatie uit te wisselen tussen twee eindpunten van het optische netwerk, er tussen deze punten een connectie opgezet wordt gedurende de duur van die uitwisseling, net zoals in klassieke telefonie. Er wordt dus bandbreedte gereserveerd, met name deze overeenstemmend met een golflengte, voor gegevensoverdracht tussen die eindpunten: men realiseert een zogenaamd lichtpad. Niet onbelangrijke verschilpunten met het voorbeeld van klassieke telefonie zijn evenwel dat (i) de bandbreedte veel hoger (2.5, 10 of 40 Gbit/s) en (ii) de duur van de verbinding veel langer is (uren, dagen, weken of zelfs maanden). Vermits een connectie overeenstemt met een golflengte, hanteert men ook wel de term golflengteschakelen (Eng. lambda switching). Verschillende nationale en internationale projecten hebben de werkbaarheid ervan aangetoond [47,48]. Het onderzoek met betrekking tot golflengteschakelen omvat tegenwoordig hoofdzakelijk het controleren en beheren van het resulterend optische netwerk [2,48]. Door de specifieke eigenschappen van de optische technologie dienen sommige van de bestaande werkwijzen uit het elektrische domein aangepast en/of uitgebreid te worden. In deze context wordt in onderzoeksmiddens vooral aandacht besteed aan het veralgemeend multi-protocol labelschakelen (Eng. Generalized Multiprotocol Label Switching, GMPLS [49–51]).

In de circuitgeschakelde aanpak van golflengteschakelen is het een heikel punt de aanwezige bandbreedte efficiënt te benutten, gelet op de hoge bandbreedte geassocieerd met een golflengte (en dus een circuit). Om aan de bandbreedte-

behoeften van de gebruikers van het optische netwerk zo kost-efficiënt mogelijk te voldoen, zal men aan de schakelaars in het optische netwerk ook elektrische schakelaars toevoegen die met een fijnere bandbreedte-granulariteit werken. Zo ontstaat het probleem van “grooming” (zie verder in Hoofdstuk 4): hoe en waar worden de (elektrische) verkeersstromen van fijnere granulariteit samengenomen in optische lichtpaden?

Om de bandbreedte aanwezig in het optische domein doeltreffend te benutten, is optisch pakketschakelen de aangewezen uitweg die het “grooming”-probleem omzeilt. In plaats van lichtpaden op te zetten, waarbij tussen de eindpunten een continue lichtstroom voor de gegevensoverdracht zorgt, zal men de gegevens over optische pakketten verdelen. Deze worden dan elk afzonderlijk tussen de eindpunten verstuurd. Zo kan één enkele golflengte gebruikt worden voor pakketten die behoren tot verschillende connecties, zolang die samen niet méér bandbreedte vereisen dan de golflengte voorziet. Om de verschillende verkeersstromen te onderscheiden, maakt men dan gebruik van de zogenaamde hoofding (Eng. header): met elke hoeveelheid gegevens die in een pakket vervat zit, correspondeert een hoofding die informatie bevat over die over te dragen gegevens, zoals begin- en eindpunt waartussen ze in het optische netwerk worden uitgewisseld. Het onderscheid met elektrisch pakketschakelen ligt in het feit dat de eigenlijke pakket-inhoud (Eng. payload) optisch behandeld wordt en dus in de schakelknopen geen conversie van en naar het elektrische domein ondergaat. Men spreekt in dit verband ook van transparante optische netwerken wanneer de eigenlijke bitrate van die inhoud willekeurig kan zijn. Merk op dat de hoofding wel elektro-optische conversie kan ondergaan, vermits de optische signaalverwerking, nodig om de hoofding te interpreteren, nog in haar kinderschoenen staat (zie verder).



Figuur 1.1: Circuit- versus pakketschakelen.

Wanneer men de term Optisch PakketSchakelen (OPS, Eng. Optical Packet Switching) hanteert [52–53], doelt men doorgaans op het geval waarbij de lengte van de pakketten constant is en aangeduid wordt met de term “slot”. De schakelaars die deze pakketten behandelen zijn in dit geval voorzien van synchronisatie-eenheden aan hun ingangen zodat pakketten, afkomstig van verschillende vezels, en/of verschillende golflengtes op een zelfde vezel, gealigneerd worden. De schakelaar kan dan repetitief werken: slot per slot de hoofdingen van de nieuw aankomende pakketten inspecteren en al deze pakketten gelijktijdig naar de correcte uitgangen doorschakelen. Omdat dergelijke synchronisatie niet voor de hand ligt en om de vereiste schakelsnelheden te beperken, heeft men het Optisch Burst Schakelen (Eng. Optical Burst Switching, OBS) bedacht [54]. Hierbij is de pakketlengte variabel en typisch langer dan de in OPS gehanteerde slotlengte. De

schakelaar heeft nu geen synchronisatie-eenheden die pakketten over verschillende ingangen heen aligneert. De voornaamste motivatie voor OBS is het afzwakken van de technologische vereisten die gesteld worden door OPS. Daarbij is men bereid sommige gekende voordelen van slot-gebaseerde aanpakken (efficiëntere benutting van bandbreedte, eenvoudiger realiseren van gedifferentieerde dienstkwaliteit, etc.) in te leveren.

In het kader van OPS en OBS is het niet oninteressant hier ook melding te maken van het zogenaamde ORION concept, wat staat voor “Overspill Routing In Optical Networks” en aan deze vakgroep bedacht werd (o.a. [55]). Het achterliggende idee is het gerbuik van een hybride aanpak die het beste van de pakket- en circuitgeschakelde concepten tracht te verenigen. Het concept stelt voor een golflengteschakelende kruisschakelaar (optisch) te combineren met een pakket-schakelaar (bv. een elektronische IP-router). De binnenkomende golflengten worden via 1×2 schakelaars verbonden met zowel de pakketschakelaar als de golflengteschakelaar. De bedoeling is het gros van de gegevens volgens het principe van golflengteschakelen te versturen en dus operaties per pakket te vermijden. Om evenwel de aanwezige bandbreedte zo volledig mogelijk te benutten, kan door het meesturen van een aangepast controlesignaal de binnenkomende golflengte omgeleid worden naar de pakketschakelaar. Het gevolg is dat de pakketschakelaar een veel lagere belasting moet verwerken dan in een puur pakketgeschakeld concept, terwijl de bandbreedte nagenoeg even efficiënt benut wordt: het aantal golflengten die nodig zijn om een gegeven hoeveelheid verkeer te dragen is beduidend lager dan in een puur golflengteschakelend concept.

De commerciële realisatie van OPS wordt nog in de weg gestaan door technologische moeilijkheden, die voornamelijk bestaan uit het ontbreken of slechts in een beperkte vorm aanwezig zijn van (i) optisch geheugen, (ii) geavanceerde, volledig optische signaalverwerking voor de hoofding van de pakketten, (iii) optische synchronisatie en (iv) optische signaalregeneratie. Momenteel wijden verschillende groepen en projecten onderzoek aan het oplossen of verhelpen van deze tekortkomingen. Volledig optische 3R-regeneratie—herstellen van vorm, versterking en timing van het signaal—is intussen meermaals gedemonstreerd [56–58]. Ook synchronisatie via optische technologie werd onder meer in [59–61] gerealiseerd, maar blijft door het gebruik van vezelvertragslijnen (Eng. Fibre Delay Lines, FDL) eerder omslachtig. Verschillende volledig optische implementaties van het verwerken van de hoofdingen in optische pakketschakelaars werden verwezenlijkt [62–65], ook al blijft de functionaliteit eerder beperkt. Echt optisch geheugen, in de zin van een tegenhanger van het elektrisch geheugen met in de tijd willekeurige toegang (Eng. Random Access Memory, RAM), is vandaag nog niet beschikbaar: het gebruik van optische flip-flops [66] om volledige pakketten te stockeren lijkt momenteel nog geen werkbare oplossing. Hoewel het stoppen van licht—of het tot zeer lage snelheden (tientallen m/s) vertragen ervan—geen science fiction meer is [67,69], lijkt het in praktijk nog niet bruikbaar, ondanks het feit dat ook bij kamertemperatuur al vertraagd licht waargenomen werd [70]. Het vooralsnog ontbreken van een optische tegenhanger van RAM kan gelukkig wel verholpen worden door het gebruik van FDL's, zoals onder meer verder in dit werk aangetoond zal worden. Een alternatief is het gebruik van elektrisch geheugen aan de rand van

het optische netwerk, waarbij de toegang tot het optisch gedeelte geregeld wordt via een zorgvuldig ontworpen protocol voor toegang tot het medium (Eng. Medium Access Control, MAC) zoals bv. uiteengezet in [71]. Zo kan het optisch netwerk zelf geheugenloos blijven.

Het mag duidelijk zijn dat al veelbelovende stappen genomen zijn in het overkomen van de huidige technologische beperkingen en we geloven dan ook dat OPS eens het onderzoeksstadium zal ontgroeien. Het resterende deel van dit hoofdstuk zullen we besteden aan een beknopte schets van OPS-architecturen in Sectie 1.2, gevolgd door de belangrijkste ontwerpskeuzes vanuit het standpunt van logische performantie (i.e. efficiëntie van bandbreedtegebruik, verlies van pakketten, kwaliteitsdifferentiatie van diensten) in Sectie 1.3. We ronden het hoofdstuk af in Sectie 1.4 met een overzicht van de in dit doctoraat gerealiseerde bijdragen, gevolgd door de obligatoire lijst van publicaties in Sectie 1.5.

1.2 Architecturen

De organisatie van een communicatienetwerk vertoont een zekere hiërarchie: men onderscheidt lokale netwerken (Eng. Local Area Networks, LAN) zoals binnen een bedrijfssite of campus, toegangsnetwerken (Eng. access network), netwerken op metropole schaal (Eng. Metropolitan Area Network, MAN) en ruggengraatnetwerken (Eng. Wide Area Network, WAN). Hoewel optische technologieën op elk van deze niveaus doordringen, zal de toepassing van OPS wellicht het snelst ingang vinden op MAN en WAN schaal: concurrentie met huidige erg kost-efficiënte LAN technologieën is moeilijk en toegangsnetwerken zijn vandaag nog zeer sterk gebaseerd op koper veeleer dan optische vezel.

De architecturen die worden voorgesteld op MAN-niveau onderstellen doorgaans eenvoudige topologieën: ze gebruiken een ster- of ring-configuratie. De motivatie hiervoor is dat de architectuur van een knoop in zo'n netwerk veel eenvoudiger—en daarom ook goedkoper—kan zijn dan wanneer men een willekeurig vermaasd netwerk zou willen construeren. De ster-gebaseerde netwerken vindt men terug in de aanpak van de zogenaamde Passieve Optische Netwerken (PON's). Op optisch niveau bevat het gros van de knopen geen enkele schakelfunctionaliteit: licht van verschillende golflengtes wordt samengenomen waar takken samenkomen, of gesplitst over verschillende uitgaande takken. Voorbeelden van PON-gebaseerde aanpakken zijn [72,73] en de eerder op toegangsnetwerken gerichte werken [74,75]. De voornaamse elementen uit een PON-architectuur zijn de optische lijnterminals (Eng. Optical Line Terminal, OLT) en optische netwerkknopen (Eng. Optical Network Unit, ONU). Deze laatste vormen de verbinding met de gebruikers van het netwerk via klassieke elektrische interfaces, terwijl de OLT-eenheden interconnetie met metro- of ruggengraatnetwerken zullen realiseren. Meerdere ONU's zullen via passieve splitters/combiners verbonden zijn met één enkele OLT. Merk op dat, hoewel zo'n PON pakketgebaseerd kan zijn, er niet echt sprake is van een OPS-netwerk, aangezien er geen schakel-elementen voorkomen en pakketten niet worden gerouteerd op basis van hun hoofding. De enige plaats waar van OPS sprake kan zijn is in een OLT.

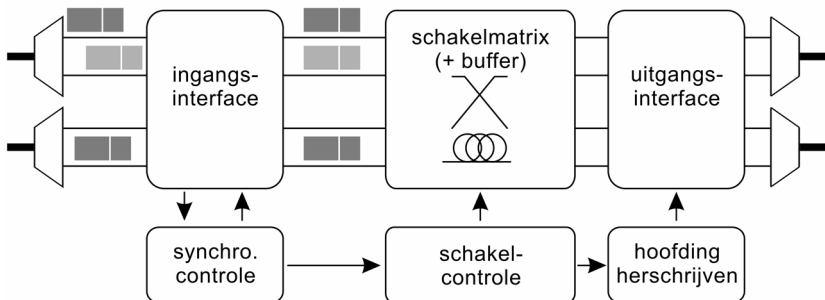
Ring-topologieën bieden een aantal inherente voordelen ten opzichte van stergebaseerde, waaronder een gemiddeld kortere vezellengte om een gegeven set knopen met elkaar te verbinden en een relatief eenvoudig te realiseren robuustheid tegen falingen zoals vezelbreuken [76]. Ring-gebaseerde OPS-architecturen werden onder meer voorgesteld in [6,77,78,79]. Sommige van de voorgestelde architecturen zijn enigszins verwant aan de PON-structuren met een centrale schakelknoop: één van de knopen op de ring heeft een speciale functie, wat zich onder meer uit in een complexere structuur. Zo'n centrale knoop zal dan bijvoorbeeld ook interconnectie van verschillende ringen voorzien [6]. In andere voorstellen zijn dan weer alle ringknopen equivalent [77–79] (maar ontbreekt bv. de mogelijkheid om meerdere optische ringen in eenzelfde knoop met elkaar te verbinden). Een studie die de complexiteit van de optische ringknopen afweegt tegenover de efficiëntie waarmee de aanwezige bandbreedte benut wordt, komt verder in dit proefschrift aan bod (zie Hoofdstuk 2, Sectie 2.2).

Het is duidelijk dat de beperkte optische functionaliteit van de knopen in dergelijke metro-netwerken zich zal vertalen in complexiteit op het logische (controle-)niveau. Het versturen van gegevens over zo'n netwerk zal worden geregeld door een aangepast MAC-protocol. Om verkeer tussen knopen op eenzelfde MAN uit te wisselen, zal aandacht besteed moeten worden aan het probleem van eerlijkheid [80]: er moet voor gezorgd worden dat één enkele knoop niet alle aanwezige bandbreedte inpalmt en daardoor andere knopen het verzenden van gegevens belet. Een typische oplossing hiervoor is het gebruiken van token-gebaseerde toegang [81,82]: een knoop kan dan slechts gegevens op de ring plaatsen wanneer hij een zogenaamd token in zijn bezit heeft, dat hij nadien zal doorgeven aan een volgende gegadigde. Wanneer het MAN opgesplitst is in verschillende deelnetwerken (bv. afzonderlijke ringen), zal ook de uitwisseling van gegevens tussen deze deelnetwerken efficiënt en eerlijk moeten verlopen. Een voorbeeld van MAC ontwerp voor geïnterconnecteerde ringen is te vinden in [71].

In een ruggengraatnetwerk zal men de eenvoudige topologieën niet langer handhaven. Men realiseert er doorgaans willekeurig vermaasde netwerken omdat die toelaten de propagatie-afstand tussen bron en bestemming te minimaliseren, eventueel de grootte van de schakelmatrices te optimaliseren, alsook de mogelijkheid bieden de belasting intelligent te verdelen (Eng. load balancing) over het netwerk. Als gevolg hiervan zal ook de functionaliteit van de knopen in dergelijke WAN's complexer zijn. Typisch vervult zo'n knoop de functie van kruisschakelaar: hij kan een willekeurige ingangspoort met een willekeurige uitgangspoort verbinden.

Een generieke structuur van een OPS-knoop is geschetst in Figuur 1.2. Die beslaat drie stadia [53]: (i) een ingangsinterface, (ii) een schakelmatrix en (iii) een uitgangsinterface. In het geval van slot-gebaseerd OPS zal de ingangsinterface synchronisatie-eenheden bevatten die inkomende pakketten aligneert. De belangrijkste functie van het eerste stadium van de OPS-schakelaar is het ontcijferen van de hoofding geassocieerd met elk pakket—een functie die ook in asynchrone concepten zoals OBS vervuld moet worden. Zoals eerder vermeld zal de hoofding informatie bevatten omtrent de bestemming van het pakket. Op die informatie zal de schakelknoop zich baseren om te beslissen naar welke uitgang het pakket zal worden doorgestuurd. De meerderheid van de OPS-concepten die tot op vandaag zijn

gedemonstreerd zet hiervoor de hoofding om van een optisch naar een elektronisch formaat [6,77,79,83,84] om de schakelmatrix te sturen. Zodoende tracht men met een pragmatische aanpak het beste van beide werelden te combineren: optische technologie omwille van bandbreedte- en schakelcapaciteiten, elektronische voor de controle- en beslissingsoperaties. Pas wanneer de functionaliteit van optische verwerking van de pakkethoofdingen rijper is, zal men de pakketten puur optisch kunnen schakelen.

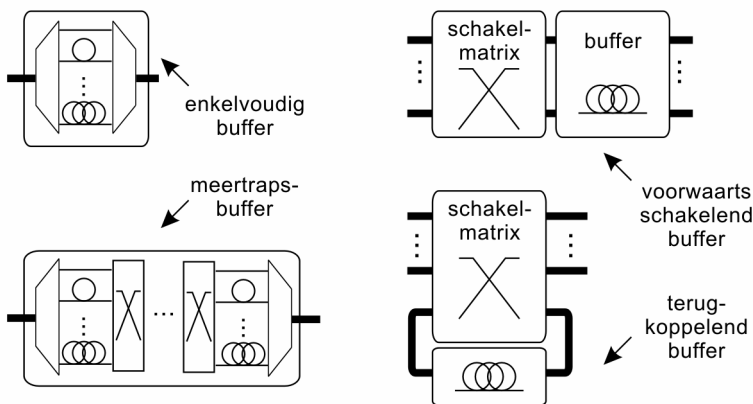


Figuur 1.2: Generieke structuur van een OPS knoop met illustratieve functies van de verschillende trappen.

De schakelmatrix in het centrale stadium realiseert de eigenlijke interconnectie van in- en uitgangen. Verschillende optische schakeltechnologieën zijn bekend. De meest tot de verbeelding sprekende zijn wellicht MEMS met microscopisch kleine twee-dimensionale opklappende [85] of zelfs drie-dimensionale kantelende [86] spiegeltjes. Voor optisch pakketschakelen zijn hun schakeltijden (in de orde van minstens enkele honderden microseconden of zelfs enkele milliseconden) evenwel te traag. Naast meer exotische voorstellen zoals bubblejet-technologie [87] en andere thermo-, electro-, of acousto-optische schakelaars [88–90], zijn de meer praktische en meest voorgestelde architecturen gebaseerd op zogenaamde “Arrayed Waveguide Gratings” (AWG) enerzijds [83,84,91] of optische versterkers in halfgeleider-technologie (Eng. Semiconductor Optical Amplifier, SOA) [92,93] anderzijds. Het passieve AWG-element heeft als kenmerkende eigenschap dat de uitgangspoort waarlangs licht het element verlaat, afhankelijk is van de frequentie, i.e. de golflengte, van dit licht. Door aan de ingangen van de AWG regelbare golflengte-convertoren te plaatsen, kan men bepalen aan welke uitgang inkomend licht wordt doorgeschakeld en op die manier een schakelaar realiseren. Het alternatief dat gebruik maakt van SOA’s is een zogenaamde verdeel-en-selecteer-aanpak (Eng. broadcast-and-select): het optische signaal wordt door splitters over de verschillende uitgangen verdeeld en de SOA wordt als schakelaar ingezet om per uitgang slechts het signaal van één ingangspoort door te laten. Een inherent voordeel van deze SOA-aanpak is dat multicast mogelijk is: hetzelfde ingangssignaal kan tegelijk naar verschillende uitgangen gezonden worden. In het voordeel van de AWG pleit dan weer dat de signaalkwaliteit aan de uitgang typisch beter is, door het ontbreken van splitters die onontkoombaar zijn in de verdeel-en-selecteer-structuur.

Het laatste stadium van de schakelaar omvat de uitgangsinterface. Die zal moeten instaan voor regeneratie van het optische signaal, waar dit door de schakeloperatie wordt aangetast, en/of het (her)schrijven van de hoofding. Ook kan in dit stadium golflengte-conversie aanwezig zijn, vermits het een belangrijk hulpmiddel is om zogenaamde contentie op te lossen [94,95]. Die contentie treedt op wanneer op verschillende ingangsvezels gelijktijdig een pakket aanwezig is dat naar een zelfde uitgang verstuurd moet worden (waarbij de inkomende golflengte bovendien dezelfde kan zijn). Door één of meerdere pakketten naar een andere golflengte te converteren, wordt dit gelijktijdig verzenden over dezelfde uitgaande vezel mogelijk gemaakt.

De klassieke oplossing voor contentie is het voorzien van bufferruimte. Optisch kan dit, zoals eerder vermeld, gerealiseerd worden met behulp van vezelvertragslijnen (FDL's). Net zoals bij elektrische schakelaars zijn er verschillende opties om buffering te voorzien [96]. Men onderscheidt voorwaarts schakelende (Eng. feed-forward) en terugkoppelende (Eng. feed-back) structuren, zoals geschetst in Figuur 1.3: in het eerste geval bevindt de buffer zich in lijn met de schakelmatrix (of is er eventueel mee verweven [92,97]), in het laatste is een deel van de uitgangspoorten van de schakelmatrix via FDL's verbonden met ingangspoorten zodat pakketten eventueel meermaals doorheen de buffer kunnen circuleren (bv. [6,84,98]). Verder kunnen de buffers gecatalogeerd worden in enkelvoudige en meervoudige stadia. Meervoudige bufferstadien worden doorgaans enkel in een voorwaarts schakelende configuratie gebruikt. In zo'n meervoudig buffer kan de vertragingstijd (i.e. de tijd die een pakket in de buffer doorbrengt) variëren: in elk stadium heeft men de keuze tussen verschillende vertragingen, zodat men met een beperkt aantal stadia toch een relatief wijd bereik aan vertragingen kan aanbieden en een FIFO buffer van bepaalde diepte implementeren [99]. In dit verband kunnen we ook opmerken dat sommigen suggereren om elektrische buffers te voorzien [100], mogelijks in combinatie met optische buffers [101]. Het mag evenwel duidelijk zijn dat dit, gezien de extra elektro-optische conversies en hoge snelheidvereisten voor het geheugen, de complexiteit van de knooparchitectuur niet ten goede komt.



Figuur 1.3: Illustratie van bufferstructuren.

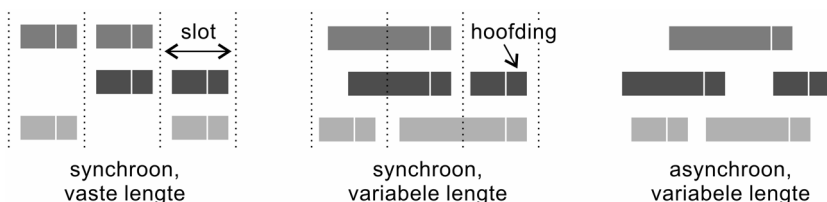
In een WAN zullen nagenoeg alle functies aangegeven in Figuur 1.2 aanwezig zijn, inclusief buffers, maar in een MAN zullen bepaalde onderdelen niet, of sterk vereenvoudigd, vertegenwoordigd zijn. In een PON-aanpak kan enkel de OLT-knoop enige gelijkenis vertonen—want randknoten zullen niet volledig optisch zijn—en dan bovendien zonder buffer. In ring-architecturen zal men in de ringknoten evenmin optische buffers voorzien en kan de functionaliteit van de schakelmatrix beperkt worden tot zogenaamde add/drop-functies. Zoals eerder aangegeven zal in zulke bufferloze architecturen het gebruik van een aangepast MAC-protocol onontbeerlijk zijn.

1.3 Ontwerpskeuzes

De ontwerpskeuzes in OPS omvatten naast de technologische uitdagingen ook eerder conceptuele vragen. De eerste cruciale vraag is uiteraard hoe het optische pakket en bijhorende hoofding er precies uitzien. Daarnaast moet worden vastgelegd hoe het algoritme geconcipteerd wordt om deze pakketten te schakelen.

1.3.1 Pakketformaat

Het idee dat aan de basis ligt van pakketschakelen is het opsplitsen van de te versturen gegevens in eenheden van beperkte lengte. Een eerste belangrijke keuze betreft het bepalen van deze lengte. In de elektrische wereld gebruikt het Internet Protocol (IP) pakketten van variabele lengte [102], terwijl men in de Asynchrone Transfer Mode technologie (ATM, [103]) gebruik maakt van eenheden van vaste lengte, die men er “cel” noemt. Ook in het optisch schakelen heeft men de keuze tussen vaste of variabele lengte pakketten, met in Figuur 1.4 de meest gangbare alternatieven.

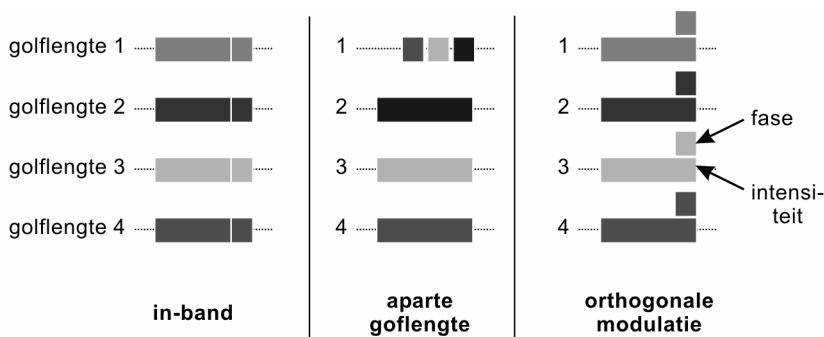


Figuur 1.4: De voornaamste alternatieven voor optisch pakketschakelen.

Het OBS concept hanteert, zoals eerder aangehaald, pakketten van variabele lengte die op asynchrone wijze worden geschakeld. Dit heeft als voordeel dat, wanneer de pakketten (of bursts) voldoende lang zijn, het absolute aantal pakketten nodig om een bepaalde hoeveelheid informatie te versturen relatief laag kan gehouden worden. Daardoor zal de belasting op de controle-eenheid in de schakelknoten beperkt gehouden worden: de schakelbeslissing wordt per pakket gemaakt en hangt niet af van de lengte ervan. Bovendien wordt dan slechts een minimale hoeveelheid bandbreedte besteed aan het versturen van de pakket-hoofding.

In OPS gebruikt men doorgaans synchrone pakketschakelaars die slot-gebaseerd schakelen. In zo'n geval ligt het voor de hand pakketten te hanteren van vaste lengte die binnen de duur van een slot vallen. Argumenten die in het voordeel pleiten van een slot-gebaseerd concept bestaan erin (i) dat het contentie-probleem efficiënter kan worden opgelost, want asynchroon schakelen verhoogt de verlieskans [104] (een fenomeen dat al eerder opdook bij vergelijken van Aloha met Slotted-Aloha in het domein van toegangsprotocollen voor gedeelde media), en (ii) dat aan de schakelmatrix minder strenge eisen worden gesteld qua blokkering, aangezien die matrix nu slechts zogenaamd herschikbaar niet-blokkerend (Eng. rearrangeable non-blocking) moet zijn [97] (zie ook verder in Hoofdstuk 2). Het voornaamste minpunt van OPS met vaste lengte pakketten is uiteraard de hoge overlast die ontstaat doordat in elk slot een aparte hoofding moet worden meegestuurd en verwerkt. Om dit te verhelpen kan men, ondanks het slot-georiënteerde concept, toch werken met pakketten van variabele lengte. In dit geval wordt het pakket opgedeeld in eenheden van één slot lang, maar wordt enkel met het eerste slot een hoofding meegestuurd die de schakelbeslissing zal bepalen. Dat de volgende slots deel uitmaken van hetzelfde pakket kan bv. worden aangegeven met een zogenaamde “continuation bit” [105]. Die beslissing wordt steeds tegelijk genomen voor alle slots die samen het pakket vormen: men behandelt als het ware “treinen van slots” [106]. Een meer uitgebreide bespreking van de performantie van de twee OPS-varianten volgt in Hoofdstuk 3.

Doorgaans wordt voor de pakket-hoofding een lagere bitrate gebruikt dan voor de eigenlijke pakket-inhoud. In principe kan de pakket-inhoud zelfs een willekeurige bitrate hebben, aangezien deze in de schakelknopen niet be- of verwerkt dient te worden. Men spreekt in dit laatste geval van transparant optisch schakelen. Hoewel dit soms als een aantrekkelijk voordeel van optisch schakelen wordt voorgesteld, vereist zo'n aanpak een groot aantal adaptieve componenten [107] en semi-transparante netwerken die ontworpen zijn voor een gegeven signaalformaat en -bitrate kunnen dan ook (kost-)efficiënter zijn.



Figuur 1.5: Alternatieven voor het transporteren van de pakket-hoofding.

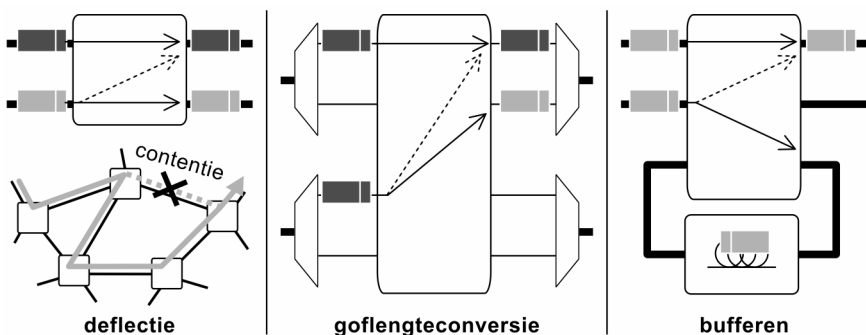
Voor het versturen van de hoofding zelf bestaan verschillende oplossingen. Een voor de hand liggende werkwijze is de hoofding vóór de pakketinhoud te zenden, op dezelfde golflengte, zoals geschetst in Figuur 1.5. Het voordeel hiervan is dat er

geen speciale maatregelen (zoals het compenseren van dispersie) getroffen moeten worden om pakket en bijhorende hoofding samen te houden. Een alternatieve aanpak is een aparte golflengte reserveren voor het doorsturen van de hoofdingen, wat typisch voorgesteld wordt in een OBS-context, waar de hoofding de naam “Burst Controle Pakket (BCP)” [108] of “Burst Hoofding Pakket (BHP)” [109] krijgt. Dit zorgt voor een erg eenvoudige afzondering van de hoofding om de schakelmatrix te sturen. Een aanpak die het voorzien van een aparte controle-golflengte vermijdt, is de hoofding in zogenaamde “sub-carrier multiplexing” (SCM) [79,110,111] of met behulp van een andere orthogonale modulatie-methode mee te sturen, zoals bv. FSK [83] of een combinatie van ASK en DPSK [112].

1.3.2 Schakel-algoritmes

Bij het eigenlijke doorschakelen van de pakketten vormt contentie het voornaamste probleem waaraan het hoofd moet worden geboden. Dit is het fenomeen waarbij op meerdere ingangen een pakket aankomt dat bestemd is voor dezelfde uitgaande poort, waarbij deze pakketten in de tijd overlappen. In het geval van optisch pakketschakelen zal de term “poort” slaan op een vezel die meerdere golflengtes in parallel draagt (cf. WDM).

Op drie manieren kan dit probleem worden opgelost [53]: (i) uitbuiten van de golflengte-dimensie door het toepassen van golflengte-conversie, (ii) buffering voorzien, en/of (iii) deflectie routing (Eng. deflection routing). In de eerste oplossing worden alle golflengtes op een vezel als equivalent beschouwd: het zijn parallele kanalen die naar dezelfde volgende schakelknoop leiden. Vermits in zo’n concept de precieze golflengte waarop een pakket wordt verstuurd geen betekenis¹ heeft, kunnen pakketten waarvoor contentie optreedt elk op een andere golflengte verstuurd worden door gepaste golflengte-conversie. (Bemerkt dat wanneer de schakelaar 3R-regeneratie voorziet aan de uitgangen van de schakelmatrix, dit regeneratie-stadium de conversie kan verzorgen.)



Figuur 1.6: Mogelijkheden tot oplossen van contentie.

¹ In de zin dat ze niet bepalend is voor de schakelbeslissing, in tegenstelling tot zogenaamde golflengte-routing die bij optisch circuitschakelen voorgesteld wordt.

Optisch bufferen betekent dat men vezelvertragslijnen zal voorzien om ervoor te zorgen dat de pakketten in de tijd gespreid, zonder overlappen, op de uitgaande vezel terecht kunnen komen. De verschillende mogelijke buffer-configuraties werden eerder besproken (zie Figuur 1.3) en vereisen extra schakel-elementen of een grotere schakelmatrix in vergelijking met bufferloze architecturen.

De zogenaamde deflectie-routing vereist geen bijkomende componenten en zal overtollige pakketten, waarvoor op de uitgaande vezel geen plaats meer is, naar een alternatieve uitgaande vezel sturen waar wel een golflengte vrij is. Het spreekt voor zich dat deze methode een bijzondere organisatie van de routing vereist, zodat een pakket ten allen tijde correct op de bestemming afgeleverd wordt en niet in een lus blijft rondcirkelen. Vergelijking van de drie methoden [98] voor het oplossen van contentie heeft trouwens uitgewezen dat deflectie enkel werkt wanneer het netwerk niet te zwaar belast wordt (cf. er moeten voldoende “alternatieve” uitgaande vezels licht belast zijn om overtollige pakketten op te vangen).

Een geschikt algoritme zal moeten beslissen welke pakketten in geval van contentie worden weerhouden om te worden doorgeschakeld en/of naar een buffer gestuurd wanneer er FDL's voorzien zijn. In het geval van buffers die een variabele vertraging kunnen aanbieden, zal het algoritme ook moeten instaan voor de beslissing over de duur van de vertraging van gebufferde pakketten. Wanneer er geen buffers zijn in de optische schakelknopen zal ofwel de belasting voldoende laag gehouden moeten worden (zodat het contentie-probleem niet optreedt), ofwel zal men contentie moeten vermijden door de toegang tot het OPS-netwerk via een gepast MAC-protocol te regelen (zie ook hoger).

Bij het ontwerp en de evaluatie van dergelijke schakel-algoritmen, zoals die in Hoofdstuk 3 in detail besproken worden, zal vanzelfsprekend ook aandacht besteed moeten worden aan differentiatie in de kwaliteit van dienstverlening [113] (Eng. Quality of Service, QoS). Het OPS-netwerk zal immers sterk heterogeen verkeer dragen, met uiteenlopende vereisten in termen van vertraging en pakketverlies.

1.4 Structuur doctoraat

Bovenstaand inleidend materiaal is gedeeltelijk gebaseerd op onze publicaties [23,31] waarvan de laatste terug te vinden is in Appendix A. In het vervolg van dit proefschrift zullen we een coherent overzicht trachten te geven van de originele bijdragen die we geleverd hebben in het domein van OPS. Hierbij zullen we vanzelfsprekend niet alle geboekte resultaten uitvoerig aan bod laten komen, maar veeleer een synthese presenteren waarbij we voor details verwijzen naar publicaties uit bijgevoegde selectie.

In het eerstvolgende hoofdstuk belichten we de onderzoeksresultaten met betrekking tot knoop-architecturen voor OPS. Een eerste deel daar behelst een evaluatie van alternatieven voor ring-gebaseerde metro-netwerken. In het bijzonder beantwoorden we de vraag of ruimtelijk hergebruik (Eng. spatial reuse) zin heeft en of het een goed idee is zogenaamde golflengtebanden te hanteren. Een tweede deel spitst zich toe op de vraag hoe schakelmatrixen zoals die typisch worden voorgesteld voor OPS-ruggengraatnetwerken geschaald kunnen worden naar grote aantallen in- en uitgangspoorten. Dit laatste deel heeft aanleiding gegeven tot de publicatie van

onze werken [9,11,26] (laatstgenoemde is gereproduceerd in Appendix C). Het eerste deel omtrent metro-netwerken wordt voorgesteld in [30] (na te lezen in Appendix B) en [33].

In Hoofdstuk 3 behandelen we de schakel-algoritmes die we ontwierpen en evalueerden voor optische schakelaars met een teruggekoppelde FDL-buffer. Daarbij bespreken we zowel het geval van pakketten van vaste lengte als variabele lengte. Het mag duidelijk zijn dat we ons hierbij richten op de logische performantie van deze algoritmes en in het bijzonder aandacht besteden aan het (beperken van) pakketverlies. Details van deze studies zijn terug te vinden in [20,24,25] voor pakketten van vaste lengte, waarvan we de eerste twee hebben opgenomen in Appendices D en E. Pakketten van variabele lengte worden behandeld in de volgende Appendices F en G (resp. [32] en [29]).

Wanneer we aldus de architectuur van afzonderlijke knopen en hun performantie besproken hebben, gaan we in Hoofdstuk 4 over tot het beschouwen van een volledig netwerk. Meer bepaald bedenken we een algoritme dat de routing van een gegeven verkeerspatroon bepaalt voor een gegeven netwerk, met het oog op het minimaliseren van het pakketverlies. Dit werk wordt uitvoerig behandeld in onze publicatie [7] die vervat is in Appendix H.

Ter afronding van de thesis klimmen we in Hoofdstuk 5 op tot net onder de applicatie-laag en bespreken de effecten van schakelacties, die in het onderliggende optische netwerk genomen worden, op het gedrag van TCP-stromen. Een deel van de resultaten van die studie werden voorgesteld in [3,16] en verder uitgewerkt in [4] die in Appendix I is opgenomen.

1.5 Lijst van publicaties

Publicaties in internationale tijdschriften:

- [1] D. Colle, P. Van Heuven, **C. Develder**, S. Van den Berghe, I. Lievens, M. Pickavet, P. Demeester, "MPLS recovery mechanisms for IP-over-WDM networks", *Photonic Network Commun. (PNET)*, Kluwer Academic Publishers, vol. 3, no. 1/2, Jan. 2001, pp. 23–40.
- [2] S. De Maesschalck, D. Colle, A. Groebbens, **C. Develder**, A. Lievens, P. Lagasse, M. Pickavet, P. Demeester, F. Saluta, M. Quagliatti, "Intelligent optical networking for multilayer survivability", *IEEE Commun. Mag.*, vol. 40, no. 1, Jan. 2002, pp. 42–49.
- [3] D. Colle, S. Demaesschalck, **C. Develder**, P. Van Heuven, A. Groebbens, J. Cheyns, I. Lievens, M. Pickavet, P. Lagasse, P. Demeester, "Data-centric optical networks and their survivability", *IEEE J. Selected Areas in Commun.*, vol. 20, no. 1, Jan. 2002, pp. 6–20.
- [4] **C. Develder**, D. Colle, S. Demaesschalck, M. Pickavet, P. Demeester, "Influence of GMPLS recovery mechanisms on TCP performance", *Photonic Network Commun. (PNET)*, vol. 4, no. 3/4, Jul.-Dec. 2002, pp. 321–345.
- [5] J. Cheyns, **C. Develder**, E. Van Breusegem, E. Baert, A. Ackaert, M. Pickavet, P. Demeester, "Routing in an AWG based optical packet switch", *Photonic Network Commun. (PNET)*, Kluwer Academic Publishers, vol. 5, no. 1, Jan. 2003, pp. 69–80.

- [6] L. Dittmann (ed.), **C. Devellder**, D. Chiaroni, F. Neri, F. Callegati, W. Koerber, A. Stavdas, M. Renaud, A. Rafel, J. Solé-Pareta, W. Cerroni, N. Leligou, L. Dembeck, B. Mortensen, M. Pickavet, N. Le Sauze, M. Mahony, B. Berde, G. Eilenberger, "The European IST project DAVID: a viable approach towards optical packet switching", *IEEE J. Selected Areas in Commun.*, vol. 21, no. 7, Sep. 2003, pp. 1026–1040.
- [7] E. Baert, **C. Devellder**, M. Pickavet, P. Demeester, "Routing strategies to minimize packet loss in an optical packet switched network with recirculating FDL buffers", zal verschijnen in *Phot. Netw. Commun. (PNET)*, vol 7, no. 2, Mar. 2004.
- [8] **C. Devellder**, B. Van Houdt, C. Blondia, M. Pickavet, P. Demeester, "Analytical MMAP-based bounds for packet loss in optical packet switching with recirculating FDL buffers", submitted to *PNET*.
- [9] J. Cheyns, **C. Devellder**, E. Van Breusegem, E. Baert, D. Colle, M. Pickavet, P. Demeester, "Clos lives on in Optical Packet Switching", aanvaard voor publicatie in *IEEE Commun. Mag.*
- [10] F. Callegati, **C. Devellder**, W. Cerroni, M. Pickavet, P. Demeester, "Scheduling algorithms for a slotted packet switch with either fixed or variable length packets", aanvaard voor publicatie in *Photonic Network Commun. (PNET)*, Kluwer Academic Publishers.
- [11] C. Matrakidis, **C. Devellder**, A. Stavdas, M. Pickavet, P. Demeester, "Multistage switches for optical packet switching based on SOA technology", to be submitted to *Optical Networks Magazine*.

Publicaties op internationale conferenties:

- [12] P. Demeester, D. Colle, S. Demaesschalck, **C. Devellder**, M. Pickavet, P. Van Heuven, "Resilience in IP over WDM based multilayer networks", (Invited) *Proc. 26st European Conf. on Optical Commun. (ECOC 2000)*, Munich, Germany, 3-7 Sep. 2000.
- [13] D. Colle, **C. Devellder**, P. Van Heuven, M. Pickavet, P. Demeester, L. Raptis, G. Chatziliass, C. Mas, Y.I. Manolessos, J. Comellas, A. Rafel, J. Prat, J. Solé-Pareta, J. Moyano, S. Brunazzi, S. Rotolo, R. Stankiewicz, A. Gladish, "Recovery techniques for IP-over-WDM networks", (Invited) *IP over DWDM Conf.*, Paris, France, 27-30 Nov. 2000.
- [14] D. Colle, P. Van Heuven, A. Groebbens, **C. Devellder**, M. Pickavet, P. Demeester, "Dimensioning reliable IP-over-WDM networks", *IP over DWDM Conf.*, Paris, France, 27-30 Nov. 2000.
- [15] P. Demeester, D. Colle, S. De Maesschalck, **C. Devellder**, M. Pickavet, P. Van Heuven, "Resilience in IP over WDM based multilayer networks", (Invited) *Optical Networks 2000*, Göteborg, Sweden, 7 Dec. 2000.
- [16] **C. Devellder**, Didier Colle, Pim Van Heuven, Steven Van den Berghe, Mario Pickavet, Piet Demeester, "Influence of recovery time on TCP behaviour", *MPLS World Congress*, Paris, France, 6-9 Feb. 2001.
- [17] D. Colle, **C. Devellder**, P. Van Heuven, S. Demaesschalck, A. Groebbens, M. Pickavet, P. Demeester, "Resilience in IP-over-WDM networks", (Invited) *Proc. 5th Working Conf. on Optical Netw. Design and Modelling (ONDM*

- 2001), Vienna, Austria, 5-7 Feb. 2001.
- [18] A. Groebbens, D. Colle, **C. Develder**, S. De Maesschalck, M. Pickavet, P. Demeester, "Use of backup trees to improve resource efficiency of MPLambdaS protection mechanisms", Proc. 3rd Intl. Workshop on Design of Reliable Commun. Netw. (DRCN 2001), Budapest, Hungary, 7-10 Oct. 2001, pp. 152-159.
 - [19] D. Colle, S. De Maesschalck, **C. Develder**, A. Groebbens, M. Pickavet, P. Demeester, "Design and dimensioning issues in survivable data-centric optical networks", (Invited) Proc. 1st COST270 Workshop on Reliability of Optical Netw., Systems and Components, Dubendorf, Switzerland, 12-13 Dec. 2001.
 - [20] **C. Develder**, M. Pickavet, P. Demeester, "Assessment of packet loss for an optical packet router with recirculating buffer", Proc. 6th IFIP Working Conf. on Optical Netw. Design and Modelling (ONDM2002), Torino, Italy, 4-6 Feb. 2002, pp. 247-261.
 - [21] M. Pickavet, A. Ackaert, E. Baert, J. Cheyns, D. Colle, S. De Maesschalck, P. Demeester, **C. Develder**, A. Groebbens, I. Lievens, E. Van Breusegem, S. Verbrugge, Q. Yan, "Design of communication networks using heuristics", (Invited) Book of abstracts of "Arbeitskreis Mathematik in Forschung und Praxis", 23rd Symp. on Mathematik in der Telekommunikation, Duisburg, Germany, 18-19 Mar. 2002.
 - [22] M. Pickavet, **C. Develder**, E. Baert, P. Demeester, "A.I. Techniques for planning telecommunication networks", Proc. Intl. Conf. on Artificial Intelligence (IC-AI'02), Las Vegas, NV, 24-27 Jun. 2002.
 - [23] **C. Develder**, J. Cheyns, E. Van Breusegem, E. Baert, A. Ackaert, M. Pickavet, P. Demeester, "Node architectures for optical packet and burst switching", Tech. Digest Int. Topical Meeting on Photonics in Switching (PS2002), (invited) paper PS.WeA1, Cheju Island, Korea, 21-25 Jul. 2002, pp. 104-106.
 - [24] **C. Develder**, M. Pickavet, P. Demeester, "Strategies for an FDL based feed-back buffer for an optical packet switch with QoS differentiation", Proc. Conf. on Optical Internet (COIN2002), paper COIN.TuD1, Cheju Island, Korea, 21-25 Jul. 2002, pp. 114-116.
 - [25] **C. Develder**, M. Pickavet, P. Demeester, "Choosing an appropriate buffer strategy for an optical packet switch with a feed-back FDL buffer", Proc. 28th European Conf. on Optical Commun. (ECOC2002), Copenhagen, Denmark, 8-12 Sept. 2002.
 - [26] **C. Develder**, J. Cheyns, M. Pickavet, P. Demeester, "Multistage architectures for optical packet switching using SOA-based broadcast-and-select switches", Tech. Digest Optical Fibre Conf. (OFC 2003), paper FS4, Atlanta, GA, USA, 23-28 Feb. 2003, pp. 794-795.
 - [27] J. Cheyns, E. Van Breusegem, **C. Develder**, A. Ackaert, M. Pickavet, P. Demeester, "Performance improvement of an internally-blocking optical packet/ burst switch", Proc. Intl. Conf. on Commun. (ICC2003), Anchorage, AK, 11-15 May 2003, vol. 2, pp. 1304-1308.
 - [28] D. Colle, J. Cheyns, **C. Develder**, E. Van Breusegem, A. Ackaert, M.

- Pickavet, P. Demeester, B. Feng, H. Lønsethagen, E. Zouganeli, J. Fernández-Palacios, J.F. Lobo, "GMPLS extensions for supporting advanced optical networking technologies", Proc. Intl. Conf. on Transparent Optical Networks (ICTON 2003), Warschau, Poland, 29 Jun. – 3 Jul. 2003.
- [29] **C. Develder**, M. Pickavet, P. Demeester, "On trains and wagons: switching variable length packets in a slotted OPS network", Proc. Conf. on Optical Internet / Australian Conf. on Optical Fibre Technology (COIN/ACOFT 2003), paper TuA2-6, Melbourne, Australia, 13-16 Jul. 2003, pp. 233–236.
 - [30] **C. Develder**, R. Van Caenegem, E. Baert, M. Pickavet, P. Demeester, "Active versus passive OPS architectures for metro rings: a network dimensioning point of view", Proc. 29th European Conf. on Optical Commun. (ECOC 2003), paper Mo.4.4.1, Rimini, Italy, 21-25 Sep. 2003, vol. 1, pp. 100–103.
 - [31] **C. Develder**, J. Cheyns, E. Van Breusegem, E. Baert, D. Colle, M. Pickavet, P. Demeester, "Architectures for optical packet and burst switches" (invited), Proc. 29th European Conf. on Optical Commun. (ECOC 2003), paper We.1.4.4, Rimini, Italy, 21-25 Sep. 2003, vol. 3, pp. 376–377.
 - [32] **C. Develder**, J. Cheyns, M. Pickavet, P. Demeester, "Service differentiation mechanisms for variable length packets in an optical switch with recirculating FDL buffer", Tech. Digest Photonics in Switching (PS2003), paper PS.Mo.C8, Versailles, France, 28 Sep.-2 Oct. 2003, pp. 92–94.
 - [33] R. Van Caenegem, **C. Develder**, E. Baert, D. Colle, M. Pickavet, P. Demeester, "Architectures for OPS metro rings: comparing active versus passive nodes / A dimensioning point of view", aanvaard voor 8th IFIP Working Conf. on Optical Network Design and Modelling (ONDM 2004).
 - [34] J. Cheyns, E. Van Breusegem, **C. Develder**, D. Colle, M. Pickavet, P. Demeester, "Evaluating cost functions for OPS node architectures / A packaging driven approach", aanvaard voor 8th IFIP Working Conf. on Optical Network Design and Modelling (ONDM 2004).

Publicaties op nationale symposia:

- [35] D. Colle, **C. Develder**, S. De Maesschalck, A. Groebbens, M. Pickavet, P. Demeester, "Design and planning of reliable communication networks", 1st FTW PhD Symposium, Gent, Belgium, 5 Dec. 2000.
- [36] **C. Develder**, J. Cheyns, E. Van Breusegem, E. Baert, M. Pickavet, P. Demeester, "Optical Packet/Burst Switching", 2nd FTW PhD Symposium, Gent, Belgium, 12 Dec. 2001.
- [37] **C. Develder**, E. Baert, M. Pickavet, P. Demeester, "Optical Packet Switched Networks with Recirculating FDL Buffers", 3rd FTW PhD Symposium, paper 9 (Proceedings available on CD-Rom), Interactive poster session, 11 Dec. 2002, Gent, Belgium.
- [38] A. Groebbens, **C. Develder**, D. Colle, S. Demaesschalck, B. Lannoo, M. Pickavet, P. Demeester, "Optimization of backup tree structures, reducing spare capacity in optical networks, while retaining protection speed", 3rd FTW PhD Symposium, Gent, Belgium, 11 Dec. 2002.

Bijdragen tot internationale onderzoeksrapporten:

- [39] R. Inkret, A. Kuchar, B. Mikac (eds.), E. Baert, S. Bauer, S. Bjørnstad, C. Bungarzenau, F. Callegati, D. Careglio, J. Cheyns, T. Cinkler, D. Colle, S. Demaesschalck, P. Demeester, **C. Develder**, V. Eramo, C. Gauger, D.R. Hjelme, G. Junyent, M. Karasek, M. Klinkowski, M. Köhn, M. Kowalewski, M. Lakovic, M. Ljolje, M. Marciniak, C. Mas, X. Masip-Bruin, F. Matera, M. Mattiello, C. Mauz, H.-P. Nolting, M. Nord, M. Pickavet, B. Puype, C. Raffaelli, S. Sanchez-Lopez, B. Sartorius, D.A. Schupke, M. Settembre, J. Solé-Pareta, N. Stol, S. Tomić, I. Tomkos, A. Tzanakak, S. Verbrugge, Q. Yan, I. Zacharopoulos, P. Zaffoni, “Advanced Infrastructure for photonic networks / Extended final report of Cost 266 action”, 2003.

Hoofdstuk 2

Knooparchitecturen

2.1 Situering

In het voorgaande inleidende hoofdstuk hebben we een kort overzicht gegeven van de verschillende functies die een OPS-schakelaar moet vervullen. In dit hoofdstuk zullen we enkele specifieke implementaties behandelen. Een eerste studie in Sectie 2.2 tracht een oordeel te vellen over een architectuur voor een OPS-schakelaar bestemd voor ringvormige metro-netwerken. We zullen er de zin van ruimtelijk hergebruik en een golflengteband-concept onderzoeken.

In ruggengraatnetwerken is de voorgestelde architectuur typisch geavanceerder dan voor metro-netwerken en situeert de toepassing zich op langere termijn [6]. In deze context bestuderen we in Sectie 2.3 hoe pakketschakelende knopen van hoge capaciteit geconstrueerd kunnen worden door een meertrapsarchitectuur te gebruiken.

De eigen publicaties waar de in dit hoofdstuk samengevatte resultaten in detail worden besproken zijn: [30] (zie Appendix B) en [33] voor sectie 2.2 en voor sectie 2.3 hoofdzakelijk [26] uit Appendix C, alsook [9,11] en in mindere mate [23,31].

2.2 Metro-netwerken

Zoals aangegeven in het inleidend hoofdstuk, loopt een typische verbinding tussen twee communicatie-eindpunten doorheen drie netwerksegmenten: het toegangsnetwerk, een metro-netwerk (MAN) en een ruggengraatnetwerk (Eng. Wide Area Network, WAN). Het toegangsnetwerk zal instaan voor het verzamelen van verkeer van eindgebruikers en beslaat doorgaans hooguit een tiental kilometer. Bij een toegangsnetwerk is de kostprijs de meest dominante factor en zal men het netwerk zo eenvoudig mogelijk houden. Vandaar dat men hiervoor vandaag typisch kopergebaseerde oplossingen (xDSL over telefoonlijnen, coaxiale kabel) vindt, of puur passieve optische netwerken [74,75] waarbij de optische vezel tot in of aan de huizen van de gebruikers kan lopen (Eng. Fiber To The Home, FTTH) [114,115]. Het metro-netwerk zal instaan voor de uitwisseling van informatie tussen verschillende toegangsnetwerken en de verbinding verzorgen met het ruggengraatnetwerk dat informatie over lange afstand (enkele honderden tot duizenden

kilometers) transporteert. De bandbreedte die in een MAN omgaat is dus van een hogere orde dan die in het toegangsnetwerk (dat voornamelijk de functie vervult van aggregatie van verkeersstromen). Hierbij is het nuttig op te merken [79] dat een niet onbelangrijk deel van dit verkeer binnen hetzelfde MAN blijft, zonder via het ruggengraatnetwerk te passeren (zie bv. de traffic-patternen gebaseerd op die van netwerkoperator Telefonica in Spanje [116]). Het metro-netwerk krijgt, vergeleken met het ruggengraatnetwerk, te maken met een verkeerspatroon dat erg in tijd en ruimte fluctueert. Daarenboven moet het diensten van uiteenlopende aard kunnen ondersteunen, gaande van gegevensstromen met gematigde vereisten qua netwerkvertraging (bv. voor Internet-surfers) tot toepassingen met strikte kwaliteitseisen (bv. bedrijfscommunicatie, videofonie). Een optisch pakketgeschakelde aanpak is in een dergelijke dynamische omgeving, die hoge bandbreedte-eisen stelt, zeer zinvol [117]. De metro-omgeving wordt dan ook gezien als de plek bij uitstek waar OPS op relatief korte termijn zijn nut kan bewijzen [6,118].

2.2.1 Alternatieven en motivatie

Een OPS-architectuur voor een metro-netwerk zal de concurrentie moeten aangaan met recente technologieën zoals Ethernet-gebaseerde optische netwerken tot zelfs 10 Gbit/s (realiseerbaarheid aangetoond in experimenten [119] en intussen in standaard IEEE 802.3ae vastgelegd [120]), of robuuste pakket-ringen (Eng. Resilient Packet Rings, RPR [121], momenteel onder standardisatie in IEEE 802.17). Vermits deze technologieën relatief goedkoop zijn, is het aangewezen de complexiteit van een OPS-alternatief beperkt te houden. Kost is hier immers een belangrijk criterium [122]. De recentste voorbeelden van op OPS-gebaseerde metro-architecturen spelen daar dan ook op in en stellen ring-gebaseerde netwerken met bufferloze schakelknoopen voor (zie bv. [77,78,79,118,123,124]).

In het Hornet-project [79] doet men beroep op de pragmatische combinatie van elektronica voor controle en optische componenten voor het eigenlijke transport van de bits en bytes. Op de vezelring worden verschillende golflengten gebruikt (in WDM), waarbij elke knoop op slechts één enkele golflengte gegevens kan ontvangen. Uiteraard kan die knoop wel alle golflengtes aanspreken om gegevens te verzenden, met behulp van een snel afstembare transmitter. Wanneer meerdere knopen dezelfde ontvangstgolflengte gebruiken, is het mogelijk dat een pakket niet rechtstreeks tot bij de uiteindelijke bestemming gezonden wordt: in een tussenliggende knoop die dezelfde golflengte ontvangt als de finale ontvangende knoop zal het pakket van de ring gehaald worden en er nadien opnieuw op worden geplaatst (via een elektronisch buffer). De hoofding van het pakket wordt in Hornet via FSK-modulatie meegestuurd (zie Hoofdstuk 1).

Een ander project, op Japanse bodem, gebruikt een 2×2 schakelaar om pakketten op/van een metro-ring te brengen/halen. In plaats van een orthogonale modulatie-techniek, wordt hier een aparte controle-golflengte benut. Zoals voorgesteld in [123], is er ook slechts één enkele golflengte voor het verzenden van de eigenlijke pakket-inhoud. Het is evenwel duidelijk dat het concept met een grotere schakelmatrix kan worden uitgebreid naar een WDM-ring met meerdere parallelle data-kanalen. De toegang wordt geregeld door een zogenaamde meesterknoop die lege slots op de ring brengt, waarbij aangeduid wordt welke bronknoop dit slot mag

gebruiken. Zogenaamd ruimtelijk hergebruik wordt toegelaten om de bandbreedte zo goed mogelijk te benutten: lege slots mogen ook gebruikt worden voor het verzenden van pakketten tussen twee knopen die gelegen zijn tussen de meesterknoop en de bronknoop waarvoor het lege slot gereserveerd is.

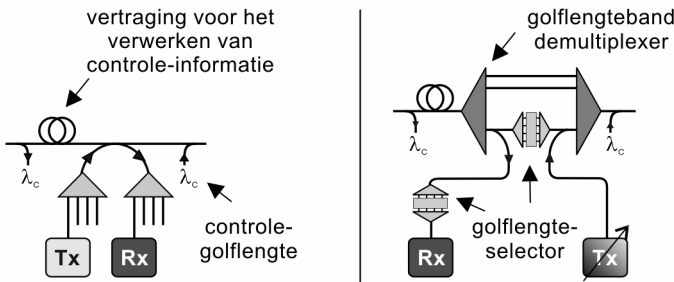
Ook door Europese onderzoekers werden ring-gebaseerde metro-netwerken voorgesteld, die conceptueel grote gelijkenissen vertonen met de zonet behandelde [77,78].

Louter op vlak van functionaliteit kunnen we twee knoop-architecturen onderscheiden: een Actieve Knoop en een Passieve Knoop. In de studie die nu volgt zullen we zo'n actieve en passieve structuur met elkaar vergelijken. De knoop-architecturen werden voorgesteld in [6,118]. Ze gebruiken een aparte golflengte als kanaal voor controlesignalen. Via dit kanaal zal toegang tot het netwerk geregeld worden door een MAC-protocol [6,71]. Die toegang gebeurt op een slot-gebaseerde wijze: het betreft een synchroon OPS-netwerk met pakketten van vaste lengte. (Merk op dat synchronisatie in een ring-structuur, waar alle data-kanalen via dezelfde vezel verstuurd worden, vrij eenvoudig te realiseren is, zeker in vergelijking met willekeurig vermaasde netwerk-structuren.)

De Passieve Knoop heeft als kenmerk dat hij niet in staat is een datastroom fysisch te termineren: het licht kan wel gedeeltelijk afgesplitst worden om de informatie te verwerken (en door te schakelen naar het toegangsnetwerk dat verbonden is met het ring-gebaseerde MAN), maar dat belet het niet verder op de ring te propageren. Met zo'n knoop komt men tot een netwerkarchitectuur die beduidend afwijkt van voorgaande geciteerde voorbeelden. Het concept dat men met zo'n passieve architectuur kan realiseren, betitelt men in [124,125] als DBORN, wat staat voor "Dual Bus Optical Ring Network". Kenmerkend is dat het golflengte-spectrum gescheiden wordt voor zenden en ontvangen. Een centrale knoop, die men Hub noemt, staat in voor de golflengte-conversie van het zend- naar het ontvangst-spectrum. Deze Hub kan ook meerdere MAN-ringen met elkaar verbinden. De "gewone" ringknoten zelf hebben de erg eenvoudige structuur die geschetst is in het linker deel van Figuur 2.1: via splitters en koppelaars wordt licht van de vezel getapt, resp. erin gekoppeld, om zogenaamde add/drop-functionaliteit te voorzien. Er zijn geen actieve schakelcomponenten in het lichtpad dat de knoop doorkruist. Door zijn eenvoud kan die architectuur gerealiseerd worden met technologie die vandaag al op punt staat en zelfs commercieel beschikbaar is. Om de capaciteit van de knoop uit te breiden, kunnen extra zend- of ontvangstelementen (Tx, resp. Rx) toegevoegd worden.

De Actieve Knoopstructuur die we beschouwen is weergegeven in de rechtse helft van Figuur 2.1. Het binnenkomende lichtsignaal wordt, na de klassieke aftapping van het controle-signaal, naar een golflengteband-demultiplexer gezonden. Deze splitst het licht in verschillende spectrale delen, de golflengtebanden. In onze studie beschouwen we golflengtebanden die elk vier afzonderlijke golflengten bevatten. Een knoop heeft toegang tot ten minste één zo'n band, waarvoor er precies één zendende laser is met een snel instelbare golflengte en één ontvanger. (In Figuur 2.1 is de ontvanger golflengte-ongevoelig en wordt er via een SOA-schakelrij slechts één golflengte tegelijk doorgelaten.) In het doorgaande pad, tussen de golflengteband-demultiplexer en -multiplexer, wordt het licht afgesplitst dat

bestemd is voor de ontvanger (Rx) en wat afkomstig is van de zender (Tx) wordt erin gekoppeld. Het cruciale verschil met de passieve architectuur is dat er tussen dit passief splitsen en koppelen een rij van SOA-schakelaars geplaatst is. Deze worden gebruikt om pakketten die voor de betreffende knoop bestemd zijn niet langer te laten verder propageren op de ring. Het geheel van Rx, Tx en de SOA-elementen die voor een bepaalde band voorzien zijn, kunnen geïntegreerd worden op wat we een baby-bord noemen [126]. De capaciteit van de knoop kan verhoogd worden door meerdere baby-borden te installeren en dus tot meerdere golflengtebanden toegang te bieden.

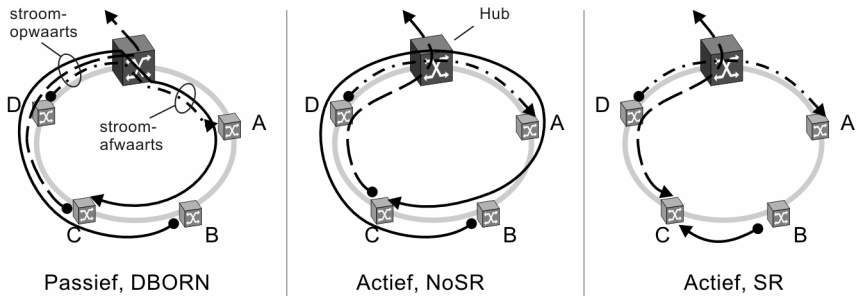


Figuur 2.1: Passieve vs Actieve Knoopstructuur.

Met een Actieve Knoopstructuur kan men, door de mogelijkheid pakketten daadwerkelijk van de ring te halen (in tegenstelling tot DBORN, waar enkel de centrale Hub daartoe in staat is), het eerder vermelde ruimtelijk hergebruik (Eng. spatial reuse, SR) uitbuiten. Dit concept impliceert dat eenzelfde golflengte op de ring gelijktijdig voor meerdere connecties kan worden benut. Dit kan zelfs wanneer die elk afzonderlijk de capaciteit van een volledige golflengte vereisen, zoals voor een connectie tussen knopen B en C enerzijds en D en A anderzijds, in het uiterst rechte deel van Figuur 2.2. In die figuur schetsen we de drie netwerkconcepten die we verder met elkaar zullen vergelijken. Het eerste is DBORN, met de intussen welbekende spectrale scheiding van zenden en ontvangen. Een tweede maakt gebruik van de Actieve Knoopstructuur, maar onderstelt een gecentraliseerd MAC-protocol waar al het verkeer doorheen de Hub moet passeren. Dit impliceert dat sommige verkeersstromen eenzelfde linksegment meermaals moeten doorkruisen, zoals de stroom tussen B en C in Figuur 2.2. Wanneer het MAC-protocol volledig gedistribueerd is, kan die beperking van een centraal doorgangspunt opgeheven worden. (Merk op dat dit de complexiteit van het toegangsprotocol aanzienlijk kan verhogen, zeker wanneer de zogenaamde “eerlijkheid” gegarandeerd moet worden zodat één knoop de toegang tot het netwerk niet zou monopoliseren en evenmin de positie van de knoop in de ring z’n bandbreedtegebruik zou beïnvloeden.) In dat geval is ruimtelijk hergebruik mogelijk en duiden we de netwerk-architectuur aan met SR, terwijl het voorgaande geval NoSR genoemd zal worden.

De bedoeling—en meteen ook de originele bijdrage—van deze studie is de drie architecturen met elkaar te vergelijken in termen van de hoeveelheid middelen die nodig zijn om een gegeven hoeveelheid verkeer op te zetten. De drijfveer voor deze studie is te achterhalen welke architectuur de meest kost-efficiënte is: de kostprijs

(i.e. kapitaal, Eng. capital expenditure, CAPEX) van een netwerk dat aan een reële vraag moet beantwoorden zal immers sterk gerelateerd zijn aan de hoeveelheid middelen.



Figuur 2.2: Drie netwerkconcepten: (i) DBORN met de Passieve Knoop, (ii) Actieve Knoop zonder ruimtelijk hergebruik, NoSR en (iii) met ruimtelijk hergebruik, SR.

2.2.2 Probleembeschrijving en oplossingsmethodiek

Het probleem dat we moeten oplossen voor de vergelijking van de onderscheiden architecturen betreft een netwerk-planningsprobleem:

- **Gegeven:**

- Een verzameling knopen V die een metro-netwerk zal vormen (dat eventueel meerdere ringen kan beslaan);
- Een verzameling (kandidaat-)linken tussen elementen v en w van V ;
- Een twee-dimensionale vraagmatrix D , waarbij elke rij i en elke kolom j overeenstemt met een knoop v , resp. w uit V zodat het element $D_{i,j}$ aangeeft hoeveel bandbreedte er van v naar w verzonden moet worden.

- **Gevraagd:**

- Welke knopen zullen door middel van welke verbindingen verbonden worden tot metro-ring(en)?
- Hoeveel capaciteit zal daarvoor moeten worden geïnstalleerd? (Hierbij wordt met capaciteit bedoeld: het aantal gebruikte golflengten op de link(en) en de zend- en ontvangstcapaciteit in de ringknopen.)

Om dit probleem op te lossen hebben we—vermits de oplossing van een ILP formulering van het probleem te tijdrovend bleek—gebruik gemaakt van heuristieken geïnspireerd op het werk voorgesteld in Hoofdstuk 6 van [76].

We hebben de oplossing onderverdeeld in twee stappen. Een eerste stap omvatte het bepalen van de benodigde capaciteit voor een gegeven vraag in één enkele gegeven ring. Een tweede stap bestond erin het volledige ring-planningsprobleem aan te pakken door mogelijke ring-topologieën te genereren en van het in de eerste stap ontwikkelde algoritme gebruik te maken om de “kost” ervan te bepalen en tot meer geschikte keuzes van ringen te komen. Voor een meer gedetailleerde beschrijving van de algoritmes verwijzen we met graagte naar [126,127].

De beschouwde indicatoren van de kost van het resulterende metro-netwerk zijn hieronder samengevat:

- **Rx/Tx-capaciteit:** Voor de Actieve Knoopstructuur (NoSR, SR) is dit het aantal geïnstalleerde baby-borden. Voor de Passieve architectuur (DBORN) telden we het totale aantal ontvangst- en zend-elementen en deelden we dit vervolgens door twee. De factor $\frac{1}{2}$ zorgt voor een min of meer eerlijke vergelijking met de actieve knoop-architectuur, aangezien één baby-bord daar precies één ontvangst- (Rx) en één zend-element (Tx) bevat.
- **Linkcapaciteit:** Dit omvatte het tellen van het aantal golflengten gebruikt op elk ringsegment (i.e. elke link) en het sommeren ervan over het gehele metro-netwerk. Bemerkt dat voor de Actieve Knoop dit steeds een viervoud zal zijn: capaciteit wordt er geïnstalleerd per golflengteband en we brengen alle vier de golflengten van zo'n band in rekening, ook al worden ze niet volledig benut.
- **Aantal gebruikte golflengten:** In tegenstelling tot de linkcapaciteit tellen we hier het aantal verschillende golflengten per vezelring (gesommeerd over alle ringen indien er meerdere zijn). Dit aantal is een indicatie voor de kost van de centrale Hub, vermits de grootte van de optische schakelmatrix erdoor bepaald zal worden. Ook hier tellen we voor de Actieve Knoopstructuur alle golflengten in een band mee.

De dimensionerings-algoritmes die we gebruiken zijn er in de eerste plaats op gericht de linkcapaciteit en het aantal gebruikte golflengten in het MAN te minimaliseren en slechts in tweede instantie de Rx/Tx-capaciteit. Onze resultaten gaven immers aan dat het precies die kostmaten waren die erg schommelden voor verschillende ring-configuraties en sterk beïnvloed werden door gepaste keuzes in de planning.

Het ligt voor de hand dat het antwoord op de vraag welke architectuur het efficiëntst is, en dus bijvoorbeeld het minst aantal golflengten zal nodig hebben, kan afhangen van het precieze verkeerspatroon waarvoor we een metro-netwerk dimensioneren. Om die invloed van de vraagmatrix te kunnen inschatten hebben we dan ook meerdere verkeerspatronen beschouwd, met name:

- **Uniform:** In dit geval zal elk element D_{ij} van de vraagmatrix gelijkgesteld worden aan dezelfde hoeveelheid bandbreedte d , waarbij we de kost van het netwerk zullen bekijken voor toenemende d -waarden. Verkeer dat door een bepaalde knoop op de ring wordt getransporteerd of ervoor bestemd is, zal hier gelijkmatig verdeeld zijn over alle andere knopen.
- **Server:** Een uitgesproken asymmetrische verdeling van het verkeer zal hier ondersteld worden: alle verkeer komt van, of is bestemd voor, één bepaalde knoop. In termen van de vraagmatrix betekent dit dat $D_{i,s}=D_{s,j}=d$ voor een zogenaamde server s , terwijl alle andere vragen $D_{i,j}=0$ (met $i,j \neq s$). Opnieuw zullen we d laten toenemen.
- **Buren:** Zoals de naam aangeeft zal een verkeersstroom in dit geval steeds tussen twee aangrenzende knopen vloeien. We beschouwen (i) het geval waar verkeer stroomt in de circulatie-richting van het licht in de ring, ($D_{i,i-1}=d$, overige $D_{i,j}=0$) en (ii) het geval waar verkeer van een knoop bestemd is voor zijn voorganger ($D_{i,i+1}=d$, overige $D_{i,j}=0$). Bemerkt dat het onderscheid tussen (i) en (ii) enkel een

rol zal spelen wanneer de uni-directionele ringen gebruikt worden, maar niet wanneer in een ring twee tegengesteld roterende vezels voorzien zullen worden (zie verder in 2.2.3).

- **3Buren:** Hier zal een knoop enkel verkeer uitwisselen met drie buurknoten. De elementen van de vraagmatrix worden in dit geval als volgt gedefinieerd: $D_{i, i+1}=D_{i, i+2}=D_{i, i+3}=d$, terwijl andere $D_{i,j}=0$ zijn.
- **Random:** We zullen elk element $D_{i,j}$ van de vraagmatrix willekeurig genereren, waarbij de bandbreedte voor een knopenpaar (i,j) genomen wordt op basis van een uniforme distributie over het interval $[0,2d]$, met als gemiddelde waarde d .
- **Reël:** Dit is een specifieke vraagmatrix die in [116] werd voorgesteld en bovendien behoorlijk realistisch bleek te zijn, in de zin dat hij zeer gelijklopend was aan een verkeerspatroon tussen ADSL concentratiepunten van de telecom-operator Telefonica in de omgeving van Madrid [116].

2.2.3 Uni- versus bidirectionele ringen

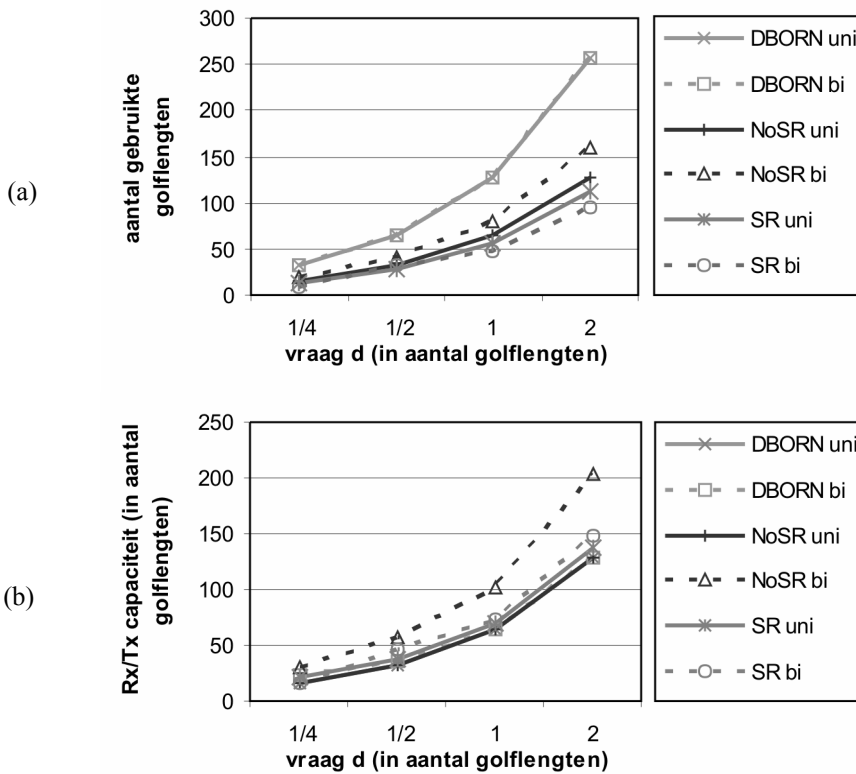
Een eerste vraag die we op basis van de dimensionerings-resultaten wilden beantwoorden was of het zin heeft twee tegengesteld roterende vezelringen te gebruiken. Dit wil zeggen dat lichtsignalen, dus gegevens, in twee richtingen doorheen de ring circuleren. Het is een techniek die zeer duidelijk nuttig is om de ring robuust te maken door gegevensoverdracht ook mogelijk te maken in geval van bijvoorbeeld een vezelbreuk [76,128,129]. Hier beogen we evenwel te achterhalen of het concept ook nuttig is, zelfs onder fout-vrije voorwaarden, om de hoeveelheid benodigde hardware te beperken.

In de grafieken in Figuur 2.3 en Figuur 2.4 wordt een selectie getoond van onze resultaten voor het Uniforme resp. Random verkeerspatroon in een enkele ring. De curves in volle lijn geven de kost aan voor unidirectionele ringen, terwijl de stippellijnen de resultaten voor bidirectionele ringen aanduiden. Wanneer in het geval van bidirectionele ringen een Tx-element, of in geval van de Actieve Knoopstructuur een baby-bord, wordt geïnstalleerd, moet beslist worden op welke van de twee tegengesteld roterende ringen het geplaatst zal worden. In de gebruikte heuristiek [127] wordt eerst Rx/Tx-capaciteit toegevoegd voor vragen tussen knopen waarvoor het kortste pad in de rotatie-richting van de ring ligt, waarna de resterende vrije capaciteit op die ring zoveel mogelijk wordt ingevuld door andere vragen. De overige te voorziene bandbreedte wordt dan langs de andere ring geïnstalleerd.

Voor het DBORN-geval leren de grafieken dat het gebruiken van bidirectionele ringen geen winst oplevert in termen van de te installeren capaciteit. Zowel de Rx/Tx-capaciteit als het aantal golflengten kan niet gereduceerd worden door tegengesteld roterende ringen te voorzien. In tegendeel zelfs²: we stellen een toename van Rx/Tx-capaciteit vast en een lichte stijging van het aantal gebruikte golflengten.

² Bemerk dat we voor het Uniforme vraagpatroon nauwelijks verschil waarnemen. Dit is evenwel te wijten aan het feit dat alle vragen precies even groot zijn en de bandbreedte van een golflengte steeds volledig kan worden ingevuld.

In geval van de Actieve Knoopstructuur in NoSR vinden we dat het installeren van bidirectionele ringen de Rx/Tx-capaciteit beduidend verhoogt en in iets minder uitgesproken mate ook het aantal gebruikte golflengten. Het enige voordeel dat bidirectionele ringen ons in dit geval opleveren, is dat de totale linkcapaciteit (i.e. het aantal gebruikte golflengten per link, gesommeerd over alle linken) licht gereduceerd kan worden, omdat een deel van het verkeer het kortste pad tussen bron en bestemming kan volgen. Het voordeel daarvan is evenwel beperkt, vermits alle verkeer doorheen de centrale Hub moet passeren. De spreekwoordelijke flessenhals wordt gevormd door de linken die verbonden zijn met de Hub. Dit is een effect waaraan niet afdoende verholpen kan worden door het verkeer in twee richtingen te laten stromen.

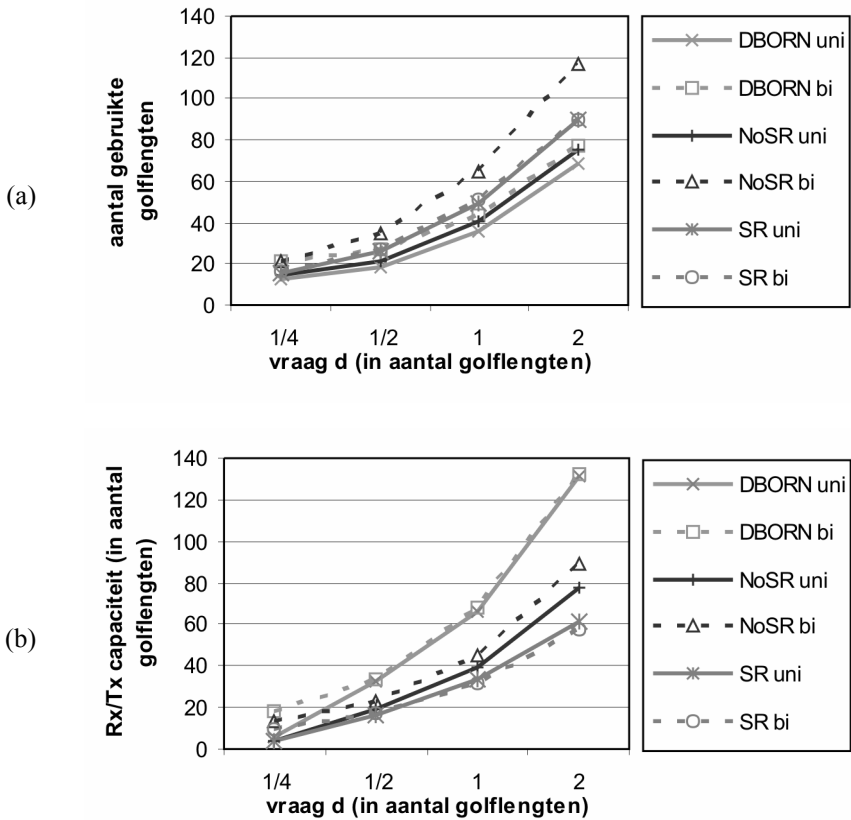


Figuur 2.3: Resultaten voor Uniform verkeer, enkele ring.

In het SR-geval bewijst het gebruik van bidirectionele ringen wél zijn voordeel. Dit is vooral het geval wanneer de fractie verkeer tussen naburige knopen binnen eenzelfde ring toeneemt (bv. het Buren patroon), eerder dan in min of meer symmetrische verdelingen van het verkeer (zoals Uniform of Random). Doordat ruimtelijk hergebruik toegelaten is, kan verkeer tussen bron en bestemming op

eenzelfde ring het kortste pad volgen. Dit laat toe de ongebruikte capaciteit ten volle te benutten.

We kunnen besluiten dat, vanuit een Rx/Tx-capaciteit of aantal gebruikte golflengten, het enkel bij ruimtelijk hergebruik (SR) voordeel kan opleveren bi-directionele ringen te installeren. En dan vooral wanneer er voldoende verkeer stroomt tussen knopen die tot dezelfde ring behoren.



Figuur 2.4: Resultaten voor Random verkeer, enkele ring.

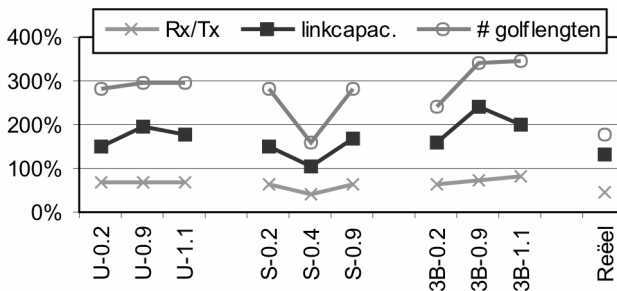
2.2.4 Ruimtelijk hergebruik

Het kenmerkendste verschil tussen de Actieve en Passieve Knoop-architectuur is dat de eerste een pakket daadwerkelijk uit de ring kan wegnemen en hierdoor ruimtelijk hergebruik toelaat. Het ligt dan ook voor de hand de vraag te stellen of dit grote voordelen oplevert. Om het antwoord daarop te achterhalen, dimensioneerden we ringen waarbij we voor de Actieve Knoopstructuur het concept van golflengtebanden achterwege lieten: we namen het aantal golflengten per band $B = 1$. De Passieve Knoopstructuur kent dit bandenconcept immers ook niet, zodat dit de

vergelijking in zekere zin eerlijker maakt. De invloed van het gebruik van het golflengtebandconcept wordt verder besproken in volgende Sectie 2.2.5.

De verkeerspatronen die we in deze studie beschouwden zijn Uniform (U), Server (S), 3Buren (3B) en Reëel. Om de vergelijking tussen Actieve en Passieve Knoopstructuren visueel te vereenvoudigen, hebben we er in Figuur 2.5 voor gekozen de kost voor de Passieve architectuur te delen door die voor de Actieve structuur (voor dezelfde vraag) en deze verhouding uit te zetten. Wanneer deze dus onder 100% ligt, betekent dit dat de Passieve structuur voordelig is, althans voor die betreffende kostmaat. Elk van de drie kostmaten (Rx/Tx-capaciteit, totale linkcapaciteit en aantal gebruikte golflengten) is uitgezet. De labels op de X-as geven het betreffende vraagscenario aan waarbij in geval van een letter- en cijfercombinatie de letter verwijst naar het vraagpatroon en het getal naar de waarde van d in de elementen van de vraagmatrix.

De resultaten leren ons dat de Passieve structuur enkel kan leiden tot een lagere Rx/Tx-capaciteit. De reden voor de hogere kost die geassocieerd is met de Actieve architectuur bestaat erin dat, om ruimtelijk hergebruik toe te laten, bron en bestemming dezelfde golflengte moeten gebruiken (cf. $B = 1$ impliceert een Tx van vaste golflengte), waardoor soms een extra Rx/Tx element voorzien moet worden. Het ruimtelijk hergebruik en dus de Actieve Knoopstructuur, bewijst enkel zijn nut wanneer de kost van het metro- netwerk gedomineerd wordt door de te installeren linkcapaciteit of het aantal golflengten per ring. Het is duidelijk dat dit voordeel volgt uit het niet spectraal scheiden³ van golflengten voor zenden en ontvangen en het beter delen van die golflengtecapaciteit door ruimtelijk hergebruik.



Figuur 2.5: Kostverhouding Passieve/Actieve architectuur.

2.2.5 Golflengtebanden

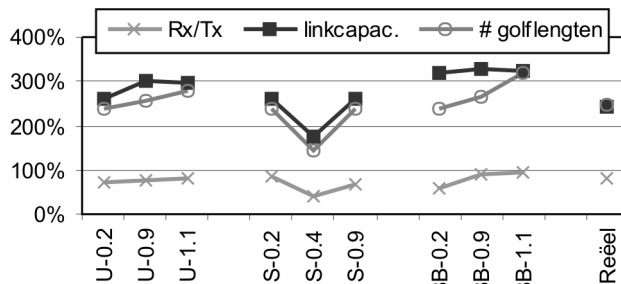
Zoals in voorgaande sectie aangegeven, is een tweede verschilpunt tussen de Actieve en Passieve architectuur dat in de eerste het gebruik van golflengtebanden wordt voorgesteld. In de vorige sectie hebben we voor de Actieve structuur

³ Merk op dat het splitsen van het spectrum—in golflengten exclusief voor ofwel zenden, ofwel ontvangen (met golflengteconversie in één centrale knoop, Hub genaamd)—betekent dat er minimaal twee golflengtes moeten worden voorzien, terwijl een actieve architectuur genoeg heeft aan één golflengte in de ring.

golflengtebanden van één enkele golflengte bestudeerd ($B=1$). Hier beschouwen we de invloed van het gebruik van golflengtebanden, door het geval $B=1$ te vergelijken met het in [116] voorgestelde gebruik van $B=4$.

In Figuur 2.6 zetten we opnieuw een verhouding van de kostmaten uit tussen de twee scenario's in kwestie: we delen de kost voor gebruik van golflengtebanden ($B=4$) door die corresponderend met het geval zonder banden ($B=1$). De vraagscenario's zijn dezelfde als vorige sectie.

Uit die grafiek blijkt dat het voornaamste voordeel van het gebruik van golflengtebanden bestaat uit de reductie van de benodigde Rx/Tx-capaciteit. Maar wanneer het belangrijkste aandeel van de kapitaalkost voortkomt uit het voorzien van linkcapaciteit of het aantal golflengten gebruikt in de ring, is een golflengtebandconcept blijkbaar niet zinvol. Het aantal gebruikte golflengten ligt immers beduidend hoger dan het band-loze concept. Dit wijst erop dat het ruimtelijk hergebruik binnen een gegeven golflengteband beperkt blijft. De achterliggende reden hiervoor is dat in het voorgestelde bandenconcept er per knoop slechts één Rx/Tx element aanwezig kan zijn.



Figuur 2.6: Kostverhouding met/zonder gebruik van golflengtebanden (met: $B=4$; zonder: $B=1$ golflengte(n) per band).

2.3 Ruggengraatnetwerk

In dit onderdeel zullen we onze aandacht toespitsen op een knoop-architectuur voor het ruggengraatnetwerk. Het aantal in- en uitgangspoorten van een knooppunt in zo'n netwerk zal ook typisch hoger zijn dan in het geval van de zonet behandelde ring-gebaseerde metro-netwerken.

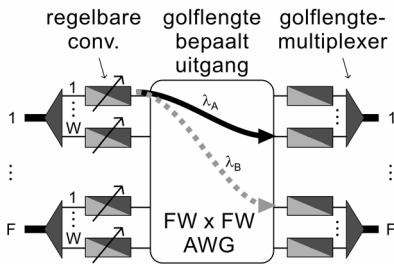
2.3.1 Alternatieven en motivatie

In het inleidende Hoofdstuk 1 gaven we al aan dat er twee architecturen zijn die in de literatuur overheersen als architectuur voor een OPS-schakelmatrix: AWG-gebaseerde enerzijds en SOA-gebaseerde "broadcast-and-select"-structuren anderzijds. De voornaamste motivatie voor de keuze van deze technologieën is dat alternatieven de voor OPS vereiste schakelsnelheden in de orde van nanoseconden niet kunnen halen. MEMS [46] hebben typische schakeltijden van de orde van een milliseconde of meer en de bubblejet-gebaseerde [87] en ook andere thermo-

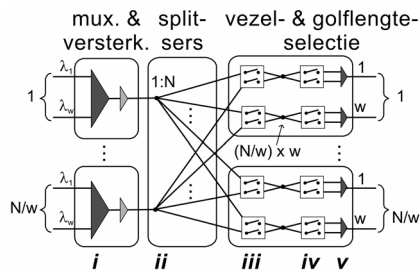
optische schakelmatrixes [88–90] situeren zich in dezelfde of nog tragere snelheidsklasse (tot een tiental milliseconden). Bovendien zijn zowel de SOA-technologie [130] als de voor een AWG-aanpak obligate golflengte-convertoren [131] behoorlijk robuust, terwijl aan de stabiliteit en levensduur van bijvoorbeeld thermo-optische realisaties wel eens getwijfeld wordt.

Een typische schakelstructuur die gebruik maakt van een AWG, is geschetst in Figuur 2.7. Door de regelbare golflengte-convertoren aan de ingang in te stellen op de gepaste golflengte zal bepaald worden langs welke uitgangspoort van de AWG het licht de schakelmatrix verlaat.

Het principe van een “broadcast-and-select”-matrix, zoals die werd voorgesteld in [93], is geschetst in Figuur 2.8. De inkomende golflengten worden in een verzamel-stadium (i) via multiplexers in groepen van w samengenomen en (ii) via een splitter (na versterking via een EDFA) verdeeld over alle uitgaande poorten. Voor elk van die poorten wordt door middel van twee stadia SOA-schakelaars één enkel ingangssignaal overgehouden. Het eerste SOA-stadium (iii) houdt een signaal over dat alle golflengten bevat die in één van de de multiplexers werden gekoppeld. Het tweede SOA-stadium (iv) laat dit signaal slechts naar één ingangspoort van een AWG-component (v) propageren, zodat aan de uitgang effectief slechts één van de oorspronkelijke ingangssignalen wordt overgehouden. Om conflicten te vermijden, zal golflengte-conversie (als deel van regeneratoren, niet getoond in Figuur 2.8) ervoor zorgen dat elk van de signalen die in eenzelfde uitgaande vezel zal worden gekoppeld een verschillende golflengte heeft.



Figuur 2.7: AWG-gebaseerde schakelmatrix.



Figuur 2.8: SOA-gebaseerde “broadcast-and-select” matrix.

2.3.2 Schalen door gebruik van Clos-netwerken

Toen de inspirerende publicatie van Clos [132] verscheen was er nog lang geen sprake van optisch pakketschakelen, maar zijn aanpak—die erop gericht was telefooncentrales met erg veel in- en uitgaande koperparen met een minimaal aantal kruisschakelaars te realiseren—vindt toch toepassing in het OPS-domein. Wanneer we de grote aantallen golflengten die per vezel kunnen worden getransporteerd [44,45] en de gestage groei van de vraag naar bandbreedte in acht nemen, is het niet ondenkbaar dat er ook voor OPS relatief grote schakelmatrixes nodig zullen zijn. Om de kost te drukken, of—zoals ook voor Clos het geval was—technologische

beperkingen te omzeilen, kan een schakelmatrix met een groot aantal poorten geconstrueerd worden door een interconnectie van kleinere schakelmatrices.

Vanzelfsprekend hebben ook andere onderzoekers dit idee om meerdere schakelstadia te hanteren in optische schakelaars bestudeerd. In [133] wordt een theoretische behandeling gegeven van schakelaars in drie domeinen: tijd, ruimte en golflengte. Er wordt besproken hoe deze gerealiseerd kunnen worden door een netwerk van schakelaars: ofwel door architecturen met componenten die in elk van de drie dimensies afzonderlijk schakelen, ofwel door een combinatie van tijd-ruimte schakelaars en ruimte-golflengte schakelaars.

Geïnspireerd op Clos behandelt MacDonald in [134] de uitbreidbaarheid van een optische kruisschakelaar (OXC). Terwijl in een typisch Clos-netwerk de grootte van de schakelmatrix in verschillende stadia verschilt, stelt MacDonald voor standaard bouwblokken te gebruiken (SKOL genoemd) en een modulaire architectuur te realiseren: indien nodig kan de capaciteit van een OXC uitgebreid worden door extra bouwblokken toe te voegen. Het nadeel van deze werkwijze is dat de maximale grootte van de OXC en de grootte van een bouwblok gerelateerd zijn, zodat in een initiële configuratie het bouwblok eventueel niet ten volle benut wordt.

In [135] wordt dan weer aandacht besteed aan de betrouwbaarheid van een MEMS circuitgeschakelde optische matrix. Net zoals bij Clos is een motivatie voor het invoeren van meerdere schakelstadia de reductie van het aantal componenten (hier MEMS spiegeltjes). Een bijkomend voordeel is hier dat zo'n schakelaar met meerdere trappen resulteert in een betere betrouwbaarheid: het aantal connecties dat verloren gaat voor een gegeven faalkans van een spiegeltje ligt beduidend lager. Eveneens toegespitst op het bepalen van de betrouwbaarheid van MEMS schakelknopen is [136]. Nog in een MEMS context worden in [137] interconnectie patronen voorgesteld om het verlies van optische signaalsterkte zo laag mogelijk te houden.

Een ander resultaat in de sfeer van circuitschakelen is [138]. Daar wordt een architectuur voorgesteld die geoptimaliseerd is voor zogenaamd full-duplex verkeer, wat betekent dat voor elke verbinding van A naar B er ook één in de omgekeerde richting loopt. De schakelmatrices, die als bouwblok worden gebruikt in een schakelaar in meerdere stadia, worden er geacht licht in twee richtingen door te laten. Door dit aspect uit te buiten kan men in de resulterende architectuur het aantal componenten die nodig zijn voor een klassiek Clos-netwerk quasi halveren.

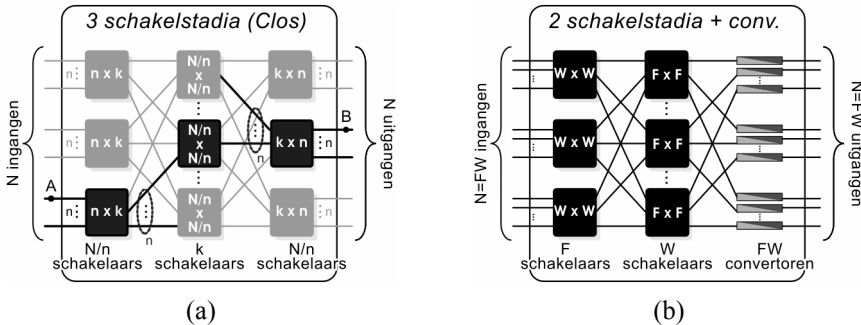
Het werk van [139] geeft aan hoe grote schakelmatrices voor golflengte-schakelen met gebruik van WDM geconstrueerd kunnen worden. De voorgestelde architectuur voor golflengte-schakelende netwerken zonder golflengte-conversie maakt gebruik van drie types bouwblokken die allen gebaseerd zijn op FBG's (Eng. Fibre Bragg Grating) en door de fundamenteel verschillende werking niet te vergelijken zijn met hoger vermelde AWG- en B&S-architecturen. Om het aantal bouwblokken te minimaliseren wordt er een Clos-architectuur voorgesteld, waarbij de maximaal haalbare dimensie (rekening houdend met de fysische karakteristieken) wordt geschat en modulariteit wordt besproken.

Onze eigen studie die in de volgende subsectie wordt voorgesteld verschilt van bovenvermelde werken door zich specifiek toe te spitsen op de "broadcast-and-select"-architectuur voor optisch pakketschakelen. We tonen aan hoe het gebruik

van meerdere schakelende trappen technologische en/of fysische beperkingen kan omzeilen. We zullen hiervoor schakelmatrixes in drie stadia gebruiken.

De generieke structuur van een Clos-netwerk met drie schakelstadia is geschetst in Figuur 2.9(a) voor een $N \times N$ schakelmatrix. De N in- en uitgangen zijn gegroepeerd in groepjes van n , waarbij voor elk van die groepen in het eerste en het laatste schakelstadium een schakelmatrix voorzien is. Het centrale schakelstadium omvat k bouwblokken die elk één verbinding hebben naar elk van de N/n blokken in het eerste en laatste stadium. De dimensies van de bouwblokken zijn dus $n \times k$, $N/n \times N/n$ en $k \times n$ in respectievelijk het eerste, het tweede en het laatste schakelstadium. De keuze van N zal de grootte van de uiteindelijke schakelaar bepalen, terwijl de keuze van n het totale aantal componenten zal beïnvloeden (zie verder). De waarde van k zal de blokkeringsgraad bepalen.

Zoals aangegeven in [132] kunnen verschillende vormen van blokkering worden onderscheiden. Een schakelmatrix wordt *strikt niet-blokkerend* (Eng. strictly non-blocking) genoemd wanneer het steeds mogelijk is een verbinding te maken tussen een nog vrije ingang en een nog vrije uitgang, onafhankelijk van en zonder te raken aan eerder gemaakte verbindingen. In *herschikbaar niet-blokkerende* matrices (Eng. rearrangeable non-blocking) kan een verbinding tussen een ingangs- en een uitgangspoort die beide nog vrij zijn ook steeds gerealiseerd worden, maar het is mogelijk dat dit het herconfigureren van eerder gemaakte verbindingen vereist. Wanneer niet kan gegarandeerd worden dat ten allen tijde een connectie gemaakt kan worden tussen vrije poorten, spreekt men van *intern blokkerende* architecturen (Eng. internally blocking).



Figuur 2.9: Schakelmatrixes in meerdere stadia: (a) een Clos-netwerk in drie stadia en (b) een structuur met twee schakeltrappen en één golf lengte-conversie-trap.

De voorwaarde⁴ opdat de structuur in Figuur 2.9a strikt niet-blokkerend zou zijn is $k \geq 2n-1$. Deze voorwaarde moet voldaan zijn wanneer de schakelaar asynchroon opereert en verbindingen worden opgezet en afgebroken op willekeurige ogenblikken ten opzichte van elkaar, zoals bij circuitschakelen of OBS. In die

⁴ Veralgemeende voorwaarden voor Clos-netwerken in drie stadia waarbij de interne verbindingen hogere transmissiesnelheden hebben en/of er parallele kanalen zijn, worden beschreven in [142].

gevallen zal het inderdaad zo zijn dat de (levens)duur van connecties (i.e. circuits, of bursts in het geval van OBS) verschillen en het begin en einde ervan niet met elkaar samenvallen. Dit impliceert dat, eenmaal beslist is hoe de verbinding gerealiseerd wordt (in het bijzonder via welke schakelmatrix in het tweede Clos-stadium), we deze keuze moeten aanhouden voor de volledige duur ervan. Het is duidelijk dat eenerschikbaar niet-blokkerende structuur niet voldoet.

Wanneer de matrix evenwel gebruikt wordt in een slot-gebaseerd regime volstaat eenerschikbaar niet-blokkerende structuur wel [140]. In dit geval kan men in elk tijdslot voor alle op te zetten verbindingen gezamenlijk beslissen hoe ze worden gerealiseerd in het Clos-netwerk van schakelaars en dit onafhankelijk van voorafgaande of volgende tijdsloten. De voorwaarde voorerschikbaar blokkeren in de Clos-structuur van Figuur 2.9a is $k=n$. Het bewijs hiervoor is een welbekend resultaat uit de grafentheorie: het probleem van het vinden van een voldoende aantal schakelmatrices in het centrale stadium is equivalent met een graafkleuringsprobleem in een bipartiete graaf van graad n en die is kleurbaar met n kleuren [141].

Wanneer het een schakelmatrix betreft voor OPS waarbij meerdere, zeg W , golflengten per vezel gebruikt worden (in WDM, zie Hoofdstuk 1), is het doorgaans toegelaten golflengte-conversie toe te passen. Dit betekent dat een inkomend pakket op een willekeurige golflengte de schakelaar mag verlaten, zo lang dit maar via de correcte uitgaande vezel gebeurt. In dit geval kunnen we het laatste schakelstadium in de Clos-structuur elimineren en vervangen door golflengte-convertoren, met elk een vaste uitgaande golflengte. De voorwaarde is wel dat het aantal poorten n (waarbij een poort overeenstemt met één enkele golflengte), dat we in het eerste en laatste stadium bundelen, zo gekozen wordt dat ze alle n bij een zelfde vezel horen. De resulterende architectuur voor de keuze $n=W$ is geschetst in Figuur 2.9b.

2.3.3 Clos-gebaseerde B&S schakelmatrix

Het voornaamste pijnpunt van de in Figuur 2.8 geschetste “broadcast-and-select”-architectuur is de aanwezigheid van een splitter die een inkomend signaal verdeelt over elk van de uitgaande poorten. Het is duidelijk dat, omwille van de resulterende verzwakking van het lichtsignaal, we het aantal uitgaande poorten niet willekeurig hoog zullen kunnen kiezen. Wanneer we heel grote schakelmatrices willen construeren zullen we daarom Clos-gewijs verschillende kleinere bouwblokken met elkaar verbinden tot een groter geheel. Vanzelfsprekend zal tussen opeenvolgende stadia dan het signaal moeten geregenereerd worden (bv. met volledig optische 3R-regeneratoren [56]).

Een bijkomend voordeel is dat zo’n structuur met meerdere stadia ook de kost kan drukken ten opzichte van een monolithische architectuur (zelfs al mocht die desondanks splitsingsverliezen werkbaar zijn). In de veronderstelling dat de SOA-schakelcomponenten de kost zullen domineren, hebben we drie architecturen met elkaar vergeleken in functie van het aantal SOA-elementen dat ze vereisen: (i) de B&S-schakelaar in één stadium, zoals Figuur 2.8, (ii) een schakelaar in drie stadia, zoals Figuur 2.9a, (iii) een schakelaar met twee schakelstadia en een derde stadium bestaande uit golflengte-convertoren zoals in Figuur 2.9b.

Voor de B&S-architectuur met één schakelstadium wordt het aantal SOA-elementen gegeven door vergelijking 2.1. Vermits de architectuur aan in- en uit-

gangen voorzien is van elementen die golflengte-conversie realiseren, kan het aantal interne golflengten w onafhankelijk gekozen worden van het aantal golflengten W dat op een vezel wordt getransporteerd. Elementaire analyse leert ons dat het minimum aan SOA-elementen bekomen wordt voor de keuze $w=N^{1/2}$. Het resulterende aantal SOA-schakel-elementen wordt dan gegeven door formule 2.2.

$$S(N, w) = N \cdot (N/w + w) \quad 2.1$$

$$S_{\text{opt}}(N) = 2 \cdot N \cdot N^{1/2} \quad 2.2$$

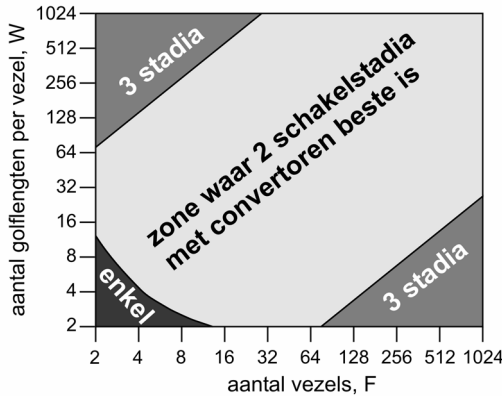
Het aantal SOA-schakel-elementen voor de Clos-structuur wordt gegeven door vergelijking 2.3. Opnieuw kunnen we dit aantal minimaliseren, aangezien door de aanwezigheid van golflengte-conversie de grootte van de groepjes n vrij gekozen kan worden. We vinden dat het minimum bereikt wordt voor de keuze $n=1/2 \cdot N^{1/2}$ en gegeven wordt door vergelijking 2.4.

$$C(N, n) = 2 \cdot N/n \cdot S_{\text{opt}}(N) + n \cdot S_{\text{opt}}(N/n) = 2 \cdot N \cdot (2 \cdot n^{1/2} + N^{1/2} \cdot n^{-1/2}) \quad 2.3$$

$$C_{\text{opt}}(N) = 4 \cdot 2^{1/2} \cdot N \cdot N^{1/4} \quad 2.4$$

In de laatste architectuur van een schakelmatrix voor F vezels met elk W golflengten, zullen we zoals aangegeven in Figuur 2.9b het laatste schakelstadium vervangen door golflengte-convertoren. Opnieuw kunnen we hiervoor het aantal SOA-schakel-elementen berekenen en vinden we vergelijking 2.5. Met de gemaakte keuze $n=F$ kan dit aantal niet verder geoptimaliseerd worden.

$$T(F, W) = F \cdot S_{\text{opt}}(W) + W \cdot S_{\text{opt}}(F) = 2 \cdot F \cdot W \cdot (W^{1/2} + F^{1/2}) \quad 2.5$$

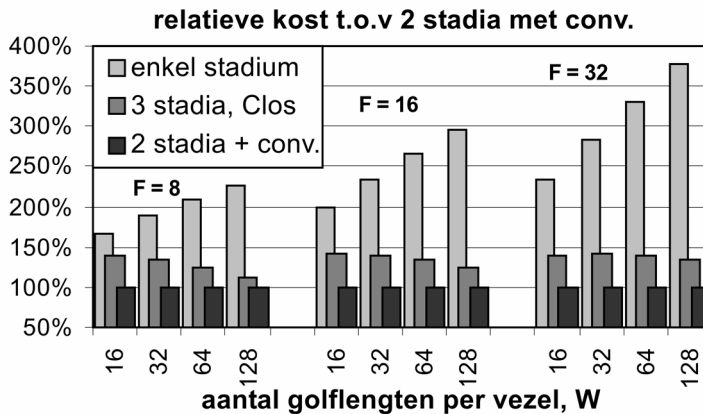


Figuur 2.10: Welke architectuur gebruikt het minst schakelende SOA's voor een gegeven grootte van de schakelmatrix?

Het antwoord op de vraag welke architectuur tot het laagste aantal SOA-schakel-elementen leidt voor gegeven dimensies van de te construeren schakelknoop kan met behulp van bovenstaande vergelijkingen gevonden worden. In Figuur 2.10 hebben we dit grafisch weergegeven voor een bereik van F en W gaande van 2 tot 1024 (merk op dat de dimensie van de knoop gegeven wordt door $F \cdot W \times F \cdot W$, dus $N=F \cdot W$ in vergelijkingen 2.1–2.4). De zone waartoe een punt met coördinaten (F, W)

behoort, is ingekleurd met een grijswaarde die aangeeft aan welke van de drie architecturen de meest voordelige is. Uit die figuur blijkt dat de B&S-structuur met een enkelvoudig schakelstadium slechts als beste uit deze vergelijking naar voor komt voor beperkte groottes van de schakelmatrix ($N=F \cdot W < 12$). Voor wat de meest praktisch bruikbare waarden van W en F lijken te zijn, vinden we dat de aanpak met twee schakelstadia en golflengte-convertoren de meest voordelige is in termen van aantal gebruikte SOA-schakel-elementen. De Clos-configuratie met drie volwaardige schakelstadia blijkt enkel nuttig voor matrices waar het aantal vezels beperkt is en het aantal golflengten per vezel zeer hoog, of vice versa.

Om een idee te geven van hoe sterk de drie architecturen van elkaar verschillen in termen van vereist aantal schakelende SOA-elementen, hebben we die aantallen relatief ten opzichte van elkaar uitgezet in Figuur 2.11. Om de vergelijking te vereenvoudigen hebben we voor een ruime selectie waarden voor F en W het aantal benodigde SOA-schakel-elementen uitgezet nodig voor elk van de drie beschouwde architecturen, gedeeld door het aantal vereist voor de meertraps-architectuur met golflengte-conversie als laatste stadium. Hieruit blijkt dat vooral voor een monolithische B&S-architectuur het aantal schakelende SOA's beduidend hoger ligt. Het voordeel van de architectuur met convertoren ten opzichte van een volwaardige Clos-structuur met drie schakelstadia is beduidend kleiner, maar nog steeds belangrijk (enkele tientallen procenten).



Figuur 2.11: Relatieve kost in termen van aantal schakelende SOA's vergeleken met de architectuur met twee schakelstadia en golflengte-conversie als finaal stadium.

2.4 Besluiten

In de context van knoop-architecturen voor metro-netwerken hebben we onze aandacht gericht op OPS ring-architecturen, vermits deze als meest geschikte kandidaat beschouwd worden en als dusdanig in verschillende projecten en onderzoeksgroepen voorgesteld worden als MAN-architecturen van de (nabije?)

toekomst [77,78,79,118,123]. De meeste van die architecturen gebruiken een Actieve Knoopstructuur, waar een optisch pakket ofwel kan doorgelaten worden, ofwel van de ring gehaald kan worden. We hebben die architectuur vergeleken met een Passieve Knoopstructuur, die op netwerk-niveau leidt tot een logische busstructuur [124]. In die vergelijking ligt precies de originaliteit van onze bijdrage tot het onderzoeksgebied omtrent MAN-architecturen.

We hebben deze gerealiseerd aan de hand van het oplossen van een netwerkplanningsprobleem: hoeveel middelen zijn nodig om een gegeven hoeveelheid verkeer tussen een gegeven verzameling eindpunten (knopen) te dragen? Uit onze resultaten blijkt dat de Actieve Knoopstructuur enkel voordelen biedt wanneer het aantal golflengten per vezel een doorslaggevende kostfactor is: op vlak van zend- en ontvangst-capaciteit in de ringknopen biedt het gebruik van een eenvoudigere Passieve Knoopstructuur geen meerkost met zich mee. En passant hebben we ook vastgesteld dat het gebruik van bidirectionele ringen, waarbij licht in twee tegengestelde richtingen doorheen de metro-ring propageert, enkel tot lagere kosten kan leiden wanneer ruimtelijk hergebruik mogelijk is en er een voldoende hoeveelheid verkeer binnen een zelfde metro-ring blijft. Het golflengteband-concept zoals voorgesteld in [116]—waarbij in elke knoop afzonderlijk per band slechts de bandbreedte geassocieerd met één golflengte aangesproken kan worden—voor een Actieve Knoopstructuur bleek weinig zin te hebben wanneer het aantal gebruikte golflengten doorslaggevend is voor de kost.

Het spreekt voor zich dat de in dit hoofdstuk gepresenteerde studie van de MAN-ring-architecturen slechts één enkel facet is van een volledige evaluatie van de Actieve en Passieve Knoopstructuren. Complementaire studies omvatten de vraag hoe efficiënt de capaciteit in een gegeven ring kan benut worden onder variërende verkeerspatronen, door gebruik te maken van een geschikt MAC-protocol [71]. Vergelijking tussen de ring-architectuur en een PON-gebaseerde metro-architectuur [143] vormt een aanvulling van onze hoger behandelde studie. Uiteraard kan ook de vergelijking gemaakt worden met andere (bestaande, commerciële technologieën) waarmee een OPS-architectuur zal moeten concurreren [116]. Een studie voorgesteld in [118] toont dat de voorgestelde optische ring-architecturen ook praktisch realiseerbaar zijn, zelfs wanneer fysische beperkingen (op bv. maximale lengte van de ringen en het aantal knopen dat in serie kan geplaatst worden) in rekening gebracht worden.

In de context van meer geavanceerde knooparchitecturen, zoals die typisch worden voorgesteld voor vermaasde ruggengraatnetwerken [6,91,92], hebben we aandacht besteed aan de schaalbaarheid ervan. Het vernieuwende van ons werk bestond erin aan te tonen hoe voor een specifieke SOA-gebaseerde “broadcast-and-select”-architectuur een Clos-geïnspireerde structuur met meerdere schakelstadia gebruikt kan worden om grote schakelmatrices te construeren. We hebben daarbij een OPS-specifieke variant bedacht op de Clos-architectuur waarbij het derde stadium gereduceerd werd van een volwaardig schakelstadium tot loutere golflengte-conversie. Er bleek overigens dat zelfs voor bescheiden groottes deze variant nuttig is met het oog op het beperken van het aantal schakelende SOA-componenten. Zo’n Clos-achtige aanpak bewijst overigens ook z’n voordelen voor andere schakeltechnologieën: in [9] tonen we bijvoorbeeld hoe we, dankzij het

invoeren van meerdere schakeltrappen, met een gelimiteerd bereik van de golflengte-convertoren toch AWG-gebaseerde schakelmatrices met een groot aantal in- en uitgangspoorten kunnen construeren.

Merk op dat deze studies dienen aangevuld te worden met analyses betreffende de fysische haalbaarheid ervan (bijvoorbeeld de maximale splitsingsfactor voor B&S-architecturen, het al dan niet vereist zijn van 2R- of 3R-regeneratie-stadia tussen verschillende schakelstadia, etc.).

Hoofdstuk 3

Schakel-algoritmes voor OPS

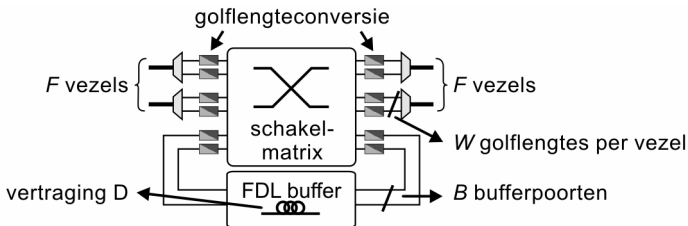
3.1 Situering

In het vorige hoofdstuk hebben we aandacht besteed aan enkele specifieke fysische knoopparchitecturen voor optisch pakketschakelen. Dit hoofdstuk zullen we wijden aan de algoritmes die in een dergelijke knoop gebruikt kunnen worden om de schakelbeslissingen te nemen.

Aan de rand van het OPS-netwerk situeert zich de interface met elektronische apparatuur, zoals bv. IP-routers. Hier zullen de optische pakketten, of bursts in het geval van OBS, worden gevormd. Afhankelijk van het gebruikte pakketformaat, en in het bijzonder de lengte ervan (vast of variabel, etc.), zullen binnenkomende gegevens moeten worden geaggregeerd of eventueel gesplitst over meerdere optische pakketten. Hier treedt dus het klassieke probleem op van segmentatie en weder-samenstelling (Eng. Segmentation And Reassembly, SAR). Een voorbeeld van zo'n interface tussen elektronische IP-routers en een optisch ring-gebaseerd metro-netwerk is te vinden in [144]. In het geval van optisch pakket- of burstschakelen gaat men er veelal van uit dat de optische pakketten of bursts (fragmenten van) meerdere IP-pakketten bevatten. Om de vullingsgraad bij pakketten van vaste lengte hoog te houden, gebruikt men doorgaans een timer: wanneer na verloop van tijd niet voldoende nieuwe gegevens zijn aangekomen om een optisch pakket volledig op te vullen, zal men de reeds aanwezige data toch versturen om de wachttijd te begrenzen. Analyse van dergelijke aggregatie-procedures zijn onder meer te vinden in [145,146,147]. Zulke tijdschakelaargebaseerde methoden hebben trouwens als bijkomend voordeel dat ze de zogenaamde zelfgelijkenis (Eng. self-similarity) reduceren [108,151,152,153], tenminste tot op bepaalde tijdschalen [154]. Zulke reductie van de zelfgelijkenis is erg welkom, gezien die het moeilijker maakt de prestatie van het schakelen hoog te houden (i.e. verlieskans en vertraging te beperken) [148,149,150].

Om de optische pakketten het OPS-netwerk in te sturen, wat typisch gebeurt in een metro-omgeving, is een toegangsprotocol vereist. Wanneer het een optisch netwerk betreft zonder buffermogelijkheden, is zo'n MAC-protocol niet triviaal. Er moet immers voor gezorgd worden dat elk toegangspunt een eerlijke kans krijgt om verkeer in het netwerk te brengen. Dit gebeurt dikwijls door krediet-gebaseerde systemen, waarbij zogenaamde tokens worden toegekend [80–82]. Een alternatief,

of mogelijke aanvulling van een krediet-gebaseerd MAC, is met reservaties te werken en knopen expliciete toegang te laten vragen aan een centrale controle-knoop [155]. Anderzijds kan men op basis van metingen van de hoeveelheid verkeer de vereiste bandbreedte inschatten [156]. Waar verschillende metro-netwerken of delen ervan samenkomen, zoals in een Hub die meerdere ringen verbindt, moet ervoor gezorgd worden dat de interconnectie niet gemonopoliseerd wordt door een beperkt aantal subnetwerken. Het verzorgen van die gegevensoverdracht tussen de verschillende ringen vertaalt zich in een scheduling probleem [157]. Ook daarin kan men een reservatie-systeem gebruiken, of continue metingen van de hoeveelheid verkeer die tussen deelnetwerken wordt uitgewisseld [71].



Figuur 3.1: OPS-schakelaar met teruggekoppeld FDL-buffer.

In een ruggengraatnetwerk zal een optische pakketschakelende knoop veelal wél buffercapaciteit bezitten, aangezien het contentieprobleem in willekeurig vermaasde netwerken typisch moeilijk anders op te lossen valt (tenzij de belasting laag is, uiteraard). In dit hoofdstuk zullen we schakel-algoritmes ontwerpen en evalueren voor een veel voorgestelde knoop-architectuur (zie bv. [6,84,98,158]) met een teruggekoppelde buffer bestaande uit vezelvertragslijnen (Eng. Fiber Delay Lines, FDL's). De structuur is geschetst in Figuur 3.1. De kern ervan wordt gevormd door een strikt niet-blokkerende matrix, bv. zoals die behandeld in Hoofdstuk 2. Hoewel de analyse in dit hoofdstuk niet afhankelijk is van de specifieke technologie waarin de matrix gerealiseerd wordt, is een cruciale onderstelde eigenschap van de schakelaar dat hij tot golflengte-conversie in staat is. We zullen die conversie immers gebruiken om contentie op te lossen. Een aantal uitgangspoorten van de matrix, aangeduid als B , zal worden verbonden met een bufferblok bestaande uit FDL's die opnieuw met de ingangspoorten geconnecteerd zijn. Merk op dat we de term poort steeds zullen gebruiken als verkorte vorm van golflengtepoort: een poort komt steeds overeen met één enkele golflengte en kan dus op elk tijdstip maximaal één optisch pakket dragen. De voor de verdere studies relevante parameters van de architectuur zijn op de figuur aangebracht.

Het vervolg van dit hoofdstuk is opgesplitst in twee delen: in Sectie 3.2 zullen we schakel-algoritmes voor vaste pakketlengtes behandelen, terwijl in Sectie 3.3 pakketten van variabele lengte aan bod komen.

Sectie 3.2 is gestoeld op onze publicaties [20, 24, 25], waarvan de eerste twee in Appendices D en E kunnen gelezen worden. De Appendix F [32] ligt aan de basis van Sectie 3.3.1, terwijl we ons voor 3.3.2 baseerden op [29] uit Appendix G.

3.2 Pakketten van vaste lengte

Vanzelfsprekend werd het probleem van het vinden van geschikte schakel-algoritmes voor het geval van een synchrone schakelaar voor pakketten van vaste lengte reeds uitvoerig bestudeerd in het kader van ATM [103]. Voorbeelden zijn [159,160], maar omdat ze betrekking hebben op gebruik van elektrisch RAM-geheugen zijn die resultaten niet van toepassing op het door ons bestudeerde geval van OPS met FDL-buffers. In het geval van een vezelvertragslijn ligt het tijdstip waarop het pakket de buffer zal verlaten immers vast op het ogenblik dat het in de buffer geplaatst wordt. Bij RAM daarentegen kan een gebufferd pakket op een willekeurig gekozen ogenblik uit de buffer gehaald worden.

In de specifieke context van synchroon pakketschakelen met gedeeld recirculerend buffer is [161] wellicht het meest relevante werk uit de ATM-gerelateerde literatuur. Daarin wordt ook een teruggekoppeld buffer met een vaste vertraging beschouwd en een ondergrens op de pakketverlies-verhouding (Eng. Packet Loss Rate, PLR) berekend. Die analyse is evenwel onnauwkeurig voor praktische groottes van een OPS-schakelaar (cf. om een analytische behandeling mogelijk te maken onderstelde [161] een oneindig aantal schakelpoorten).

Beroemd vroeg werk in het kader van OPS vinden we in [97], maar de architectuur die er beschouwd werd betrof een voorwaarts schakelend buffer, zonder golflengte-multiplexering (WDM). In het kader van het Europese project KEOPS [92], toonde onderzoek van een niet-gedeeld uitgangsbuffer, maar wel met WDM, het voordeel aan van golflengte-conversie om contentie op te lossen [94]. Verder werk bewees dat ook met beperkte golflengte-conversie [162], of een beperkte verzameling van convertoren [163], verlies ten gevolge van contentie sterk gereduceerd kan worden.

De schakel-algoritmes in bovenvermelde OPS studies herleiden zich eenvoudig gesteld tot het bepalen van de kortste, in een FDL beschikbare, vertraging voor een pakket dat door contentie niet zonder wachten kan worden doorgeschakeld. Hoewel ze duidelijk gerelateerd zijn aan ons werk dat zo meteen behandeld wordt, beschouwen die studies geen teruggekoppelde buffer en evenmin een buffer die gedeeld wordt door verkeer bestemd voor alle uitgaande poorten. Recent werk dat wel zo'n recirculerende buffer beschouwde situeert zich onder meer in de context van fotonisch slotschakelen (Eng. Photonic Slot Routing, PSR) [164]. Maar dit PSR-concept verschilt sterk van het OPS-concept dat wij beschouwen: in PSR worden alle golflengtes steeds tesamen geschakeld. Pakketten op verschillende golflengten vormen er immers samen een multi-slot met één welbepaalde bestemming. Studies voorgesteld in onder meer [165,166] kaderen wel in het OPS-domein en beschouwen een gedeeld teruggekoppeld buffer, maar dan zonder WDM. Het gecombineerd gebruik van golflengte-conversie en vertraging door FDL's komt er dus niet aan bod. Bovendien wordt ook geen aandacht besteed aan het probleem dat zich stelt wanneer niet alle recirculerende bufferpoorten dezelfde vertraging kennen.

3.2.1 Een geschikt schakel-algoritme

In het geval van OPS met pakketten van vaste lengte gebruikt men een synchrone, slot-gebaseerde schakelaar (herinner u Figuur 1.4). Pakketten die via de inkomende vezels de schakelaar bereiken worden in de tijd gealigneerd aan de slotgrenzen. In elk tijdslot moet de controle-logica van de schakelaar beslissen wat er met de aankomende pakketten gebeurt. In het algemeen geval van schakelen in een WDM omgeving omvat die beslissing drie aspecten: (i) het bepalen van de uitgaande vezel waarlangs het pakket de schakelaar zal verlaten, (ii) het bepalen van de golflengte, (iii) het bepalen van de eventuele vertraging die het pakket moet krijgen.

Het bepalen van de uitgaande vezel gebeurt op basis van de hoofding van het pakket en behelst wat men in het Engels “forwarding” noemt. Dit kan bijvoorbeeld MPLS-gebaseerd zijn [28]. We zullen onderstellen dat de bestemming van een pakket eenduidig uit die hoofding volgt en we zullen geen deflectie toepassen. Dit betekent dat we niet toelaten dat een pakket naar een andere uitgaande vezel geschakeld zal worden om contentie op te lossen. De motivatie hiervoor is dat deze techniek enkel werkt wanneer het netwerk licht belast is [98].

Op welke golflengte een pakket doorgestuurd zal worden, kunnen we vrij kiezen, vermits golflengte-conversie in de onderstelde architectuur mogelijk is en die golflengte geen invloed zal hebben op de behandeling van het pakket in een volgende schakelaar: schakelbeslissingen worden enkel genomen op basis van de inhoud van de pakket-hoofding (in tegenstelling tot bv. bij golflengte-routing [167]). Welke golflengte precies gekozen wordt zal evenmin impact hebben op latere pakketten, want in het volgende tijdslot zal het pakket de schakelaar volledig verlaten hebben.

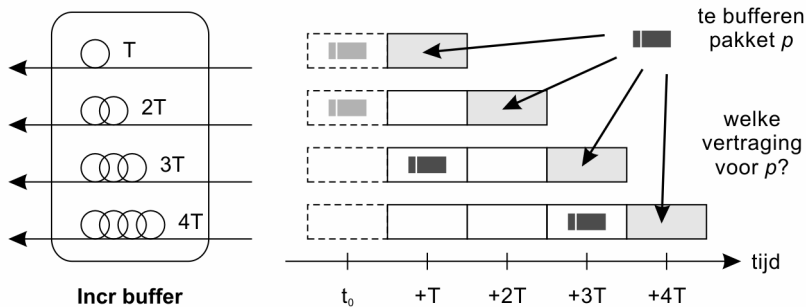
Het bepalen van welke vertraging aan een pakket gegeven wordt komt neer op de keuze van een recirculerende bufferpoort, en/of het aantal circulaties. Een beslissing over het aantal keren dat een pakket doorheen een FDL wordt gestuurd, zullen we niet a priori nemen bij aankomst van het pakket in de schakelaar. We zullen voor een aankomend pakket enkel kiezen of we het direct naar een uitgaande vezel, ofwel naar de FDL-buffer sturen. Wat er nadien mee gebeurt wordt beslist wanneer het pakket opnieuw aan de ingang van de schakelmatrix verschijnt. In [168] wordt dit gecatalogeerd als een PostRes strategie. Er blijft dus enkel nog de keuze over welke van de recirculerende bufferpoorten we gebruiken. Die keuze is vanzelfsprekend enkel van belang wanneer de bufferpoorten niet allemaal leiden naar een FDL van dezelfde lengte. We zullen daarom twee structuren beschouwen:

- **Fix:** In dit geval leidt elke teruggekoppelde bufferpoort tot een zelfde vertraging van één enkel tijdslot: voor elke poort $i = 1 \dots B$ is de bijhorende vertraging $D_i = T$ (waarbij T de duur van het tijdslot is).
- **Incr:** Hier gebruiken we oplopende vertragingen voor elk van de B recirculerende bufferpoorten, i.e. $D_i = i \cdot T$.

Het schakel-algoritme dat we toepassen omvat twee selectie-fases: (i) selecteer voor elke van de F uitgaande vezels afzonderlijk maximaal W pakketten die rechtstreeks zullen worden doorgeschakeld, (ii) selecteer uit de overblijvende pakketten maximaal B pakketten die naar de recirculerende buffer gezonden worden. Pakketten die niet geselecteerd werden in één van deze fases zullen verloren gaan,

omdat golflengte-conversie en de aanwezige buffer niet volstaan om de contentie en overbelasting op te lossen. De selectie in elk van de fasen is willekeurig, maar voor de eerste fase geven we voorrang aan pakketten die al eens doorheen een FDL vertraagd werden tegenover pakketten die net de schakelaar bereiken via een ingangsvezel.

Wanneer de buffer FDL's van verschillende lengte bevat, zal in de tweede fase beslist moeten worden welke vertragslijn voor een gegeven pakket gekozen wordt. Het gestelde vraagstuk wordt in Figuur 3.2 geïllustreerd voor een Incr-bufferstructuur. Gezien de schakelaar geen kennis heeft over welke pakketten in de toekomst via de ingangsvezels zullen aankomen, kunnen we niet a priori bepalen welke keuze van vertragslijn tot de minste contentie zal leiden.



Figuur 3.2: Illustratie van keuze van vertraging wanneer niet alle FDL's dezelfde vertraging realiseren. De grijs gekleurde posities zijn mogelijke keuzes om het pakket p te bufferen. De pakketten aanwezig in de buffer voor t_0+T en t_0+3T zijn bestemd voor dezelfde uitgaande vezel als p .

De enige beschikbare kennis over de toekomst omvat eerder genomen buffer-beslissingen. Om te beoordelen of gebruik van die kennis meetbare voordelen (i.h.b. lagere PLR) oplevert hebben we volgende vier strategieën beschouwd:

- **MinV (minimale vertraging):** Voor elk pakket dat we toelaten in de buffer kiezen we de nog vrije bufferpoort met de kleinste vertraging. Deze strategie zal dus de FDL poorten sequentieel opvullen indien deze volgens toenemende vezellengte geordend zijn. In het voorbeeld van Figuur 3.2 zal de vertragslijn met lengte T gekozen worden.
- **GeenO (geen overbelasting):** Om een pakket p te bufferen kiezen we de nog vrije bufferpoort verbonden met de vertragslijn van lengte L , waarvoor geldt dat het aantal pakketten dat de buffer zal verlaten op tijdstip t_0+L , bestemd voor dezelfde uitgaande vezel als p , niet meer bedraagt dan $W-1$ (waarbij t_0 het tijdstip is waarop de keuze voor p gemaakt wordt). Als er geen vrije bufferpoort gevonden kan worden die aan deze voorwaarde voldoet, gaat het pakket p verloren. Zodoende zal de *GeenO* strategie ervoor zorgen dat overbelasting van een uitgaande vezel nooit veroorzaakt wordt door gebufferde pakketten alleen. Wanneer er slechts één golflengte per uitgaande vezel zou zijn, zou voor Figuur 3.2 de keuze vallen op vertraging $2T$.

- **VermijdO (vermijd overbelasting):** Zoek voor een pakket p een bufferpoort zoals in strategie *GeenO*. Wanneer dit niet lukt, kies dan de nog vrije bufferpoort met kleinste vertraging (zoals in *MinV*). Ook hier trachten we dus te vermijden dat overbelasting van een uitgaande vezel veroorzaakt wordt door alleen maar gebufferde pakketten, maar kiezen ervoor een pakket toch nog toe te laten in de buffer indien er een vrije bufferpoort is.
- **Balans:** We zullen pakketten bestemd voor dezelfde uitgaande vezel zo goed mogelijk spreiden in de tijd. Om een bufferpoort te kiezen voor een pakket p , tellen we voor elke beschikbare vertraging L het aantal pakketten dat de buffer op t_0+L zal verlaten voor dezelfde uitgaande vezel als waarvoor p bestemd is en noteren dit aantal N_L (met opnieuw de notatie t_0 voor het tijdstip waarop we de buffer-keuze maken). We sturen p dan naar de vrije bufferpoort die leidt naar de kortste vertragingstijd met laagste N_L . Voor Figuur 3.2 zou dit de poort met vertraging $2T$ zijn.

Wat betreft de volgorde waarin we de keuze maken voor pakketten toegelaten tot de buffer wordt opnieuw voorrang verleend aan pakketten die al eerder een doorgang door een FDL gemaakt hebben, maar verder is die volgorde willekeurig. Merk op dat de *GeenO* en *VermijdO* enkel verschillen van *MinV* wanneer het aantal recirculerende bufferpoorten het aantal golflengten per vezel overstijgt ($B > W$).

3.2.2 Methodologie

Om de prestatie van de verschillende algoritmes voor verschillende waarden van de buffer-parameters te onderzoeken hebben we geopteerd voor simulatie. Analytische of numerieke methoden voor een teruggekoppeld buffer zijn immers beperkt tot relatief kleine schakelaars (bv. [160,161,165]), of geven enkel ondergrenzen [159,160]. Desalniettemin werd, zoals in [8] getoond, ons vertrouwen in de simulator wel gesterkt door vergelijking van simulatieresultaten met waarden voor de pakketverlies-verhouding bekomen met numerieke methoden, en dit voor verkeer beschreven door een zogenaamd MMAP-model [169].

Tabel 3.1: Gebruikte parameters van de pakquetschakelaar voor de studies omtrent schakelen van pakketten van vaste lengte.

<i>Symbol</i>	<i>Waarde</i>	<i>Betekenis</i>
F	6	Aantal in- en uitgangsvezels
W	32	Aantal golflengten per vezel
B	0...64	Aantal golflengtepoorten van de schakelmatrix dat gebruikt wordt voor circulatie doorheen het FDL-buffer
D	1...64	Lengte van een FDL, uitgedrukt als de vertraging in aantal slots

Om de betrouwbaarheid van de resultaten te bewerkstelligen [170], hebben we gebruik gemaakt van de beroemde Mersenne Twister generator [171] voor de waarden van toevalsveranderlijken en een methode gebaseerd op “batch means” [172] om relatieve fouten kleiner dan 10% met 98% betrouwbaarheid te bekomen. De gemeenschappelijke parameters die we gebruikten staan in Tabel 3.1 samengevat en zijn gebaseerd op de 256×256 schakelmatrix zoals die voorzien was voor de demonstratie beschreven in [173]. Voor een meer gedetailleerde beschrijving van de simulator verwijzen we naar Appendix D.

Tabel 3.2: Gebruikte verkeersmodellen.

<i>Model</i>	<i>Distributie en parameters</i>	<i>Toelichting</i>
<i>Poisson</i>	$p(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \quad x \in \mathbb{N}$	x is het aantal pakketten gegenereerd in een slot λ is de verkeers-intensiteit (in pakketten/slot)
<i>GeoAanUit</i>	$p(d) = \frac{1}{m+1} \cdot \left(1 - \frac{1}{m+1}\right)^{d-1}, \quad d \in \mathbb{N} \setminus \{0\}$ $m_{aan} = 4$ $m_{uit} = m_{aan} \cdot \left(\frac{1}{\lambda} - 1\right)$	d is de duur van de te genereren aan- of uit-periode in slots m is de gemiddelde waarde van die duur
<i>ParetoAanUit</i>	$p(d) = \frac{\alpha \cdot b^\alpha}{d^{\alpha+1}}, \quad d \in [b, +\infty)$ $\alpha_{aan} = 1.35$ $\alpha_{uit} = 1.5$ $b_{aan} = 1$ $b_{uit} = b_{aan} \cdot \frac{1 - 1/\alpha_{uit}}{1 - 1/\alpha_{aan}} \cdot \left(\frac{1}{\lambda} - 1\right)$	d is de duur van de te genereren aan- of uit-periode in slots b is de minimale duur van die periode α is de zogenaamde vormparameter: voor grotere waarden is de staart van de distributie nauwer, i.e. hogere d waarden hebben lagere probabiliteiten

Om de prestatie voor voldoende uiteenlopende verkeersprofielen te vatten hebben we resultaten verzameld voor drie beschouwde modellen. Het eerste is het welbekende Poisson-proces, waarbij pakketten aankomen met exponentieel verdeelde tijden tussen opeenvolgende aankomsten. De twee andere modellen voor de aankomsten van pakketten gebruiken een zogenaamd aan/uit-model: aan-perioden, waarin in elk tijdslot een pakket aankomt, worden afgewisseld met uit-perioden, tijdens dewelke geen pakketten arriveren. Voor het GeoAanUit-model hebben we de geometrische distributie gebruikt voor de duur van zowel de aan- als uit-perioden en voor het ParetoAanUit-model de Pareto-verdeling. Dit laatste model werd gekozen omdat het welbekend is dat een aggregaat van zulke ParetoAanUit-

bronnen zelfgelijkend (Eng. self-similar) verkeer genereert [148]. De analytische gedaante van de gebruikte verdelingen en hun parameters, zijn samengevat in Tabel 3.2. Voor de ParetoAanUit-bronnen hebben we een minimale duur van de aanperiode van één enkel slot gekozen, terwijl de waarden van de vormparameter α gebaseerd zijn op metingen van reëel verkeer [149]. De gemiddelde duur van de aanperiode voor de GeoAanUit-modellen werd zo gekozen dat deze overeenkwam met die van het gebruikte ParetoAanUit-verkeer, zodat het verschil tussen deze modellen enkel ligt in de variantie van de duur van de aan- en uit-perioden.

De bestemmingen van de pakketten waren uniform verdeeld over alle uitgaande vezels. We hebben dit gerealiseerd door voor elke combinatie van een in- en uitgaande poort een afzonderlijke verkeersstroom te genereren voor een gegeven verkeersklasse (i.e. prioriteit, zie verder sectie 3.2.5). Voor het GeoAanUit-model hebben we zelfs 32 individuele bronnen voor elke (in,uit,prioriteit)-combinatie gebruikt, omdat het model enkel $\lambda < 1$ kan genereren, wat met $W=32$ golflengten per vezel niet altijd volstaat om de gewenste belasting te bereiken. Onder de term belasting verstaan we in het vervolg de input-belasting, i.e. het gemiddeld aantal pakketten per tijdseenheid dat aan de schakelaar wordt aangeboden via zijn ingangen gedeeld door het maximale aantal dat er zou kunnen aankomen (i.e. $F \cdot W$ per tijdslot).

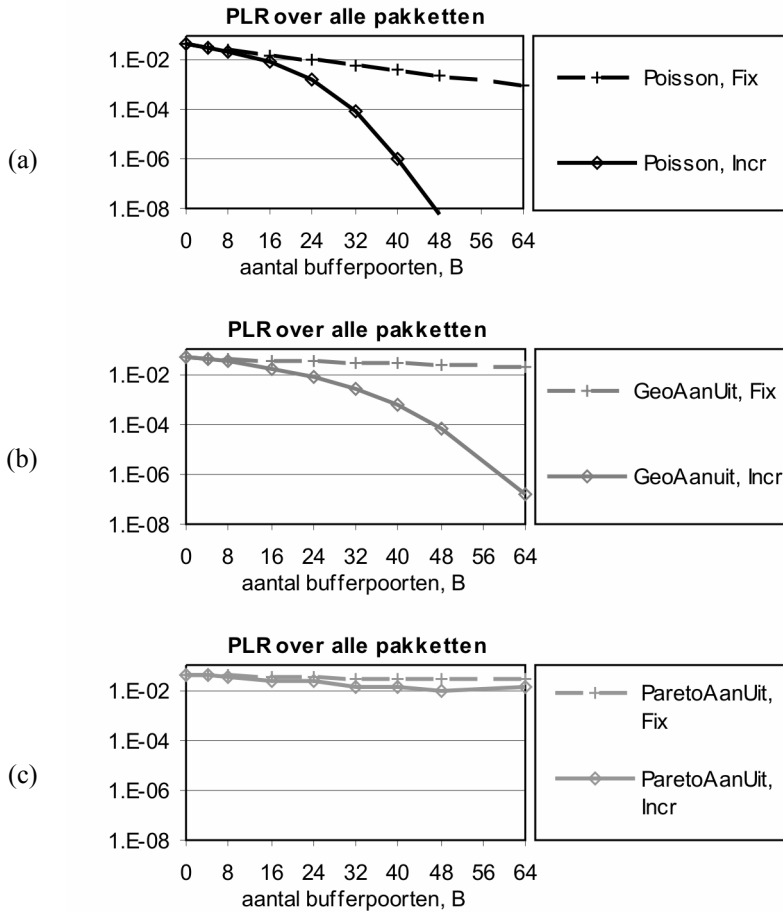
3.2.3 Keuze van de buffer-architectuur

Zoals hoger aangegeven kan voor de teruggekoppelde buffer geopteerd worden elk van de recirculerende poorten te verbinden met vertraginglijnen van verschillende lengte. Een eerste experiment dat we met onze simulator uitvoerden bestond dan ook uit het onderzoek naar het verschil in performantie tussen een schakelaar met een Fix-buffer, waarbij elk van de B bufferpoorten een vertraging van één enkel slot realiseert, en een Incr-buffer waarbij de bufferpoorten een toenemende vertraging veroorzaken (zie hoger).

In Figuur 3.3 tonen we de pakketverlies-verhouding (Eng. Packet Loss Rate, PLR) voor een toenemend aantal recirculerende bufferpoorten B . Zoals we intuïtief verwachten, zien we dat de Incr structuur beduidend lagere verliezen veroorzaakt voor toenemende B -waarden. Voor een hoger aantal bufferpoorten neemt immers het verschil in opslagcapaciteit tussen de Incr- en Fix-structuren toe: een Incr-buffer kan weliswaar maar evenveel pakketten opnemen in een bepaald tijdslot (namelijk B), maar het totale aantal pakketten dat potentieel aanwezig kan zijn in de buffer ligt hoger dan voor de Fix-buffer ($B \cdot (B+1)/2$ tegenover B voor Fix).

Wat de verschillen tussen de verkeerspatronen betreft, wordt hier het resultaat uit de klassieke wachttijtheorie bevestigd: het toevoegen van bufferruimte levert meer op voor verkeer zonder correlatie in de tijd, zoals voor het Poisson-model. Maar ook wanneer er correlatie op relatief korte tijdschaal is, blijkt een FDL-gebaseerde buffer vrij efficiënt te zijn voor voldoende bufferdiepte (cf. het GeoAanUit-model met een Incr-buffer). Wanneer er correlatie op lange tijdschalen optreedt, zoals verkeer met zelfgelijkenis waar ParetoAanUit een voorbeeld van is, helpt het duidelijk niet een buffer te hebben dat slechts op relatief korte tijdschaal werkt. Dit resultaat was al bekend voor wachttijnsystemen in de context van elektrische buffers (zie bv. [150] en referenties daarin).

Het verrast ons niet dat de prijs die betaald wordt voor de lagere verlieskans met een Incr-buffer erin bestaat dat de gemiddelde vertraging die pakketten ondergaan verhoogt. Desalniettemin zijn deze vertragingen behoorlijk klein⁵ (enkele tientallen slots) en kunnen ze verwaarloosd worden tegenover vertraging die in regionale of toegangsnetwerken voorkomt. Een wellicht belangrijker nadeel van de Incr-structuur is dat de volgorde van de pakketten niet langer behouden blijft. Wanneer we voorrang verlenen aan pakketten die uit de buffer komen tegenover



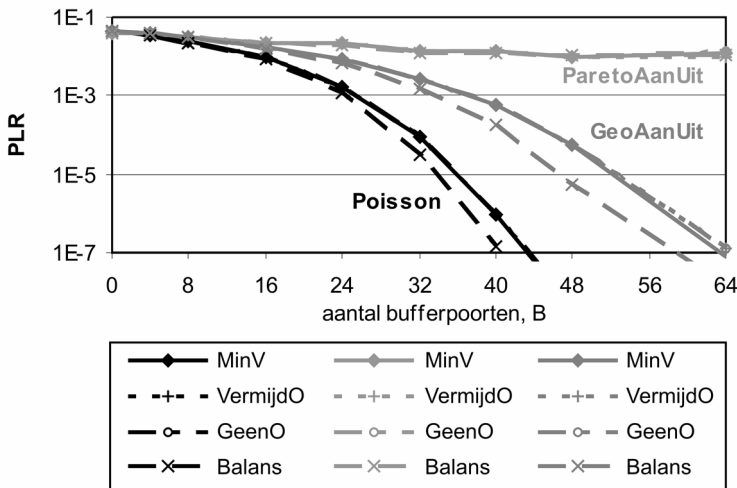
Figuur 3.3: Verlieskans (PLR) voor toenemend aantal bufferpoorten B: vergelijking Incr en Fix-bufferstructuur. De belasting van de knoop was 0.95.

⁵ In [6] bijvoorbeeld wordt een duur van een tijdsloot voorgesteld van de orde van een microseconde, zodat een vertraging in de orde van duizend slots nog steeds slechts in het milliseconden-bereik ligt; vgl. met eind-tot-eind vertragingen in het Internet van vandaag in de buurt van een paar honderd milliseconden [174].

pakketten van een inkomende vezel, zal een schakelaar met Fix-buffer de volgorde van pakketten bewaren. In het geval van een Incr-buffer is dit niet langer waar: een pakket dat later aankomt kan een voorganger voorbijsteken en eerder aan de uitgang van de schakelaar (of zelfs de buffer) verschijnen. Voor een Incr-buffer met $B=64$ poorten blijkt bijvoorbeeld dat bij een belasting van 0.95 net iets minder dan 10% van alle pakketten uit volgorde⁶ aankomt.

3.2.4 Bufferstrategieën

Om de hoger geïntroduceerde bufferstrategieën met elkaar te vergelijken hebben we het experiment uit de vorige sectie (waar we de MinV strategie hanteerden) herhaald voor de overige strategieën. De resultaten hiervan zijn te zien in Figuur 3.4. De schakel-algoritmes waarbij overbelasting vermeden wordt (GeenO, VermijdO) leiden tot quasi dezelfde verlieskansen als MinV, ook daar waar ze verschillende beslissingen kunnen nemen (voor $B > W$). De Balans-strategie bewijst wel z'n nut, tenminste toch voor die verkeerstypes waar buffercapaciteit het contentieprobleem kan verzachten (Poisson, GeoAanUit).



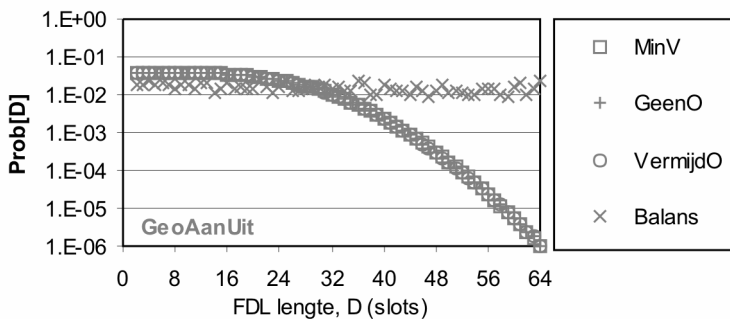
Figuur 3.4: Verlieskans (PLR) voor toenemend aantal bufferpoorten B: vergelijking van de bufferstrategieën. De belasting van de knoop was 0.95. Merk op dat de curves voor MinV, VermijdO, GeenO telkens zo goed als samenvallen.

De verklaring voor Balans' betere performantie in termen van PLR ligt precies in het feit dat pakketten over de tijd gespreid worden: daar waar de overige strategieën systematisch voor de kortst mogelijke vertragingstijd kiezen wanneer een

⁶ Een pakket p wordt beschouwd als "uit volgorde" wanneer een ander pakket p' , dat tussen dezelfde in- en uitgaande vezel geschakeld wordt, de schakelaar later binnenkomt maar toch vóór p verlaat.

pakket gebufferd wordt, zal Balans ze eerder verdelen over de beschikbare vertragingen. Het gevolg is dat minder pakketten gelijktijdig de buffer verlaten voor een zelfde uitgaande vezel en daardoor contentie met nieuw aankomende pakketten beperkt blijft. Dat pakketten effectief beter gespreid worden blijkt wanneer we op Figuur 3.5 kijken naar het aantal keer dat een bepaalde poort van de Incr-buffer gebruikt werd om een pakket te recirculeren. Aangezien het spreiden in de tijd impliceert dat lange vertraginglijnen vaker gebruikt worden, vinden we dat voor de Balans-strategie de gemiddelde vertraging typisch hoger ligt dan voor de overige buffer-algoritmes.

Tot slot willen we nog opmerken dat voor een Fix-buffer de hoger vermelde strategieën gereduceerd worden tot twee gevallen: MinV, VermijdO en Balans maken dan dezelfde bufferbeslissingen en enkel GeenO onderscheidt zich ervan wanneer $B > W$. Het gevolg van GeenO is dat een pakket nooit meer dan één keer doorheen de buffer circuleert. De spreekwoordelijke prijs die ervoor betaald wordt is een uiterst beperkte toename van de PLR. Voor een Fix-buffer volstaat het dus de meest eenvoudige strategie te volgen: zoveel pakketten tot de buffer toelaten als er bufferpoorten zijn.



Figuur 3.5: Verdeling van het aantal keer dat een bepaalde bufferpoort met een vertraginglijn van lengte D gebruikt wordt. Illustratie voor GeoAanUit in een knoop met Incr-buffer van $B=64$ poorten. De waarden voor MinV, GeenO en VermijdO vallen zo goed als samen.

3.2.5 Kwaliteitsdifferentiatie

Zoals al aangegeven in het inleidende Hoofdstuk 1 zal een OPS-netwerk het transport verzorgen van een heterogene verzameling verkeersstromen. Omdat elk van die types stromen verschillende eisen stelt aan het transportnetwerk in termen van bijvoorbeeld verlieskans en vertraging, moet het OPS-netwerk gediversifiëerde diensten kunnen aanbieden. Vermits het aanzienlijke aantal individuele stromen dat een typisch ruggengraatnetwerk doorkruist, denkt men doorgaans aan kwaliteitsdifferentiatie op basis van verkeerstype, eerder dan aan het garanderen van

bepaalde verlieskansen e.d. voor individuele stromen⁷: veeleer een DiffServ-aanpak [175] dan een IntServ-gebaseerde [176] oplossing.

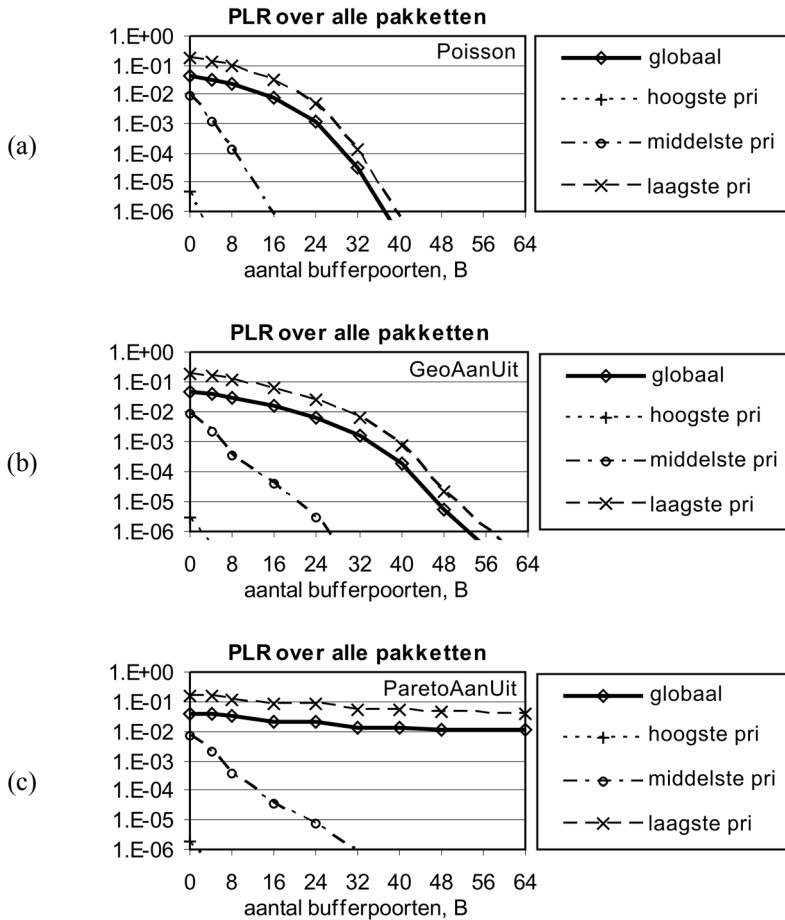
In die DiffServ-filosofie hebben wij onderzocht of de eerder voorgestelde schakel-algoritmes makkelijk uit te breiden zijn naar het aanbieden van gediversifiëerde diensten. De meest voor de hand liggende oplossing was het voorzien van een strikt prioriteitsschema. Elk pakket zal behoren tot een bepaalde prioriteitsklasse en we geven strikte voorrang aan pakketten van hogere prioriteit. Dit doen we door bij de selectie voor rechtstreekse doorschakeling of buffering eerst de pakketten van hogere prioriteiten te beschouwen en de lagere prioriteitsklassen pas toegang te geven tot een uitgaande vezel of een bufferpoort wanneer er nog plaats overblijft. De bufferstrategieën kunnen worden uitgebreid met een strikt prioriteitsschema door voor de buffering van een bepaald pakket p enkel rekening te houden met eerder gebufferde pakketten van dezelfde of hogere prioriteit.

De vraag of zo'n prioriteitsschema ook sterke diversifiëring in termen van verlieskans (PLR) oplevert wordt beantwoord door de resultaten in Figuur 3.6. Net zoals voorgaande resultaten wordt de verlieskans uitgezet voor toenemend aantal bufferpoorten, maar nu voor elk van de prioriteitsklassen. We beschouwden drie klassen: 50% verkeer van hoogste prioriteit, 25% verkeer van gemiddelde prioriteit en 25% van de laagste prioriteit. De PLR-curves tonen aan dat zelfs onder deze behoorlijk extreme voorwaarden (cf. groot aandeel hoogste prioriteit en totale belasting van 0.95) de diversifiëring uitgesproken is: zodra er enige bufferruimte is ($B > 0$) duikt de PLR voor de hoogste prioriteitsklasse onder 10^{-6} . Met uitzondering van het notoire ParetoAanUit-model kunnen we gelijkwaardige verlieskansen bereiken voor de laagste prioriteit door voldoende buffer te voorzien.

Uiteraard weerspiegelt het prioriteiten-mechanisme zich ook in de vertraging. Analyse van de simulatieresultaten leert dat enkel in de laagste prioriteitsklasse een aanzienlijk deel (30–40%) van de pakketten vertraagd wordt. Dankzij de aanwezigheid van golfenlengte-convertoeren hoeft slechts een kleine fractie (0.5–1%) van de pakketten van gemiddelde prioriteit en een verwaarloosbaar deel (minder dan één op 10^6) van de hoogste prioriteit gebufferd worden. Vermits herordening enkel veroorzaakt kan worden door bufferen, duikt dit fenomeen—op de laagste prioriteitsklasse na—nauwelijks op.

Men zou kunnen vermoeden dat de succesvolle differentiatie ten koste gaat van de globale performantie en in het bijzonder leidt tot een verhoging van de globale verlieskans. Vergelijking van de PLR met die wanneer prioriteitsinformatie genegeerd wordt en alle klassen dezelfde behandeling ondergaan, leert evenwel dat dit niet het geval is. We hebben tussen simulatieresultaten van beide gevallen, waarbij het aankomstenpatroon van de pakketten identiek was, geen significante verhoging van de totale gemiddelde verlieskans vastgesteld.

⁷ Voor een aanzet tot het bepalen van omstandigheden waarin garanties voor een aggregaat ook gelden voor individuele stromen verwijzen we naar [177].

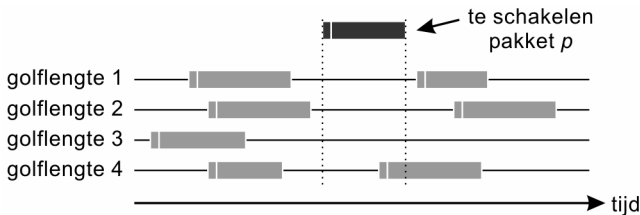


Figuur 3.6: Verlieskans (PLR) voor toenemend aantal bufferpoorten B: differentiatie door een strikt prioriteitsschema. De belasting van de knoop was 0.95, waarvan 50% hoogste, 25% gemiddelde en 25% laagste prioriteit.

3.3 Pakketten van variabele lengte

Het probleem van het schakelen van variabele lengte pakketten, ook in een optisch pakketgeschakelde context, is al grondig aan analyse onderworpen. Een specifieke moeilijkheid die ontspruit aan de niet langer vaste grootte van de schakeleenheden is het invullen van de zogenaamde leegtes die ontstaan tussen opeenvolgende pakketten (Eng. void filling, VF). Vermits er asynchroon geschakeld wordt en de pakketlengte niet constant is, is het beduidend moeilijker de bandbreedte volledig op te vullen (vergelijk met het fragmentatieprobleem bij geheugen-allocatie met

variabele groottes van geheugenblokken [178]). De leegtes worden gecreëerd doordat de tijd dat een pakket gebufferd wordt slechts gekozen kan worden uit een discreet aantal vertragingen die bepaald worden door de vezellengtes van FDL's. Daardoor is het niet altijd mogelijk een gebufferd pakket direct te laten aansluiten op een voorafgaand pakket op dezelfde uitgaande vezel en golflengte [179]. Het werk [180] beschrijft het “void filling”-algoritme voor asynchroon pakkettschakelen en toont aan dat dit aanzienlijke voordelen oplevert tegenover algoritmes die de onstane leegtes niet invullen. Het algoritme is gelijkaardig aan eerder gerapporteerde resultaten zoals “horizon scheduling” [181], of LAUC-VF [109]: er wordt gezocht naar dat kanaal waar een pakket p zonder overlap met een ander kan worden geschakeld en de tijd tussen p zelf en diens voorganger zo klein mogelijk is. Zo zal in het voorbeeld geschetst in Figuur 3.7 gekozen worden voor golflengte 2. Andere strategieën om leemtes te vullen bleken minder efficiënt in het opkrikken van de performantie [182,183]. Het voornaamste nadeel van leemte-vullende algoritmes is dat ze het bijhouden van nogal wat toestandsinformatie vereisen (lijsten van leegtes per uitgaande golflengte). Voorbeelden van minder veeleisende, maar desalniettemin behoorlijk performante algoritmes zijn te vinden in [10], specifiek voor de architectuur die we in dit hoofdstuk beschouwen (Figuur 3.1). Wanneer de golflengte niet vrij gekozen mag worden, maar vast is voor een gegeven verkeersstroom (zoals in [179]) treedt het probleem van leemte-vullen alleen op als de aankomsten van een pakket niet altijd even lang op voorhand gekend zijn (zoals bij OBS met gedifferentieerde offsets [184]).



Figuur 3.7: Het beslissingsprobleem voor pakketten met variabele lengte: op welke golflengte zullen we pakket p doorschakelen?

Algoritmes voor zo een knoop met een gedeelde buffer van teruggekoppelde FDL's worden in [168] geanalyseerd, in het bijzonder voor een OBS-concept. Er wordt een classificatie gemaakt van de strategieën in PostRes- en PreRes-algoritmes. In de laatstgenoemde zal, wanneer een pakket tot de buffer wordt toegelaten, meteen ook een uitgaande vezel en golflengte worden gereserveerd waarnaar het pakket geschakeld zal worden eenmaal het uit de FDL komt. Nadeel van zo'n PreRes-strategie is dat deze in bijvoorbeeld een OBS context kan interfereren met kwaliteitsdifferentiatie gebaseerd op verschillende offsets van de pakkethoofding.

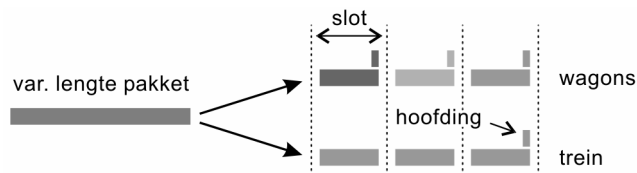
De voornoemde onderzoekswerken handelen allemaal over het schakelen van pakketten zonder deze in kleinere eenheden op te splitsen. In de slot-gebaseerde schakelaar zal men evenwel pakketten in slot-eenheden moeten verdelen, waardoor we de vraag kunnen stellen of het in sommige gevallen niet beter kan zijn deze slots

onafhankelijk te behandelen, eerder dan het pakket dat ze samen vormen als één schakeleenheid te beschouwen. Die vraag wordt in volgende sectie beantwoord.

3.3.1 Treinen versus wagons

Zoals aangeduid in Figuur 3.8 zijn er twee fundamenteel verschillende mogelijkheden om in een slot-gebaseerde, synchrone pakketschakelaar met pakketten van variabele lengte om te gaan. De eerste is de “trein van slots” [106] als één geheel te behandelen en een beslissing te nemen voor de volledige trein wanneer het eerste slot aankomt. In dit geval is het duidelijk dat enkel dit eerste slot een volwaardige hoofding moet bevatten met alle informatie nodig om te beslissen langs welke uitgaande poort de trein z’n weg doorheen het netwerk zal vervolgen. Een alternatief is elk van de onderscheiden wagons onafhankelijk van elkaar te schakelen. In dit geval dient elk van de wagonnetjes een eigen hoofding mee te krijgen en kan de pakketschakelaar opereren zonder zich van hun samenhang bewust te zijn. De algoritmes uit vorige Sectie 3.2 kunnen dan zonder wijziging worden toegepast. Vanzelfsprekend dienen aan de uitgangen van het OPS-netwerk de pakketten wel opnieuw worden samengesteld (SAR).

Een vergelijking van de trein- versus wagonbehandeling werd al eerder uitgevoerd in [105] voor het specifieke geval van een gedeelde bus-structuur voor lokale of regionale netwerken van hoge snelheid. De auteurs analyseerden er de reductie van de belasting op de controle-eenheid in het geval van een treinconcept (cf. minder te behandelen hoofdingen) en bestudeerden het verloop van vertraging in functie van de hoeveelheid doorgeschakeld verkeer (Eng. throughput). Ze concludeerden dat vanuit het standpunt van benutting van de aanwezige bandbreedte de wagon-aanpak voordelig kan zijn wanneer de gemiddelde treinlengte beperkt is.



Figuur 3.8: Schakelen van pakketten van variabele lengte in slot-gebaseerd OPS: treinen of wagonnetjes?

In het kader van OPS werd al aandacht besteed aan de vergelijking tussen het slot-gebaseerd schakelen van pakketten van vaste lengte en een asynchroon schakelconcept met pakketten van variabele lengte [185,186]. Die studies vergeleken evenwel de prestatie van een OPS-netwerk voor twee situaties met elk zowel een ander verkeersmodel als verschillende parameters voor de dimensies van de schakelknopen. Voor het slot-gebaseerde model werd gewerkt met aankomsten van pakketten van vaste lengte en men vergeleek de pakketverliezen met die voor een asynchroon verkeersmodel met pakketten van variabele lengte. In het laatste geval werden ook schakelknopen met meer bufferruimte gebruikt. Ondanks die uiteenlopende onderstellingen besluiten de auteurs van [185] dat slot-gebaseerd schakelen tot lagere pakketverliezen leidt. Ze tonen ook dat door extra golflengtes per vezel te

gebruiken voor een asynchrone schakelaar lagere pakketverlies-verhoudingen bereikt kunnen worden dan een slot-gebaseerde (met dus minder golfengten per vezel), of toch tenminste voor lichte belasting.

Hier richten we onze aandacht op een slot-gebaseerde OPS-schakelaar en we trachten uit te vinden of een wagon- dan wel trein-gebaseerd schakel-algoritme het best presteert in termen van logische performantie (i.e. pakketverlies, vertraging, kwaliteitsdifferentiatie). Hierbij zullen we—in tegenstelling tot [185,186]—voor beide gevallen hetzelfde verkeersmodel en pakketschakelaars met exact dezelfde dimensies beschouwen.

3.3.1.1 Methodologie

Om de trein- en wagonconcepten tegenover elkaar af te wegen, beschouwen we één enkele pakketschakelaar en zoals in de overige studies in dit hoofdstuk bepalen we de performantie door middel van simulatie. De beschouwde dimensies van de pakketschakelaar staan in Tabel 3.3 samengevat.

Het verkeersmodel dat we gebruikten is vanzelfsprekend een treinmodel, gelijkaardig aan [187]. Voor zowel de lengtes van de treinen als de tijd tussen twee opeenvolgende treinen gebruikten we een gediscretiseerde negatieve exponentiële verdeling, waarbij tijdsloten behorend tot eenzelfde trein steeds in opeenvolgende slots komen. Hoewel de exponentiële distributie niet noodzakelijk de meest realistische is wanneer treinen bv. als IP-pakketten geïnterpreteerd worden, geloven we toch dat de kwalitatieve conclusies die verder getrokken worden vrij algemeen geldend zijn.

Tabel 3.3: Gebruikte parameters in de trein vs wagons studie.

<i>Symbol</i>	<i>Waarde</i>	<i>Betekenis</i>
<i>F</i>	6	Aantal in- en uitgangsvezels
<i>W</i>	8	Aantal golfengten per vezel
<i>B</i>	0...64	Aantal golfengtepoorten van de schakelmatrix dat gebruikt wordt voor circulatie doorheen de FDL-buffer
<i>D</i>	$2 \cdot L$	Vertraging van de FDL-buffer
<i>L</i>	1.5...20	Gemiddelde treinlengte (eenheid=tijdslot)

Om de logische performantie te evalueren hebben we vanzelfsprekend vooral belang gehecht aan het bepalen van het pakketverlies (dat optreedt wanneer de voorziene buffering en golfengte-conversie falen in het oplossen van contentie). Aangezien we een trein als een ondeelbare gegevenseenheid beschouwen, rekenen we hem als volledig verloren zodra één enkele wagon de uitgaande vezel niet bereikt. Het is welbekend dat schakel-algoritmes niet altijd eerlijk zijn, in de zin dat ze pakketten discrimineren op basis van hun lengte: in het asynchrone pakket-

schakelen met leemte-vulling [180] maken korte pakketten meer kans om succesvol doorgeschakeld te worden dan lange [188]. Een ander belangrijk aspect is de mate waarin het algoritme kwaliteitsdifferentiatie kan realiseren. We zullen ook daaraan de nodige aandacht besteden. De belasting van de controle-eenheid die de schakelaar stuurt is een criterium dat we gemakkelijk kunnen evalueren: het aantal beslissingen dat moet worden genomen is gelijk aan het aantal hoofdingen die de schakelaar binnenkomen. Vermits dit slechts één hoofding per variabele lengte pakket betreft in de trein-aanpak tegenover één per slot voor het wagon-gebaseerd schakelen, zal het aantal operaties van het wagonconcept hoger liggen met een factor evenredig aan de gemiddelde treinlengte in slots gemeten. Merk overigens op dat we ondersteld hebben dat de hoofdingen op een orthogonaal kanaal verstuurd worden (zoals geschetst in Figuur 3.8): het feit dat in een wagon-aanpak er meer hoofdingen verstuurd moeten worden heeft geen repercussies op de treinlengte gemeten in slots.

Het schakel-algoritme dat we gebruiken is in beide gevallen hetzelfde. Zowel wanneer we op basis van wagons als treinen beslissen, ondernemen we in elk tijdslot de volgende stappen: (i) selecteer voor elk van de F uitgaande vezels zoveel pakketten als er op de vezel in het beschouwde tijdslot t_0 vrije golflengtes zijn en schakel deze rechtstreeks door; (ii) selecteer uit de overblijvende pakketten die op t_0 aankomen er zoveel als er op dit ogenblik t_0 vrije poorten naar het recirculerend buffer zijn en schakel ze naar de corresponderende FDL's. Bemerkt dat in het geval van een wagon-gebaseerde aanpak het aantal vrije uitgangen op een uitgaande vezel steeds W zal bedragen en het aantal vrije bufferpoorten B : de wagons uit het vorige tijdslot zijn dan al volledig doorgeschakeld. De controle-eenheid dient dus niks te onthouden over beslissingen die genomen werden in het verleden⁸. Dit is duidelijk wel het geval voor het in hun geheel schakelen van treinen: wanneer een trein wordt doorgeschakeld naar een bepaalde poort, dient men het ogenblik te onthouden waarop die poort terug vrijkomt. We kunnen dan ook verwachten dat, gezien de mindere complexiteit, de controle-eenheid in het geval van wagonschakelen sneller en/of met minder (dure) hardware de vereiste beslissingen kan nemen.

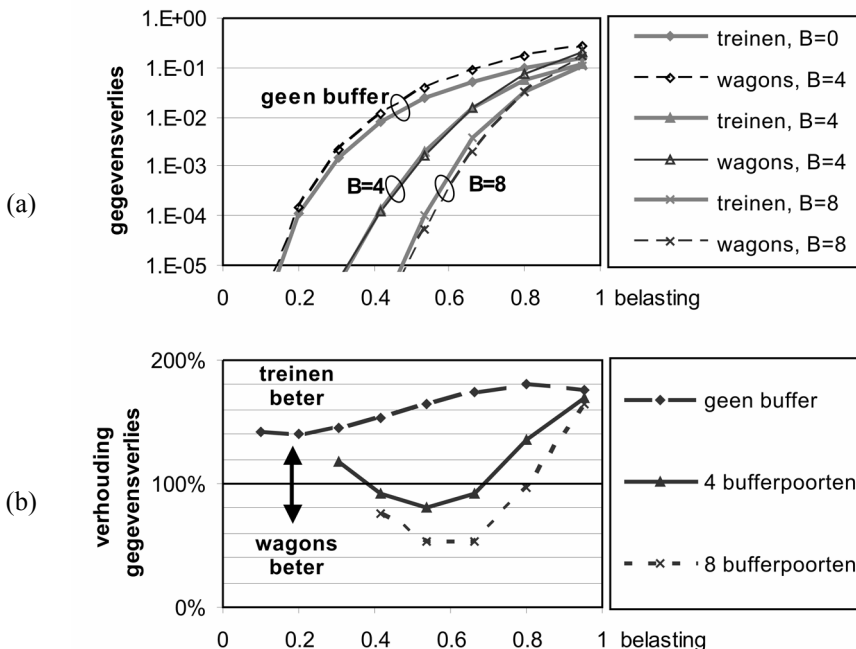
3.3.1.2 Invloed van de belasting

Met een eerste set van simulatieresultaten trachten we te achterhalen of er een bepaalde belasting bestaat waarvoor één van de schakelconcepten beter presteert. De studie [185] leerde immers dat een schakelconcept met variabele pakketten voor lage belastingen tot lagere pakketverliezen kan leiden. Aangezien de verlieskans duidelijk zal afhangen van de hoeveelheid buffer hebben we een staalkaart van drie buffergroottes beschouwd: geen buffer, vier, of acht bufferpoorten ($B = 0, 4, 8$). In Figuur 3.9a is voor toenemende ingangsbelasting de fractie verloren gegevens uitgezet (i.e. de gesommeerde lengte van de verloren treinen gedeeld door de

⁸ We opteren immers opnieuw voor een PostRes strategie [168] waarbij pakketten die de buffer verlaten op dezelfde manier behandeld worden als nieuw aankomende: er wordt op het ogenblik van bufferen geen golflengte op een uitgaande vezel gereserveerd. Bovendien hebben in het onderzochte geval alle FDL's dezelfde lengte.

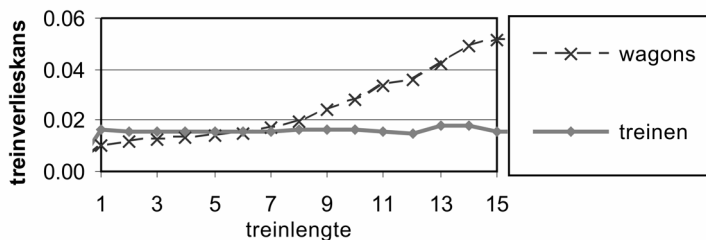
gecumuleerde lengte van treinen die bij de schakelaar aankwamen), die uiteraard een stijgend verloop kent.

Om de verschillen in gegevensverlies tussen trein- en wagonconcepten duidelijker te onderscheiden, hebben we in Figuur 3.9b de verhouding van het gegevensverlies gerealiseerd met de wagon-aanpak gedeeld door die in geval van het treinconcept. Wanneer er geen buffer is ($B=0$), zien we dat de wagon-aanpak nooit beter presteert: de vernoemde verhouding ligt hoger dan 100%. Wanneer er een recirculerende buffer voorzien wordt, stellen we evenwel vast dat wagon-gebaseerd schakelen wel voordeel oplevert. Maar dit is slechts waar voor een bepaald bereik van de belasting: voor erg hoge of erg lage belasting blijft het trainschakelen doeltreffender. De verklaring voor het slechter presteren van een wagon-aanpak luidt dat, wanneer er verschillende treinen in de tijd overlappen, het zo zal zijn dat van meerdere treinen een (beperkt) aantal slots verloren zal gaan, zodat het aantal treinen dat we als verloren beschouwen typisch hoger zal liggen dan in het trainschakelen. Voor een schakelaar met een buffer is dit evenwel niet altijd het geval, vermits een concept met vaste schakeleenheden die bufferruimte efficiënter kan benutten dan één met variabele. Wanneer de belasting niet te hoog is, zodat de contentie in grote mate kan worden opgelost, leidt dit tot lagere verliezen voor het wagonconcept. Voor lage belastingen is het zo dat er voldoende bufferruimte is zodat zelfs een treingeschakeld concept er genoeg aan heeft om de meeste contentieproblemen op te lossen: daar levert het efficiënter uitbaten van bufferruimte door de wagon-aanpak geen netto voordeel op.



Figuur 3.9: Wagens vs treinen: (a) gegevensverlies, (b) verhouding wagon/trein.

Zoals eerder aangegeven kan een schakel-algoritme voor pakketten van variabele lengte van oneerlijkheid blijf geven en langere pakketten discrimineren. Om te evalueren of dit voor de trein- en wagon-gebaseerde schakel-algoritmes ook zo is, hebben we in Figuur 3.10 de verlieskans uitgezet in functie van de lengte van de trein in geval van een buffer met $B=4$ poorten en een belasting van 0.62. Doordat de diepte van de buffer, i.e. de lengte van de FDL's, zo gekozen is dat vrijwel alle pakketten (ongeveer 95% om preciezer te zijn) in een FDL passen, zien we dat in het geval van een treinconcept geen noemenswaardige oneerlijkheid voorkomt. Voor een wagonschakelend algoritme vinden we evenwel een duidelijke discriminatie van de langere treinen, wat het evidente gevolg is van het feit dat de individuele wagons onafhankelijk van elkaar behandeld worden. Het schakel-algoritme geeft alle wagons die in een zelfde tijdslot aankomen immers dezelfde kans om door-geschakeld te worden: gegeven een bepaalde wagonverlieskans is het duidelijk dat de treinverlieskans min of meer evenredig zal zijn met de lengte gemeten in aantal slots of wagons.



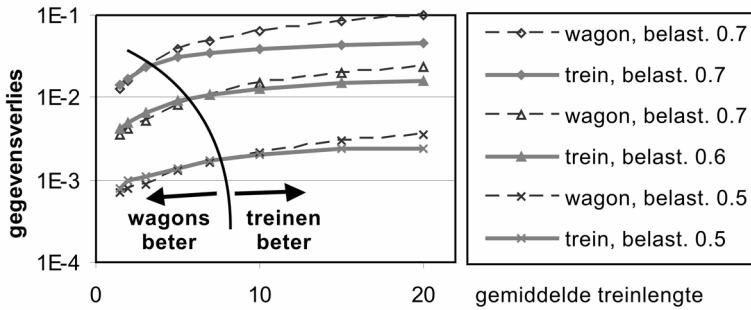
Figuur 3.10: Illustratie van discriminatie: verlieskans in functie van de treinlengte, voor een belasting van $l=0.62$ in een schakelaar met $B=4$ bufferpoorten.

3.3.1.3 Invloed van de granulariteit

Zoals een studie voor bus-structuren met een MAC-protocol aantoonde [105], kan het antwoord op de vraag of trein-gebaseerd dan wel wagon-gebaseerd schakelen te verkiezen is, afhangen van de gemiddelde lengte van de treinen. Anders geformuleerd komt dit erop neer dat voor een gegeven absolute lengte (gemeten in bits, of dus seconden voor een gegeven bandbreedte) het al dan niet beter zijn van een wagon-gebaseerde aanpak kan afhangen van de keuze van de lengte van een wagon, of met andere woorden van de (slot-)granulariteit. De keuze van een bepaalde slotlengte zal uiteraard bepalend zijn voor het in meer of mindere mate verspillen van bandbreedte door zogenaamde “padding” die ervoor zorgt dat een trein een geheel aantal slots bedraagt (men spreekt ook soms van “excess load” [179]). Maar zelfs wanneer we dit effect buiten beschouwing laten, zal ook de prestatie in termen van gegevensverlieskans beïnvloed worden.

Om dit te onderzoeken hebben we in Figuur 3.11 het gegevensverlies uitgezet voor toenemende gemiddelde treinlengte. Hierin zien we het intuïtief verwachte resultaat dat een wagon-aanpak enkel voordeel kan opleveren wanneer de gemiddelde treinlengte voldoende klein is. Wanneer we naar de verliescurves kijken voor

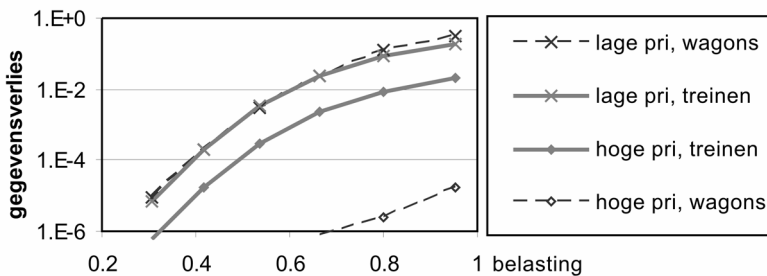
verschillende belastingen (we tonen ze voor $\rho = 0.5, 0.6, 0.7$), stellen we—niet algeheel tot onze verbazing—vast dat het keerpunt licht verschuift naar langere treinlengtes (dus fijnere granulariteit) met lagere belasting.



Figuur 3.11: Wagons vs treinen: gegevensverlies voor toenemende gemiddelde treinlengte in een schakelaar met B=4 bufferpoorten.

3.3.1.4 Kwaliteitsdifferentiatie voor treinen vs wagons

In onze studie omtrent het synchroon schakelen van pakketten met een vaste lengte is gebleken dat een eenvoudig prioriteitsmechanisme volstaat om uitgesproken differentiatie te realiseren in termen van verlieskans en vertraging. In deze sectie onderzoeken we of eenzelfde techniek ook voor het slot-gebaseerd schakelen van pakketten van variabele lengte voldoening kan schenken. Hiertoe passen we het schakel-algoritme zo aan dat bij de selectie voor doorschakelen naar een uitgaande vezel of naar de buffer onvoorwaardelijk voorrang gegeven wordt aan pakketten van hogere prioriteit. Hierbij zullen we eerder genomen beslissingen niet herroepen en niet aan zogenaamde preëemptie doen. Dit betekent dat in het geval van het treinschakelen we de trein steeds in zijn geheel laten: ofwel wordt hij volledig doorschakeld, ofwel gaat hij volledig verloren. Enkele wagons aan de staart kunnen niet weggelaten worden ten voordele van een later aangekomen trein van hogere prioriteit.



Figuur 3.12: Kwaliteitsdifferentiatie voor wagons vs treinen: gegevensverlies per prioriteitsklasse (40% hoogste prioriteit), in functie van toenemende belasting voor een schakelaar met B=4 buffer poorten.

In Figuur 3.12 hebben we de verlieskans uitgezet in functie van de belasting voor een schakelaar met $B=4$ bufferpoorten, waarvoor we het verkeer in twee prioriteitsklassen opdeelden, met 40% verkeer van de hoogste prioriteit. We stellen vast dat net zoals in het schakelen van pakketten van vaste lengte dit voor de hand liggend prioriteitsschema een sterke differentiatie oplevert. Ook in het treinconcept slagen we erin de verliezen voor de prioriteitsklassen duidelijk te scheiden, maar veel minder uitgesproken: waar de verlieskansen voor een wagon-aanpak tot zowat vier grootteordes verschillen, is de differentiatie in geval van treinschakelen beperkt tot één grootteorde of minder.

3.3.2 Kwaliteitsdifferentiatie voor asynchroon pakketschakelen

Uit voorgaande studie omtrent treinen en wagons blijkt dat een heel eenvoudig prioriteitsmechanisme slechts een beperkte graad van kwaliteitsdifferentiatie bereikt. In deze sectie zullen we enkele alternatieven voor differentiatie bestuderen voor het schakelen van pakketten van variabele lengte die asynchroon aankomen.

In de literatuur bestaat er een behoorlijk brede waaier aan mogelijkheden tot het verlenen van gedifferentieerde diensten in een IP-over-WDM-scenario [189]. Reservatie is een welgekende manier om bepaalde kwaliteitsgaranties te kunnen realiseren: golflengte-convertoren [163] of buffers [190] kunnen exclusief voorbehouden worden voor bepaalde verkeersklassen, waarbij de hoogste prioriteit tot meer middelen toegang krijgt. Zulke reservaties voor bepaalde klassen kunnen zowel statisch vastliggen, als dynamisch aangepast worden [190]. We zullen ons in de hierna volgende studie evenwel beperken tot mechanismen zonder reservatie, waarbij het duidelijk is dat die mechanismen eenvoudig kunnen worden uitgebreid met een reservatiesysteem.

In een OBS-aanpak maakt men gebruik van gedifferentieerde offsets: voor verkeer van hogere prioriteit worden de hoofdingen langer op voorhand uitgestuurd voor de eigenlijke pakket-inhoud [184]. Er is ook aangetoond dat zogenaamde segmentatie van de bursts helpt om de performantie (i.h.b. gegevensverlies) te verbeteren [191]. Hierbij is het toegelaten om langere bursts slechts gedeeltelijk door te sturen en om contentie te vermijden een portie aan de kop kan weglaten. De auteurs van [192] merkten op dat kwaliteitsdifferentiatie in zo'n geval gerealiseerd kan worden door gegevens van hogere prioriteit in de staart van een burst te positioneren, omdat die minder kans maakt verloren te gaan. Het is evenwel duidelijk dat dit een zeer impliciete manier is om differentiatie te bekomen. Bovendien is ze niet bruikbaar om differentiatie te verkrijgen tussen verkeersstromen met verschillende bron- en eindknoten. Dit verkeer zal immers nooit in dezelfde optische pakketten terechtkomen, zelfs al valt een deel van de gevolgde route samen. In onze studie zullen we segmentatie als middel tot gedifferentieerde performantie dan ook niet beschouwen. We zullen zoals in de hoger uiteengezette trein-studie steeds pakketten in hun geheel schakelen of laten verloren gaan.

Een andere aanpak om op basis van prioriteitsindicaties verschillende dienstkwaliteit te bekomen is het met opzet laten verloren gaan van gegevens van lagere prioriteit—ook als er op dat ogenblik voldoende ruimte is om eventuele contentie op

te lossen—om op die manier meer middelen vrij te maken voor verkeer van hoge prioriteit. Deze filosofie wordt onder meer toegepast in de techniek van proportionele dienstkwaliteit (Eng. *proportional QoS*), zoals in [193]. Hoewel dit inderdaad lagere verliezen oplevert voor verkeer van hoge prioriteit is het zo dat de verlieskans, gemiddeld genomen over de totale hoeveelheid geschakeld verkeer, toeneemt. We zullen die techniek van opzettelijk verliezen van gegevens hier verder niet in aanmerking nemen. Het is evenwel evident dat, wanneer de hieronder bestudeerde algoritmes voor kwaliteitsdifferentiatie niet volstaan om voor verkeer van hoge prioriteit voldoende lage verliezen te bereiken, we deze algoritmes eenvoudigweg met intentioneel verliezen kunnen uitbreiden.

Samengevat zullen we in onze studie enkel methodes beschouwen zonder segmentatie, zonder reservaties en zonder opzettelijk gegevensverlies. Verder zal het zo zijn dat eenmaal een pakket op een uitgaande vezel werd toegelaten, deze beslissing niet zal herroepen worden. We vervolgen met een beschrijving van de bestudeerde alternatieven.

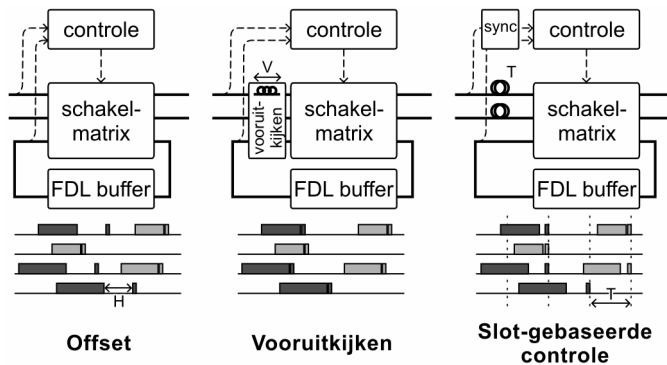
3.3.2.1 Drie alternatieven voor kwaliteitsdifferentiatie

Het schakel-algoritme dat we toepassen bestaat net als in vorige studies uit verschillende stappen en is opnieuw een PostRes-mechanisme dat voor gebufferde pakketten niet meteen een golflengte op de uiteindelijke uitgaande vezel reserveert: (i) gebruik LAUC-VF [109] om een vrije golflengte te vinden op een uitgaande vezel, (ii) als dit niet lukt, pas het LAUC-algoritme [109] toe om de meest geschikte recirculerende bufferpoort te vinden, (iii) laat het pakket verloren gaan wanneer ook dit mislukt. De drie onderzochte technieken voor differentiatie zijn:

- **Offset differentiatie (offset):** Dit is de hoger vermelde techniek die courant is voor zogenaamd burstschakelen (OBS-JET). Doordat hoofdingen van pakketten van hogere prioriteit eerder aankomen dan die van lagere prioriteit zal verkeer van hogere prioriteit meer kans hebben om te worden doorgeschakeld zonder verloren te gaan of zelfs gebufferd te worden.
- **Vooruitkijken (voortuit):** In dit geval wordt er geen onderscheid gemaakt in offset van de pakkethoofding. De schakelaar krijgt na het ontvangen van de hoofding evenwel een zekere tijd V om de uiteindelijke schakelbeslissing te nemen. Dit kan worden gerealiseerd door, zoals in Figuur 3.13 geschetst, een vertragingsslijn te plaatsen tussen het afsplitsen van de hoofdinginformatie naar de controle-eenheid en de ingang van de schakelmatrix. Op die manier kan een pakket van lagere prioriteit verdrongen worden door een broertje van hoger belang dat tot V later aankomt. (Dezelfde filosofie van het inbouwen van een zekere bedenktijd om de schakelbeslissing te nemen wordt gevolgd in [27], maar wordt daar toegepast in een totaal andere context.)
- **Slot-gebaseerde controle (slot):** Het laatste alternatief neemt, opnieuw zonder onderscheid in hoofding-offset, de schakelbeslissingen in groep voor alle pakketten die tijdens het zelfde zogenaamde slot van duur T aankomen en wel op het einde van dat slot. Op die manier kan aan een pakket van hogere prioriteit inderdaad voorrang gegeven worden ten opzichte van tot een tijd T eerder aangekomen verkeer van lagere prioriteit. Om dit te bewerkstelligen volstaat een

gesynchroniseerd controle-kanaal dat de hoofdingen bevat, of minstens een (elektronische) synchronisatie-eenheid voor de schakelcontrole, zoals in de figuur geschetst.

Bemerk dat een studie voorgesteld in [194] voor één enkele bufferloze schakelaar dezelfde performantie zal bekomen als Vooruitkijken: daar heeft men geen vertraginglijnen om de bedenktijd in te bouwen (zoals in Figuur 3.13), maar onderstelt men dat voor alle pakketten, ook die van lage prioriteit, de hoofding voldoende ver vooruit wordt gestuurd zodat een genomen schakelbeslissing even later kan worden herzien. Merk ook op dat dit vereist dat men via het controlekanaal hoofdingen, die al werden doorgelaten voor lage prioriteitspakketten, kan annuleren ten voordele van later aankomend verkeer van grotere prioriteit.



Figuur 3.13: Drie alternatieven voor kwaliteitsdifferentiatie voor asynchroon geschakelde pakketten van variabele lengte.

3.3.2.2 Methodologie

Net zoals in vorige studies zullen we de voorgestelde alternatieven vergelijken door middel van simulatie. De structuur van de schakelknoop is ook hier die van Figuur 3.1, met als waarden voor de parameters: $F=6$ in- en uitgaande vezels met $W=8$ golflengtes elk en een variërend aantal bufferpoorten $B=0\dots 64$ die allen dezelfde vertraging D realiseren.

Voor het verkeersmodel hebben we Poisson-aankomsten gebruikt met een links-getrunceerde exponentiële verdeling voor de pakketlengtes met gemiddelde lengte L en minimale lengte $L/2$. De bestemming van het aankomend verkeer was uniform gespreid over alle uitgaande vezels. We beschouwden twee prioriteitsklassen, met 40% verkeer van hoogste prioriteit, waarbij de klasse werd aangegeven in de hoofding van de onderscheiden pakketten. Die hoofding wordt al dan niet vooruit gestuurd op de eigenlijke pakket-inhoud, afhankelijk van het beschouwde mechanisme voor kwaliteitsdifferentiatie: de tijd tussen hoofding en inhoud is voor Offset-differentiatie afhankelijk van de prioriteit nul of H , bij Vooruitkijken steeds nul en gelegen in het half-open interval $[0, T)$ voor Slot-gebaseerde controle. We verwaarlozen de tijd nodig voor het verwerken van de hoofding.

Om het verschil in performantie tussen de onderscheiden QoS-mechanismen te evalueren gaan we opnieuw vooral aandacht besteden aan de verlieskans, maar we

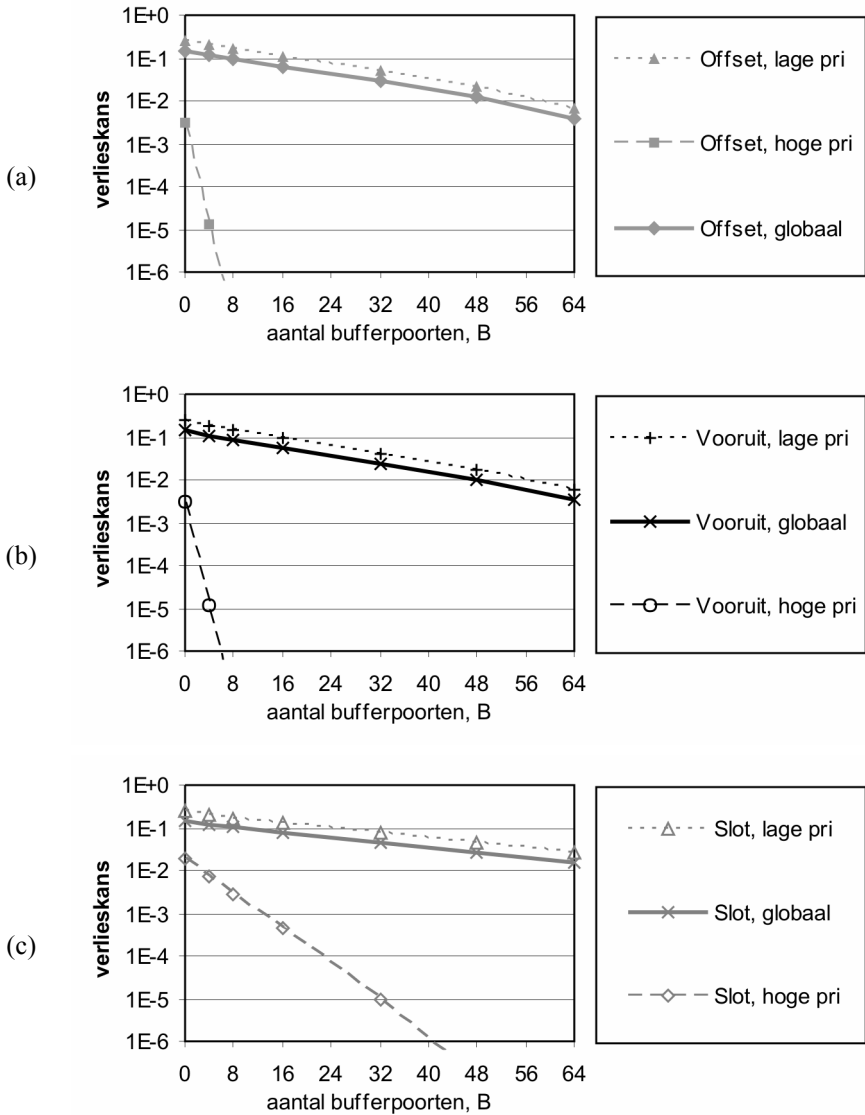
zullen ook de gemiddelde vertraging bekijken. Het doel van de mechanismen is uiteraard duidelijke verschillen in verlieskans per prioriteitsklasse te bereiken. Maar ook binnen een zelfde prioriteitsklasse kan de verlieskans variëren en is discriminatie op basis van pakketlengte niet ondenkbaar [188]. We zullen dan ook ruimte maken voor het kijken vanuit dit perspectief van eerlijkheid.

3.3.2.3 Invloed van het aantal bufferpoorten

Ons eerste experiment bestond erin voor een vaste belasting van 0.8 te kijken naar het verloop van de performantie-maten voor een toenemend aantal bufferpoorten. Het resultaat hiervan is geschetst in Figuur 3.14. De verlieskans is er uitgezet voor volgende parameters van de verschillende QoS-technieken: (i) in Offset was de offset van de hoogste prioriteitsklasse $H=2\cdot L$, (ii) bij Vooruitkijken stelden we de $V=2\cdot L$ en (iii) voor de slotresolutie bij Slot namen we $T=2\cdot L$. De diepte van de buffer, i.e. de gerealiseerde vertraging werd gelijk gesteld aan $D=2\cdot L$ (omdat vanaf die waarde de laagste verlieskans gerealiseerd kan worden vergeleken met minder grote bufferdieptes [168]).

Uit de grafiek in Figuur 3.14 blijkt duidelijk dat het gebruik van slot-gebaseerde controle in hogere verlieskansen resulteert dan Offset of Vooruit, vooral bij een toenemend aantal bufferpoorten B . Waar het verschil in globale verlieskans nog enigszins beperkt blijft, vinden we dat de verlieskans van verkeer van de hoogste prioriteit beduidend hoger ligt (verschillende grootteordes zodra er buffer-ruimte is) dan bij de Offset- en Vooruitkijkende technieken. Desalniettemin vinden we toch een duidelijke differentiatie en zelfs een verlieskans die aanvaardbaar kan zijn wanneer er voldoende bufferruimte is (cf. in de orde van 10^{-6} en lager voor meer dan $B=32$ bufferpoorten).

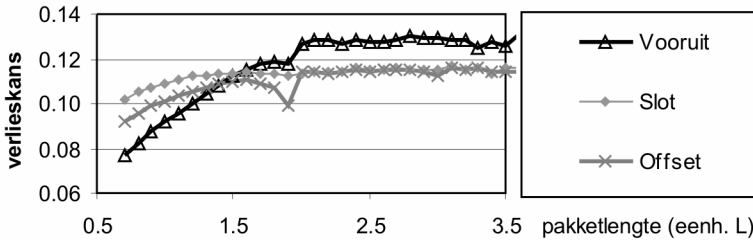
Vergelijken we Offset en Vooruitkijken, dan nemen we slechts een miniem verschil waar: op de logaritmische grafieken valt het relatief kleine voordeel van Vooruitkijken (tot ca. 20% lagere waarden voor de verlieskans) nauwelijks af te lezen. Wanneer we daarentegen de discriminatie van lange pakketten bekijken, stellen we wel een duidelijk onderscheid vast. In Figuur 3.15 hebben we de globale verlieskans in functie van de pakketlengte uitgezet. Deze grafiek bevestigt de verwachting dat korte pakketten in een asynchrone schakelaar bevoordeeld worden: ze ondervinden een lagere verlieskans. In geval van de Vooruitkijkende strategie is de discriminatie op basis van pakketlengte het meest uitgesproken: er is een bijna lineaire toename van de verlieskans tot pakketgroottes gelijk aan de tijd dat we vooruitkijken $V=2\cdot L$. Deze sterkere discriminatie is te wijten aan het preëmptieve karakter van het Vooruitkijken: er wordt een schakelbeslissing genomen op het ogenblik dat een pakket p aankomt. Voor later aankomende pakketten q van dezelfde of lagere prioriteit wordt die beslissing voor p gerespecteerd. Wanneer evenwel later een pakket van hogere prioriteit toekomt kan de schakelbeslissing voor pakket p aangepast worden, wat tot een andere configuratie van de vrije ruimtes op de uitgaande vezels kan leiden dan degene op basis waarvan beslissingen voor q genomen werden. Het gevolg is dat de allocatie van pakketten op golflengtes van uitgaande vezels niet optimaal is (in termen van de vrije ruimte die tussen pakketten overblijft). De langere pakketten zijn de eerste die daaronder lijden.



Figuur 3.14: Pakketverlies-verhouding voor een toenemend aantal recirculerende bufferpoorten B : (a) Offset-gebaseerde differentiatie, (b) Vooruitkijken, (c) Slot-gebaseerde controle.

In Appendix F [32] hebben we ook het gemiddeld aantal circulaties doorheen de buffer in kaart gebracht voor toenemende buffergroottes. Eén circulatie correspondeert uiteraard met een vertraging van $D=2 \cdot L$, de bufferdiepte. Vanzelfsprekend neemt de gemiddelde vertraging toe wanneer er meer bufferruimte

is: pakketten worden vertraagd in plaats van verloren te gaan. Net zoals voor de verlieskans is de rangorde volgens dalende performantie (hier: toenemende vertraging): (i) Vooruitkijken, (ii) Offset, (iii) Slot-gebaseerde controle, waarbij het onderscheid tussen (i) en (ii) heel beperkt is.



Figuur 3.15: Pakketverlies in functie van pakketlengte, voor een pakquetschakelaar met $F=6$ in- en uitgaande vezels van $W=8$ golflengten elk en $B=8$ bufferpoorten, met een belasting van 0.8.

3.3.2.4 Invloed van de klasse-offset

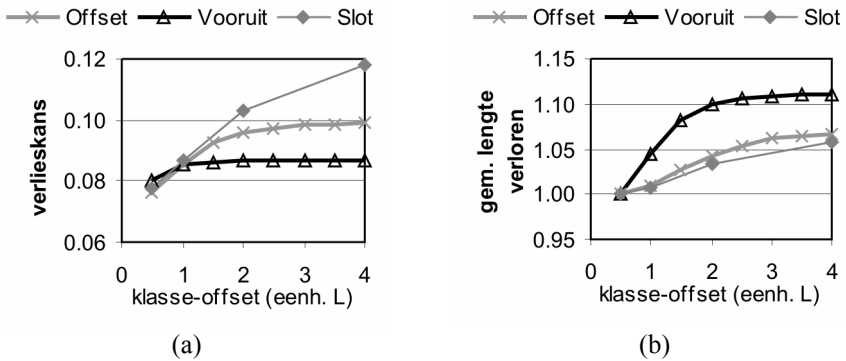
Vanuit het standpunt van kwaliteitsdifferentiatie is de mate waarin de performantie-maten verschillen voor elk van de prioriteitsklassen de belangrijkste graadmeter. De parameter waarmee die separatie beïnvloed kan worden is afhankelijk van het gebruikte differentiatie-mechanisme, maar we zullen die hier met een gemeenschappelijke noemer “klasse-offset” aanduiden. Voor de Offset-gebaseerde methode gaat het dan over het verschil H in offset van de hoofding voor elk van de prioriteitsklassen, terwijl het bij Vooruitkijken de duur V van de vooruitkijk-periode betreft. In geval van slot-gebaseerde controle is het de duur T van het tijdslot.

Om de invloed van de klasse-offset te evalueren, hebben we een reeks simulaties uitgevoerd met toenemende waarden voor die klasse-offset. De resulterende globale gemiddelde verlieskans en vertraging zijn in Figuur 3.16a uitgezet voor een belasting van 0.8 van een schakelaar met $B=8$ bufferpoorten, elk met vertraging $D=4 \cdot L$. De klasse-offset varieëde tussen $L/2$ en $4 \cdot L$.

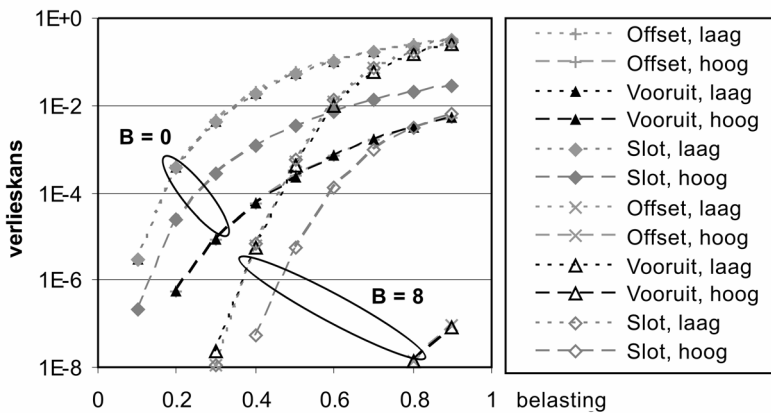
Door de klasse-offset te verhogen vinden we een toenemende globale verlieskans: door pakketten van hoge prioriteit meer voorrang te geven zal hun verlies wel gereduceerd worden, maar dit gaat ten koste van pakketten van lage prioriteit, die de globale verlieskans domineren. Uiteraard is er een grens aan de mate waarin de performantie kan worden verbeterd: zodra pakketten van hoge prioriteit die van lagere prioriteit als het ware niet meer “zien”, i.e. totale separatie is bereikt, verandert de verlieskans niet meer. Wanneer die totale separatie gerealiseerd wordt hangt af van de distributie van de pakketlengte. Voor de gebruikte exponentiële distributie is minder dan 5% van de pakketten langer dan $2 \cdot L$. We zien dan ook een stagnerende verlieskans vanaf ongeveer die waarde. Bij Vooruitkijken treedt die stagnering het vroegst op, terwijl voor een Slot-gebaseerd schakelconcept er meer ruimte lijkt te zijn voor variaties in de klasse-offset.

De invloed van de klasse-offset op discriminatie van langere pakketten blijkt uit Figuur 3.16b. Daar wordt voor dezelfde toenemende waarden van de klasse-offset

de gemiddelde grootte van de verloren gegane pakketten uitgezet. Met toenemende klasse-offset en dus klassenseparatie, verergert ook de discriminatie (voornamelijk pakketten langer dan gemiddeld gaan verloren). Zoals eerder vastgesteld zien we ook hier dat Vooruitkijken het sterkst discrimineert.



Figuur 3.16: Invloed van de klasse-offset: (a) globale verlieskans en (b) gemiddelde grootte van de verloren gegane pakketten voor toenemende klasse-offset $H=A=T$ in een schakelaar met parameters $F=6$, $W=8$, $B=8$, $D=4L$. De klasse-offset en pakketlengte zijn uitgezet in veelvoud van L .



Figuur 3.17: Pakketverlies-verhouding voor toenemende belasting voor een schakelaar met parameters $F=6$, $W=8$, $B=0$ of 8 , $D=2L$ en klasse-offset $H=A=T=2L$. Er was 40% verkeer van hoogste prioriteit. (Merk op dat curves voor laagste prioriteit voor gegeven B zo goed als samenvallen, alsook die van Vooruit en Offset voor de hoogste prioriteit.)

3.3.2.5 Invloed van de belasting

Het liefst zouden we differentiatie bewerkstelligen voor een zo groot mogelijk bereik aan belasting. Om na te gaan of de beschouwde technieken daar ook in slagen, hebben we in Figuur 3.17 de verlieskans uitgezet voor een belasting variërend tussen 0.1 en 0.9. Die curves kunnen bijvoorbeeld gebruikt worden om de “aanvaardbare belasting” te bepalen, met name: hoe zwaar mag een schakelaar belast worden om een gegeven verlieskans niet te overschrijden?

Wanneer we de resultaten voor hoge en lage prioriteitsklassen met elkaar vergelijken, observeren we een afnemend verschil in verlieskans met toenemende belasting voor zowel de Offset-gebaseerde differentiatie als het Vooruitkijken. Voor het bufferloze geval ($B=0$) bijvoorbeeld daalt de verhouding tussen de verlieskansen van lage en hoge prioriteit van een factor van circa 700 tot zowat 60. Voor de Slot-gebaseerde controle vinden we opnieuw een beduidend zwakkere differentiatie, maar ze neemt minder uitgesproken af met stijgende belasting. Ook blijkt de Slot-gebaseerde aanpak minder te kunnen profiteren van extra bufferruimte om daarmee de verlieskans van de hoogste prioriteit sterk te reduceren.

Als we de grafieken bekijken vanuit het standpunt van het bepalen van de aanvaardbare belasting, noteren we dat het voorzien van enige bufferruimte een aanzienlijk hogere last toelaat, zeker wat de hoogste prioriteiten betreft. Maar ook voor de lagere prioriteiten vinden we bijvoorbeeld dat voor een verlieskans van 10^{-3} de toelaatbare belasting verschuift van ca. 0.2 tot iets meer dan 0.5.

3.4 Besluiten

We hebben in dit hoofdstuk een brede waaier aan mogelijke schakel-algoritmes besproken voor een optische pakketschakelaar met een recirculerend buffer gedeeld door verkeer voor alle uitgaande vezels.

Voor het schakelen van pakketten van vaste lengte in een slot-gebaseerde schakelaar hebben we algoritmes bedacht en vergeleken voor het benutten van een buffer dat verschillende vertragingen kan bieden (het Incr-buffer). Het originele aspect van deze studie ligt in de nieuwe Balans-strategie, die zoveel mogelijk vermijdt dat pakketten die bestemd zijn voor dezelfde uitgang de buffer gelijktijdig verlaten en daardoor lagere verlieskansen realiseert dan de overige strategieën. De resultaten gekend voor (elektrische) schakelaars met conventioneel geheugen (i.e. RAM) werden terloops bevestigd: meer bufferruimte (i.e. meer bufferpoorten, of een Incr-buffer in plaats van de Fix-buffer) leidt tot lagere verliezen en voor verkeer met zelfgelijkenis (zoals het gebruikte ParetoAanUit-model) kan een beperkte buffer-ruimte weinig soelaas brengen.

Om gedifferentieerde diensten te kunnen aanbieden, bleek in het geval van vaste pakketten in een synchrone schakelaar het invoeren van een eenvoudig prioriteitsmechanisme te volstaan om duidelijk lagere verlieskansen te realiseren voor verkeer van hogere prioriteit. En dit zonder te leiden tot een hogere globale verlieskans.

Wanneer dezelfde slot-gebaseerde schakelaar gebruikt wordt voor pakketten van variabele lengte, bleek een trein-gebaseerde aanpak—waarbij beslissingen voor

slots die één pakket vormen tesamen genomen worden, namelijk wanneer het eerste slot van zo'n zogenaamde trein aankomt—over het algemeen beter te presteren dan de wagon-aanpak waarbij elk slot afzonderlijk behandeld wordt. Enkel voor een bepaald bereik aan middelgrote belasting van de schakelaar kan de wagon-aanpak iets lagere verlieskansen realiseren, als de gemiddelde lengte van de treinen niet te lang wordt (een aantal slots). Wanneer evenwel differentiatie voor verschillende verkeersklassen gerealiseerd moet worden, blijkt dit eenvoudiger voor de wagon-aanpak: een eenvoudig prioriteitsmechanisme, dat strikte voorrang geeft aan pakketten van hogere prioriteit, volstaat. In de trein-aanpak blijkt zo'n eenvoudige prioritaire behandeling veel minder doeltreffend. Deze studie is, voor zover we in de literatuur hebben kunnen terugvinden, de eerste in zijn soort met betrekking tot een schakelaar voor een vermaasd OPS-netwerk.

Het realiseren van kwaliteitsdifferentiatie voor variabele pakketten werd verder onderzocht voor asynchroon schakelen waarbij dezelfde filosofie als in de trein-aanpak gevolgd werd: schakelbeslissingen worden voor het gehele pakket genomen dat ofwel in zijn volledigheid doorgeschakeld wordt, ofwel niet. Een eenvoudige slot-gebaseerde controle, waarbij schakelbeslissingen gelijktijdig genomen worden voor pakketten die in eenzelfde slot aankomen, blijkt tot vergelijkbare globale verlieskansen te leiden. Het is echter niet in staat dezelfde sterke differentiatie te bereiken als bv. het uit OBS bekende mechanisme van Offset-differentiatie. Een aanpak waarbij de offset niet verschilt voor verschillende pakketten is het Vooruitkijken: in termen van verlieskansen en vertraging is die strategie evenwaardig of zelfs een tikkeltje beter dan Offset-differentiatie. Het minpunt van die Vooruitkijkstrategie is wel dat langere pakketten sterker gediscrimineerd worden. Het is de verdienste van onze studie om de welbekende Offset-gebaseerde differentiatie te vergelijken met het Slot-gebaseerde en een nieuw Vooruitkijkend alternatief waarvoor we in de literatuur geen performantie-resultaten terugvonden.

We hebben van de verschillende technieken voor differentiatie ook de robuustheid onderzocht door de voornaamste parameters voor separatie van de prioriteitsklassen te variëren: (i) de "klasse-offset", (ii) belasting, (iii) het aantal bufferpoorten en (iv) de bufferdiepte. We kunnen besluiten dat (i) de klasse-offset nodig om voldoende separatie te bekomen afhangt van de distributie van de pakketlengte, waarbij voor Vooruitkijken een iets kortere klasse-offset volstaat; (ii) de separatie licht afneemt voor toenemende belasting; (iii) extra bufferpoorten de verlieskans beduidend kunnen reduceren, maar de Slot-gebaseerde aanpak de bufferruimte minder efficiënt kan benutten dan de overige differentiatietechnieken; (iv) toenemende bufferdiepte helpt om verlieskans te reduceren tot op een zeker punt, gerelateerd aan de pakketlengte-distributie.

Hoofdstuk 4

Routing in een OPS-netwerk

4.1 Situering

Nu we de architectuur van een individuele schakelaar en de performantie ervan geanalyseerd hebben, lijkt het nuttig onze blik te verruimen en een netwerk van dergelijke schakelaars te beschouwen. Zo'n netwerk wordt geïnstalleerd om te voldoen aan een gegeven vraag naar bandbreedte. In deze context situeren zich twee gerelateerde deelproblemen: (i) hoeveel bandbreedte moet worden voorzien om aan een gegeven vraag te voldoen en (ii) hoe dienen de gegeven verkeersstromen doorheen dit netwerk gerouteerd te worden? Om deze problemen gezamenlijk op te lossen voor ring-gebaseerde metro-netwerken hebben we een heuristisch plannings-algoritme beschouwd en met behulp daarvan in Hoofdstuk 2 metro-architecturen met elkaar vergeleken. In dit hoofdstuk daarentegen zullen we enkel het routerings-probleem (ii) behandelen in vermaasde ruggengraatnetwerken opgebouwd uit OPS-schakelaars die we in het vorige hoofdstuk nader onderzochten.

Het bepalen van de routing in een vermaasd netwerk is een welbekend probleem. Vooral in het kader van golflengteschakelen werd het grondig onderzocht en herleidt het zich tot het toekennen van routes en golflengtes aan elk van de verkeersstromen (Eng. routing and wavelength assignment, RWA) [195]. Omdat het een NP-compleet probleem betreft [196], neemt men er veelal zijn toevlucht tot heuristieken [195,197–200].

In de pakketgeschakelde aanpak die wij hier beschouwen vervalt de golflengte-toekenning: bandbreedte aanwezig op verschillende golflengtes van een vezel kan gedeeld worden over alle connecties die deze vezel doorkruisen. In die optiek is routing in een OPS-netwerk nauwer verwant aan routing van IP-stromen. Routing in IP wordt typisch bepaald via gedistribueerde routeringsprotocollen zoals OSPF [201] of IS-IS [202]. Die methoden gebruiken kortste-pad algoritmes om de routes tussen knopenparen te bepalen en IP routingstabellen in te vullen. Door het gebruik van een kortste-pad algoritme worden doorgaans een beperkt aantal linken zwaar belast.

Om de verkeerslast gelijkmatiger te verdelen, heeft men zogenaamde last-verdeling (Eng. load balancing) bedacht. Een eerste voor de hand liggende manier bestaat erin de kostmaten aan te passen [203] en ze niet langer te laten bepalen door de fysische afstand. Meer flexibiliteit bekomen we door af te stappen van een

kortste-pad routing en de routes zorgvuldig te kiezen door uitgekiende methodes die men aanduidt als “traffic engineering”. MPLS-gebaseerde netwerken blijken zich hiertoe uitstekend te lenen [205,206]. Wanneer men daarenboven zoals in [204] een stroom tussen gegeven bron en bestemming gaat opsplitsen in deelstromen, slaagt men erin een quasi-optimale spreiding van de belasting over een gegeven netwerk te bereiken.

In de hier behandelde studie zullen we ook zo’n labelschakelend concept onderstellen, maar we splitsen gegeven vragen naar bandbreedte niet op in deelstromen. We trachten een algoritme te vinden dat de verkeersstromen zodanig verdeelt dat de verlieskans geminimaliseerd wordt. Hiertoe zullen we gebruik maken van een schatting van die verlieskans. Die schatting zal gebaseerd zijn op performantie-resultaten van één enkele schakelaar, zoals bekomen met de simulator gebruikt in vorig hoofdstuk.

Merk op dat het beschouwde routeringsprobleem gesteld wordt in een netwerk met gegeven capaciteit. Wanneer het de bedoeling is zowel routing als het bepalen van de benodigde netwerkcapaciteit gekoppeld op te lossen, vervalt men in het geval van WDM-gebaseerde netwerken veelal in het zogenaamde “grooming”-probleem [207]. Voorbeelden van algoritmes voor zulke capaciteitsplanning zijn onder meer te vinden in [208], het daarop geïnspireerde [209] voor asymmetrisch⁹ IP-verkeer, of [210] waar ook de vereiste bandbreedte voor robuustheid tegen mogelijke falingen in rekening gebracht wordt.

In dit hoofdstuk zullen we ons evenwel tot routing beperken en de netwerkcapaciteit als gegeven onderstellen. In de volgende Sectie 4.2 zullen we de gebruikte methodiek, inclusief twee bestudeerde algoritmische varianten, beschrijven. De voorgestelde algoritmen worden geëvalueerd in 4.3. Vervolgens zullen we in Sectie 4.4 het nut aantonen van het schatten van het pakketverlies om op basis daarvan de routing te bepalen. Besluiten staan samengevat in Sectie 4.5.

Eigen publicaties die in dit hoofdstuk verwerkt zijn: [7] uit Appendix H en [21,22] voor de gebruikte heuristische oplossingsmethoden.

4.2 Methodologie

Het routeringsprobleem dat we behandelen luidt: vind in een gegeven netwerk de routes voor een gegeven set verkeersstromen tussen knopenparen zodat het pakketverlies, meer bepaald de maximale waarde ervan gevonden over alle stromen, zo laag mogelijk blijft. Het netwerk bestaat hierbij uit OPS-knopen met gekende performantie (i.h.b. in termen van pakketverlies) en tussenliggende verbindingen met gekende bandbreedte. Een meer formele beschrijving is:

- **Gegeven:**

- $G=(N,V)$ is een gegeven netwerk bestaande uit knopen $n \in N$ en verbindingen $v \in V$, ook wel takken of linken genoemd.

⁹ Het is immers typisch voor IP dat de hoeveelheid verkeer die van een knoop A naar een knoop B stroomt kan verschillen van de hoeveelheid die in tegenovergestelde richting vloeit [211].

- $c(v)$ is de bandbreedte van verbinding $v \in V$
- $D_{b,e}$ is hoeveelheid verkeer die tussen beginpunt $b \in N$ en eindpunt $e \in N$ verstuurd moet worden
- $\text{PLR}(L_f | L_1, \dots, L_F)$ is de verlieskans die optreedt voor een verbinding, met belasting L_f vertrekkend uit een knoop met in totaal F linken waarvoor de overige uitgaande linken een belasting L_i hebben ($i = 1 \dots F$ en $i \neq f$)
- **Gevraagd:**
 - vind een routing $x_{ij}(v)$, waarbij $i, j \in N$ en $x_{ij}(v) = 1$ wanneer de stroom tussen i en j langsheen verbinding $v \in V$ passeert en anders $x_{ij}(v) = 0$ (wat impliceert dat we een vraag niet zullen opsplitsen in deelstromen) en wel
 - zo dat $\max \{\text{PLR}(L_v | \dots), v \in V\}$ minimaal is.
- **Onder randvoorwaarden:**
 - bandbreedte respecteren:

$$L_v = \frac{\sum_{b,e} f_{b,e}(v)}{c(v)} \leq 1, \quad \forall v \in V,$$

$$\text{met } f_{b,e}(v) = D_{b,e} \cdot x_{b,e}(v)$$

- alle vraag wordt gerouteerd tussen de gevraagde knopen:

$$\sum_{v \in I(n)} f_{b,e}(v) - \sum_{v \in I'(n)} f_{b,e}(v) = \begin{cases} D_{b,e} & \text{voor } n = b \\ 0 & \text{voor } n \neq b, e \\ -D_{b,e} & \text{voor } n = e \end{cases}$$

waarbij $I(n)$ de verzameling takken is die vertrekken uit knoop n , en $I'(n)$ de verzameling takken toekomend in n .

Het bepalen van de verlieskans, i.e. de formulering van de PLR-functie wordt behandeld in een volgende Sectie 4.4. Eerst zullen we de algoritmes bespreken die we uitdachten om voor bovenstaand probleem een oplossing te vinden.

4.2.1 Heuristische routeringsalgoritmes

Als oplossing van bovenstaand probleem hebben we gekozen voor een heuristische aanpak. We hanteren een zogenaamde zoom-in filosofie [22], waarbij in een eerste fase vanuit een *globaal* perspectief het volledige netwerk bekeken wordt. De tweede, *lokale*, fase zoomt in op een beperkt deel van het netwerk om de oplossing te verfijnen. Elk van deze fases is iteratief: vertrekkend van de voorlopige oplossing, die bekomen werd in een voorafgaande stap, wordt de route voor een of meerdere vragen aangepast om een lagere verlieskans te bekomen. De initiële oplossing waarvan we vertrekken is de kortste-pad routing op basis van het aantal knopen die doorkruist worden om een gevraagde verkeersstroom te realiseren.

In de *globale fase* wordt gezocht naar de link tussen twee knopen waar het hoogste pakketverlies optreedt. Door het aanpassen van het taggewicht van enkel die link en het herberekenen van alle routes, zullen we trachten de maximale verlieskans die optreedt over het gehele netwerk te reduceren. Het spreekt voor zich dat na het

herrouteren de link met aangepast gewicht niet langer diegene hoeft te zijn waar het maximale verlies—dat nu lager ligt—optreedt. Het aanpassen van het takgewicht, wat origineel 1 bedraagt voor alle verbindingen, gebeurt door dit gewicht te verhogen met een zogenaamde strafmaat P . De waarde van P wordt zo bepaald dat een maximale reductie van de globale verlieskans gerealiseerd wordt. We herhalen het procédé van zoeken naar de link met maximale verlieskans en aanpassen van zijn gewicht, tot we er niet meer in slagen de maximaal optredende verlieskans nog te verlagen.

Na onderzoek voor een aantal test-scenario's bleek de reductie van het pakket-verlies met een globale aanpak slechts beperkt. Dit ontspruit aan de gelimiteerde flexibiliteit die geboden wordt door routing van alle vragen met een zelfde kortstepad algoritme. Daarom gaan we vervolgens over tot een *lokale fase*, waarin niet langer alle vragen tegelijk geherrouteerd worden. We beschouwden twee algoritmische varianten hiervoor: SlechtsteLink en SlechtstePad.

In de SlechtsteLink-variant gaan we net als in de globale fase op zoek naar de verbinding met hoogste verlieskans. Maar nu zullen we de vragen die deze link doorkruisen één per één (in plaats van allemaal tesamen zoals in de globale stap), in een willekeurige volgorde, trachten te herrouteren. We doen dit opnieuw door het takgewicht te verhogen, maar nu enkel voor de op dat ogenblik beschouwde vraag.

Het SlechtstePad-algoritme gaat op zoek naar de verkeersstroom die het meeste pakketverlies ondervindt en sorteert de stromen daartoe volgens dalende PLR. Voor deze stroom wordt een ander pad gezocht door (enkel voor die stroom) aan linken van het originele pad hogere takgewichten toe te kennen. We beginnen hierbij met de link die op dat pad de hoogste verlieskans vertoont, om zo de linken te ontlasten die het ergst lijden aan pakketverlies. Wanneer we er na een vast aantal pogingen niet in slagen de maximale PLR optredend in het netwerk te verlagen, gaan we over tot de volgende verkeersstroom in de lijst.

We beëindigen¹⁰ de zoektocht naar geschikte routes wanneer we er niet in slagen de maximaal optredende verlieskans nog te verlagen. Voor de SlechtsteLink-variant zal dit gebeuren wanneer, na het herrouteren van sommige vragen die de verbinding met hoogste verlieskans doorkruisen, alle andere verbindingen nog steeds een lagere verlieskans vertonen. In geval van SlechtstePad zal het algoritme eindigen wanneer alle stromen aan bod zijn gekomen.

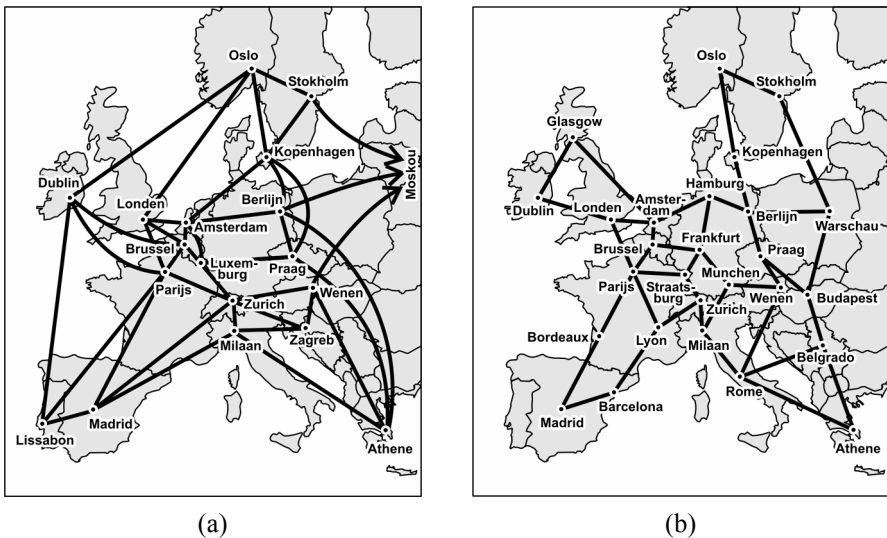
4.2.2 Netwerken en vraagmatrices

Om de prestatie van de algoritmes te evalueren zullen we de routes bepalen in twee pan-Europese netwerken. Het eerste, geschetst in Figuur 4.1a, verbindt negentien steden met veertig linken. Het tweede netwerk in Figuur 4.1b heeft een topologie ontleend aan een studie [212] kaderend in een samenwerking tussen twee Europese onderzoeksprojecten (Lion en Cost266) en gebruikt evenveel verbindingen, maar

¹⁰ Bemerk dat het algoritme nooit in oneindige lussen verzeild zal geraken: we zullen nooit stromen verleggen van pad A naar pad B en dan terug van B naar A. Immers, wanneer we van A naar B overgaan, gebeurt dit enkel als daarmee de verlieskans strikt daalt.

tussen een groter aantal steden. De vraagmatrices die we beschouwen voor onze resultaten zijn:

- **Strikt uniform (uni):** Elke verbinding heeft dezelfde bandbreedte: alle $D_{b,e}$ waarden zijn dezelfde.
- **Random uniform (rnd):** We laten elke vraag $D_{b,e}$ genereren op basis van een uniforme distributie tussen 0 en een maximale waarde die afhangt van de gewenste netwerkbelasting.
- **Random niet-uniform (rndG):** We splitsen de vragen $D_{b,e}$ willekeurig op in twee groepen, waarbij we 96% van de vragen genereren op basis van een uniforme distributie over $[0,7]$ en de overige 4% uniform kiezen uit $[8,11]$. Deze laatste vragen zijn de zogenaamde “grote vragen” (vandaar G). Om een bepaalde netwerkbelasting te bekomen vermenigvuldigen we de aldus bekomen waarden met een gepaste constante.
- **Reëel (rl):** Dit is de verkeersmatrix uit studies die kaderen in het Lion project, met name de waarden voor het jaar 2001 [212], die we met een constante vermenigvuldigen om een gewenste gemiddelde netwerkbelasting te verkrijgen.



Figuur 4.1: Twee pan-Europese netwerken: (a) 19 knopen, (b) 27 knopen.

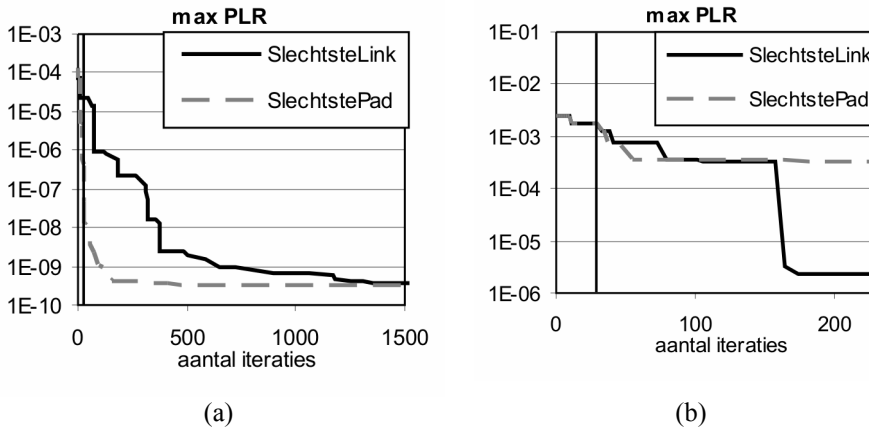
4.3 Evaluatie van de lokale fase algoritmes

Het spreekt voor zich dat we, vooraleer het hoofd te buigen over de voordelen van een PLR-gebaseerd routeringsalgoritme, hebben achterhaald welk van de twee varianten voor de lokale fase het efficiëntst bleek.

Een illustratie van de evolutie van de maximaal optredende verlieskans in het netwerk is in Figuur 4.2 gegeven. De verlieskans wordt er gedurende opeenvolgende

iteraties¹¹ uitgezet voor twee vraagmatrices van het type Random Niet-Uniform. We merken vooreerst op dat het aandeel in de PLR-reductie gerealiseerd door de globale fase beperkt is (aangegeven door een verticale lijn en in beide gevallen slechts een paar tiental iteraties). Wat het verschil in evolutie tussen SlechtsteLink en SlechtstePad betreft, stellen we vast dat SlechtstePad sneller tot een lagere verlieskans komt. Dit vindt zijn oorsprong in het feit dat dit algoritme zijn aandacht richt op die stromen waarvoor de verlieskans het hoogst is. Voor SlechtsteLink is dit niet noodzakelijk zo: alle stromen die de verbinding passeren waar verlies het meest optreedt komen in aanmerking om geherrouteerd te worden.

Het finale resultaat, i.e. de maximale PLR die aan het einde bereikt wordt, is in de meerderheid van de gevallen gelijk voor beide varianten (bv. Figuur 4.2a). In een klein aantal gevallen evenwel slaagt SlechtsteLink erin een reductie van de verlieskans te bekomen die soms een grootte-orde lager ligt, zoals in Figuur 4.2b. De verklaring hiervoor is dat in het SlechtstePad-algoritme een verkeersstroom slechts één keer beschouwd wordt om er een andere route voor te vinden, namelijk wanneer de stroom de hoogste verlieskans ondervindt. Bij SlechtsteLink kan een zelfde stroom meermaals van route veranderen. Of SlechtstePad er al dan niet in slaagt dezelfde lage verlieskans te bereiken, hangt af van het beschouwde netwerk, de vraagmatrix en de willekeurige volgorde waarin de stromen ge(her)routeerd worden.



Figuur 4.2: Evolutie van de maximale verlieskans waargenomen in opeenvolgende iteraties.

Het mag ook duidelijk zijn dat, zelfs wanneer de uiteindelijke maximale verlieskans die optreedt voor beide varianten gelijk is, de gevolgde route voor een gegeven vraag niet noodzakelijk in beide gevallen dezelfde is. Wanneer we de gebruikte capaciteit, zijnde de gebruikte bandbreedte gesommeerd over alle linken, berekenen, vinden we dat de SlechtstePad-variant er gemiddeld iets meer nodig heeft (tussen de twee en vijf procent).

Wanneer we het aantal iteraties beschouwen, zien we dat SlechtstePad er meer nodig heeft om tot een einde te komen. Het algoritme zal immers alle stromen een

¹¹ Met één iteratie bedoelen we één keer (pogen tot) herrouteren van een vraag.

aantal keer proberen herrouteren, terwijl het aantal iteraties in SlechtsteLink bepaald wordt door het aantal linken. Het is duidelijk dat het aantal linken doorgaans beduidend lager zal liggen dan het aantal verkeersstromen (dat van de orde $|N|^2$ is, met $|N|$ het aantal elementen in N). Anderzijds is het wel zo dat SlechtstePad binnen een relatief klein aantal iteraties al zijn laagste verlieskans bereikt, waar dit voor SlechtsteLink iets meer stappen vereist (cf. Figuur 4.2).

Uit deze vergelijking kunnen we besluiten dat de SlechtsteLink variant op basis van de kwaliteit van de gevonden oplossing te verkiezen is: we bereiken er de laagste verlieskans mee en de geconsumeerde bandbreedte in het netwerk ligt een tikkeltje lager.

4.4 Voordelen van het schatten van het pakketverlies

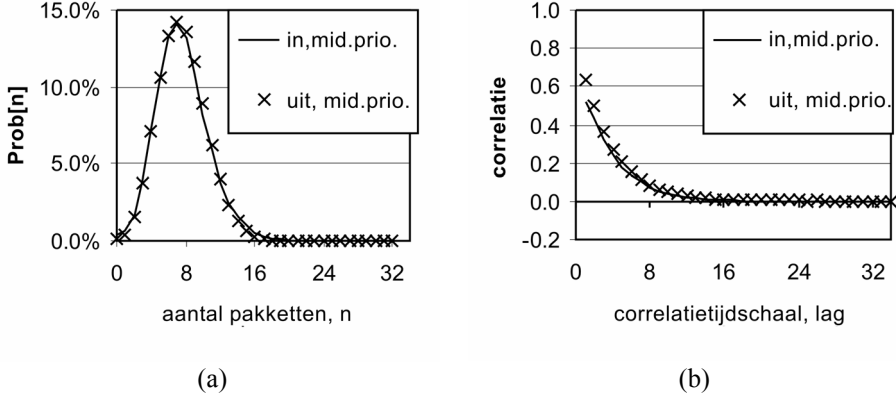
4.4.1 Schatting van pakketverlies

Zoals mag blijken uit de probleemstelling gaan we ervan uit dat we beschikken over een formule die ons het pakketverlies geeft dat optreedt op elk van de uitgaande linken van een optische pakketschakelaar, in functie van de belasting op elk van die linken. Aangezien die formule in elke iteratie van ons heuristisch algoritme opnieuw het pakketverlies op elke link moet geven, is een analytische gedaante gewenst. De performantie op netwerk-niveau in elke iteratie opnieuw bepalen door simulatie zou immers te tijdrovend zijn. Desalniettemin zijn de simulatieresultaten van een alleenstaande pakketschakelaar, zoals in vorig hoofdstuk, wel bruikbaar om tot zo'n formule te komen.

In de beschouwde netwerken is het uiteraard zo dat verkeer aan de ingang van zo'n pakketschakelaar minstens ten dele afkomstig is van de uitgaande vezels van een andere schakelaar. Opdat de resultaten van een enkelvoudige knoop nuttig zouden zijn voor de netwerk-scenario's, mag het verkeersprofiel van een stroom die de schakelaar doorkruist dus niet significant wijzigen. We hebben dan ook geverifieerd of dit zo is. Voor de verschillende trafiektypes uit Hoofdstuk 3 hebben we per prioriteitsklasse het aantal pakketten bijgehouden dat in elk tijdslot de schakelaar binnenkomt en verlaat via elk van de vezels. We hebben dit aantal pakketten dat in een zelfde tijdslot de schakelaar binnenkomt n_T genoemd. Om het aankomstenproces te karakteriseren hebben we onze aandacht gevestigd op de probabiliteitsverdeling van n_T en de correlatie tussen het aantal pakketten in tijdsloten die *lag* uiteen liggen: tussen n_T en n_{T+lag} .

De schakelaar zal het profiel van een verkeersstroom op twee manieren kunnen beïnvloeden: door pakketten verloren te laten gaan en door pakketten te vertragen via de buffer. Pakketverlies zal voornamelijk impact hebben op de probabiliteits-distributie van n_T , terwijl vertraging eerder de correlatiestructuur veranderen kan: buffering kan pakketten meer in de tijd spreiden. Laatstgenoemd effect zal het sterkst optreden voor de Balans-strategie met een Incr-buffer. Een voorbeeld van de opgetekende histogrammen voor n_T en de correlatie is geschetst in Figuur 4.3 voor het GeoAanUit-model in een schakelaar met Incr-buffer van $B=64$ bufferpoorten,

gebruik makend van de Balans-strategie. Onze intuïtieve verwachtingen worden er bevestigd: het pakketverlies zorgt voor een beperkte afname van de probabiliteiten voor hogere n_T waarden en de buffer zorgt voor een lichte wijziging van de correlatie tussen n_T en n_{T+lag} . De impact van de schakelaar is evenwel beperkt en we kunnen met voldoende vertrouwen stellen dat hetzelfde model gebruikt kan worden om zowel in- als uitgaand verkeer te beschrijven: noch de vorm van histogrammen, noch die van de correlatie worden grondig veranderd.



Figuur 4.3: Illustratie van het aankomstenproces en de corresponderende karakteristieken van het uitgaande verkeer: (a) aantal pakketten n_T dat aankomt per tijdslot; (b) correlatie tussen n_T en n_{T+lag} .

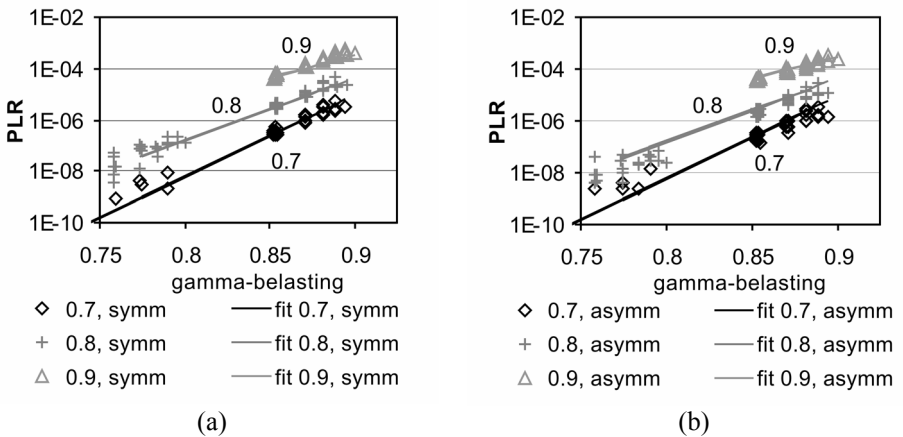
Nu de zin van simulaties voor een enkelvoudige knoop bevestigd is, willen we zulke simulaties gebruiken om tot een gesloten gedaante te komen voor een functie die het verlies geeft in functie van de belasting van de uitgaande linken. En dit voor een voldoende wijd bereik aan belastingen, die uiteraard niet voor alle uitgaande linken dezelfde hoeven te zijn. Het is duidelijk dat de belasting op een bepaalde uitgaande link van doorslaggevend belang zal zijn voor het pakketverlies dat op die link optreedt. Maar aangezien het verkeer bestemd voor die link de buffer zal moeten delen met andere verkeersstromen, zal er toch een zekere afhankelijkheid zijn van de belasting op andere linken. Om die wisselwerking te vatten hebben we de empirische gedaante van een verlieskansformule voorgesteld zoals gegeven in vergelijking 4.1, waar L_f staat voor de belasting op uitgaande vezel f . De maat L_γ is een indicator van de gemiddelde belasting van de schakelaar, waarbij meer gewicht gehecht wordt aan hoog belaste vezels. Het is immers het verkeer dat bestemd is voor die vezels dat intensief van de buffer gebruik zal maken en daardoor de verlieskans op andere vezels kan beïnvloeden.

$$\text{PLR}(L_f | L_1 \dots L_F) \approx g(L_f, L_\gamma) \quad \text{met } L_\gamma = \sqrt{\frac{\sum_{i=1}^F L_i^\gamma}{F}} \quad 4.1$$

Om in te schatten of een dergelijke gedaante voor de verlieskansfunctie zinvol is, en zo ja, een idee te krijgen over een mogelijke analytische vorm van de functie g

uit bovenstaande vergelijking (4.1), hebben we een reeks simulaties uitgevoerd voor elke mogelijke combinatie van lasten L_f genomen uit $\{0, 0.7, 0.8, 0.9\}$. We hebben de verlieskansen bepaald voor elk van die combinaties, gegenereerd door twee types trafficmatrices. In de eerste, *symm* genaamd, draagt elke ingangsvezel i in dezelfde mate bij tot de belasting van een gegeven uitgangsvezel f : de fractie van de belasting L_f gegenereerd door ingangsvezel i bedraagt L_f/F . Het tweede type trafficmatrix *asymm* wordt gekarakteriseerd door asymmetrische bijdragen van de onderscheiden ingangsvezels: het verkeer gegenereerd voor een bepaald koppel in- en uitgangsvezels (i, f) wordt gegeven door vergelijking 4.2.

$$L_{i,f} = \begin{cases} L_f / 2^{1+(i-f) \bmod F}, & i \neq f \\ L_f / 2^{F-1}, & i = f \end{cases} \quad 4.2$$



Figuur 4.4: Verlieskansen en gefitte curves in functie van L_γ : (a) voor een *symm* verkeerspatroon waar elke ingangsvezel in dezelfde mate bijdraagt tot de belasting op een uitgaande vezel; (b) voor een *asymm* verkeerspatroon waar dit niet zo is.

De resulterende grafieken die de verlieskans uitzetten in functie van de gewogen last L_γ zijn gegeven in Figuur 4.4 voor een schakelaar met $F=6$ in- en uitgangsvezels die elk $W=32$ golflengten dragen en een Fix-buffer met $B=32$ bufferpoorten. Door γ te variëren zagen we dat voor relatief hoge waarden (orde 20 en meer) de verlieskansen voor verschillende L_f waarden sterk uiteen lagen, terwijl die voor gelijke (L_γ, L_f) waarden sterk geconcentreerd waren. Vermits de curves voor constante L_f zich in een logaritmische grafiek quasi tot een rechte herleiden, zijn we gekomen tot de gedaante beschreven in vergelijkingen 4.3–4.5 voor de verlieskansfunctie. De waarden voor de onderscheiden parameters werden bekomen met de kleinste kwadratenmethode en zijn samengevat in Tabel 4.1. We gebruikten de resulterende verlieskansfunctie al voor de resultaten uit vorige sectie en zullen dat blijven doen in het resterende deel van dit hoofdstuk.

$$\text{PLR}(L_f | L_1 \dots L_F) \approx p(L_f) \cdot 10^{r(L_f) \cdot (L_\gamma - L_f)} \quad 4.3$$

$$p(L_f) = \text{PLR}(L_f | \text{uniform}) = a \cdot L_f^b \cdot 10^{c - L_f} \quad 4.4$$

$$r(L_f) = u \cdot L_f + v \quad 4.5$$

Tabel 4.1: Gefitte waarden van de parameters voor vergelijkingen 4.1, 4.3–4.5.

<i>Parameter</i>	<i>Waarde</i>	<i>Parameter</i>	<i>Waarde</i>
γ	32.97	a	62.59
u	71.58	b	189.15
v	-82.77	c	-63.81

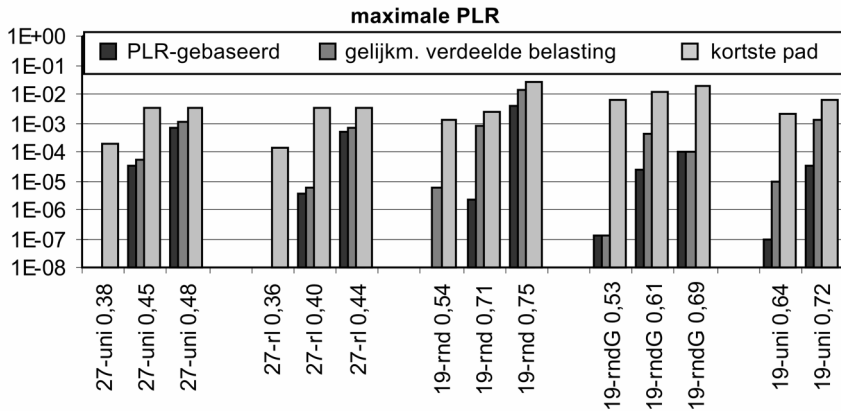
4.4.2 Vergelijking met kortste-pad routing en gelijkmatig verdeelde belasting

Om de zin of onzin van het beschouwde routingsalgoritme te achterhalen hebben we de resulterende maximale verlieskans vergeleken met minder intelligente strategieën. Het ligt voor de hand de reductie van het pakketverlies te beschouwen ten opzichte van zuivere kortste-pad routing. Met de aldus gevonden routes starten we ons streven naar lager pakketverlies.

Zoals aangegeven bij het zoeken naar een analytische gedaante van het pakketverlies voor een gegeven link, is een doorslaggevende factor de belasting op die link zelf. Dit is uiteraard de reden waarom het gelijkmatig verdelen van het verkeer (Eng. load balancing) dikwijls wordt nagestreefd. Merk op dat de algoritmes opgesteld in Sectie 4.2.1 aangepast kunnen worden om zulke gelijkmatige verdeling te bekomen. Het volstaat in de beschrijving “verlieskans” te vervangen door “belasting”: we zullen dan met het algoritme het minimaliseren van de maximaal voorkomende linkbelasting nastreven. Vergelijking van de hiermee gerealiseerde verlieskansen met diegene die bereikt waren door het oorspronkelijke algoritme zal dan leren hoe belangrijk een nauwkeurige schatting van de verlieskans is om die zo laag mogelijk te houden. De linkbelasting kan immers gezien worden als benadering van eerste orde voor de verlieskans.

In Figuur 4.5 zijn de waarden geschetst van de maximale verlieskans die optreedt wanneer de routes respectievelijk met het PLR-gebaseerde algoritme, het gelijkmatig verdelen van de last en de kortste-pad routing bepaald worden. We stellen tot onze opluchting vast dat de reductie van de verlieskans door een intelligente routing tegenover kortste-pad routing aanzienlijk is. Het verschil tussen kortste-pad routing en ons PLR-gebaseerd algoritme kan tot meerdere grootteordes bedragen. Het is duidelijk dat het meeste voordeel behaald kan worden wanneer de gemiddelde netwerkbelasting eerder laag is: er zijn dan voldoende linken met nog vrije capaciteit om door herrouteren de links waar veel verkeer doorstroomt te ontlasten. Om dezelfde reden zien we ook dat bij niet-uniforme verkeerspatronen (Random Niet-Uniform, rndG) de reductie zelfs bij relatief hoge

gemiddelde netwerkbelasting nog aanzienlijk is. Vanzelfsprekend zal een intelligente routing nauwelijks baten wanneer het netwerk nagenoeg ten volle belast is.

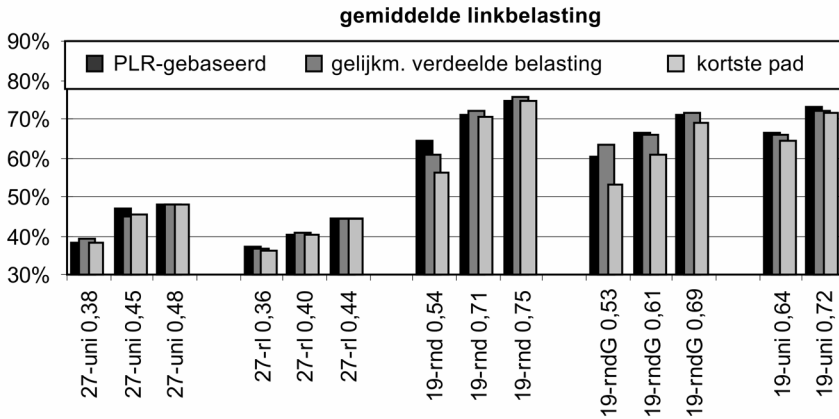


Figuur 4.5: Vergelijking van de maximaal optredende PLR voor de verschillende routeringsalgoritmes. De labels op de X-as geven het scenario aan: het aantal knopen in het netwerk, het verkeerspatroon, de gemiddelde netwerkbelasting.

De vergelijking met de resultaten die behaald werden door gelijkmatige verdeling van de belasting leert dat het wel degelijk zinvol is om van een nauwkeurigere schatting van de verlieskans gebruik te maken. Hoewel ook de verdeling van de belasting, vergeleken met kortste-pad routing, significante reducties van de verlieskans oplevert, stellen we vast dat de kloof met PLR-gebaseerde routing, hoewel vaak beperkt (factor 1.5), soms nog aanzienlijk kan zijn (tot twee grootteordes toe in de onderzochte gevallen). De verklaring ligt in het feit dat enkel de belasting van de link zelf als indicatie van pakketverlies niet volstaat: op die manier wordt interactie met andere stromen, veroorzaakt door het delen van een zelfde (beperkt) buffer, verwaarloosd. Wanneer de belasting laag is (cf. de gevallen met gemiddelde netwerkbelasting van de orde 0.35 à 0.5), is het verschil met een nauwkeurige PLR-schatting beperkt tot een factor 1.5 à 2: in dit geval wordt de buffer niet zwaar gebruikt, met een nauwelijks merkbare correlatie tussen de verlieskansen op verschillende uitgaande vezels tot gevolg. Wanneer de netwerkbelasting evenwel hoger oploopt, zal de buffer zo vaak benut worden dat de interactie niet langer verwaarloosbaar is. Zoals eerder gesuggereerd zal de discrepantie opnieuw verdwijnen voor uitermate hoge netwerkbelasting, vermits intelligente routing daar nauwelijks nog baat.

Bemerkt dat het verschil in bereikte verlieskansen tussen PLR-gebaseerd routeren en gelijkmatig verdelen van de last vooral in het oog springt voor de verkeerstypes Random (rnd) en Uniform (uni). De verklaring is dat in deze gevallen de beschouwde netwerkbelasting betrekkelijk hoog ligt, waardoor verwaarlozing van de interactie tussen stromen op verschillende uitgaande vezels duidelijk nadeel oplevert. Voor het Random Niet-Uniforme (rndG) type ligt de belasting ook vrij

hoog, maar aangezien de individuele verkeersstromen meer bandbreedte vragen kunnen ze niet eenvoudig geherrouteerd worden naar lager belaste linken.



Figuur 4.6: Vergelijking van de gemiddelde linkbelasting voor de verschillende routeringsalgoritmes.

De prijs die betaald wordt door intelligente routing om lagere pakketverliezen te bereiken is uiteraard dat afgeweken wordt van kortste-pad routing en de totale gebruikte bandbreedte hoger komt te liggen. Om een indicatie te krijgen van die toename, hebben we in Figuur 4.6 de benutte fractie van de aanwezige capaciteit uitgezet—wat equivalent is met de gemiddelde linkbelasting. Het blijkt dat in de door ons beschouwde gevallen de toename in bandbreedtegebruik relatief beperkt blijft (een verhoging met maximaal 7% belasting). Het is niet verwonderlijk dat in het bijzonder voor het Random Niet-Uniform (rndG) vraagpatroon die toename hoog is, vermits daar de geherrouteerde stromen de meeste bandbreedte verbruiken. Of een PLR-gebaseerd algoritme dan wel een gelijkmatige lastverdeling leidt tot de hoogste toename, hangt af van het beschouwde geval, vermits de volgorde waarin vragen geherrouteerd worden kan verschillen.

4.5 Besluiten

We hebben in dit hoofdstuk ons werk omtrent het intelligent routeren in een OPS-netwerk uiteengezet. Gebaseerd op simulaties hebben we allereerst een methode ontwikkeld om het pakketverlies nauwkeurig te schatten voor verkeersstromen in zo'n netwerk. We toonden aan hoe een routeringsalgoritme van deze schattingen gebruik kan maken om routes te vinden voor een gegeven set verkeersstromen met het oog op het minimaliseren van de verlieskansen. De reductie van dat verlies ten opzichte van kortste-pad-routing was aanzienlijk (verschillende grootteordes). Vergelijking met een routeringsalgoritme om de belasting zo uniform mogelijk te spreiden over het netwerk leerde dat ook dit een beduidende verlaging van het verlies oplevert. Maar aangezien de linkbelasting slechts een indicator is van de

verlieskans en de wisselwerking—door het delen van een buffer in de schakelknopen—tussen verschillende linken er niet door in rekening wordt gebracht, blijkt een meer nauwkeurige PLR-schatting toch nog voordeliger, soms zelfs tot een paar grootteordes.

Bemerkt dat we in dit hoofdstuk de vragen tussen knopenparen als ondeelbaar en allen van dezelfde prioriteit hebben beschouwd. Voor een bespreking van het geval met meerdere verkeersstromen tussen eenzelfde knopenpaar en meerdere prioriteiten, verwijzen we naar Appendix H.

Hoofdstuk 5

Interactie met het Transport Control Protocol

5.1 Situering

In vorig hoofdstuk behandelden we algoritmes die bepaalden hoe verkeer best doorheen een netwerk gerouteerd wordt. Voor een relatief statische verkeerssituatie zoals we die er impliciet onderstelden zal die routing lange tijd onveranderd blijven. Een ogenblik waarop ze drastisch veranderen kan op korte tijdschaal, is wanneer er in het netwerk een fout optreedt, bijvoorbeeld doordat bij graafwerken een kabel met optische vezels gebroken wordt. Gezien de omvang van de hoeveelheid verkeer die hierdoor verloren kan gaan, heeft men verschillende herstelmechanismen bedacht [206,213,214]. Zo'n mechanisme zal ervoor zorgen dat verkeer de falende netwerkdelen omzeilt en er zo toch nog in slaagt z'n bestemming te bereiken.

Wat er precies gebeurt met verkeer dat moet omgeleid worden van een primair pad naar een herstelpad zal afhangen van de gebruikte technologie en in het bijzonder of die een circuit- dan wel een pakketgeschakeld concept implementeert. Bemerkt dat we bij onze bespreking optische netwerken onderstellen, maar de resultaten kunnen even goed gebruikt worden om bv. elektrische pakketschakelende netwerken met circuitschakelende te vergelijken (zie Appendix I). Aan de fundamentele verschilpunten tussen die circuit- en pakketschakelende concepten zullen we hier bijzondere aandacht besteden vanuit het standpunt van herstel.

De herstelacties hebben hun weerslag op de performantie die gebruikers van het netwerk ervaren. Zoals eerder aangehaald vormt IP-verkeer het grootste deel van wat de gebruikers door het netwerk willen transporteren. Binnen IP domineert één subtype, met name het TCP-verkeer, wat volgens recente metingen verantwoordelijk is voor meer dan 95% van al het IP-verkeer gemeten in bytes [215]. Nu is het TCP-protocol reactief van aard, wat betekent dat het zich aanpast aan de beschikbare bandbreedte. Herstel- en andere acties die de routes voor zo'n TCP-verbindingen veranderen, kunnen dan reactie uitlokken van de adaptieve mechanismen eigen aan TCP. We zullen in dit hoofdstuk zulke effecten behandelen.

In de volgende Sectie 5.2 geven we een beknopt overzicht van de reactieve natuur van TCP, wat zich situeert op de zogenaamde transportlaag in het klassieke

gelaagde OSI-model [216]. Een netwerk zal instaan voor het verzorgen van de vereiste connectiviteit tussen eindpunten van communicatiepartners op de toepassingsslaag. Twee fundamenteel verschillende schakeltechnologiën voor zo'n netwerk zullen we behandelen in Sectie 5.3, met bijzondere aandacht voor de gebruikelijke herstelmechanismen. De eigenlijke resultaten met betrekking tot de effecten van die mechanismen op het TCP-gedrag worden uiteengezet in 5.4. Besluiten staan samengevat in Sectie 5.5.

De eigen publicaties waarop dit hoofdstuk gestoeld is, zijn [4,16] (waarvan de eerste te lezen staat in Appendix I) en [1,3] voor enkele kanttekeningen.

5.2 TCP/IP

Het Transmissie Controle Protocol (TCP) is connectie-georiënteerd: vóór het versturen van de eigenlijke gegevens wordt een verbinding opgezet via een driedovoudig proces van spreekwoordelijk handjes schudden (Eng. three-way handshake) [102,216]. Belangrijker voor de studie in dit hoofdstuk is het feit dat TCP betrouwbaar is, in de zin dat de zender vereist dat hij van de ontvanger een bevestiging (Eng. acknowledgement, ACK) krijgt. De zender zal daarom, na een bepaald aantal nieuwe segmenten te hebben doorgestuurd, wachten tot hij ofwel de correcte ontvangst door de andere partij bevestigd krijgt via een ACK, ofwel bij dat wachten op een ACK z'n geduld verliest (bv. er gaat een timer af). In het laatste geval zal hij aannemen dat de gegevens verloren gingen en opnieuw proberen hetzelfde segment te zenden. De hoeveelheid gegevens die de bron mag versturen—vooraleer er gepauseerd wordt tot ontvangst van een ACK—hangt af van het congestievenster *cwnd* (Eng. congestion window), dat nooit de door de ontvanger bepaalde waarde *rwnd* (Eng. receiver window) zal overschrijden.

Hoewel er nog steeds onderzoek naar en ontwikkeling van TCP-varianten is (zie bv. een recent overzicht in [217]), werden de fundamenteen ervan geruime tijd geleden vastgelegd in RFC 793 [218]. Ze kunnen worden samengevat in vier componenten [217]: (i) het additief toenemen, multiplicatief afnemen (Eng. additive increase, multiplicative decrease, AIMD) dat het congestievenster met één segment ter grootte van SMSS (Eng. Sender Maximum Segment Size) verhoogt per zogenaamde RTT (Eng. Round Trip Time), maar het venster halveert wanneer een segment verloren blijkt te zijn gegaan; (ii) het gebruik van een timer om segmenten opnieuw te versturen wanneer hun ontvangst niet bevestigd wordt binnen een zogenaamd RTO-interval (Eng. Retransmission Timeout); (iii) het trage start mechanisme (Eng. slow start) om bij aanvang van de communicatie af te tasten hoeveel bandbreedte beschikbaar is, tot het *cwnd* een grens bereikt die aangeduid wordt als *ssthresh* (Eng. slow start threshold); en (iv) het principe van ACK-klokken, waarmee verwezen wordt naar het zenden van (nieuwe) segmenten als respons op de ontvangst van een ACK. In onze studie zullen we gebruik maken van de TCP NewReno versie, die beschreven wordt in de meest recente RFC omtrent TCP [219].

Het bovenstaande summiere lijstje van de voornaamste karakteristieken van TCP bevat aanwijzingen over hoe het TCP-gedrag beïnvloed kan worden door het falen van verbindingen en de daarop volgende herstel-acties die de route gevolgd door TCP-stromen veranderen. Een eerste effect is dat door het omschakelen naar

een andere route de RTT, dit is de tijd nodig om de weg van bron naar bestemming en terug af te leggen, verandert. De omleiding rond het falende netwerkgedeelte kan die RTT plotsklaps verhogen, waardoor in TCP de RTO-timer kan afgaan met het nodeloos heruitzenden van verloren gewaande segmenten tot gevolg. De wachttijd RTO wordt bekomen op basis van een schatting van de RTT die SRTT genoemd wordt (Eng. smoothed round trip time) [220,221]. Om zich voldoende snel aan herroutering door herstel-acties te kunnen aanpassen zou die SRTT dus niet al te traag de werkelijke RTT moeten volgen.

Een tweede effect op TCP is uiteraard dat door de faling een reeks segmenten verloren kan gaan, tot de vereiste herstel-acties voltooid zijn. Het verder gebruikte NewReno algoritme zal bij detectie van de verliezen terugvallen op een mechanisme van zogenaamd snel heruitzenden/snel herstel (Eng. fast retransmit/fast recovery). Maar in tegenstelling tot wat de naam laat uitschijnen is dit, afhankelijk van de hoeveelheid verloren gegane segmenten, niet altijd de snelste methode om de verliezen te boven te komen. In sommige gevallen, met name wanneer relatief veel gegevens verloren gaan, kan het voordeliger zijn om de RTO-timer te laten afgaan en op het trage start mechanisme te vertrouwen om de verkeersuitwisseling te herstellen. Wanneer we één enkele TCP-verbinding beschouwen, kunnen we de voorwaarde gegeven in vergelijking 5.1 afleiden voor de hoeveelheid verloren gegane gegevens G opdat het “fast retransmit/fast recovery”-mechanisme sneller de verliezen te bovenkomt dan het afgaan van de RTO-timer. (Zie Appendix I voor details.)

$$\frac{\max(0, sstresh - G) + (sstresh - SMSS)}{2} \cdot \frac{\min(sstresh, G)}{SMSS} \geq 2^{G+1} \cdot SMSS \quad 5.1$$

In de praktijk zal een falende verbinding uiteraard een bundeling van vele TCP-stromen dragen, zodat het bepalen van de hoeveelheid verloren gegevens per individuele stroom niet eenvoudig bepaald kan worden. De reactie van zo’n bundel op falingen en herstel-acties is dan ook niet langer te vatten in analytische formules. Daarom zullen we, zoals verder beschreven, simulaties gebruiken om toch het gedrag van een aggregaat van TCP-stromen te kunnen evalueren.

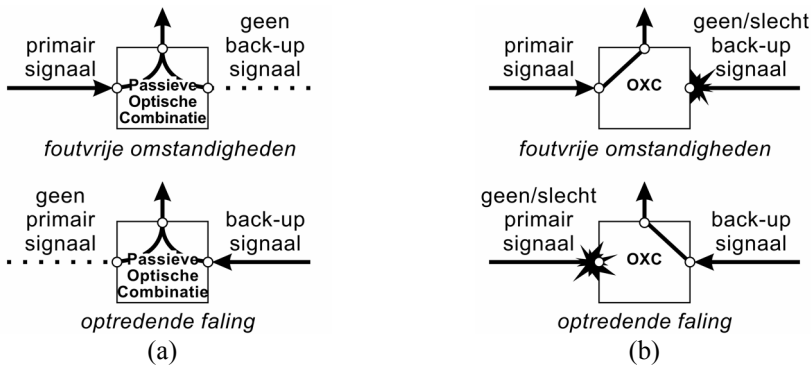
5.3 Herstel: circuit- versus pakketschakelen

Zoals uiteengezet in het inleidende Hoofdstuk 1 zijn voor het transportnetwerk twee concepten mogelijk: pakketschakelen ofwel circuitschakelen. Origineel werden de herstelmechanismen in het kader van MPLS ontwikkeld voor een (elektrisch) pakketgeschakeld netwerk. Terwijl de vertaling ervan naar optisch pakketschakelen evident is, zijn er voor het circuitgeschakelde alternatief van golflengteschakelen toch een aantal niet onbelangrijke verschillpunten [222].

Waar de routes in een pakketgeschakeld concept probleemloos kunnen overlappen en dezelfde golflengten kunnen gebruiken, is het samennemen van twee inkomende circuits in één enkel uitgaand signaal (van dezelfde bitrate) doorgaans niet mogelijk. Nochtans is dit wat typisch gebeuren moet in een schakelaar waar een omleiding voor herstel en een primair werkend pad samenkomen. In het geval van

optisch schakelen kan dit gerealiseerd worden door optische koppelaars zoals schematisch aangegeven in Figuur 5.1a. Een vereiste opdat dit werkbaar zou zijn is dat het werkende pad en het herstelpad nooit gelijktijdig een signaal dragen. Wanneer dit niet gegarandeerd kan worden is een volwaardige schakelaar nodig (Figuur 5.1b), of moeten voor werkende paden en herstelpaden afzonderlijke golflengten gebruikt worden.

Dit brengt ons bij een tweede belangrijk onderscheid tussen pakketschakelende en circuitschakelende technologieën. Om snelle omschakeling in geval van falingen mogelijk te maken, kunnen herstelpaden op voorhand worden opgezet. Men spreekt dan van protectie (in tegenstelling tot restauratie [223]). In circuitschakelen impliceert dit het reserveren van bandbreedte, in het bijzonder een volledige golflengte, voor hersteldoeleinden. Voor pakketschakelen gebaseerd op MPLS verbruikt het op voorhand installeren van de herstelpaden evenwel geen bandbreedte: er worden in de schakelknopen enkel routingstabellen ingevuld met de correcte afbeeldingen van inkomende labels op uitgaande poorten en eventueel nieuwe uitgaande labels. In [222] wordt getoond dat hierdoor in een circuitgeschakeld concept een grotere hoeveelheid bijkomende capaciteit voor protectie-doeleinden geïnstalleerd moet worden.



Figuur 5.1: Optisch samenvloeien van werkend en herstelpad: (a) met passieve koppelaars, of (b) een volwaardig schakel-element.

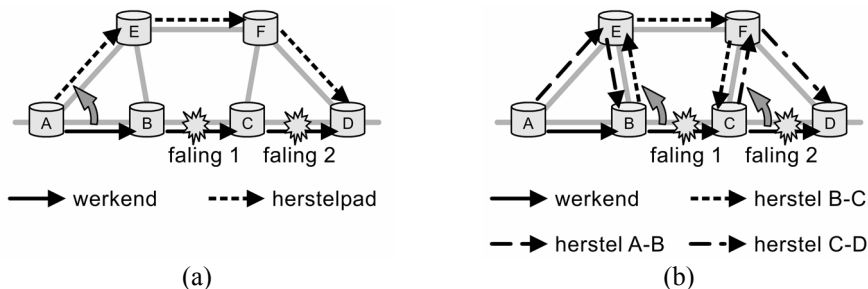
De GMPLS-protectiemechanismen die we verderop zullen beschouwen zijn allen gebaseerd op het vooraf opzetten van labelgeschakelde paden (waarbij voor circuitschakelen een label zal overeenkomen met een golflengte). Zo'n herstelpad zal ofwel een enkele link of knoop overspannen, ofwel de hele weg tussen bron en bestemming: in het eerste geval heeft men het over lokale protectie, terwijl het laatste padprotectie genoemd wordt [224,225]. Deze mechanismen zijn geïllustreerd in Figuur 5.2, waaruit blijkt dat in geval van lokale protectie een herstelpad voorzien moet worden per mogelijke link- of knoofout¹² (de figuur beschouwt enkel linkfouten). Padprotectie vereist slechts één herstelpad per werkend pad, maar

¹² Merk op dat, wanneer zogenaamde labelstapeling (Eng. label stacking) mogelijk is, een herstelpad per beschouwde knoop- of linkfout kan gedeeld worden voor alle werkende paden die het falende netwerkelement doorkruisen [226].

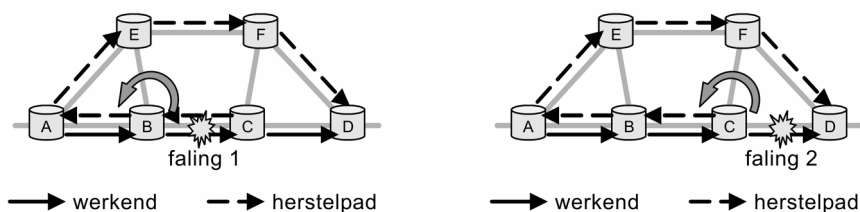
vertrouwt wel op signalisatie om de bronknoop van de verbindingen op de hoogte te brengen van de gewenste omschakeling naar het herstelpad. Lokale protectiemechanismen laten toe dat de schakelacties genomen worden op de plaatsen waar de fout gedetecteerd wordt. Het vermijden van signalisatie kan tijd besparen en er zo voor zorgen dat minder gegevens verloren gaan omwille van een snellere reactietijd.

Een derde protectieschema, geïllustreerd in Figuur 5.3, werd gesuggereerd door Haskin en Krishnan [227] en tracht de voordelen van voorgaande schema's te verenigen: lokale teruglussing (Eng. local loop-back). De kerngedachte die aan de grondslag ligt is slechts één herstelpad per werkend pad te voorzien waarbij toch lokale omschakelacties volstaan (i.e. zonder signalisatie naar andere knopen te vereisen). Figuur 5.3 toont dat de omschakelactie zal verschillen naargelang van de optredende fout, maar steeds hetzelfde herstelpad volgt: eerst terugkerend naar de bronknoop (in tegenovergestelde richting als het werkende pad) en vervolgens via een van het primaire pad gescheiden route naar de bestemming.

Ter afronding van dit korte overzicht van de beschouwde herstelmechanismen nog deze eindnoot: we zullen in onze studies enkel protectie beschouwen op één enkele laag, met name de optische transportlaag. Vanzelfsprekend kan er ook gedacht worden om de taak van het opvangen van netwerkfalingen te spreiden over meerdere lagen (bv. een elektrische IP/MPLS laag en een optische circuit-geschakelde laag), maar daar zullen we hier verder geen gewag van maken: we verwijzen met genoegen naar bv. [2] voor een discussie daaromtrent.



Figuur 5.2: Illustratie van (a) padprotectie en (b) lokale linkprotectie.



Figuur 5.3: Illustratie van lokaal teruglussen voor twee verschillende falingen.

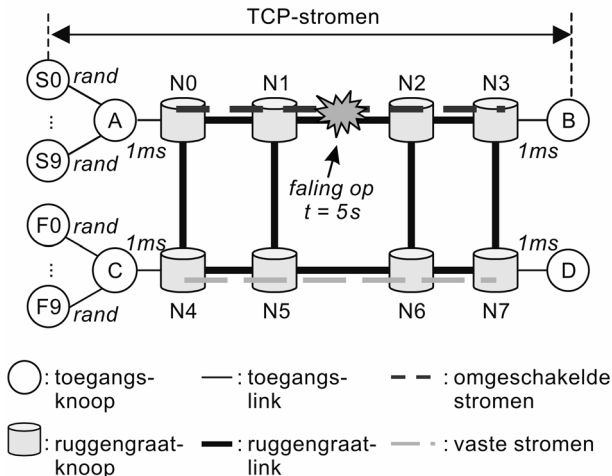
5.4 Effect van herstel-acties op TCP verkeer

5.4.1 Bestudeerde facetten

De hamvraag die we in dit hoofdstuk proberen te beantwoorden, is: hoe goed worden herstelacties, en de neveneffecten die ze teweeg brengen, verteerd door het belangrijkste protocol uit de transportlaag, TCP? We hebben al aangegeven dat het voornaamste effect van een falen en het daaropvolgende herstel is dat er een hoeveelheid verkeer verloren gaat en dat de gewijzigde route leidt tot een verandering van de RTT. We zullen het gedetailleerde onderzoek naar de invloed op een bundeling TCP-connecties opsplitsen in een aantal facetten: (i) de invloed van de omschakel-actie naar het herstelpad, (ii) de invloed van de veranderende RTT en (iii) het effect van het terugschakelen naar het oorspronkelijke pad eenmaal de faling hersteld is. Tot slot zullen we ook een gevalsstudie beschrijven die verifieert of de conclusies ook vertaald kunnen worden naar de protectiemechanismen beschreven in voorgaande sectie. Vooraleer de resultaten daarvan voor te stellen, zetten we nu eerst de gebruikte methodologie uiteen.

5.4.2 Methodologie

Het antwoord op de geformuleerde vragen en gevalsstudies zullen we zoeken door gebruik te maken van de behoorlijk populaire Netwerk Simulator, kortweg ns-2 [228]. Dit is een kosteloos beschikbare discrete-gebeurtenissen-simulator specifiek gericht op onderzoek van (computer)netwerken en ondersteunt onder meer TCP, IP-routing en multicast protocollen over zowel bedrade als draadloze netwerken.



Figuur 5.4: Schets van de simulatie-opstelling.

Om een bepaald herstelmechanisme, of één van de parameters die het TCP-gedrag zal bepalen, te bestuderen, zullen we steeds gebruik maken van dezelfde

generische topologie (of een deel daarvan), zoals geschetst in Figuur 5.4. De gebruikte parameters gemeenschappelijk aan alle simulaties staan opgesomd in Tabel 5.1. We hebben ervoor gekozen de flessenhals—die de verliezen veroorzaakt die de bandbreedte ingenomen door een TCP-verbinding beperkt—in het toegangsnetwerk te lokaliseren (tenminste wanneer geen falingen optreden). Daartoe gaven we toegangslinken een lagere bandbreedte. De propagatietijd voor de ruggengraatlinken werd genomen ergens tussen 1 en 60 ms, afhankelijk van het beschouwde scenario, zodat de totale RTT over de verschillende TCP-connecties varieerde van een paar tien- tot een paar honderdtal milliseconden.

De pakketten zullen vooraf bepaalde routes volgen, waarvoor we de MPLS-voorzieningen van de simulator gebruikt hebben—het spreekt voor zich dat de precieze route voor de herstelpaden zal afhangen van het beschouwde herstelmechanisme. Langs deze paden zullen we een aggregaat van TCP-stromen sturen. Vermits zulke TCP-verbindingen hun oorsprong vinden buiten het ruggengraatnetwerk en zo duidelijk verschillende eind-tot-eind vertragingen kunnen hebben, laten we ze vertrekken in verschillende toegangsknopen, verbonden met de rest van het netwerk met linken van verschillende lengte en dus propagatievertraging: we kozen die willekeurig volgens een uniforme distributie over [10ms,100ms]. Daarenboven lieten we de TCP-bronnen met zenden beginnen op willekeurige ogenblikken, om kunstmatige synchronisatie te vermijden. Zodoende vulden we het netwerk met een mix van TCP-bronnen met ongecorrleerde starttijdstippen en verschillende RTT's. De TCP-stromen zullen, gezien de relatie tussen RTT en zowel de RTO als het ACK-klokken, met verschillende snelheid reageren op de falingen en herstelacties.

Tabel 5.1: Gemeenschappelijke simulatieparameters.

<i>Parameter</i>	<i>Waarde</i>
TCP	TCP NewReno [219]
aantal bronknopen	10 voor elk pad
aantal TCP-stromen	10 voor elke bronknoop (dus 100 per pad)
bandbreedte toegangslink	80% van de bandbreedte van een link in het ruggengraatnetwerk (90% voor de studie van tijdseffecten in Sectie 5.4.3)
<i>rand</i> vertraging toegangslink	random gegenereerd o.b.v. een uniforme distributie in het interval [10ms, 100ms]

De twee belangrijkste parameters die we zullen variëren om de TCP-effecten te beoordelen zijn:

- **Omschakeltijd (δ):** Dit is de tijd die verstrijkt tussen het optreden van een link-faling en de omschakeling van de getroffen verkeersstromen naar het herstelpad.
- **RTT-verandering (ΔRTT):** Deze is gerelateerd aan de extra propagatietijd die de TCP-stromen zullen ondervinden wanneer ze van het originele werkende pad naar

het herstelpad omgeschakeld worden. Voor de propagatietijden zullen we Latijnse letters gebruiken, typisch L . Het verschil in RTT tussen het originele en het herstelpad zal dan ΔRTT genoteerd worden.

Het scenario dat we simuleren om de invloed van deze parameters op het TCP-gedrag te onderzoeken zal bestaan uit een opeenvolging van drie periodes, elk vijf seconden lang, zoals aangegeven in Tabel 5.2. Tijdens de eerste periode zullen de TCP-bronnen het zenden beginnen en geleidelijk aan het netwerk vullen met verkeer. Op $t = 5s$ zal de link N1-N2 falen. Vervolgens, op $t = 5s + \delta$, zal de herstelschakeling uitgevoerd worden en zullen de getroffen stromen voortaan het herstelpad volgen. Om δ exact te regelen zullen we de schakeling als het ware handmatig instellen door de MPLS-routing op precies het gewenste tijdstip aan te passen. Bemerk dat dit betekent dat zowel de gegevens in doorgang op link N1-N2 als diegene die in N1 (of N2 voor de ACK's) aankomen in $[5s, 5s + \delta]$. De laatste vijf seconden van het simulatie-scenario zal de link N1-N2 terug werken en het verkeer wordt teruggeschakeld naar het originele werkende pad. Elk scenario hebben we 150 keer herhaald met andere willekeurig gegenereerde waarden voor de starttijden van de TCP-bronnen en de propagatietijden van de toeganslinken.

Tabel 5.2: Gemeenschappelijke simulatiescenario.

<i>Tijd</i>	<i>Gebeurtenis</i>
$t = 0s$	Begin van de simulatie
$0.5s < t < 1s$	TCP-bronnen beginnen verkeer te genereren
$t = 5s$	de link N1-N2 faalt
$t = 5s + \delta$	de herstelschakeling wordt uitgevoerd
$t = 10s$	de faling van link N1-N2 wordt opgeheven
$t = 15s$	einde van de simulatie

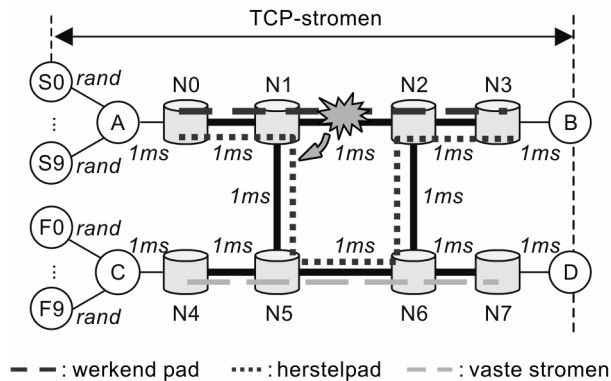
Het criterium dat we hanteren om de herstelmechanismen op hun prestatie te evalueren is TCP-goodput. Dit is het aantal unieke bytes dat succesvol van bron naar bestemming werd overgebracht binnen een bepaald tijdsinterval. Het mag klaar wezen dat het deze goodput is, uitgedrukt in bv. bytes per seconde, waar de eindgebruikers belang aan hechten. In de verder getoonde grafieken zullen we goodput relatief uitzetten tegenover de linkbandbreedte die beschikbaar is voor het aggregaat van TCP-stromen. We zullen dus het totale aantal succesvol overgebrachte bytes nemen in intervallen van duur T en dit delen door het maximale aantal dat we hadden kunnen ontvangen (gegeven door T vermenigvuldigd met de bandbreedte).

5.4.3 Invloed van de omschakelsnelheid

Het eerste aspect dat we behandelen is de impact van de snelheid waarmee na de faling wordt overgeschakeld naar het herstelpad. Het hiertoe bestudeerde scenario is geschetst in Figuur 5.5. We hebben twee sets van TCP-stromen beschouwd: (i) de omgeschakelde stromen die vertrekken uit knopen S0-S9 en die omwille van een

link-faling zullen worden omgeschakeld naar een herstelpad en (ii) een verzameling vaste stromen die ontspringen aan knopen F0-F9 en de falende link niet doorkruisen. Deze laatste zullen dus voor de volledige duur van de simulatie dezelfde route volgen. Het interval waarop we onze aandacht richten is de periode waarin de link faalt, het interval $[5s, 10s]$.

In een *pakketgeschakeld* netwerk kan zoals aangegeven een herstelpad opgezet worden zonder daarvoor bandbreedte te reserveren. In dit geval kan het zo zijn dat de omgeschakelde stromen langs hun omweg een link moeten delen met andere verkeersstromen. Op het ogenblik van de omschakeling zullen de geschakelde stromen de strijd om bandbreedte (die TCP zo volledig mogelijk tracht te vullen) aangaan. Hoe die strijd verloopt, zal afhangen van de omschakeltijd δ . Wanneer de omschakeling ogenblikkelijk plaatsgrijpt, zullen de omgeschakelde TCP-stromen nog aan een vrij hoog tempo aan het zenden zijn, omdat het verlies van de pakketten in doorgang op de falende link nog niet gedetecteerd zal zijn en TCP het zenden nog niet vertraagd zal hebben. Wanneer ze dan de knoop N5 bereiken, waar ze samenkomen met andere TCP-stromen die ook veel bandbreedte innemen, kan dit in die knoop tot verliezen leiden door een tekort aan bufferruimte om de plotse toevloed aan gegevens op te vangen. Zo kunnen heel wat segmenten verloren gaan voor beide categorieën van stromen, zowel de vaste als de omgeschakelde. Als reactie daarop zullen alle TCP-bronnen hun zend-tempo behoorlijk drastisch reduceren. Het ligt dus niet voor de hand dat de snelste omschakeltijd in termen van goodput ook de beste resultaten zal opleveren.

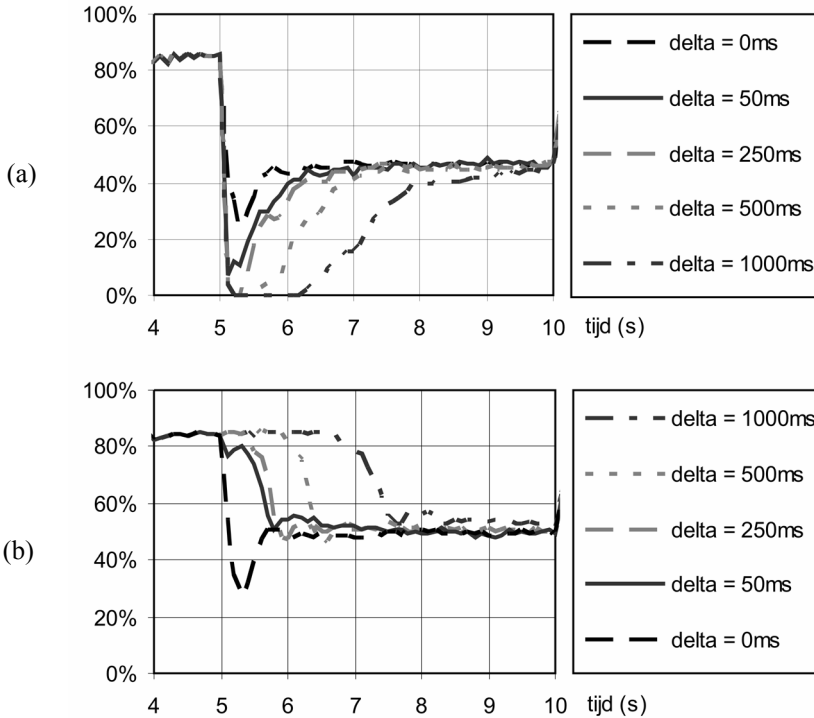


Figuur 5.5: Invloed van de omschakelsnelheid: simulatie-opstelling.

In Figuur 5.6 hebben we de evolutie van die goodput uitgezet voor zowel de vaste als omgeschakelde stromen in $[4s, 10s]$, voor het geval van paketschakelen zonder reservatie van bandbreedte voor herstel. Het beschreven effect van een zeer snelle omschakeltijd ($\delta=0$) is er duidelijk uit af te lezen. Merk ook op dat het een aantal seconden kan duren eer de TCP-effecten tot een nieuwe min of meer stabiele verdeling van de beschikbare bandbreedte hebben geleid.

Om in te schatten welke δ het meest voordelig is in termen van totale goodput, vergeleken we de verhouding $f(\delta) = \text{Good}(\delta) / \text{Good}(0)$, waarbij $\text{Good}(\delta)$ de goodput

is die bereikt wordt met omschakeltijd δ voor het geheel van omgeschakelde en vaste stromen, gedurende de eerste anderhalve seconde na het optreden van de linkfaling. Het beschouwde interval¹³ van $(5s, 6.5s]$ werd gekozen omdat dit de relevante periode is voor de kleinere δ -waarden. Voor de 150 beschouwde experimenten per δ -waarde tonen we in Figuur 5.7 het histogram en gefitte normale verdelingen van de $f(\delta)$ -waarden. Uit die gegevens blijkt dat alle waarden $\delta > 0$ door de band genomen een betere goodput realiseren dan ogenblikkelijk herstel ($\delta=0$): de gemiddelde waarde voor $f(\delta)$ ligt boven de 100%.

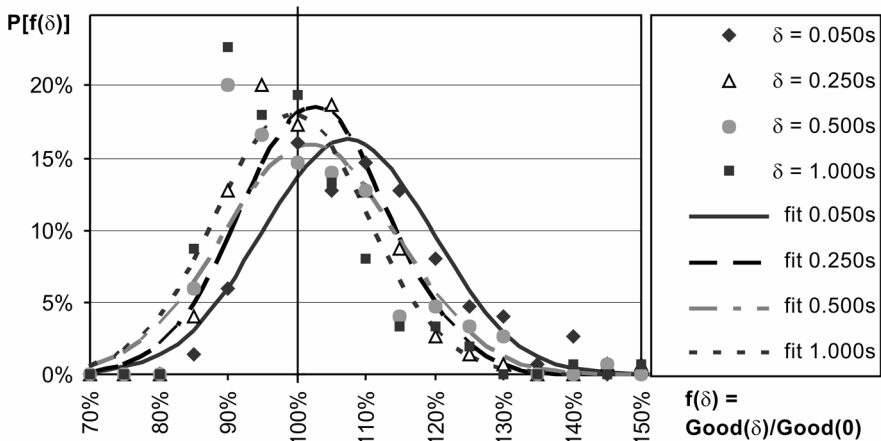


Figuur 5.6: Invloed van de omschakelsnelheid: goodput evolutie voor (a) de omgeschakelde stromen tussen Sx en B, (b) de vaste stromen tussen Fx en D.

Zo snel mogelijk herstellen blijkt dus niet noodzakelijk de beste oplossing: het introduceren van een kleine vertraging zorgt ervoor dat de omgeschakelde stromen hun verbruikte bandbreedte al getemperd hebben vooraleer ze de vaste stromen beginnen te vervoegen in knoop N5. We zien dan een geleidelijkere aanpassing van beide stromen aan de nieuwe netwerksituatie. Bepalen welke δ -waarde het beste

¹³ Voor de grenzen van een interval zal een rond haakje gebruikt worden wanneer het een open randpunt betreft en een recht wanneer het gesloten is: $(a, b]$ betekent dus het interval van a tot en met b , exclusief a , maar inclusief b .

netto resultaat oplevert is echter niet eenvoudig (zie Appendix I voor details). Dit zal afhangen van onder meer de linkbelasting (in het voorgestelde geval was die begrensd tot 90% door de bandbreedte op de toegangslinken, zodat de last op verbinding N5-N6 tot 180% kon oplopen), de RTT die de TCP-bronnen ervaren (grotere RTT impliceert tragere reactie op veranderingen in netwerktopologie) en het aantal TCP-stromen dat een link deelt (een groter aantal zal tot op zekere hoogte tot snellere stabilisatie leiden omdat het absolute gegevensverlies per stroom zal dalen).



Figuur 5.7: Invloed van de omschakelsnelheid: histogram van de waargenomen goodput in $[5s, 6.5s]$, relatief ten opzichte van die voor $\delta=0$.

Vanuit een pragmatisch oogpunt kunnen we voorzichtig besluiten dat snel herstel (orde van tiental milliseconden) niet zo slecht is. Deze conclusie zal wellicht des te meer van toepassing zijn op gevallen waar linken massaal veel TCP-stromen dragen (meer dan de 100 of 200 uit het bestudeerde geval), en/of lichter belast zijn. Zo bijvoorbeeld vonden we voor toegangslinken die slechts 60% van de bandbreedte in het ruggengraatnetwerk bedroegen, dat de δ -waarde met beste goodput resultaten duidelijk verschoof richting $\delta=0$.

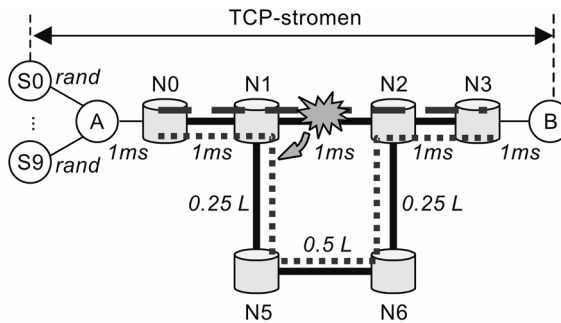
Wanneer we een *circuitgeschakeld* onderliggend netwerk beschouwen zal er geen interactie optreden tussen omgeschakelde en vaste stromen: voor elk apart zal er een circuit, i.e. een golflengte voorbehouden zijn. Het is duidelijk dat dynamische TCP-effecten, gelieerd aan het verlies van pakketten op de falende link en de veranderende RTT, wel zullen blijven optreden. In het circuitgeschakelde geval besluiten we op basis van simulaties—gelijklopend aan degene voor het pakketgeschakelde geval—that des te sneller de omschakeling naar het herstelpad gemaakt wordt, des te betere performance in termen van goodput. Voor een set van 140 simulaties (met verschillende waarden voor de willekeurig gegenereerde starttijden van TCP-bronnen en vertragingen van toeganslinken) voor elk van de δ -waarden uit $\{0, 5, 10, 20, 30, 40, 50, 250, 500, 1000ms\}$ vonden we dat in 94% van de gevallen

$\delta=0$ tot de beste goodput leidde. Het voordeel van circuitschakelen met betrekking tot het TCP-gedrag is duidelijk dat door reservatie van bandbreedte de interactie tussen vaste en omgeschakelde stromen vermeden wordt, waardoor TCP-stromen minder sterk verstoord worden (aangezien ze zich niet moeten aanpassen aan een gewijzigde hoeveelheid beschikbare bandbreedte). Bemerk dat men hiervoor wel een prijs betaalt in termen van te installeren netwerkcapaciteit [1].

5.4.4 Invloed van de veranderende RTT

In vorige sectie hebben we geanalyseerd wat de impact is van tragere of snellere herstel-acties, waarbij we ook aandacht besteedden aan mogelijke interacties tussen verschillende TCP-stromen. In deze sectie ligt de focus op het effect van de verandering in RTT die kan optreden wanneer het herstelpad niet dezelfde lengte heeft (in hops en/of vezellengte en dus propagatietijd). Om de invloed van die ΔRTT te isoleren zullen we hier de voornoemde interacties achterwege laten en voor het vergelijken van verschillende ΔRTT -waarden de herstelsnelheid δ constant houden.

Het beschouwde scenario wordt in Figuur 5.8 verduidelijkt. De totale propagatietijd van de overspanning N1-N5-N6 werd gelijkgesteld aan L en gevarieerd over $\{1\text{ms}, 2\text{ms}, 10\text{ms}, 50\text{ms}, 100\text{ms}\}$. De falende verbinding N1-N2 heeft een vertraging van 1ms , zodat de beschouwde variaties in RTT gegeven zijn door $\Delta RTT = 2 \cdot (L - 1\text{ms})$. Voor $L = 100\text{ms}$ betekent dit quasi een verdrievoudiging van de RTT die, gelet op de waarden aangegeven in Figuur 5.8 (en Tabel 5.1), voor de stromen tussen bronknoten S_x en eindknoop B gemiddeld 120ms bedraagt. Dit is een situatie die zou kunnen optreden wanneer lokale teruglissing als herstelmechanisme wordt gebruikt bij een falen dicht bij een begin- of eindpunt van een verbinding (in het bijzonder in een netwerk met lage connectiviteit): in dat geval zal een herstelpad inderdaad beduidend langer zijn dan het originele werkende pad, vermits dit zal bestaan uit bijna het volledige werkende pad, diezelfde route in tegenovergestelde richting en een disjunct alternatief pad tussen bron- en eindknoop (herinner u Figuur 5.3).



Figuur 5.8: Invloed van veranderende RTT: simulatie-opstelling.

Een voorbeeld van de goodput-evolutie na de faling, meer bepaald in $(5s, 10s]$, is voor het geval van een omschakeltijd van $\delta = 5\text{ms}$ gegeven in Figuur 5.9. Het bevestigt onze verwachtingen op basis van de karakteristieken van TCP: een toe-

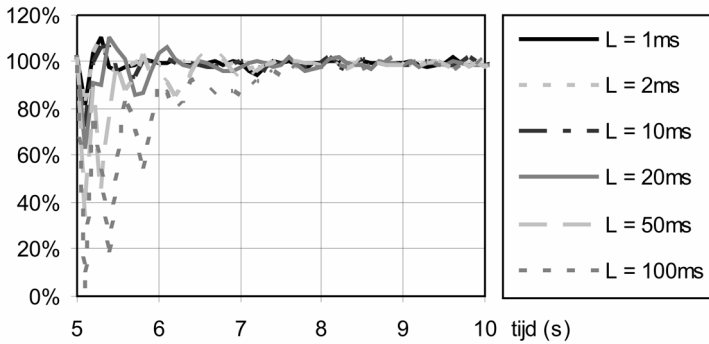
name in RTT, veroorzaakt door omschakeling naar het herstelpad, resulteert in een tijdelijke afname van de goodput. Het langere pad kan ervoor zorgen dat de timer voor heruitzenden (RTO) afgaat, zelfs al wordt snel omgeschakeld ($\delta = 5\text{ms}$). Bovendien impliceert een langere RTT ook een trager herstel van het door de faling veroorzaakte verlies aan pakketten: de snelheid waarmee het zend-venster in TCP verhoogd wordt is gerelateerd aan de RTT, zodat de stromen meer tijd nodig hebben om de beschikbare bandbreedte opnieuw op te vullen.

Om een min of meer rigoureuze vergelijking mogelijk te maken, hebben we, net als in vorige sectie, de goodput over het interval kort na de faling beschouwd, meer bepaald tijdens $[5s, 6.5s]$. Opnieuw hebben we het histogram van de goodput voor een gegeven L -waarde opgesteld, gedeeld door die voor $L = 1\text{ms}$. De vergelijking van de gevonden waarden, opnieuw over een set van 150 simulaties per waarde van L , staat samengevat in Tabel 5.3. Uit die numerieke gegevens kunnen we besluiten dat voor de beschouwde herstelsnelheid van $\delta = 5\text{ms}$ een langer herstelpad, dat leidt tot een toename van RTT met ongeveer 50%, wellicht te aanvaarden is: we vinden een resulterende goodput die niet meer dan 10% lager ligt dan wanneer de RTT onveranderd zou blijven.

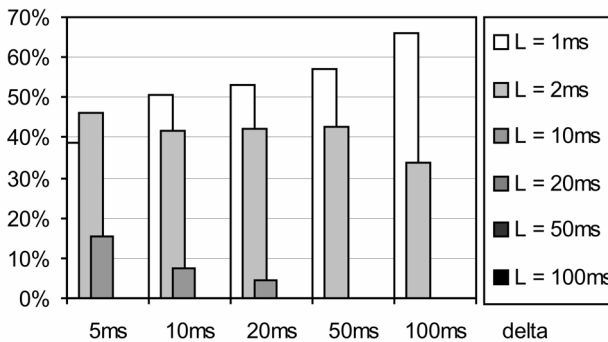
De vraag rijst uiteraard of we geen grotere verschillen in RTT kunnen accepteren voor tragere herstelmechanismen. Om daarop te antwoorden hebben we onze experimenten herhaald voor δ -waarden genomen uit $\{5\text{ms}, 10\text{ms}, 20\text{ms}, 100\text{ms}\}$. Wanneer de tijd δ tussen faling en omschakeling naar het herstelpad voldoende groot wordt, verwachten we dat de goodput tot nul terugvalt, ongeacht het RTT-verschil. Wat zoals eerder uitgelegd wél altijd zal afhangen van die RTT is de snelheid waarmee de bandbreedte opnieuw geleidelijk aan wordt gevuld (i.e. de aangroei van goodput richting 100%). De evolutie van de goodput (met curves gelijkaardig aan Figuur 5.9, zie Appendix I) heeft deze kwalitatieve voorspellingen bevestigd: voor grotere waarden van L vinden we een trager herstel van het goodput-niveau.

Tabel 5.3: Invloed van veranderende RTT: vergelijking van de goodput in het interval $[5s, 6.5s]$.

L	ΔRTT	<i>relatief verschil in goodput tegenover $L=1\text{ms}$</i>	<i>fractie van de resultaten waar L lagere goodput realiseert dan $L=1\text{ms}$</i>	<i>fractie van de resultaten waar L tot de hoogste goodput leidt</i>
1 ms	0.00%	0.00%	0.00%	38.67%
2 ms	+1.67%	+0.13%	42.00%	46.00%
10 ms	+15.00%	-1.03%	75.33%	15.33%
20 ms	+31.67%	-3.13%	94.00%	0.00%
50 ms	+81.67%	-13.66%	100.00%	0.00%
100 ms	+165.00%	-31.48%	100.00%	0.00%



Figuur 5.9: Invloed van veranderende RTT: goodput evolutie voor $\delta=5\text{ms}$.



Figuur 5.10: Invloed van veranderende RTT: fractie van de simulatieresultaten waar een bepaalde L-waarde tot de hoogste goodput leidde in $(5\text{s}, 6.5\text{s}]$.

Om een accurate vergelijking te maken tussen de onderscheiden veranderingen in RTT (evenredig zijnde met L), hebben we ook hier de behaalde goodput berekend over het interval¹⁴ $(5\text{s}, 6.5\text{s}]$. Figuur 5.10 toont aan dat wanneer de omschakeltijd δ verhoogt, de “beste” L -waarde verschuift naar lagere waarden. Dit wijst erop dat hoe langer we wachten om naar het herstelpad om te schakelen, met hoe meer zekerheid we kunnen stellen dat een kleinere verandering in RTT tot betere goodput zal leiden. Een analyse van de goodput gegevens bevestigt dit (zie Tabel I.5 in Appendix I), alsook Figuur 5.10. Figuur I.14 uit de appendix leert trouwens ook dat een verhoging van de padlengte (en dus RTT) met $L = 1\text{ms}$ ongeveer even nefast is voor de goodput als een verhoging van de omschakeltijd δ van dezelfde grootteorde (maar met behoud van RTT): vergelijk bijvoorbeeld de goodput-waarden voor $(L, \delta) = (10\text{ms}, 100\text{ms})$ en $(100\text{ms}, 10\text{ms})$. Dit suggereert dat het niet eenvoudig is te

¹⁴ Bemerk dat dit voor sommige situaties korter is dan de tijd nodig om opnieuw tot een stabiel goodput niveau te komen, maar een langer interval verandert de relatieve posities voor de verschillende L -waarden niet.

beslissen wat beter is: een snelle omschakeling maken naar een langer herstelpad (denken we bv. aan teruglussen), of eerder een tragere omschakeling naar een pad van vergelijkbare lengte (wat o.a. bij padprotectie zou kunnen gebeuren).

5.4.5 Invloed van terugschakelen naar werkend pad

De vorige analyses spitsten zich toe op wat er gebeurt wanneer een faling optreedt. Wanneer evenwel de faling wordt opgelost (bv. herstel van gebroken vezels), zal in het geval van omkerende herstelmechanismen het verkeer teruggeschakeld worden naar het originele pad. Hier zullen we de effecten van dit terugschakelen onderzoeken in twee gevallen, kortweg Pakket en Circuit genoemd, en hun prestatantie in termen van goodput vergelijken met een niet-omkerend mechanisme dat ook na het opheffen van de faling de stromen het herstelpad laat volgen. Het geval Pakket verwijst naar pakketschakelen, waar het samenvloeien van herstel- en werkend pad geen probleem vormt, terwijl het Circuit-geval een circuitschakelende laag onderstelt die niet toelaat dat in een schakelaar twee inkomende circuits worden samen genomen in één uitgaand circuit.

De beschouwde topologie is dezelfde als in voorgaande secties. Waar we voorheen het gebeuren tijdens de faling onderzochten, kijken we nu wat er in [10s,15s] gebeurt eenmaal de falende link hersteld is. De mogelijke acties die daar ondernomen worden zijn:

- **Pakket:** De routes die gevolgd worden door de pakketten worden teruggeschakeld naar het originele werkende pad op tijdstip $t = 10s$. In dit pakketgeschakelde scenario zal de resterende stroom pakketten, die nog onderweg zijn langs het (langere) herstelpad, in knoop N2 samenvloeien met pakketten die opnieuw de link N1-N2 volgen. Hetzelfde gebeurt in knoop N1 voor de ACK-pakketten die in de omgekeerde richting bewegen.
- **Circuit:** Hier is samenvloeien van de stroom langs het opnieuw werkende pad en de resterende pakketten langs het herstelpad niet mogelijk. Zodra het eerste pakket de gerepareerde link doorkruist en in knoop N2 aankomt, zullen pakketten die nog langs het herstelpad van N5 komen, verloren gaan. Hetzelfde geldt, mutatis mutandis, voor de ACK's in N1.
- **GeenActie:** In dit geval blijven pakketten na het opnieuw functioneren van link N1-N2 het herstelpad volgen.

De enige parameter die hier nog van belang is, is het verschil in padlengte (dus RTT). Net als in voorgaande sectie hebben we om het effect ervan te bestuderen L gevarieerd over {1ms, 2ms, 10ms, 20ms, 50ms, 100ms}.

Vooraleer de resultaten in meer detail te bekijken, kunnen we al een kwalitatieve voorspelling maken van de verschillen die zullen optreden:

- **Pakket:** Na de faling krijgt knoop N2 twee pakketstromen te verwerken, die samen dezelfde uitgaande link naar N3 moeten volgen. Wanneer L groot is betekent dit dat de duur van dit samenvloeien relatief lang kan worden (want er zijn dan nog veel pakketten onderweg langs het alternatieve pad, dat $L - 1ms$ langer is). In zo'n geval verwachten we dus dat verliezen kunnen optreden door gebrek aan bufferruimte in N2 om de tijdelijke overbelasting op te vangen. Een tweede effect, dat ook optreedt voor kleinere L -waarden, is het feit dat pakketten

niet langer in de correcte volgorde worden afgeleverd op de bestemming B. Wanneer een pakket langs het opnieuw werkende pad een voorganger (nog onderweg langs het herstelpad) voorbijgestoken heeft, zal TCP reageren als was er pakketverlies opgetreden en de verloren gewaande pakketten opnieuw verzenden. We verwachten dan ook voor lage L -waarden een zekere terugval in goodput.

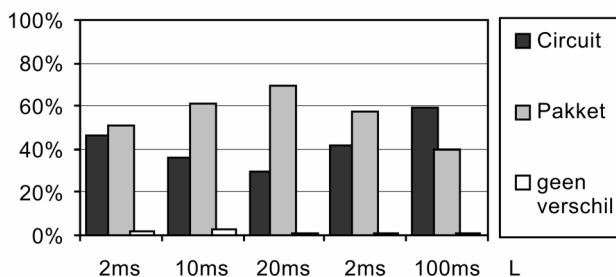
- **Circuit:** In dit geval zullen steeds pakketverliezen optreden zodra het herstelpad in lengte verschilt van het originele: eenmaal de faling opgelost is, gaan de pakketten die nog onderweg zijn op het herstelpad verloren. Hoe groter L , des te meer pakketten er verloren gaan. Opnieuw verwachten we daarom een terugval in goodput. Maar in tegenstelling tot het pakketgeschakelde concept zullen pakketten niet in een verkeerde volgorde bij de bestemming toekomen, waardoor de terugval mogelijks ietwat sneller kan worden goedge maakt.
- **GeenActie:** Vermits er hier niets verandert aan de netwerksituatie zoals de TCP-stromen die ervaren, zal de goodput hetzelfde niveau behouden.

Uit deze korte analyse blijkt dat GeenActie de beste goodput resultaten zal kunnen voorleggen, maar of bij een terugschakeling naar het originele pad een Circuit-geschakeld dan wel een Pakketgeschakeld concept beter is, valt moeilijk a priori in te schatten.

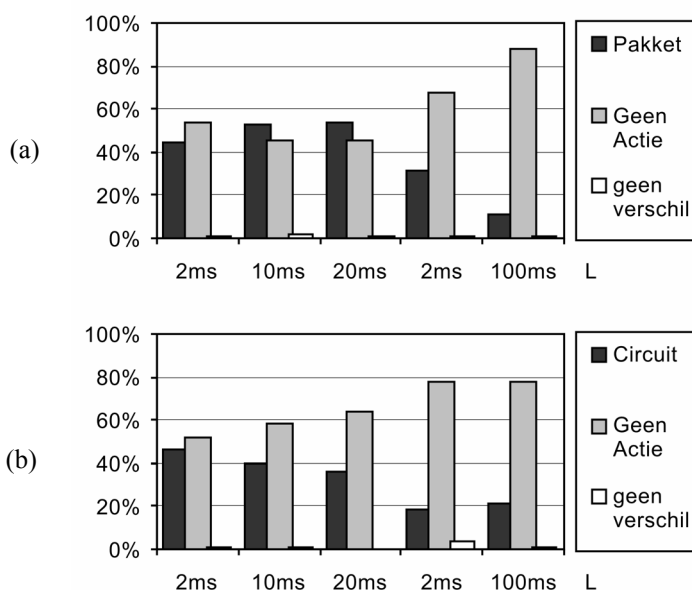
Een studie van de goodput-evolutie in het relevante interval na $t = 10s$ leert dat de drie beschouwde gevallen zich voor kleine RTT-verschillen ($L = 1ms, 2ms$) nauwelijks van elkaar onderscheiden in termen van TCP-goodput: voor het Pakketgeschakelde geval zijn er geen verliezen door het samenvloeien en slechts een uiterst beperkt aantal herordeningen van pakketten; voor Circuitschakelen is het aantal verloren pakketten miniem. Maar wanneer we L verhogen tot 10 à 20ms vinden we wel een duidelijk waarneembaar verschil tussen Pakket- en Circuitschakelen. Het Pakketgeschakeld geval vertoont een kleine terugval in goodput (ca. -5%) doordat een beperkt aantal TCP-connecties terugvalt naar zogenaamd snel heruitzenden/snel herstel (zie hoger, Sectie 5.2). Omdat niet pakketverlies, maar wel een herordening daarvoor aan de basis ligt, kan het TCP-mechanisme zeer snel het oorspronkelijke goodput-niveau herstellen. Voor circuitschakelen daarentegen treden zulke verliezen wel op en observeren we een grotere terugval in goodput. Voor voldoende grote waarden van L zullen zoals gezegd verliezen optreden zelfs in het pakketgeschakelde geval en er de goodput-degradatie, veroorzaakt door het uit volgorde afleveren van pakketten, versterken. Voor het bestudeerde geval $L = 100ms$ bijvoorbeeld is het gezamenlijke effect van beide fenomenen dat het globale goodput-niveau slechter is dan voor Circuitschakelen.

Om een meer nauwkeurige vergelijking te maken hebben we alweer de goodput over een bescheiden interval ($10s, 11.5s$] na de schakelactie beschouwd. Figuur 5.11 toont in hoeveel van de 150 beschouwde simulaties per L -waarde elk van de onderscheiden strategieën (Pakket, Circuit) de hoogste goodput realiseert. Op basis van deze goodput-gegevens de keuze te maken tussen een Pakket- en een Circuit-geschakelde aanpak is niet zo evident. De resultaten wijzen uit dat voor beperkte verschillen in RTT het Pakketgeschakelen beter presteert, terwijl voor grotere verschillen het Circuitschakelen hoger scoort. Het exacte punt waar Pakketgeschakelen slechter wordt dan Circuit-schakelen zal in het algemeen afhangen van de

hoeveelheid bufferruimte¹⁵ in de schakelknopen, in het bijzonder die waar herstelpad en origineel pad samenvloeien.



Figuur 5.11: Invloed van terugschakelen na opheffing van de faling: circuit- vs pakketschakelen. We zetten de fractie uit van de simulatieresultaten waar een bepaalde terugschakel-strategie de hoogste goodput opleverde in (10s,11.5s].



Figuur 5.12: Invloed van terugschakelen na opheffing van de faling: al dan niet terugschakelen? We zetten de fractie uit van de simulatieresultaten waarbij terugschakelen of het onveranderd laten van de routes (GeenActie) de hoogste goodput opleverde tijdens (10s,11.5s], in (a) voor pakketschakelen en (b) voor circuitschakelen.

¹⁵ In de simulaties werd een output-buffer gebruikt met SFQ-strategie, ter grootte van 50 pakketten.

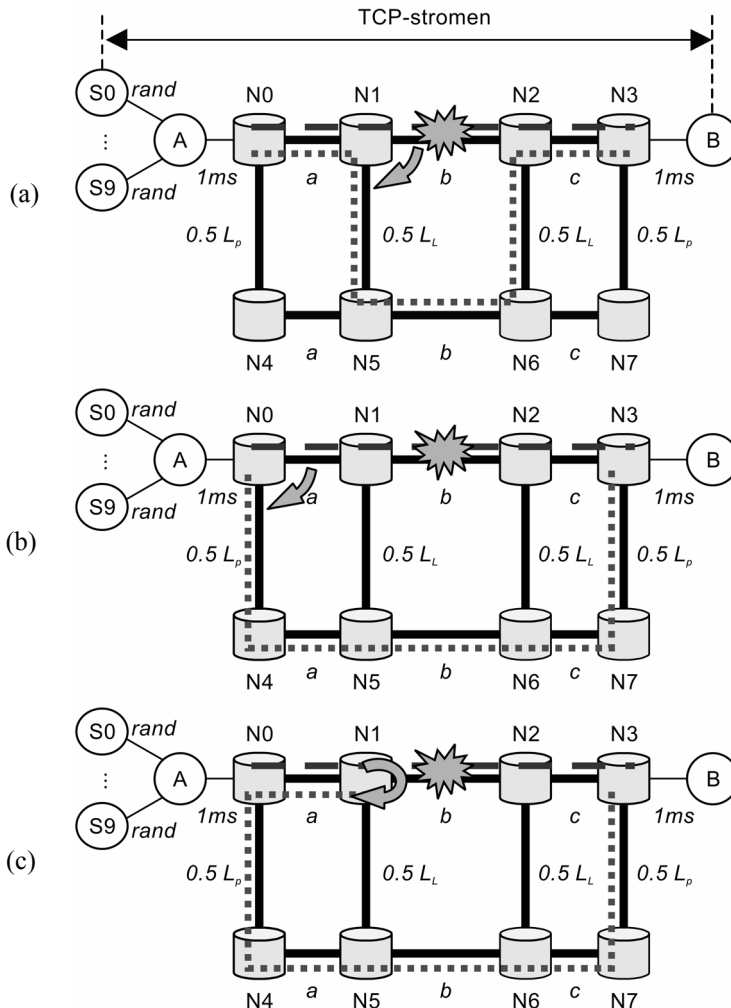
In Figuur 5.12 beogen we de terugval in goodput, veroorzaakt door de herstelacties, in te schatten. We hebben hiervoor voor Pakket- en Circuitschakelen apart gekeken in hoeveel van de simulatiegevallen het beter was Geen Actie te ondernemen bij herstel van het primaire pad. Voor kleine L -waarden is er geen overtuigende winnaar aan te duiden en vinden we dat er nauwelijks enige terugval in goodput is. Voor grote L is de trend duidelijk: GeenActie is ontegensprekelijk beter. We kunnen besluiten dat alleen wanneer het verschil in padlengte groot wordt er een noemenswaardige performantiedaling is, in vergelijking met het onaangeroerd laten van de gevolgde route.

5.4.6 Gevalsstudie: MPLS-protectiemechanismen

Met voorgaande simulatiestudies hebben we inzicht verworven in de invloeden van omschakelsnelheden en veranderende RTT op het gedrag van TCP-stromen. In de nu volgende gevalsstudie zullen we de vergaarde kennis toepassen op de belangrijkste protectiemechanismen in GMPLS voor een pakketschakelende technologie. De topologie die we zullen beschouwen is erg gelijkaardig aan voorgaande situaties en wordt geschetst in Figuur 5.13. We zullen in totaal vier scenario's beschouwen, overeenkomend met de drie protectiemechanismen uitgelegd in Sectie 5.3 en een variant van het teruglussen. De onderscheiden herstelacties, die gevallen na het falen van de link tussen N1 en N2 ondernomen zullen worden, zijn:

- **Lokale protectie (Lokaal):** Hier wordt lokale linkprotectie toegepast. Na het falen van de link, wat opnieuw na vijf seconden gebeurt, zal zoals aangegeven in het bovenste deel van Figuur 5.13 op $t = 5s + \delta$ een omschakeling gebeuren in knoop N1 naar het herstelpad dat L_L langer is dan het originele. Na het opheffen van de faling op $t=10s$ zullen pakketten door schakelaar N1 (resp. N2) pakketten opnieuw langs het originele pad gestuurd worden en zo in N2 (resp. N1) samenvloeien met de pakketten die nog langs het herstelpad vertrokken waren.
- **Padprotectie (Pad):** Wanneer we padprotectie toepassen is signalisatie nodig om knoop N0 van het falen van verbinding N1-N2 op de hoogte te brengen. Hierdoor zal de omschakeling, nu in N0 (resp. N3), op een later tijdstip gebeuren dan voor Lokale protectie, met name op $t = 5s + \delta + a$ (resp. $t = 5s + \delta + c$). Opnieuw zal het herstelpad langer zijn dan het originele, met nu een verhoging ter grootte van L_P en zal, na opheffen van de faling, verkeer langsheen het originele pad en het herstelpad samenvloeien, nu in knoop N3 (resp. N0). Ook die terugschakeling naar de oorspronkelijke route vereist signalisatie, maar dat de schakelactie later optreedt dan voor bv. lokale protectie is hier minder erg: verkeer dat even na $t=10s$ nog het herstelpad volgt gaat immers niet verloren, in tegenstelling tot verkeer dat we even na $t=5s$ nog langs het werkende pad sturen.
- **Lokale teruglussing (Teruglus):** Dit is het Lokaal Teruglussen zoals origineel voorgesteld in [227] en eerder beschreven in Sectie 5.3. Hier is, zoals in lokale protectie, geen signalisatie vereist en kan de omschakeling naar het herstelpad gebeuren op $t = 5s + \delta$, in een knoop die een eindpunt vormt van de falende link. Het herstelpad van N0 naar N3 is nu $L_P + 2 \cdot a$ langer dan het originele; de terugweg (gevolgd door de ACK's) is $L_P + 2 \cdot c$ langer. Nadat de faling is opgeheven, schakelen N1 en N2 opnieuw over op de originele route.

- **Variant op lokale teruglusing (TeruglusVar):** Het nadeel van lokaal teruglussen is dat het verkeer nodeeloos een rondje maakt tussen knopen N0 en N1 (tussen N2 en N3 voor de ACK's). In deze variant van teruglussen zullen we het originele teruglussen mengen met padprotectie: op $t = 5s + \delta$ schakelen we over naar het terugluspad en zendt schakelaar N1 een controlesignaal naar knoop N0 (ook N2 naar N3). Bij ontvangst ervan op $t = 5s + \delta + a$ schakelt N0 pakketten niet meer langs het onnodige lusje, maar onmiddellijk naar N4 zoals in padprotectie (idem voor N3 en N7).



Figuur 5.13: Gevalsstudie van MPLS-protectiemechanismen: schets van de simulatie-opstelling: (a) lokale protectie, (b) padprotectie, (c) lokale teruglusing. Merk op dat voor de eenvoud enkel de routes lopend van N0 naar N3 getoond zijn en niet die in omgekeerde richting.

Tabel 5.4: Gevalsstudie van MPLS protectiemechanismen: vier netwerkscenario's en de corresponderende parameters.

<i>Scenario</i>	L_L	L_P	a	b	c
I	1ms	3ms	2ms	1ms	48ms
II	5ms	15ms	10ms	1ms	40ms
III	10ms	30ms	20ms	1ms	30ms
IV	20ms	60ms	40ms	1ms	10ms

Tabel 5.5: Gevalsstudie van MPLS-protectiemechanismen: de tijdsparameters die bepalend zijn voor het TCP-gedrag.

<i>Protectie-mechanisme</i>	<i>Omschakeltijd</i>	ΔRTT
Lokaal	δ	$2 \cdot L_L$
Padprotectie	$\delta + a$	$2 \cdot L_P$
Teruglussen	δ	$2 \cdot L_P + 2 \cdot a + 2 \cdot c$

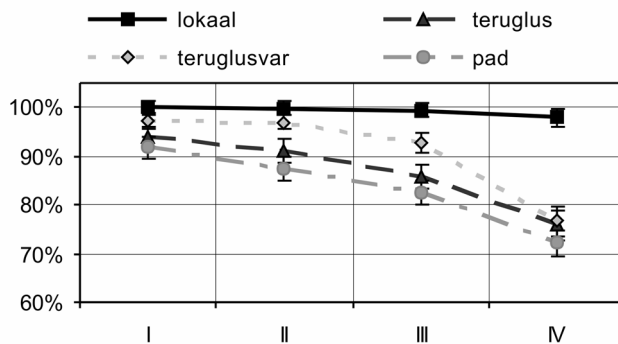
Voor elk van deze vier scenario's hebben we simulaties uitgevoerd voor vier sets topologie-parameters, zoals opgesomd in Tabel 5.4. Met $a + c = 50\text{ms}$ en $b = 1\text{ms}$ leidt dit tot een gemiddelde RTT van 216ms. De overige parameterwaarden gaven we dezelfde waarden als besproken in 5.4.2.

Uit de gebruikte parameterwaarden en het discours gevoerd in voorgaande secties, kunnen we meteen afleiden dat lokale protectie het best zal scoren in termen van goodput onmiddellijk na de falings: de omschakeltijd naar het herstelpad is het kortst (namelijk δ) en ook het verschil in RTT is het kleinst. De waarden van deze twee tijdsparameters voor de onderscheiden scenario's staan samengevat in Tabel 5.5. Voor padprotectie is de toename in RTT hoger dan lokale protectie. Daarenboven zal de omschakeling ook langer op zich laten wachten, met hoger pakketverlies tot gevolg: beide effecten zullen tot een reductie van goodput leiden kort na de falings. Lokale teruglissing vermijdt de relatief trage omschakeltijd, maar verhoogt de toename in RTT tegenover padprotectie. Aangezien we in 5.4.4 hebben vastgesteld dat een stijging van RTT even schadelijk is voor de goodput als een verhoging van de omschakeltijd van vergelijkbare grootteorde, verwachten we dat het netto voordeel van lokale teruglissing tegenover Padprotectie niet aanzienlijk zal zijn. De variant die de schakelsnelheid behoudt, maar de onnodige lus in de routing nadien verwijdert, zou hierdoor een hogere goodput kunnen realiseren. Maar de omschakeling van de route met lus naar die zonder heeft een gelijkaardig effect als het in 5.4.5 beschouwde omschakelen van herstel- naar werkend pad: een plotse afname van de RTT en het niet in volgorde afleveren van de pakketten. Of er ook een netto voordeel zal zijn is dus zeer de vraag.

De bovenstaande kwalitatieve analyse wordt bevestigd door de geobserveerde goodput-evolutie, waarvoor we verwijzen naar Appendix I. Een vergelijking van de

bereikte goodput over het interval (5s,7.5s], waarin de verschillende mechanismen duidelijk verschillen in goodput gedrag, is samengevat in Figuur 5.14. We kunnen eruit besluiten dat Lokale Protectie tot de beste goodput-resultaten leidt voor de beschouwde netwerkparameters. Omdat kiezen tussen Padprotectie en Lokale Teruglussing neerkomt op het inruilen van langere omschakeltijd voor grotere RTT-toename, vinden we dat ze niet al te veel van elkaar verschillen qua netto goodput. De variant van het lokaal teruglussen blijkt enkel duidelijk voordeel op te leveren wanneer de lengte van de lus niet al te groot is (scenario's I, II en III).

Voor het gedrag na de faling, met name wanneer terug overgeschakeld wordt naar het originele pad, is zoals uiteengezet in 5.4.5 enkel het verschil in RTT van belang. We vinden dan ook (zie Appendix I) dat lokale protectie het beste goodput-resultaat oplevert, gevolgd door padprotectie en pas dan lokale teruglussing. De variant van teruglussing valt in de periode na het opheffen van de faling uiteraard samen met padprotectie.



Figuur 5.14: Gevalsstudie van MPLS-protectiemechanismen: goodput in (5s,7.5s] voor elk van de verschillende scenario's.

5.5 Besluiten

In dit hoofdstuk zijn we opgeklommen tot het niveau van de gebruikers van een (al dan niet optisch pakketgeschakeld) netwerk, meer bepaald de transportlaag TCP, om er de invloed van plotse omschakelacties in de routing naar aanleiding van netwerkfalings te bestuderen. Voor zover we hebben kunnen vaststellen uit het doorsnuisteren van de literatuur, mag dit gerust als een originele bijdrage bestempeld worden¹⁶.

We hebben in onze studie de interactie beschouwd tussen omgeschakelde TCP-stromen en andere die al aanwezig waren in (delen van) de herstelroute en de impact van de omschakeltijd daarop geëvalueerd. Uit die analyse leidden we af dat het tot

¹⁶ Wat we ook wel kunnen afleiden uit het geciteerd worden in de rubriek "Scanning the literature" van Optical Networks Magazine in het nummer van november–december 2002.

het uiterste opdrijven van de herstelsnelheid (tot ruim onder de 50ms) niet noodzakelijk voordelig is: vanuit het standpunt van TCP-goodput kan een iets langere vertraging betere resultaten opleveren.

Verder hebben we ook het effect van het verschil in RTT tussen werkend en herstelpad bestudeerd voor verschillende omschakelsnelheden. Niet algeheel tot onze verrassing hebben we geconstateerd dat die verhoging in RTT een negatieve invloed heeft op de goodput en des te sterker voor grotere RTT-toename. Onze resultaten toonden ook dat geen duidelijk uitsluitel kan worden gegeven over de vraag of het beter is snel naar een duidelijk langer herstelpad om te schakelen dan trager naar een pad van ongeveer dezelfde lengte als het origineel werkende: de verschillen in goodput zijn klein en hangen af van de precieze tijds- en netwerkparameters.

Ook hebben we de reactie van TCP onderzocht op een terugschakeling naar het werkende pad eenmaal de faling opgelost is. Hierbij hebben we gewezen op het verschillende gedrag voor circuit- en pakketschakelende technologieën en hun uiteenlopende performantie in termen van goodput. Het feit dat in circuitschakelen stromen niet kunnen samenvloeien—wat wenselijk zou zijn waar herstelpad en een origineel werkend pad samenkomen—blijkt enkel in mindere prestaties te resulteren voor erg kleine verschillen in padlengte (klein ten opzichte van de aanwezige buffercapaciteit voor pakketschakelen). In de andere gevallen vinden we evenwel geen sterke discrepanties in bereikte goodput voor circuit- of pakketschakelen. Vanzelfsprekend heeft een circuitgeschakeld concept wel andere nadelen, zoals het feit dat meer extra bandbreedte moet worden voorzien om het netwerk resistent te maken tegen falingen. Voor een gedetailleerde studie daarvan verwijzen we naar [222].

We hebben het hoofdstuk afgerond met een gevalsstudie omtrent de invloed van herstelacties op TCP-gedrag voor vier protectiemechanismen zoals voorgesteld in het GMPLS-raamwerk. Die analyse leerde dat lokale protectie beter presteert in termen van TCP-goodput. Lokale teruglissing levert niet veel voordeel op tegenover padprotectie: een kleine winst aan goodput bij de omschakeling na de faling, maar licht verlies bij terugschakeling eenmaal de faling opgelost is. Het vermijden van de lus in het herstelpad bij lokale teruglissing levert in termen van goodput enkel voordeel op wanneer die lus betrekkelijk kort is in vergelijking met de totale padlengte. Opnieuw dienen we hier de kanttekening te maken dat deze rangschikking in termen van TCP-performantie genuanceerd moet worden met andere criteria, zoals in [1], waar we de duurte—in termen van vereiste bandbreedte—aantonen van lokale protectie in vergelijking met andere herstelmechanismen.

Hoofdstuk 6

Samenvattend besluit

Het opzet van dit werk was een zo breed mogelijk spectrum te belichten van de conceptuele vraagstukken die optreden bij het beschouwen van optisch pakketgeschakelde netwerken. We zijn begonnen met het scheppen van een referentiekader op basis van een literatuurstudie, om zo tot een overzicht te komen van wat een optisch pakketgeschakeld netwerk precies is, welke keuzes gemaakt kunnen worden op vlak van technologie, pakketformaat en architectuur van de schakelknoten.

Qua architecturen hebben we vastgesteld dat er typisch ingespeeld wordt op de geografische schaal van het netwerk. Zo wordt voor regionale, zogenaamde metro-netwerken in veel gevallen geopteerd voor een ringnetwerk. Onze bijdrage heeft twee varianten van schakelarchitecturen voor zo'n metro-omgeving met elkaar vergeleken en geëvalueerd in termen van de hoeveelheid apparatuur (componenten) die moet worden geïnstalleerd om te voldoen aan een gegeven vraag naar bandbreedte binnen een verzameling verkeersknooppunten, waar de optische pakketschakelaars gelokaliseerd zijn. We bespraken een zogenaamde passieve knooparchitectuur, die omwille van zijn zeer eenvoudige structuur op basis van vandaag commercieel beschikbare technologieën zou kunnen worden geconstrueerd. We hebben getoond dat die nauwelijks moet onderdoen voor een complexere actieve architectuur: in termen van te installeren zend- en ontvangstcapaciteit brengt de passieve architectuur geen meerkost met zich mee. Wel is het zo dat een actieve architectuur minder golflengten nodig heeft, vooral wanneer er voldoende vraag is naar bandbreedte voor verbindingen die binnen één metro-ring blijven en deze ruimtelijk niet al te veel overlappen. Dit laatste geval vereist wel dat het controleprotocol dat de toegang tot het metro-netwerk regelt zogenaamd ruimtelijk hergebruik (Eng. spatial reuse) toelaat. Daarnaast hebben we ook aandacht besteed aan het al dan niet voordelig zijn van bidirectionele ringen en een zogenaamd golflengtebandconcept.

Voor schakelarchitecturen voor vermaasde ruggengraatnetwerken hebben we aandacht besteed aan het schalen van schakelmatrices naar grote dimensies. Geïnspireerd op het baanbrekende werk van Clos hebben we geïllustreerd dat hiervoor een architectuur met meerdere stadia zijn nut kan bewijzen. Specifiek voor optische pakketschakelaars hebben we aangetoond dat het derde schakelstadium uit Clos' originele werk geëlimineerd kan worden en vervangen kan worden door golflengte-conversie. Voor een "broadcast-and-select"-schakelmatrix gebaseerd op SOA-technologie hebben we in een gevalstudie aangetoond dat dit een sterkere

reductie oplevert van het aantal vereiste schakelende SOA's dan een Clos-gebaseerde matrix met drie volwaardige schakelstadia.

We hebben onze studie vervolgd met onderzoek naar schakel-algoritmen voor een optische pakketschakelaar met een gedeelde recirculerende buffer opgebouwd uit vertraginglijnen. Voor synchroon, slot-gebaseerd schakelen van pakketten van vaste lengte hebben we de Balans-strategie ontworpen die de aanwezige bufferruimte zo efficiënt mogelijk benut. Die nieuwe strategie slaagt erin duidelijk lager pakket-verlies te bereiken voor dezelfde hoeveelheid vertraginglijnen.

Voor het schakelprobleem zonder differentiatie vonden we in de literatuur duchtig uitgekende oplossingen. In geval van pakketten van variabele lengte hebben we daarom bijzondere aandacht besteed aan het realiseren van kwaliteits-differentiatie. Waar voor vaste pakketlengte een eenvoudige aanduiding van prioriteiten in pakketten volstaat (zoals aangetoond met eigen werk), moeten voor variabele lengte bijkomende technieken gebruikt worden. De uit OBS bekende strategie van offset-gebaseerde differentiatie—waar men de pakkethoofding van prioritair verkeer eerder voorop stuurt—werd vergeleken met twee zelf bedachte strategieën: vooruitkijken en slot-gebaseerde controle. Hoewel deze laatste techniek er ook wel in slaagt differentiatie te realiseren, is ze beduidend minder uitgesproken dan de overige mechanismen. Het voordeel ervan is dat, door het hanteren van een slot-concept, de complexiteit van de controle-eenheid gereduceerd kan worden. Het vooruitkijken bleek op vlak van pakketverlies even goed of een tikkeltje beter te scoren dan offset-gebaseerde differentiatie, maar zorgt voor sterkere discriminatie van langere pakketten. De klassieke offset-differentiatie blijkt dus moeilijk te overtreffen.

Ander vernieuwend werk werd geleverd door het vergelijken van de performantie van twee mechanismen om in slot-gebaseerde schakelmatrices toch pakketten van variabele lengte te behandelen. Een trein-concept, waarbij alle slots die samen één pakket vormen steeds als een geheel worden geschakeld, vergeleken we met een zogenaamd wagon-concept dat voor elk van die slots onafhankelijk beslissingen neemt. Het trein-concept kwam hierbij als winnaar uit de bus: de performantie in termen van gegevensverlies ligt in de meerderheid van de gevallen lager. Enkel voor gemiddelde belasting en wanneer de pakketlengte gemeten in slots beperkt blijft, kan een wagon-concept voordeel opleveren doordat de bufferruimte dan efficiënter benut wordt. Wel blijkt een wagon-concept erin te slagen behoorlijk sterke kwaliteitsdifferentiatie te realiseren louter op basis van prioriteits-aanduidingen in de pakkethoofding, waar dit voor het trein-concept slechts in beperkte differentiatie resulteert.

Ook op netwerkniveau hebben we een onderzoeksbijdrage geleverd: we hebben een routeringsalgoritme bedacht om het pakketverlies (PLR) te minimaliseren. Daartoe gebruikten we een vrij nauwkeurige schatting van het pakketverlies in de optische pakketschakelaars met gedeelde recirculerende buffer. Routering op basis van de schatting van het verlies bleek uiteraard beduidend lagere PLR te bereiken dan pure kortste-pad routering. Belangrijker in deze is dat ook ten opzichte van een algoritme dat zich enkel baseert op het spreiden van de linkbelasting (Eng. load balancing) duidelijke voordelen geboekt werden.

Nog dichterbij de gebruiker die zijn verkeer over een (al dan niet optisch pakketgeschakeld) netwerk stuurt, situeert zich onze uitgebreide studie van TCP-gedrag. Hier hebben we de invloed bestudeerd van het omschakelen van de routes gevolgd door een aggregaat van zulke TCP-stromen (die het gros van het IP-verkeer vormen, dat tegenwoordig qua volume het meer traditionele verkeer zoals klassieke telefonie overtreft). In het bijzonder hebben we de route-veranderingen bestudeerd die optreden naar aanleiding van falingen. Het criterium dat we hanteerden was de zogenaamde goodput, zijnde de overdracht van unieke gegevens tijdens een bepaalde periode. We hebben de individuele invloeden van de omschakelsnelheid naar een herstelpad geanalyseerd, alsook de toename in lengte van de gevolgde route.

In een pakketgeschakelde netwerktechnologie kan interactie ontstaan tussen omgeschakelde TCP-stromen en andere die al aanwezig zijn langsheen (delen van) het herstelpad. Er blijkt dat in zo'n geval het tot een absoluut minimum herleiden van de omschakelsnelheid niet noodzakelijk beter is. Wanneer die interactie niet kan optreden, blijken de circuit- en pakketgeschakelde technologieën vergelijkbare goodput-prestaties te leveren. Enkel bij het terugschakelen naar het originele werkende pad na opheffen van de faling is er dan onderscheid, waarbij het afhangt van de verandering in padlengte of het pleit wordt beslecht in het voordeel van circuit- dan wel pakketschakelen. Voor zeer kleine verschillen lijkt pakketschakelen (met voldoende buffer) iets beter te scoren, terwijl circuit-schakelen doorgaans hogere goodput bereikt voor grotere afwijkingen tussen de lengte van werkend- en herstelpad.

Met een gevalsstudie voor de bekende GMPLS-protectiemechanismen toetsten we onze besluiten. De hoogste goodput-resultaten werden er behaald met lokale linkprotectie. Lokale teruglussing presteerde niet veel beter dan padprotectie: we vonden een klein voordeel qua goodput bij omschakeling naar het herstelpad, maar een licht verlies bij terugschakelen na het opheffen van de faling. De lus wegwerken die eigen is aan—jawel, u raadt het al—lokale teruglussing blijkt niet echt de moeite te lonen: enkel wanneer die lus betrekkelijk kort is met betrekking tot de totale padlengte, hebben we een verbetering van de goodput vastgesteld tegenover het laten bestaan van het nodeloze rondje.

Bij deze TCP-studies hebben we erop gewezen dat de evaluatie van circuit-versus pakketschakelen, en ook die van de GMPLS-protectiemechanismen, moet worden gebalanceerd met studies van de bandbreedtevereisten zoals uiteengezet in [1,222]: circuitschakelen vereist meer capaciteit om het netwerk robuust te maken tegen mogelijke falingen dan pakketschakelen. Lokale protectie blijkt daar de slechtst scorende strategie te zijn (tegenover de betere op vlak van TCP-goodput).

We mogen besluiten dat de collectie van deze studies heel wat vragen oplost die rijzen bij het nadenken over optisch pakketschakelen. Maar even duidelijk is het dat er minstens nog even zoveel niet beantwoord werden. Voor de knoop-architectuur hebben we bijvoorbeeld geen gedetailleerde analyse gemaakt van de technologische haalbaarheid en het nagaan van bijvoorbeeld fysische beperkingen van de voorgestelde Clos-architectuur op basis van SOA's. Maar daarvoor zullen we

in aansluitend onderzoek beroep doen op de expertise van collegae¹⁷ uit Athene. Wat de realisatie van de bestudeerde metro-architectuur betreft, zijn intussen in de literatuur al onderzoeksresultaten te vinden [93,118]. Omtrent de AWG-gebaseerde schakelmatrix, die we hier slechts uiterst beknopt bespraken, zal u ongetwijfeld veel meer vernemen in het nakende doctoraatsproefschrift van ir. J. Cheyns.

Meer geavanceerde routeringsalgoritmes dan degene die we in Hoofdstuk 4 bedachten zal u kunnen vinden in de doctoraatsthesis van lic. S. Vandenberghe, in het bijzonder voor IP/MPLS netwerken met kwaliteitsdifferentiatie. Voor de routing in meerlaagse netwerken en in het bijzonder het zogenaamde “grooming”-probleem verwijzen we u naar de recente proefschriften [229,230].

Een aspect dat we in dit proefschrift bewust niet hebben behandeld is het netwerkbeheer. Maar daarin onderscheidt een OPS-netwerk zich niet wezenlijk van actuele netwerktechnologieën, waaraan in onze vakgroep al meer dan één volledig doctoraat werd gewijd [231,232]. Ook de controle-aspecten eigen aan optische netwerken genoten binnen INTEC onderzoeks aandacht (bv. het vierde hoofdstuk in [230]).

¹⁷ Met wie we trouwens al een uitbreiding van de hier voorgestelde resultaten voorbereid hebben [11].

Bibliografie

- [1] D. Colle, P. Van Heuven, C. Develder, S. Van den Berghe, I. Lievens, M. Pickavet, P. Demeester, "MPLS recovery mechanisms for IP-over-WDM networks", *Photonic Network Commun. (PNET)*, Kluwer Academic Publishers, vol. 3, no. 1/2, Jan. 2001, pp. 23–40.
- [2] S. De Maesschalck, D. Colle, A. Groebbens, C. Develder, I. Lievens, P. Lagasse, M. Pickavet, P. Demeester, F. Saluta, M. Quagliatti, "Intelligent optical networking for multilayer survivability", *IEEE Commun. Mag.*, vol. 40, no. 1, Jan. 2002, pp. 42–49.
- [3] D. Colle, S. Demaesschalck, C. Develder, P. Van Heuven, A. Groebbens, J. Cheyns, I. Lievens, M. Pickavet, P. Lagasse, P. Demeester, "Data-centric optical networks and their survivability", *IEEE J. Selected Areas in Commun.*, vol. 20, no. 1, Jan. 2002, pp. 6–20.
- [4] C. Develder, D. Colle, S. Demaesschalck, M. Pickavet, P. Demeester, "Influence of GMPLS recovery mechanisms on TCP performance", *Photonic Network Commun. (PNET)*, vol. 4, no. 3/4, Jul.-Dec. 2002, pp. 321–345.
- [5] J. Cheyns, C. Develder, E. Van Breusegem, E. Baert, A. Ackaert, M. Pickavet, P. Demeester, "Routing in an AWG based optical packet switch", *Phot. Netw. Commun. (PNET)*, vol. 5, no. 1, Jan. 2003, pp. 69–80.
- [6] L. Dittmann (ed.), C. Develder, D. Chiaroni, F. Neri, F. Callegati, W. Koerber, A. Stavdas, M. Renaud, A. Rafel, J. Solé-Pareta, W. Cerroni, N. Leligou, L. Dembeck, B. Mortensen, M. Pickavet, N. Le Sauze, M. Mahony, B. Berde, G. Eilenberger, "The European IST project DAVID: a viable approach towards optical packet switching", *IEEE J. Selected Areas in Commun.*, vol. 21, no. 7, Sep. 2003, pp. 1026–1040.
- [7] E. Baert, C. Develder, M. Pickavet, P. Demeester, "Routing strategies to minimize packet loss in an optical packet switched network with recirculating FDL buffers", zal verschijnen in *Phot. Netw. Commun. (PNET)*, vol 7, no. 2, Mar. 2004.
- [8] C. Develder, B. Van Houdt, C. Blondia, M. Pickavet, P. Demeester, "Analytical MMAP-based bounds for packet loss in optical packet switching with recirculating FDL buffers", submitted to *PNET*.
- [9] J. Cheyns, C. Develder, E. Van Breusegem, E. Baert, D. Colle, M. Pickavet, P. Demeester, "Clos lives on in Optical Packet Switching", aanvaard voor publicatie in *IEEE Commun. Mag.*
- [10] F. Callegati, C. Develder, W. Cerroni, M. Pickavet, P. Demeester,

- “Scheduling algorithms for a slotted packet switch with either fixed or variable length packets”, aanvaard voor publicatie in Photonic Network Commun. (PNET), Kluwer Academic Publishers.
- [11] C. Matrakidis, C. Develder, A. Stavdas, M. Pickavet, P. Demeester, “Multistage switches for optical packet switching based on SOA technology”, to be submitted to Optical Networks Magazine
 - [12] P. Demeester, D. Colle, S. Demaesschalck, C. Develder, M. Pickavet, P. Van Heuven, “Resilience in IP over WDM based multilayer networks”, (Invited) The 26st European Conference on Optical Communication (ECOC 2000), Munich, Germany, 3-7 Sep. 2000.
 - [13] D. Colle, C. Develder, P. Van Heuven, M. Pickavet, P. Demeester, L. Raptis, G. Chatziliass, C. Mas, Y.I. Manolossos, J. Comellas, A. Rafel, J. Prat, J. Solé-Pareta, J. Moyano, S. Brunazzi, S. Rotolo, R. Stankiewicz, A. Gladish, “Recovery techniques for IP-over-WDM networks”, (Invited) IP over DWDM Conf., Paris, France, 27-30 Nov. 2000.
 - [14] D. Colle, P. Van Heuven, A. Groebbens, C. Develder, M. Pickavet, P. Demeester, “Dimensioning reliable IP-over-WDM networks”, IP over DWDM Conf., Paris, France, 27-30 Nov. 2000.
 - [15] P. Demeester, D. Colle, S. Demaesschalck, C. Develder, M. Pickavet, P. Van Heuven, “Resilience in IP over WDM based multilayer networks”, (Invited) Optical Networks 2000, Göteborg, Sweden, 7 Dec. 2000.
 - [16] C. Develder, Didier Colle, Pim Van Heuven, Steven Van den Berghe, Mario Pickavet, Piet Demeester, “Influence of recovery time on TCP behaviour”, MPLS World Congress, Paris, France, 6-9 Feb. 2001.
 - [17] D. Colle, C. Develder, P. Van Heuven, S. Demaesschalck, A. Groebbens, M. Pickavet, P. Demeester, “Resilience in IP-over-WDM networks”, (Invited) Proc. 5th Working Conf. on Optical Netw. Design and Modelling (ONDM 2001), Vienna, Austria, 5-7 Feb. 2001.
 - [18] A. Groebbens, D. Colle, C. Develder, S. De Maesschalck, M. Pickavet, P. Demeester, “Use of backup trees to improve resource efficiency of MPLambdaS protection mechanisms”, Proc. 3rd Intl. Workshop on Design of Reliable Commun. Netw. (DRCN 2001), Budapest, Hungary, 7-10 Oct. 2001, pp. 152–159.
 - [19] D. Colle, S. De Maesschalck, C. Develder, A. Groebbens, M. Pickavet, P. Demeester, “Design and dimensioning issues in survivable data-centric optical networks”, (Invited) Proc. 1st COST270 Workshop on Reliability of Optical Netw., Systems and Components, Dubendorf, Switzerland, 12-13 Dec. 2001.
 - [20] C. Develder, M. Pickavet, P. Demeester, “Assessment of packet loss for an optical packet router with recirculating buffer”, Proc. 6th IFIP Working Conf. on Optical Netw. Design and Modelling (ONDM2002), Torino, Italy, 4-6 Feb. 2002, pp. 247–261.
 - [21] M. Pickavet, A. Ackaert, E. Baert, J. Cheyns, D. Colle, S. De Maesschalck, P. Demeester, C. Develder, A. Groebbens, I. Lievens, E. Van Breusegem, S. Verbrugge, Q. Yan, “Design of communication networks using heuristics”, (Invited) Book of abstracts of “Arbeitskreis Mathematik in Forschung und

- Praxis", 23rd Symp. on Mathematik in der Telekommunikation, Duisburg, Germany, 18-19 Mar. 2002.
- [22] M. Pickavet, C. Develder, E. Baert, P. Demeester, "A.I. Techniques for planning telecommunication networks", Proc. Intl. Conf. on Artificial Intelligence (IC-AI'02), Las Vegas, NV, 24-27 Jun. 2002.
- [23] C. Develder, J. Cheyns, E. Van Breusegem, E. Baert, A. Ackaert, M. Pickavet, P. Demeester, "Node architectures for optical packet and burst switching", Proc. Int. Topical Meeting on Photonics in Switching (PS2002), (invited) paper PS.WeA1, Cheju Island, Korea, 21-25 Jul. 2002, pp. 104-106.
- [24] C. Develder, M. Pickavet, P. Demeester, "Strategies for an FDL based feed-back buffer for an optical packet switch with QoS differentiation", Proc. Conf. on Optical Internet (COIN2002), paper COIN.TuD1, Cheju Island, Korea, 21-25 Jul. 2002, pp. 114-116.
- [25] C. Develder, M. Pickavet, P. Demeester, "Choosing an appropriate buffer strategy for an optical packet switch with a feed-back FDL buffer", Proc. 28th European Conference on Optical Communication (ECOC2002), Copenhagen, Denmark, 8-12 Sept. 2002.
- [26] C. Develder, J. Cheyns, M. Pickavet, P. Demeester, "Multistage architectures for optical packet switching using SOA-based broadcast-and-select switches", Tech. Digest Optical Fibre Conference (OFC 2003), paper FS4, Atlanta, GA, USA, 23-28 Feb. 2003, pp. 794-795.
- [27] J. Cheyns, E. Van Breusegem, C. Develder, A. Ackaert, M. Pickavet, P. Demeester, "Performance improvement of an internally-blocking optical packet/ burst switch", Proc. Intl. Conf. on Commun. (ICC2003), Anchorage, AK, 11-15 May 2003, vol. 2, pp. 1304-1308.
- [28] D. Colle, J. Cheyns, C. Develder, E. Van Breusegem, A. Ackaert, M. Pickavet, P. Demeester, B. Feng, H. Lønsethagen, E. Zouganeli, J. Fernández-Palacios, J.F. Lobo, "GMPLS extensions for supporting advanced optical networking technologies", Proc. Intl. Conf. on Transparent Optical Networks (ICTON 2003), Warschau, Poland, 29 Jun. - 3 Jul. 2003.
- [29] C. Develder, M. Pickavet, P. Demeester, "On trains and wagons: switching variable length packets in a slotted OPS network", Proc. Conference on Optical Internet / Australian Conference on Optical Fibre Technology (COIN/ACOFT 2003), paper TuA2-6, Melbourne, Australia, 13-16 Jul. 2003, pp. 233-236.
- [30] C. Develder, R. Van Caenegem, E. Baert, M. Pickavet, P. Demeester, "Active versus passive OPS architectures for metro rings: a network dimensioning point of view", Proc. 29th European Conf. on Optical Commun. (ECOC 2003), paper We.1.4.4, Rimini, Italy, 21-25 Sep. 2003, vol. 3, pp. 376-377.
- [31] C. Develder, J. Cheyns, E. Van Breusegem, E. Baert, D. Colle, M. Pickavet, P. Demeester, "Architectures for optical packet and burst switches" (invited), Proc. 29th European Conf. on Optical Commun. (ECOC 2003), paper Mo.4.4.1, Rimini, Italy, 21-25 Sep. 2003, vol. 1, pp. 100-103.
- [32] C. Develder, J. Cheyns, M. Pickavet, P. Demeester, "Service differentiation

- mechanisms for variable length packets in an optical switch with recirculating FDL buffer”, Tech. Digest Photonics in Switching (PS2003), paper PS.Mo.C8, Versailles, France, 28 Sep.-2 Oct. 2003, pp. 92–94.
- [33] R. Van Caenegem, C. Develder, E. Baert, D. Colle, M. Pickavet, P. Demeester, “Architectures for OPS metro rings: comparing active versus passive nodes / A dimensioning point of view”, aanvaard voor 8th IFIP Working Conf. on Optical Network Design and Modelling (ONDM 2004).
 - [34] J. Cheyns, E. Van Breusegem, C. Develder, D. Colle, M. Pickavet, P. Demeester, “Evaluating cost functions for OPS node architectures / A packaging driven approach”, aanvaard voor 8th IFIP Working Conf. on Optical Network Design and Modelling (ONDM 2004).
 - [35] D. Colle, C. Develder, S. De Maesschalck, A. Groebbens, M. Pickavet, P. Demeester, “Design and planning of reliable communication networks”, 1st FTW PHD Symposium, Gent, Belgium, 5 Dec. 2000.
 - [36] C. Develder, J. Cheyns, E. Van Breusegem, E. Baert, M. Pickavet, P. Demeester, “Optical Packet/Burst Switching”, 2nd FTW PHD Symposium, paper 100 (Proceedings available on CD-Rom), Interactive poster session, , Gent, Belgium, 11 Dec. 2001.
 - [37] C. Develder, E. Baert, M. Pickavet, P. Demeester, “Optical Packet Switched Networks with Recirculating FDL Buffers”, 3rd FTW PHD Symposium, Gent, Belgium, 11 Dec. 2002.
 - [38] A. Groebbens, C. Develder, D. Colle, S. Demaesschalck, B. Lannoo, M. Pickavet, P. Demeester, “Optimization of backup tree structures, reducing spare capacity in optical networks, while retaining protection speed”, 3rd FTW PhD Symposium, Gent, Belgium, 11 Dec. 2002.
 - [39] R. Inkret, et al., “Advanced Infrastructure for photonic networks / Extended final report of Cost 266 action”, 2003.
 - [40] R.W. Wagner, “Communications Network Trends and Opportunities”, Proc. Conf. On the Optical Internet/ Australian Conf. On Optical Fibre Technology (COIN/ACOFT 2003), Melbourne, Australia, 13-16 Jul. 2003.
 - [41] K.G. Coffman, A.M. Odlyzko, “Growth of the internet”, in: I.P. Kaminow, T. Li (eds.), “Optical fiber telecommunications IV B: systems and impairments”, Academic Press, 4th ed., April 2002, pp. 17–56.
 - [42] C.K. Kao, G.A. Hockham, “Dielectric-fibre surface waveguides for optical frequencies”, Proc. IEE, vol. 113, no.7, Jul. 1966, pp. 1151–1158.
 - [43] ITU Recommendation G.652, “Characteristics of a single mode optical fibre cable”, 1997.
 - [44] J. Kim, O. Boyraz, M.N. Islam, “150+ channel ultra-DWDM source with $N \times 10$ GHz spacing utilizing longitudinal mode slicing of supercontinuum”, Tech. Digest Optical Fibre Commun. Conf. and Exhibition (OFC 2000), Baltimore, MD, 7-9 Mar. 2000, vol. 3, pp. 5–7.
 - [45] Y. Hida, et al., “400-channel 25-GHz spacing arrayed-waveguide grating covering a full range of C- and L-bands”, Tech. Digest Optical Fibre Commun. Conf. and Exhibition (OFC 2001), Anaheim, CA, 17-22 Mar. 2001, paper WB2.
 - [46] A. Neukermans, R. Ramaswami, “MEMS technology for optical networking

- applications", *IEEE Commun. Mag.*, vol. 39, no. 1, Jan. 2001, pp. 62–69.
- [47] L. Garret, et al., "The MONET New Jersey Network Demonstration", *IEEE J. on Selected Areas in Commun.*, vol. 16, no. 7, Sep. 1998, pp. 1199–1219.
 - [48] K. Shimano, C. Cavazzoni, A. D'Allessandro, A. Manzalini, "Architecture and functional requirements of control planes for Automatic Switched Optical Networks: experience of the IST project LION", *Proc. Workshop on High Performance Switching and Routing (HPSR 2002)*, Kobe, Japan, 26–29 May 2002.
 - [49] G. Bernstein, E. Mannie, V. Sharma, "Framework for MPLS-Based control of optical SDH/SONET networks", *IEEE Network*, vol. 15, no. 4, Jul./Aug. 2001, pp. 20–26.
 - [50] K. Sato, N. Yamanaka, Y. Takigawa, M. Koga, S. Okamoto, K. Shiimoto, E. Oki, W. Imajuku, "GMPLS-based photonic multilayer router (Hikari router) architecture: an overview of traffic engineering and signaling technology", *IEEE Commun. Mag.*, vol. 40, no. 3, Mar. 2002, pp. 96–101.
 - [51] L. Haché, L. Li, "Unified control infrastructure for carrier network evolution", *IEEE Commun. Mag.*, vol. 38, no. 11, Nov. 2000, pp. 74–77.
 - [52] S. Yao, B. Mukherjee, S. Dixit, "Advances in photonic packet switching: an overview", *IEEE Commun. Mag.*, vol. 38, no. 2, Feb. 2000, pp. 84–94.
 - [53] D.K. Hunter, I. Andonovic, "Approaches to Optical Internet Packet Switching", *IEEE Commun. Mag.*, Sep. 2000, pp. 116–122.
 - [54] C. Qiao, M. Yoo, "Optical Burst Switching (OBS) – a New Paradigm for an Optical Internet", *J. High Speed Networks*, vol. 8, no. 1, Jan. 1999.
 - [55] E. Van Breusegem, J. Cheyns, D. Colle, M. Pickavet, P. Demeester, "Overspill Routing In Optical Networks: a new architecture for future proof IP-over-WDM networks", *Proc. Conf. on Optical Netw. and Communications (OptiComm 2003)*, Dallas, TX, 15–48 Oct. 2003.
 - [56] D. Chiaroni, et al., "New 10 Gbit/s 3R NRZ optical regenerative interface based on semiconductor optical amplifiers for all-optical networks", *Proc. Int. Conf. Integrated Optics and Optical Fiber Commun. / European Conf. on Optical Commun. (IOOC/ECOC 97)*, Edinburgh, UK, 22–25 Sept. 1997, pp. 41–45.
 - [57] B. Lavigne, et. al., "Full validation of an optical 3R regenerator at 20 Gbit/s", *Tech. Digest Optical Fibre Commun. Conf. and Exhibition (OFC 2000)*, Baltimore, MD, 7–9 Mar. 2000, vol. 3, pp. 93–95.
 - [58] B. Sartorius, "3R all-optical signal regeneration", *Proc. European Conf. on Optical Commun. (ECOC 2001)*, Amsterdam, The Netherlands, 30 Sep. – 4 Oct. 2001, pp. 98–125.
 - [59] D. Chiaroni, B. Lavigne, A. Jourdan, L. Hamon, C. Janz, M. Renaud, "Feasibility assessment of a synchronization interface for photonic packet-switching systems", *Proc. Int. Conf. Integrated Optics and Optical Fiber Commun. / European Conf. on Optical Commun. (IOOC/ECOC 97)*, Edinburgh, UK, 22–25 Sept. 1997, vol. 3, pp. 148–151.
 - [60] M. Burzio, P. Cinato, R. Finotti, P. Gambini, M. Puleo, E. Vezzoni, L. Zucchelli, "Optical cell synchronization in an ATM optical switch", *Proc. European Conf. on Optical Commun (ECOC 94)*, Firenze, Italy, 25–29 Sep.

- 1994, pp. 581–584.
- [61] T. Sakamoto, et al., “Optical packet synchronization for optical packet switches”, Proc. Conf. on Optical Internet (COIN2002), paper COIN.TuA3, Cheju Island, Korea, 21–25 Jul. 2002, pp. 78–80
 - [62] M. C. Cardakli, S. Lee, A. E. Willner, V. Grubsky, D. Starodubov, J. Feinberg, “Reconfigurable optical packet header recognition and routing using time-to-wavelength mapping and tunable fiber Bragg gratings for correlation decoding”, IEEE Photon. Technol. Lett., vol. 12, no. 5, May 2000, pp. 552–554.
 - [63] T. Fjelde, et al., “Novel scheme for simple label-swapping employing xor logic in an integrated interferometric wavelength converter,” IEEE Photon. Technol. Lett., vol. 13, no. 7, Jul. 2001, pp. 750–752.
 - [64] K. Kitayama, K. Onohara, M. Murata, “Capability of optical code-based MPLS (OC-MPLS)”, Proc. 6th IFIP Working Conf. on Optical Netw. Design and Modelling (ONDM2002), Torino, Italy, 4–6 Feb. 2002, pp. 263–278.
 - [65] H.J.S. Dorren, et al., “Optical packet switching and buffering by using all-optical signal processing methods”, IEEE J. Lightwave Techn., vol. 21, no. 1, Jan. 2003, pp. 2–12.
 - [66] M.T. Hill, H. De Waardt, G.D. Khoe, H.J.S. Dorren, “All-optical flip-flop based on coupled laser diodes”, IEEE J. Quantum Electron., vol. 37, Mar. 2001, pp. 157–159.
 - [67] C. Liu, Z. Dutton, C.H. Behroozi, L.V. Hau, “Observation of coherent optical information storage in an atomic medium using halted light pulses”, Nature, vol. 409, Jan. 2001, pp. 490–493 .
 - [68] D. F. Phillips, A. Fleischhauer, A. Mair, R.L. Walsworth, “Storage of Light in Atomic Vapor”, Physical Review Letters, vol. 86, Jan. 2001, pp 783–786.
 - [69] A. V. Turukhin, et. al., “Observation of ultraslow and stored light pulses in a solid”, Physical Review Letters, vol. 88, Dec. 2001.
 - [70] M.S. Bigelow, N.N. Lepeshkin, R.W. Boyd, “Observation of Ultraslow Light Propagation in a Ruby Crystal at Room Temperature”, Physical Review Letters, vol. 90, March 2003.
 - [71] A. Bianco, G. Galante, E. Leonardi, F. Neri, M. Rundo, “Access Control Protocols for Interconnected WDM Rings in the DAVID Metro Network”, Proc. 2001 Tyrrhenian Intern. Workshop on Digital Communications (IWDC 2001), Taormina, Italy, Sept. 2001.
 - [72] S. De Maesschalck, P. Demeester, “A functional model for the SONATA switchless optical network”, Proc. 4th IFIP Working Conf. on Optical Netw. Design and Modeling (ONDM 2000), Athens, Greece, 7–8 Feb. 2000, pp. 151–162.
 - [73] A. Bianco, E. Leonardi, M. Mellia, F. Neri, “Network controller procedures in SONATA, a large scale all-optical WDM network”, Proc. 4th IFIP Working Conf. on Optical Netw. Design and Modeling (ONDM 2000), Athens, Greece, 7–8 Feb. 2000, pp. 247–264.
 - [74] M.Pousa, C. Steenbergen, T. Almeida, J. Wellen, B. Vermeulen, “Harmonics – Hybrid Access Reconfigurable Multi-wavelength Optical Networks for IP-based Communication Services”, Proc. 5th World Multiconf. on Systemics,

- Cybernetics and Informatics (SCI 2001), Orlando, FL, 22-25 Jul. 2001.
- [75] G. Kramer, G. Pesavento, "Ethernet Passive Optical Network (EPON): building a next-generation optical access network", *IEEE Commun. Mag.*, vol. 40, no. 2, Feb. 2002, pp. 66–73.
 - [76] P. Arijs, "Planning of ring-based telecommunication networks", PhD. Thesis, Ghent University, 2000-2001.
 - [77] A. Carena, V. Ferrero, R. Gaudino, V. De Feo, F. Neri, P. Poggiolini, "RINGO: A demonstrator of WDM optical packet network on a ring topology", *Proc. 6th IFIP Working Conf. on Optical Netw. Design and Modeling (ONDM 2002)*, Torino, Italy, 4-6 Feb. 2002, pp. 183–197.
 - [78] D. Dey, A. van Bochove, A. Koonen, D. Geuzebroek, M. Salvador, "FLAMINGO: a packet-switched IP-over-WDM all-optical MAN", *Proc. European Conf. on Optical Commun. (ECOC 2001)*, Amsterdam, The Netherlands, 30 Sep. – 4 Oct. 2001, pp. 480–481.
 - [79] K. Shrikhande, et al., "HORNET: A packet-over-WDM multiple access metropolitan area ring network", *IEEE J. Selected Areas in Commun.*, vol. 18, no. 10, Oct. 2000, pp. 2004–2016.
 - [80] M.A. Marsan, A. Bianco, E. Leonardi, M. Meo, F. Neri, "MAC-protocols and fairness control in WDM multi-rings with tunable transmitters and fixed receivers", *IEEE J. Selected Areas in Commun.*, vol. 14, no. 6, June 1996, pp. 1230–1244.
 - [81] I. Cidon, Y. Ofek, "MetaRing – A full-duplex ring with fairness and spatial reuse", *IEEE Trans. on Commun.*, vol. 40, no. 1, Jan. 1993, pp. 110–120.
 - [82] A. Marsan, A. Bianco, E. Leonardi, A. Morabito, F. Neri, "All-optical WDM multi-rings with differentiated QoS", *IEEE Commun. Mag.*, vol. 37, no. 2, Feb. 1999, pp. 58–66.
 - [83] K. Vlachos, et al., "An optical IM/FSK coding technique for the implementation of a label-controlled arrayed waveguide packet router", *IEEE J. Lightwave Techn.*, vol. , no., Nov. 2003, pp. xx–xx.
 - [84] D.K. Hunter, et al., "WASPNET: A wavelength switched packet network", *IEEE Commun. Mag.*, vol. 37, no. 3, Mar. 1999, pp. 120–129.
 - [85] Y. Lin, "Micromachined free space switches with sub millisecond switching times", *Tech. Digest Optical Fiber Commun. Conf. and Exhibition (OFC 1998)*, 5-10 Mar. 1998, San Jose, CA, pp. 147–148.
 - [86] R. Ryf, et al., "1296-port MEMS transparent optical crossconnect with 2.07 Petabit/s switch capacity", *Tech. Digest Optical Fiber Commun. Conf. and Exhibition (OFC 2001)*, 17-22 Mar. 2001, Anaheim, CA, postdeadline paper PD28.
 - [87] J.F. Fouquet, "Compact optical cross-connect switch based on total internalreflection in a fluid-containing planar lightwave circuit", *Tech. Digest Optical Fiber Commun. Conf. and Exhibition (OFC 2000)*, 5-10 Mar. 2000, Baltimore, MA, pp. 204–206.
 - [88] L. Eldada, R. Norwood, R. Blomquist, L.W. Shacklette, M.J. McFarland, "Thermo-optically active polymeric photonic components", *Tech. Digest Optical Fiber Commun. Conf. and Exhibition (OFC 2000)*, 5-10 Mar. 2000, Baltimore, MA, pp. 124–126.

- [89] D. Enguang, W. Deming, X. Anshi, "High speed integrated acousto-optic switch with high extinction ratio", Tech. Digest Optical Fiber Commun. Conf. and Exhibition (OFC 2000), 5-10 Mar. 2000, Baltimore, MA
- [90] A. Dugan, J.-C. Chiao, "The Optical Switching Spectrum: A Primer on Wavelength Switching Technologies", Telecommunication Magazine, May 2001.
- [91] S. J. B. Yoo, et al., "Rapidly switching all-optical packet routing system with optical-label swapping incorporating tunable wavelength conversion and a uniform-loss cyclic frequency AWGR", Photonics Technology, vol. 14, no. 8, Aug. 2002, pp. 1211–1213.
- [92] C. Guillemot, et al., "Transparent optical packet switching: the European ACTS KEOPS project approach", IEEE J. Lightwave Techn., vol. 16, no. 12, Dec. 1998, pp. 2117–2134.
- [93] D. Chiaroni, et al., "First demonstration of an asynchronous optical packet switching matrix prototype for multi-terabit class routers/switches", Proc. European Conf. on Optical Commun. (ECOC 2001), Amsterdam, The Netherlands, 30 Sep. – 4 Oct. 2001, postdeadline paper ThA4.11.
- [94] S.L. Danielsen, C. Joergensen, B. Mikkelsen, K.E. Stubkjaer, "Analysis of a WDM packet switch with improved performance under bursty traffic conditions due to tuneable wavelength converters", IEEE J. Lightwave Technology, vol. 16, no. 5, May 1998, pp. 729–735.
- [95] V. Eramo, M. Listanti, "Packet loss in a bufferless optical WDM switch employing shared tunable wavelength converters", IEEE J. Lightwave Techn., vol. 18, no. 12, Dec. 2000, pp. 1818–1833.
- [96] D.K. Hunter, M.C. Chia, I. Andonovic, "Buffering in optical packet switches", J. Lightwave Techn., vol. 16, no. 12, Dec. 1998, pp. 2081–2094.
- [97] Z. Haas, "The 'staggering switch': an electronically controlled optical packet switch", J. Lightwave Techn., vol. 11, no. 5, May-Jun 1993, pp. 925–936.
- [98] S. Yao, B. Mukherjee, S. J. Ben Yoo, S. Dixit, "All-Optical Packet-Switched Networks: A Study of Contention-Resolution Schemes in an Irregular Mesh Network with Variable-Sized Packets", Proc. Conf. on Optical Netw. and Commun. (OptiComm 2000), 22-26 Oct. 2000, Plano, TX.
- [99] D.K. Hunter, et al., "SLOB: a switch with large optical buffers for packet switching", J. Lightwave Techn., vol. 16, no. 10, Oct. 1998, pp. 1725–1736.
- [100] S. Bjornstad, D.R. Hjelm, N. Stol, "An optical packet switch design with shared electronic buffering and low bit rate add/drop inputs", Proc. of Int. Conf. on Transparent Optical Netw. (ICTON 2002), Warsaw, Poland, April 2002, pp. 69–72.
- [101] J. Masip, J. Solé-Pareta, "Providing differentiated service categories in optical packet networks", Proc. 16th Intl. Teletraffic Congress (ITC 1999), Edinburgh, UK, 7-11 Jun. 1999, vol. 3, pp. 1115-1126.
- [102] R.W. Stevens, "TCP/IP Illustrated, Volume 1: The Protocols, Addison-Wesley, 1994.
- [103] M. De Prycker, "Asynchronous Transfer Mode: solution for broadband ISDN", Prentice Hall Intl. Ltd., UK, 1995.
- [104] P.B. Hansen, S.L. Danielsen, K.E. Stubkjaer, "Optical packet switching

- without packet alignment”, Proc. European Conf. on Optical Commun. (ECOC 1998), Madrid, Spain, 22-24 Sep. 1998, pp. 591–592.
- [105] B. Mukherjee, A.E. Kamal, “Scheduling variable-length messages on slotted, high-speed fiber optic LANs/MANs using the continuation-bit approach”, Proc. 10th IEEE Conf. on Computer Commun. (Infocom 1991), Bal Harbour, FL, 7-11 Apr. 1991, vol. 2, pp. 678–687
- [106] F. Callegati, W. Cerroni, “Time-wavelength exploitation in optical feedback buffer with trains of packets”, Proc. Conf. on Optical Netw. and Communications (OptiComm 2002), Boston, MA, 29 Jul. – 2 Aug. 2002, pp. 274–285.
- [107] M. Nord, S. Bjornstad, C. Gauger, “OPS or OBS in the core network?”, Proc. 7th IFIP Working Conf. on Optical Netw. Design and Modeling (ONDM 2003), Budapest, Hungary, 3-5 Feb. 2003.
- [108] A. Ge, F. Callegati, L.S. Tamil, “On optical burst switching and self-similar traffic”, IEEE Commun. Letters., vol. 4, no. 3, Mar. 2000.
- [109] Y. Xiong, M. Vandenhouste, H.C. Cankaya, “Control architecture in optical burst-switched WDM networks”, IEEE J. Selected Areas in Commun., vol. 18, no. 10, Oct. 2000, pp. 1838–1851
- [110] S.F. Su, R. Olshansky, “Performance of multiple access WDM networks with subcarrier multiplexed control channels”, IEEE J. Lightwave Techn., vol. 11, no. 5, May-June 1993, pp. 1028–1033.
- [111] C. L. Lu, D. J. M. Sabido, P. Poggiolini, R. T. Hofmeister, L.G. Kazovsky, “CORD—A WDMA optical network: Subcarrier-based signaling and control scheme”, IEEE Photon. Technol. Lett., vol. 7, no. 5, May 1995, pp. 555–557.
- [112] N. Chi, et al., “Optical label swapping and packet transmission based on ASK/DPSK orthogonal modulation format in IP-over-WDM networks”, Tech. Digest Optical Fiber Commun. Conf. and Exhibition (OFC 2003), 23-28 Mar. 2003, Atlanta, GA, pp. 792–794.
- [113] X. Xiao, L.M. Ni, “Internet QoS: the big picture”, IEEE Netw. Mag., vol. 13, no. 2, March 1999, pp. 8–18.
- [114] D. Kettler, H.Kafka, D. Spears, “Driving fiber to the home”, IEEE Commun. Mag., no. 38, no. 11, Nov. 2000, pp. 106–110.
- [115] J. Kawataka, et al., “Novel optical fiber cable for feeder and distribution sections in access network”, IEEE J. Lightwave Techn., vol. 21, no. 3, Mar. 2003, pp. 789–796.
- [116] N. Le Sauze, ed., “Network concepts validation and benchmarking”, Project Deliverable D101 of the IST-project DAVID, Sep. 2003.
- [117] S. Yao, S. J. Ben Yoo, B. Mukherjee, S. Dixit, “All-optical packet switching for metropolitan area networks: Opportunities and challenges”, IEEE Commun. Mag., vol. 39, no. 3, March 2001, pp. 142–148.
- [118] A. Stavdas, et al., “IST-DAVID: concept presentation and physical layer modeling of the metropolitan area network”, IEEE J. Lightwave Techn., vol. 21, no. 2, Feb. 2003, pp. 372–383.
- [119] T.K. Woodward, A.L. Lentine, J.D. Fields, G. Giaretta, R. Limacher, “First demonstration of native Ethernet optical transport system prototype at 10 Gb/s based on multiplexing of gigabit Ethernet signals”, IEEE Photonics

- Techn. Lett., vol. 12, no. 8, Aug. 2000, pp. 1100–1102.
- [120] D.G. Cunningham, “The status of the 10-Gigabit Ethernet standard”, Proc. European Conf. on Optical Commun. (ECOC 2001), Amsterdam, The Netherlands, 30 Sep. – 4 Oct. 2001, pp. 364–367.
 - [121] —, “An Introduction to Resilient Packet Ring Technology”, white paper by the RPR Alliance, online: <http://www.rpralliance.org/articles/Whitepaper10-01.pdf>
 - [122] A. Jourdan, D. Chiaroni, E. Dotaro, G.J. Eilenberger, F. Masetti, M. Renaud, “The perspective of optical packet switching in IP dominant backbone and metropolitan networks”, IEEE Commun. Mag., vol. 39, no. 3, Mar. 2001, pp. 136–141.
 - [123] A. Takada, J.H. Park, “Architecture of ultrafast optical packet switching ring network”, IEEE J. Lightwave Techn., vol. 20, no. 12, Dec. 2002, pp. 2306–2315.
 - [124] N. Le Sauze, et al., “A novel, low cost optical packet metropolitan ring architecture”, Proc. European Conf. on Optical Commun. (ECOC 2001), Amsterdam, The Netherlands, 30 Sep. – 4 Oct. 2001, vol.6, pp. 66–67.
 - [125] N. Le Sauze, et al., “DBORN: a shared WDM Ethernet bus architecture for optical packet metropolitan networks”, Tech. Digest Int. Topical Meeting on Photonics in Switching (PS2002), paper PS.TuC2, Cheju Island, Korea, 21–25 Jul. 2002, pp. 89–91.
 - [126] H. Skouffis, ed., “Results of prospective studies and modelling support”, Project Deliverable D203 of the IST-project DAVID, Sep. 2003.
 - [127] R. Van Caenegem, “Planning van een OPS metro-network”, Master Thesis, Ghent University, 2002–2003.
 - [128] P. Arijs, P. Demeester, “The merit of shared and dedicated protection WDM rings in a hybrid network design”, Tech. Digest Optical Fibre Commun. Conf. and Exhibition (OFC 2000), Baltimore, MD, 7–9 Mar. 2000, vol. 4, pp. 93–95.
 - [129] J.-K. Rhee, I. Tomkos, M.-J. Li, “A broadcast-and-select OADM optical network with dedicated optical-channel protection”, IEEE J. Lightwave Techn., vol. 21, no. 1, Jan. 2003, pp. 25–31.
 - [130] D. Chiaroni, N. Le Sauze, T. Zami, J.-Y. Emery, “Semiconductor optical amplifiers: a key technology to control the packet power variation”, Proc. European Conf. on Optical Commun. (ECOC 2001), Amsterdam, The Netherlands, 30 Sep. – 4 Oct. 2001, vol. 3, pp. 314–315.
 - [131] D. Wolfson, T. Fjelde, A. Kloch, “Technologies for all-optical wavelength conversion in DWDM networks”, Proc. Conf on Lasers and Electro-Optics (CLEO/Pacific Rim 2001), 15–19 Jul. 2001, Chiba, Japan, pp. II-574 – II-575.
 - [132] C. Clos, “A study of non-blocking switching networks, Bell System Technical Journal”, vol. 32, 1953, pp. 406–424.
 - [133] R. A. Thompson, D. K. Hunter, “Elementary photonic switching modules in three divisions”, IEEE J. of Selected Areas in Commun., vol. 14, no. 2, Feb. 1996, pp. 362–373.
 - [134] R. I. MacDonald, “Large modular expandable optical switching matrices”,

- Photonic Techn. Lett., vol. 11, no. 6, Jun. 1999, pp. 668–670.
- [135] I. Lelić, “Large multi-stage OXC”, Proc. European Conf. on Optical Commun. (ECOC 2001), Amsterdam, The Netherlands, 30 Sep. – 4 Oct. 2001, vol.4, pp. 540–541.
 - [136] L. Wosinska, L. Thylen, R. Holmstrom, “Large-capacity strictly nonblocking optical cross-connects based on microelectrooptomechanical systems (MEOMS) switch matrices: reliability performance analysis”, IEEE J. Lightwave Techn., vol. 19, no. 8, Aug. 2001, pp. 1065–1075.
 - [137] G. Shen, T. H. Cheng, S. K. Bose, C. Lu, T. Y. Chai, “Architectural design for multistage 2-D MEMS optical switches”, IEEE J. Lightwave Techn., vol. 20, no. 2, Feb. 2002, pp. 178–187.
 - [138] O. Gerstel, P. Lin, “Efficient architecture for a duplex multi-stage optical non-blocking switch”, Tech. Digest Optical Fibre Commun. Conf. and Exhibition (OFC 2000), Baltimore, MD, 7-9 Mar. 2000, vol. 2, pp. 350–352.
 - [139] Y.-K. Chen, C.-C. Lee, “Fiber-Bragg Grating-based large nonblocking multiwavelength cross-connects”, IEEE J. Lightwave Techn., vol. 16, no. 10, Oct. 1998, pp. 1746–1756.
 - [140] H.J. Chao, C.H. Lam, E. Oki, “Broadband packet switching technologies”, John Wiley & Sons, Inc., 2001.
 - [141] F.T. Leighton, “Introduction to parallel algorithms and architectures: Arrays – Trees – Hypercubes”, Morgan Kaufmann, 1992.
 - [142] M. Collier, T. Curran, “The strictly nonblocking condition in three-stage networks”, Proc. 14th International Teletraffic Congress (ITC 14), Antibes, 6-10 Jun. 1994.
 - [143] A. M. Hill1, D. Careglio, J. Solé-Pareta, A. Rafel, “Relative costs of WDM rings and PONs for metro optical packet networks”, Proc.
 - [144] B.B. Mortensen, M.S. Berger, “Optical Packet Switching Demonstrator”, Proc. European Conf. on Optical Commun. (ECOC 2002), Copenhagen, Denmark, 8-12 Sep. 2002.
 - [145] S. Yao, F. Xue, B. Mukerjee, S.J.B Yoo, S. Dixit, “Electrical ingress buffering and traffic aggregation for optical packet switching and their effect on TCP-level performance in optical mesh networks”, IEEE Commun. Mag., vol. 40, no. 9, Sep. 2002, pp. 66–72.
 - [146] B.B. Mortensen, M.S. Berger, “Estimating timeout parameters for packet aggregation”, Proc. Conf. on the Optical Internet / Australian Conf. on Optical Fibre Techn. (COIN/ACOFT 2003), Melbourne, Australia, 13-16 Jul. 2003, pp. 225–228.
 - [147] M. Düser, P. Bayvel, “Performance of a dynamically wavelength-routed optical burst switched network”, IEEE Photonic Techn. Lett., vol. 14, no. 2, Feb. 2002, pp. 239–241
 - [148] W. Willinger, M. Taqqu, R. Sherman, D. Wilson, “Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level”, IEEE/ACM Trans. Netw., vol. 5, no. 1, Jan. 1997, pp. 71–86.
 - [149] M.E. Crovella and A. Bestavros, “Self-similarity in World Wide Web traffic: evidence and possible causes”, IEEE/ACM Trans. Netw., vol. 5, no. 6, Dec. 1997, pp. 835–846.

- [150] K. Park, G. Kim, M. Crovella, "On the effect of traffic self-similarity on network performance", *Proc. SPIE Int. Conf. on Perf. and Control of Netw. Sys.*, Nov. 1997, pp. 296–310.
- [151] Y. Zhou, H. Sethu, "A simulation study of the impact of switching systems on self-similar properties of traffic", *Proc. Workshop on Statistical Signal and Array Processing*, 14-16 Aug. 2000, pp. 500–504.
- [152] F. Xue, S. Yao, B. Mukherjee, S.J.B. Yoo, "The performance improvement in optical packet-switched networks by traffic shaping of self-similar traffic", *Tech. Digest Optical Fiber Commun. Conf. and Exhibit (OFC 2002)*, Anaheim, CA, 17-22 Mar. 2002.
- [153] H. Christiansen, "Using OPNET to compare and analyse different traffic bundling schemes", *Proc. OPNETWork 2001*, Washington, DC, 27-31 Aug. 2001.
- [154] M. Izal, J. Aracil, "On the influence of self-similarity on optical burst switching traffic", *Proc. Global Telecommun. Conf. (Globecom 2002)*, Taipei, Taiwan, 17-21 Nov. 2002, vol. 3, pp. 2308–2312.
- [155] J.D. Angelopolous, N. Leligou, H. Linardakis, A. Stavdas, "A QoS-sensitive MAV for slotted WDM metropolitan rings", *Proc. 6th IFIP Working Conf. on Optical Netw. Design and Modeling (ONDM 2002)*, Torino, Italy, 4-6 Feb. 2002, pp. 3–15.
- [156] A. Bianco, G. Galante, E. Leonardi, F. Neri, "Measurement based resource allocation for interconnected WDM rings", *Photonic Netw. Commun. (PNET)*, vol. 5, no. 1, Jan. 2003, pp. 5–22.
- [157] T. Inukai, "An Efficient SS/TDMA Time Slot Assignment Algorithm", *IEEE Trans. on Commun.*, vol. 27, no. 10, Oct. 1979, pp. 1449–1455.
- [158] F. Callegati, D. Careglio, W. Cerroni, J. Sole-Pareta, "Assessment of packet loss for an optical feedback buffer node using slotted variable-length packets and heavy-tailed traffic", *Proc. Int. Conf. on Transparent Optical Networks (ICTON 2002)*, Warschau, Poland, 21-25 Apr. 2002, pp. 51-56.
- [159] G. Bianchi, J.S. Turner, "Improved queueing analysis of shared buffer switching networks", *IEEE/ACM Transactions on Networking*, no. 4, August 1993, pp. 482-490.
- [160] A. Monterosso, A. Pattavina, "Performance analysis of multistage interconnection networks with shared-buffered switching elements for ATM switching", *Proc. 11th IEEE Conf. on Computer Commun. (Infocom 1992)*, Florence, Italy, May 1992, pp. 124-131.
- [161] M.G. Hluchyj, M.J. Karol, "Queueing in High-Performance Packet Switching", *IEEE J. Selected Areas in Commun.*, no. 9, December 1988, pp. 1587–1597.
- [162] G. Shen, S.K. Bose, T. Hiang Cheng, C. Lu and T. Yoong Chai, "Performance study on a WDM packet switch with limited-range wavelength converters", *IEEE Commun. Letters*, vol. 5, no. 10, October 2001, pp. 432–434.
- [163] F. Callegati, G. Corazza, C. Rafaelli, "Exploitation of DWDM for optical packet switching with quality of service guarantees", *IEEE J. Selected Areas in Commun.*, vol. 20, no. 1, Jan. 2002, pp. 190–201.

- [164] H. Zang, J.P. Jue and B. Mukherjee, "Photonic slot routing in all-optical WDM mesh networks", Proc. IEEE Global Telecommun. Conf. (Globecom 1999), vol. 1, Rio de Janeiro, Brazil, 5-9 Dec. 1999, pp. 1449-1453.
- [165] A. Kushwaha, Sanjay K. Bose and Y. N. Singh, "Analytical modeling for performance studies of an FLBM-based all-optical packet switch", IEEE Commun. Letters, vol. 5, no. 5, May 2001, pp. 227-229.
- [166] P.D. Bergstrom Jr., M.A. Ingram, A. J. Vernon, J.L.A. Hughes, P. Tetali, "A Markov Chain Model for an Optical Shared-Memory Packet Switch", IEEE Trans. on Commun., vol. , no. 10, October 1999, pp. 1593-1603
- [167] C.A. Brackett, et al., "A scalable multiwavelength multihop optical network: a proposal for research on all-optical networks", IEEE J. Lightwave Techn., vol. 11, no. 5, May/Jun. 1993, pp. 736-753.
- [168] C. Gauger, "Dimensioning of FDL buffers for optical burst switching nodes", Proc. 6th IFIP Working Conf. on Optical Netw. Design and Modeling (ONDM 2002), Torino, Italy, 4-6 Feb. 2002, pp. 117-132.
- [169] Q. He, M.F. Neuts, "Markov chains with marked transitions", Stochastic Processes and their Applications, vol. 74, no. 1, Oct. 1998, pp. 37-52.
- [170] K. Pawlikowski, H.-D.J. Jeong, J.-S.R. Lee, "On credibility of simulation studies of telecommunication studies", IEEE Commun. Mag., vol. 40, no. 1, Jan. 2002, pp. 132-139.
- [171] M. Matsumoto, T. Nishimura, "Mersenne-twister: a 623-dimensionally equidistributed uniform pseudo-random number generator", ACM Trans. Modeling and Comp. Sim., vol. 8, no. 1, Jan. 1998, pp. 3-30
- [172] A.M. Law, W.D. Kelton, "Simulation modeling and analysis", 3rd ed., McGraw Hill, Singapore, 2000.
- [173] W. Cerroni, ed., "Traffic models for optical packet networks with quality of service differentiation", Project Deliverable D121 of the IST-project DAVID, Jul. 2001.
- [174] C.J. Bovy, H.T. Mertodimedjo, G. Hooghiemstra, H. Uijterwaal, P. Van Mieghem, "Analysis of end-to-end delay measurements in Internet", Proc. Passive and Active Measurement Workshop (PAM2002), Fort Collins, CO, 25-26 Mar. 2002.
- [175] S. Blake, et al., "An architecture for differentiated services", Internet RFC 2475, Informational, Dec. 1998.
- [176] R. Braden, D. Clark, S. Shenker, "Integrated services in the Internet architecture: an overview", Internet RFC 1633, Informational, Jun. 1994.
- [177] Y. Xu, R. Guérin, "Individual QoS versus aggregate QoS: A loss performance study", Proc. 21th IEEE Conf. on Computer Commun. (Infocom 2002), New York, NY, 23-27 Jun. 2002.
- [178] D.E. Knuth, "The art of computer programming, volume 1: fundamental algorithms", Third Edition, Addison-Wesley, 1997.
- [179] F. Callegati, "Approximate modeling of optical buffers for variable length packets", Photonic Netw. Commun. (PNET), vol. 3, no. 4, Oct. 2001, pp. 383-390.
- [180] L. Tancevski, S. Yegnanarayanan, G. Castanon, L. Tamil, F. Masetti, T. McDermott, "Optical routing of asynchronous, variable length packets",

- IEEE J. Selected Areas in Commun., vol. 18, no. 10, Oct. 2000, pp. 2084-2093.
- [181] J.S. Turner, "Terabit burst switching", J. High Speed Netw., vol. 8, no. 1, Jan. 1999, pp. 3-16.
 - [182] B. Lannoo, J. Cheyns, E. Van Breusegem, A. Ackaert, M. Pickavet, P. Demeester, "A performance study of different OBS scheduler implementations", Proc. 7th Annual Symp. of the IEEE/LEOS Benelux Chapter, Amsterdam, The Netherlands, 9 Dec. 2002, pp. 191-194.
 - [183] B. Lannoo, "Studie van optical burst switching (OBS) aan de hand van simulaties", Master Thesis, Ghent University, 2001-2002.
 - [184] M. Yoo, C. Qiao, "Supporting multiple classes of services in IP over WDM networks", Proc. IEEE Global Telecommun. Conf. (Globecom 1999), Rio de Janeiro, Brazil, 5-9 Dec. 1999, vol. 1b, pp. 1023-1027.
 - [185] S. Yao, B. Mukherjee, S. J. B. Yoo, S. Dixit, "A unified study of contention-resolution schemes in optical packet-switched networks", IEEE J. Lightwave Techn., vol. 21, no. 3, Mar. 2003, pp. 672-683.
 - [186] S. Yao, S. J. B. Yoo, B. Mukherjee, "A comparison study between slotted and unslotted all-optical packet-switched network with priority-based routing", Tech. Digest Optical Fibre Commun. Conf. and Exhibition (OFC 2001), Anaheim, CA, 17-22 Mar. 2001.
 - [187] R. Jain, S. Routhier, "Packet Trains—Measurements and a New Model for Computer Network Traffic", IEEE J. Selected Areas in Commun., vol. 4, no. 6, Sep. 1986, pp. 986-995.
 - [188] K. Dolzer, C. Gauger, J. Späth, S. Bodamer, "Evaluation of reservation mechanisms for optical burst switching", AEÜ Intl. J. Electron. Commun., vol. 55, no. 1, Jan. 2001.
 - [189] A. Kaheel, T. Khattab, A. Mohamed, H. Alnuweiri, "Quality-of-service mechanisms in IP-over-WDM networks", IEEE Commun. Mag., vol. 40, no. 12, Dec. 2002, pp. 38-43.
 - [190] J. Liu, N. Ansari, "Class-based dynamic buffer allocation for optical burst switching networks", Proc. Workshop on High Performance Switching and Routing (HPSR 2002), Kobe, Japan, 26-29 May 2002.
 - [191] A. Detti, V. Eramo, M. Listanti, "Performance evaluation of a new technique for IP support in a WDM optical network: optical composite burst switching (OCBS)", IEEE J. Lightwave Techn., vol. 20, no. 2, Feb. 2002, pp. 154-165.
 - [192] A. Maach, G.V. Bochmann, "Segmented burst switching: enhancement of optical burst switching to decrease loss rate and support quality of service", Proc. 6th IFIP Working Conf. on Optical Netw. Design and Modelling (ONDM2002), Torino, Italy, 4-6 Feb. 2002, pp. 69-84.
 - [193] Y. Chen, M. Hamdi, D.H.K. Tsang, "Proportional QoS over OBS networks", Proc. IEEE Global Telecommun. Conf. (Globecom 2001), San Antonio, TX, 25-29 Nov. 2001, vol. 3, pp. 1510-1514.
 - [194] C. Loi, W. Liao, D. Yang, "Service differentiation in optical burst switched networks", Proc. Globecom 2002.
 - [195] R. Ramaswami, K. Sivarajan, "Routing and wavelength assignment in all-optical networks", IEEE/ACM Trans. on Netw., vol. 3, no. 5, Oct. 1995, pp.

- 489–500.
- [196] I. Chlamtac, A. Ganz, G. Karmi, “Lightpath communications: an approach to high bandwidth optical WANs”, *IEEE/ACM Trans. on Commun.*, vol. 40, no. 7, Jul. 1992, pp. 1171–1182.
 - [197] N. Wauters, P. Demeester, “Design of the optical path layer in multiwavelength cross-connected networks”, *IEEE J. Selected Areas in Commun.*, vol. 14, no. 5, Jun. 1996, pp. 881–892.
 - [198] N. Nagatsu, S. Okamoto, K. Sato, “Optical path cross-connect system scale evaluation using path accommodation design for restricted wavelength multiplexing”, *IEEE J. Selected Areas in Commun.*, vol. 14, no. 5, Jun. 1996, pp. 893–902.
 - [199] S. Baroni, P. Bayvel, “Wavelength requirements in arbitrarily connected wavelength-routed optical networks”, *IEEE J. Lightwave Techn.*, vol. 15, no. 2, Feb. 1997, pp. 242–251.
 - [200] B. Van Caenegem, W. Van Parys, F. De Turck, P. Demeester, “Dimensioning of survivable WDM networks”, *J. Selected Areas in Commun.*, vol. 16, no. 7, Sep. 1998, pp. 1146–1157.
 - [201] J. Moy, “OSPF version 2”, *Internet RFC 2328*, Standards Track, Apr. 1998.
 - [202] D. Oran, “OSI IS-IS Intra-domain routing protocol”, *Internet RFC 1142*, Feb. 1990.
 - [203] B. Fortz, M. Thorup, “Internet traffic engineering by optimizing OSPF weights”, *Proc. 19th IEEE Conf. on Computer Commun. (Infocom 2000)*, Tel Aviv, Israel, 26–30 Mar. 2000, pp. 519–528.
 - [204] A. Elwalid, Low C. Jin, I. Widjaja, “MATE: MPLS adaptive traffic engineering”, *Proc. 20th IEEE Conf. on Computer Commun. (Infocom 2001)*, Anchorage, AK, 22–26 Apr. 2001, pp. 1300–1309.
 - [205] D. Awduche, “MPLS and traffic engineering in IP networks”, *IEEE Commun. Mag.*, vol. 37, no. 12, Dec. 1999, pp. 42–47.
 - [206] G. Swallow, “MPLS advantages for traffic engineering”, *IEEE Commun. Mag.*, vol. 37, no. 12, Dec. 1999, pp. 54–57.
 - [207] A. Lardies, R. Gupta, R. Patterson, “Traffic grooming in a multi-layer network”, *Optical Netw. Mag.*, vol. 2, no. 3, May/Jun. 2001, pp. 91–99.
 - [208] M. Gryseels, P. Demeester, “A multi-layer planning approach for hybrid SDH-based ATM networks”, *Proc. 6th Intl. Conf. on Telecommun. Systems (ICTS)*, Nashville, TN, 5–8 Mar. 1998, pp. 461–474.
 - [209] S. De Maesschalck, M. Pickavet, D. Colle, P. Demeester, “Multi-layer traffic grooming in networks consisting of an IP/MPLS Layer on top of a meshed optical network layer”, accepted for *IEEE Global Telecommun. Conf. (Globecom 2003)*, 1–5 Dec. 2003, San Francisco, CA.
 - [210] A. Groebbens, D. Colle, S. De Maesschalck, M. Pickavet, P. Demeester, “Logical topology optimisation for dynamic multi-layer recovery schemes”, accepted for *4th Intl. Workshop on Design of Reliable Commun. Netw. (DRCN 2003)*, Alberta, Canada, 19–22 Oct. 2003.
 - [211] S. De Maesschalck, M. Pickavet, D. Colle, P. Demeester, “Influence of the IP traffic asymmetry on the cost of the optical network layer”, *Photonic Netw. Commun. (PNET)*, vol. 4, no. 2, May 2002, pp. 133–150.

- [212] S. De Maesschalck, et al., "Pan-European optical transport networks: an availability-based comparison", *Photonic Netw. Commun. (PNET)*, vol. 5, no. 3, May 2003, pp. 203–225.
- [213] N. Ghani, "Lambda-labeling: a framework for IP-over-WDM using MPLS", *Optical Netw. Mag.*, vol. 1, no. 2, April 2000, pp. 45–48.
- [214] Y. Ye, S. Dixit, M. Ali, "On joint protection/restoration in IP-centric DWDM based optical transport networks", *IEEE Commun. Mag.*, vol. 38, no. 6, Jun. 2000, pp. 174–183.
- [215] S. Leinen, cited on <http://www.cs.columbia.edu/~hgs/internet/traffic.html>, Feb. 2001.
- [216] A.S. Tanenbaum, "Computer networks", 4th ed., Prentice Hall, Aug. 2002.
- [217] S. Floyd, "A report on recent developments in TCP congestion control", *IEEE Commun. Mag.*, vol. 39, no. 4, Apr. 2001, pp. 84–90.
- [218] J. Postel, "Transmission control protocol", Internet RFC 793, Standards Track, Sep. 1981.
- [219] S. Floyd, T. Henderson, "The NewReno modification to TCP's fast recovery algorithm", Internet RFC 2582, Experimental, Apr. 1999.
- [220] R. Braden (ed.), "Requirements for internet hosts – communication layers", Internet RFC 1122, Standards Track, Oct. 1989.
- [221] V. Paxson, M. Allman, "Computing TCP's retransmission timer", Internet RFC 2988, Standards Track, Nov. 2000.
- [222] D. Colle, A. Groebbens, P. Van Heuven, S. De Maesschalck, M. Pickavet, P. Demeester, "Porting MPLS-recovery techniques to the MPLS paradigm", (Invited), *Optical Netw. Mag.*, vol. 2, no. 4, Jul./Aug. 2001, pp. 29–47.
- [223] P. Demeester, et al., "Resilience in multilayer networks", *IEEE Commun. Mag.*, vol. 37, no. 8, Aug. 1999, pp. 70–76.
- [224] V. Sharma, F. Hellstrand (eds.), "Framework for MPLS-based recovery", Internet RFC 3469, Informational, Feb. 2003.
- [225] C. Huang, V. Sharma, K. Owens, S. Makam, "Building reliable MPLS networks using a path protection mechanism", *IEEE Commun. Mag.*, vol. 40, no. 3, Mar. 2002, pp. 156–162.
- [226] R. Goguen, G. Swallow, "RSVP label allocation for backup tunnels", Internet Draft, *draft-swallow-rsvp-bypass-label-01.txt*, Work in Progress, Nov. 2000.
- [227] D. Haskin, R. Krishnan, "A method for setting an alternative label switched paths to handle fast reroute", Internet Draft, *draft-haskin-mpls-fast-reroute-05.txt*, Work in Progress, Nov. 2000.
- [228] L. Breslau, et al., "Advances in network simulation", *IEEE Computer*, vol. 33, no. 5, May 2000, pp. 59–67.
- [229] S. Demaesschalck, "Planning and evaluation of multi-layer optical networks", PhD. Thesis, Ghent University, 2003.
- [230] D. Colle, "Design and evolution of data-centric optical networks", PhD. Thesis, Ghent University, Feb. 2002.
- [231] F. De Turck, "Design of a generic software architecture for efficient management of integrated broadband networks and scheduled multimedia services", PhD. Thesis, Ghent University, May 2002.

- [232] F. Vandermeulen, “On the design of a generic, scalable, and open distributed software architecture for control and management of integrated broadband networks and multimedia services”, PhD. Thesis, Ghent University, May 2001.

Appendices

- [A] C. Develder, J. Cheyns, E. Van Breusegem, E. Baert, D. Colle, M. Pickavet, P. Demeester, “Architectures for optical packet and burst switches”, invited for 29th European Conference on Optical Communications (ECOC 2003), Rimini, Italy, 21-25 Sep. 2003
- [B] C. Develder, R. Van Caenegem, E. Baert, M. Pickavet, P. Demeester, “Active versus passive OPS architectures for metro rings: a network dimensioning point of view”, accepted for 29th European Conference on Optical Communications (ECOC 2003), Rimini, Italy, 21-25 Sep. 2003
- [C] C. Develder, J. Cheyns, M. Pickavet, P. Demeester, “Multistage architectures for optical packet switching using SOA-based broadcast-and-select switches”, Tech. Digest Optical Fibre Conference (OFC 2003), paper FS4, Atlanta, GA, USA, 23-28 Feb. 2003, pp. 794–795.
- [D] C. Develder, M. Pickavet, P. Demeester, “Assessment of packet loss for an optical packet router with recirculating buffer”, Proc. 6th IFIP Working Conference on Optical Network Design and Modelling (ONDM2002), Torino, Italy, 4–6 Feb. 2002, pp. 247–261.
- [E] C. Develder, M. Pickavet, P. Demeester, “Strategies for an FDL based feed-back buffer for an optical packet switch with QoS differentiation”, Proc. Int. Conf. on Optical Internet (COIN2002), paper COIN.TuD1, Cheju Island, Korea, 21–25 Jul. 2002, pp. 114–116.
- [F] C. Develder, J. Cheyns, M. Pickavet, P. Demeester, “Service differentiation mechanisms for variable length packets in an optical switch with recirculating FDL buffer”, accepted for PS 2003, Paris, France, 28 Sep.-2 Oct. 2003
- [G] C. Develder, M. Pickavet, P. Demeester, “On trains and wagons: switching variable length packets in a slotted OPS network”, Proc. Conference on Optical Internet / Australian Conference on Optical Fibre Technology (COIN/ACOFT 2003), paper TuA2-6, Melbourne, Australia, 13-16 Jul. 2003, pp. 233–236.
- [H] E. Baert, C. Develder, M. Pickavet, P. Demeester, “Routing strategies to minimize packet loss in an optical packet switched network with recirculating FDL buffers”, accepted for PNET.
- [I] C. Develder, D. Colle, S. Demaesschalck, M. Pickavet, P. Demeester, “Influence of GMPLS recovery mechanisms on TCP performance”, Photonic Network Commun. (PNET), Vol. 4, No. 3/4, Jul.-Dec. 2002, pp. 321–345.

Inleiding	1
1.1 Optisch PakketSchakelen	1
1.2 Architecturen	5
1.3 Ontwerpskeuzes.....	9
1.3.1 Pakketformaat.....	9
1.3.2 Schakel-algoritmes	11
1.4 Structuur doctoraat	12
1.5 Lijst van publicaties.....	13
Knooparchitecturen	19
2.1 Situering	19
2.2 Metro-netwerken	19
2.2.1 Alternatieven en motivatie	20
2.2.2 Probleembeschrijving en oplossingsmethodiek.....	23
2.2.3 Uni- versus bidirectionele ringen	25
2.2.4 Ruimtelijk hergebruik.....	27
2.2.5 Golflengtebanden	28
2.3 Ruggengraatnetwerk.....	29
2.3.1 Alternatieven en motivatie	29
2.3.2 Schalen door gebruik van Clos-netwerken.....	30
2.3.3 Clos-gebaseerde B&S schakelmatrix	33
2.4 Besluiten	35
Schakel-algoritmes voor OPS.....	39
3.1 Situering	39
3.2 Pakketten van vaste lengte.....	41
3.2.1 Een geschikt schakel-algoritme	42
3.2.2 Methodologie	44
3.2.3 Keuze van de buffer-architectuur	46
3.2.4 Bufferstrategieën	48
3.2.5 Kwaliteitsdifferentiatie.....	49
3.3 Pakketten van variabele lengte	51
3.3.1 Treinen versus wagons	53
3.3.1.1 Methodologie.....	54
3.3.1.2 Invloed van de belasting.....	55
3.3.1.3 Invloed van de granulariteit.....	57
3.3.1.4 Kwaliteitsdifferentiatie voor treinen vs wagons.....	58
3.3.2 Kwaliteitsdifferentiatie voor asynchroon pakketschakelen	59
3.3.2.1 Drie alternatieven voor kwaliteitsdifferentiatie	60
3.3.2.2 Methodologie.....	61
3.3.2.3 Invloed van het aantal bufferpoorten.....	62
3.3.2.4 Invloed van de klasse-offset	64
3.3.2.5 Invloed van de belasting.....	66
3.4 Besluiten	66
Routing in een OPS-netwerk	69
4.1 Situering	69
4.2 Methodologie.....	70
4.2.1 Heuristische routeringsalgoritmes	71

4.2.2	Netwerken en vraagmatrices.....	72
4.3	Evaluatie van de lokale fase algoritmes.....	73
4.4	Voordelen van het schatten van het pakketverlies.....	75
4.4.1	Schatting van pakketverlies.....	75
4.4.2	Vergelijking met kortste-pad routing en gelijkmatig verdeelde belasting.....	78
4.5	Besluiten.....	80
	Interactie met het Transport Controle Protocol.....	83
5.1	Situering.....	83
5.2	TCP/IP.....	84
5.3	Herstel: circuit- versus pakketschakelen.....	85
5.4	Effect van herstel-acties op TCP verkeer.....	88
5.4.1	Bestudeerde facetten.....	88
5.4.2	Methodologie.....	88
5.4.3	Invloed van de omschakelsnelheid.....	90
5.4.4	Invloed van de veranderende RTT.....	94
5.4.5	Invloed van terugschakelen naar werkend pad.....	97
5.4.6	Gevalsstudie: MPLS-protectiemechanismen.....	100
5.5	Besluiten.....	103
	Samenvattend besluit.....	105
	Bibliografie.....	109
	Appendices.....	127