AFSTUDEERVERSLAG VOOR FONTYS HOGESCHOOL INFORMATICA

Project Heart of Gold
(Publieke Versie)

Pim B. van Pelt (83010)
Hogere Informatica

2 december 2002

Bedrijfsbegeleider:	Dhr. M. van Osenbruggen (CEO) Business Internet Trends, Ede Paraaf:
Schoolmentor:	Drs. H. H. Schaaf Fontys Hogeschool, Eindhoven Paraaf:

In de periode van 11/02/2002 tot en met 29/11/2002 (100 dagen)

Samenvatting

Dit document is geschreven in het kader van mijn afstudeerproject aan de Fontys Hogeschool te Eindhoven. Binnen de zakelijke internet provider Business Internet Trends BV (Ede, Gelderland, NL) is een nieuwe generatie data- en telecommunicatie netwerk gerealiseerd. Nadat een Plan van Aanpak is opgesteld waarin probleem- en doelstelling van het project worden neergezet, is een veldonderzoek uitgevoerd om leveranciers van apparatuur te vinden die het best bij de opdrachtgevers wensen en eisen aansluiten (Hoofdstuk 1). Met de uitkomst van dit onderzoek is een ontwerp gemaakt voor de nieuwe generatie backbone van de ISP waarbij de nadruk op redundantie en hoge beschikbaarheid wordt gelegd (Hoofdstuk 2). Daarna is een migratieplan opgesteld om de aangeboden diensten van het bedrijf zo efficiënt mogelijk te migreren naar de gewenste toestand (Hoofdstuk 3). Tenslotte is het reeds ontworpen netwerk concreet gebouwd en is de productie hierop overgezet. De voorgestelde redundantietechnieken worden geïmplementeerd en van de gemaakte (router)configuratie wordt verslag gedaan (Hoofdstuk 4). Aangezien er enkele dagen over waren, is in dit project nog aandacht besteed aan het YAPS programma – wat door de ISP gebruikt wordt aan de basis van haar billing systeem en wat door mij is geschreven (Hoofdstuk 5).

Abstract

This document was written during an internship for a Bachelors Degree with Fontys Hogeschool in Eindhoven, the Netherlands. A new generation data- and telecoms network was created for Business Internet Trends, a business to business ISP based in Ede, Gelderland, the Netherlands. After constructing an Approach Plan (*"Plan van Aanpak"*) which outlines the problems and goals for the project, extensive research was done into the possible vendors and resellers of equipment which bests suits the company's demands (Chapter 1). From the results of this research a design was made for the next generation backbone for the ISP. The main goals for this design were redundancy and high availability for customers (Chapter 2). Then, a migration howto was created to transform the current services of the company into the desired situation as efficiently as possible (Chapter 3). Finally, the designed network was actually built and taken into production. The proposed redundancy techniques were implemented and the used (router)configurations are presented (Chapter 4). There was some time left in the project, so my previous work on YAPS was included. YAPS is a packet analyser program written and used as a corner stone of a billing system for the ISP (Chapter 5).

Dankwoord

Aan een afstudeeropdracht als deze werken veel mensen mee. Ik wil hier van de gelegenheid gebruik maken om de mensen die mij met raad en daad hebben bijgestaan tijdens de afstudeerwerkzaamheden te bedanken.
Allereerst wil ik Michel van Osenbruggen, de eigenaar en directeur van Business Internet Trends BV, bedanken voor de mogelijkheid om zulks een complexe en uitgebreide taak op mij te laten nemen. Daarnaast is gebleken dat Alex Bik niet alleen een prima techneut is, maar ook een bijzonder gastvrij mens. Ik wil hem en zijn vriendin Rose Jeronimus bedanken voor de talloze malen dat zij hun huis voor me hebben geopend; ik hoop dat het eten een beetje lekker was!
Dankzij de prima begeleiding van Bert Schaaf, die me vrij liet in mijn werkwijze en waarmee ik goede afspraken kon maken, is dit project grotendeels gerealiseerd. Hopelijk heb ik enigszins bijgedragen in je algemene netwerkkennis.
Ik wil mijn lieve vriendin Marina Hombroeckx om meer dan één reden mijn dank betuigen. Vooral omdat ze me liet gaan, soms wel voor 48 uur non–stop, zonder thuis te komen en zonder iets van me te laten horen. Soms ging ik erg op in mijn werkzaamheden. Zij begrijpt mij en heeft me altijd gesteund zodat ik kon opgaan in mijn werk.
Ik heb bijna altijd naast mijn studie gewerkt. Hierdoor leerde ik enorm veel van Software Engineering dankzij mijn collega's bij Track en Wegener, later Wiseguys. Ook de mensen bij Freeler en Intouch hebben me de mogelijkheden geboden om me te verdiepen in netwerkgerelateerde zaken, waardoor ik ervan overtuigd ben dat ik een dergelijke opdracht kan voltooien. Bedankt Toine, Rob, Frank, Adrianus en Rager!
Het document wat voor U ligt is een product wat tot stand is gekomen door de inspanningen van mijzelf, maar ook door proeflezers. Zij hebben vele tips en aanwijzingen gegeven over de woordkeuze (soms Vlaams?), maar ook over grammaticale fouten en simpele typefoutjes. Daarom wil ik Peter van Dijk en Joost Vunderink van harte bedanken voor hun inspanning tijdens het proeflezen.
Tenslotte wil ik mijn familie bedanken voor de steun die ze mij, doorheen mijn acht studiejaren hebben getoond. Zonder me te pushen hebben ze mij vrij gelaten om de zaken aan te pakken zoals ik dit wilde. Dit heeft voor mij geresulteerd in een succesvol afgeronde studie en tegelijk een jaar of zes aan relevante bedrijfservaring. Bedankt Pieter, Ingrid, Marjolein en Paul.

Inhoudsopgave

Verklarende Woordenlijst

Term	*Betekenis*
ATM	Asynchronous Transfer Mode – een op cellen gebaseerd netwerk waarin communicatiestromen asynchroon ten op zichte van de andere stromen worden gemultiplexed over het transmissie kanaal. Zie ook: STM
BDC	Backup Domain Controller – een server in de Windows NT netwerkomgeving die vanaf de PDC (zie PDC) de gebruikersrechten en gedeelde bronnen opvraagt. Als de PDC uitvalt, kan de BDC worden geraadpleegd omtrent rechten en instellingen voor het Windows NT Domain. Zie ook PDC en DC.
BGP	Border Gateway Protocol – een routerings protocol waarbij verschillende routers van verschillende bedrijven elkaar op de hoogte houden van welke routes zij kennen, een EGP (zie: EGP)
BRI	Basic Rate Interface – een bundel van 2 ISDN kanalen. Dit type verbinding is wat men gewend is als men spreekt over 'een ISDN aansluiting'. Zie ook PRI.
CIDR	Classless Inter Domain Routing – Een methode om subnets van variabele lengte te creëren en te noteren als prefixlengte zoals 213.136.0.0/19
CWDM	Coarse WDM – een "grove" invulling van maximaal 8 kanalen op één vezelpaar. Zie ook WDM.
DC	Domain Controller – een server in de Windows 2000 netwerkomgeving die de authenticatie verzorgt zodat alle andere servers hun gebruikers en gedeelde bronnen vanuit één centraal punt kunnen beheren. Zie ook PDC en BDC.
DWDM	Dense WDM – een "dichte" invulling van tussen de 8 en 64 kanalen op één vezelpaar. Een aanzienlijk duurdere technologie dan CWDM, maar DWDM biedt wel meer bandbreedte en kanalen. Zie ook WDM.
EAPS	Ethernet Automatic Protection System – een door Extreme Networks ontwikkeld protocol wat in staat is om Ethernet in ringen te bouwen, waarbij de hersteltijd bij een vezel– of switchbreuk lager is dan 50 milliseconden.
EGP	Exterior Gateway Protocol – een verzamelnaam voor externe routingprotocollen, zie BGP.
FE	Fast Ethernet – met een transmissiesneiheid van 100 Mbps de meest gebruikte koperen koppeling.
GE (GigE)	Gigabit Ethernet – met een transmissie snelheid van 1000 Mbps een defacto standaard in hedendaagse netwerken. Ook wel GE genoemd.
IGP	Interior Gateway Protocol – een verzamelnaam voor interne routingprotocollen, zie OSPF.

... vervolgd op volgende pagina

Term	*Betekenis*
OCx	Optical Carrier – een standaard transmissiekanaal op een SONET (zie: SONET) digitaal glasvezelnetwerk. Een OC kanaal is 51.84 Mbps breed. Drie OC kanalen worden gebundeld tot één STM (zie: STM) kanaal.
OSI	Open Standards Interconnect – een theoretische protocolbundel die communicatie over een datanetwerk implementeert. Hiervan zijn de Internet Protocollen (zie: IP en IPv6) afgeleid.
OSPF	Open Shortest Path First – een IGP die gebruikt wordt om de routers in netwerken van bedrijven en ISPs dynamisch de beste routes naar hun bestemmingen te kunnen laten berekenen (zie: IGP)
PDC	Primary Domain Controller – een server in de Windows NT netwerkomgeving die de rechten en instellingen van het Windows NT Domain opslaat. Zie ook PDC en DC.
PDH	Plesiochronous Digital Hierarchy – Een voorloper van SDH die ook digitale subkanalen door middel van TDM (zie: TDM) over één verbinding stuurt.
PRI	Primary Rate Interface – een bundel van 30 ISDN kanalen (Europa) of 24 ISDN kanalen (Japan en Amerika). Zie ook BRI.
SDH	Synchronous Digital Hierarchy – de onderliggende transmissietechnologie in glasvezel verbindingen. Bovenop SDH wordt SONET (zie: SONET) getransporteerd.
SMF	Single Mode Fiber – een type glasvezel dat 9 micrometer dik is en signalen kan vervoeren tussen 1310 en 1600 nm
SONET	Synchronous Optical NETwork – een industriestandaard die gebruik maakt van SDH (zie: SDH) transport technologie. Er worden Optical Carrier levels (zie: OCx) gedefinieerd van ieder 51.84 Mbps, waarop ATM of STM (zie: ATM, STM) kan worden toegepast
STM	Synchronous Transport Module – de technologie om drievouden van Optical Carrier levels (zie: OCx) te transporteren over SONET (zie: SONET). Snelheden zijn STM1 (OC3), STM4 (OC12) STM16 (OC48) en STM64 (OC192).
TDM	Time Division Multiplexing – een techniek die gebruikt wordt door PDH technologie waarbij kleine (2 Mbit/s) kanalen om en om worden uitgezonden op het glasvezelmedium.
WDM	Wavelength Division Multiplexing – een techniek die gebruikt wordt om de signalen van meerdere glasvezelverbindingen te combineren tot één signaal en die uit te zenden over één vezel of vezelpaar.

Hoofdstuk 1

Netwerk onderzoek

1.1 Inleiding

Om ons een beeld te kunnen vormen van het huidige productienetwerk van Business Internet Trends, zullen we een netwerkschets opmaken van de situatie zoals die aan het begin van het project is. Vervolgens zetten we de voorlopige eisen voor de nieuwe verbindingen in termen van bandbreedte en doorvoer op papier.
Om het nieuwe communicatie netwerk van het bedrijf te kunnen bouwen volstaat de huidige apparatuur niet. De huidige routers zijn niet in staat om snel genoeg de data door te voeren die wordt gegenereerd in het nieuw te bouwen netwerk.
Er worden enkele leveranciers van router apparatuur benaderd. Hen wordt een lijst van kenmerken gegeven en de leveranciers beoordelen hun eigen product. Deze beoordelingen worden vergeleken en hieruit rolt een kandidaat-leverancier voor het nieuwe netwerk.

1.2 Probleem- en doelstelling

Voor het onderzoek gedeelte van het project definiëren we drie problemen en in dit hoofdstuk worden oplossingen gepresenteerd voor deze problemen.

1.2.1 Probleemstelling

1. Het huidige netwerk is nog onbekend en zal in het project opgenomen moeten worden. Het netwerk wat nieuw gebouwd moet worden zal (zonder in te gaan op lokale details) in ruwe vorm gespecificeerd moeten worden.

2. Er is momenteel geen overzicht van de mogelijke nieuwe apparatuur. Business Internet Trends is enigszins bekend met Cisco Systems, maar niet met andere fabrikanten zoals Juniper en Extreme Networks. Nochtans hebben deze fabrikanten in de laatste jaren een aanzienlijk marktaandeel weten te verkrijgen.

3. Als we om ons heen kijken naar andere Internet Exchange projecten, blijkt dat het succes maar matig is. We vragen ons af waar dit aan ligt en willen graag weten hoe we onze Ede Internet Exchange (*EdIX*) zo kunnen opzetten, dat we een groter succes kunnen boeken.

1.2.2 Doelstelling

Het huidige netwerk zal volledig in kaart worden gebracht. Hierbij wordt melding gemaakt van de verschillende routers en switches in zowel Ede als Amsterdam en alle onderlinge verbindingen, VLANs en gebruikte interfaces.
We stappen actief af op leveranciers en vragen hen om tekst en uitleg te geven over hun producten. We gaan in een initiëel gesprek naar de fabrikant toe om het bedrijf en onszelf voor te stellen. Daarna nodigen we de technische teams van de leverancier uit om ons te overtuigen van de kwaliteit en doordachtheid van hun routers en switches.
We nemen ons tenslotte voor om inzicht te verkrijgen in het opzetten van een technisch model voor een Internet Exchange.

1.3 Technieken en Methodes

1.3.1 Netwerkschema's

Voor het tekenen van netwerk schema's wordt een standaard methode gebruikt, berustend op het software programma *dia* [1]. Er zullen tekeningen gemaakt worden van het huidige locale netwerk, de langeafstands verbindingen naar Amsterdam, alsook de beoogde nieuwe situatie (in termen van verbindingen en lokaal netwerk).

1.3.2 Vergelijkend onderzoek

Om nieuwe apparatuur te kiezen, worden eerst de problemen van het huidige netwerk in kaart gebracht. Vervolgens wordt aan de engineering afdeling gevraagd welke aspecten van het netwerk vereist dan wel optioneel zijn en deze worden ieder van een score tussen 0 en 10 toegekend. Hierbij betekent 0 "totaal irrelevant" en 10 "uiterst relevant".
De leveranciers krijgen deze lijst met kenmerken zonder score toegestuurd en zij worden gevraagd een rapportcijfer tussen 0 ("absoluut onmogelijk") en 10 ("uitstekend geschikt") te geven. Als het cijfer lager is dan een 4 of hoger dan een 7, wordt een korte maar overtuigende redenering, al dan niet aangevuld met documentatie, gevraagd.

1.3.3 Internet Exchange ontwerpen

Om meer kennis op te doen van het opereren van een Internet Exchange, zal gesproken worden met de AMS-IX netwerk operatoren. Hiervoor zal een afspraak gemaakt worden en aan hen gevraagd welke struikelblokken er te verwachten zijn. Deze kennis wordt gedeeld met andere beginnende exchanges zoals de Enschede Internet Exchange (*NDIX*) en de Almere Internet Exchange (*AlmIX*) en dit zal leiden tot een aanzet tot een technisch ontwerp van de eigen Ede Internet Exchange.

[1] Een GPL based tekenprogramma voor schema's. De homepage voor het project is http://www.lysator.liu.se/~alla/dia/

1.4 Uitvoering

1.4.1 De huidige situatie

Het huidige productienetwerk

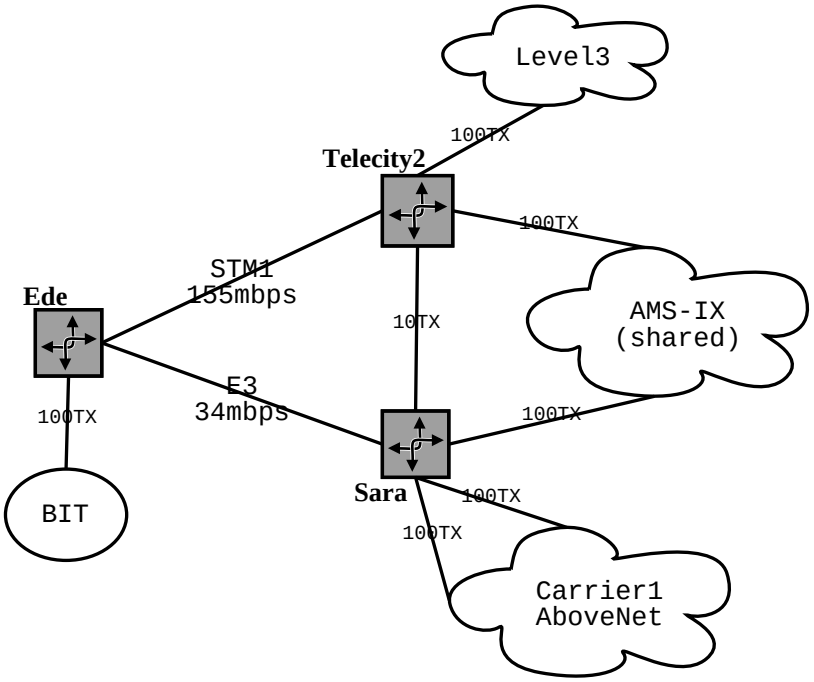
Het netwerk van Business Internet Trends wordt *top-down* onderzocht. We beginnen bij de Amsterdam Internet Exchange, met de koppelingen naar het hoofdgebouw in Ede. Vervolgens bespreken we in detail het huidige lokale netwerk van het bedrijf.

De nationale verbindingen Amsterdam-Ede

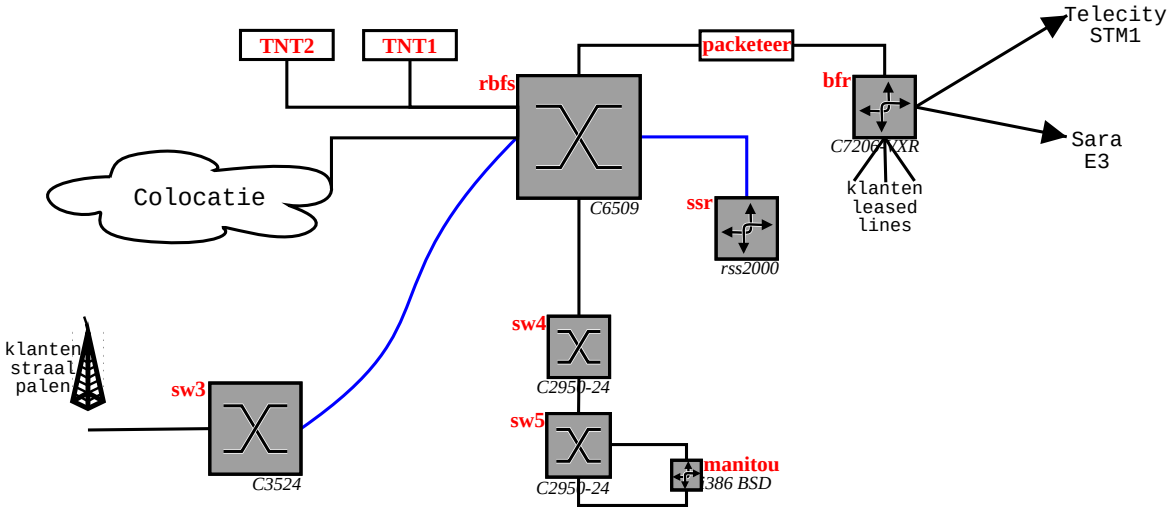
De AMS-IX bestaat uit vier locaties (Nikhef, Sara, GlobalSwitch en Telecity2), waarvan wij presentie hebben op de Sara en de Telecity2 locatie. Op beide plaatsen staat een Cisco router. In het Sara gebouw op de Kruislaan (Amsterdam Watergraafsmeer) staat een Cisco 7206 die een E3 (34 Mbps ATM) naar Ede opvangt. Deze router hangt ook aan het shared medium van de AMS-IX en voert transit af naar twee providers, te weten Carrier1 en AboveNet. In Telecity2 (Amsterdam Zuid-Oost) staat een Cisco 7505 die een OC3 (155 Mbps STM1) verbinding naar Ede opvangt. Deze router is met 10 Mbps verbonden met de Sara router en met één transit provider, te weten Level3. Tevens hangt de Telecity2 router aan het shared medium. In Ede komen zowel de E3 uit Sara als de STM1 uit Telecity2 samen in één router. Deze is met 100 Mbps verbonden aan het locale netwerk van het bedrijf. Samen omvatten deze drie routers de backbone van Business Internet Trends. Hun onderlinge relatie wordt weergegeven in figuur 1.1(a).

Het lokale netwerk in Ede

Het lokale netwerk in Ede verdient wat meer aandacht. Een schematische voorstelling is gemaakt in figuur 1.1(b).
De router in Ede (met de verbindingen naar Amsterdam) is een Cisco 7206 en heeft de volgende poorten: de E3 naar Sara, de OC3 naar Telecity2, 16 huurlijnen van E1 (elk 2 Mbps) en tenslotte een FE poort (100 Mbps) naar het lokale netwerk. Er valt op dat de volledige capaciteit van de uitgaande verbindingen, niet geheel ontkoppeld wordt in de Ede locatie, immers $E3 + OC3 > 16 * E1 + FE$ waarbij het overschot $189 - 132 = 57Mbps$ (30%) blijkt. Dit is zonde van de investering en zal moeten worden opgelost.
De grens van het Edese netwerk is BFR, de Cisco 7206. De FE koppeling gaat via een Packeteer naar een flink uit de kluiten gewassen Cisco 6509 switch met ruim 400 FE poorten (RBFS, een joviale afkorting voor *really big fucking switch*). De Packeteer is in staat om uitgaande datastromen op een vantevoren ingestelde bandbreedte te limiteren en kan ook realtime statistieken van de flows bijhouden en weergeven in een webinterface.
Links boven in figuur 1.1(b) staan twee TNT inbelservers. Deze bieden landelijke inbelfunctionaliteit (voor ISDN zowel als analoge modems) aan de klanten. Op iedere TNT worden 7 PRIs afgehandeld, dus de totale inbelcapaciteit van het bedrijf bedraagt 420 verbindingen.
De colocatie faciliteit van Business Internet Trends omvat drie afzonderlijke ruimten. In iedere ruimte staan 19- rekken met daarin de servers van de klanten



(a) Het backbone netwerk tussen Amsterdam en Ede.



(b) Het lokale netwerk in de colocatie faciliteit te Ede

Figuur 1.1: *Het gehele Business Internet Trends netwerk aan het begin van de opdracht (februari 2002)*

alsook de eigen servers. Sommige van de klanten hebben in de colocatie een eigen virtueel netwerk (VLAN).

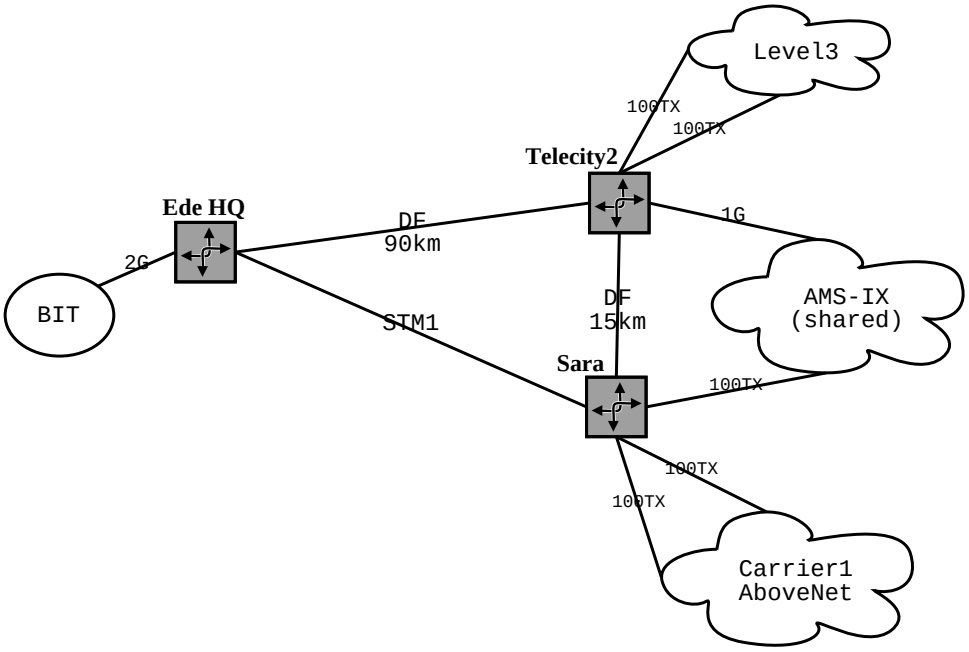
VLAN id	*Naam*	*Interface*
1	DEFAULT	213.136.0.4/24
2	Wireless	213.136.6.254/24
3	Princen	213.136.15.1/24
6	Akamai	213.136.9.177/28
7	Van Beekum	213.136.14.49/28
8	Commotio	213.136.14.65/27
9	Bart Smit	213.136.9.65/28
10	Bandwidth Brokers	213.136.13.1/25
11	RCS	213.136.9.233/29
12	Terrazur	213.136.14.145/28
13	Testnetwerk	213.136.9.241/28
14	Diva	213.136.18.1/27
15	Rodi	213.136.9.217/29
16	PrimeXS	213.136.14.161/28
17	ABC	213.136.9.113/28
51	Colocatie1	213.136.19.1/26
52	Colocatie2	213.136.19.65/26
53	Colocatie3	213.136.19.129/26
54	Colocatie4	213.136.19.193/26
102	Lukkien2	213.136.14.1/27

Tabel 1.1: *De VLAN indeling op* SSR *bij aanvang van het project.*

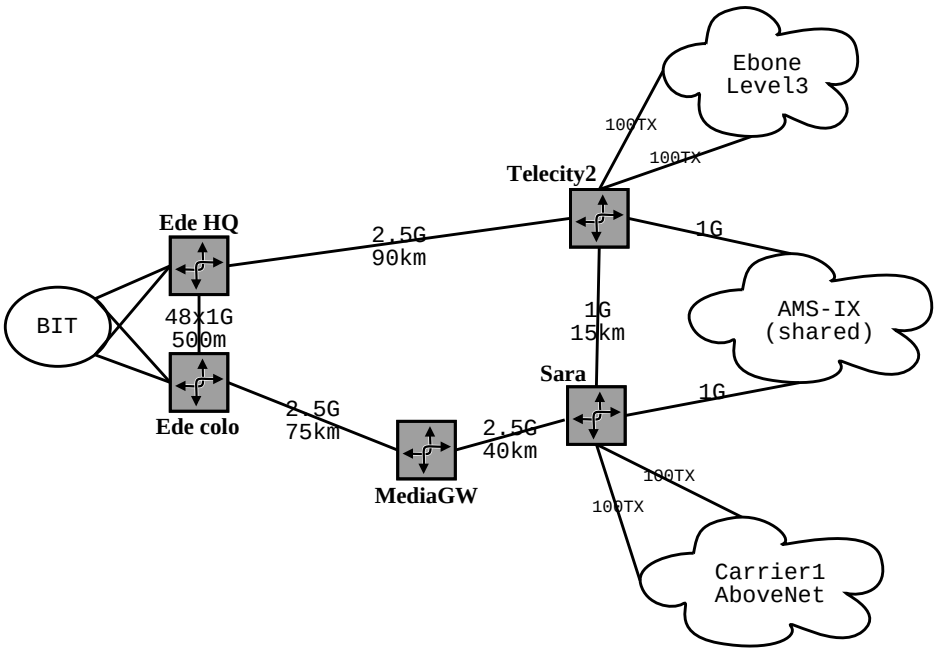
Tussen RBFS en SW3 is een glasvezel GE koppeling gemaakt. De taak van deze switch is om een aantal straalverbindingen op te vangen. Deze verbindingen worden tot stand gebracht met klanten in de regio door middel van antennes op het dak van het kantoorgebouw.
Naar het eigen kantoornetwerk wordt een FE koppeling gemaakt. In het engineering kantoor staan twee switches SW4 en SW5. Het kantoornetwerk wordt door middel van NAT afgescheiden van het Internet. Dit verhoogt de veiligheid van de werkstations van de engineers. De machine die deze firewalling en NAT verzorgt heet MANITOU en draait als bestuuringssysteem FreeBSD (een unix variant).
De routering tussen alle bovengenoemde netwerken wordt verzorgd door een Riverstone router (RSS). Deze is met RBFS verbonden door middel van een glasvezel GE koppeling. Alle VLANs worden hier gedefinieerd en de router biedt zich aan als default gateway voor nagenoeg elk VLAN. De VLANs en interfaces worden in tabel 1.1 uiteengezet.

1.4.2 De beoogde situatie

Een van de aanleidingen van deze opdracht was het verwerven van een dark fiber tussen Amsterdam en Ede. De eigenaar van dit vezelpaar is Level3 Communications. De term 'dark' fiber slaat op het feit dat er aan weerszijden geen transmissie apparatuur aangesloten staat en dat de vezel onderweg nergens wordt versterkt. Dit heeft als nadeel dat Business Internet Trends zelf voor de apparatuur zorg zal moeten dragen maar als voordeel dat er practisch geen beperking is op de snelheid waarmee kan worden gecommuniceerd. Hierbij is vooral het budget een beperkende factor.



(a) De initiële backbone vereisten, korte termijn



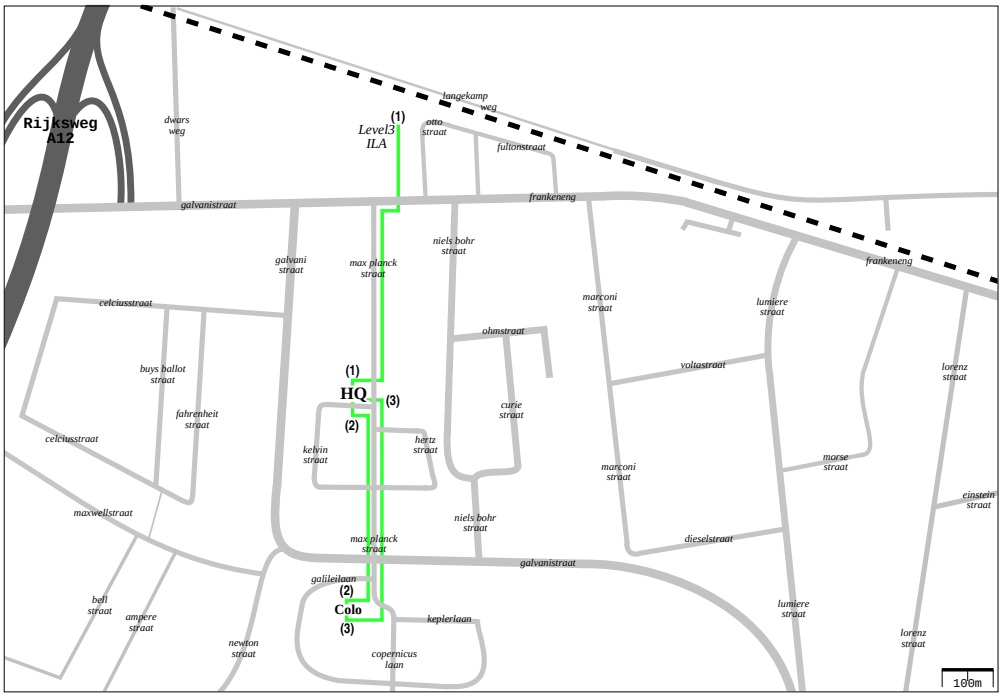
(b) De geschatte backbone vereisten over twee jaar

Figuur 1.2: *De evolutie van de Business Internet Trends backbone gedurende dit project (a) en daarna (b).*

Figuur 1.2 laat de backbone verwachtingen zien op twee tijdstippen. Figuur 1.2(a) toont het doel van deze afstudeeropdracht en figuur 1.2(b) toont de verwachting binnen twee jaar. Wij zullen ons vanaf nu bezighouden met het onderzoeken en ontwerpen van het netwerk in figuur 1.2(a).

1.4.3 De glasvezel Ede-Amsterdam

De vezel die we huren maakt deel uit van de Europese ring van Level3 Communications. Zo'n glasvezel ring wordt normaal gezien om de 85 kilometer optisch versterkt in een zogenaamde *ILA site*. Deze 'inline amplification site' is doorgaans een zeecontainer (bovengronds) waarin de glasvezels in een versterker (zie verderop in dit hoofdstuk) worden geleid. Het signaal wordt vervolgens verder op de ring getransporteerd. Level3 heeft zo'n ILA site op nog geen 700 meter van ons pand. Vanaf deze locatie naar het eigen pand draagt Business Internet Trends zorg voor het graven van een buis, ook wel 'duct' genaamd. Door deze buis wordt een glasvezel geblazen van de ILA site naar het hoofdkantoor op de Kelvinstraat. Twee soortgelijke ducts worden gecreëerd tussen de Kelvinstraat en de nieuwbouw, zodat in de toekomst een redundant pad tussen beide gebouwen kan worden gemaakt.



Figuur 1.3: *Het stukje Ede waarin onze ducts worden gemaakt. (1) is duct van de ILA naar de Kelvinstraat (HQ), (2) en (3) zijn de ducts van de Kelvinstraat naar het nieuwe pand (Colo).*

Bij de ILA site wordt de vezel uit onze duct ingekoppeld in de ring van Level3. We spreken hierbij met Level3 af, dat ons signaal niet door hen opgepakt en versterkt

zal worden. De ring wordt aan de Amsterdamse kant van de ILA dus verbroken (dit heet een “splice”). Vervolgens worden de vezels uit onze duct aan de Level3 vezels vastgelast (een “weld”) en ontstaat er een glasvezelverbinding van 86.6 km van ons pand naar het eindstation van Level3 in Amsterdam. Alhier zal Level3 een koppeling verzorgen met een laatste stukje glasvezel van hun Amsterdamse site, de Level3 Gateway genoemd, naar het Telecity gebouw in Zuid-Oost. Deze verbinding is nog eens 3 km, wat de totale verbinding 89.6 km maakt.
Omdat de vezel geen enkele electrische of optische machine doorloopt, zal er, totdat wij apparatuur in Amsterdam en Ede aansluiten, geen licht door de vezels schijnen. Hierdoor heet zo’n vezel, waarbij dus geen tussenliggende apparatuur wordt gebruikt, een *Dark Fiber*. Bij traditionele telefonie en datacommunicatie diensten (zoals de STM1 of E3 die Business Internet Trends nu heeft) wordt de klant een dienst aangeboden waarvan de telecom provider de apparatuur op de vezels reeds heeft aangesloten. Hierdoor kan men geen willekeurige apparatuur meer aansluiten, maar alleen de door de telco gespecificeerde types.
Bij een Dark Fiber is dit niet het geval. Omdat de keuze voor transmissie apparatuur hier bij de klant ligt, kan men kiezen voor een STM1 (155 Mbps) verbinding, maar ook voor een Ethernet (100 Mbps) verbinding. Ook staat de volledige capaciteit van de vezels nu aan de klant ter beschikking. Men kan dus al naar gelang het budget een keuze maken tussen 100 Mbps (en oplopend tot zelfs 2.5 Tbps!).
Gedacht wordt aan een transmissiesnelheid tussen de 2Gbps en 10Gbps. Er blijken drie belangrijke motivaties voor deze (redelijk hoge) snelheid:

1. De ISP activiteiten van Business Internet Trends groeien en hebben binnenkort behoefte aan meer bandbreedte dan er momenteel voorradig is. Geschat wordt dat binnen 12 maanden de gigabit grens wordt bereikt.

2. Een onderdeel van het dienstenpakket van het bedrijf wordt het verhuren van layer2 bandbreedte tussen de AMS-IX en de nieuwe locatie in Ede. Hierbij valt te denken aan de Internet Exchange, maar ook aan colocatie van andere ISPs in Ede.

3. Klanten worden steeds veeleisender. Het aanbieden van een contract waarin gigabit en hogere snelheden kunnen worden gegarandeerd met strenge service level agreements zullen bijdragen tot grotere omzet. Vanuit sales perspectief is een snelle verbinding dus wenselijk.

1.4.4 Het leveranciers onderzoek

Omdat Business Internet Trends op dit moment klant is bij Cisco Systems en er steeds meer bedrijven zich positief uitspreken over Juniper Networks, zijn deze twee bedrijven bij voorbaat toegevoegd aan de potentiële leverancierslijst (tabel 1.2).
CeBIT Hannover is ’s werelds grootste beurs voor informatie en telecommunicatie technologie. Deze beurs, met elk jaar ruim 2 miljoen bezoekers, geeft een bijzonder compleet beeld van de stand van zaken in de ICT branche en biedt tevens een uitstekende mogelijkheid om in contact te treden met potientiële leveranciers. Dit jaar (2002) was de CeBIT van 13 tot en met 20 Maart in de Hannover Messe. Wij zijn met vier personen naar de beurs afgereisd om ons te laten informeren over de laatste stand van zaken in hogesnelheids verbindingen over dark fiber. Samen

met MICHEL VAN OSENBRUGGEN (bedrijfsmentor), ALEX BIK (domeindeskundige) en SABRI BERISHA (engineering) heb ik me twee dagen lang laten onderdompelen in de high-tech maar ook de in mainstream beschikbare transmissie apparaten.

Bedrijf	*Branche*	*Naam contactpersoon*	*Gevonden via*
Foundry	IP/SDH	Ralf Gotsche	AMS-IX
Tellabs	DWDM	Kimmo Pajunen	CeBIT
MicroSens	CWDM	Dirk Herppich	CeBIT
Sorrento	CWDM	Stefan Nothdurft	CeBIT
Alcatel	DWDM	Ulrich Knörr	CeBIT
Cisco	IP/DPT	Jochem Steman	(reeds klant)
Lucent	IP/SDH	Avsarhan Tanir	CeBIT
RiverStone	IP/SDH	John Schaap	vorige werkgever
Nortel	IP/SDH	Wolfgang Meinolf	CeBIT
Juniper	IP/SDH	Rene Roersma	reeds bekend
Optical Access	CWDM	Bram v/d Kade	WWW
Fiber Driver	CWDM	Ronald v/d Meer	WWW
FNE	CWDM	Brian Sinfield	WWW
Luminous	DWDM	Steve Wines	vorige werkgever

Tabel 1.2: *Alle leveranciers die benaderd zijn in het leveranciers onderzoek.*

1.4.5 Mogelijke oplossingen

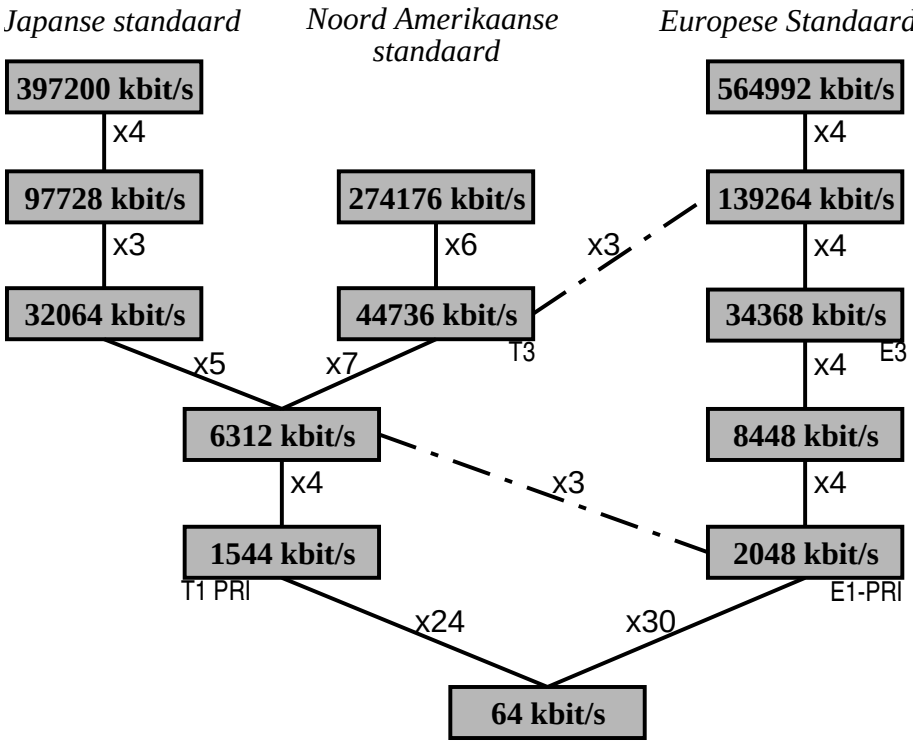
Op de CeBIT bleek dat we niet enkel kunnen volstaan met leveranciers die SDH technologie verkopen (dus STM16 of STM64, de “klassieke oplossing”). De bedrijven die WDM apparatuur verkopen, blijken een uiterst competitieve oplossing te kunnen aanbieden. We besloten daarom om de mogelijke oplossingen niet enkel met SDH, maar ook met WDM apparatuur te onderzoeken.
In onze gesprekken met Cisco Systems kwam naar voren dat zij een duidelijke voorkeur hebben voor een nieuwe technologie die gebaseerd is op SDH in ringstructuren. Cisco noemt deze technologie *Resilient Packet Ring* (RPR) maar de generieke term voor een dergelijke ring is *Dynamic Packet Transport* (DPT).

SDH Technologie

Om reguliere digitale telefoniekanalen (ieder 64 kbit/s breed) goed te kunnen combineren is door de telecom-industrie het *plesiochronous* [2] *digital hierarchy* (PDH) ontworpen. Aan de basis hiervan staan bundels van 24 bovengenoemde kanalen in de Verenigde Staten of 30 kanalen in Europa, het Primary Rate Interface van in totaal 1544 kbit/s (VS) respectievelijk 2048 kbit/s (EU) breed.
Voor telecom operatoren zijn deze kanalen in backbones te bundelen door middel van multiplexing tot grotere kanalen van 8, 34, 140 Mbit/s en verder tot aan 10 Gbit/s. Hierbij worden op hoge snelheid om en om frames van de PRI kanalen uitgezonden op dezelfde glasvezel. Dit heet *Time Division Multiplexing* (TDM).

[2] Plesiochroon staat voor “pseudo” synchroon. Het tijdsaspect van de onderlinge signalen wijkt enigszins af.

Door de onnauwkeurige timing van datastromen, die niet exact wordt gegarandeerd door PDH, wordt het samenvoegen van kanalen bij toenemende bandbreedte erg inefficiënt en dus kostbaar. Figuur 1.4 toont de samenhang van digitale kanalen in de PDH standaarden.

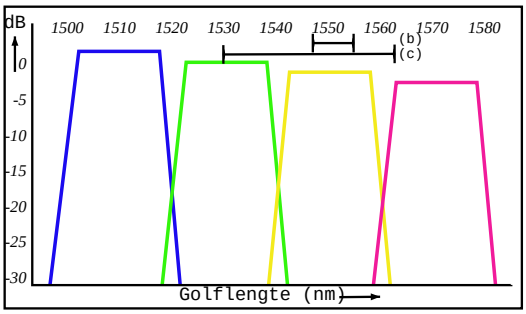


Figuur 1.4: *De onderlinge samenhang van de wereldwijd toegepaste PDH technologie*

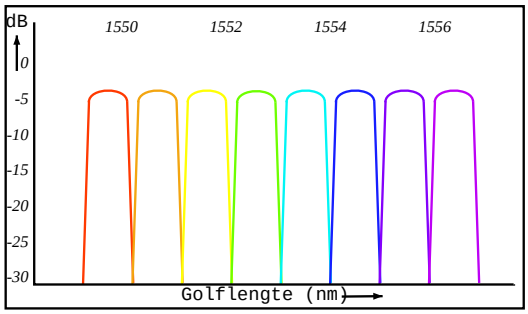
In de datacomunicatie industrie is de *synchronous digital hierarchy* SDH standaard ontwikkeld en zijn de basis-kanalen vastgesteld op 51840 kbit/s. Deze kanalen worden *Optical Carriers* (OCx) genoemd. Er kunnen nu in Europa 21 PRIs (van 2048 kbit/s) en in Amerika 28 PRIs (van 1544 kbit/s) in een OC1 worden omgezet. Hierbij wordt er voor beide signaaltypen een speling gehanteerd van 7 Mbit/s om de tijdsverschillen in PDH op te vangen en exact synchroon te maken in het SDH signaal. Dit principe wordt SONET genoemd, een afkorting voor *Synchronous Optical NETwork*.

WDM Technologie

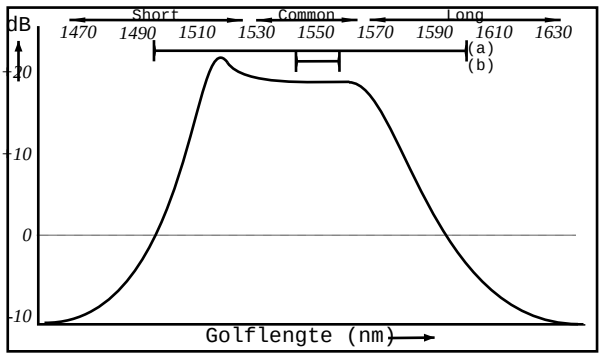
WDM staat voor *Wavelength Division Multiplexing*, een methode om simultaan over één vezel meerdere signalen uit te zenden. Dit wordt bewerkstelligd door meerdere inkomende signalen op te pakken en ieder met een eigen golflengte over dezelfde glasvezel uit te sturen.
De laserdiodes die worden gebruikt bij WDM zenden ieder in een bepaalde golflengte uit. De golflengte van de transmissie hangt echter af van de temperatuur. Voor elke graad celcius warmte stijging, verschuift de golflengte van de laser met 0.08 nm. Bij een groot temperatuursverschil tussen de zender en ontvanger kan het



(a) Coarse WDM, kanaalbreedte: 20nm. Lijnstuk (b) is de bandbreedte die voor DWDM wordt gebruikt, (c) is het versterkbare spectrum.



(b) Dense WDM, kanaalbreedte: 0.8nm. DWDM kan versterkt worden door optische versterkers (figuur c)



(c) Optische versterkers. CWDM (a) is te breedbandig om versterkt te worden, DWDM (b) kan wel versterkt worden.

Figuur 1.5: *Het belangrijkste verschil tussen Coarse en Dense WDM is de gebruikte bandbreedte per kanaal. Opmerkelijk is ook dat DWDM versterkt kan worden, CWDM niet.*

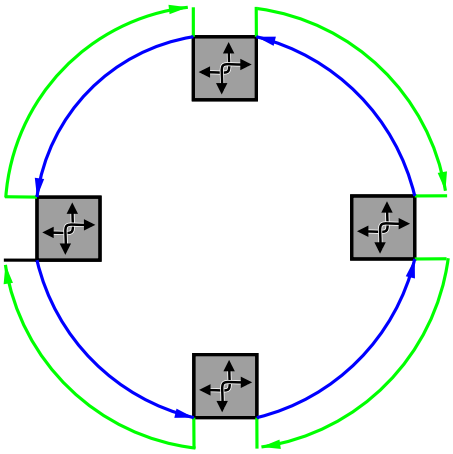
verloop derhalve zó groot worden, dat de kanalen door elkaar gaan verschuiven en dit moet uiteraard voorkomen worden.
Kanaalscheiding kan in principe op twee manieren. Enerzijds kan de kanaalbreedte dusdanig groot worden gekozen, dat het verschuiven met 10 nm getolereerd wordt. Anderzijds kan men de lasers zeer nauwkeurig koelen, waardoor ze niet afhankelijk zijn van de buiten-temperatuur en dus immer licht van dezelfde golflengte uitzenden.
Deze gedifferentiëerde technieken staan aan de basis van wat men Coarse (ongekoeld, brede kanalen) en Dense (gekoeld, smalle kanalen) Wavelength Devision Multiplexing noemt. Figuur 1.5 toont deze kanalen voor beide types WDM.

CWDM vs DWDM

CWDM biedt als voordeel dat de lasers niet goed gekoeld moeten opereren, wat een kostenbesparing is in termen van warmte dissipatie alsook aanschafprijs. Uiter-

aard volgt uit een ruime kanalenkeuze dat er in standaard SMF (tussen de 1470nm en 1610nm) maximaal 8 kanalen kunnen worden voorzien. De kanalen zijn dan ongeveer 20nm breed.
DWDM daarentegen biedt een veelvoud aan kanalen tussen de 1530nm en 1562nm. Al naar gelang de kwaliteit van de laser diodes, kan een kanaalbreedte van 0.8nm (figuur 1.5(b)) of zelfs 0.25nm worden gekozen. Hierdoor kunnen er met de huidige stand van techniek 128 kanalen worden uitgezonden.
Een belangrijk aspect bij de overweging welke WDM technologie te kiezen is de reikwijdte van de laser diodes (en bijbehorende ontvangers). Zonder versterking kan zo'n laser maximaal 150km overbruggen. In figuur 1.5(c) is het gedrag van een optische versterker te zien die wordt gebruikt om WDM signalen te versterken. We merken op dat deze versterkers een beperkte bandbreedte kunnen versterken. Het gedeelte wat versterkt kan worden heet het *Common Spectrum*. Eronder en -boven functioneert de versterker niet of werkt hij zelfs dempend op het signaal.
In figuur 1.2(b) stelden we reeds vast dat de verbinding tussen het hoofdkantoor in Ede en Telecity in Amsterdam 90 km is. Als de verbinding tussen Telecity en Sara uit valt (vezelbreuk) zal het verkeer tussen die twee lokaties de omgekeerde richting over de ring moeten volgen en deze afstand (180 km) overschreidt de kritische grens van het vermogen van (onversterkbare) CWDM lasers. De keuze tussen CWDM en DWDM zal dus niet alleen afhangen van de hoeveelheid benodigde kanalen maar ook van de wens voor redundantie.
Aangezien het in het OSI model op steeds hogere lagen ook steeds bewerkelijker wordt om redundantie te kunnen garanderen, zal er een afweging gemaakt moeten worden of de routers in layer3 (IP) redundancy zullen aanbrengen door middel van omrouteren van IP packets, of dat de onderliggende transmissie apparatuur voor de redundantie zal zorg dragen.

DPT Technologie



Figuur 1.6: *De Dynamic Packet Transport (ook: Resilient Packet Ring) technologie.*

DPT wordt altijd in ringen gecreëerd. Hierbij zijn de routers allen uitgerust met (dezelfde) SDH lijnkaarten. De glasvezelring (die uit een vezel*paar* bestaat) wordt in twee ringen verdeeld. Hierbij gaat de data van de ene ring in de tegenovergestelde richting ten opzichte van de andere ring (zie figuur 1.6).

De inhoud van de datacommunicatie bestaat enerzijds uit commando's voor de routers onderling en anderzijds uit de daadwerkelijke datapakketjes. Nu gaan de commando's voor de buitenste (groene) ring over de binnenste (blauwe) ring en omgekeerd. Hierdoor kunnen, bij overbelasting of zelfs bij uitval van één vezelverbinding, de routers steeds blijven communiceren over de andere ring.
Het DPT protocol verzorgt dynamisch herstel van een glasvezelbreuk. Na constatering van een verbroken verbinding tussen twee routers, zullen deze twee routers als het ware de buitenste en binnenste ringen aan elkaar koppelen, zodat er één grote fysieke ring ontstaat. De data zal dan wellicht langer onderweg zijn, maar wegens de ring eigenschap kan bewezen worden dat de data daadwerkelijk aan zal komen.
Aanvankelijk heeft Cisco een eigen protocol ontworpen om in een ringstructuur een dergelijk hersteltijd te realiseren. Deze heette Resilient Packet Ring (RPR). Later heeft Cisco dit standaard (in iets gewijzigde vorm) vrijgegeven en is er de RPR technologie aan ontleend. Deze laatste is een open standaard. Voor het maken van de DPT controller chip die nodig is in elk DPT apparaat, heeft Cisco echter nog steeds enkele patenten waardoor iedere leverancier die dergelijke routers wil aanbieden, licentiekosten zal moeten afdragen aan Cisco.

1.4.6 Het veldonderzoek

In tabel 1.2 zijn de bedrijven en contactpersonen opgesomd die we op de CeBIT en tijdens ons vooronderzoek hebben ontmoet. Met alle leveranciers is contact gezocht na de beurs en aan de hand daarvan zijn vervolg afspraken gemaakt.
De wensen van de sales- en engineering-afdelingen zijn verwerkt in een concreet engelstalig voorstel (zie bijlage C) en naar geselecteerde leveranciers gestuurd. Dit stelde hen in staat om goed na te denken over onze wensen en zo tot een zo goed mogelijk aanbod te komen. Tabel 1.3 toont de bedrijven die we het voorstel stuurden en de afspraak die hieruit voortvloeide.

Bedrijf	*Gestuurd*	*Contactpersoon*	*Meeting*
IP			
Cisco	22/03	Michiel Bührs	03/04
Juniper	25/03	Rene Roersma	02/04
Riverstone	25/03	John Schaap	08/04
Extreme	25/03	Fred Noordam	16/04
Foundry	27/03	Bernard Schep	-
Lucent	28/03	Stefan van Hal	08/04
WDM			
Sorrento	25/03	Paul Rombouts	26/04
ONI	25/03	Peter Reinders	09/04
FNE	25/03	Brian Sinfield	-
N-Base	25/03	Bram van de Kade	17/04
Optical Access	25/03	Bram van de Kade	17/04
Luminous	25/03	Steve Wines	-
Alcatel	28/03	Wouter van Willenswaard	16/04
Tellabs	29/03	Jan van Hees	-

Tabel 1.3: *De bedrijven die aangeschreven zijn (kolom 2) en de afspraken die daaruit volgden (kolom 4). '-' betekent: geen reactie.*

Alle bovenstaande bedrijven werd ons document gestuurd. Helaas heeft slechts

één bedrijf serieus gekeken naar het document. Dit was Cisco Systems. Zij kwamen echter met een derde oplossing aan, de DPT ring technologie.
Er kan dus gesteld worden dat een vergelijkingsonderzoek gebaseerd op een mathematisch principe zoals een vergelijkingsmatrix (waarbij we de score van de punten vermenigvuldigen met ons eraan gehechte belang en dit sommeren) niet haalbaar is. Derhalve is onze taktiek in overleg met de opdrachtgever veranderd in een waarbij we veel waarde hechten aan het gevoel wat we krijgen omtrent de apparatuur, de overtuigingskracht van de engineers en de sales heren.
In het publieke document zijn de offertes en prijsbepalingen weggelaten. Indien deze informatie noodzakelijk is, gelieve contact op te nemen met de directie van Business Internet Trends BV.

Cisco Systems (CWDM)

Contactpersonen: Robert Palse en Michiel Bürhs
Voor een gigabit gebaseerd CWDM netwerk stelt Cisco voor gebruik te maken van standaard Cisco GBICs die in de verschillende 20nm kanalen zijn ingesteld, samen met een optical add/drop module (de OADM). Het powerbudget per GBIC is 32 dB, een forse jongen dus. De GBIC wordt ondersteund in de 2948G, 2980G-A, 2950, 3550, 4000 en 6500 platforms. De (passieve!) OADMs, kunnen één, vier of acht lambda's van de SMF vezel affilteren.
Voor de IP laag wordt een Catalyst4006 aanbevolen. Deze volwaardige L2/L3/L4 router bevat 2xGigE poorten (één naar de WDM, één naar het locale netwerk)

Cisco Systems (DWDM)

Contactpersonen: Robert Palse en Michiel Bürhs
Voor een DWDM oplossing biedt Cisco Systems één oplossing, gebaseerd op het ONS15454 systeem, een DWDM machine. Er wordt voorzien in een dedicated OC48 (2.5 Gbps) IP verbinding, 6 GigE verbindingen en 24 FE verbindingen. De oplossing voorziet een DWDM verbinding tussen Ede en Telecity2, en tussen Telecity2 en Sara. Er worden dus drie machines ingezet. De DWDM trunk bevat een power budget van 26 dB, een enigszins kritische waarde.
Op elke locatie wordt een Cisco 12404 router geplaatst voor de IP connectiviteit. De router bevat een 3-ports GigE lijnkaart, waarvan één GigE aan de WDM machine wordt gekoppeld en twee poorten vrij zijn voor het locale netwerk.

Cisco Systems (DPT)

Contactpersonen: Robert Palse en Michiel Bürhs
Het paradepaardje van de Cisco engineers, is een Dynamic Packet Transport (*DPT*) ring. Met deze oplossing komen twee problemen. Allereerst zijn de lijnkaarten van de DPT apparaten uitgerust met een powerbudget van 25 dB, welke zeer kritisch is voor onze toepassing. Ten tweede, een zeer belangrijk aspect, is de DPT technologie enkel op STM4 en hoger van toepassing. Wij hebben op twee van de drie lengtes dark fiber, maar de derde is een STM1, welke dan geupgrade moet worden naar STM4 om een totale doorvoer over de ring van 622 Mbps te verkrijgen. Het ontwerp voorziet op elke locatie twee routers, een 'core router' (een GSR12404) en een 'edge router' (een c7606). Een en ander hangt af van het standpunt van Cisco dat men

geen klanten zou moeten koppelen aan de core, iets wat voor ons geen probleem is. Omdat de prijs van deze oplossing dermate hoog is, wordt hij verworpen.

Juniper (IP)

Contactpersonen: Rene Roersma en Jean-Marc Haye
De m-series routers van Juniper hebben hun kwaliteit bewezen. Belangrijk in deze zijn veld-referenties van door ons bekende netwerk engineers die al klant zijn. De architectuur is volledig losgekoppeld en de software met zijn features draait gegarandeerd op elke denkbare interface. Ook wordt MPLS volledig ondersteund, waardoor linkredundantie en her-routering kan worden bereikt binnen 50 ms.
Omdat alle machines van Juniper dezelfde features hebben, is voor ons alleen de hoeveelheid beschikbare 'PIC' (physical interface connector) slots van belang. De M10 bevat 8 slots en is derhalve ruim genoeg.
Gedurende het onderzoeksproces is veel waarde gehecht aan de bereidheid van Juniper sales, alsook collega engineers, om tekst en uitleg te geven over de producten, waardoor er veel vertrouwen in Juniper is.

Riverstone (IP)

Contactpersonen: John Schaap en Roy Weijburg
Riverstone is een bedrijf wat zich positioneert in de metro en datacenter markt. De machines zijn, volgens henzelf, niet bedoeld voor core-routing activiteiten, die men vaak ziet bij ISPs. Wel zijn ze uitermate geschikt voor interne netwerken bij datacenters, metropole bedrijven en andere toepassingen zoals Internet Exchanges. Dit wordt genoteerd voor de keuze van leverancier van ons IX project. Voor de core routing in het nieuwe netwerk ontwerp, geeft Riverstone zelf aan niet de beste keuze te zijn. Bovendien is de support voor features die we erg belangrijk achten nagenoeg niet aanwezig. Zo kunnen de machines geen GRE, geen tunneling, geen IPv6 en geen SSH.

Extreme (Ethernet)

Contactpersonen: Fred Noordam en Luuk Dries
Extreme is een bedrijf dat zich in de ethernet-markt positioneert. Van Noordam komt het idee om het paar vezels tussen Ede en Telecity2 te gebruiken om twee onafhankelijke circuits te vormen, gebruik makend van een innovatief splitter/combiner apparaat welke van een singlemode 1550nm GigE verbinding de twee inkomende vezels combineert tot één uitgaande vezel. Met behulp van twee high-power GBICs kunnen dan op ons vezelpaar twee GigE verbindingen tot stand gebracht worden. Een en ander wordt vergemakkelijkt door het gebruik van GBICs die in allerlei apparatuur passen, maar er is geen groeimogelijkheid. We hebben wel een goed idee van de beschikbare apparatuur van Extreme en het bijbehorende kostenplaatje.

Lucent (SDH/Ethernet)

Contactpersonen: Peter Reinders en Stefan van Hal (Landis)

Ondanks het motiverende gesprek van enkele Lucent engineers op de CeBIT, heeft Landis (als Reseller) zijn plicht voldaan en gemeld dat volgens hen Lucent geen goede partij is voor dit specifieke netwerk ontwerp. Dit hangt samen met de telco-insteek van Lucent, die Ethernet over SDH infrastructuur mogelijk wil maken. Aangezien onze darkfiber paren voor elke mogelijke technologie kunnen worden ingezet, hebben ze ons aangeraden te kijken naar WDM aanbiedingen. Zelf brachten ze ONI als potentiële leverancier naar voren. Dat aanbod wordt verderop in deze sectie besproken.

Optical Access (WDM)

Contactpersonen: Bram van der Kade en Ronald van der Meer
Optical Access is een amerikaans bedrijf dat zich in de ethernet en WDM markt positioneert. Van hen hebben we twee aanbiedingen voor CWDM connectiviteit gebaseerd op een eigen product (LD800) en dat van een zusterbedrijf, NBase-Xyplex (WDM84).
Optie 1 - Optical Access LD800 - CWDM oplossing
Er wordt gebruik gemaakt van een 8-slots chassis, waarin een management module, een multiplexer en een demultiplexer voorzien zijn. Er is ruimte voor acht GigE transponders. Het optisch budget ligt tussen de 25 en 26 dB en is dus vrij kritisch.
Optie 2 - FiberDriver WDM84 - CWDM oplossing.
In deze oplossing wordt gebruik gemaakt van een vaste WDM kaart, die niet modulair uitbreidbaar is, maar wel over 4 GigE kanalen beschikt en een voldoende ruim powerbudget heeft voor 100km.
Een en ander zal afhangen van het meetrapport van Level3. Deze meting wordt verricht zodra de verbinding met het pand in Ede tot stand is gebracht.

Alcatel (WDM)

Contactpersonen: Wouter van Willenswaard en Marcel Derogée
Alcatel biedt vier mogelijkheden, allen gebaseerd op de 1696 SM terminal. Elke optie biedt ten opzichte van zijn voorganger een extra toekomstgerichte uitbreiding.
Optie 1 - Onversterkt, zonder mux/demux en met een 4-any board
Er worden twee GigE connecties gerealiseerd op één lambda. Als de behoefte aan gigabit kanalen bij twee blijft, is er geen behoefte aan een lambda multiplexer (die dus meerdere lambda's over één glasvezelpaar stuurt). Tevens is de reikwijdte van deze machine zonder mux/demux kaart ruim 100 km.
Optie 2 - Onversterkt, met mux/demux en met een 4-any board
In deze optie wordt het signaal uit het 4-any board (met daarin dus 2x GigE) door een mux/demux kaart geleid waardoor op een later tijdstip nog kanalen bijgeschakeld kunnen worden. Het signaal wordt hierdoor enigszins verzwakt en zal een power budget van 23 dB aanhouden. Deze optie wordt verworpen wegens linkbudget.
Optie 3 - Versterkt, met mux/demux en met een 4-any board
Hier worden de twee GigE connecties uit optie 1 eerst door een mux/demux kaart geleid, waardoor er later zonder problemen nog lambda's bijgeschakeld kunnen worden. Aangezien de mux/demux het signaal enigszins verzwakt, zal er ook een versterker aangebracht worden, waardoor de 90km met een ruim budget zal kunnen worden overbrugd.

Optie 4 - Onversterkt, met mux/demux maar zonder 4-any board
Hier wordt er op één lambda ook maar één GigE connectie verzorgd, welke door een mux/demux kaart wordt uitgezonden op de langeafstandsvezel. Hierdoor wordt op een later tijdstip het bijschakelen van gigabit kanalen vergemakkelijkt, maar er wordt begonnen met slechts één gigabit kanaal. Deze oplossing wordt verworpen omdat hij maar één GigE koppeling voorziet.

ONI (WDM)

Contactpersonen: Peter Reinders en Stefan van Hal (Landis)
Door Landis worden vier opties voorgesteld, allen gebaseerd op ONI WDM technologie. Deze leverancier en technologie heeft volgens Landis de voorkeur boven de SDH framing technologie van Lucent, welke zij ook leveren.
Optie 1 - Online2500 met 2xGigE datamux zonder OADM (CWDM)
In deze oplossing worden twee GigE signalen direct door de datamux kaart getransporteerd zonder gebruik te maken van een optische add/drop multiplexer (OADM). In deze oplossing kunnen maximaal twee GigE signalen getransporteerd worden omdat er maar één lambda beschikbaar is. Er wordt dan gebruik gemaakt van een GBIC met golflengte van 1550nm.
Optie 2 - Online2500 met 4xGigE datamux en een 4OADM (CWDM)
In deze oplossing worden vier GigE signalen getransporteerd op twee lambda's. Hiervoor zijn optische add drop multiplexers (OADM) nodig. Het 1OADM kan één lambda uit de WDM bundel splitsen en derhalve zouden er om twee lambda's uit te splitsen twee 1OADMs nodig zijn. Er is ook een 4OADM kaart, die 4 lambda's uitsplitst. De oplossing heeft een powerbudget van 27.2 dB
Optie 3 - Online7000 met 2xGigE datamux zonder OADM (DWDM)
Deze oplossing is gebaseerd op de Online 7000 apparatuur van ONI Systems. De datamux kaart biedt 2xGigE en kan in de toekomst worden uitgebreid door meerdere datamux kaarten toe te passen, die ieder op een aparte lambda van 2.5Gbps communiceren. De demping van de verbinding van Level3 levert geen problemen op doordat de Online7000 gebruik maakt van een variabele post amplifier. Deze vangt fluctuaties in het optische vermogen door herconfiguratie dynamisch op.
Optie 4 - Online7000 met 4xGigE datamux en een OADM (DWDM)
Deze oplossing is gelijk aan optie 3, waarbij er in plaats van één, twee datamux kaarten worden gebruikt, zodat er 4 GigE kanalen kunnen worden getransporteerd op twee lambda's. Ook hier draagt de Online7000 zorg voor een zuivere transmissie over 90km.

1.4.7 De uitkomsten

De IP laag

Voor de IP afhandeling stelden we stricte eisen. De engineering afdeling heeft te kennen gegeven veel waarde te hechten aan MPLS, IPv6 en de mogelijkheid tot ssh en ip(v6)ip tunneling. Om deze redenen kwamen er maar twee leveranciers door de initiële kennismakingsronde heen: Cisco en Juniper.
Beide leveranciers hebben we goed aan de tand gevoeld en zelf met een voorstel laten komen. Cisco kwam met een indrukwekkend document waarin ze ons een

DPT ring aanboden (welke in de praktijk nauwelijks uitvoerbaar is). Hun IP aanbiedingen waren gebaseerd op drie Catalyst4006 routers, welke ons onvoldoende krachtig lijken. Als alternatief op het c4000 platform stelde men drie GSR12404s voor. Deze bevatten dan drie Gigabit kaarten en zijn typisch bedoeld voor een ring configuratie
Juniper stelt drie M5s voor, ieder uitgerust met GigE modules en twee ervan met STM1 modules. Helaas komen deze modules enkel per 2 op een PIC, waardoor de prijs voor één poort prijzig is. De GigE poorten zijn vergelijkbaar in prijs (ten opzichte van de 12404 van Cisco) en omwille van de geboden featureset (in het bijzonder IPv6 en MPLS) verdient het Juniper platform de voorkeur.

De transport laag

Losstaand van de keuze voor een leverancier van de IP hardware, moet een leverancier gekozen worden voor de transportlaag. We hebben gezien dat de DWDM systemen per GigE kanaal flink duurder zijn dan de CWDM systemen en derhalve gaat onze voorkeur uit naar de laatste.

Deze tabel is in de publieke versie weggelaten.

Tabel 1.4: *De prijsverhouding van de verschillende WDM leveranciers, de 'per GigE' is de prijs per verbinding die betaald moet worden in een volledig chassis*

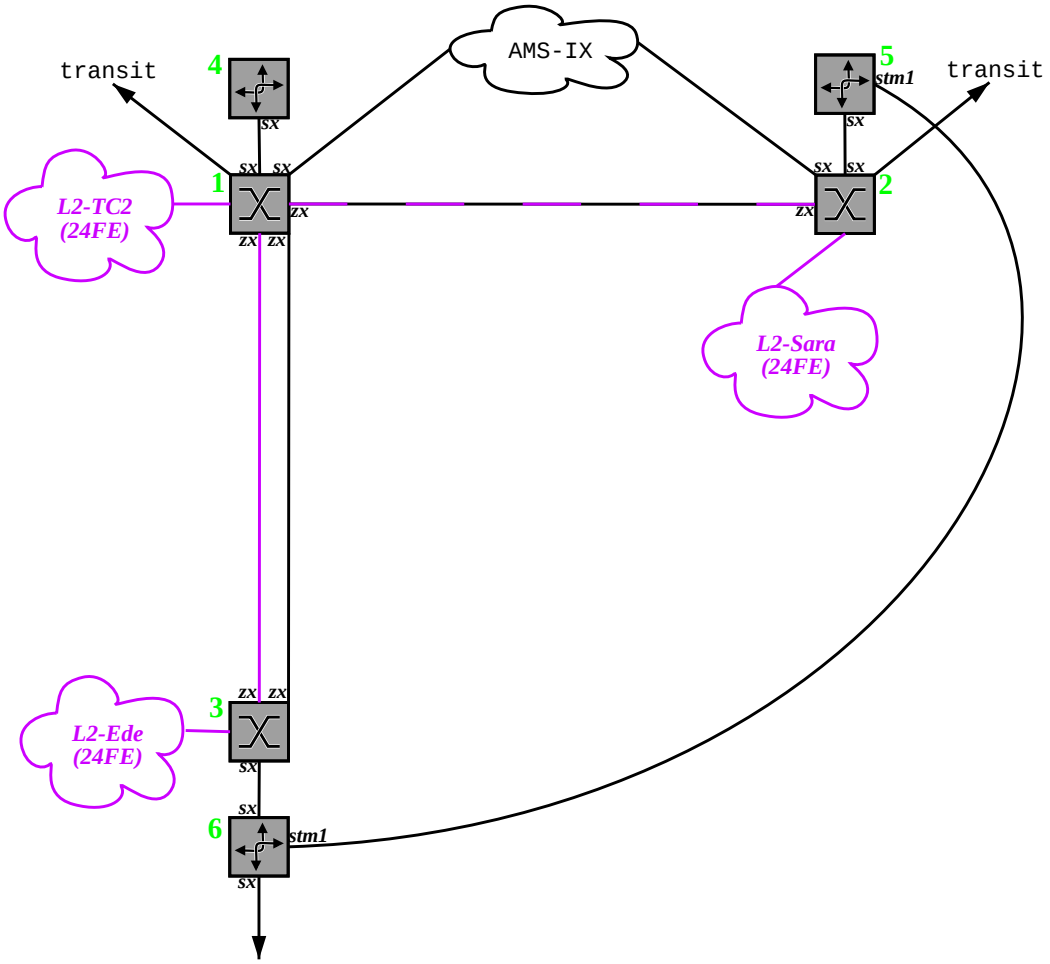
Voor de Ede-Telecity2 verbinding is het niet nodig om een DWDM systeem te ontplooien. De system zijn erg duur en bovendien is DWDM gespecificeerd voor 32 of meer kanalen, die we niet zullen benutten in dit netwerkontwerp.
De CWDM oplossing van Cisco voldoet prima aan link specificaties, maar heeft GBICs, waarvoor de Juniper geen interface modules heeft. Als we deze oplossing kiezen, zullen we alsnog een switch moeten aanschaffen met twee GBICs erin (de WDM GBIC en een standaard 1000baseSX GBIC). Dit brengt twee nadelen met zich mee. Enerzijds is het een duurder ontwerp, anderzijds introduceren we hierdoor een extra kritiek punt. Immers als deze switch uitvalt, ontkoppelt de gehele verbinding.
De CWDM oplossing van Optical Access heeft ten opzichte van voorgaande als voordeel dat de uitgangen standaard van het type 1000baseSX zijn. Hierdoor kunnen we de Juniper direct aan de machine hangen. Dit lost het kritieke punt hierboven op. Een ander voordeel is dat de LD800 in staat is om maximaal 8 GigE kanalen op te zetten. Hierdoor verdient deze oplossing de voorkeur, mits we CWDM kiezen en blijkt dat de splitter/combiner truuk, die Optical Access ook kan leveren, niet wenselijk is of niet implementeerbaar is.

1.5 Ontwerpvoorstellen

Uit alle mogelijke ontwerpen die de revue gepasseerd zijn, in totaal over de 15, blijken er drie de moeite waard te zijn om voor te dragen aan de directie van Business Internet Trends. Allereerst diepen we het ontwerp uit gebaseerd op splitter/combiner technologie, waarmee we zeer goedkoop twee GigE kanalen kunnen transporteren.
De CWDM oplossing van Cisco is gebaseerd op GBICs, die van OpticalAccess niet. Hierdoor verschillen beide oplossingen behoorlijk in termen van benodigde switchpoorten en connectiviteit.
We spelen eerst met drie verschillende ontwerpen, die gedreven zijn vanuit verschillende perspectieven. Daarna sommen we de plus- en minpunten op van deze ontwerpen en dragen we een ontwerp voor, wat volgens ons zo goed mogelijk de functionaliteit biedt die we zoeken in het ontwerp.
Hierbij zal niet alleen het budget een rol spelen, ook de ervaringen van de engineers van Business Internet Trends, verwoord door de domeindeskundige (ALEX BIK), zullen hierbij een rol spelen.
We beperken ons in deze ontwerpen tot enkele leveranciers van switches en routers. In voorgaande paragrafen trokken we reeds conclusies over deze leveranciers, veelal door formele en informele gesprekken met techneuten, vertegenwoordigers en collegae uit het veld.
Ondanks dat het een moeilijke taak blijkt om één voorstel te kiezen hebben we toch geprobeerd om, door middel van technische en gevoelsmatige afwegingen, een beredeneerde voordracht te doen.
Eventueel kan hierbij een herberekening gemaakt worden op basis van andere leveranciers, maar ons voorstel voor een layer2 deel gebaseerd op CWDM, alsmede een layer3 netwerk gebaseerd op Juniper blijft staan.

1.5.1 Gebaseerd op Ethernet

Dit netwerk maakt gebruik van de splitter/combiner technologie, die in staat is om een vezelpaar te combineren tot één single mode vezel. Hierdoor kunnen we ons vezelpaar tussen Ede en Telecity2 gebruiken voor twee GigE verbindingen. Het ontwerp bestaat uit een layer2 gedeelte, gebaseerd op Extreme Alpine switches, met daarop Juniper routers. Alle koppelingen zijn gigabit. De switches kunnen ook gebruikt worden om de Junipers te trunken (zodat we maar één 1000baseSX module nodig hebben per router).



Figuur 1.7: *Een mogelijk ontwerp gebaseerd op splitter/combiner technologie van NBase Xyplex, met een demping van 6.4 dB. We gebruiken 30 dB GBICs waardoor 23.6 dB aan linkbudget overblijft.*

In dit ontwerp hebben we drie switches (1, 2 en 3) en drie routers (4, 5 en 6). We hebben de volgende bouwstenen nodig voor de switches en routers die we in figuur 1.7 zien.

Deze tabel is in de publieke versie van het document weggelaten.

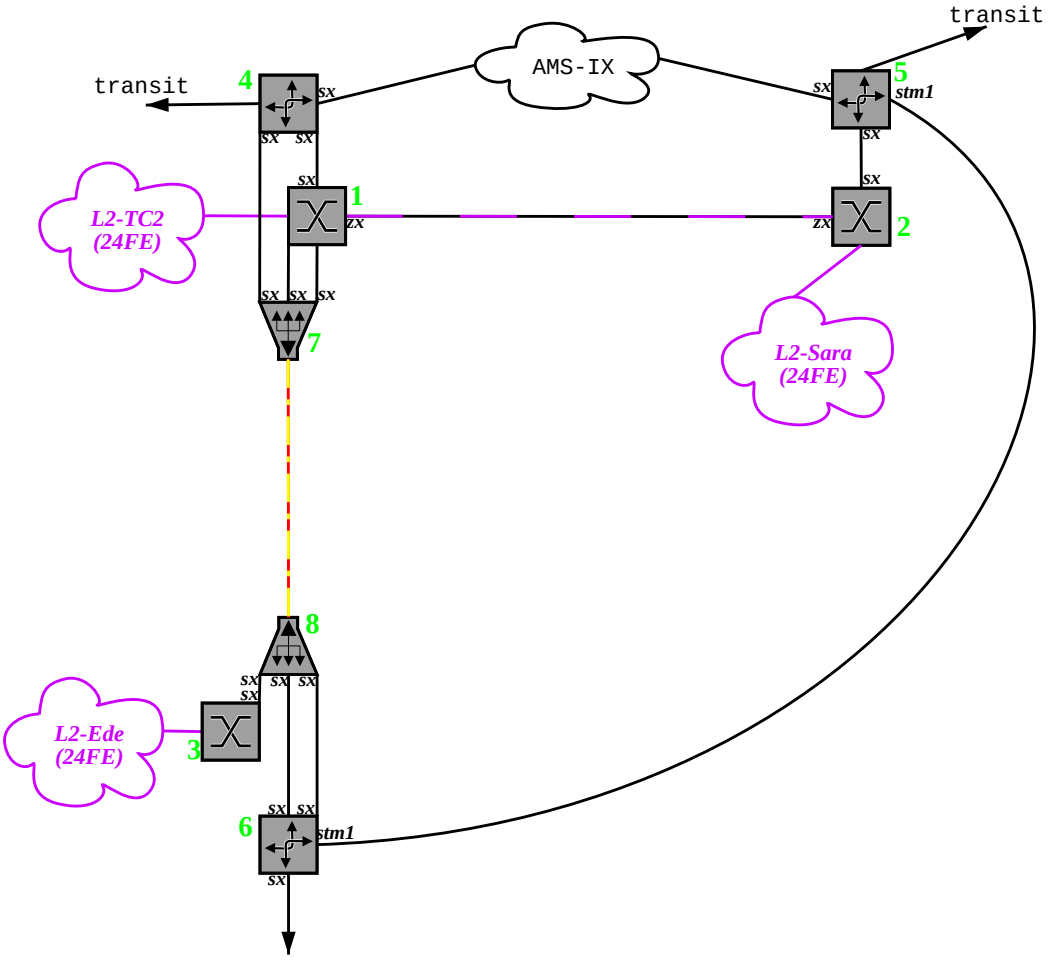
Tabel 1.5: *De benodigde hardware voor het netwerk, gebaseerd op splitter/combiner oplossing.*

Voor- en nadelen

Het duidelijke voordeel van deze oplossing is dat het relatief goedkope GBICs gebruikt om de transmissie te verzorgen. Uiteraard is de totale hoeveelheid GigE kanalen tussen Ede en Telecity2 fysiek begrensd op twee, omdat er maar twee SMF vezels zijn. Bovendien is het vermogen van de GBICs (30 dB) verminderd met de demping van de splitter/combiners (6.4 dB per verbinding) behoorlijk kritisch op onze verbinding. Hierdoor zijn we terughoudend met het voorstel en draaien we liever met iets meer marge. De splitter/combiner truuk kunnen we echter prima uithalen in Amsterdam zelf, mochten we daar behoefte hebben aan twee GigE kanalen tussen Sara en Telecity2.

1.5.2 Gebaseerd op CWDM, zonder switch afhankelijkheid

Dit netwerk maakt gebruik van de LD800 CWDM oplossing van Optical Access. Er worden drie GigE kanalen (van de acht mogelijke) ingezet en de Junipers worden direct aan de WDM transponder gekoppeld, wat de kans op storing vermindert.



Figuur 1.8: *Een mogelijk ontwerp gebaseerd op CWDM van Optical Access, waarbij we de Junipers direct op de WDM transponder, het shared medium en de transit providers aansluiten.*

In dit ontwerp hebben we drie switches (1, 2 en 3) en drie routers (4, 5 en 6). Voor de langeafstandsverbinding Ede-Telecity2 hebben we twee actieve WDM chassis nodig (7 en 8) waar aan iedere zijde 3 1000baseSX worden getermineerd. We hebben de volgende bouwstenen nodig voor de switches en routers die we in figuur 1.8 zien.

Deze tabel is in de publieke versie van het document weggelaten.

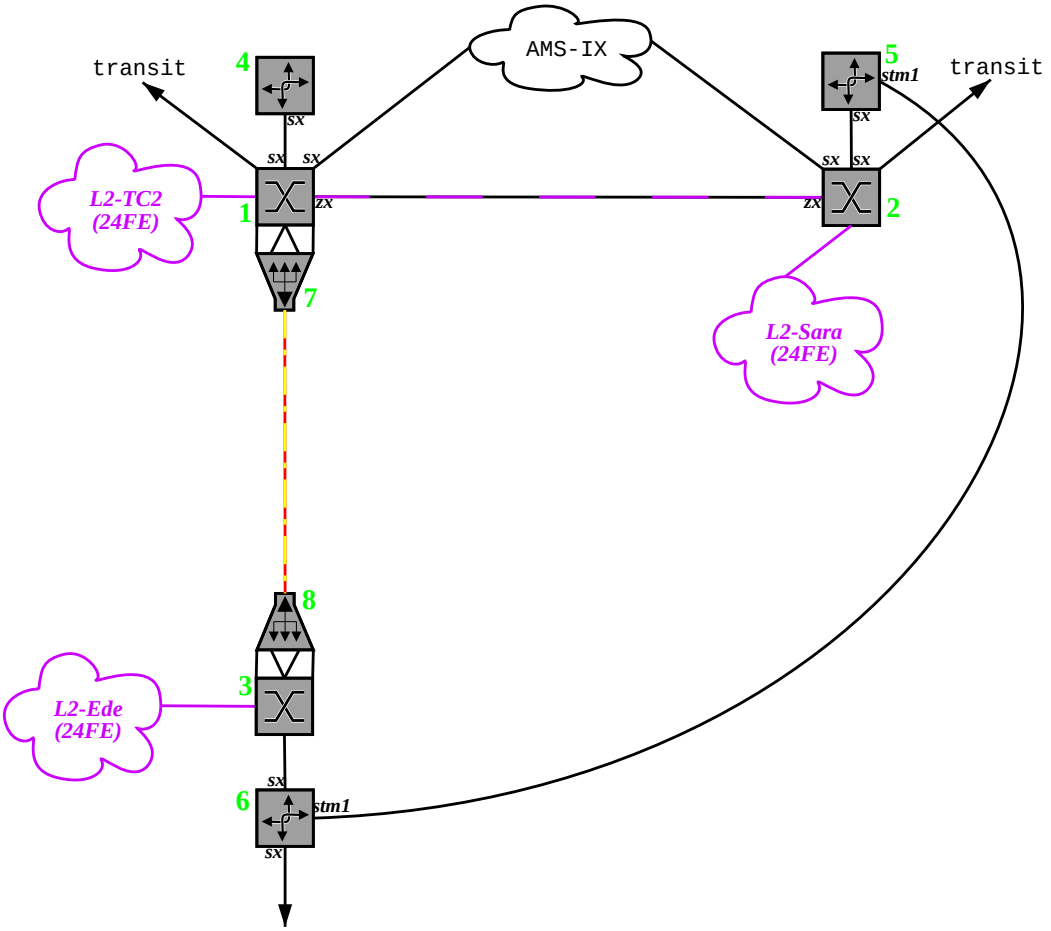
Tabel 1.6: *De benodigde hardware voor het netwerk, gebaseerd op de CWDM oplossing van Optical Access*

Voor- en nadelen

Het gebruik van de LD800 biedt de mogelijkheid tot 8 GigE kanalen, meer dan de Cisco oplossing (die er 4 biedt). De routers staan in directe verbinding over de CWDM infrastructuur. Hierdoor kunnen ze elkaar, het shared medium en de transit providers bereiken ondanks enige uitval van de switches.
Het aansluiten van de Junipers aan het shared medium en de transit providers (4xFE per locatie) is prijzig omdat de interfaces voor Junipers veel duurder zijn dan die van de switches. Hierdoor vervalt de switch als single point of failure, maar wordt het ontwerp fors duurder.

1.5.3 Gebaseerd op CWDM, met switch afhankelijkheid

Dit netwerk maakt gebruik van de CWDM GBICs van Cisco. Hier wordt aan beide zijden (Ede en Telecity2) een passief WDM-frame met 4 GBICs getermineerd op een Alpine switch, waardoor er een 4 Gbps verbinding tot stand gebracht wordt.



Figuur 1.9: *Een mogelijk ontwerp gebaseerd op CWDM van Cisco, waarbij we 4 paar GBICs implementeren.*

In dit ontwerp hebben we drie switches (1, 2 en 3) en drie routers (4, 5 en 6). Voor de langeafstandsverbinding Ede-Telecity2 hebben we twee passieve WDM chassis nodig (7 en 8) waar aan iedere zijde 4 GBICs worden getermineerd in de Extreme switches. We hebben de volgende bouwstenen nodig voor de switches en routers die we in figuur 1.9 zien.

Deze tabel is in de publieke versie van het document weggelaten.

Tabel 1.7: *De benodigde hardware voor het netwerk, gebaseerd op de CWDM oplossing van Cisco*

Voor- en nadelen

Het voordeel en tevens nadeel van de Cisco WDM oplossing is het gebruik van de GBICs. Deze worden door een passief frame gemultiplexed op de SMF verbinding, wat vrij storings ongevoelig is. De totale hoeveelheid kanalen is begrensd op vier omdat de OADM maar vier kanalen aankan. Echter, de GBICs kunnen niet in de Juniper, waardoor de switch een duidelijke single point of failure zullen zijn in het ontwerp. Het is maar de vraag of we dit kunnen accepteren, hoewel in de praktijk switches (mits als layer2 device gebruikt) maar weinig storingen blijken te hebben. Eventueel kan overwogen worden om de LD800 van Optical Access in te zetten in de plaats van het (passieve) Cisco frame. In prijs maakt dit niet zo veel uit. Er zijn met de LD800 meer GigE kanalen inzetbaar, terwijl er een electrisch component bijkomt.

1.5.4 Verdediging

Na overleg met de domeindeskundige is het vermoeden bevestigd dat de switches in de praktijk minder storingsgevoelig zijn. We moeten dan wel de switches enkel voor layer2 toepassingen gebruiken.
Hierdoor is een ontwerp dat berust op een layer2 transportlaag op alle drie de locaties verdedigbaar. Als de switch in Telecity2 of Ede uitvalt, is er nog een backuppad via de STM1. Als de switch in Sara uitvalt, is er het reguliere pad tussen Telecity2 en Ede nog (en het backup pad met de STM1). Bij uitval van een switch zal enkel de layer2 connectiviteit in gevaar komen. We besluiten dit risico te nemen, om de kosten van het project met ongeveer €60000 te verlichten.
We komen uit op ontwerp 3, getekend in figuur 1.9 en beschreven in tabel 1.7. We merken hierbij het volgende op:

- De WDM oplossing van Cisco kan vervangen worden door die van Optical Access. Het prijsverschil is miniem, maar de mogelijkheden tot uitbreiding zijn groter met de LD800.
- Als de WDM oplossing van Cisco vervangen wordt, kan hij nog prima dienst doen tussen Telecity2 en Sara of tussen Sara en de Mediagateway.
- Als tussen Telecity2 en Sara meer bandbreedte vereist is, kan men overwegen om hier een splitter/combiner paar in te voegen. Dit zal dan 4 splitters en 2 nieuwe ZX100 GBICs vereisen.

Hierdoor lijkt ons dit ontwerp een prima beginpunt voor het nieuwe netwerk.

1.6 De Internet Exchange ideëen

Het starten van een Internet Exchange wordt steeds vaker gedaan. Deze exchanges hebben het moeilijk om zich goed te vestigen omdat er al een redelijk grote exchange operationeel is in Nederland: de AMS-IX.
Hierdoor is het moeilijk om klanten te overtuigen zich aan te sluiten bij het eigen netwerk, in plaats van in Amsterdam.
Er is een afspraak gemaakt met het technisch personeel van de AMS-IX. In hun kantoor op het Westeinde te Amsterdam is gesproken over de mogelijke modellen, waarbij er drie mogelijkheden de revue passeerden:

1. Het starten van een satelliet locatie van AMS-IX in Ede. Hiervoor zullen redundante vezels moeten worden gelegd, Foundry switches moeten worden aangeschaft en alles moet worden geschonken aan de AMS-IX gebruikers vereniging, die beheerd wordt door AMS-IX BV. Dit is vooralsnog niet haalbaar.
2. Het verhuren van poorten in Ede, die door onze ethernet infrastructuur kunnen worden doorgestuurd naar Amsterdam. Hierbij neemt Business Internet Trends poorten af op AMS-IX, die worden gepatched in de switch in Telecity2 en doorgestuurd naar Ede. De klant plugt in Ede zijn apparatuur in de switch en ziet het AMS-IX shared medium via deze poort.
3. Er is een provinciaal project in ontwikkeling waarbij een glasvezelring door enkele Gelderse steden wordt gelegd. Hierbij wordt gedacht aan Apeldoorn,

Arnhem, Nijmegen en Ede/Wageningen. In deze laatste locatie kan Business Internet Trends zich als partner aanbieden en zelfs een gedeelte van de glasvezelring aandragen. We hebben immers al een vezel waarop we met WDM een verbinding kunnen aanbieden.

Deze mogelijkheden moeten nog rijp worden. Vooral de laatste, zeer interresante mogelijkheid moet nog voorbereid worden door de provincie. In ons ontwerp is de tweede mogelijkheid in orde en zowel het technisch personeel alsmede de directie van AMS-IX zijn akkoord met deze zogenaamde *backhaul* van poorten naar Ede. Er zullen wel goede afspraken omtrent service levels moeten worden afgesproken tussen de klant, Business Internet Trends, en de AMS-IX. Optie 1 wordt vooralsnog verworpen wegens de kostprijs van de eisen voor dit business model. Er zal niet gemakkelijk winst kunnen worden gemaakt met deze oplossing, met de andere twee wel.

1.7 Conclusies en aanbevelingen

Bij het onderzoeken van de huidige situatie in het Business Internet Trends netwerk, zijn de volgende zaken opgemerkt.

- De totale capaciteit van de huidige uitgaande verbindingen (189Mbps) kan niet worden ontkoppeld in het lokale netwerk (132Mbps).
- Het netwerk in Ede is niet redundant gekoppeld, waardoor het uitvallen van *bfr*, de Cisco in Ede, het gehele netwerk ontkoppelt van het Internet.
- Het netwerk moet beter in VLANs worden opgedeeld. Vooral het gebruik van hetzelfde VLAN voor eigen productieservers en klantenmachines in de colocatie is onverstandig.
- De servers van het bedijf zelf kunnen wellicht beter in een fysiek andere ruimte worden aangesloten, in verband met overzicht en veiligheid.

We pakken deze zaken aan in een herontwerp van het netwerk in een volgend hoofdstuk.
Bij het veldonderzoek spraken met verschillende leveranciers (in totaal zestien stuks). De volgende apparaten zijn na zorgvuldige analyse preferent bevonden in het ontwerp voor het nieuwe netwerk.

- Het Juniper platform voor IP connectiviteit
- Extreme Networks Alpine 3800 series switch voor ethernet connectiviteit
- De Optical Access LD800 machine, of het passieve Cisco systeem voor CWDM verbindingen
- De NBase Xyplex optische Splitter/Combiner voor SMF verbindingen over éń vezel

Wij stellen het ontwerp voor wat getekend is in figuur 1.9 en waarvan de apparatuurlijst (met kosten) in tabel 1.7 beschreven is.

Hoofdstuk 2

Netwerk ontwerp

2.1 Inleiding

In het vorige hoofdstuk is uitgebreid aandacht besteed aan de vereisten van het nieuwe netwerk, alsook de potentiële leveranciers. Uit dit onderzoek kwamen enkele merken naar voren. In dit hoofdstuk wordt een concreet netwerk ontworpen aan de hand van deze leveranciers. Het resultaat is een fysiek ontwerp, inclusief een financiëel budget, waarmee de nieuwe generatie backbone kan worden gebouwd.
Daarna zal gekeken worden naar de logische indeling van het netwerk. Hiermee bedoelen we de interne en extrene routingprotocollen, de numberplans en de gebruikte VLANs in het backbone netwerk.
Het resultaat is een netwerkontwerp voor de kern van het nieuwe netwerk, de backbone. Hoe we het huidige netwerk in de nieuwe backbone zullen schuiven, stellen we uit tot een volgend hoofdstuk (het migratieplan).
De configuratie van de netwerkbouwstenen zal geschieden in de implementatie fase. Hiervoor is uiteraard de behoefte aan de apparaten, welke in deze fase nog niet besteld zijn.

2.2 Probleem- en doelstelling

Voor dit gedeelte van het project definiëren we vijf problemen, die structureel worden aangepakt in dit hoofdstuk. Het doel van dit gedeelte is het oplossen van de hierna gestelde problemen.

2.2.1 Probleemstelling

- Het huidige netwerk is niet redundant omdat er maar één router in Ede aanwezig is. Deze vormt een single point of failure.
- Er is geen kennis aanwezig over het interne routeringsprotocol OSPF. Deze zal nodig zijn in het nieuwe ontwerp.
- Er is nog geen concreet budget opgesteld voor de benodigde apparatuur. Deze is nodig voor een financieringsaanvraag bij de bank.

- Sommige klanten stellen extreem hoge eisen aan de beschikbaarheid van het Business Internet Trends LAN. We hebben momenteel geen oplossing voor redundantie van interne netwerken.
- Sommige klanten wensen een hoge beschikbaarheid van hun webpages. Hiervoor zal ook server-redundantie benodigd zijn. We hebben momenteel geen oplossing voor redundantie van (klanten)servers.

2.2.2 Doelstelling

Er wordt een corenetwerk ontworpen dat een robuuste dienstverlening kan bieden voor internetverkeer tussen Ede en Amsterdam. Het uitvallen van één der routers moet automatisch worden hersteld.
We zullen kijken naar het meest gebruikte interne routeringsprotocol (OSPF).
We bieden een oplossing voor de netwerk-redundantie die sommige klanten van het bedrijf eisen (VRRP).
We zullen een oplossing presenteren voor de server-redundantie die gewenst is bij enkele van de klanten van Business Internet Trends (LSNAT).
De diensten van de ISP worden nogmaals in kaart gebracht en in VLANs ingedeeld. Deze VLANs worden structureel in zogenaamde numberplans gevat. Hierdoor is er geen ambiguïteit in de switchconfiguratie mogelijk. Tevens wordt een plan gemaakt voor de IP indeling van de nieuwe backbone en locale diensten (zoals mail-, news-, en nameservers).
Uiteraard zal het nieuwe ontwerp alle functionaliteit van het huidige netwerk moeten kunnen bieden. Daarnaast zal het ontwerp de activiteiten van een Internet Exchange moeten kunnen ondersteunen en de klantenwensen in de toekomst kunnen implementeren.

2.3 Technieken en Methodes

Zoals in het gehele project, zal *dia* worden gebruikt voor het tekenen van netwerk schema's. De tekeningen die gemaakt zijn in het onderzoek zullen verder gedetailleerd worden in het definitieve ontwerp. Hierdoor zijn niet alleen de lokale en langeafstands verbindingen in kaart gebracht, maar ook de concrete netwerk bouwstenen opgesomd.
Van de benodigde apparatuur wordt beredeneerd waarom er behoefte aan is en de prijzen worden opgezocht. Hierdoor ontstaat een verwacht budget dat gebruikt kan worden door de directie en financiële dienst van Business Internet Trends in hun voorstel naar de bank.
Voor de concrete indeling van de switches en routers maken we gebruik van numberplanning. Dit zijn tabellen met VLANs en IP indeling. Deze worden aan het einde van het ontwerp gepresenteerd en ze vergemakkelijken het implementeren van het netwerk in een volgende fase.

2.4 Resultaten

We delen het ontwerp dus in twee delen in:

1. De basale layer2 en layer3 backbone. Deze is noodzakelijk en zal meteen worden opgebouwd. Dit zijn de POPs in Amsterdam (Telecity2 en Sara) en die in Ede-HQ. Ze zijn genummerd van 1-3. Daarnaast zal, als het nieuwe gebouw af is, de vierde POP moeten worden gemaakt (Ede-Colo). De optionele toevoegingen

2. . Deze sites zullen worden bijgebouwd naar behoefte van bepaalde klanten. Wanneer dit gebeurt is nog onduidelijk. Wel zullen we nu al rekening houden met het aankoppelen van dergelijke projecten in ons core netwerk.

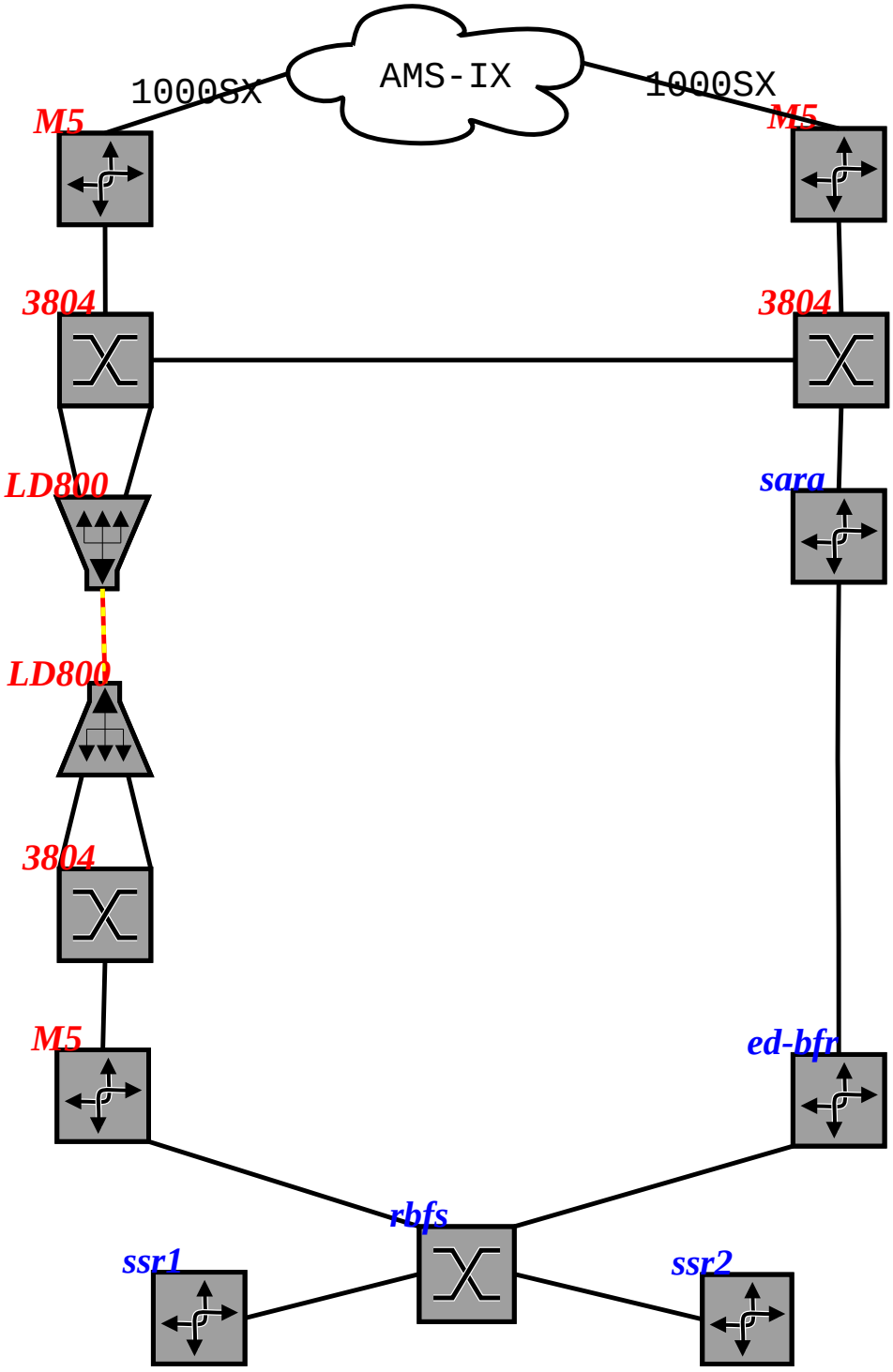
Met HENK VAN BEEK, de financiële directeur van Business Internet Trends, wordt afgesproken wat hij nodig heeft voor een financieringsplan. Dit komt redelijk nauw, omdat in juni 2002 de conjunctuur voor ISPs laag staat en in de directe omgeving van de ISP/ICT branche veel partijen omvallen. De banken zijn dus op dit moment niet zo happig op leningen en financieringen.
Enerzijds zullen de kosten dus zo laag mogelijk moeten worden gehouden, anderzijds moeten we zo lang mogelijk aanschaffen uitstellen. Zo wordt er op 12 juni afgesproken dat we het backbone netwerk daadwerkelijk zullen bouwen in twee stappen. In october komen de eerste drie POPs, met de CWDM verbinding. Een half jaar later, in februari 2003, zal de vierde POP (Ede-Colo) en een tweede CWDM verbinding worden aangeschaft.
Er blijkt inmiddels namelijk meer interesse te zijn vanuit de telecom hoek om een tweede dark fiber van Ede naar Amsterdam te leveren. In juni heeft Telecom Utrecht een offerte gedaan aan Business Internet Trends met een geldigheid van zes maanden, voor een fiberpaar tussen Sara en de nieuwbouw.
Het verkrijgen van een tweede fiberpaar zal de STM1 verbinding overbodig maken. We hebben dan immers de beschikking over een ring van dark fiber. Hierdoor verandert op het laatste moment het ontwerp een beetje. We stellen vast dat het netwerk in twee delen zal worden aangeschaft. Deze stadia zijn in figuur 2.1 en 2.2 te zien. Om deze reden is het aanschaffen van STM modules voor de nieuwe Junipers zonde van het geld. We zullen dus proberen de huidige Cisco's te blijven gebruiken voor de STM1 verbinding (fig 2.1). Later, als we de vezelring rond hebben, kunnen deze Cisco's verwijderd worden en zal het backup pad voor de ISP via deze vezels lopen (fig 2.2). De STM1 van Priority Telecom kan dan worden opgezegd.

2.5 Het fysieke ontwerp

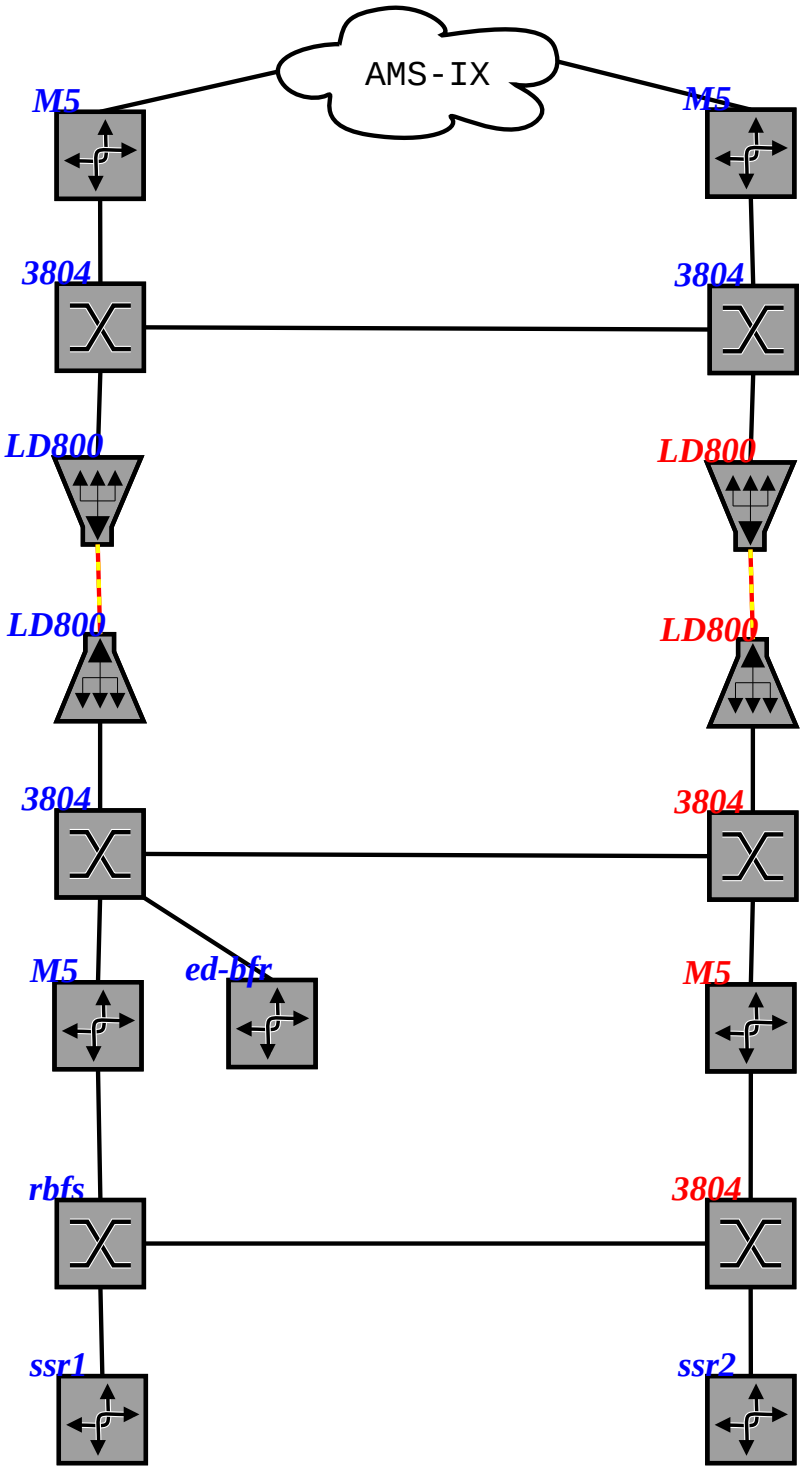
Omdat we in twee delen het netwerk bouwen, zijn er twee relevante figuren van de backbone situatie. In figuur 2.1 zien we het eerste deel, welke gerealiseerd moet worden in october van 2002. Figuur 2.2 toont het netwerk als de werken voltooid zijn, wat verwacht is in februari 2003.

2.5.1 De Core Backbone, fase 1

We gaan ervan uit dat we per direct beginnen met de bouw van de eerste drie locaties. Het nieuwe pand is nog niet af en die apparatuur wordt pas in februari 2003 besteld. We hebben echter nu al wel behoefte aan de STM1 verbinding als backup, mocht de CWDM verbinding falen. We besluiten hiervoor de twee huidige Cisco 7206's te gebruiken. Dit bespaart ons een investering in STM1 modules voor



Figuur 2.1: *Het core netwerk ontwerp, deel 1. De gestippelde vierkanten stellen fysieke locaties voor. In groen de locatie- en apparaatnummer, in rood de typenummers van de aan te schaffen apparatuur (blauw is reeds bestaande apparatuur). Het wolkje stelt AMS-IX voor.*



Figuur 2.2: *Het core netwerk ontwerp, deel 2. De gestippelde vierkanten stellen fysieke locaties voor. In groen de locatie- en apparaatnummer, in rood de typenummers van de aan te schaffen apparatuur (blauw is reeds bestaande apparatuur). Het wolkje stelt AMS-IX voor.*

de Junipers, die in 2003 overbodig zouden worden als het tweede vezelpaar wordt opgeleverd.

2.5.2 De Core Backbone, fase 2

Bij het opbouwen van de nieuwe colocatie faciliteit, zal de vierde POP worden geconstrueerd als verlengstuk van de derde. De planning van de directie van Business Internet Trends voorziet tegen die tijd een tweede vezelpaar tussen het nieuwe pand en het Sara gebouw in Amsterdam. Hier zal dus transmissie apparatuur moeten worden voorzien. Zie figuur 2.2.

2.5.3 De optionele POPs

Er zijn, naast de vier hoofdlocaties, nog vijf andere locaties denkbaar. De belangrijkste van deze locaties is de Internet Exchange die we in Ede gaan bouwen. Daarnaast is er sprake van uitbreiding van ons netwerk naar andere Internet Exchanges, zoals de BNIX in Brussel, en LoNAP in Londen.

2.5.4 Samenvatting

In totaal zullen we apparatuur moeten aanschaffen voor het core netwerk, voor twee CWDM verbindingen tussen Telecity2 en Ede-HQ, respectievelijk Sara en Ede-Colo, en 30 glasvezelpatches om alle routers en switches lokaal op elkaar aan te kunnen sluiten.
We bouwen dit netwerk in twee delen. Het grootste deel zal in october 2002 worden opgebouwd. Als het nieuwe pand af is (februari '03), zal daar het netwerk worden ingericht. Tevens wordt dan een tweede langeafstands darkfiber in productie gebracht en redundantie met EAPS aangebracht.

2.6 Het logische ontwerp

In dit gedeelte bespreken we welke protocollen we zullen gebruiken op ons nieuwe netwerk. We merken op dat Extreme een eigen protocol heeft om Ethernet in ringstructuren aan te brengen.
We onderscheiden dan twee types verkeer. Als verkeer naar een bestemming buiten ons netwerk moet worden getransporteerd, zal aan de rand van ons netwerk een beslissing moeten worden genomen. Dit gebeurt door middel van een External Gateway Protocol, een EGP. Wij kiezen voor BGP en schetsen de werking hiervan.
Als verkeer in ons netwerk binnenkomt met een eindbestemming bij een van onze IP nummers, zal besloten moeten worden hoe deze eindbestemming het beste intern bereikt kan worden. Dit geldt ook voor verkeer wat binnen ons netwerk blijft; er zal besloten moeten worden welke router verantwoordelijk is voor de eindbestemming. Een route naar de desbetreffende router moet dan worden gevonden. Dit gebeurt door middel van een Internal Gateway Protocol, een IGP. We tonen de werking van OSPF, de meest gebruikte IGP.
Om hoge beschikbaarheid van het lokale netwerk te kunnen garanderen, beschrijven we tenslotte twee redundantie-protocollen, VRRP en LSNAT.

2.6.1 Ethernet Automatic Protection System (EAPS)

Onze glasvezelring, waarop we Ethernet draaien, wordt een gesloten ring. Dit brengt een probleem met zich mee. Omdat Ethernet eigenlijk niet is ontworpen om ringstructuren mee te bouwen, zal het spanning tree protocol ervoor zorgen dat één van de verbindingen in de ring, te allen tijde geen verkeer zal dragen. Bij het uitvallen van een verbinding die wèl verkeer draagt, zal het spanning tree protocol een herberekening maken en dit duurt erg lang (één minuut) en is dus onacceptabel. Extreme Networks, de fabrikant die we kozen voor onze core switches, heeft een eigen protocol ontwikkeld dat binnen 50 milliseconden een verbroken verbinding kan detecteren *en* omrouteren. Bovendien voorziet EAPS in het gebruiken van alle verbindingen in de ring, zonder dat daarbij één van de verbindingen zal moeten worden uitgeschakeld. Deze technologie heeft wel een prijs. Er is namelijk een extra licentie voor nodig, die per switch betaald moet worden.

2.6.2 Border Gateway Protocol (BGP)

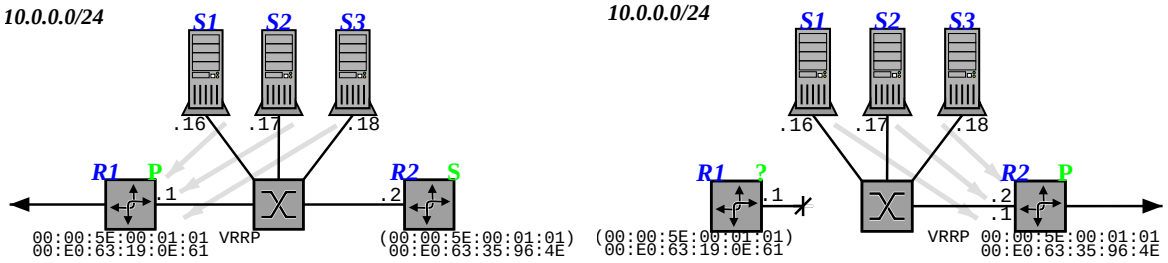
De Juniper routers in Amsterdam (1.1 en 2.1 in figuren 2.1 en 2.2) zullen een route moeten hebben voor elke mogelijke bestemming in het Internet. Hiervoor wordt een External Gateway Protocol gebruikt.
De routers wisselen met routers van andere ISPs gegevens uit over welke routes zij kunnen bereiken. Zo zullen onze routers tegen al hun buren op AMS-IX annonceren dat ze een route hebben voor 213.136.0.0/19 en 193.109.122.0/24 voor IPv4, 3FFE:8350::/28 en 2001:7B8::/35 voor IPv6. Dit zijn de netwerken die door RIPE aan Business Internet Trends zijn toegekend.
De andere routers op AMS-IX (van zo'n 120 andere ISPs in Nederland) zullen op hun beurt hun eigen netwerken annonceren aan onze routers. Zo leren we als het ware waar we het verkeer heen moeten routeren. Het verkeer over de AMS-IX kost de ISPs niets en het is dus wenselijk om zo veel mogelijk van deze buren (*peers* genaamd) te hebben. Het uitwisselen van verkeer naar deze peers is namelijk gratis.
Voor alle routes die we niet op AMS-IX kunnen leren, schaffen we zogenaamde transit contracten aan. Transit providers zijn providers die op vrijwel alle Internet Exchanges aanwezig zijn en derhalve alle routes ter wereld kennen. Omdat zij hebben moeten investeren in transatlantische en pan-europese verbindingen, betaalt Business Internet Trends deze providers om verkeer van ons netwerk weg te zetten in de rest van de wereld. Het bedrijf heeft op dit moment drie transitproviders. We leren dus van elk netwerk ter wereld van drie verschillende providers een route, zodat we de beste daaruit kunnen selecteren (BGP doet dit voor ons). Bij het uitvallen van één van de transit providers (bijvoorbeeld bij een storing of omdat zij failliet gaan) kan een alternatief pad worden gekozen bij een andere transit provider.
Omdat dit in de regel gebeurt aan de rand van het netwerk van een ISP (immers, ons netwerk houdt in Amsterdam op), wordt dit het Border Gateway Protocol genoemd. BGP is het de facto Internet routing protocol.
Beide routers in Amsterdam zullen we dus BGP laten spreken met AMS-IX peers, met transit providers en met elkaar. Hierdoor kunnen we garanderen dat als de routers in Amsterdam de data ontvangen, ze altijd weten hoe ze het moeten weg-routeren uit ons netwerk.

2.6.3 Open Shortest Path First (OSPF)

Binnen ons netwerk is het zaak om vanuit elke router een route te hebben naar alle machines binnen ons netwerk. Omdat we intern de beste route willen kunnen gebruiken, moeten we een protocol aanwenden dat een dergelijke kortste route berekent. Een internet standaard voor interne routering is OSPF - het Open Shortest Path First protocol.
OSPF is een zogenaamd *link-state* protocol, waarbij alle routers elkaar op de hoogte houden van de netwerken waaraan zij geconnecteerd zijn. Hierdoor heeft elke router hetzelfde beeld van het netwerk, de topologie.
Van elke interface wordt een kostprijs gegeven, die omgekeerd evenredig is met de bandbreedte die beschikbaar is op dat interface. Zo zal een FastEthernet (100 Mbps) verbinding een waarde van 10 krijgen, een E1 (2 Mbps) zal een waarde van 500 verkrijgen en een Gigabit Ethernet zal goedkoop zijn, met waarde 1.
Routers worden gegroepeerd in *areas*, waarbij de routers enkel de routers in hetzelfde area op de hoogte zullen stellen van veranderingen in de topologie van die area.
Nu zal iedere router in dezelfde OSPF area door middel van het “kortste pad-algoritme” van Dijkstra bij het veranderen van de topologie een herberekening doen en van elke andere router in de area de kostprijs berekenen.
Bij een routeringsbeslissing wordt nu het pad gekozen met de laagste kostprijs. Bij het uitvallen van dit pad, zal een herberekening van het kortste pad geschieden en het nieuwe kortste pad gebruikt worden. Bij meer paden van gelijke kostprijs, wordt het verkeer evenredig verdeeld over de paden waardoor loadbalancing mogelijk is.

2.6.4 Virtual Redundant Router Protocol (VRRP)

Naarmate de klanten van Business Internet Trends hogere eisen stellen aan de beschikbaarheid van hun diensten op het Internet, groeit het belang van een zeer beschikbaar netwerk mee. Het uitvallen van een router in het lokale netwerk kan tot gevolg hebben dat de server(s) van de klanten tijdelijk niet meer bereikbaar zijn en is derhalve onacceptabel.
Om de doelstelling van een zo hoog mogelijke uptime in ons netwerk te kunnen halen, passen we VRRP toe. De werking maken we duidelijk aan de hand van een voorbeeld (zie ook figuur 2.3).
Twee routers, R1 en R2, zitten beiden in hetzelfde netwerk. De IP nummers zijn respectievelijk 10.0.0.1 voor R1 en 10.0.0.2 voor R2. De servers S1…S3 stellen hun default gateway in op 10.0.0.1.
We spreken af dat R1 primair (’P’) is en R2 secundair (’S’). VRRP zorgt er voor dat er een denkbeeldig MAC adres wordt gegenereerd en dat R1 naar dit MAC address zal luisteren. R2 is ook op de hoogte van dit MAC adres, maar zal hier vooralsnog niets mee doen. Als servers in het LAN door middel van ARP vragen “who-has 10.0.0.1”, zal R1 met zijn VRRP MAC adres antwoorden met “10.0.0.1 is-at 00:00:5E:00:01:01”. R2 zal niets doen.
R1 zal elke seconde een multicast pakketje sturen naar R2. Aan de hand van dit zogenaamde keepalive pakketje, kan R2 zien dat R1 in orde is. Het verkeer wordt zo dus via R1 gerouteerd (zie grijze pijlen in figuur 2.3(a)).
Als nu R1 mocht crashen of anderszins onbereikbaar blijken, zullen de multicast pakketjes R2 niet meer bereiken en zorgt VRRP ervoor, dat R2 direct het IP nummer



(a) VRRP in normale toestand: R1 handelt het verkeer af

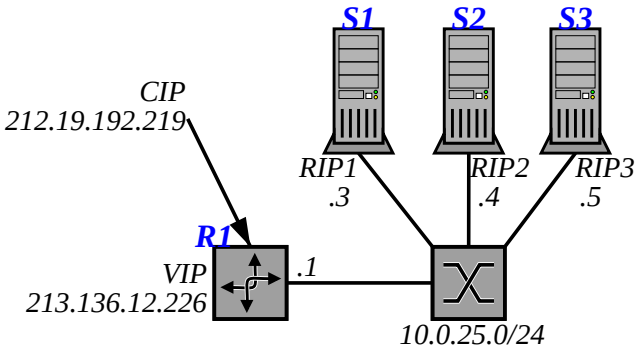
(b) VRRP in backup toestand: R2 handelt het verkeer voor R1 af

Figuur 2.3: *Het VRRP protocol, waarbij twee routers hetzelfde IP nummer met een gedeeld MAC adres bedienen. Dit verhoogt aanzienlijk de beschikbaarheid van de servers S1...S3.*

10.0.0.1 van R1 begint te gebruiken en dus de primaire router wordt. Als er nog Ethernet verkeer wordt gestuurd naar 00:00:5E:00:01:01, zal R2 deze nu ontvangen en routeren. Ook zal R2 ARP antwoorden geven voor 10.0.0.1 en de servers S1...S3 merken niets van de downtime. Voor hen blijft namelijk de default gateway gewoon 10.0.0.1 op het bij hen bekende MAC adres.

2.6.5 Load Sharing Network Address Translation (LSNAT)

Naast het netwerk is er nog een andere beperking op beschikbaarheid van diensten van de klanten van Business Internet Trends. Zo kan een server crashen, of overbelast raken. Hierdoor is de dienst van de klant ook niet optimaal bereikbaar. Eén mogelijke oplossing voor dit probleem is het inzetten van meerdere servers met dezelfde dienst, bijvoorbeeld drie servers met precies dezelfde content.
We zouden DNS kunnen gebruiken om "www.klant.nl" te laten wijzen naar de drie IP nummers van S1, S2 en S3. Echter, bij het uitvallen van één van deze servers, zal een derde van de aanvragen op de website door DNS worden verwezen naar de kapotte server en dit is niet wenselijk. Ook is het (tijdelijk) down brengen van een server voor bijvoorbeeld een upgrade niet gemakkelijk.



Figuur 2.4: *Het LSNAT protocol, waarbij een router een serverpool (S1...S3) presenteert aan de buitenwereld als één IP nummer (VIP).*

Een oplossing wordt geboden door het LSNAT principe, geschetst in figuur 2.4. Hierbij laten we de router R1 naar een virtueel IP nummer luisteren, het VIP (213.136.12.226). In de configuratie van de router geven we de drie servers S1...S3 op. Omdat dit de echte servers zijn (in tegenstelling tot de virtuele server op R1), heten deze IP nummers RIP1, RIP2 en RIP3 (dit komt van de engelse term Real IP).
LSNAT werkt als volgt. Een client (CIP) bezoekt de website van "www.klant.nl" en vraagt aan zijn DNS server het IP van deze site. Het antwoord is VIP dus de client zal hiernaar connecteren. De router ontvangt het pakket en zal hier een netwerk address translatie (NAT) op toepassen, en het bestemmings adres van de aanvraag omschrijven tot één van de drie servers, bijvoorbeeld S2. Het pakket wordt dus omgeschreven van $(CIP \rightarrow VIP)$ naar $(CIP \rightarrow RIP2)$. De router levert het pakket af bij S2. Deze accepteert de connectie en zal data sturen naar CIP, omdat hij het pakket daarvandaan zag komen. Omdat op S2 de default gateway naar R1 verwijst, wordt de data hierheen gestuurd. R1 op zijn beurt, ontvangt het pakket en schrijft het om van $(RIP2 \rightarrow CIP)$ naar $(VIP \rightarrow CIP)$, zodat de client de data weer ziet verschijnen van VIP.
Nu zal de router om de zoveel tijd kijken of de dienst op S1...S3 nog wel up is, door de machines te pingen en door te connecteren naar de webserver en te kijken of de machine nog reageert. Zo niet, wordt de server uit de loadbalance pool gehaald totdat hij weer goed bevonden wordt. Alle nieuwe connecties worden dan niet meer naar die betreffende server gestuurd, maar nog wel naar de andere twee.
Met LSNAT kan een zeer hoge beschikbaarheid nagestreefd worden en dit zal dus niet enkel voor de klanten van Business Internet Trends worden ingezet, maar zeker ook voor de eigen diensten, zoals het mailplatform, en de web-, news- en name-servers.

2.7 De nummering

Om tot een goede implementatie te kunnen komen in een volgende fase van dit project, zullen we een degelijke planning moeten maken van welke VLANs, OSPF areas en welke IP nummers we gaan gebruiken in het nieuwe ontwerp.
In deze sectie lichten we de VLANs toe die we zullen gebruiken in het nieuwe netwerk. We maken een scheiding tussen de core en de klanten. De Junipers in Ede vormen deze scheiding. De verbinding tussen de colo switches (*rbfs* en in fase 2 ook de Alpine switch in de nieuwe colo) en de Juniper routers in Ede, wordt een "untagged" verbinding, waarover enkel IPv4 en IPv6 verkeer zal gaan. Hierdoor kunnen we VLAN *n* zowel in het lokale LAN als in de core gebruiken en hoeven we ons geen zorgen te maken dat deze zaken door elkaar gaan lopen.
Achtereenvolgens tonen we in tabel 2.1 de benodigde VLANs in het core netwerk, in tabel 2.2 de OSPF areas voor het gehele netwerk en in tabel 2.3 de planning voor toekenning van IP nummers aan de apparaten in het nieuwe netwerk.

2.7.1 De VLAN indeling

In tabel 2.1 komen geen colocatie servers of klanten servers. De genummerde VLANs (100-199, 200-299, 300-399), worden één voor één toegekend aan de gebruikers ervan. In principe zijn de VLANs boven de 100 layer2, dus alleen ethernet

Nummer	*Naam*	*Doelstelling*
1	v-default	Default VLAN, ongebruikt
2-9	–	Ongebruikt
10	v-bb-v4	IPv4 backbone verkeer
11	v-bb-v6	IPv6 backbone verkeer
12	v-admin	Administratieve machines
13	v-amsix	AMS-IX koppelingen
14-99	–	Gereserveerd voor later gebruik
100-199	v-bitNN	BIT VLANs, genummerd 00, 01, . . . , 99
200-299	v-custNN	Klanten VLANs, genummerd 00, 01, . . . , 99
300-399	v-edixNN	Ede IX VLANs, genummerd 00, 01, . . . , 99

Tabel 2.1: *De VLAN indeling de Alpine switches en Juniper routers.*

verkeer. Er zal dan in de regel geen gebruik gemaakt worden van IP connectiviteit van de Junipers en andere routers in de backbone.
Voor de *v-bit* VLANs kunnen we bijvoorbeeld een PC in Amsterdam aansluiten en in hetzelfde netwerk laten deelnemen als PCs in Ede. De *v-cust* VLANs kunnen dienen om klanten van Business Internet Trends de gelegenheid te geven om hun datacenter uit te breiden met een colocatie bij ons bedrijf. Hiervoor maken ze dus geen gebruik van IP diensten van ons en zullen in de regel hun eigen routers en IP nummering hanteren. Zo behoudt de klant zijn autonomiteit en kan hij toch servers in Ede huizen. De *v-edix* VLANs zijn specifiek voor het transporteren van ethernet verkeer van Amsterdam naar Ede. We zullen dan de fast-ethernet poorten van deze klanten in Amsterdam direct in de AMS-IX switch koppelen, en in Ede aan de Ed-IX switch (POP 5). Hierdoor kan de klant in Ede een router in de Ed-IX koppelen en hierdoor een directe layer2 koppeling hebben met de AMS-IX.

2.7.2 De OSPF areas

Area	*Beschrijving*	*Deelnemers*
0.0.0.0	Backbone	Alle Junipers, beide Cisco 7206s (fase 1)
1.0.0.0	Sara	De Juniper in Sara, de Cisco 7206 in Amsterdam
2.0.0.0	TC2	De Juniper in Telecity2
2.0.0.1	TC2-Dialup	De Junipers in Amsterdam, de dialup Ciscos in TC2
3.0.0.0	Ede-HQ	De Junipers in Ede, de Cisco 7206 in Ede, beide SSRs
3.0.0.1	Ede-Dialup	De Junipers in Ede, Cisco in Ede, de TNTs

Tabel 2.2: *De indeling in OSPF areas van het gehele netwerk*

Het indelen van een ISP netwerk gebeurt in principe door eerst het netwerk in delen op te splitsen en aan de hand daarvan routers te groeperen die in de verschillende subnetwerken deelnemen. Deze subnetwerken heten in OSPF areas. Area 0.0.0.0 heeft een speciale betekenis, de zogenaamde *backbone* area. Alle routers die in deze area zijn aangesloten, hebben kennis van alle topologie updates van alle andere areas en kunnen waar nodig updates doorgeven. Alle andere areas worden *stub* areas genoemd. Deze hebben enkel zicht op de lokale topologie van hun eigen area en zullen zich niet bekommeren over de rest van het netwerk. Hiervoor gebruiken ze een default route naar een router die wèl in de backbone area deel neemt.
We definiëren voor elke fysieke locatie een area. De nummering volgt hier de POP

nummering. We zorgen er dus voor, dat topologie-veranderingen in Ede-HQ niet zullen propageren door het gehele netwerk, maar slechts tot de Juniper, die aan de andere backbone routers dit door geeft. Dit heeft een gunstig effect op de hoeveelheid routers die een herberekening moeten maken van het kortste-pad algoritme, bij het wijzigen van de lokale topologie.

2.7.3 Het IPv4 numberplan

netwerk	*naam*	*doel*
213.136.31.0/27	backbone	routers, wdm en switches in de core
213.136.31.32/27	ede-bb	ssrs, junipers, switches in Ede
213.136.31.64/30	pos-1-2	stm1 interface van kelvin naar tc2
213.136.31.68/30	atm-1-2	e3 interface van kelvin naar sara
213.136.31.72/30	-	gereserveerd voor uitbreiding
213.136.31.76/30	-	gereserveerd voor uitbreiding
213.136.31.80/28	-	gereserveerd voor uitbreiding
213.136.31.96/27	-	gereserveerd voor uitbreiding
213.136.31.128/25	-	gereserveerd voor uitbreiding
213.136.12.0/28	mail	pop, smtp en virusscan servers
213.136.12.16/28	dialin	tnt2 dialin servers, ede
213.136.12.32/28	nntp	newsservers en –feeders
213.136.12.48/29	dns1	nameservers
213.136.12.56/29	dns2	backup nameservers
213.136.12.64/27	web	Unix (web)servers
213.136.12.96/27	windows	Microsoft (web)servers
213.136.12.128/26	admin	administratieve servers
213.136.12.192/27	misc	overige diensten
213.136.12.224/27	loadbalance	IP pool voor loadbalancers

Tabel 2.3: *Het Numberplan voor IPv4 met betrekking tot het nieuwe netwerk.*

Voor de routers van het nieuwe netwerk is door de netwerkbeheerders van Business Internet Trends 213.136.31.0/24 gereserveerd (256 IP nummers). Hierin maken we een verdeling in backbone en lokale routers, die we *ede-bb* noemen. Daarnaast hebben we in fase 1 nog de STM1 verbinding, waarvoor een IP loopback netwerkje nodig is.
De vier POPs die we hebben gecreëerd krijgen ieder een doopnaam. De Amsterdamse POPs heten Sara respectievelijk Telecity2, de Ede-HQ is gelegen aan de Kelvinstraat en krijgt aldus de naam Kelvin. De Colocatie faciliteit in Ede zal liggen aan de Galileistraat, en krijgt dus de naam Galilei. Dit weerspiegelt in de subdomain onder netwerk.bit.nl zodat we zeker weten waar de betreffende machine staat.
Voor de naamgeving van de machines zelf, houden we de volgende conventie aan. Allereerst het type machine, JUN voor Juniper, CIS voor Cisco, ALP voor Alpine en WDM voor Optical Access, gevolgd door het POP nummer en een volgnummer binnen die POP. De eerste Juniper in de Ede-HQ locatie zal dus JUN-3-1.KELVIN.NETWORK.BIT.NL heten.

2.8 Conclusies en aanbevelingen

We hebben enkele beslissingen genomen met betrekking tot het logische ontwerp van het netwerk. We concluderen dat OSPF als IGP de geschikte keuze is. BGP zal nog steeds worden gebruikt als EGP om routes in Amsterdam te leren. De routers in Amsterdam zullen dan een default route in het OSPF backbone area injecteren, zodat zij al het externe verkeer naar zich toe trekken.
We bevelen het gebruik van EAPS aan, ondanks de kostprijs van de licenties hiervoor.

Hoofdstuk 3

Migratieplan

3.1 Inleiding

Bij het ontwerpen van het nieuwe netwerk voor Business Internet Trends, zijn beslissingen genomen die er voor zorgen dat het interconnecteren van het huidige netwerk bemoeilijkt wordt.
In het licht van groei van het IP verkeer en beschikbaarheid van de aangeboden diensten, zullen we het nieuwe netwerk voorzien van loadbalancers en redundante interne routing door middel van OSPF.
Dit houdt in dat de diensten van Business Internet Trends op een andere manier in het netwerk opgenomen dienen te worden en dat deze in sommige gevallen worden uitgebreid met redundante servers.
We zullen deze migratie uitvoeren vanuit het perspectief van de gebruiker. Er zullen voor hem een aantal zaken wijzigen en we achten het verstandig om alle klanten op voorhand in te lichten over onze veranderingen. Hiervoor zullen we een mailing sturen.

3.2 Probleem- en doelstelling

Voor dit gedeelte van het project definiëren we vier problemen, die structureel worden aangepakt in dit hoofdstuk. Het doel van dit gedeelte is het oplossen van de hierna gestelde problemen.
Wanneer er gesproken wordt van het 0-netwerk, bedoelen we het netwerk in Ede welke als IP range 213.136.0.0/24 heeft en in het default VLAN bestaat.

3.2.1 Probleemstelling

- Het huidige netwerk maakt enkel gebruik van de Cisco 7206 om verkeer van Ede naar Amsterdam te transporteren. We zullen de Cisco moeten ontlasten en het verkeer door de gigabit Junipers laten afhandelen.
- De servers van Business Internet Trends en van haar klanten zitten veelal in hetzelfde VLAN. Er zal een logische scheiding gemaakt moeten worden tussen deze twee types machines.

- Voor de verschillende diensten is nog geen overzicht aanwezig dat beschrijft hoe we het beste de servers kunnen verplaatsen naar hun eigen VLAN. Dit overzicht moet gemaakt worden.
- Onze klanten zijn nog niet op de hoogte van de interne veranderingen in het Business Internet Trends netwerk. Ze zullen op de hoogte gesteld moeten worden.

3.2.2 Doelstelling

Naar aanleiding van de hierboven gespecificeerde problemen brengen we hier de doelstellingen voor deze fase in kaart.

- We onderzoeken welke gevolgen de migratie van servers heeft voor onszelf en voor de klanten en brengen dit in kaart in dit document.
- We sturen al de klanten (ruim 1400) van Business Internet Trends een brief waarin we de migratie aankondigen. Dit doen we voor we beginnen met de implementatiefase.
- Het netwerk zal moeten worden gewijzigd zodat gebruik kan worden gemaakt van de nieuwe Juniper routers. In het bijzonder moet de Cisco uit worden gefaseerd. Dit wordt besproken in paragraaf 3.4.
- De servers in het default VLAN (213.136.0.0/24) zullen moeten worden verhuisd naar de VLANs die in het ontwerp zijn gespecificeerd. Dit wordt besproken in paragraaf 3.5.

Van de daadwerkelijke uitvoering van de serververhuizing wordt in een volgend hoofdstuk tijdens de implementatiefase verslag gemaakt.

3.3 Technieken en Methodes

We zullen een brief sturen aan alle betalende klanten van Business Internet Trends. Dit zijn voornamelijk resellers van onze producten, bedrijven die inbel- of DSL diensten afnemen en mensen die bij ons servers colocaten. In totaal zijn er 1455 bedrijven en instellingen aangeschreven. De brief is te vinden in bijlage D.
Er zal vooral 's nachts worden gewerkt. We spreken af in ploegen van minimaal twee engineers te werken. Er wordt een mailinglist gemaakt waarop alle betalende klanten zich gratis kunnen abboneren. We zullen er zorg voor dragen dat alle onderhoudswerkzaamheden die een serviceonderbreking ten gevolge (kunnen) hebben, worden aan- en afgemeld op deze mailinglist (onderhoud@bit.nl).

3.4 Netwerkmigratie

In het nog draaiende netwerk speelt de Cisco 7206 VXR-300 (*ed-bfr*) een centrale rol. Hij wordt in principe als default gateway gebruikt in het 0-netwerk. Dit is om een aantal redenen niet wenselijk meer:

1. Business Internet Trends heeft de beschikking over twee Riverstone routers die perfect geschikt zijn als interne (datacenter) routers. Er wordt al enigszins gebruik gemaakt van deze routers (*ssr1* en *ssr2*) maar het 0-netwerk gebruikt nog de Cisco.

2. Deze Cisco wordt vervangen door een Juniper en het gebruik ervan zal dus moeten worden afgebouwd.

3. De machine heeft een 100 mbps verbinding naar het lan in Ede. Deze verbinding wordt ook gebruikt voor uitgaand verkeer. Verkeer wat van het 0-netwerk naar een ander VLAN gaat, zal over deze router geleid worden en derhalve gaat al het verkeer wat eigenlijk intern blijf (van 0-netwerk naar een ander VLAN) de capaciteit van de uitgaande verbinding knijpen.

We zullen dus een werkwijze moeten opstellen die ons in staat stelt om de Cisco uit het 0-netwerk te verhuizen, zonder dat hierbij de connectiviteit down gaat. Immers, totdat de Junipers zijn geïnstalleerd en goedgekeurd, moet de Cisco nog wel al het verkeer van de ISP afhandelen.
De Cisco heeft in het 0-netwerk het IP adres 213.136.0.2. Verreweg de meeste machines in dit netwerk stellen zijn IP nummer in als de default gateway. De oudste Riverstone (*ssr1*) heeft het IP adres 213.136.0.4 en recentelijk heeft de nieuwste Riverstone (*ssr2*) het IP adres 213.136.0.1 gekregen. Het lijkt ons verstandig om de eerste ssr het adres van de Cisco te laten overnemen. We hebben dan een netwerk waarin .1 en .2 beiden routers zijn, die door middel van VRRP (zie het hoofdstuk “Ontwerp”) samen de connectiviteit van het netwerk kunnen garanderen.

3.4.1 Router Migratie

Om *ed-bfr* uit het default VLAN te krijgen en het routeren bij Business Internet Trends geheel te laten verlopen over de Riverstone routers *ssr1* en *ssr2*, zijn de volgende stappen nodig.

Stap 0 – Connectiviteit testen

We gebruiken het programma “fping” om alle machines in het 0-netwerk te pingen. We noteren welke servers er up zijn en welke down. In het bijzonder tellen we de servers goed na, zodat we na de migratie kunnen vaststellen dat de connectiviteit voor alle servers geheel hersteld is. De uitvoer van “fping -g 213.136.0.0 213.136.0.255” wordt bewaard. Over de IPv6 connectiviteit in het 0-netwerk hoeven we ons in principe geen zorgen te maken, omdat we dit nog niet aan de klanten aangeboden hebben en derhalve niet verwachten dat er gebruik van gemaakt wordt in het 0-netwerk. De enige uitzondering hierop is manitou, maar dit is een server van onszelf.

Stap 1 – Ede Backbone activeren (213.136.31.32/28)

In Ede zal een lokale “backbone” worden geactiveerd. De routers in Ede zullen hieraan deelnemen. Omdat we de Junipers nog niet binnen hebben, kunnen we alleen nog maar de SSRs en de Cisco’s activeren. We reserveren wel alvast IP nummers voor de Junipers. De lijst wordt aldus:

IP nummer	*Router naam*
213.136.31.33	jun1.kelvin.network.bit.nl
213.136.31.34	jun2.kelvin.network.bit.nl
213.136.31.35	ssr1.kelvin.network.bit.nl
213.136.31.36	ssr2.kelvin.network.bit.nl
213.136.31.37	cis2.kelvin.network.bit.nl (*IPv6*)
213.136.31.38	cis1.kelvin.network.bit.nl (*ed-bfr*)

Deze configuratie kan gedurende de dag geschieden. Het toevoegen van dit VLAN en de interfaces op de routers zal geen invloed hebben op de connectiviteit van de klanten van Business Internet Trends.

Stap 2 – OSPF op Ede Backbone activeren

Op de Riverstones en Cisco's zal OSPF moeten worden geactiveerd op de interfaces waarmee zij aan v-edebb zijn gekoppeld. Dit zal geen onderbreking in de dienstverlening hebben. Op *ed-bfr* zal tijdelijk een static route voor 213.136.31.32/28 gezet worden naar één van de SSRs. Als dit niet gebeurt, is het ede-bb netwerk niet bereikbaar voor de buitenwereld. We zetten een static route naar ssr1 op IP adres 213.136.31.35.

Stap 3 – *ed-bfr* verhuizen naar Ede Backbone

De Cisco router is default gateway voor alle machines in het 0-netwerk. Als we deze dus weghalen, zal tijdelijk de connectiviteit wegvallen van het gehele ISP netwerk. Dit moet dus zorgvuldig gebeuren.
We zullen *ed-bfr* omnummeren van 213.136.0.2 naar zijn plaats in het v-edebb netwerk: 213.136.31.38. Daarna zal OSPF moeten worden geconfigureerd, analoog als dit ging met de IPv6 router in Stap 2. *Ed-bfr* zal dan routes leren voor alle netwerken die aan de SSRs zijn gekoppeld - in de regel zijn dit alle lokale netwerken in de colocatie faciliteit.

Stap 4 – Op *rbfs, ssr1, ssr2* VLAN v-colo0 configureren

Nu de connectiviteit verdwenen is, zal de *rbfs* switch moeten worden geconfigureerd zodat de ethernet poorten die nu in het default VLAN zitten, verhuizen naar het v-colo0 VLAN (50). We zullen alle poorten die nog geen VLAN toegewezen hebben (dit zijn er nogal wat), maar waar daadwerkelijk een server aan hangt, veranderen naar VLAN 50.
ssr1 en *ssr2* hebben op dit moment een IP adres in het 0-netwerk. Deze zullen ze behouden, maar niet in het default VLAN. We verhuizen het interface i-default op beide routers naar VLAN v-colo0 en noemen ze derhalve i-colo0.
Wat het IPv6 betreft verandert er nu iets in het 0-netwerk. Aanvankelijk bestond het 0-netwerk in het default VLAN waardoor de IPv6 allocatie hiervoor 2001:7b8:3:0::/64 was. Het 0-netwerk verhuist nu naar v-colo0, met nummer "50", waardoor het IPv6 netwerk meeverhuist naar 2001:7b8:3:32::/64. De enige machine die hier last van zal hebben is Manitou, en deze zullen we hier dus handmatig omnummeren. De IPv6 Cisco zal een dot1q interface moeten aanmaken voor dit specifieke /64 netwerk en zichzelf weerhouden van het aanbieden van router-advertisements in dit VLAN,

omdat we nog steeds niet klaar zijn om onze klanten hier gebruik van te laten maken. Manitou kan echter wel handmatig een IPv6 adres instellen en derhalve deelnemen aan het IPv6 netwerk van Business Internet Trends.

Stap 5 – *ssr2* verhuizen naar oude plaats *ed-bfr*

Omdat 213.136.0.2 nu geen machine meer heeft en men wel juist dit IP nummer gebruikt als default gateway in het 0-netwerk, zullen we *ssr2* verhuizen van 213.136.0.4 naar 213.136.0.2. De machines in het oude default VLAN zullen nu allen in VLAN v-colo0 staan en weer een gateway vinden op 213.136.0.2 (i-colo0 op *ssr2*).

Stap 6 – VRRP configureren op de SSRs

De gateways voor het v-colo0 netwerk zijn nu 213.136.0.1 en 213.136.0.2. We zullen voor deze IP nummers twee VRRP groepen opzetten, zodat de gebruikers van het adres op ssr1, ssr2 als backup router krijgen en andersom.

Stap 7 – Connectiviteit testen

We pingen met 'fping' alle servers het het 0-netwerk en verwachten dat alle servers die up waren in Stap 0, nu nog steeds up zijn. Eventueel kunnen we 5 minuten wachten of de arp-cache van *rbfs* wissen, zodat we zeker weten dat alle servers de juist MAC adressen van de routers geleerd hebben. We sluiten deze operatie af als blijkt dat er evenveel servers up zijn vóór de migratie, als daarna. De uitvoer van "fping -g 213.136.0.0 213.136.0.255" wordt vergeleken met de bewaarde uitvoer in Stap 0.

3.4.2 Switch Migratie

Bij Business Internet Trends zijn in principe vijf switches verantwoordelijk voor de layer2 connectiviteit in Ede. Aan de basis hiervan staat *rbfs*. Op enkele van de (trunk) switchpoorten van *rbfs* zijn andere Cisco switches aangesloten. Drie hiervan staan in het engineering kantoor, de laatste staat op de hoek van de Kelvinstraat in het Internet Opleidingscentrum.
Nadat we het default VLAN naar VLAN v-colo0 (nummer 50) hebben omgenummerd, moeten we er zorg voor dragen dat de overige drie switches hiermee in overeenstemming zijn. Ze zullen nu namelijk verkeer voor het 0-netwerk ontvangen met VLAN 50 op hun trunkpoort, dus de overige poorten die daar nog in het default VLAN zitten moeten ook omgenummerd worden.
Tevens moeten we de IP nummers van de switch zelf, waarmee we inloggen op de machines en waarmee we SNMP monitoring bedrijven, omnummeren. Deze staan namelijk nog steeds in het 0-netwerk. We besluiten hiervoor weer het default VLAN te gebruiken. De achterliggende gedachte is, dat zelfs als alle VLAN informatie om de een of andere manier verloren zou gaan, we nog steeds alle switches in het default VLAN zullen zien.
We maken dus een interface in het default VLAN aan op beide SSRs. Daarnaast geven we alle switches een nieuw IP nummer in een voor switches gereserveerd netwerk: 213.136.31.48/28. We verwachten binnen één jaar niet meer dan 12 switches in gebruik te nemen.

We nummeren dus als volgt om:

IP nummer	*Switch naam*
213.136.31.49	default-gw.network.bit.nl (*ssr1*)
213.136.31.50	default-vrrp.network.bit.nl (*ssr2*)
213.136.31.51	switch-3.network.bit.nl (*IOC*)
213.136.31.52	switch-4.network.bit.nl (*Engineering*)
213.136.31.53	switch-5.network.bit.nl (*Engineering*)
213.136.31.54	rbfs.network.bit.nl (*Colo I*)
213.136.31.55	switch-6.network.bit.nl (*Bakkenbouwers*)

Het verhuizen van de IP nummers en VLANs van de switches heeft geen inpact op de klantendiensten. Het enige wat van belang is, is dat de DNS na de migratie zo snel mogelijk moet worden gesynchroniseerd met de nieuwe situatie, omdat ons MRTG statistieken programma elke vijf minuten inlogt op de switches om van elke ethernet poort het verkeer te tellen. We zullen dus voordat we de switches omnummeren, de DNS in orde brengen en daarna de nameservers en de caching resolvers herstarten zodat overal de hostnames voor switch-X.network.bit.nl naar het juist IP nummer verwijzen. De statistieken zullen daarna gewoon door gaan met de nieuwe informatie.

3.5 Servermigratie

Aan het begin van de opdracht werd vastgesteld dat er een “klanten” migratie plan moest komen zodat aan de servers van de klanten van Business Internet Trends in het nieuwe netwerk een goede dienstverlening kon worden gegarandeerd. Als we de migratie echter nader beschouwen, blijkt dat het gemakkelijker zal zijn om, in plaats van de klantenservers, de *eigen* servers te verhuizen naar andere VLANs. Dit heeft twee motivaties:

1. Er zijn minder eigen servers (58) dan klantenservers (123).
2. We zullen toch VLAN based services gaan aanbieden (dus mail, dns, webservers), dus onze eigen servers moeten in een later stadium worden omgenummerd.

Er wordt dus besloten om een *server*migratieplan te schrijven, waarbij we als uitgangspunt nemen dat aan het einde van de migratie zo veel mogelijk van onze eigen servers uit het 0-netwerk verhuizen naar het in het ontwerp vastgestelde 213.136.12.0/24 netwerk.
Er zijn enkele servers met een zeer duidelijk profiel. We proberen zo veel mogelijk een indeling te maken aan de hand van de aan te bieden dienst (webserver, newsserver, nameserver, ...) en komen tot een achttal groepen die we in de volgende secties bespreken.
Van elke categorie wordt in kaart gebracht welke servers hiertoe behoren, of zij IP afhankelijkheden hebben (en zo ja: welke) en hoe we zonder kleerscheuren deze servers kunnen verhuizen uit het 0-netwerk naar een geschikt netwerk in 213.136.12.0/24.

3.5.1 Nameservers

Het DNS-mechanisme bestaat uit een vraag en antwoord spel. Een client vraagt aan een resolver het IP adres van een hostname (of de hostname die bij een IP adres hoort) en deze resolver gaat dit uitzoeken. De resolver doorloopt de DNS boomstructuur vanuit de wortel "." en komt uiteindelijk bij juist die nameserver, die verantwoordelijk is voor de betreffende zone. Het antwoord wat hij krijgt van de verantwoordelijke nameserver, zal hij doorgeven aan de client.
Enerzijds heb je dus de zogenaamde *resolver* en anderzijds de *authoritative nameserver*. Bij Business Internet Trends draaien beide diensten op dezelfde server. De nameserver die de klanten instellen als resolver is dus tevens de authoritative nameserver. Dit is niet wenselijk om de volgende reden. Stel, een klant van Business Internet Trends, verhuist zijn domein – ipng.nl – naar een andere provider en past daarbij de hostname voor www.ipng.nl aan. Als Business Internet Trends hier niet of niet goed van op de hoogte is, zullen de nameservers die onze klanten gebruiken, de verkeerde antwoorden voor www.ipng.nl geven omdat we hen hebben ingesteld als authoritative nameserver voor ipng.nl. Hierdoor is dit specifieke domein onbereikbaar voor onze klanten.
Een tweede motivering om de DNS diensten te scheiden is de grootte van de resolver. Bij het draaien van een nameserver die ook resolver is, groeit het programma behoorlijk in termen van geheugengebruik waardoor de server trager wordt.
Het is dus om deze redenen wenselijk om de resolver en de authoritative nameserver te scheiden en we besluiten derhalve om dit te doen. We hebben in ons netwerk reeds een resolver draaien: 213.136.0.64. Deze horen de klanten te hebben ingesteld als hun 'default nameserver'.
Omdat het wenselijk is (zelfs in een RFC gedocumenteerd) om de nameservers redundant uit te voeren, besluiten we om twee DNS VLANs aan te maken (v-dns1 en v-dns2). In beide VLANs, die door VRRP beschermd zijn, wordt één resolver en één authoritative nameserver gedraaid. De gebruikte programmatuur zal *djb-dnscache* zijn voor de resolver functionaliteit, en *isc-bind8* voor de authoritative nameservers.

Nieuwe Resolvers

Terwijl we de huidige servers (213.136.0.66 en 213.136.0.77) nog steeds laten resolven, loggen we wie daar nog gebruik van maakt. We verwittigen deze klanten dat ze beter de nieuwe resolvers kunnen gebruiken op nscache1.bit.nl (213.136.12.52, v-dns1) en nscache2.bit.nl (213.136.12.60, v-dns2). Gedurende langere tijd zal onze huidige primaire resolver op 213.136.0.64 blijven werken. Tezijnertijd kunnen we met behulp van de logs ook deze server ontlasten, maar hier is voorlopig geen haast bij.

Nieuwe Nameservers

Terwijl we de huidige servers (213.136.0.66 en 213.136.0.77) nog steeds antwoorden laten geven voor de ruim 13000 domeinen die we in beheer hebben, maken we twee nieuwe nameservers (in v-dns1 en v-dns2) in orde. We maken deze slave aan onze huidige ns.bit.nl machine. Er zal nu bij alle registries een verhuizing van servers moeten worden opgestart. Voor de .nl domeinen is dit de Stichting IDNL, waarvoor

procedures bekend zijn. Voor de overige domeinen zal een andere nameserver verhuizing in gang gezet moeten worden.
We zetten nu een uitgebreide logging aan op de oorspronkelijke servers, in de verwachting dat verreweg de meeste zones niet meer worden geraadpleegd op de oude nameservers (en dus wel op de nieuwe). Na enkele weken bekijken we de logs van de oude nameservers nog eens en ondernemen actie op alle domains die nog opgevraagd worden op onze oude servers.
De nieuwe nameservers noemen we nsauth1.bit.nl 213.136.12.51 en nsauth2.bit.nl 213.136.12.59. De huidige ns3.bit.nl kunnen we zonder problemen omnummeren naar nsauth3.bit.nl en tevens de recursie (de optie die een resolver in bind8 aanzet) uitzetten omdat hier niemand gebruik van hoort te maken. Als er nog mensen gebruik maken van de resolver functionaliteit van ns1.bit.nl en ns2.bit.nl, zullen we deze verwittigen dat ze de nscache1.bit.nl en nscache2.bit.nl moeten gebruiken.
In december zullen we dan de servers ns1.bit.nl en ns2.bit.nl kunnen ontmantelen en inzetten voor een ander doel.

3.5.2 E-Mail diensten

De huidige mailserver installatie bij Business Internet Trends splitst zich in vier subsystemen: SMTP, MX, Virusscanners en POP servers. Een mailsysteem van een grote ISP is dusdanig complex dat we van alle subsystemen een aparte sectie maken en uitgebreid bij de huidige en nieuwe situatie stil staan, alvorens het daadwerkelijke migratieplan uit te werken.

SMTP servers

Allereerst is er de uitgaande mailserver, die het versturen van klantenmail verzorgd. Deze server maakt gebruik van het Simple Mail Transfer Protocol (*SMTP*) en is niet in staat om mail op te slaan, enkel om mail door te sturen naar de eindbestemming. De klant stelt “smtp.bit.nl” in zijn e-mail programma in en daardoor wordt alle uitgaande mail hierop afgeleverd. In het oorspronkelijke netwerk was er één SMTP server; bij uitval zouden klanten niet of nauwelijks mail kunnen versturen.
We stellen voor om de beschikbaarheid van deze (cruciale) dienst te vergroten door middel van een LSNAT oplossing. Het einddoel is het opereren van twee identieke SMTP servers die zich met behulp van de LSNAT router (van Riverstone) aanbieden aan het Internet op één IP nummer. We kiezen als SMTP programma het qmail pakket van Daniel Bernstein.
We zullen een nieuwe server gereedmaken en in een speciaal VLAN (v-mail) opstellen. We geven de machine het IP adres 213.136.12.3 en beschermen dit VLAN met VRRP waarbij 213.136.12.1/28 en 213.136.12.2 de gateway vormen. Aanvankelijk is er dus één server in de serverpool aanwezig. We veranderen nu de DNS zodat “smtp.bit.nl” verwijst naar de loadbalance VIP 213.136.12.226 en houden de oude mailserver in de gaten. Klanten die om de een of andere reden nog het oude IP nummer gebruiken zullen verwittigd worden. We verwachten niet dat deze verhuizing lang duurt (binnen enkele dagen) omdat alle klanten gebruik maken van de e-mail faciliteiten. We kunnen vervolgens de oude server omnummeren naar 213.136.12.4 en toevoegen aan de loadbalance.
In de toekomst is het – zonder service downtime – mogelijk om elk van de smtpservers af te sluiten, onderhoud te plegen of te vervangen alsook naadloos

de capaciteit van onze SMTP dienst uit te breiden met meerdere servers.

MX servers

De huidige setup van Business Internet Trends voor de inkomende e-mail (Mail eXchange ofwel *MX*) is als volgt geïnstalleerd. De inkomende mail wordt naar één van de drie MX servers van Business Internet Trends gestuurd. Elk van deze servers heeft een prioriteit. De twee servers met de laagste prioriteit zullen, als ze een email ontvangen, deze doorsturen naar de server met de hoogste prioriteit *mx1.bit.nl*. Deze zogenaamde fallback servers dienen als tijdelijke buffer bij het uitvallen van de primaire server. Alle ontvangende mailservers draaien het mailpakket SendMail, de oudste van alle mailservers en veelal berucht om zijn cryptische en allesomvattende configuratiebestanden. Heerlijke software dus.
Bij ontvangst van de e-mail door mx1.bit.nl, wordt nagegaan of de betreffende klant zijn mail gescanned wil hebben op virussen en zo ja, doorgestuurd naar één van de twee virusscan machines (zie hierover ook de volgende sectie). Daarna wordt gekeken of de klant een Spam Filter wenst (dit houdt ongewenste, ongeaddreseerde en dikwijls commerciële boodschappen tegen). Hiervoor wordt het programma SpamAssassin gebruikt.
Als de viruscheck en de spamcheck succesvol afregond zijn, wordt de mail in een mailfolder voor de klant opgeslagen. Dit gebeurt momenteel in het traditionele “mbox” formaat op een speciale fileserver van het merk Network Appliances (een Filer F740).
We merken op dat LSNAT oplossingen hier voor een grotere beschikbaarheid kunnen zorgen maar dat het SMTP protocol zelf al voorziet in een behoorlijke hoeveelheid redundantie. Wel kunnen we met LSNAT een grotere doorvoer halen door meerdere MX servers parallel te schakelen. De engineers vinden dit echter momenteel nog niet van belang en we laten LSNAT op de mailservers achterwege.
Uit jarenlange ervaring met mailservers is gebleken dat het mbox formaat, die alle mails in één bestand opslaat, niet zo goed bestand is tegen NFS (netwerk filesystemen). Als namelijk de gebruiker zijn mailfolder raadpleegt en er tegelijk nieuwe mail binnenkomt, kunnen er zich zogenaamde locking problemen voordoen waardoor er mail verloren raakt.
Voor dit concurrency probleem is een oplossing gevonden die beter werkt over NFS terwijl er meerdere processen de mails raadplegen, het “Maildir” formaat. Bij deze strategie wordt voor elke gebruiker een directory aangemaakt op de fileserver en elk mailtje wordt in zijn eigen afzonderlijke bestand opgeslagen. Hierdoor zal nieuwe inkomende mail geen invloed hebben op de te raadplegen mail.
Vanuit het oogpunt van schaalbaarheid en onderhoudbaarheid in de toekomst wordt besloten om het Maildir formaat te implementeren. Dit heeft gevolgen voor twee delen van ons mailsysteem. Ten eerste dient het opslaan van de mail door de zogenaamde *local delivery* in Maildir formaat te geschieden. Onze local delivery programmatuur heet ProcMail. Ten tweede zal het programma wat de mail weer ophaalt voor de gebruiker als hij dit wenst – de zogenaamde *popper* – ook van het Maildir formaat moeten afweten.

Virusscan servers

Er zijn twee machines ingericht als virusscanner. Beide machines hebben dezelfde prioriteit en zullen dus het werk gelijk verdelen. Wellicht kan hier beter een loadbalance gemaakt worden, als blijkt dat de servers het werk niet meer goed kunnen verzetten en hierdoor vertragingen optreden in het bezorgen van mail. Voorlopig is hier echter geen probleem dus we schuiven LSNAT vooruit. Wel maken we de kanttekening dat ook de virusscanner beter beschikbaar zou zijn als we het werk met LSNAT verdelen.
De taak van de virusscanmachine is om mail met het programma Amavis te doorzoeken op bekende virussen. Bij het vinden van een dergelijke mail wordt hij per direct vernietigd en houdt de verdere verwerking op.
Eventueel zal de machine, mits de klant hierom gevraagd heeft, de mail op Spam content scannen en bij het vinden van een dergelijke mail wederom het bericht vernietigen. De klant krijgt dus geen ongewenste Spam als hij dit niet wil. De mail wordt voorzien van een extra “header” en teruggestuurd naar mx1.bit.nl. Deze herkent de header en zal nu de mail lokaal afleveren en in de mailfolder van de klant bijschrijven.
Een inkomend mailtje kan dus op vier verschillende manieren behandeld worden:

1. Geen virusscan, geen spamcheck. De mail wordt aan het localdelivery programma gegeven en opgeslagen.

2. Geen virusscan, wel spamcheck. De mail wordt door mx1.bit.nl op spam gecontroleerd en bij een goedkeuring aan het localdelivery programma gegeven en opgeslagen.

3. Wel virusscan, geen spamcheck. De mail wordt naar één van de virusscanners gestuurd, gescanned, en bij goedkeuring teruggestuurd naar mx1.bit.nl met als optie “geen virusscan, geen spamcheck”.

4. Wel virusscan, wel spamcheck. De mail wordt naar één van de virusscanners gestuurd, gescanned, gechecked op spam en bij goedkeuring teruggestuurd naar mx1.bit.nl, met als optie “geen virusscan, geen spamcheck”.

In alle gevallen is mx1.bit.nl verantwoordelijk voor de initiële ontvangst en de uiteindelijke aflevering op de fileserver.

POP server

Naast het ontvangen van mails voor klanten, zullen deze mails ook opgehaald moeten worden door de klanten zelf. Dit gebeurt met het Post Office Protocol (*POP*), ook wel poppen genoemd. De POP server zal de gebruiker om zijn usernaam en wachtwoord vragen, de opgeslagen mail lezen, naar de gebruiker sturen en vervolgens wissen van de server.
De huidige situatie staat niet toe dat er meerdere POP servers zijn. Dit in verband met de concurrencyproblemen op de fileserver. Tijdens het migreren van mbox naar Maildir moeten we niet meer dan één POP server inzetten, om corruptie van de bestanden te kunnen voorkomen. Nadat Maildir is geïmplementeerd, kunnen we meerdere POP servers inzetten en ook voor deze dienst een veel hogere beschikbaarheid garanderen.

De migratie

Het lijkt mogelijk om een migratie van mbox naar Maildir formaat uit te voeren. Hierbij wordt de mailserver stil gezet, alle mailfolders van de klanten door een programma omgezet naar Maildir formaat, de local delivery en POP programma's vervangen door varianten die Maildir begrijpen, en het mailsysteem tenslotte weer gestart. De geschatte downtime van een dergelijke operatie is vijf uur.
De directie vindt het onacceptabel dat er service veranderingen of downtime in de e-mail afhandeling geschieden. Met hen wordt zorgvuldig een alternatief afgewogen, zodat de e-mail diensten naar de klanten toe geen noemenswaardige downtime zullen hebben.
We passen zelf de local delivery aan, zodanig dat bij het wegschrijven van een mail gekeken wordt of de betreffende klant een mbox of Maildir formaat heeft (dit is gemakkelijk te doen door te zien of de gebruiker een bestand (mbox) of een folder (Maildir) gebruikt). We bekijken regelmatig of de bestanden in onze mailserver nul bytes groot zijn – dit gebeurt wanneer de klant zijn mail juist opgehaald heeft. Hierna wordt het (lege) bestand weggehaald en een directory aangemaakt. De volgende mail die wordt afgeleverd, wordt in het Mailbox formaat weggeschreven. We hoeven nu enkel nog een POP server te creëren die beide formaten ondersteunt.
Dit doen we door zelf in de pen te klimmen en uit de beschikbare broncode voor POP servers (qmail-pop3d voor Maildir en qpopper voor mbox) de juiste code te selecteren en samen te voegen tot één programma. We houden hierbij drie modules aan. Allereerst is er een authenticatiemodule, die de username en wachtwoord afhandeling verzorgt. Als dit gelukt is wordt gekeken of de mail van de gebruiker in een directory of een bestand opgeslagen is en derhalve één van de POP programma's geselecteerd en uitgevoerd. In het geval van mbox (de qpopper variant) wordt na de transactie gekeken of de mbox van de gebruiker leeg is. Zo ja, wordt deze verwijderd en omgezet in een lege Maildir waardoor de volgende mails voor deze gebruiker in Maildir formaat binnenkomen. De programmeerwerkzaamheden worden in overleg met de directie uitbesteed aan een andere werknemer, Peter van Dijk.
Na verloop van tijd zullen alle klanten naadloos overgestapt zijn op de Maildir strategie. Het is vervolgens mogelijk om met meerdere programma's tegelijk in de mail omgeving van een en dezelfde klant te werken.
Dit is een prettige situatie, omdat we voor de POP servers nu een soortgelijke verhuizing als de SMTP server kunnen uitvoeren. We installeren een nieuwe server op 213.136.12.5 in v-mail, maken een loadbalance aan op 213.136.12.227, voegen de nieuwe server aan de loadbalance toe en verhuizen de DNS voor "pop.bit.nl" van de oude server naar de nieuwe lokatie. Als dit gebeurd is, wordt de oude server na enkele weken bekeken op hardnekkige bezoekers en deze worden verwittigd dat ze de nieuwe server moeten gebruiken. Vervolgens kan ook de oude server omgenummerd worden naar 213.136.12.6 en aan de loadbalance worden toegevoegd.
De gebruikte IP nummers worden in het nieuwe mailserver ontwerp dus:

- 213.136.12.1 Nieuwe gateway voor alle mailservers
- 213.136.12.2 VRRP gateway voor alle mailservers
- 213.136.12.3 oneida.bit.nl – SMTP-server 1
- 213.136.12.4 cherokee.bit.nl – SMTP-server 2

- 213.136.12.5 pomo.bit.nl – POP-server 1
- 213.136.12.6 elwha.bit.nl – Virusscan-server 1
- 213.136.12.7 hopi.bit.nl – Virusscan-server 2
- 213.136.12.8 crow.bit.nl – Localdelivery-server 1
- 213.136.12.226 smtp.bit.nl – Loadbalance VIP
- 213.136.12.227 pop.bit.nl – Loadbalance VIP

3.5.3 Dialin (TNT)

Aan het begin van de migratie is de situatie als volgt. Er zijn twee TNT Max dialin machines van Lucent geïnstalleerd. Deze hangen beiden in het 0-netwerk (213.136.0.30 en 213.136.30.31) met als default gateway de ed-bfr router op 213.136.0.2. Het is wenselijk om een apart VLAN te maken voor dialin en dus kennen we er een netwerk aan toe: 213.136.12.16/28. Hierin zullen de beide TNTs zitten en beide SSRs.
Alle machines zullen elkaar updaten door middel van OSPF, zoals in een vorig hoofdstuk is besproken. Hierdoor kunnen de TNTs, wanneer zij een inbeller een IP nummer verschaffen, een /32-route hiervoor aanmaken zodat alle andere routers weten hoe ze deze dialin gebruikern kunnen bereiken. Aangezien we meer dan één TNT hebben, moet dit dynamisch gebeuren. Onze telefonieprovider, Energis, heeft ons één inbelnummer toegekend. Als mensen inbellen bij Business Internet Trends, komen ze willekeurig uit op één van de twee TNTs uit.
Het is zaak om de migratie zo spoedig mogelijk uit te voeren. Omdat we onze dialin dienst niet zomaar kunnen omzetten, wordt besloten dit in een nachtelijk onderhoudsvenster aan te pakken. Er moeten dan vier zaken geschieden:

1. Er moet een VLAN v-dialin worden gemaakt. Op de SSRs moet een interface worden gemaakt in dit VLAN. SSR1 krijgt het IP nummer 213.136.12.17 en SSR2 wordt 213.136.12.18.
2. De TNTs moeten ieder nieuwe IP nummers krijgen. TNT1 wordt 213.136.12.19 en TNT2 wordt 213.136.12.20.
3. De FastEthernet poorten waarin TNT1 en TNT2 op RBFS inkoppelen moeten uit het default VLAN en in v-dialin worden geschakeld.
4. Er moet OSPF worden geconfigureerd op beide SSRs en beide TNTs.

De migratie van het dialin platform naar zijn eigen VLAN zal naar schatting 20 minuten downtime opleveren.

3.5.4 Newsservers

Voor de newsserver van Business Internet Trends wordt het software pakkets “Dnews” gebruikt. Deze NNTP (Network News Transfer Protocol) server is goed configureerbaar en draait sinds geruime tijd betrouwbaar. Zijn adres aan het begin van de migratie is 213.136.0.65, seminole.bit.nl. De newsserver gebruikt ditzelfde

IP nummer voor zowel de readers (de clients) als de feeds (de newspeerings met andere ISPs). Hierdoor zien we geen onderscheid in het verkeer naar andere ISPs en naar onze klanten.
Het is wenselijk om deze server naar zijn eigen VLAN te verplaatsen omdat we dan alle news-gerelateerde servers kunnen groeperen. We definiëren drie IP nummers voor de newsserver:

- 213.136.12.35 - news.bit.nl (de readers)
- 213.136.12.36 - newsfeed-in.bit.nl (de ingaande feeds)
- 213.136.12.37 - newsfeed-out.bit.nl (de uitgaande feeds)

We kunnen dan dus gemakkelijk aan de hand van het verkeer op de drie IP adressen zien hoeveel client-verkeer, hoeveel ingaand en hoeveel uitgaand nieuws we gegenereerd hebben.
Het migreren van de newsserver (seminole.bit.nl), houdt in dat we al onze newspeers (dit zijn er ten tijde van de migratie acht) per e-mail op de hoogte stellen van de veranderingen. Daarna zal seminole nieuwe IP adressen krijgen in v-nntp en zullen we uitgaande newsfeeds gaan opzetten die komen van 213.136.12.37 in plaats van 213.136.0.65.
Alle newspeers moeten dus het nieuwe adres in hun configuratie aanpassen zodat 213.136.12.37 toegang heeft tot hun server, en tevens de connectie met onze newsserver niet meer op 213.136.0.65 maken, maar op 213.136.12.36.
De verwachte downtime van de newsserver is maximaal 10 minuten, tijdens het herconfigureren van de IP nummers. Daarna kunnen één of meerdere newsfeeds tijdelijk (tot maximaal één dag) down zijn, omdat onze newspeers hun configuratie nog moeten aanpassen. Dit heeft echter geen impact op de dienstverlening naar de klanten toe (immers, news.bit.nl op 213.136.12.35 levert ook news uit als er geen newspeerings zijn).

3.5.5 Timeservers

Het is wenselijk dat de klokken van alle servers in het serverpark gelijk lopen. Hiervoor kan gebruik worden gemaakt van NTP (Network Time Protocol), om de servers onderling hun klok te laten synchroniseren met een externe bron.
Bij Business Internet Trends draait een speciale server van RIPE voor een netwerktest (tt52.ripe.net) en deze bevat een GPS. Omdat voor het uitrekenen van de geografische positie via GPS een erg nauwkeurige klok vereist is, heeft deze server de beschikking over wat men noemt een “stratum 0 clock” (de GPS). De tijd van deze server loopt gelijk met de universele standaard-tijd, tot in de nanoseconden nauwkeurig.
De RIPE server is voorzien van een ntp-server. Dat betekent dat andere computers via hem een nauwkeurige tijdsmeting kunnen opvragen. De servers die het aan tt52.ripe.net vragen zijn dan “stratum 1 clock”.
Om alle servers van Business Internet Trends, evenals alle klantenservers op gelijke tijd te krijgen, besluiten we twee nieuwe timeservers te plaatsen. Deze kunnen meedraaien op de reeds aanwezige twee nameservers (besproken in paragraaf 3.5.1). We noemen de machines ntp1.bit.nl (213.136.12.53) en ntp2.bit.nl 213.136.12.61) en zullen de huidige verwijzing van ntp.bit.nl naar tt52.ripe.net omzetten naar

ntp1.bit.nl, zodat alle clients die aanvankelijk de RIPE server gebruikten, nu de eigen timeservers zullen gebruiken.

3.5.6 Het NT platform

Bij de aanvang van de afstudeerwerkzaamheden was er een Windows NT4.0 serverpark geïnstalleerd. Het NT beheersteam acht het raadzaam om over te stappen op een Windows 2000 serverpark en dit heeft enkele consequenties voor de servers.
Uiteraard zullen alle servers een nieuw IP adres verkrijgen in de reeks 213.136.12.96/28. Hierbij worden ze verhuisd uit het 0-netwerk en in hun eigen VLAN gezet: v-windows (met ID 47).
We brengen eerst de verschillende servers in kaart. De gegeven IP nummers zijn de nummers die de servers *na* de verhuizing zullen krijgen.

- Testdoos – 213.136.12.106 – Aanvankelijk was dit een testserver, zoals de naam doet vermoeden. Later echter ging deze server statistieken van webserver bezoeken bijhouden en publiceren. Klanten kunnen via deze server de drukte van hun eigen website bekijken. Deze server zal tevens een nieuwe naam krijgen uit de lijst met Indianenstammen: Bannock wordt zijn nieuwe naam.

- Pawnee – 213.136.12.99 – Dit is een oude Windows NT4 server die een MSSQL 7.0 server omvat. De klanten die hierop draaien worden zo spoedig mogelijk overgenomen door Eputary, een nieuwere versie MSSQL server.

- Eputary – 213.136.12.102 – Deze server draait MSSQL 2000, een ODBC database server. Alle databases van de Windows hostingklanten draaien op deze machine.

- Cheyenne – *geen* – Dit is de Primary Domain Controller (PDC) voor het Windows NT4 domain. Alle gebruikers en hun rechten worden op deze server aangemaakt en andere servers zullen vragen omtrent rechten aan de PDC stellen. Bij het uitvallen zal Ohlone deze taak overnemen.

- Ohlone *geen* – Dit is de Backup Domain Controller (BDC) voor het Windows NT4 domain. Alle gebruikers en hun rechten worden vanaf de PDC gekopieerd en hier opgeslagen om redundantie te kunnen bieden bij het uitvallen van de Primary Domain Controller.

- Geronimo – 213.136.12.104 – Deze Windows 2000 server verzorgt samen met Eskadi een loadbalanced webserver voor het “nieuwe” webhosting platform.

- Eskadi – 213.136.12.105 – Deze Windows 2000 server verzorgt samen met Geronimo een loadbalanced webserver voor het “nieuwe” webhosting platform. Beide servers maken gebruik van een Windows 2000 domain, een SQL server (Eputary) en de fileserver (NetApp).

- Wintun – 213.136.12.144 – Deze server wordt gebruikt als VNC en RDP[1] server voor het Windows platform. Beide protocollen stellen gebruikers in staat om

[1]Remote Desktop Protocol – een Microsoft protocol om over het Internet de desktop van een Windows2000 server te kunnen raadplegen en via deze server te kunnen werken, een tegenhanger van het X Window System voor Unix

over het Internet verbinding te maken met interne Windows servers om ze zodoende te kunnen beheren.

- Wichita – 216.136.12.108 – Dit is een dedicated Windows98 server die maar één doel heeft: het uitvoeren van een MS-DOS applicatie die een financieringsberekening doet voor een klant die een autoverzekering website uitbaat.
- Cayuga – 213.136.12.107 – Dit is het "oude" webhosting platform gebaseerd op één NT4 server die in totaal 115 IP nummers heeft waarbij op ieder IP adres één specifieke website draait.

NB. De twee servers cheyenne en ohlone zullen niet worden omgenummerd. Zodra we op het Windows 2000 domain productie draaien zullen ze worden ontmanteld en hergebruikt.

De NetApp

De grootste uitdaging bij deze migratie is de fileserver. We gebruiken hiervoor een hardware oplossing, geboden door Network Appliances. Deze maakt ook een verbinding met de PDC van een NT4 domain en zal hiermee zogenaamde *security IDs* maken, kortweg SIDs genoemd.
Om de NetApp Filer op een juiste manier te laten deelnemen in het Windows 2000 domain, blijkt dat er een software upgrade van het besturingssysteem van de Filer zal moeten plaatsvinden. Voor de migratie draaide de machine Data ONTAP versie 6.2R1, maar er is minimaal versie 6.2.1 nodig om op een goede manier met het Windows 2000 domain te kunnen werken.
Dit blijkt erg gemakkelijk te kunnen door de versie te downloaden (Business Internet Trends heeft een service contract) en de NetApp te rebooten. Volgens de handleiding zal dit zo'n twee minuten duren en dit is tevens de tijd waarin de Filer geen bestanden meer zal uitleveren aan zowel het Windows als het Unix serverpark (die via NFS met de Filer communiceert).
Bij het overschakelen naar een Windows 2000 domain, zullen alle SIDs ongeldig worden en zal de gehele directorystructuur moeten worden nagelopen en de bestanden worden voorzien van nieuwe SIDs. Helaas blijkt er geen geautomatiseerde mogelijkheid te zijn om dit klusje te klaren. Er wordt dus afgesproken dat we dit in de nachtelijke uren doen tijdens het onderhoudsvenster wat hiervoor geplanned staat.

Wintun

Wintun wordt door engineers van Business Internet Trends gebruikt om op in te loggen. De server wordt maximaal vertrouwd door alle interne servers en door klanten en vanaf Wintun kan men dan doorloggen naar deze servers. Hierdoor hoeven we niet voor *alle* IP nummers van thuisPCs (die doorgaans bij kabelmodem providers zoals @Home en Chello draaien) een toegangsregel toe te voegen in de firewalls van de te beheren servers.
We besluiten dat deze server niet in v-windows thuishoort, maar in v-admin. Van dit VLAN hebben we besloten dat het in elke firewall als maximaal betrouwbaar wordt beschouwd. De wintun.bit.nl server kennen we dus een IP nummer toe uit v-admin, te weten 213.136.12.144.

Loadbalanced Webplatform

Het loadbalanced webplatform (op Geronimo en Eskadi) is reeds in productie; er draaien zo'n 25 websites op. De servers melden zich echter (net zoals de andere servers) aan het Windows NT4 domain aan. Dit zal moeten veranderen en de servers zullen zich moeten aanmelden aan Comanche (de nieuwe Windows 2000 DC). Zij zullen moeten kunnen communiceren met Eputary, de database server. Veel van de websites worden namelijk gegenereerd door middel van ASP[2].
Op dit moment is er in de DNS een CNAME aangemaakt voor ntvhost-1.bit.nl, die verwijst naar de beide servers. Webgebruikers die naar ntvhost-1.bit.nl connecteren, zullen in principe om en om naar Geronimo en Eskadi gestuurd worden. Dit is onwenselijk, omdat als één van de servers uitvalt, de helft van de webbezoeken zal falen. Dit is een uitstekend moment om LSNAT toe te passen, zoals we dat ook doen voor de SMTP en POP servers.
We kennen het IP adres 213.136.12.228 toe uit de loadbalance range. We zullen hiervoor intern het adres 10.0.80.3 aan Geronimo geven en 10.0.80.4 aan Eskadi. Na de migratie zullen we ntvhost-1.bit.nl niet meer via round-robin DNS naar de beide servers laten verwijzen, maar simpelweg naar het loadbalance VIP op de Riverstone router. Dit heeft een geweldig beschikbaarheidsvoordeel voor de klanten van dit loadbalanced platform.

Cayuga

Op deze server draait een Microsoft IIS server die op een behoorlijk aantal IP nummers dedicated website hosting verzorgt. Dit houdt in dat iedere website zijn eigen IP nummer heeft. Bij het verhuizen van de server zelf naar een nieuw adres zal er dus geen probleem ontstaan omdat we alle IP nummers van de websites die er op draaien, in de router statisch kunnen routeren van het oude adres naar het nieuwe adres van de server.

Winroute

Alle Windows servers van Business Internet Trends, zowel als van de klanten, draaien een Windows based firewalling programma wat Winroute heet. Bij het omnummeren van de servers moeten we er ter dege rekening mee houden dat we op alle betrokken servers de nieuwe reeks IP nummers voor v-windows instellen (dit is 213.136.12.96/28). Ook is het verstandig om het v-admin netwerk (dit is 213.126.12.128/26) toe te staan in de firewalls van de klanten (en van onszelf). Er zal voor iedere server nagelopen moeten worden of de server firewall regels heeft en waar nodig zullen aanpassingen gemaakt moeten worden. Anders lopen we de kans dat we ergens niet meer in kunnen nadat we de IP nummers van (vooral Wintun) onze servers aangepast hebben.

[2] Active Server Pages – een Microsoft webserver scripting taal

De migratie

Allereerst zal er van het PDC/BDC[3] principe afgestapt moeten worden en zal er over gegaan worden op een Windows 2000 domain met bijbehorende DC. Hiervoor worden twee nieuwe DCs aangemaakt, comanche.bit.nl 213.136.12.100 en ebola.bit.nl 213.136.12.101 en het nieuwe domain zal "BIT.NL" heten.
Er zal 's nachts een reboot van de NetApp moeten gebeuren om hem te upgraden van 6.2R1 naar 6.2.1. Zowel het implementeren van de Windows 2000 DCs als het upgraden van de NetApp zullen we enkele dagen van te voren doen.
Daarna zullen de rechten van de bestanden op de NetApp in kaart worden gebracht. Het betreft zo'n 30 websites (waarvan er enkelen nog geen productie draaien), een honderdtal directories en enkele duizenden bestanden.
We zullen beginnen met Testdoos (de server die statistieken maakt). Deze server zal zich namelijk aanmelden aan de DC, maar verder draait de server geen cruciale taken en kunnen we op basis van deze server de anderen wellicht gemakkelijker verhuizen. We geven de server vanaf nu de naam bannock.bit.nl 213.136.12.106.
In een nachtelijk onderhoudsvenster zullen we de migratie uitvoeren. We beginnen met alle services uit te zetten. Hierdoor komen de servers tot rust en zijn ze wat handelbaarder. We voeren dan voor iedere server het volgende uit:

- Creëer op Comanche een computer account voor de om te nummeren server. Hij zal zich hiermee aanmelden op het Windows 2000 domain zodra hij in het v-windows VLAN actief wordt.
- Geef op de om te nummeren server zijn nieuwe IP nummer op. Stel in, dat hij zich moet aanmelden aan het Windows 2000 domain "BIT.NL". Reboot de server.
- Zet op RBFS de switchpoort van de server om van VLAN 50 (v-colo0) naar VLAN 47 (v-windows).
- Pas de DNS aan en verander de hostname naar het nieuwe IP. Laat de DNS versneld propageren door beide secondary DNS servers ook aan te passen en de resolvers te herstarten.

Nadat de server gereboot is, zal hij zich hebben aangemeld in het nieuwe domain. Achtereenvolgens voeren we dit uit voor Wintun, Bannock, Wichita, Cayuga, Eputary, Pawnee, en tenslotte Geronimo en Eskadi.
Op Geronimo en Eskadi zullen we de loadbalance RIPs toekennen (10.0.80.3 respectievelijk 10.0.80.4) en de DNS aanpassen zodat de websites verwijzen naar het loadbalance VIP 213.136.12.228.

3.5.7 Administratieve servers

Aan het begin van de afstudeeropdracht werd er vanaf enkele servers in het Business Internet Trends netwerk het beheer uitgevoerd. In principe zijn er zes servers:

- manitou.bit.nl – 213.136.0.33 – deze server dient als office-gateway voor alle medewerkers in de Kelvinstraat (Ede-HQ). Vanaf hier wordt veel dagelijks be-

[3] PDC/BDC – Primary en Backup Domain Controllers, Windows NT4 authenticatieservers

heer gepleegd dus Manitou staat in de firewall van vrijwel elke klantenserver of -router als “trusted”.

- omaha.bit.nl – 213.136.0.38 – omaha dient als lokale fileserver voor alle medewerkers in de Kelvinstraat (Ede-HQ). Vroeger diende deze server ook als router/firewall, maar is vervangen door Manitou. De oudste routers en servers hebben omaha nog als “trusted” in hun firewall staan.
- linux.bit.nl – 213.136.0.64 – deze server was de eerste all-purpose server voor de medewerkers van Business Internet Trends. Tevens draait hij als resolver (en dit hebben veel klanten zo ingesteld).
- chinook.bit.nl – 213.136.0.81 – chinook draait een voice-pakket welke in staat is om server-storingen via de (normale spraak) telefoon kenbaar te maken aan de storingsdienst engineer. Omdat hij veel checks moet doen op (klanten)servers, is hij veelal trusted.
- saginaw.bit.nl – 213.136.0.86 – deze server wordt gebruikt als http://noc.bit.nl/ waarop alle verkeers-statistieken verkrijgbaar zijn. Derhalve is hij “trusted” op elke switch of router in ons netwerk.
- pokagon.bit.nl – 213.136.0.130 – deze server draait het pakket NetSaint, dat door middel van periodieke checks ervoor waakt dat alle diensten en servers van zowel Business Internet Trends als haar klanten up zijn en goed draaien. De checks zijn soms op niet-publieke diensten waardoor de meeste firewalls deze server doorlaten.

Na uitvoerig overleg met de opdrachtgever en de domeindeskundige wordt besloten om manitou.bit.nl en linux.bit.nl niet om te nummeren. Deze zullen voor onbepaalde tijd in het 0-netwerk blijven bestaan. Wel zullen we de overige servers verhuizen naar een eigen VLAN voor administratieve servers (v-admin).
De medewerkers van Business Internet Trends kunnen dan gaandeweg overal firewalls updaten zodat de nieuwe reeks IP nummers ook toegang heeft. Als er vanuit v-admin een server niet bereikbaar blijkt te zijn, kan altijd nog gebruik worden gemaakt van manitou.bit.nl en linux.bit.nl.
In principe zullen we op alle (klanten)servers en routers in het nieuwe netwerkdesign de reeks 213.136.12.128/26 beschouwen als hoogwaardig betrouwbaar. Hierin mogen dan enkel goedbeveiligde servers komen. Alle servers die nu in de firewall rules van onze klanten en onszelf staan, worden omgenummerd naar v-admin en het gehele VLAN wordt dan als “trusted” ingesteld op de servers.

3.5.8 Overigen

Binnen de ISP zijn er behoorlijk wat servers die niet in een bepaalde categorie vallen. Voorbeelden hiervan zijn:

- pala.bit.nl – 213.136.0.91 – PostgreSQL databaseserver.
- yurok.bit.nl – 213.136.0.121 – MySQL databaseserver.

- hoh.bit.nl – 213.136.0.103 – deze server wordt gebruikt om andere servers in de colocatie faciliteit hardwarematig te resetten. Er ligt een seriële bus door de colo, die aan de hand van rack en positie van één bepaalde server de resetknop één seconde lang kan indrukken.
- viejas.bit.nl en elem.bit.nl – 213.136.0.115 en 213.136.0.89 – deze servers worden gebruikt om webstatistieken te tonen aan de klanten.
- camserv-1.bit.nl t/m camserv-4.bit.nl – 213.136.0.92 - 213.136.0.95 – dit zijn vier embedded linux servertjes die ieder vier analoge videocamera's digitaliseren en via een web-interface tonen.
- lummi.bit.nl – 213.136.0.117 – met deze server tellen we al het verkeer wat tussen Ede en Amsterdam wordt gestuurd. We gebruiken hier een een zelfgemaakt programma, YAPS. In hoofdstuk 5 wordt dit programma uitvoerig beschreven.

Alle bovengenoemde servers worden naar een VLAN voor "overige" servers verhuisd: v-misc. Ze worden omgenummerd en krijgen ieder een adres in 213.136.12.192/27.
Verder zijn er nog tijdelijke servers en servers die niet meer gebruikt worden. Deze worden simpelweg verwijderd uit het netwerk en opgeslagen binnen de afdeling engineering. Sommigen worden opnieuw ingezet bij nieuwe projecten.

Hoofdstuk 4

Implementatie

4.1 Inleiding

Nadat het netwerkontwerp is goedgekeurd en het migratieplan is geschreven, is het tijd om het ontworpen netwerk te implementeren en het huidige netwerk van Business Internet Trends te migreren naar de gewenste situatie.
Dit hoofdstuk beschrijft de genomen stappen om tot de daadwerkelijke implementatie van het nieuwe netwerk ontwerp te komen. Omdat de financiering van het project niet in één keer rond was, is besloten de implementatie in drie subfasen op te splitsen:

1. Constructie van een gedeelte van het core netwerk met Alpine switches, beschreven in paragraaf 4.4.
2. Verhuizing van de ISP servers en diensten naar hun eigen VLAN, beschreven in paragraaf 4.6.
3. Uitbreiding van het corenetwerk met Juniper routers, beschreven in paragraaf 4.5.

4.2 Netwerkmigratie

Het netwerk is op drie tijdstippen aangepast. In een vroeg stadium is de Cisco 7206 uit het 0-netwerk gehaald. Daaropvolgend zijn de verschillende firewalls, loadbalancers en VRRP groepen aangemaakt.
In een tweede stap zijn de Alpine switches ingezet, waarbij de dark fiber van Level3 in productie is genomen.
Tenslotte is het netwerk verrijkt met twee (van de vier) Juniper routers zodat de Cisco routers ontmanteld en elders ingezet kunnen worden.

4.3 De Riverstone routers

In de nacht van 11 op 12 juli 2002 hebben we om 04:00 de Cisco router met de SSR2 router omgewisseld conform het stappenplan in paragraaf 3.4.1. De Riverstones verzorgen vanaf 12 juli de ontsluiting van het 0-netwerk en de Cisco 7206 VXR draait als backup router in het Ede Backbone VLAN (v-edebb). Omdat nu al het

lokale verkeer in Ede door één van de twee SSR routers moet, is de volgende stap het aanmaken van filters, loadbalancers en VRRP groepen op die SSRs, zodat het interne netwerk van Business Internet Trends beschermd wordt en de gewnste redundantie verkrijgt.

4.3.1 Riverstone: IP filters creëren

Omdat *ed-bfr* aanvankelijk al het IP verkeer verwerkte, was hij de preferente plaats om firewalls aan te maken. Er kleefden twee nadelen aan deze situatie. Ten eerste is de Cisco router niet in staat om "wire-speed" te filteren. Bovendien heeft de Cisco niet een handige interface om de filters te specificeren. Riverstone kan in hardware filteren en heeft een handige interface om filters te specificeren. Om een IP filter (ook wel Access Control List ofwel ACL) te implementeren, wordt het volgende getypt:

```
1 : acl a-dapje permit ip 193.109.122.224/28 any
2 : acl a-dapje permit udp any 193.109.122.224/28 any >1024
3 : acl a-dapje permit tcp any 193.109.122.224/28 any any established
4 : acl a-dapje permit tcp 193.109.122.0/24 193.109.122.224/28 any 22
5 : acl a-dapje permit tcp 213.136.0.0/19 193.109.122.224/28 any 22
6 : acl a-dapje permit tcp any 193.109.122.224/28 any 25
7 : acl a-dapje permit tcp any 193.109.122.224/28 any 80
8 : acl a-dapje permit tcp any 193.109.122.224/28 any 113
9 : acl a-dapje permit icmp
10: acl a-dapje deny ip
11: acl a-dapje apply interface i-dapje output
```

Uitleg bij deze firewall

We noemen vanaf nu de IP reeks 193.109.122.224/28 gewoonweg v-dapje. Op regel 1 staan we al het IP verkeer (dit is UDP, TCP en ICMP) toe die komt uit v-dapje en gaat naar elk willekeurige IP adres (het woordje "any"). Op regel 2 staan we al het UDP verkeer toe die eender waar vandaan komt en gaat naar v-dapje. Het mag hierbij van elke poort komen, maar alleen gaan naar een UDP poort boven de 1024. Op regel 3 staan we al het TCP verkeer toe die overal vandaan mag komen, en van elke willekeurige poort op zowel de server buiten, als de server binnen v-dapje, maar waarbij we alleen maar "established TCP" toestaan.
Dit verdient wellicht enige uitleg. In het TCP/IP model worden verbindingen als volgt opgezet. Een server A die een connectie aanvraagt met poort 80 van server B zal hier een willekeurige nog ongebruikte poort boven de 1024 voor kiezen (Engels: empherial port), bijvoorbeeld 1234. Het eerste pakket wat gestuurd wordt door A naar B heeft een speciale vlag gezet, de Syn vlag genoemd. Als B de connectie wenst te accepteren wordt er een pakket terug gestuurd met twee vlaggen gezet Syn en Ack. Alle volgende pakketjes zullen het Ack vlaggetje gezet hebben. Wezenlijk is het hier om te onderkennen dat er slechts één pakket géén Ack bit gezet heeft: de originele aanvraag.
Terug naar onze filter. Het established commando in regel 3 zorgt er voor dat er enkel pakketjes worden toegestaan die de Ack vlag gezet hebben. Alle pakketjes die enkel het Syn vlaggetje hebben, worden hierdoor dus geweerd. In essentie zorgt

regel 3 er voor dat er geen inkomende connecties worden doorgelaten. Alle uitgaande connecties worden reeds toegestaan door regel 1.
Regel 4 en 5 staan inkomende aanvragen voor de Secure Shell (ssh) service toe van 193.109.122.0/24 en 213.136.0.0/19 naar v-dapje. Regel 6, 7 en 8 staan inkomende aanvragen toe voor poort 25 (smtp), 80 (http) en 113 (identd) vanaf elk IP adres op het Internet naar v-dapje. Regel 9 staat alle mogelijke ICMP verkeer toe van het Internet naar v-dapje. Regel 10 tenslotte verbiedt elke vorm van communicatie.
In de laatste regel wordt deze ACL toegepast op het interface i-dapje (die de default gateway is voor servers in het VLAN "v-dapje" in het IP netwerk 193.109.122.224/28 zitten).

4.3.2 Riverstone: Loadbalancer aanmaken

Een van de voordelen van het inzetten van Riverstone routers in het netwerk van Business Internet Trends is hun mogelijkheid tot hardware loadbalancing, gebruik makend van het LSNAT protocol welke besproken is in paragraaf 2.6.5 op pagina 36.
Bij het instellen van de loadbalance kan men als volgt te werk gaan:

```
1: interface create ip i-mail address-netmask 213.136.12.1/28 vlan v-mail
2: interface add ip i-mail address-netmask 10.0.25.1/24
3: load-balance create group-name lb-smtp virtual-ip 213.136.12.226 \
      virtual-port 25 protocol tcp
4: load-balance add host-to-group 10.0.25.3 port 25 group-name lb-smtp \
      weight 100
5: load-balance add host-to-group 10.0.25.4 port 25 group-name lb-smtp \
      weight 100
6: load-balance set group-options lb-smtp app-int 20 ping-int 30
```

Uitleg bij deze loadbalancer

Allereerst verklaren we het gebruik van de RFC1918 space in bovenstaande implementatie. Als we de 'echte' IP nummers van onze te loadbalancen servers gebruiken werkt de loadbalance naar behoren, maar zijn we niet meer in staat om te connecteren naar de servers! Dus als we op cherokee en oneida poort 25 willen connecteren lukt dit niet meer. Daarom hebben we geprobeerd om een tweede IP netwerk in het v-mail VLAN te maken. Hierbij hebben we de logische nummering van 10.0.25/24 gekozen (omdat poort 25 de well-known poort is voor SMTP).
Regel 1 maakt het primaire interface aan voor het v-mail VLAN 213.136.12.1/28. Regel 2 maakt een tweede interface aan voor dit VLAN op 10.0.25.1/24. Op de SMTP servers cherokee en oneida maken we nu ook een tweede interface aan op 10.0.25.3 respectievelijk 10.0.25.4. Op regel 3 creëren we de virtuele servergroep lb-smtp op een IP nummer in ons VLAN voor loadbalancing. We kiezen het IP nummer 213.136.12.226 als VIP en laten de router luisteren op TCP poort 25.
Op regel 4 en 5 voegen we cherokee en oneida toe aan de servergroep lb-smtp. We geven aan dat de router aan beide servers hetzelfde gewicht moet koppelen, dus dat allebei de servers evenveel aanvragen krijgen.
In regel 6 specificeren we dat de router elke 30 seconden de servers in de groep moet pingen (met ICMP echo pakketjes) en dat hij elke 20 seconden moet control-

eren of de SMTP dienst op de servers in de groep nog draait. Als dit niet het geval is, zal de router automatisch de server uit de groep verwijderen en pas terugzetten als de server weer pingt en de SMTP dienst twee minuten achtereen weer operationeel is.

4.3.3 Riverstone: VRRP activeren

Om het Virtual Router Redundancy Protocol te activeren hebben we twee routers nodig. De beide routers krijgen interfaces met de laagste twee IP nummers van het subnet. De primaire router wordt de naam "vlan-gw.network.bit.nl" gegeven, en de backuprouter wordt "vlan-vrrp.network.bit.nl" genoemd. We configureren VRRP als volgt op de SSR routers:

```
1: ip-redundancy create vrrp 1 interface i-ic
2: ip-redundancy create vrrp 2 interface i-colo0
3: ip-redundancy create vrrp 2 interface i-colo1
4: ip-redundancy associate vrrp 1 interface i-ic address 213.136.31.65/28
5: ip-redundancy associate vrrp 2 interface i-colo0 address 213.136.0.1/24
6: ip-redundancy associate vrrp 2 interface i-colo1 address 213.136.19.1/26
7: ip-redundancy start vrrp 1 interface i-ic
8: ip-redundancy start vrrp 2 interface i-colo0
9: ip-redundancy start vrrp 2 interface i-colo1
```

Uitleg bij deze VRRP commando's

Voor beide routers wordt achtereenvolgens VRRP groepen aangemaakt in de commando's 1, 2 en 3. Groep 1 bevat één interface, 'i-ic'. Groep 2 bevat er twee, i-colo0 en i-colo1. Vervolgens worden de addressen van de VRRP routers aangegeven (4–6) en tenslotte in 7–9 worden deze VRRP groepen actief gemaakt.
Op SSR1 is er daadwerkelijk een interface voor 213.136.31.65, 213.136.0.1 en 213.136.19.1. Op SSR2 is dit niet het geval (die zit op 213.136.31.66, 213.136.0.2 en 213.136.19.2 respectievelijk). SSR2 zal dus voor deze interfaces de backup router worden.
Door op SSR1 het commando `ip-redundancy show vrrp id 1 interface i-ic` te typen, zien we de status van de VRRP groep.

```
VRRP Virtual Router 1 - Interface i-ic
----------------------------------------
Uptime                  5 days, 21 hours, 19 minutes, 0 seconds.
State                   Master
Priority                255 (default value)
Virtual MAC address     00005E:000101
Advertise Interval      1 sec(s) (default value)
Preempt Mode            Enabled (default value)
Authentication          None (default value)
Primary Address         213.136.31.65
Associated Addresses    213.136.31.65
```

Bij de implementatie van VRRP in het netwerk van Business Internet Trends leken we tegen een fysieke beperking van het aantal mogelijke interfaces aan te lopen.

Het aantal was vijf. Omdat dit voor ons netwerk niet voldoende blijkt, hebben we contact gezocht met Riverstone. Er bleek dat de routers wel degelijk meer VRRP groepen aankonden, maar dat ze in totaal maar vijf MAC adressen (van de vorm 00:00:5E:00:01:**xx**) aan kunnen maar dat er geen bezwaar is tegen het hergebruiken van hetzelfde MAC adres voor meerdere VRRP groepen. We gebruiken dus nu twee VRRP id's (dus twee MAC adressen), één voor de eigen VLANs en één voor de klanten VLANs.

4.4 De Alpine coreswitches

Allereerst wordt de in het migratieplan beschreven verhuizing van *ed-bfr* uit het 0-netwerk verzorgd. Daarna zullen de nieuwe switches in Amsterdam worden aangesloten op de darkfiber en de internet koppeling worden geupgrade van 100 Mbps naar 1 Gbps. Tenslotte wordt de routering van de ISP van de STM1 en E3 ontkoppeld en op de gigabit darkfiber ingekoppeld. Hiervoor zullen nog altijd de bestaande Ciscos worden ingezet.
Het configureren van de Alpine switches bestaat over het algemeen uit drie taken. Allereerst moeten we de EAPS ringen aanmaken die gebruikt zullen worden. Daarna zullen we VLANs moeten maken en deze koppelen aan één van de ringen. Tenslotte zullen we de Gigabit en FastEthernet poorten moeten configureren in de VLANs. In de regel maken we trunkpoorten van alle EAPS ringpoorten, zodat we daar meer dan één VLAN over kunnen transporteren.

4.4.1 Extreme: EAPS configureren

Het aanmaken van een EAPS ring gebeurt op een Extreme als volgt, aangenomen dat deze switch de master switch is en poort 1:1 de primary poort is en 1:2 de backup (dus initiëel geblokkeerde) poort. Zie verder paragraaf 2.6.1 voor een uitleg van EAPS.

```
1: enable eaps
2: create vlan v-ring1-control
3: config v-ring1-control qosprofile Qp8
4: config v-ring1-control tag 123
5: create eaps e-ring1
6: config e-ring1 mode master
7: config e-ring1 e-ring1 primary 1:1
8: config e-ring1 e-ring1 secondary 1:2
9: config e-ring1 control vlan v-ring1-control
10: enable eaps e-ring1
```

Uitleg bij deze EAPS commando's

Bij stap 1 zetten we de EAPS feature op de switch aan. We creëren een VLAN die gebruikt wordt voor de interne communicatie met stap 2. In 3 zorgen we er voor dat het verkeer van dit VLAN de allerhoogste prioriteit krijgt in de switch. Hierdoor zullen EAPS berichten altijd voorrang krijgen boven 'normaal' verkeer. Stap (4) stelt voor het VLAN v-ring1-control het id 123 in. We maken vervolgens de ring aan (5)

en maken deze switch met commando (6) de *master* switch. Alle andere switches in de ring behoren nu *transit* switches te zijn. Het verschil is dat de master switch zijn primaire poort (ingesteld met 7) open heeft, en zijn secundaire poort (zie 8) gesperd. Hierdoor kan het verkeer niet in cirkels rondrennen over de ring. Met (9) geven we aan dat we v-ring1-control gebruikt moet worden om besturingsberichten over de EAPS status te sturen. Vervolgens wordt de EAPS ring aangezet in (10).
De master switch zal dan elke tiende seconde een bericht samenstellen en naar zijn buurman op de primaire poort sturen (dus het bericht wordt verstuurd op poort 1:1). Alle transitswitches zullen alle berichten die ze in het VLAN v-ring1-control ontvangen meteen doorsturen naar de andere poort, net zolang totdat de laatste switch in de ring het pakketje weer uitstuurt en het bij poort 1:2 op de master switch terug komt. Deze switch weet nu dat de ring naar behoren functioneert.
Als een transit switch een buurman ziet downgaan, zal hij een alarmbericht samenstellen en naar zijn 'levende' buurman sturen. Deze bereikt uiteindelijk ook de master switch, die vervolgens constateert dat er een ringbreuk is ontstaan en zal per direct de secundaire poort ontsperren zodat de ring nu is verworden tot één ethernet lijn.

4.4.2 Extreme: VLANs configureren in de EAPS ring

Het aanmaken van een VLAN "v-test" en die toevoegen aan een EAPS ring, alsmede het toevoegen van poorten (al dan niet 802.1q getagged) aan een VLAN gebeurt als volgt:

```
1: create vlan v-test
2: config v-test tag 200
3: config e-ring1 add protect vlan v-test
4: config v-test add port 2:1-2:8,2:16 untagged
5: config v-test add port 1:3 tagged
```

Uitleg bij deze VLAN commando's

We maken een VLAN die we v-test noemen in stap (1) en in (2) kennen we aan dit VLAN de tag 200 toe. Om het VLAN redundant over de EAPS ring te laten transporteren, koppelen we het VLAN aan de ring in (3). Als we access poorten in v-test willen zetten (poort 2:1 tot en met 2:8 en 2:16), gebruiken we een statement zoals in (4). Om het verkeer, voorzien van een 802.1q VLAN tag naar buiten te laten sturen op poort 1:3 (bijvoorbeeld omdat daar een router aanhangt), gebruiken we (5).

4.5 De Juniper routers

Tenslotte zal het resterende gedeelte van de netwerkmigratie worden uitgevoerd waarna het gehele bedrijf productie draait op de Juniper routers in samenwerking met de SSRs. De Cisco's zullen nu als backup routers dienen totdat de laatste darkfiber wordt opgeleverd en de layer2 ring kan worden geïmplementeerd.

4.5.1 Juniper: BGP sessies

In de rand van het netwerk van Business Internet Trends, op de locaties Sara en Telecity2 in Amsterdam, moeten er routes worden uitgewisseld met peers op AMS-IX en transit providers in beide locaties. Hiervoor wordt BGP gebruikt.
Om op een Juniper BGP te configureren, moeten we een aantal zaken onderscheiden. In principe maken we een onderscheid tussen drie verschillende sessies:

- Peers op AMS-IX, waarvan we maar één of enkele routes leren,
- Transitproviders, waarvan we de volledige internet routetable aangeboden krijgen (dit zijn ongeveer 120.000 routes), en
- Onze eigen routers, waaraan we met iBGP alle routes die we geleerd hebben uitwisselen.

Met de Juniper kunnen we onze BGP sessies groeperen. We creëren enkele groepen met daarin peers met dezelfde parameters en we noemen ze 'AMSIX', 'Transit' en 'iBGP'.
Het configureren van BGP op een Juniper gaat als volgt:

```
edit routing-options
1: set autonomous-system 12859

edit protocols bgp
2: set group AMSIX type external
3: set group AMSIX local-address 193.148.15.200
4: set group AMSIX neighbor 193.148.15.166 peer-as 3265
5: set group AMSIX neighbor 193.148.15.166 description "XS4ALL"
```

Bij stap 1 wordt het lokale AS nummer ingesteld (het AS nummer van Business Internet Trends is dus 12859). In stap 2 wordt de groep AMS-IX gedefinieerd als externe groep. Alle uitgaande connecties zullen worden gemaakt vanuit het adres 193.148.15.200, wat is ingesteld in stap 3. Er wordt in 4 een "neighbor" gemaakt voor AS3265, die bereikbaar is op het IP adres 193.148.15.166. Tenslotte wordt in stap 5 aan deze sessie een omschrijving meegegeven. Dit is gemakkelijk als we later in de configuratie wijzigingen willen doorvoeren.

4.5.2 Juniper: BGP filtering

Zonder nadere configuratie zullen routers alle routes die ze leren van hun peers aan alle andere routers, waarmee ze een BGP sessie hebben, doorgeven. Dat zou betekenen dat onze routers voor alle anderen het verkeer zou gaan willen afhandelen. Dit is natuurlijk niet de bedoeling dus het is zaak dat we goede filters installeren op de sessies.
Voor alle sessies geldt, dat we enkel onze eigen routes en dat van onze klanten willen sturen. Op dit moment zijn dat drie autonome systemen: AS12589 (wijzelf met 3 routes), AS3333 (RIPE, 1 route) en AS1200 (AMS-IX, 2 routes). We zullen dus in totaal 6 routes gaan annonceren aan onze peers.
We besluiten gebruik te maken van een kenmerk van BGP wat "community tagging" heet. Aan elke route die we leren, kunnen we één of meer nummers toekennen

van de vorm 12859:XXXXX. Aan de hand van deze communities kunnen we dan de uitgaande filters maken. We onderscheiden de volgende communities:

- 12859:1000 – alle routes die we geleerd hebben op AMS-IX
- 12859:2000 – routes die we leerden van transit providers
- 12859:3000 – onze eigen routes (van AS12859 dus)
- 12859:4000 – routes van onze klanten (AS3333 en AS1200 dus)

Op de Juniper maken we nu zogenaamde policy-options. Een dergelijke policy is als een klein programma'tje, die kan bepalen of een gegeven route wordt geaccepteerd of gewijgerd. Het is voor de hand liggend om de routes met community tag 12859:3000 en 12859:4000 aan onze peers te annonceren en dat we aan alle geleerde routes op AMS-IX de community 12859:1000 moeten toekennen.
We maken de volgende routing-policies:

```
edit policy-options
1: set community LOCAL 12859:3000
2: set community CUSTOMER 12859:4000
3: set as-path ASBIT "^1200$|^3333$|^12859|^$"

edit policy-options policy-statement Send-AMSIX
4: set term one from community [ LOCAL CUSTOMER ]
5: set term one then accept
6: set term else then reject

edit policy-options policy-statement Recv-ASBIT
7: set term one from as-path ASBIT
8: set term one then accept community add CUSTOMER
9: set term else then reject
```

In stap 1 en 2 definieren we eerst de community tags. We geven ze de logische namen LOCAL respectievelijk CUSTOMER. Daarna definieren we een as-path door middel van een reguliere expressie (3), die alle routes van AS1200, AS3333 en AS12859 accepteert.
In stap 4–6 maken we een filter die alle routes toestaat die voorzien zijn van de community tag LOCAL of CUSTOMER, en verwerpen we de rest (6).
In stap 7–9 maken we een filter die aan alle routes die voldoen aan het as-path ASBIT (dus onze routes en die van onze klanten), de community tag CUSTOMER (12859:4000) toevoegt. Alle andere routes worden verworpen.
Nadat we de filters hebben gecreëerd, moeten we ze nog toepassen op de sessies. Dit doen we in de BGP configuratie als volgt:

```
edit protocols bgp group AMSIX
1: set export Send-AMSIX
2: set neighbor 193.148.15.1 peer-as 1200 import Recv-ASBIT
3: set neighbor 193.148.15.68 peer-as 3333 import Recv-ASBIT
4: set neighbor 193.148.15.166 peer-as 3265
```

In 1 zorgen we er voor dat voor alle neighbors in de group AMSIX dezelfde uitgaande updates worden gedaan: Send-AMSIX. Voor alle peers in de group geldt, dat we alle inkomende routes accepteren. In 2 en 3 zullen we de filter Recv-ASBIT, die een communitytag toevoegt op de ontvangen routes van RIPE (AS3333) en AMS-IX (AS1200). Deze routes zullen dan weer door de Send-AMSIX filter worden doorgelaten, omdat dit filter alle routes doorlaat met community LOCAL of CUSTOMER.
Dan is er nog maar één aspect over: hoe geven we onze eigen routes een community tag 'LOCAL'? Dit doen we in de routing-options sectie van de configuratie:

```
edit routing-options
1: set aggregate route 213.136.0.0/19 passive community 12859:3000
2: set aggregate route 193.109.122.0/24 passive community 12859:3000
3: set aggregate route 194.153.157.128/27 passive community 12859:3000
```

In 1–3 geven we aan met de commando 'aggregate route' dat de daaropvolgende routes door onszelf worden geannonceerd en dat er de community 12859:3000 (die in de policy-options configuratie 'LOCAL' heet) aan toegekend dient te worden.

4.6 Servermigratie

Nadat het eerste deel van het nieuwe core netwerk is gebouwd en de ISP hier productie mee draait hebben we besloten alle servers te verhuizen naar hun eigen VLAN. Deze migratie heeft verreweg de grootste inspanning gevergd, omdat het van cruciaal belang is dat de diensten tijdens en na de verhuizing consistent blijven draaien. Een service-onderbreking kan immers niet of nauwelijks getolereerd worden.

4.6.1 Stappenplan serververhuizing

Van alle servers die we willen omnummeren, geldt in principe het volgende stappenplan:

Stap 1 – Opstartscripts omzetten

Voor Linux wordt de netwerkconfiguratie in het bestand /etc/rc.d/rc.inet1 gewijzigd in de nieuwe situatie. De eventuele firewall rules in /etc/rc.d/rc.firewall worden aangepast. Daarna wordt het opstartscript /etc/rc.d/rc.inet1 gedraaid waardoor de server op zijn netwerkkaart het nieuwe IP nummer instelt. De verbinding blijft hangen (omdat we zijn ingelogd via het oude, dus door deze actie niet meer bestaande IP nummer).
Voor FreeBSD wordt de netwerkconfiguratie in het bestand /etc/rc.conf opgeslagen. Eventuele firewall rules worden hier in /etc/rc.firewall.local aangepast. Nadat de bestanden zijn aangepast wordt de situatie actief gemaakt door middel van het ifconfig(1) programma.

Stap 2 – Op *rbfs* het VLAN omnummeren

We zoeken uit welke ethernet-poort de server op de switch gebruikte. Deze poort staat in VLAN 50 (het v-colo0 VLAN waarin we elke server in het 0-netwerk hebben

omgenummerd). Hij zal in het juiste VLAN gezet moeten worden. Na enkele seconden zal de switch de verandering doorgevoerd hebben en is de server op zijn nieuwe adres te bewonderen.

Stap 3 – De DNS aanpassen

Op de primary nameserver moeten de hostname in 'bit.nl' en het IP adres in een reversed zonefile worden aangepast. Hiertoe wordt de server uit 0.136.213.in-addr.arpa verwijderd en in 12.136.213.in-addr.arpa toegevoegd.

Stap 4 – De DNS versneld propageren

Omdat gebruikers op het internet wellicht nog de oude naam of het oude adres van de server zullen hebben in hun cache, is het raadzaam onze eigen nameservers en caches te herstarten, zodat wij te allen tijde de juiste informatie uitleveren.
We zullen dus onze drie authoritatieve nameservers (ns1.bit.nl, ns2.bit.nl en ns3.bit.nl) aanpassen en tevens onze drie caching resolvers (linux.bit.nl, nscache1.bit.nl, nscache2.bit.nl) herstarten.

Stap 5 – Pokagon en Chinook bijwerken

De server bewakingssystemen werken veelal IP gericht. Zo heeft Netsaint van elke dienst die we aanbieden een referentie naar het IP nummer waarop deze draait. De configuratiebestanden van Netsaint zullen dus op pokagon.bit.nl moeten worden omgezet. Voor het voiceresponse-systeem op chinook.bit.nl geldt hetzelfde.

4.6.2 Volgorde serververhuizing

Aan de basis van het netwerk staat het DNS (nameservers). Daarnaast is het e-mail platform (zowel inkomend als uitgaand) van het grootste belang voor het juist functioneren van een ISP. Tabel 4.1 toont welke servers in welke onderhoudsvensters worden omgenummerd. In tabel 2.3 op pagina 39 bespreken we de VLANs en het voorgestelde numberplan.

Datum	*Dienst*	*Betrokken servers*
15-08-2002	DNS	nsauth1, nsauth2, ns1, ns2
26-08-2002	MAIL	oneida, cherokee
26-08-2002	YAPS	lummi
02-09-2002	MAIL	pomo, elwha, hopi, crow, seneca
03-09-2002	MISC	yurok, hoh, elem, camserv-1, camserv-2, camserv-3, camserv-4, cam-1, viejas
03-09-2002	NNTP	seminole
20-09-2002	Windows	cherokee, ebola, cheyenne, ohlone, pawnee, eputary, wichita, geronimo, eskadi, wintun, bannock
15-10-2002	Admin	iowa, pokagon, kickapoo

Tabel 4.1: *De volgorde waarin we de diensten van Business Internet Trends hebben omgenummerd en de datum waarop we deze verhuizingen hebben gerealiseerd.*

4.7 Conclusies en aanbevelingen

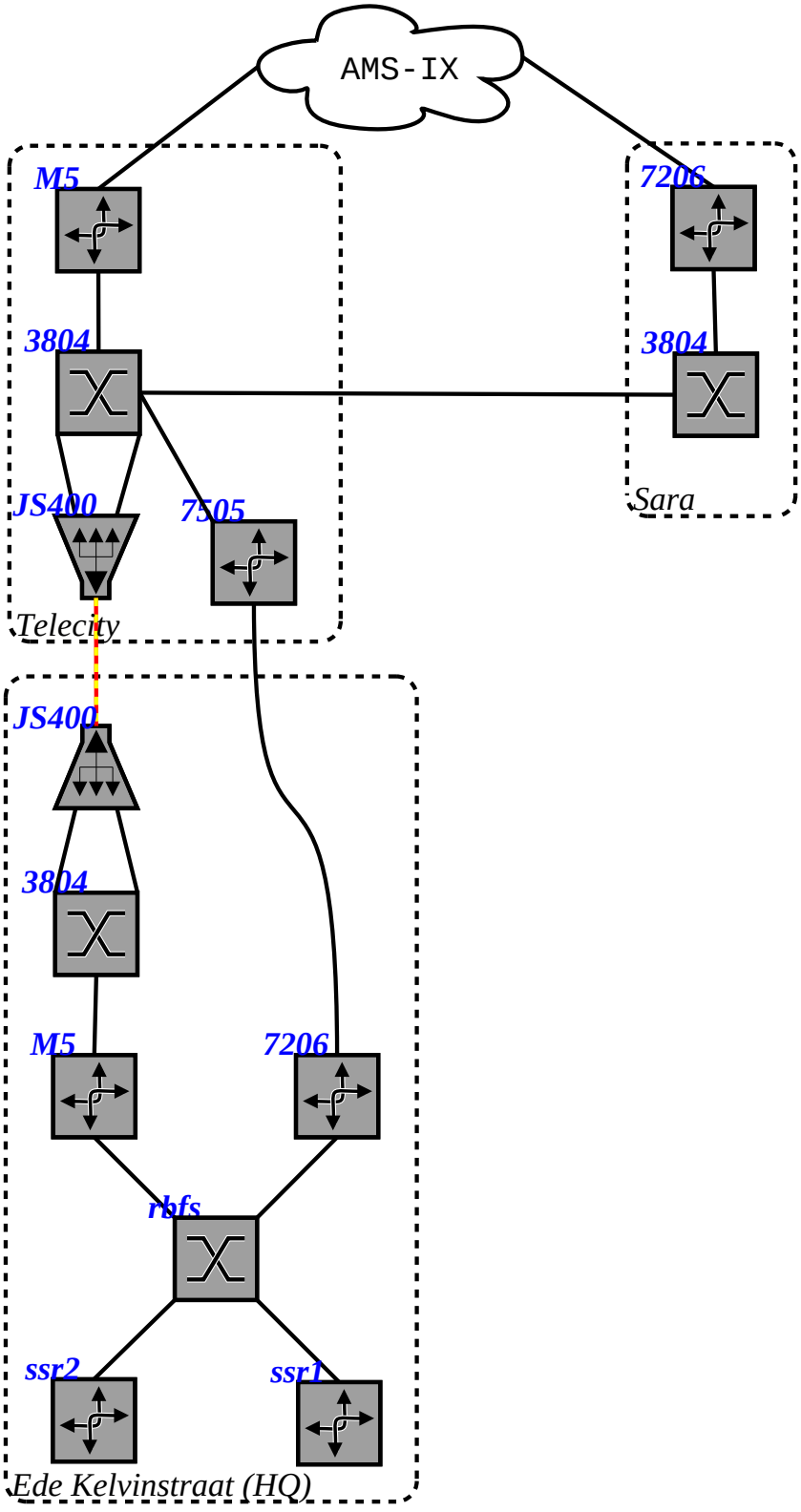
Nadat verreweg de meeste servers van Business Internet Trends, zowel als het transmissienetwerk, de layer2 switches en de layer3 IP routers op hun plek staan, komt de afstudeeropdracht tot zijn einde.
Gedurende de opdracht bleek het onmogelijk te zijn om per direct alle apparatuur te bestellen. Dit is ook aannemelijk, omdat het totale budget voor het project zo'n €275.000 is! Daarom is besloten het project stapsgewijs te implementeren. De helft per augustus/oktober van dit jaar, de andere helft in februari van 2003.
Daarom wordt het netwerk "half" opgeleverd. In figuur 1.1 op pagina 4 werd geschetst hoe het netwerk er bij lag aan het begin van de opdracht. Ter referentie schetsen we nu het netwerk zoals die er bij ligt aan het einde van de stageopdracht.

4.7.1 Conclusies

In figuur 4.1 wordt het netwerk schematisch weergegeven. De WDM verbinding tussen Ede Kelvinstraat (het hoofdkantoor) en de Telecity2 verbinding, gebruikmakend van de darkfiber van Level3, is up. De switches in alle drie de locaties zijn geïnstalleerd en draaien volop productie.
De darkfiber tussen Telecity2 en Sara in de Watergraafsmeer is ook operationeel. Hierop draaien twee 1000baseLX GBICs tussen de Extreme switches.
Er zijn twee Juniper routers ontplooid. In Telecity2 wordt op de AMS-IX met 1000baseSX ingekoppeld. Hierover wordt de primaire IPv4/IPv6 productie gedraaid. De Edese Juniper vormt de scheiding tussen het colocatie netwerk van Business Internet Trends, en de layer2 core van het bedrijf.
We concluderen dat het implementatie traject, dankzij de goede voorbereiding in de vorm van het migratie document en het ontwerp van het core netwerk, uitstekend is verlopen. De klanten waren te allen tijde op de hoogte van de werkzaamheden en ondervonden niet of nauwelijks hinder van de omschakelingen van de oude Cisco naar de nieuwe Extreme/Juniper core.

4.7.2 Aanbevelingen

De Cisco routers (Cisco 7505 in Telecity en Cisco 7206 in Ede) zouden nog gebruikt kunnen worden om een backup verbinding te hebben in de vorm van een STM1 (155 Mbps), mocht de darkfiber of de transmissie apparatuur blijken offline te gaan. Uiteraard is het zaak om zo snel mogelijk de derde darkfiber te installeren, zodat de layer2 ring afgemaakt kan worden en EAPS voor de bescherming kan zorgen.
De overige twee Juniper routers zouden het beste, conform de planning, in februari van 2003 worden besteld en ontplooid. Hierbij kan dit document als leidraad dienen voor de implementatie van de beide machines. Daarnaast kan worden gekeken naar de reeds aanwezige configuratie van de switches en routers in het netwerk.



Figuur 4.1: *Het Business Internet Trends netwerk, in intermediair stadium, aan het einde van de opdracht (november 2002).*

Hoofdstuk 5

YAPS – Yet Another Packet Sniffer

5.1 Inleiding

Business Internet Trends BV is een zakelijke internet provider die zich vooral in colocatie en zakelijke access specialiseert. Klanten kunnen voor wat het de colocation betreft kiezen tussen twee hoofdvormen. Enerzijds kunnen zij per maand een vaste bandbreedte afnemen, bijvoorbeeld 2 Mbps, tegen een vast tarief per maand. Anderzijds kunnen ze een hoeveelheid datavolume afnemen, waarbij de gebruikte bandbreedte mag variëren.
Voor de laatstgenoemde categorie is het belangrijk een goed overzicht te houden van het gebruikte datavolume per klant. Hiervoor wordt sinds geruime tijd een shareware programma ipcount[1] gebruikt.
Tegen het einde van elk kwartaal zal de debiteurenafdeling door middel van een lange lijst van trafficstatistieken, uit rekenen wat het verbruik per klant is. Dit is een kostbare bewerking, aangezien er van 120 dagen met een calculator moet worden opgeteld hoeveel verkeer er is verbruikt door elke klant. Sommige klanten hebben meerdere IP nummers. Soms wil de klant dat al zijn IP nummers samen worden gefactureerd. Soms is er afgesproken dat één of meer van de IP nummers op een ander tarief verkeer kopen, bijvoorbeeld omdat het een preferente klant betreft.

5.2 Probleem- en Doelstelling

Er blijken drie belangrijke problemen te zijn met de huidige oplossing:

- Het ipcount programma blijkt niet zo erg te schalen. Het is bedoeld om te draaien in kleine netwerken in de orde van grote 256 IP nummers en minder dan of gelijk aan 10 Mbps netwerk-belasting. Bij Business Internet Trends BV wordt een transitie van 155 Mbps naar 1 Gbps voorbereid.
- IPcount is niet erg stabiel. Het crasht dagelijks een aantal malen, waarbij de historie van het IP gebruik verloren gaat. Dit gebeurt meestal 's avonds rond negenen, wanneer de ISP zijn topdrukte qua verkeer beleeft.
- Er is behoefte aan een opsplitsing van het gebruikte IP verkeer per klant, wat niet wordt geboden door IPcount.

[1] Geen idee waar dit vandaan komt

We stellen ons ten doel om een programma te ontwerpen, die grote hoeveelheden IP verkeer (in de orde van grootte 800 Mbps) kan verwerken. Het programma moet robuust en configureerbaar zijn, en nauwkeurige tellingen van het verbruikte IP verkeer per klant kunnen bijhouden in een database.
Het programma moet portable gemaakt worden, zodat het kan draaien onder de volgende besturingssystemen: FreeBSD, NetBSD, OpenBSD en Slackware Linux.

5.3 Methoden en Technieken

We zullen het programma baseren op andere opensource projecten. We zullen het configuratiesysteem in flex[2] schrijven. De rest van het systeem zal in ANSI C worden geschreven en gecompileerd met de GNU C compiler *gcc* voor Unix.
We gebruiken CVS[3] om verschillende versies van de geïmplementeerde routines te kunnen bijhouden. Om het programma portable te maken (en te houden), zullen we gebruik maken van het software pakket GNU Autoconf[4].

5.4 Ontwerp

In deze sectie zullen we uitgebreid stilstaan bij de genomen beslissingen en het implementatietraject om tot het gewenste programma te komen. We beginnen bij de opdrachtgever die zijn eisen en randvoorwaarden formuleert. Aan de hand van deze specificatie, zullen we een executiemodel bouwen en hieraan invulling geven.
Er zullen zich een aantal uitdagingen voordoen in verband met het geheugen gebruik en de algoritmes die nodig zijn om de records van de IP nummers efficiënt te kunnen benaderen. Hiervoor wordt een hashfunctie gepresenteerd.
Daarnaast zullen we een algoritme moeten bedenken die kan afleiden of een gegeven IP pakketje een inkomend, of juist een uitgaand pakketje is. Deze berekening zullen we voor elk pakketje wat we ontvangen moeten uitvoeren, dus het is zaak dat dit algoritme snel is.
We staan even stil bij de PCAP bibliotheek. PCAP staat voor **P**acket **CAP**turing en wordt veel gebruikt door programma's die rauwe (Ethernet) data moeten lezen en verwerken.

5.4.1 De gewenste uitvoer

De opdrachtgever heeft maar één uitgesproken wens: Bouw het ipcount programma en zijn omgeving na.
Bij nadere inspectie blijkt dat het systeem alle IP verkeer wat de ISP in- en uitgaat realtime moet kunnen verwerken en dagelijks een rapport moet samenstellen van

[2] Fast LEXical scanner – Een utility om programmacode te genereren die patroon-herkenning op text doet. Het leest *.l* files en genereert *.c* files.

[3] Concurrent Versioning System – een stuk software wat geschreven is om bestanden vanaf verschillende lokaties door verschillende mensen te laten bewerken en centraal op te slaan in een zogenaamde *repository* (depot)

[4] Autoconf – Een set hulpmiddelen die de verschillende library- en kernelcalls in C includefiles kan zoeken en typedeclaraties kan forceren zodat alle operating systemen op dezelfde manier tegen de programmacode aankijken

de verbruikte bandbreedte per intern IP adres.
De opdrachtgever geeft te kennen dat een database koppeling gemakkelijker is omdat dan de tellingen per kwartaal geautomatiseerd kunnen worden.
De opdrachtgever wil graag elke nacht een statusrapport krijgen van het door de ISP verbruikte datavolume van de afgelopen dag. Daarnaast is het wenselijk om de hoeveelheid TCP, UDP en ICMP verkeer te meten en te rapporteren.
We definiëren dus de volgende deliverables voor dit software project:

- Het YAPS programma wat realtime de tellingen verzorgt
- Documentatie bij dit programma (dit document)
- Een rapportage programma wat een CSV[5] kan exporteren met de relevante data
- Een SQL database waarin we de CSV bestanden kunnen importeren

5.4.2 Het OSI-model revisited

Zoals we later zullen zien, is het erg gemakkelijk om verder te gaan met verkeerstellingen op een hoger niveau in het OSI model. We kunnen dus tellingen maken van al het in- en uitgaande UDP, TCP en ICMP verkeer per IP nummer. Tegelijk zullen we tellers laten meelopen voor de layer2 en layer3 protocollen die voorkomen op het netwerk van Business Internet Trends. We zullen in deze paragraaf verder in gaan op de IP stack zoals die is geïmplementeerd volgens het bekende ISO/OSI model.
Het OSI model bestaat uit zeven lagen. Het IP protocol heeft een invulling gegeven voor de onderste vier lagen. We doorlopen iedere laag en zoeken de bijbehorende definities in de header files van een Unix operating systeem. Uit de header definities destilleren we de gewenste informatie.

Layer1 – Transmissie

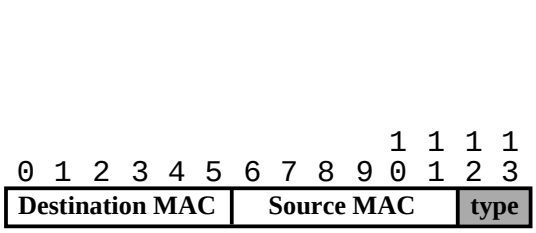
De eerste laag van het OSI model wordt voorzien door dedicated hardware op de Ethernet (netwerk)kaart van de PC. Deze laag voorziet het fysiek uitzenden en ontvangen van Ethernet pakketjes op een netwerkkabel. We hebben verder niet zo veel met deze functionaliteit te maken; de driver van onze kernel verzorgt deze in samenwerking met de netwerkkaart zelf.

Layer2 – Medium Access Control

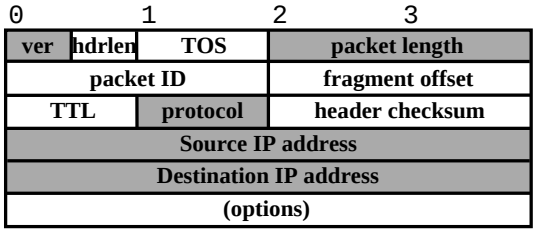
Voor Ethernet geldt, dat frames worden verzonden van en naar een zogenaamd MAC adres. Een MAC adres is 48 bits groot (6 bytes). De header van een Ethernet frame bestaat uit een destination MAC adres gevolgd door een source MAC adres en tenslotte het Ethernet type (16 bits unsigned integer). De headerlengte wordt daarmee dus 14 bytes. Zie figuur 5.1(a).
Aan de hand van het Ethernet type kan de kernel onderscheiden wat voor verkeer dit ethernet frame bevat. PCAP verschaft ons rauwe Ethernet frames, dus inclusief

[5] Comma Seperated Values – een standaard bestandsformaat voor SQL gebaseerde databasesystemen

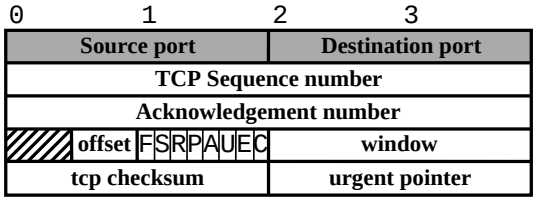
header en type. Het is aan ons om een verdere analyse te doen. We zoeken op dat het typenummer voor IPv4 0x800 is. Als er in het type veld dus een andere waarde staat, is het geen IP verkeer en zullen we het frame niet verder behandelen.



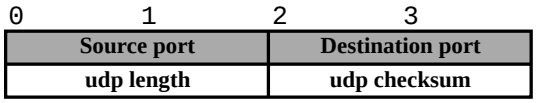
(a) De header van een Ethernet frame (14 bytes)



(b) De header van een IPv4 pakket (20 bytes, zonder opties)



(c) De header van een TCP datagram (20 bytes)



(d) De header van een UDP datagram (8 bytes)

Figuur 5.1: *De verschillende headerformaten die het IP protocol biedt. De velden die we met YAPS bekijken zijn grijs gemaakt.*

Layer3 – Internet Protocol

In de C include file `/usr/local/netinet/ip.h` vinden we de header voor een IP pakketje. Zie figuur 5.1(b). Aan de hand van het “ver” veld zien we dat het een IPv4 pakket betreft. De lengte van de inhoud van het IP pakket wordt in het 16 bits unsigned integer “packet length” opgeslagen (de maximale grootte van een IP pakket is dus 65535 bytes). Aan de hand van het protocol veld kan worden gezien of het een TCP, UDP, ICMP, of anderssoortig IP pakket betreft. Er zijn 256 verschillende layer4 protocollen mogelijk (omdat het protocol veld een 8 bits unsigned integer is).
We besluiten een histogram te maken van het verbruik van alle mogelijke IP protocollen. We kunnen dan later zien wat het totale verbruik is van bijvoorbeeld UDP of TCP aan de hand van dit histogram. We zullen hiervoor $256 \times 3 \times 16$ bytes nodig hebben (12.288).

Layer4 – Transport Control Protocol

In de C include file `/usr/local/netinet/tcp.h` vinden we de header voor een TCP datagram. Zie figuur 5.1(c). Veel poorten onder de 1024 hebben betrekking op een vooraf gedefinieerde dienst. Zo dient poort 25 voor SMTP (mail) en poort 80 voor

HTTP (web). Het kan interresant zijn om een histogram van al het verbruikte IP verkeer per poort bij te houden. Zo zien we hoeveel in- en uitgaand e-mail of webverkeer Business Internet Trends doet. Er zijn 65536 TCP poorten mogelijk (een 16 bits unsigned integer). We zullen dus voor in- en uitgaand, zowel als lokaal verkeer, voor alle poorten tellertjes bijhouden. Dit kost ons $65536 \times 3 \times 16$ bytes (3.145.728).
Overigens valt het ons op dat er in het TCP datagram 8 bits ongebruikt blijven. Deze zijn in figuur 5.1(c) gearceerd weergegeven. De letters F,S,R,P,A,U,E en C stellen vlaggetjes voor (ieder één bit).

Layer4 – Unreliable Datagram Protocol

In de C include file `/usr/local/netinet/udp.h` vinden we de header voor een UDP datagram. Zie figuur 5.1(d). We houden een grote lijst bij van alle mogelijke poorten (dit zijn er 65536). Net zoals bij TCP, hebben veel UDP poorten (vooral die onder de 1024) een speciale betekenis. Zo heeft het DNS systeem poort 53 en het NTP protocol poort 123. We maken in YAPS een histogram van de verbruikte bandbreedte per UDP poort, voor zowel inkomend als uitgaand verkeer. Ook dit kost ons, net zoals voor TCP, 3.145.728 bytes aan opslagcapaciteit.
We merken op dat, in tegenstelling tot TCP, de UDP header wèl voorziet in een datagram lengte veld. Dit is merkwaardig, aangezien de lengte van inhoud van het UDP datagram kan worden afgeleid uit het veld “packet length” uit de IP header. Dit gebeurt ook daadwerkelijk zo bij TCP.

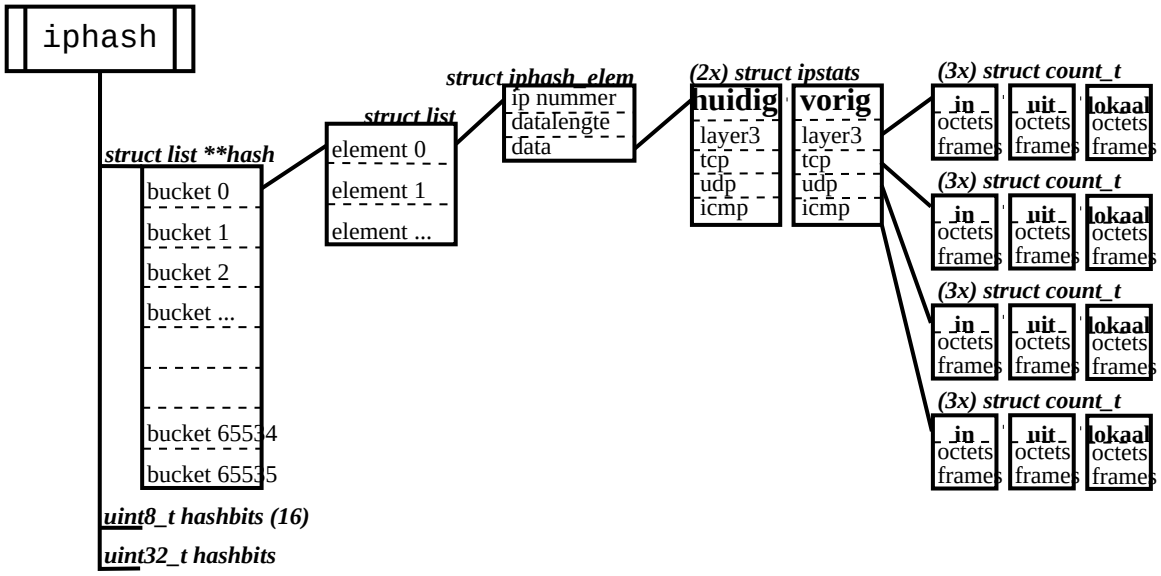
Layer4 – Internet Control Message Protocol

In de C include file `/usr/local/netinet/ip_icmp.h` vinden we de header voor een ICMP datagram. Elk ICMP bericht heeft een bepaald type. We houden een lijst van alle mogelijke types ICMP berichten bij (dit zijn er 256) en zullen bij ontvangst van een pakket van een bepaald type in deze tabel de tellers ophogen. Dit kost ons $256 \times 3 \times 16$ bytes (12.288) aan geheugen.

5.4.3 De hash functie

In het YAPS programma zullen we rekening moeten houden met veel IP nummers. Op dit moment opereert Business Internet Trends een IPv4 allocatie van RIPE-NCC ter grootte van een /19. Dit zijn 8192 IP nummers. We moeten in staat zijn om alle IP nummers in constante tijd in het geheugen terug te vinden. We besluiten dus om een hash tabel te implementeren.
IP nummers zijn geordend. We maken dankbaar gebruik van deze geordendheid met onze hashfunctie. In de IP allocatie 213.136.0.0/19 kunnen we gebruik maken van de IP nummers 213.136.0.0 t/m 213.136.31.255 waarbij opvalt dat de bovenste 19 bits constant blijven en de onderste 13 bits kunnen variëren. Als we de laatste 13 bits nu gebruiken om onze hashfunctie op te bouwen, kunnen we aan de hand van deze laatste bits voor elk IP nummer een eigen bucket maken, door minimaal 2^{13} (8192) buckets te hebben.
Als we nu meerdere netwerken zouden hebben, bijvoorbeeld naast 213.136.0.0/19 ook nog 193.109.122.0/24, valt op dat in de eerste 256 buckets twee IP nummers zouden zitten. Om deze reden voorzien we ieder bucket van een linked list met

elementen. We zouden nu met een hashtabel van bijvoorbeeld 8 bits breed (hierbij ontstaan 256 buckets), maximaal 33 IP nummers per bucket hebben. Met een breedte van 16 bits (waarbij we 65536 buckets creëren) daalt dit al tot 1 IP nummer per bucket en kunnen we in constante tijd IP nummers opzoeken.
Over het algemeen kunnen we stellen dat het vergroten van de breedte van de hashtabel leidt tot een grotere hoeveelheid buckets. We moeten er echter rekening mee houden dat er bij een *n* bits breede hashtabel en dus 2^n buckets, resulteert in evenzoveel pointers naar linkedlists die ieder weer geheugen in gebruik nemen. Als we bijvoorbeeld n=20 kiezen, hebben we $2^{20} \times 4$ (4.194.304 bytes) nodig voor de hash tabel alleen al. Als iedere linked list 64 bytes aan geheugen nodig heeft, wordt de totale behoefte al gauw $2^{20} \times 4 + 2^{20} \times 64$ (71.303.168, 71MB) ! We moeten dus de waarde voor *n* niet te groot kiezen, maar ook niet te klein. Voor Business Internet Trends is n=13 op dit moment de ideale waarde.



Figuur 5.2: *De interne memorylayout van de tellertjes in YAPS. We gebruiken een hash tabel met meerdere posities per bucket door middel van een linked list. Ieder IP heeft twee* ipstats *structs om bandbreedteberekeningen te kunnen doen.*

In figuur 5.2 zien we de hashtabel met in iedere bucket een linked list weergegeven.

5.4.4 De interne representatie

Omdat het hele principe van YAPS op het tellen van pakketten en bytes berust, lijkt het logisch om een struct te definiëren voor een teller. We doen dit door middel van struct count_t, die twee 64 bits waarden bevat: octets en frames. De tellers zijn 64 bits. We willen immers niet dat de bovengrens wordt vastgelegd op 2^{32} (4 Gigabyte), want veel servers in het netwerk kunnen potentieël meer dan 4 Gigabyte per dag aan verkeer verstoken.
Voor elk van de protocollen TCP, UDP en ICMP, alsmede het totale aantal pakketten en bytes voor een IP nummer, willen we tellers maken. Voor elk van deze tellergroepen willen we dan weer drie richtingen (ingaand, uitgaand en lokaal) bijhouden.

We omvatten alle tellers in het struct ipstats. Hierin zitten dus $4 \times 3 \times 2$ (24) tellertjes van ieder 64 bits groot (totaal 192 bytes). Voor elk IP nummer houden we twee versies van de struct ipstats bij: de huidige versie en de vorige versie. Hierover wijden we verder uit in paragraaf 5.5.2.
Omdat we voor elk IP adres twee ipstats structs bijhouden, wordt onze totale geheugenbehoefte 384 bytes per IP adres. Voor 8.192 IP adressen komen we dan uit op 3.145.728 bytes wat acceptabel is.

5.4.5 De PCAP bibliotheek

Deze library is beschikbaar voor het UNIX en Windows besturingssysteem en omvat een verzameling functies (de API) om rauwe ethernet data te kunnen ontvangen met een PC. De API bestaat uit een grote hoeveelheid functies, waarvan wij er maar enkele zullen gebruiken. Omdat YAPS in alle ethernet frames geïnteresseerd is, zullen we geen gebruik maken van de selectie algoritmen van PCAP.
Uiteraard moet aan de library worden aangegeven welke netwerkkaart er gebruikt moet worden. We gebruiken hiervoor de functie `pcap_open_live()`. Als we niet expliciet opgeven welke netwerkkaart we willen gebruiken, kan PCAP er ook eentje selecteren. Hij zal dan in de regel de eerste kaart kiezen die geconfigureerd is. We laten hem een netwerkkaart selecteren met `pcap_lookupdev()`.
We maken dan zelf een functie `pcap_callback()`, die alle frames zal krijgen toegespeeld door PCAP. We noemen dit een callbackfunctie, omdat we vanuit een oneindige lus (`pcap_loop()`) PCAP frames laten vragen aan de kernel en voor elk frame één aanroep van de callback functie laten uitvoeren, met als argument het frame wat PCAP net van de kernel gekregen heeft.
Het fijne van PCAP is dat het een abstractielaag vormt tussen hoe de kernel daadwerkelijk met ethernet frames omgaat, en de applicatie die we schrijven daarmee wenst om te gaan. We zorgen er dankzij PCAP voor, dat YAPS zal kunnen draaien op elke computer waarop de PCAP bibliotheek is geïnstalleerd.

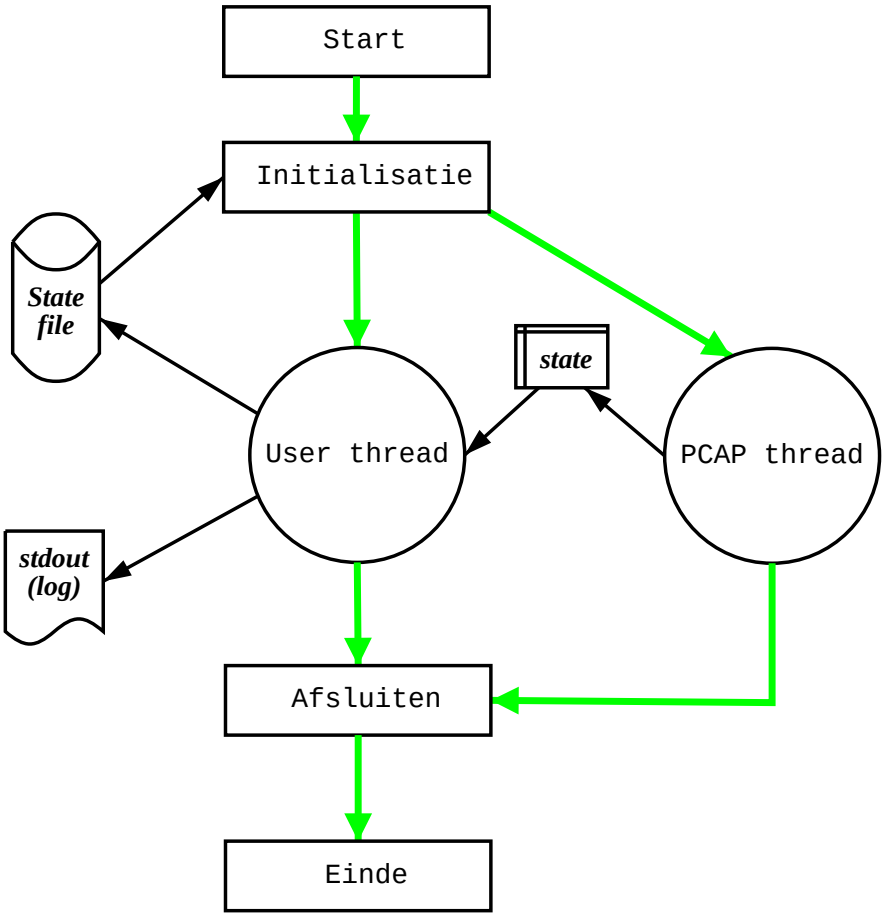
5.5 Het executiemodel van YAPS

Bij het opstarten van het YAPS programma, zal (als de user dit wenst) gezocht worden naar een state file. Als deze bestaat wordt hij ingelezen zodat de historie behouden blijft. Daarna worden er twee threads gestart:

- De User Thread – in deze thread worden periodiek berekeningen gedaan van de verbruikte bandbreedte per IP en wordt de interne state weggeschreven naar de harde schijf. Mocht YAPS of de server waarop hij draait crashen, zal bij een restart de state file weer worden ingelezen.
- De PCAP Thread – in deze thread worden alle frames die door de kernel op een Ethernet interface worden gelezen, in hun geheel gelezen (inclusief layer2 informatie) en verwerkt. Echtereenvolgens wordt er naar layer3 en layer4 informatie gekeken en aan de hand daarvan de juiste tellertjes bijgewerkt.

Wanneer de User Thread het besluit neemt op de executie te staken, zal dit worden doorgegeven aan de PCAP thread. Deze zal zichzelf opruimen waarna de User

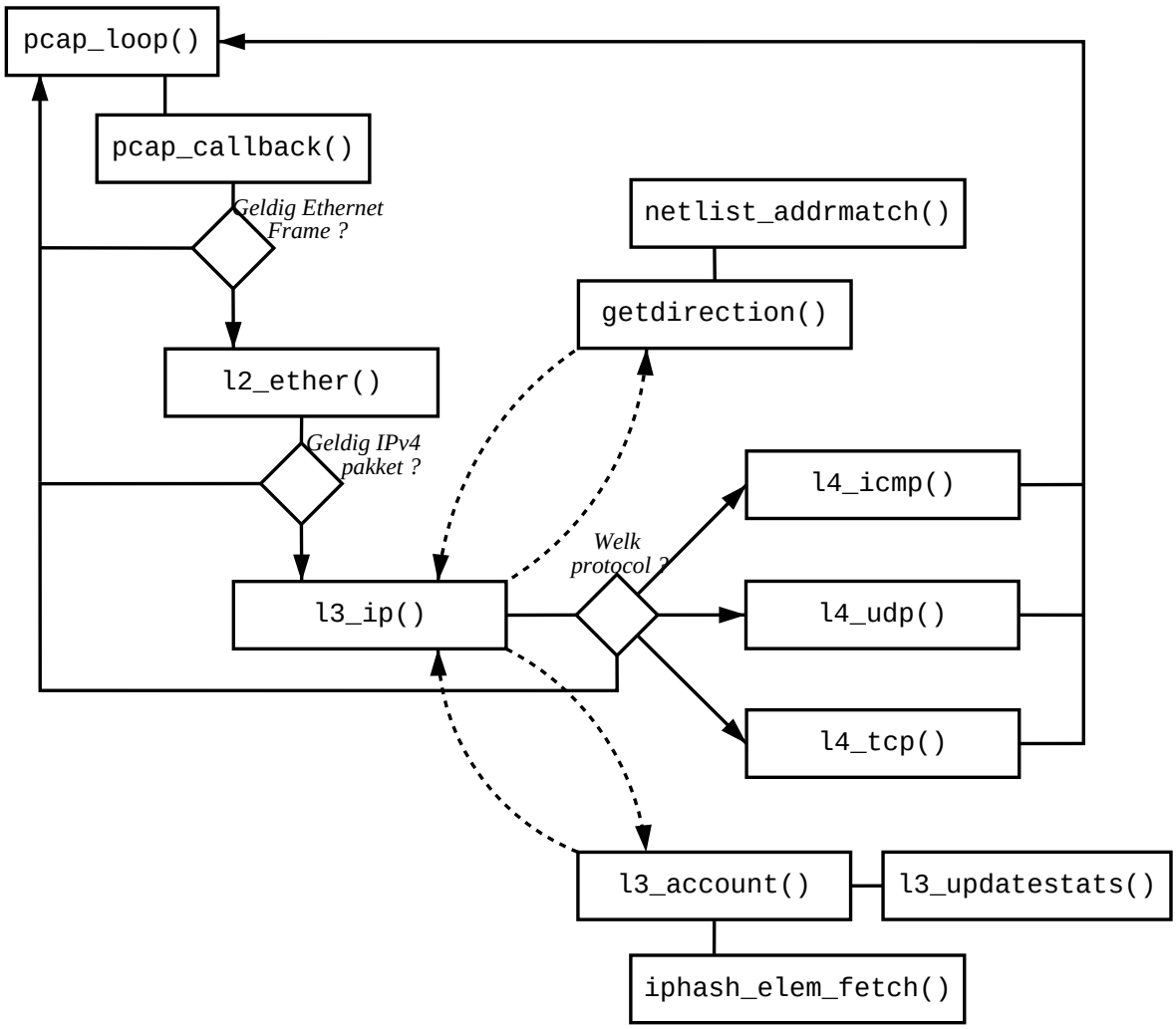
Thread de gebruikte memory structuren kan opruimen en tenslotte zichzelf kan stoppen.
Een globaal overzicht van de het executiemodel wordt gegeven in de flowchart van figuur 5.3.



Figuur 5.3: *YAPS maakt gebruik van één externe statefile om periodiek een backup van de interne memorystate te maken. Het bestaat uit twee threads, één voor het verwerken van PCAP data, en één voor het periodiek uitvoeren van checks (die naar stdout worden gerapporteerd) en het maken van state backups naar de harde schijf.*

5.5.1 De PCAP Thread

De hoofdlus van deze thread zal met behulp van de library-call `pcap_loop()` Ethernet frames aan de kernel vragen. We initialiseren de functie zó, dat voor elk ontvangen frame een call-back functie `pcap_callback()` wordt uitgevoerd met een door PCAP gezette variabele en een pointer naar het gehele Ethernet frame als argumenten.
Ofschoon de kernel voor PCAP Ethernet frames buffert, is het zaak om zo snel mogelijk het frame te verwerken, omdat de buffer van de kernel eindig is. Om 50.000 frames per seconde te kunnen meten, moeten we ernaar streven dat de gemiddelde verwerkingstijd van één frame korter is dan 0,02 μs (200 ns per frame dus).



Figuur 5.4: *De PCAP thread nader bekeken*

De `pcap_callback()` call-back functie zal na een consistentiecheck op het frame `l2_ether()` aanroepen met als argumenten een pointer naar de inhoud van het frame en de lengte daarvan. Er wordt nu gekeken naar het *Ethernet type* van het frame. Als dit een IPv4 frame is, wordt er verder gegaan met `l3_ip()`. De Ethernet header wordt van het frame afgestript en het overgebleven IP pakket wordt doorgegeven. Dit betekent dat de lengte met 14 bytes afneemt (6 bytes source-MAC, 6 bytes destination-MAC en 2 bytes ethertype).
Het eerste wat `l3_ip()` doet is vaststellen of het versienummer van het IP pakket wel gelijk is aan 4 (kijkend naar ip_v) en of de header lengte minimaal 40 bytes lang is (kijken naar ip_hl). Zo nee, wordt dit frame als corrupt beschouwd en derhalve niet verder geanalyseerd.
Nu we vastgesteld hebben dat we een geldig IPv4 pakket vasthebben, wordt er berekend of dit een in- of uitgaand pakket betreft. We doen dit door te kijken naar het source en destination IP nummer. De interne functie `getdirection()` kijkt aan de hand van de zelfgemaakte library-functie `netlist_addrmatch()` of een gegeven IP adres in een lijst van “lokale netwerken” zit. We hebben hier drie mogelijkheden:

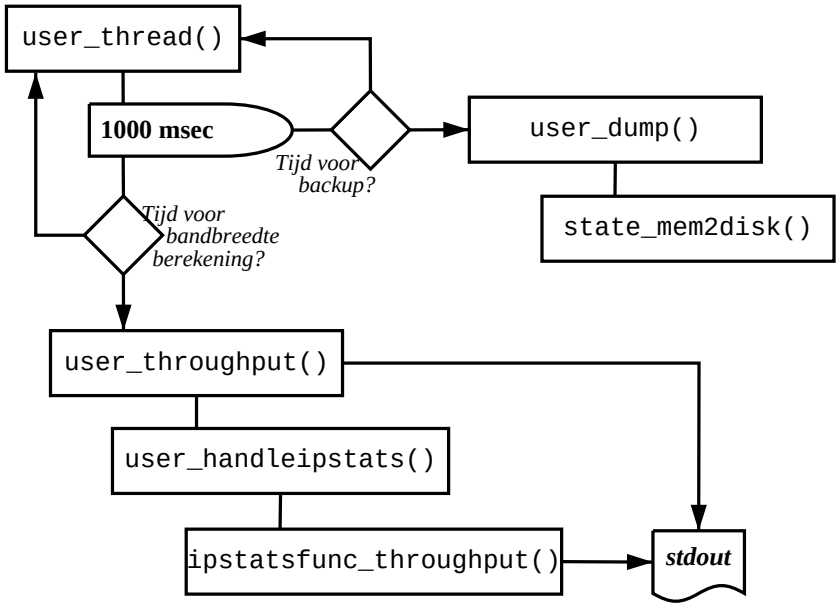
- Het Source IP ligt binnen het netwerk en het Destination IP ligt buiten het netwerk: Dit is een uitgaand pakket.
- Het Source IP ligt buiten het netwerk en het Destination IP ligt binnen het netwerk: Dit is een inkomend pakket.
- Het Source IP en het Destination IP liggen binnen ons netwerk: Dit is een lokaal pakket.

Nu we de richting van het IP nummer weten, kunnen we de records voor de betreffende IP nummers opzoeken en hun tellers ophogen. We doen dit in de functie `l3_account()`.
De functie `l3_account()` zal met behulp van de zelfgeschreven library call `iphash_elem_fetch()` opzoeken of we de betreffende IP nummers wensen bij te houden. Als de entry in de hashtabel bestaat is dit het geval en zal een pointer naar het record worden onthouden.
Zondermeer worden eerst de totalen voor het in- en uitgaande verkeer bijgewerkt en wordt aan de hand van het layer3 protocol (TCP, UDP, ICMP, ...) de histogram bijgewerkt die de totale verkeersstromen per protocol voorstelt. Daarna worden de statistieken van de betreffende IP nummers bijgewerkt door de functie `l3_updatestats()` aan te roepen, mits de IP nummers een entry in de hashtabel hebben.
Als we nu kijken naar het protocol-nummer (in ip_p) van de IP header (bijvoorbeeld 6 voor TCP, 17 voor UDP of 1 voor ICMP), kunnen we een verdere analyse doen op layer4. We roepen voor de drie belangrijkste IP protocollen dus respectievelijk `l4_tcp()`, `l4_udp()` of `l4_icmp()` aan, waarbij we de volgende argumenten meegeven: De gevonden IP header, de berekende richting van het pakket, de lengte van het layer4 PDU en natuurlijk een pointer naar de inhoud van de layer4 PDU zelf.
Omdat we graag een overzicht willen hebben van de distributie van TCP en UDP verkeer per servicetype (oftewel per poort), en de distribute van het ICMP verkeer in termen van types ICMP berichten, zullen we voor UDP en TCP een tabel bijwerken van alle 65.536 mogelijke poorten. Voor ICMP zullen we een tabel bijwerken met de 256 mogelijke ICMP types. We houden voor elke tabel drie varianten bij: de inkomende, de uitgaande en de lokale variant.
Als we alle tellertjes hebben bijgewerkt, geven we het oorspronkelijke layer2 frame weer vrij en wachten we op de volgende aanroep van `pcap_callback()`.

5.5.2 De User Thread

De User thread heeft twee taken. Zijn primaire taak is het bijwerken van lopende statistieken over de verbruikte bandbreedte per IP. We doen dit door periodiek (bijvoorbeeld iedere 60 seconden) een kopie te maken van alle tellertjes en het verschil te nemen tussen de huidige toestand en de vorige. We zien dan de hoeveelheid bytes en pakketjes van alle tellers in de afgelopen periode. Hiermee kunnen we per IP nummer het bandbreedte verbruik in bits per seconde en pakketten per seconde uitrekenen.
Zijn tweede taak is het periodiek wegschrijven van de tellertjes, zodat externe programma's hier bij kunnen. Uiteraard zorgt het backuppen van de interne data naar hardeschijf er ook voor, dat als onze server of YAPS zelf mocht crashen, we geen historie kwijt raken. We zullen er naar streven om elke vijf minuten de tellertjes

te backuppen naar hardeschijf en schrijven hiervoor een aparte module die bestaat uit twee hoofdfuncties: `state_mem2disk()` en `state_disk2mem()`.



Figuur 5.5: *De User Thread nader bekeken*

De hoofdlus van `user_thread()` bestaat uit een klok die elke seconde kijkt of een stopwatch is verstreken. De stopwatch stellen we altijd in op *n* seconden in de toekomst, dus wanneer we de volgende backup of bandbreedte berekening willen uitvoeren. Als deze *n* seconden zijn verstreken wordt de functie `user_dump()` aangeroepen en daarna de stopwatch voor het backuppen weer met *n* seconden vooruitgeschoven.
We bekijken daarna (ook elke seconde dus) of het tijd is om een bandbreedte berekening te doen. Zo nee, gaan we weer een seconde slapen, zo ja springen we naar de `user_throughput()` functie. Deze zal voor ieder IP in de hash tabel dezelfde functie aanroepen: `ipstatsfunc_throughput()`. Deze functie kijkt in het geheugen naar de kopie van de vorige iteratie van de bandbreedteberekening. Als deze nog niet is gezet (dus bij de eerste keer dat deze functie voor dit IP wordt aangeroepen), zal de structuur worden geïnitialiseerd met de huidige tellers. De tweede en volgende keer kunnen we de “vorige” data (dus van 60 seconden geleden) met de huidige data vergelijken, berekeningen doen, en tenslotte de huidige data over de oude heen kopiëren voor de volgende iteratie.
Voor elk IP adres dat meer dan 15 Mbps (Megabit per seconde) of 1 Kpps (pakketjes per seconde) aan verkeer genereert, wordt hier melding gemaakt. We kunnen hiermee later gemakkelijk DDoS[6] aanvallen detecteren.

[6]Distributed Denial of Service – een manier om enorme hoeveelheden verkeer op een IP nummer af te vuren.

5.6 Implementatie

Het programma bestaat, naast de twee hoofdmoten (de pcap- en userthread zoals die reeds beschreven zijn), ook nog uit de volgende subsystemen.

5.6.1 Configuratie parser

Er zijn behoorlijk wat parameters die van invloed zijn op de werking van een dergelijk traffic accounting systeem. Denk hierbij aan de te gebruiken netwerkkaart, welke IP adressen er als intern moeten worden beschouwd en voor welke IP adressen er statistieken moeten worden bijgehouden.
Omdat het parseren van een configuratie bestand bepaald niet gemakkelijk te implementeren is in C, is gekozen om het programma flex te gebruiken. Dit programma is in staat om een reguliere taal te herkennen en aan de hand van de gevonden tokens C code uit te voeren.
Er worden eerst een aantal definities gemaakt, allen aan de hand van reguliere expressies. Enkele voorbeelden van definities zijn:

```
SEP             [\r\n \t\v\f]
SEPS            {SEP}+
ALPHA           [A-Za-z]
DIGIT           [0-9]
DIGITS          {DIGIT}+
```

In bovenstaand voorbeeld kan men zien dat het SEP token wordt gedefinieerd als de set van de volgende tekens: carriage return, newline, spatie, tab, vertikale tab en tenslotte een form feed. De token `SEPS` wordt dan gedefinieerd als één of meer keer het `SEP` token (aan de hand van de `+` operator). Als men nul of meer keer het `SEP` token zou willen, schrijft men `SEP*`.
Flex kan gaandeweg een state bijhouden. Aan het begin heet deze state “INITIAL” maar nadat een token gevonden is, kan men besluiten naar een andere state te gaan (en de huidige op te slaan), om met een andere set herkenningsregels te scannen. Op deze manier wordt de scanner context gevoelig. Er kan van state worden veranderd door de commando’s `yy_push_state()` en `yy_pop_state()`.
Zoals gezegd kan flex patronen ontdekken in een tekstbestand. Als hij een patroon heeft gescanned kan hij hiermee een stukje C code uitvoeren. In ons geval lezen we dus als het ware de gehele configuratiefile, herkennen we de configuratiecommando’s daarin, en zullen we YAPS aan de hand van deze configuratiecommando’s instellen als gewenst.
Een voorbeeld. YAPS kan, als de gebruiker dit wenst, na het opstarten zoeken naar een state-file en hieruit de laatst opgeslagen tellers lezen. Als men dit gedrag wenst, kan men “keepstate;” definiëren in het configuratie bestand. Wil men dit gedrag niet, kan men “no-keepstate;” definiëren. De scanner voor dit gedeelte ziet er dan als volgt uit:

```
<INITIAL>{
    {SEP}*"keepstate"{SEP}*";" {
        conf->keepstate = 1;
    }
    {SEP}*"no-keepstate"{SEP}*";" {
```

```
        conf->keepstate = 0;
    }
}
```

Voor een lijst van alle configuratie mogelijkheden wordt verwezen naar appendix E.

5.6.2 IP Hashtable module

Het doel van deze module is het verzorgen van een zo snel mogelijke manier om willekeurige IP nummers op te zoeken in een hash tabel. We stellen ons ten doel om in constante tijd het record behorend bij een IP op te kunnen zoeken.
We geven hier de header file van de module, zodat inzichtelijk wordt welke datatypes en functies we definiëren.

```
struct iphash_elem {
        uint32_t ip;
        uint32_t datalen;
        void *data;
};

struct iphash {
        uint8_t hashbits;
        struct list **hash;
        uint32_t elements;
};

struct iphash *iphash_create (uint8_t bits);
void iphash_destroy (struct iphash **iphash);
void *iphash_elem_create (struct iphash *iphash, const uint32_t ip);
void *iphash_elem_fetch (struct iphash *iphash, const uint32_t ip);
void iphash_elem_destroy (struct iphash *iphash, const uint32_t ip);
```

De functies zijn als volgt:

- `iphash_create();` Deze functie initialiseert een iphash structuur en retourneert er een pointer naar.
- `iphash_destroy();` Deze functie ruimt aan de hand van een pointer naar het adres van een bestaande iphash het gebruikte geheugen op.
- `iphash_elem_create();` Hiermee kunnen we een nieuw IP adres introduceren in de hash tabel. De juiste bucket wordt opgezocht en er wordt gekeken of het IP daar reeds in bestaat. Zo nee, wordt hij achteraan de linked list van de bucket bijgeschreven en wordt een pointer naar de nieuw gemaakte `struct iphash_elem` geretourneerd.
- `iphash_elem_fetch();` Aan de hand van een IP adres wordt het record uit de hash tabel opgevraagd. Als het IP adres niet in de bucket voorkwam, wordt de NULL pointer geretourneerd, anders een pointer naar de gevonden `struct iphash_elem`.

- `iphash_elem_destroy()`; Hiermee wordt het record van een IP adres uit de hash tabel verwijderd. Als het IP adres niet bestond, wordt er niets gedaan. Als het wel bestaat in de hash, wordt de bijbehorende `struct iphash_elem` opgeruimd. Aangenomen wordt, dat het data veld van dit element reeds is vrijgegeven.

Zoals te zien is, zal `iphash_elem_fetch()`; een void pointer retourneren. Dit is feitelijk een pointer naar een eigen datastructuur die `struct iphash_elem` heet. Deze bevat een veldje `data` die op zijn beurt weer verwijst naar een tellerstructuur (`struct ipstats`) voor het betreffende IP adres.

5.6.3 Networklist module

Het doel van deze module is het verzorgen van een snelle manier om te zien of een IP nummer in een bepaalde lijst met netwerken zit. We doen dit door in een linked list van IP netwerken in CIDR[7] notatie te zoeken.
We geven hier de header file van de module, zodat inzichtelijk wordt welke datatypes en functies we definiëren.

```
#define NETLIST_NAMELEN 64
struct prefix {
        uint8_t family;
        uint8_t prefixlen;
        uint32_t prefix;
        uint32_t mask;
};

struct netlistnode {
        uint32_t action;
        struct prefix prefix;
};

struct netlist {
        uint8_t family;
        struct list *list;
        char name[NETLIST_NAMELEN];
};

struct netlist *netlist_create ();
int netlist_destroy (struct netlist **nl);
int netlist_setname (struct netlist *nl, const char *name);
int netlist_setfamily (struct netlist *nl, const int af);
int netlist_prefix_add (struct netlist *nl,  const char *prefix,
        const uint8_t prefixlen, const uint32_t action);
struct netlistnode *netlist_addrmatch (const struct netlist *nl,
        const uint32_t ip);
```

De functies zijn als volgt:

[7] Classless Inter Domain Routing – Een methode om subnets van variabele lengte te creëren en te noteren als prefixlengte zoals 213.136.0.0/19

- `netlist_create()`; Deze functie initialiseert een netlist en retourneert er een pointer naar.
- `netlist_destroy()`; Deze functie ruimt aan de hand van een pointer naar het adres van een bestaande netlist het gebruikte geheugen op.
- `netlist_setname()`; Deze functie kent aan een netlist een naam toe. Deze naam mag maximaal 64 tekens lang zijn.
- `netlist_prefix_add()`; Met deze functie kan men een bepaalde prefix met een bepaalde prefixlengte (bijvoorbeeld 213.136.0.0 met prefixlengte 19) aan de lijst toevoegen. Elk toegevoegd netwerk wordt in de linked list als `struct netlistnode` opgeslagen.
- `netlist_addrmatch()`; Dit is de functie waarmee één IP adres kan worden vergeleken met de lijst. Het eerste netwerk in de lijst die het IP adres overkoepelt, wordt geretourneerd. Als het IP in geen van de netwerken voorkwam, wordt de NULL pointer geretourneerd.

We moeten wel opmerken dat het in principe een slecht idee is om IPv4 adressen op te slaan als 32 bits integers. Er bestaat in C een andere tructuur (`struct addrinfo`) waar dit protocol onafhankelijk kan worden gedaan. Uiteraard is het gebruiken van integers het snelst, en omdat voor elk frame twee maal zo'n `netlist_addrmatch()` functie moet worden aangeroepen (één keer voor het source IP en één keer voor het destination IP), vinden we de beslissing gerechtvaardigd.

5.7 Installatie

Om YAPS te installeren, moet men eerst de source code bemachtigen. De source wordt in de regel verspreid als zogenaamde tar-file. Zo'n bestand is net zoals een zip-file voor Windows. Het installeren kan als volgt gebeuren:

```
$ tar xzf yaps.tar.gz
$ cd yaps
$ ./configure --prefix=/home/yaps
$ make
```

Het configure programma zal aan de hand van het systeem achterhalen waar de benodigde C headers zijn, en eventueel enkele ontbrekende definities alsnog maken. Zo raakt de programmeur eraan gewend om altijd voor dezelfde omgeving code te schrijven. We geven één argument mee met het configure script. Deze zal er voor zorgen dat de installatie van het programma zal geschieden in de directory `/home/yaps/`. Het make programma zal de compilatie verzorgen, aan de hand van de meegeleverde Makefile.
Er worden nu drie bestanden opgeleverd: `yaps`, `report` en `yaps.conf`. In het configuratiebestand kunnen alle run-time variabelen worden ingesteld. Zie hiervoor verder appendix E.

5.8 Productie

Binnen Business Internet Trends is een server ingericht voor het YAPS programma. Het is een Pentium III op 866 Mhz, met 512 MB geheugen en een 18 GB SCSI harde schijf. De specificaties van dit systeem zijn dus niet exorbitant hoog. De machine is uitgerust met drie netwerkkaarten. Op de eerste kaart (eth0) wordt het 'gewone' IP adres van de machine gezet (lummi.bit.nl (213.136.12.131), in het VLAN v-misc). Op de tweede kaart (eth2) wordt het NFS netwerk ingekoppeld. We kunnen dan de resultaten van de traffic metingen dagelijks kopieren naar de fileserver. De derde kaart (eth3) is enigszins bijzonder. Het is namelijk geen gewone netwerkkaart maar een glasvezel gigabit netwerkkaart van Intel.
De switch die het verkeer van en naar Amsterdam transporteert via de Extreme EAPS ring en het CWDM netwerk, zal een kopie maken van al het verkeer en deze aflveren op een glasvezelpoort. Aan deze poort sluiten we eth3 van Lummi aan, zodat al het verkeer wat het netwerk in Ede in- en uitgaat door Lummi kan worden geanalyseerd.
YAPS starten we als volgt:

```
$ /home/yaps/bin/yaps -f -c /home/yaps/etc/yaps.conf --state
```

YAPS zal het volgende melden:

```
Using configuration directives from etc/yaps.conf
Adding 213.136.0.0/19 to netlist 'Local networks'
Adding 193.109.122.0/24 to netlist 'Local networks'
The list 'Local networks' now contains 2 networks
Adding 213.136.0.0/19 to netlist 'Accounted networks'
Adding 193.109.122.0/24 to netlist 'Accounted networks'
The list 'Accounted networks' now contains 2 networks
Adding ports 1-1024 to portlist 'Accounted TCP ports'
Adding ports 6660-6670 to portlist 'Accounted TCP ports'
The list 'Accounted TCP ports' now contains 1038 ports
Adding ports 1-1024 to portlist 'Accounted UDP ports'
The list 'Accounted UDP ports' now contains 1024 ports
State directory is /home/yaps/var
Setting up IPHASH structure
Traversing acctnets list
STATE_DISK2MEM: Starting state retrieval from '/home/yaps/var/yaps.state'
STATE_DISK2MEM: * Last write: Thu Oct 10 17:40:34 2002
STATE_DISK2MEM: * Last clear: Thu Oct 10 00:05:43 2002
STATE_DISK2MEM: * IPHASH stats (8448 IPs retrieved)
STATE_DISK2MEM: Read 1871840 bytes in 142 ms
*** Using device 'gx0'
Starting pcap thread
Performing user interaction in this thread
USER: Dumping every 300 seconds
USER: Calculating throughput every 60 seconds
USER_THROUGHPUT: Initializing throughput tracker
USER_HANDLEIPSTATS: Initializing per-IP bandwidth tracker
PCAP: RECV:  10000, DROP:      0
```

```
PCAP: RECV:  20000, DROP:      0
```

Allereerst wordt de oude statefile opgehaald. Deze bevatte 8.448 IP nummers (de som van één /19 en één /24). Hiermee worden alle tellertjes geïnitialiseerd.
De PCAP bibliotheek begint dan aan de kernel frames te vragen en zal voor ieder frame een analyse uitvoeren. Om de 10.000 frames zal hij melding maken van zijn voortgang. Er worden op de piek-drukte van Business Internet Trends zo'n 30.000 frames per seconde gelezen.
Na een minuut (de tijd die gespecificeerd is in de configfile met het configuratie commando `throughputinterval 60;`), zal YAPS van de hele ISP de doorvoer berekenen en daarna van alle IP adressen die een grens van 15 Mbps of 1 Kpps bereiken een melding maken:

```
USER_THROUGHPUT: bits/sec(in,out,local)=(56976621,52143716,212433)
USER_THROUGHPUT: frames/sec(in,out,local)=(8656,7609,86)
USER_HANDLEIPSTATS: Calculating per-IP bandwidth
USER_PRINTIP: 213.136.0.147 bits/sec(in,out,local)=(402736,6819947,0)
USER_PRINTIP: 213.136.0.147 frames/sec(in,out,local)=(519,673,0)
USER_PRINTIP: 213.136.9.182 bits/sec(in,out,local)=(218055,8257442,0)
USER_PRINTIP: 213.136.9.182 frames/sec(in,out,local)=(390,710,0)
USER_PRINTIP: 213.136.9.184 bits/sec(in,out,local)=(462398,7715667,18)
USER_PRINTIP: 213.136.9.184 frames/sec(in,out,local)=(606,786,0)
USER_PRINTIP: 213.136.12.36 bits/sec(in,out,local)=(47703783,619313,0)
USER_PRINTIP: 213.136.12.36 frames/sec(in,out,local)=(3609,1288,0)
USER: Scheduling next throughput calculation in 60 seconds
```

Een poosje later, zo'n 5 minuten na het begin (gedefinieerd in het configuratie commando `dumpinterval 300;`), zal YAPS een backup maken van alle tellers naar de harde schijf. We zien dan:

```
STATE_MEM2DISK: Starting state dump in '/home/yaps/var/yaps.state'
STATE_MEM2DISK: * Last write: Thu Oct 10 17:30:34 2002
STATE_MEM2DISK: * Last clear: Thu Oct 10 00:05:43 2002
STATE_MEM2DISK: * Ethernet counters (16)
STATE_MEM2DISK: * IP counters (48)
STATE_MEM2DISK: * IP protocol histogram (12288)
STATE_MEM2DISK: * ICMP protocol histogram (12288)
STATE_MEM2DISK: * Portlist 'Accounted TCP ports', 1038 items
STATE_MEM2DISK: * Portlist 'Accounted UDP ports', 1024 items
STATE_MEM2DISK: * IPHASH stats (8448 elements in 65536 buckets)
STATE_MEM2DISK: Wrote 1871840 bytes in 81 ms
```

We kunnen het wegschrijven van de state ook nog op een andere manier forceren. Door het programma een zogenaamd signaal te sturen van een bepaald type, zal hij de state wegschrijven. We implementeren dit door het HUP signaal af te vangen binnen YAPS.
Naast het wegschrijven van de tellertjes, kan het ook wenselijk zijn om de tellertjes te resetten. We implementeren dit door het USR1 signaal af te vangen binnen YAPS. Hierdoor zal YAPS alle tellers op nul zetten.

5.9 Rapportage

Omdat YAPS elke 5 minuten een kopie van de interne state (met daarin alle relevante tellertjes) kunnen we zonder dat de productie in gevaar komt, de statefile periodiek bekijken.
Hiervoor is een klein programma'tje geschreven die simpel weg alle tellertjes uitleest en print. Dit programma'tje heet `report` en kan op twee manieren worden aangeroepen.
De eerste methode produceert een voor de mens leesbaar resultaat. We roepen het als volgt aan:

```
$ /home/yaps/bin/report /home/yaps/var/yaps.state
```

De (enorm lange) uitvoer is dan als volgt. We knippen de meeste IP nummers weg, om het overzicht enigszins te behouden.

```
------------------- Generate global statistics dump --------------------
* Stats valid Thu Oct 10 00:05:43 2002 thru Thu Oct 10 17:55:34 2002
* Unix timestamps: 1034201143 1034265334
* Timespan: 64191 seconds

* Traffic analysis Layer3
direction  |           octets    |           frames
-----------|---------------------|---------------------
  Incoming |421040752586 ( 421G) |   503479838 ( 503M)
  Outgoing |281072806611 ( 281G) |   406563493 ( 406M)
  Local    |  2118128436 (   2G) |     5616753 (   5M)
  Total    |702113559197 ( 702G) |   910043331 ( 910M)


The following tables depict the total amount of octets and frames for each
service, if the sum of the incoming, outgoing and local frames exceeds 50.

* Traffic analysis per IP protocol
protocol   |                 Octets    |  |                Packets
-----------|---IN-|--OUT-|--LOC-|--TOT-|  |---IN-|--OUT-|--LOC-|--TOT-
icmp       | 590M | 636M |   7M |   1G |  |   1M |   1M | 129K |   3M
igmp       |    0 |  14K |    0 |  14K |  |    0 |  514 |    0 |  514
ipencap    |  13M | 641K |   8M |  14M |  | 148K |   6K |  52K | 155K
tcp        | 416G | 277G |   2G | 694G |  | 477M | 380M |   4M | 857M
udp        |   3G |   2G |  41M |   5G |  |  22M |  24M | 467K |  47M
ipv6       |  696 |   4K |    0 |   5K |  |    6 |   58 |    0 |   64
gre        |  22M |  11M |   8M |  33M |  | 140K | 108K |  21K | 248K
esp        | 541M |  76M |    0 | 617M |  |   1M | 608K |    0 |   1M
ospf       |    0 |  15M | 324K |  15M |  |    0 | 146K |   2K | 146K

* Traffic histogram for the TCP protocol
protocol   |                 Octets    |  |                Packets
-----------|---IN-|--OUT-|--LOC-|--TOT-|  |---IN-|--OUT-|--LOC-|--TOT-
ftp-data   |   1G | 444M |  33M |   1G |  |   1M | 883K |  56K |   2M
```

```
ftp         |  14M |  16M |  10M |  30M |   | 271K | 231K | 151K | 502K
ssh         | 320M | 685M |   6M |   1G |   |   1M |   2M |  53K |   3M
telnet      |   2K |  960 | 316K |   3K |   |   56 |   24 |   6K |   80
smtp        |   3G | 128M | 439M |   3G |   |   3M |   2M | 649K |   5M
time        |  11K |   9K |   6K |  20K |   |  257 |  219 |  155 |  476
whois       |   4M |  18M |   80 |  23M |   |  94K |  90K |    2 | 184K
domain      | 106M | 656M |   4M | 762M |   |   2M |   5M |  85K |   7M
mtp         |   9K |   8K |   80 |  17K |   |  209 |  201 |    2 |  410
finger      |  17K |  13K | 677K |  30K |   |  358 |  338 |   4K |  696
www         |  13G | 222G | 858M | 236G |   | 151M | 206M |   1M | 358M
iso-tsap    |   2M |  52K |   80 |   2M |   |   2K |   1K |    2 |   3K
pop3        | 105M |   1G | 135M |   1G |   |   2M |   2M | 377K |   5M
auth        | 562K | 383K |  90K | 945K |   |  10K |   8K |   1K |  18K
nntp        | 363G |   4G |  49K | 367G |   | 247M |  83M |  495 | 330M
netbios-ssn|  30M |  23M |   6M |  53M |   | 201K | 169K |  24K | 370K
imap        | 410K |   4M |  10M |   5M |   |   5K |   6K |  67K |  12K
bgp         | 200K | 478K |  77K | 679K |   |   4K |   4K |   1K |   8K
ipx         |   2K |   1K |   80 |   4K |   |   45 |   45 |    2 |   90
ldap        |   7K |   4K |   1K |  11K |   |  148 |  106 |   38 |  254
https       |  35M | 122M | 622K | 157M |   | 183K | 189K |   1K | 372K
shell       |  468 |  360 |   6M |  828 |   |    9 |    9 |  29K |   18
printer     |   1M |  400 |   80 |   1M |   |   3K |   10 |    2 |   3K
pop3s       | 799K |   2M |   80 |   2M |   |   9K |   7K |    2 |  16K
ircd        |  67K | 173K |   80 | 241K |   |  950 |   1K |    2 |   2K

* Traffic histogram for the UDP protocol
protocol    |                 Octets    |  |                Packets
------------|---IN-|--OUT-|--LOC-|--TOT-|  |---IN-|--OUT-|--LOC-|--TOT-
domain      | 564M |   1G |  34M |   1G |  |   7M |   7M | 408K |  15M
ntp         | 823K | 866K |   1M |   1M |  |  10K |  11K |  17K |  22K
netbios-ns  |  17M |  10M | 613K |  28M |  | 221K | 124K |   7K | 345K
netbios-dgm| 144K |   1K | 197K | 145K |  |  704 |    4 |  814 |  708
snmp        |   5K |  600 |   6M |   5K |  |   49 |    6 |  41K |   55
snmp-trap   |   46 |    0 | 162K |   46 |  |    1 |    0 |  656 |    1
isakmp      |   1M |  59K |    0 |   1M |  |   4K |  318 |    0 |   4K
syslog      |   92 |    0 | 247K |   92 |  |    2 |    0 |   2K |    2

* Traffic histogram for the ICMP protocol
protocol    |                 Octets    |  |                Packets
------------|---IN-|--OUT-|--LOC-|--TOT-|  |---IN-|--OUT-|--LOC-|--TOT-
echorep     | 487M |  38M | 870K | 526M |  | 615K | 313K |  10K | 929K
destunr     |  27M |  21M |   7M |  48M |  | 389K | 256K | 123K | 645K
sourceq     | 151K |    0 |    0 | 151K |  |   2K |    0 |    0 |   2K
redir       |   6M |  972 |    0 |   6M |  | 118K |    9 |    0 | 118K
echoreq     |  41M | 573M |   1M | 615M |  | 334K | 753K |  16K |   1M
time-exd    |  25M |   2M |   6M |  28M |  | 435K |  47K | 107K | 482K
-------------------- List Bandwidth consumption per IP --------------------

* Traffic breakdown per local IP
```

```
IP number       |                  Octets    |  |                Packets
----------------|---IN-|--OUT-|--LOC-|--TOT-|  |---IN-|--OUT-|--LOC-|--TOT-
213.136.0.0     | 112K |    0 |    0 | 112K |  |   2K |    0 |    0 |   2K
213.136.0.1     |   5K |   3K |    0 |   8K |  |  144 |   68 |    0 |  212
213.136.0.2     |  35K | 665K |    0 | 701K |  |  596 |   7K |    0 |   7K
213.136.0.6     |   2K |    0 |  26K |   2K |  |   42 |    0 |  408 |   42
213.136.0.33    |   2G | 346M |   1M |   3G |  |   2M |   1M |  20K |   4M
```

Als we dagelijks deze rapporten creëren, zien we van dag tot dag het verbruik binnen de ISP. We kunnen zien hoeveel TCP, UDP en ICMP verkeer er is en welke IP nummers binnen het netwerk de grootverbruikers zijn.
Om deze getallen makkelijker verwerkbaar te maken, kiezen we er voor om ook een CSV uitvoer te kunnen laten maken en dat brengt ons tot de tweede modus van het `report` programma'tje. Deze zal voor elk IP adress in de statefile één regel printen. De uitvoer ziet er wat soberder uit, maar kan wel linea recta een SQL database ingeduwd worden. We roepen het programma dan als volgt aan:
`$ /home/yaps/bin/report -csv /home/yaps/var/yaps.state`

In de praktijk zullen we 's nachts om vijf minuten na middernacht het volgende stappenplan uitvoeren:

1. Zend YAPS het HUP signaal (`killall -HUP yaps`). Hij zal state wegschrijven. Deze bevat de totalen van de afgelopen 24 uur.
2. Kopieer de statefile naar een andere plek.
3. Zend YAPS het USR1 signaal (`killall -USR1 yaps`). Hij zal alle tellertjes op nul zetten.
4. Zend YAPS het HUP signaal (`killall -HUP yaps`). Hij zal state wegschrijven, waarbij alle tellers nu op nul staan.
5. Voer de bewaarde state file aan het `report` programma. Mail de uitvoer hiervan naar de beheerders van Business Internet Trends.
6. Voer de bewaarde state file nogmaals aan het `report` programma, maar laat het een CSV uitvoer maken. Voer deze uitvoer aan een SQL server voorzien van de dag van meting.

Voor de veiligheid worden per nacht drie bestanden gearchiveerd. Allereerst natuurlijk de statefile zelf. Daarnaast wordt het leesbare rapport alsook de CSV uitvoer opgeslagen. We doen dit in respectievelijk `/home/yaps/archive/` (de state file), `/home/yaps/report/` (het leesbare rapport) en `/home/yaps/csv/` (het comma seperated values bestand).
Tevens wordt er om 3 uur 's nachts een schaduwkopie van de gehele YAPS directory `/home/yaps/` gemaakt naar een andere hardeschijf op Lummi en op de NetApp.
We zijn hiermee dus verzekerd van een goedwerkend systeem.

Lijst van figuren

Lijst van tabellen

Literatuurlijst

Boekenlijst

Art and Science of C, The
– Eric S. Roberts (Addison-Wesley, 1995)

Complete Reference Juniper Networks Routers, The
– Jeff Doyle et al (Osborne, 2002)

Data Communications, Computer Networks and Open Systems
– Fred Halsall (Addison-Wesley, 1996)

Whitepapers

CWDM Whitepaper
– Fiber Network Engineering (313 Earhart Way, Livermore CA 94550)

ATM Pocket Guide
SDH Pocket Guide
Sonet Pocket Guide
– Wandel and Goltermann

EAPS Whitepaper
Extremeware User Guide, version 6.2.1
Extremeware Commandline Configuration Guide
Virtual Metropolitan Area Networks Whitepaper
– Extreme Networks Inc (3583 Monroe Street, Santa Clara CA 65051)

Bijlage A

Plan van Aanpak

Inleiding

Deze afstudeeropdracht wordt uitgevoerd in het kader van de opleiding Hogere Informatica van Fontys Hogeschool bij het bedrijf Business Internet Trends te Ede. Dit bedrijf is een hosting en Internet service provider die als doelgroep de zakelijk markt heeft. Om de groei van het bedrijf te kunnen ondersteunen, wordt een nieuw pand gebouwd, waarin een data- en telecommunicatie netwerk moet worden ontworpen en aangelegd. Het doel van dit netwerk is tweeledig. Enerzijds moet er een colocatie faciliteit worden gebouwd waarin bestaande en nieuwe klanten hun servers kunnen plaatsen. Anderzijds moet het mogelijk zijn voor derden om netwerken te koppelen aan het eigen netwerk en aan de AMS-IX [1].

Het bedrijf

De vennootschap Business Internet Trends BV is een zakelijke Internet service provider met een breed dienstenpakket voor bedrijven. De aangeboden diensten varieren van hosting tot Internet toegang, van het ontwerpen van websites tot het bouwen van database oplossingen. Bij al deze diensten ligt de nadruk op service en kwaliteit, zodat klanten hier op kunnen bouwen.
Enkele kenmerken van de dienstverlening zijn een eigen landelijk inbelnetwerk, eigen vaste verbindingen naar de Amsterdam Internet Exchange (ATM34 en STM1), die worden afgenomen van twee onafhankelijke telecom operators, geheel eigen apparatuur, veelal van Cisco Systems, waarop dagelijks beheer in eigen handen is. Voor alle klanten, dialup zowel als colocatie is er 24 uurs support, een real time monitoring systeem en een team met leden van een zeer uitgebreid kennisniveau om de klant te dienen en zodoende een optimale presentie op het Internet te kunnen garanderen.

[1] Amsterdam Internet Exchange - het grootste koppelpunt van bedrijfs- en onderzoeksnetwerken in Nederland

Aanleiding tot de opdracht

Business Internet Trends is voornemens haar huidige verbindingen naar de AMS-IX te vervangen door dark fiber. Momenteel zijn de verbindingen layer3 en wenselijk is een en ander naar layer2 om te zetten en te gaan werken met VLANs. Tevens wordt een tweede locatie geopend. Deze beide vestigingen dienen voorzien te worden van een redundant netwerk met load balancing. Verder zal een groot gedeelte van de bestaande apparatuur, zowel van het bedrijf als van haar klanten, verhuisd moeten worden naar dit nieuwe pand. Als laatste zal op de nieuwe locatie een nieuw bedrijf gevestigd worden dat een lokale Internet Exchange gaat opzetten en exploiteren. Hier moeten de klanten zowel lokaal peering op kunnen zetten met andere klanten, alswel een aansluiting kunnen krijgen op de AMS-IX via de eerder genoemde VLANs.

Probleemstelling

In het bedrijf heerst een nagenoeg volledige bezetting van de huidige werknemers. Ondanks de behoefte aan uitbreiding en documentatie van het netwerk, is er bij hen weinig tijd beschikbaar. Op het gebied van VLAN routing is verminderd kennis aanwezig binnen het bedrijf en er zal moeten worden gezocht naar uitvoerbare oplossingen voor de huidige routerings problemen.
Bij het opzetten van een Internet Exchange (IX) moet rekening gehouden worden met de gangbare normen voor een degelijke bedrijfsvoering. Hiervoor kan het beste worden gekeken naar en gesproken met de reeds draaiende AMS-IX.
Om klanten tevreden te houden en om een hoge kwaliteit van de diensten te kunnen blijven garanderen, zal een goed migratieplan moeten worden geschreven. Hiervoor is een documentatie inspanning nodig, welke wellicht niet goed geleverd kan worden door het huidige team, dit in verband met de constant hoge werkdruk.
Om deze redenen is de hulp van een stagaire ingeroepen, die de coördinatie en planning van de activiteiten, in samenspraak met de directie en uitvoerenden op zich zal nemen. Ook zal hij kennis moeten inbrengen in het bedrijf, op de de volgende vlakken.

- Router leveranciers, hun producten, hun prestaties (OSI, laag 1)
- Het splitsen van het netwerk in logische segmenten (OSI, laag 2)
- Het implementeren van nieuwe routingprotocollen in alle segmenten op alle locaties (OSI, laag 3)
- Het inbouwen van redundantie in het netwerk, door middel van load balancing in soft- of hardware (OSI, laag 7)
- Het vinden van een migratieplan dat de klantendiensten zo weinig mogelijk onderbreekt

Projectbeschrijving

In dit hoofdstuk wordt een beschrijving gegeven van de tien kenmerken die het project zullen definiëren. Per onderdeel wordt schuingedrukt een gangbare definitie

van het kenmerk gegeven. Hierdoor valt er niet te twisten over de interpretatie van de kenmerken en dit kan vooral voor de opdrachtgever van belang zijn.

De opdrachtgever

De opdrachtgever is diegene die de opdracht geeft tot het uitvoeren van een project en de daarvoor benodigde middelen ter beschikking stelt aan de opdrachtnemer.
De opdrachtgever is Dhr. MICHEL VAN OSENBRUGGEN namens Business Internet Trends BV te Ede (Gelderland).

De opdrachtnemer

De opdrachtnemer is diegene die zich verbindt tot het uitvoeren van de projectactiviteiten en het opleveren van het overeengekomen projectresultaat.
De opdrachtnemer is de stagiar, PIM VAN PELT, student aan de Fontys Hogeschool Informatica te Eindhoven.

Beginsituatie en uitgangspunten

Uitgangspunten zijn wezenlijke aspecten van een project die het resultaat kunnen beïnvloeden. Zaken waarmee het gehele project rekening zal moeten houden.
Uitgegaan wordt van het volgende.

- Er is een vaste werkplek voor de duur van het project. Dit werkstation heeft volledige Internet toegang via een vaste verbinding waardoor telewerken mogelijk is
- De stagiar heeft de beschikking over een vast telefoonnummer en e-mail adres waarmee hij kan communiceren met leveranciers en medewerkers
- Er zijn vaste tijdstippen waarop teruggekoppeld kan worden met de opdrachtgever
- Er is ruggespraak met Dhr. H.H. SCHAAF, de schoolmentor die de opdrachtnemer procesmatig begeleidt naar een bevredigend eindresultaat
- Er is een domeindeskundige aanwezig, die verregaande kennis heeft over het huidige netwerk ontwerp
- Er zijn 800 uren (100 dagen) gereserveerd voor het uitvoeren van de afstudeer werkzaamheden
- Alle betrokken partijen zijn in staat om het .PDF document formaat te lezen
- Er is de beschikking over een postscript printer, om alle documenten af te kunnen drukken

Doelstellingen

De doelstelling is een beschrijving van hetgeen men met het project wil bereiken. De projectdoelstellingen zijn niet de persoonlijke doelstellingen van de betrokkenen.
Het afstudeerproject heeft de volgende doelen.

1. Het vernieuwen van het fysieke netwerk, ook in het nieuwe pand
2. Het toekomstgericht herontwikkelen van het huidige productienetwerk
3. Het opzetten van een technisch model voor een Internet Exchange
4. Het opleveren van een migratieplan voor het (technische) team zodat de klanten tussen de huidige en nieuwe situatie kunnen worden omgezet

Probleemstelling

De opdrachtgever formuleert zijn problemen. Wat zijn de aanleidingen voor het starten van dit project. Bij welke punten ligt de prioriteit.
Zoals reeds aangegeven moet er een oplossing gevonden worden voor de volgende problemen.

- De kennis over de mogelijk in te zetten apparatuur is slechts deels aanwezig
- De indeling van het huidige netwerk is evoluerend gedaan en mist derhalve een zekere structuur
- Er is nog geen werkend plan voor een Internet Exchange, vooral omdat er geen mankracht voor beschikbaar is
- Er is geen migratieplan wat zorgt dat de ingebruikname van het te ontwerpen netwerk soepel verloopt

Projectresultaat

Het projectresultaat is de beschrijving van hetgeen het project na uitvoering concreet oplevert.
Na uitvoering van deze afstudeerwerkzaamheden zullen de volgende projectresultaten gerealiseerd zijn.

1. Er is een onderzoek gedaan naar de prestaties van verschillende hardware leveranciers waarbij de prijs-prestatie verhouding vergeleken wordt
2. Er is een ontwerp voor een nieuw productienetwerk binnen het bedrijf, inclusief de verbindingen tussen Ede en Amsterdam
3. Het netwerk in het nieuwe gebouw is geïmplementeerd en het huidige netwerk is daarin grotendeels getransformeerd. Hiervan is verslag gemaakt
4. Er is een technisch rapport welke beschrijft hoe de klanten kunnen worden opgenomen in de nieuwe productieomgeving
5. Er is een concrete Internet Exchange gebouwd in de nieuwe locatie

Wat behoort *wel* tot het projectresultaat

De beschrijving van de zaken die tot het project resultaat behoren. Dit zijn concrete zaken die wel tot de taken van de opdrachtnemer behoren.
Tot het resultaat van dit afstudeerproject behoren:

- De “deliverables” van de onderdelen die worden opgesomd in paragraaf 13:
    1. Het onderzoeksrapport van leveranciers en hun producten
    2. Het ontwerp van de nieuwe productieomgeving, met de verbindingen Amsterdam - Ede, de *backbone*
    3. Het ontwerp van de Internet Exchange faciliteit in Ede
    4. Het verslag van de implementatie van het nieuwe netwerk
    5. Het migratiedocument met aanbevelingen voor het verhuizen van de huidige klanten
- Het aansluiten en configureren van de routers voor de nieuwe *backbone* verbindingen
- Het koppelen van het huidige netwerk aan de nieuwe infrastructuur
- Het inrichten van de ruimte voor de Internet Exchange in het nieuwe gebouw

Wat behoort *niet* tot het projectresultaat

De beschrijving van de zaken die niet tot het project resultaat behoren. Dit zijn concrete zaken die niet tot de uiteindelijke taken van de opdrachtnemer behoren.
Tot het resultaat van dit afstudeerproject behoren niet:

- Het fysiek leggen van de glasvezelverbinding van Ede naar Amsterdam, de koppelpunten aan weerszijden, de interne bekabeling van het nieuwe pand
- Direct contact met de klanten, omtrent de taken van customer care en sales
- De daadwerkelijke verhuizing van klanten van de huidige naar de nieuwe situatie (het migratiedocument dient als ondersteuning van het team wat dit zal gaan uitvoeren)

Randvoorwaarden

Relevante omstandigheden, die de activiteiten en de resultaten van het project tijdens de uitvoering daarvan kunnen beïnvloeden. Randvoorwaarden hebben betrekking op de toekomst en hebben een beheersingskarakter. Met het begrip beheersingskarakter wil men duidelijk maken dat de voorwaarden allen te maken hebben met geld, expertise, kwaliteit, informatie en organisatie. Deze voorwaarden beheersen het project zeer duidelijk en zijn daarom belangrijk.
De project randvoorwaarden zijn de volgende.

- Er zullen voldoende middelen beschikbaar moeten worden gesteld om nieuwe hardware aan te schaffen

- De domeindeskundigen en opdrachtgever ruimen voldoende tijd in om afspraken met leveranciers te maken
- Het volledige traject alsook de specifieke afstudeerverplichtingen dienen eind 2002 afgerond, respectievelijk vervuld te zijn. Hiermee wordt aan de verplichting van minstens 100 werkdagen voldaan
- De te produceren documenten dienen aan de binnen Business Internet Trends gedefinieerde richtlijnen te voldoen
- Voor de visualisatie (typesetting) van documenten wordt TeX gebruikt, en alle partijen moeten om kunnen gaan met PDF files

Risico's

Een project risico is een onzekerheid uit de omgeving, die het verloop van de projectactiviteiten en het te bereiken projectresultaat in negatieve, maar ook in positieve zin kunnen beïnvloeden.

- Uitstel van de bouw en inrichting van het nieuwe pand. Als dit lang uitloopt, kunnen enkele projectresultaten niet of niet op tijd opgeleverd worden
- Er is ondanks de hulp van experts nog te weinig kennis over een bepaald onderdeel van het project
- Domeindeskundigen, zoals engineers van leveranciers, zijn niet beschikbaar, bijvoorbeeld door ziekte of nalatigheid

Projectfasering

In dit hoofdstuk wordt een overzicht gegeven van de verschillende zaken waaraan aandacht zal worden geschonken tijdens het afstudeerproject. In principe biedt het project een oplossing voor een vernieuwing van de netwerkfaciliteiten van het bedrijf. Derhalve zullen we in ons model in vijf fases een evolutie maken van huidige naar gewenste situatie.

Fase 1 - Projectinitialisatie

In deze beginfase zal de opdrachtnemer de tijd nemen om kennis te maken met de werknemers in het bedrijf, eerste afspraken maken over het vervolg en de invulling van de opdracht. Er zal een rondleiding door het bedrijf worden gegeven. Ook zal de opdrachtgever meer invulling geven aan zijn wensen en eisen voor het verdere verloop van het project. De opgedane kennis in deze fase wordt verwerkt tot het Plan van Aanpak, dit document. Na goedkeuring van dit document wordt aangevangen met de tweede fase.

Fase:	Initialisatiefase
Tijdspanne:	11 februari 2002 - 08 maart 2002
Ingeplande dagen:	8 (over 4 weken)
Opgeleverd document:	Plan van Aanpak (dit document)

Fase 2 - Onderzoek en Ontwerp

Deze fase bestaat uit twee onderdelen. Allereerst zal een staat van bevinding opgesteld worden. Er wordt gedocumenteerd hoe het netwerk nu is ontworpen. Daarna wordt nagegaan aan welke functionaliteit de opdrachtgever en de domeindeskundigen van het bedrijf behoefte hebben. Met deze kennis gewapend wordt een ontwerp gemaakt voor het nieuwe netwerk en deze wordt teruggekoppeld aan de directie en de afdeling engineering.

Fase 2a - Huidige situatie

Bij de beschrijving van de huidige situatie wordt nagegaan welke infrastructuur elementen, denk aan routers en switches, het netwerk vormen. Van deze hardware wordt een inventarisatie gemaakt, de specificaties worden opgezocht en goed doorgenomen. Van de Internet verbindingen wordt een schema getekend.

Daarna zal het lokale netwerk in Ede onder de loep genomen worden. Dit geldt niet alleen voor het fysieke aspect, zeker zo belangrijk is ook de configuratie. Van de routers en switches wordt nagegaan hoe ze zijn ingezet en waarom dit destijds is besloten.

Fase 2b - Gewenste situatie

Samen met het sales team wordt gekeken welke functionaliteit (in natuurlijke taal) de klanten zoal willen afnemen. Vervolgens wordt met het engineering team gekeken welke technische eisen deze functionaliteit aan het netwerk opleggen, en of dit haalbaar is. Uiteraard is een gedeelte van de klantenwensen reeds in productie gebracht en er wordt gezorgd dat deze worden meegenomen in het verdere traject. Aan de hand van deze specificaties worden potentiële leveranciers uitgenodigd om te spreken over een oplossing gebaseerd op hun producten. Hierbij is het belangrijk een goed overzicht van de mogelijkheden van de verschillende merken en types te behouden. De apparatuur van de leveranciers zal worden geëvalueerd in het test laboratorium van Business Internet Trends of bij de leverancier in zogenaamde *Proof of Concept* opstellingen. Hieruit zal een groter begrip van de apparatuur worden verkregen.

Fase 2c - Internet Exchange faciliteit

Het derde en laatste onderdeel van deze fase is het in kaart brengen van de wensen en eisen voor de te bouwen Internet Exchange faciliteit. Hiertoe zullen afspraken gemaakt worden met andere uitbaters van dergelijke faciliteiten in Nederland en met name de Amsterdam Internet Exchange. Deze afspraken hebben ten doel een geaccepteerde werkomgeving te kunnen creëren voor geïntereseerde klanten.

Fase:	Situatie beschrijving
Tijdspanne:	11 maart 2002 - 24 mei 2002
Geplande dagen:	28 (over 11 weken)
Opgeleverde documenten:	Onderzoeksrapport leveranciers
	Ontwerp nieuw netwerk
	Aanzet tot technisch ontwerp IX

Fase 3 - Klanten migratieplan

Nadat goedkeuring van de documenten in fase 2 wordt verleend, zal de netwerk apparatuur worden besteld. Dit brengt vaak lange wachttijd met zich mee en hierop wordt ingespeeld door aandacht te besteden aan een soepele transitie van klantensystemen tussen de twee netwerk generaties.
Uiteraard zullen zo weinig mogelijk klanten hinder moeten ondervinden van de migratie. Ook zal er voorbereidend werk gedaan moeten worden om de eigen servers te migreren naar het nieuwe netwerk.

Fase:	Klanten Migratieplan
Tijdspanne:	3 juni 2002 - 28 juni 2002
	13 augustus 2002 - 23 augustus 2002
Geplande dagen:	20 (over 6 weken)
Opgeleverd document:	Het migratieplan

Fase 4 - Implementatie

Eind augustus, wanneer de zomer een luwte voor ISPs met haar meebrengt, dient zich de mogelijkheid aan om het nu draaiende netwerk over te brengen in de nieuwe infrastructuur. Hierbij zal zorg moeten worden gedragen dat de dienstverlening zo weinig mogelijk in het gevaar komt en dat de klanten goed op de hoogte gebracht worden van de omzetting.
Deze omzetting gebeurt zonder de servers te verplaatsen naar de nieuwe locatie, waaraan tijdens deze fase nog volop gebouwd wordt. Aan het begin van het eerste kwartaal van 2003 zal het nieuwe gebouw worden opgeleverd. Daarvoor al, zullen de netwerk- en stroomvoorzieningen moeten worden aangelegd om direct productie te kunnen draaien met de nieuwe colocatieruimten. Hiervoor zal in het huidige gebouw al rekening worden gehouden, tijdens deze implementatiefase.

Fase:	Implementatie
Tijdspanne:	26 augustus 2002 - 18 oktober 2002
Geplande dagen:	24 (over 8 weken)
Opgeleverd document:	Implementatierapport
	Ontwerp van de IX

Fase 5 - Overdracht

Nadat in de vierde fase de eigen servers en diensten zijn gemigreerd naar de nieuwe infrastructuur, kan worden begonnen aan het afstudeerverslag. Hierin komt het hele traject samen tot één document voor de opdrachtgever. Dit afstudeerverslag is tevens het laatste in te leveren document voor de instelling. Alle opgedane kennis en expertise wordt zo nauwkeurig mogelijk omschreven in het document, welke ook als naslagwerk voor het bedrijf kan dienen. Het uiteindelijke verslag zal een bundel zijn van alle deliverables, genoemd in paragraaf 13.

Fase:	Overdracht
Tijdspanne:	21 oktober - 29 november
Geplande dagen:	20 (over 6 weken)
Opgeleverd document:	Het afstudeerverslag
	Presentatie van de werkzaamheden

Beheersaspecten

Om de beheersaspecten in dit project zo goed mogelijk te belichten, wordt gebruik gemaakt van het zogenaamde MOSQUITO-model. De letters in deze engelse term staan voor de zes beheersaspecten, namelijk: MOney (geld), Safety (veiligheid), QUality (kwaliteit), Information (informatie), Time (tijd) en tenslotte Organization (organisatie).
Uit sommige subtaken van dit afstudeerproject komt een beslisdocument naar voren. In het hoofdstuk over beheersaspecten bespreken we een onderverdeling in termen van tijd, expertise en kwaliteit. Ieder document brengt een eigen inspanning met zich mee, sommigen zijn gemakkelijker te realiseren dan anderen.

Aspect Geld

In het publieke document is geen referentie aan personen of gelden te vinden. Indien deze informatie noodzakelijk is, gelieve contact op te nemen met de directie van Business Internet Trends BV.

Aspect Veiligheid

Aan veiligheid in dit project is op het vlak van data integriteit gedacht. Alle relevante documenten worden met de grootste zorg gearchiveerd door de opdrachtnemer. Het volgende is voorzien.

1. De bestanden worden gecreëerd op een werkstation van de opdrachtnemer, BFIB.COLO.BIT.NL. Deze hangt in Amsterdam direct aan de Internet Exchange en zal gedurende het gehele project beschikbaar zijn.

2. Op het werkstation is CVS[2] geïnstalleerd. Alle bestanden worden na elke revisie in dit systeem opgeslagen zodat wijzigingen in de loop der tijd kunnen worden geraadpleegd. Tevens dient CVS als primair backup medium.

3. Elke nacht om 03.30 wordt een kopie gemaakt van het CVS en op een andere server, op een fysiek andere locatie, weggeschreven. Deze server zal in de colocatie ruimte in Ede komen te hangen. Voor deze backup zal rsync[3] gebruikt worden.

[2] Concurrent Versioning System, een stuk software wat geschreven is om bestanden vanaf verschillende lokaties door verschillende mensen te laten bewerken en centraal op te slaan in een zogenaamde *repository* (depot)

[3] Een utility om over het netwerk filesystemen te synchroniseren. We gebruiken dit om op een andere server een exacte kopie te maken van onze CVS repository

Deze maatregelen zullen er voor zorgen dat er geen verslagen, rapporten en referenties kunnen kwijtraken.

Aspect Kwaliteit

Voor Business Internet Trends en de opdrachtgever is het van het grootste belang dat er betrouwbare rapporten worden opgeleverd. Dit dient voor het bedrijf twee doelen.
Enerzijds zullen de rapporten bijdragen aan een onderhoudbaar netwerk voor de toekomst, ook als de opdrachtnemer na zijn afstudeerproject niet in dienst treedt. Uiteraard zullen de documenten in positieve zin het bouwen van het netwerk ondersteunen. Het is over het algemeen beter om dit soort netwerken eerst grondig te onderzoeken en ontwerpen, alvorens ze daadwerkelijk te implementeren.
Anderzijds ondersteunen deze rapporten de professionele uitstraling van het bedrijf naar zijn klanten. Er wordt zichtbaar waarde gehecht aan goed opgezette onderzoeken en ontwerpen door deze zakelijke markt. Dit project kan een belangrijke bijdrage leveren aan de sales prestaties van het bedrijf.

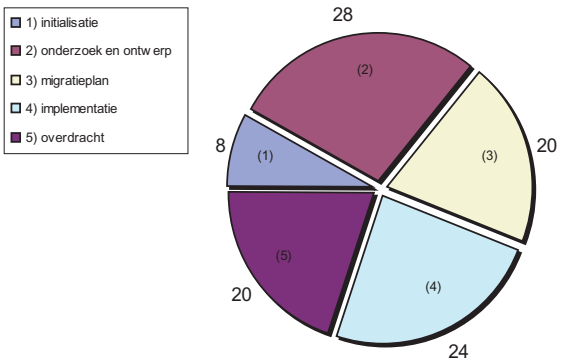
Aspect Informatie

Een belangrijk punt bij het verrichten van dit onderzoek is dat een groot gedeelte van de specificaties en handleidingen van de apparatuur die bekeken wordt in digitale vorm wordt aangeboden door de leveranciers.
Dit houdt in dat de referenties in de op te leveren documenten dikwijls websites of gedeelten daarvan zijn. Als het mogelijk is, zal de opdrachtnemer de bestanden downloaden en meenemen in het eindrapport. Hiervoor zal te zijner tijd een website worden opgezet waarop ook de voortgang van het project te zien zal zijn.
Er wordt veel gebruik gemaakt van email in dit project. Het communicatieplan in hoofdstuk 13 zal daarom van alle betrokkenen een werk- en privé email adres bevatten.

Aspect Tijd

Dit afstudeerproject, waarvoor 100 dagen gereserveerd zijn door de opdrachtnemer, wordt niet in nominale tijd (dus 20 weken) gerealiseerd. Derhalve zal niet in juni het project afgerond en opgeleverd worden. In plaats daarvan, is besloten om het project op te leveren in de maand december.
Ten grondslag aan deze beslissing, ligt het besef dat het bouwen van de colocatie ruimte in het nieuwe gebouw tijdrovend is. Als in april wordt begonnen met constructie, is per oktober de locatie op zijn vroegst af. Omdat de looptijd lang is, moet er voor gezorgd worden dat de opdrachtnemer zijn tijd goed in de gaten houdt.
Enerzijds is er een verlichtte tijdsdruk omdat, als drie dagen per week niet voldoende is om een fase af te handelen, er meer tijd vrij gemaakt kan worden. Anderzijds moet men goed de tijd in de gaten houden. Door middel van een dagenverantwoording zal de besteedde tijd worden verantwoord. Een voorlopige schatting van

de tijd die besteed is in de verschillende onderdelen van het project is weergegeven in figuur A.1.

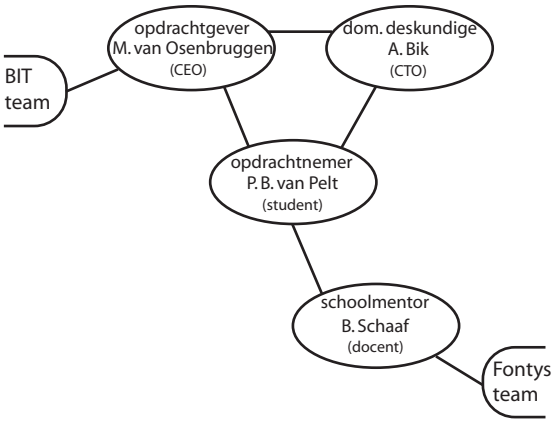


Figuur A.1: *De besteedde tijd, in dagen, per fase van het project*

De looptijd van het project is van 11 februari 2002 tot 29 november 2002 en bestrijkt 43 weken. Aangezien het project gefaseerd verloopt, is het toch mogelijk om een goede beheersing van de tijd te handhaven. Er zal afwisselend twee of drie dagen per week worden gewerkt aan het afstuderen. De overige dagen wordt besteed aan het afronden van andere modulen van Fontys en het verrichten van andere nuttige werkzaamheden binnen het bedrijf.

Aspect Organisatie

Zoals bij elk project, bestaat deze afstudeeropdracht uit meerdere mensen die op een bepaalde manier met elkaar omgaan. In figuur A.2 wordt deze samenhang weergegeven.



Figuur A.2: *De direct betrokken personen in dit afstudeer project, tussen haakjes hun functie*

De opdrachtgever, ook wel bedrijfsmentor genoemd, werkt samen met de op-

drachtnemer om tot een zo goed mogelijk project resultaat te komen. De domeindeskundige is een medewerker van Business Internet Trends, die direct in contact staat met de student. Opdrachtgever en domeindeskundige zullen samen over de kwaliteit van de documenten (de producten) oordelen.
Alle interactie met de andere medewerkers van het bedrijf, de sales afdeling, de afdeling engineering en de klanten, zal via de opdrachtgever verlopen. Deze zal zijn toestemming moeten geven alvorens de opdrachtnemer andere medewerkers van het bedrijf werkzaamheden kan laten uitvoeren.
De begeleider van Fontys, ook de schoolmentor genoemd, oordeelt over de procesmatige aanpak van de opdrachtnemer. Samen met deze begeleider zal een zo goed mogelijk resultaat voor Fontys worden nagestreefd. Achter deze mentor staat het Fontys team. Met hen, in het bijzonder de voorzitter van de afstudeercommissie, wordt via de mentor contact gehouden.

Communicatieplan

Belangrijke adressen

In het publieke document is geen referentie aan personen of gelden te vinden. Indien deze informatie noodzakelijk is, gelieve contact op te nemen met de directie van Business Internet Trends BV.

Terugkoppeling

Er wordt afgesproken tussen de schoolmentor en de opdrachtnemer om minimaal eens per kalendermaand contact te hebben. Er wordt gevraagd om een regelmatige email communicatie. In de maand augustus wordt de schoolmentor uitgenodigd om een kijkje te komen nemen op een evenement wat georganiseerd wordt door de werknemers van Business Internet Trends. Meer informatie is te vinden op http://www.megabit.nl/
De opdrachtnemer en opdrachtgever hebben wekelijks contact via de mail. Er wordt afgesproken dat het werk deels thuis kan worden gedaan, maar dat er minimaal één dag per week in Ede doorgebracht zal worden. Ook zal de opdrachtnemer, in overleg met de domeindeskundige en opdrachgever, op locatie (bij leveranciers) gaan om hen een goed beeld te kunnen geven van het project, primair vanuit het perspectief van het bedrijf.
Er wordt afgesproken dat elk document in minimaal twee versies wordt voorgedragen. De eerste, voorlopige versie, wordt aan beide mentoren (dus de opdrachtgever en de schoolmentor) gestuurd. Deze zullen binnen 10 werkdagen een reactie geven op dit document. Eventuele wijzigingen worden in de tweede versie doorgebracht en deze wordt ter goedkeuring aan de opdrachtgever voorgelegd. Dit proces gaat zo lang door als nodig is, waarbij het eindresulatat een document is welke de goedkeuring kan wegdragen van zowel de opdrachtgever als de schoolmentor.
Om alle partijen op de hoogte te houden, zal de student in elke fase een visuele presentatie geven van zijn werkzaamheden. Hierdoor blijft niet alleen het proces duidelijk, ook de producten van de afstudeerwerkzaamheden worden dan toegelicht.

Bijlage B

Achtergrondinformatie over IXPs

Internet Service Providers

Als een consument of een bedrijf aansluiting wil vinden op het Internet, wenden zij zich tot een Internet Service Provider, een *ISP*. Zij zorgen ervoor, dat de website van het bedrijf voor alle andere deelnemers van het Internet bereikbaar zijn. Tevens zorgen zij er voor dat de eigen klanten het gehele Internet kunnen bereiken.

Vraagbundeling

Een ISP voorziet in vraagbundeling. De gebruikers, die via xDSL, kabelmodem of ISDN bij hen verbinding zoeken, genereren IP verkeer. De ISP zal het gezamenlijke verkeer van de klanten bundelen en naar het Internet wegzetten. Net zoals consumenten, zullen ISPs ook hun verkeer moeten betalen. De vraagbundeling zorgt er voor dat de ISP tegen gunstigere tarieven dataverkeer kan inkopen.

Tier-1 en tier-2

Traditioneel zal een ISP bij een andere ISP verkeer inkopen, ofwel tegen een vast tarief per maand, ofwel tegen een vast tarief per weggezette volume verkeer. Er is een gelaagdheid in ISPs, wat met het engelse woord *tier* wordt weergegeven.
Een tier-1 ISP is een wereldwijd actieve backbone provider. Deze providers hebben geïnvesteerd in (glasvezel)verbindingen over grote gebieden van de wereld en kunnen dus gedeelten van deze verbindingen verder verhuren aan hun klanten. Dit type ISP wordt doorgaans een *carrier* genoemd.
Een tier-2 ISP is klant bij één of meer tier-1 ISPs. Zij zullen de tier-1 ISP betalen om hun dataverkeer te transporteren naar andere delen van het Internet. Dit type ISP is het meest voorkomend en is wat de meeste mensen verstaan onder de gewone term *internet service provider*.
Het leidt geen twijfel dat het wegzetten van verkeer tussen twee nabijgelegen (tier-2, dus “gewone”) ISPs niet altijd hoeft te geschieden met tussenkomst van een carrier. Zo is het mogelijk dat er twee ISPs zijn die in de zelfde stad een vestiging hebben. Zij zouden kunnen afspreken om het verkeer, wat gericht is aan elkaar, zelf uit te wisselen. Hierdoor hoeven zij geen verkeer meer in te kopen bij de carrier, voor het verkeer wat ze zelf kunnen wegzetten.

Internet Exchanges

Bij de opkomst van het Internet in Nederland, zijn bedrijven zich gaan realiseren dat het voor hen kosteneffectief is om zo veel mogelijke directe verbindingen te hebben met andere ISPs. Zo kan een webhosting ISP baat hebben bij een verbinding naar een dialup-ISP, zodat de surfende klanten van de dialup ISP snel en efficiënt de websites binnen het eigen bedrijf kunnen bezoeken.
Allerlei ISPs begonnen in Nederland rond de Universiteit van Amsterdam in de Watergraafsmeer samen te klonteren. Hier werd hen de mogelijkheid geboden om via een gedeelde *switch* verkeer uit te wisselen. Men moest dan zorgen dat men samen de kosten van de switch terug verdiende, maar het uitwisselen van verkeer was verder kostenloos. Het bundelen van vraag en aanbod door middel van gemeenschappelijke infrastructuur wordt een Internet Exchange Point (*IXP*) ofwel *IX* genoemd. De IX in Amsterdam werd de “AMS-IX” genoemd. Een andere term voor dergelijke initiatieven is Network Access Point (*NAP*), wat vooral door de oudgediende amerikaanse netwerk engineers wordt gebruikt.

Historische blik op Internet

In het oorspronkelijke model, waarbij een ISP alle verkeer inkoopt bij een carrier ISP, moet de ISP dikwijls een huurlijn aanleggen naar het dichtstbijzijnde inkoppelpunt van de carrier. Doorgaans hebben de Nederlandse carriers een redelijk fijnmazig netwerk waardoor een huurlijn naar de dichtstbijzijnde grote stad voldoende blijkt. Alle verkeer wordt nu door de carrier, meestal tegen een vast tarief per maand, of tegen een variabel tarief per volume eenheid verkeer per maand, weggezet op het Internet. De ISP hoeft zich geen zorgen te maken over de routering van het verkeer buiten zijn netwerk. Dit is een dienst die de carrier levert aan de ISP.

Hedendaagse blik op Internet

De samenklontering van bedrijven, ISPs en carriers in Amsterdam heeft ervoor gezorgd dat hier de vraag het aanbod kan ontmoeten. De aan de IXP deelnemende bedrijven kunnen hier een groot gedeelte van hun volume (van 30–50 procent) zelf wegzetten.
Voor een Nederlandse ISP, met een Nederlandse doelgroep, kan het financiëel wenselijk zijn om een aansluiting te vinden bij een dergelijke IX.
Met de opkomst van de (vrije) telecom-markt in Nederland, blijkt dat het aanleggen van een vaste verbinding (bijvoorbeeld van Arnhem naar Den Haag) tegenwoordig veel goedkoper is. Ook zakken de prijzen voor kortere huurlijnen, zoals bijvoorbeeld van Arnhem naar Nijmegen.
Men kan er dus voor kiezen, om de huurlijn direct naar het IXP te leggen en hierop een aansluiting te nemen.

Benodigdheden voor een IXP aansluiting

Een vaste verbinding

Allereerst heeft de ISP een vaste verbinding (huurlijn) nodig naar het IXP. Dit kan een klassieke telefoniedienst zijn, zoals een E1 (2 Mbps) of een E3 (34 Mbps) of

een glasvezelverbinding, zoals een STM1 (155 Mbps) of zelfs STM16 (2.5 Gbps). De prijzen van dergelijke verbindingen verschillen behoorlijk van telco tot telco en het is zaak om alle mogelijke leveranciers offertes te laten doen.

Een router

Daarnaast heeft het bedrijf minimaal één router nodig. De router zal ervoor zorgen dat het verkeer van de ISP naar de juiste plaats wordt weggezet. Omdat deze router aan de rand van het eigen netwerk staat, wordt hij dikwijls een Border Router genoemd.

Een AS nummer, een IP netwerk

Een verzameling van switches, routers en servers die onder beheer van één instantie valt, wordt een autonoom systeem (*AS*) genoemd. Een AS heeft één routing policy, die geldt voor het AS in zijn geheel. Er zal een registratie van dit systeem bij een erkende autoriteit moeten worden voorzien. Voor Nederlandse bedrijven is dit het Netwerk Coördinatie Centrum van het Réseaux Internet Protocol Européens, afgekort *RIPE-NCC*. Deze instantie verzorgt (onder andere) de uitgifte van AS nummers en IP adressen.
Zowel AS nummers als IP adressen kunnen worden verkregen op twee manieren.

1. Allereerst kan men er voor kiezen om lid te worden van RIPE. Men wordt dan een Local Internet Registry (*LIR*). Hieraan verleent RIPE-NCC bepaalde diensten met betrekking tot de registratie van AS nummers en de toekenning van unieke Europese IP adressen.

2. Daarnaast kan men kiezen om het AS nummer en de IP adressen aan te vragen via een bestaande LIR, bijvoorbeeld een carrier. Voor deze diensten zal de carrier dan waarschijnlijk een bedrag in rekening brengen. Hoe hoog dit bedrag is, verschilt behoorlijk van carrier tot carrier.

Een contract met een IXP

Er moet een IXP gekozen worden. Voor Nederlandse bedrijven is dit oorspronkelijk de AMS-IX gebleken. Hier vinden inmiddels ruim 130 bedrijven aansluiting en kunnen zij verkeer uitwisselen.
Er zal een connectietype moeten worden gekozen. Deze kan varieren van 10 Mbps voor een kleine ISP, 100 Mbps voor een middelgrote en 1000 Mbps (één gigabit) voor ISPs met een groot datavolume.

Verkeer uitwisselen

Als de fysieke verbinding tot de IXP tot stand is gebracht, zal met de andere deelnemers contact moeten worden gezocht. Men maakt dan afspraken over het uitwisselen van verkeer over de infrastructuur van het IXP. Dit gebeurt meestal via e-mail, alhoewel sommige (met name grotere) ISPs de afspraken in een schriftelijke Bi-Laterale Peering Afspraak (*BLPA*) willen vastleggen.

Niet alle bedrijven hebben dezelfde strategie wat het betreft peering. De ene zal met iedere partij verkeer willen uitwisselen, de andere niet. Deze strategie voor een ISP wordt omvat in zijn peering policy.
De gedachte van veel ISPs is dat er met zo veel mogelijk andere ISPs verkeer moet worden uitgewisseld. Dit scheelt immers in de kosten van het gebruik van een carrier netwerk, verkort de route naar het bestemmingsnetwerk en verhoogt de snelheid en de betrouwbaarheid van datauitwisseling.
Er kan zelfs een Multi-Laterale Peering Afspraak (*MLPA*) opgesteld worden. Hierin stemt de ISP toe in het peeren van elke willekeurige andere ISP die ook een dergelijke afspraak heeft gedeponeerd. Deze werkwijze heet OpenPeering. Het voordeel hiervan is dat een nieuwkomer aan het IXP, bij het tekenen van de MLPA, peering opzet met alle andere deelnemers tegelijk en dus niet apart met elke deelnemer afspraken hoeft te maken.

Peering

Het op een IXP verkeer uitwisselen met andere deelnemende ISPs, door een openpeering initiatief (MLPA) ofwel door een aantal één–op–één peering afspraken (BLPAs), wordt peering genoemd. Dit komt van het engelse woord *peer*, wat gelijke betekent. Het uitwisselen van verkeer naar peers gebeurt in de regel met gesloten beurzen.

Transit

Bedrijven die geen aansluiting hebben op de dezelfde IXP, of geen peering agreement hebben, zullen elkaar toch moeten kunnen bereiken. Dit gebeurt door de traditionele carriers. Zij hebben verbindingen aangelegd tussen de verschillende europese steden en zijn bereid, om tegen betaling, het verkeer tussen de verschillende netwerken te transporteren.
Deze providers laten dus verkeer toe van het ene netwerk, dóór hun eigen netwerk, naar een ander netwerk. Hierdoor worden deze providers ook wel transit-providers genoemd en het inkopen van volumes verkeer door hun netwerk gewoonweg transit kopen.

Kosten/Baten

Er moet een router worden gekocht en een AS nummer worden aangevraagd. Ook moet er een huurlijn naar de dichtstbijzijde Internet Exchange worden gelegd, of een andere ISP worden gevraagd om transit te leveren naar die Exchange.
Daarnaast wordt het goedkoper gemaakt om verkeer op het IXP uit te wisselen. Zo kan in transitkosten worden bezuinigd. Voor een ISP met voornamelijk Nederlands verkeer, bijvoorbeeld een webhosting bedrijf met voornamelijk Nederlandse klanten, kan tot 50 procent aan kosten voor dataverkeer worden bespaard.

Bijlage C

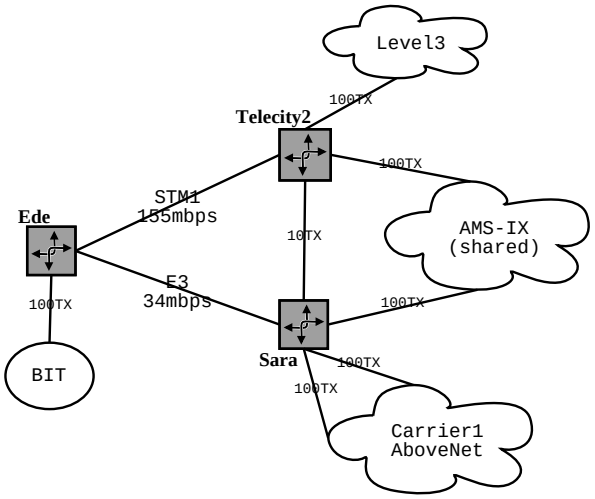
Voorstel naar Leveranciers

Introduction

Before we start, it is wise to take a short look at the current IP production network of the Business Internet Trends company. We have constructed a triangle consisting of three locations:

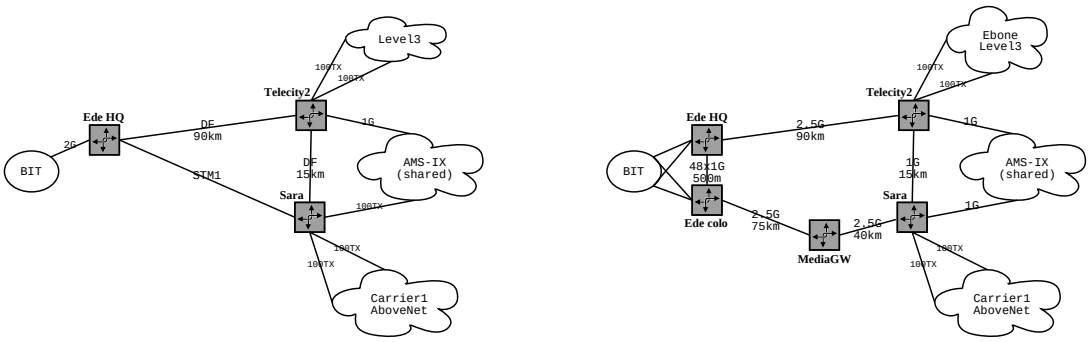
- Ede Head Quarters
- Amsterdam Internet Exchange: Sara location
- Amsterdam Internet Exchange: Telecity location

We operate three separate circuits. The oldest is an E3 running an ATM pvc between Ede and Sara. It operates at 34 mbps. The second is an STM1 link from Ede to Telecity, running at 155 mbps. The two Amsterdam locations are crossconnected via a 10 mbps VLAN, leased from TeleCity. The current network layout is shown in figure C.1.



Figuur C.1: *The current network situation for Business Internet Trends*

Recently, Business Internet Trends acquired a pair of dark fibers from Level3. It runs from our main site (Ede HQ) to the Telecity location and will be operational



(a) The intermediate situation, starting today

(b) The new backbone situation, planned for Q1-2004

Figuur C.2: *The Business Internet Trends backbone evolution during the next 24 months*

in Q2-2002. A second pair of dark fibers was rented from Energis, to connect both Amsterdam locations.
To sustain current growth, we are in the process of erecting a new colocation facility which will house some 1400 square meters of floor space, devided into 10 rentable suites of approximately 100 square meters.
There will be a duct between the HQ and this location, with 24 pairs of single mode and 24 pairs of multimode fiber.
In addition to this, we will acquire a second pair of dark fiber from some telco, connecting this new colocation facility to the Amsterdam Sara location. An attempt will be made to connect the MediaGateway in Hilversum to either the Ede location or the Amsterdam location. Then we will have an additional two locations:

- Ede colocation facility
- Hilversum: The mediagateway

Once the first darkfiber (90km) is operational, we will be able to migrate the current STM1 between Ede and Telecity to run from Ede to the Sara building. The E3 can then be dismantled.
This will result in a backbone evolution depicted in figure C.2 which is set to start today and reach maturity in early 2004.

Problems and Goals

The current network, consisting of two Cisco 7206VXR-300s (Ede and Sara) and one Cisco 7505 (Telecity) is not capable of high speed data transfer. Therefore, we must invest in a new infrastructure from layer1 and up.
We have not yet decided whether we should use transmission technology and L2/L3 functionality in the same box, or spread this out into two machines. Using a DWDM solution and transporting several gigabit ethernet trunks over this might be a more cost effective solution than having an OC48 connection. Especially the cheaper

CWDM equipment, with a throughput of up to 1.25 gigabit per channel (and 8 channels per fiber) seems like a fine choice.
We will define operational needs on the equipment and challenge several vendors to come forward with a cost effective solution based on their available equipment.
We have not yet built the colocation facility, nor acquired fiber from that location to Amsterdam, so today our demands are to erect a highspeed link from Ede to Telecity and from there to the Sara building. Near-future plans (2003) will extend the network from Sara back to Ede, possibly touching the Mediagateway in Hilversum.

Parameter matrix

We define a specific set of features we would like to see implemented in the new situation, keeping in mind the ISP and colocation facilities are primary business to the company. We would like to present value added services to our customers, lease layer2 VLANs between amsix and Ede to other companies (up to gigabit), and offer IP connectivity in the HQ and COLO locations. This list of features is given in tables C.1 and C.2.

Feature	*Quantity*	*Details*
Transmission		
OC48	2	Only needed in non-WDM setup, power budget at least 28 dB.
OC192	2	Only needed in non-WDM setup, power budget at least 28 dB.
DWDM	4 channels	Each channel at least 2.5 gbps
CWDM	4 channels	Each channel at least 1 gbps
GigE	4	Minimum 3, maximum 8
Copper FE	4	Minimum 2, maximum 24
Performance		
packets/second	5M	
bits/second	5G	
nonblocking	-	routing, filtering, ratelimiting, accounting
memory	256MB	4 BGP full views

Tabel C.1: The list of hardware features (1)

We do not expect all of the functionality to be in one machine. We do however wish to see one IP machine with enough switching capacity, also under modest (gigabit throughput) loads. For transmission, we accept the possibility to attach two or more gigabit trunks to a DWDM transponder. In that case, we can use gigabit ethernet to transport VLANs from Amsterdam to Ede and switch to STM-16 (or STM-64) when we reach a critical load of the gigabit trunk ports. We'll need at least 4 wavelengths in this case.

Features	*Details*
IPv6	Native, non-software implementation
Filtering	ACL L3 (extended): protocol, src dst srcprt dstprt For both IPv4 and IPv6!!
Accounting	IP (L3) aggregate per IP/network
Sampling	high datarate (gbps range)
Flow logging	to loghost or local disk
ssh to terminal	ssh1 or ssh2 instead of telnet
IPv6-in-IPv4	standard proto-41 tunnel
IPv4-in-IPv4	standard proto-94 tunnel
GRE	Generic tunneling
MPLS	
Link aggregation	Combining GE ports to a larger trunk
Hotswappable	interface and/or linecards
Routing	
BGP4+	Both: demand
ISIS	IPv4: demand
OSPF	IPv4: demand
RIP(ng)	optional

Tabel C.2: The list of features we expect from the backbone network as a whole (2)

Possible solutions

As said previously, we are investigating either one of two options. In the first case, we would invest in one box per location, performing the core and edge functionality, and taking care of the long haul.
Alternatively, we may lay down a seperate circuit via means of Dense or perferrably Coarse WDM systems, and coupling one or more gigabit ethernet ports on either side of the link.
In any case, we will start with a total combined bandwidth of 2.5 gbps from day one. Any given channel in the system should support a wirespeed data throughput of 1 gbps, and no less.
In order to gain complete confidence in any given solution, we would like to be able to test-drive the equipment in a real life setup, thus connecting a box at either end of the link and seeing them prove link stability and throughput. The price of the system as a whole plays a major role in the decision making regarding the new backbone.

Using SDH transport

Should the vendor choose to supply a solution based on SDH technology, we should take into account the following aspects:

- The distance between Ede-Amsterdam is 89 km in total. Preliminary calculations, based on worst case, dictate a minimal power budget of 28 dB.
- The minimal supplied bandwidth is STM16.

- The circuit should be able to cross connect layer2 VLANs at gigabit speed. We cannot tolerate one STM channel per gigabit port, but need to see 8 STM channels committed per gigabit port
- We would like to be able to exchange an STM16 module for an STM64 module without the need for a new chassis. Otherwise, we would like to see a scenario where the vendor has provisions for exchanging the chassis.

Using WDM transport

In this case, the requirements for the router boxes will be substantially lower. We will take into account the following:

- The WDM box should be able to transmit channels in excess of 1 gbps. It should have at least 4 channels. We do not have a need for 32 or more channels at this point so high grade DWDM equipment is not preferrable over lower yield CWDM equipment.
- The WDM box should have gigabit ethernet connectors (multimode and/or singlemode).
- The router box should have multiple gigabit ports (at least 6)

Vendor specific

We would like to suggest the following course of action from your party.

1. Study the wishlist in tables C.1 and C.2 closely.
2. Chose a setup that can meet our wishes and present this using either one of two solutions described in this document.
3. For this setup, please present a setup that will conquer the tasks to build the “intermediate stage” network, depicted in figure C.2(a), delivering a long haul bandwidth of 2.5 gbps
4. Idem, but with long haul bandwidth of 10 gbps from day one
5. Idem, but starting today with 2.5 gbps, and some ideas on migrating to a 10 gbps platform in the timeframe of 12-24 months.
6. List papers, datasheets and pricing of these three solutions.

Building these scenario’s, please bear in mind the continuing efforts to expand the Business Internet Trends network to Hilversum and back to Ede, creating a completed ring structure topology.
If there are any requests, comments or other information, please do not hesitate to contact us. The engineer handling this project can be reached at:

Name:	Pim van Pelt
Email:	**pim@bit.nl**
Phone:	+31 318 648688

Bijlage D

Brief naar de Klanten

Geachte relatie,

Om de constante groei van Business Internet Trends BV te kunnen ondersteunen, is besloten om in een nieuwe generatie netwerk te investeren. Dit netwerk zal gebaseerd zijn op eigen zogenaamde “dark fiber” glasvezels, die via geografisch gescheiden routes van Ede naar Amsterdam lopen. De capaciteit zal hierbij van 155 Mbps naar 2 Gbps gaan, een vertwintigvoudiging!
Het in gebruik nemen van deze nieuwe backbone zal gepaard gaan met een grootschalige migratie van onze huidige diensten. Hiervoor willen we u op voorhand verwittigen, aangezien de nieuwe situatie in sommige gevallen kan afwijken van de huidige situatie. Om te voorkomen dat bepaalde zaken ongewenst niet (meer) werken, presenteren we u een overzicht van de diensten zoals die ingesteld zouden moeten zijn. Let er op dat u alleen voor de DNS servers IP nummers gebruikt en voor alle andere diensten de namen.

Inkomende e-mail	pop.bit.nl	DNS Server 1	213.136.12.52
Uitgaande e-mail	smtp.bit.nl	DNS Server 2	213.136.12.60
Web Proxy	proxy.bit.nl	News Server	news.bit.nl

We raden aan om deze instellingen even door te lopen teneinde onnodige problemen tijdens of na de migratie te voorkomen.
Hiervoor zijn eind augustus en begin september enkele onderhoudswerkzaamheden gepland binnen Business Internet Trends BV. Tijdens deze werkzaamheden zullen delen van de dienstverlening kort onderbroken worden, vooral ’s nachts.
Voor dit doel is een (e-mail) mailinglist opgezet waarop de engineers van Business Internet Trends BV hun werkzaamheden aankondigen en regelmatig statusupdates sturen. Het is raadzaam u op deze mailinglist te abboneren. Dit kan door middel van het bezoeken van de volgende website: **http://mailman.bit.nl/mailman/listinfo/onderhoud**
Voor vragen voor of tijdens de migratiewerkzaamheden, kunt u zich altijd wenden tot onze medewerkers, per e-mail naar (**migratie@bit.nl**) of per telefoon tijdens kantooruren naar **0318-648688**.

Met vriendelijke groet,

Pim van Pelt

Bijlage E

YAPS Configuratie commando's

YAPS configuratie

In deze appendix sommen we de verschillende configuratie mogelijkheden van het YAPS programma op. We geven bij elk configuratiecommando een woordje uitleg.
Allereerst merken we op dat alle regels die beginnen met `#` worden beschouwd als commentaar. Tevens wordt alles na het voorkomen van `//` als commentaar gezien. Dit laatste geldt ook in een regel, maar dan weer niet binnen enkele of dubbele quotes (' of ").
Het 'interface' commando vertelt YAPS welke netwerkkaart hij moet gebruiken om verkeer op te samplen. Dit interface wordt dan in "promiscuous" mode gezet zodat hij al het verkeer ontvangt, en niet alleen het verkeer naar zijn eigen MAC adres. Bijvoorbeeld:

```
interface "fxp0";
```

Het programma kan zichzelf als daemon draaien, of in de voorgrond draaien. Hiervoor hebben we twee mogelijke commando's:

```
daemonize;
```

en

```
no-daemonize;
```

Het 'localnets' commando creëert een lijst met CIDR netwerken die beschouwd worden als 'intern' voor de ISP. Aan de hand van deze lijst kunnen we dan voor elk frame zien of hij inwaards, uitwaards, of lokaal gericht was. Bijvoorbeeld:

```
localnets {
    "213.136.0.0/19";
    "193.109.122.0/24";
};
```

YAPS houdt intern de tellertjes van de IP nummers bij in een hash tabel. De sleutel van deze functie is de laatste 'n' bits van het IP adres. Hoe groter we 'n' kiezen, hoe meer buckets we krijgen. We kunnen de waarde echter niet te groot maken, omdat geheugengebruik exponentiëel groeit met 'n'. Een ideale waarde is tussen de 4 en 20 bits. Het geheugen wat hierbij gebruikt wordt is 2^{n+2} bytes. Met $n = 16$ zou dit een gebruik van 256 Kilobyte opleveren. Bijvoorbeeld:

```
iphash-bits 16;
```

Het ‘acctnets’ commando vertelt YAPS voor welke IP nummers (of netwerken) we tellers wensen bij te houden. Voor elk IP nummer wordt een entry in de iphash tabel gemaakt. In principe mogen ook externe IP adressen worden ingevuld. In dat geval wordt bijgehouden hoeveel verkeer daar van- en naartoe wordt gestuurd vanuit het netwerk. Meestal echter zullen de eigen netwerken worden genomen. Van elk IP adres worden de volgende zaken bijgehouden:

- Het aantal IP frames ingaand, uitgaand en lokaal.
- Het aantal TCP, UDP en ICMP pakketjes ingaand, uitgaand en lokaal.

Een voorbeeld zou kunnen zijn:

```
acctnets {
    "213.136.0.0/19";    // Een intern netwerk (zie localnets)
    "193.109.122.0/24";  // Nog een intern netwerk (zie localnets)
    "212.19.192.216/29"; // Een extern netwerk
};
```

Met het ‘acctports’ commando wordt YAPS duidelijk voor welke TCP en UDP poorten we wensen op te slaan in de statefile. YAPS zal altijd van alle 65536 UDP en TCP poorten bijhouden hoeveel verkeer er op geconstateerd is (wederom voor zowel in- als uitgaand verkeer en lokaal verkeer), maar slechts van die poorten melding maken in de statefile, die hier worden opgegeven. Hierbij neemt de diskaccess aanzienlijk af (omdat we dikwijls alleen maar geïnterreseerd zijn in poorten onder de 1024). Voorbeeld:

```
acctports {
    tcp { 1-1024; 1080; 3128; 6660-6670; 8080; };
    udp { 1-1024; };
};
```

De statefile kan worden opgeslagen in een door de gebruiker aangegeven directory. Wordt die niet opgegeven, dan wordt de default waarde `/usr/local/yaps/var/` gebruikt. Voorbeeld:

```
statedir "/home/pim/yapsstate/";
```

Elke ‘dumpinterval’ seconden wordt een backup gemaakt van alle tellertjes naar de disk. Hierbij wordt de statefile in de `statedir` directory geplaatst en worden alle IP nummers in `acctnets` en de UDP/TCP poorten in `acctports` weggeschreven naar de disk. Standaard staat deze waarde op 300, bijvoorbeeld:

```
dumpinterval 300;
```

Het al dan niet inlezen van de statefile na het opstarten (zodat YAPS verder kan tellen waar hij gebleven was), wordt gestuurd door:

```
keepstate;
```

en

```
no-keepstate;
```

Als we willen dat YAPS bandbreedte berekeningen doet, zullen we hem twee setjes tellers moeten laten gebruiken. Hij zal dan periodiek het verschil tussen de oude en nieuwe waarden berekenen en delen door de periode, zodat gemiddeld bandbreedte gebruik per IP kan worden berekend. Het al dan niet uitvoeren van deze berekening hangt dus af van:

```
ipstats-diffs;
```

en

```
no-ipstats-diffs;
```

Vervolgens kan elke ‘throughputinterval’ seconden de bandbreedte berekening gebeuren. Het is verstandig deze niet te laag te kiezen. Een waarde tussen de 15 en 600 lijkt op zijn plaats. Zijn standaard waarde is 60 seconden. Bijvoorbeeld:

```
throughputinterval 60;
```