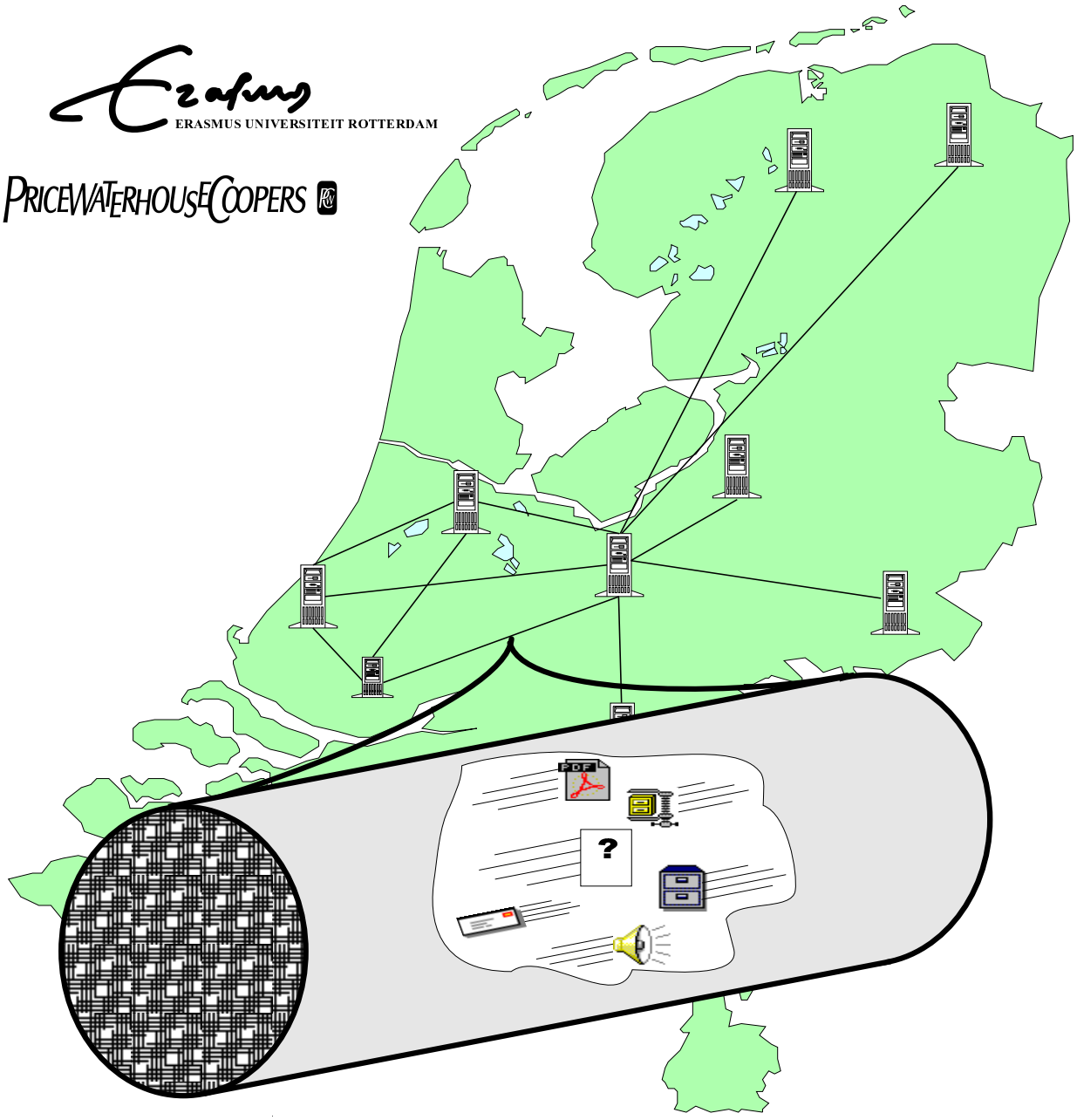
Network Control

Edward van der Jagt

ERASMUS UNIVERSITEIT ROTTERDAM

PRICEWATERHOUSECOOPERS



Network Control

Doctoraalscriptie
Bestuurlijke Informatica
Erasmus Universiteit Rotterdam
Edward v/d Jagt

15 mei 2001

Voorwoord

Na ietwat te veel jaren studie en diverse pogingen is het dan eindelijk zo ver: hier ligt mijn scriptie ter afsluiting van de studie Bestuurlijke Informatica.

Na een aantal voortijdig afgebroken pogingen waarin ik onderzoek was begonnen met als uiteindelijk doel een afstudeerscriptie, kwam ik via een omweg bij Pricewaterhouse-Coopers terecht. Na (te) lang diverse anderen te hebben geholpen met hun werk en ook enkele onderwerpen na korte tijd afgebroken te hebben, kwam daar in een van onze vele discussies ineens Anne Bakker met een idee voor een onderwerp opdraven. De stelling die we na enige discussie namen was dat de meeste organisaties geen idee hebben waar hun (WAN) netwerk voor wordt gebruikt, terwijl er toch een flinke zak geld aan wordt uitgegeven. En dan komt vervolgens de vraag hoe je daar dan achter kan komen. En zie daar, één scriptie-onderwerp.

Bij het schrijven loop je dan ondanks een hoop moeite toch tegen de zelfde problemen als anderen hebben bij het schrijven: gebrek aan inspiratie en geen idee hoe je verder moet.

De eerste keer was het Frank van Vonderen die mijn probleem inzag en de juist tips gaf. Zelfs al was het reeds diep in de nacht en mochten we niet meer achter het stuur kruipen. De tweede keer was het Paul de Jong die het gebrek aan inspiratie op wist te lossen met een suggestie die zo helder en voor de hand liggend was, dat je jezelf afvraagt hoe je dat niet hebt kunnen zien. Hierbij wil ik beide dan ook bedanken voor hun input. Daarnaast wil ik ook Jos Geluk, mijn begeleider bij PwC, bedanken, vooral voor de input aan het begin van het schrijven van deze scriptie.

En uiteraard scriptiebegeleider Jan van den Berg voor zijn bruikbare kritiek, suggesties en voorspellingen (de doelstelling blijf je tot het laatst aan veranderen, let maar op). En niet te vergeten:

- Mijn ouders voor de ondersteuning op allerlei fronten
- Iedereen die maar bleef vragen: *Wanneer ben je nou klaar ?*
- Linus Torvalds, voor het maken van Linux
- Leslie Lamport, voor het maken van LaTeX (Bye, bye, Microsoft Word)
- Ferdinand Porsche, voor het ontwerpen van mijn auto's, passie nummer 1
- The Lego Group, voor de andere (geldverslindende) passie.

Edward van der Jagt
Mei 2001

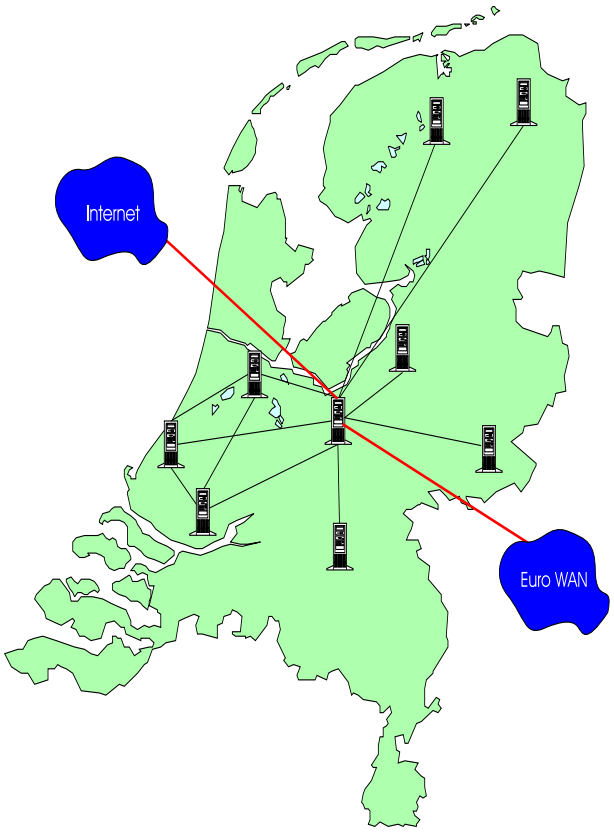
II

Inhoudsopgave

Lijst van figuren

Hoofdstuk 1

Inleiding



Reeds vele jaren zijn er organisaties die groeien, zowel door acquisities als door zelf uit te breiden. Door deze groei bevinden deze organisaties zich vaak niet meer op één lokatie, maar zijn dan verspreid over de regio, het land, of zelfs de hele wereld. Om beter samen te kunnen werken wordt dan de beslissing genomen om de (computer)netwerken van die diverse vestigingen aan elkaar te koppelen door middel van een Wide Area Network (WAN). Om zo'n WAN tot stand te brengen wordt er gebruik gemaakt van zogenaamde backbone providers die de gehele (lange afstands-) infrastructuur bezitten en aanbieden. De kosten voor het gebruik van die backbones kunnen nogal oplopen.

De meeste organisaties krijgen voor hun WAN verbindingen één grote telecomrekening waarop een vast bedrag staat of grofweg alleen tarief maal eenheid (tijd of hoeveelheid) vermeld staan. Het product van tarief en eenheid zijn dan de gemaakte kosten. Bij het lokale netwerk (LAN) is de kostenstructuur iets anders. Daar bestaan de kosten voornamelijk uit de eenmalige aanschaf van en de afschrijving op de hardware, in combinatie met de (personeels)kosten voor het beheer.
Het op deze manier specificeren van kosten voor het WAN is om een aantal redenen een onwenselijke situatie:

1. Uit een managementperspectief is dit een ongewenste situatie, omdat men **a)** informatie wil hebben om beslissingen te kunnen nemen en **b)** men graag kosten toerekend aan de eenheid waar die gemaakt worden. Indien alles als één grote overhead kostenpost wordt neergezet, kan er niet geanalyseerd en gestuurd worden. Dit is onacceptabel voor het management.

2. Afdelingen of gebruikers zullen het niet prettig vinden dat hun kosten omhoog gestuwd worden door een andere afdeling die de WAN verbinding veel meer gebruikt (of misbruikt).

3. Een service provider (ISP, NSP, ASP, zie Bijlage C) wil en moet exact weten wat een klant verbruikt, zodat deze hierop kan afrekenen.

4. Uit het oogpunt van bandbreedtebeheer is dit een onwenselijke situatie. Vaak wordt er bandbreedte verbruikt door niet-kritieke of niet-essentiële applicaties waardoor belangrijke applicaties (bijvoorbeeld ERP) niet goed of niet vlot genoeg werken. Essentiële activiteiten kunnen hier dus onder lijden.

Organisaties hebben dus baat bij een bruikbare methode om grip op bandbreedte gebruik en bijbehorende kosten te kunnen krijgen. In dit document wordt uitgelegd welke stappen moeten worden genomen om deze kosten te kunnen specificeren en te beheersen. Daarbij komen diverse problemen aan het licht waarvoor mogelijke oplossingen worden voorgesteld. Een aantal van deze problemen kunnen worden opgelost door een goede inrichting van het LAN waarop gemeten moet worden.

1.1 Doelstelling

In deze scriptie wordt onderzoek gedaan naar wat nodig is om de benodigde informatie te verkrijgen waarmee het management kan sturen op het WAN gebruik teneinde het gebruik en de kosten te beheersen. Deze *stuurinformatie* bestaat uit gespecificeerde gebruiksgegevens, geaggeregeerd en geanalyseerd naar de componenten zoals het management dat wenst.

Het doel van het scriptie onderzoek is het samenstellen van een methodiek ter beheersing van gebruik en kosten van Wide Area Network verbindingen.

Aan het einde van het onderzoek resulteert dit in een *stappenplan* en *methodes* om de diverse problemen aan te pakken.

1.2 Methodologie

De gehanteerde aanpak voor dit onderzoek wijkt voor wat betreft volgorde sterk af van de volgorde waarin het een en ander in deze scriptie wordt gepresenteerd.

Bij het bepalen van het onderwerp voor dit onderzoek was de eerste vraag of het technisch gezien wel mogelijk is om de gewenste informatie te verkrijgen. De eerste stap was dan ook om te onderzoeken wat de technische (on)mogelijkheden hierbij zijn. Ook kon veel uit de eigen jarenlange ervaring met netwerkbeheer en -onderhoud worden geput. Daarnaast is een stuk state-of-the-art onderzoek gedaan om uit te vinden welke technieken en producten er bestaan om gebruiksmeting of beheer van WAN verbindingen mogelijk maken of ondersteunen. Hierbij is, zoals in deze tijden vaker gebruikelijk is, het Internet als grootste informatiebron gebruikt. Standaarden van het IETF (de zogenaamde Request For Comments van de Internet Engineering Task Force) en produkt specificaties van zowel hardware als software blijken hierbij de grootste informatiebron.

Met de eerste invulling van het technische gedeelte kwamen er reeds diverse vragen tevoorschijn waarvan de invulling op management niveau dient te gebeuren. Met deze opgedane kennis werd het nu mogelijk om een stappenplan op te stellen. Dit stappenplan bleek dermate helder dat hier de indeling en invulling van de hele scriptie van kon worden afgeleid.

Het management gedeelte van dit onderzoek kreeg zijn invulling in beginsel door een stuk theorie (Starreveld [20]), aangevuld en aangepast na gesprekken met een aantal managers op dit gebied. Deze managers gaven het gewenste inzicht in de wensen van het management.

Als laatste gedeelte van het onderzoek kreeg het zijn toetsing aan de praktijk door middel van een aantal praktijksituaties. Deze praktijkgedeeltes dienden eveneens voor het ontdekken en invullen van eventuele hiaten in het verhaal.

1.3 Opbouw scriptie

In deze scriptie wordt uitgegaan van een lezer die weinig weet van de onderliggende technieken en protocollen die in netwerken worden gebruikt. Na een introductie in de materie en de techniek wordt verder reeds in grove lijnen het stappenplan gevolgd:

Hoofdstuk 2 gaat dieper in op de doelstelling en de te verwachten problemen. Tevens wordt hier een stappenplan geïntroduceerd.

Hoofdstuk 3 legt een aantal technische termen en begrippen uit die verderop in het onderzoek worden gebruikt.

Hoofdstuk 4 bekijkt het identificeren van de informatie en gegevensbehoefte.

Hoofdstuk 5 behandelt de stappen die nodig zijn om de gegevens uit het netwerk te verkrijgen.

Hoofdstuk 6 gaat in op het verwerken van de verkregen gegevens tot stuurinformatie.

Hoofdstuk 7 Geeft een aantal situaties uit de praktijk weer.

Hoofdstuk 2

Stuurinformatie

2.1 De noodzaak

Zoals in de inleiding al is geschetst heeft het management voor het uitvoeren van zijn taken stuurinformatie nodig. Deze stuurinformatie bestaat uit gegevens over de activiteiten, resultaten en kosten die een afdeling, unit, of welke bedrijfseenheid dan ook maakt. Op basis van deze gegevens kunnen er beslissingen worden genomen en uitspraken worden gedaan over het al dan niet efficiënt werken van die eenheid.

Zolang een kostencomponent een directe relatie heeft met de activiteiten van die eenheid valt hier op te sturen. Wanneer echter deze directe relatie er niet is, wordt het erg lastig, zo niet onmogelijk om er op te sturen. Kosten die bijvoorbeeld als overhead worden geboekt zijn niet gewenst. Kosten die via een vaste verdeelsleutel worden toegerekend aan een eenheid zijn dus lastig om op te sturen.

Bij kosten die geen directe relatie hebben zou het wenselijk zijn om ze afhankelijk te maken van het daadwerkelijke gebruik. Dat is iets dat bijvoorbeeld wordt gedaan bij een boekhoudkundige methode die Activity Based Costing heet. Bij die methode worden kosten toegerekend aan eenheden op basis van hun component (lees: daadwerkelijk gebruik) in het geheel. Dit is een zeer gunstige positie uit het oogpunt van stuurinformatie.

In veel (vooral grote) organisaties wordt er gebruik gemaakt van een WAN om netwerken met elkaar te koppelen. De kosten voor die koppelingen bestaan uit:

- Kosten voor apparatuur; een eenmalige component (aanschaf) of een vaste component (bijvoorbeeld bij lease)
- Kosten voor de verbinding; de kabel en de toegang, beiden een vaste component
- Kosten voor het gebruik; welke bij sommige technologiëen een variabele component is (afhankelijk van daadwerkelijk gebruik) en bij anderen een vaste component is. De trend is naar een vaste component (zie sectie 3.5).

De kosten voor zo'n WAN zijn dus veelal vast, zonder dat er direct componenten uit kunnen worden gehaald. Het zou beter zijn als deze opgesplitst konden worden naar eenheden binnen de organisatie.

2.2 Het verkrijgen van stuurinformatie

Wat kan het management doen om de gewenste stuurinformatie te verkrijgen ?

Als eerste dienen zij te bepalen in welke vorm en uit welke gegevens de stuurinformatie dient te bestaan. Nadat vorm en inhoud is bepaald, kan er worden gezocht naar methoden om de gewenste gegevens te verkrijgen.

Een eerste gedachte kan dan zijn om de gewenste gegevens via de provider van de verbinding te verkrijgen. Immers, deze levert de verbinding, heeft allerlei apparatuur op en om die verbinding staan en zal zelf ook al het een en ander meten in verband met zaken als Service Level Agreements (SLA, zie pagina 70) en troubleshooting.

In de meeste gevallen zal de provider deze gegevens niet willen of kunnen leveren, omdat het hem misschien te veel kost om de gegevens boven water te halen en in een bruikbare vorm te gieten. Een andere reden kan zijn dat zijn tariefstelling onafhankelijk is van het gebruik, en de provider daarom geen gebruiksgegevens verzamelt of wil leveren.

Zelfs al zou de provider de gegevens leveren, dan nog zijn die gegevens meestal niet genoeg. De gegevens zullen bijna altijd te weinig detail leveren, en vaak ook niet nauwkeurig genoeg zijn voor het eigen billing systeem. De diverse redenen hiervoor worden uiteen gezet op sectie 5.6.4.

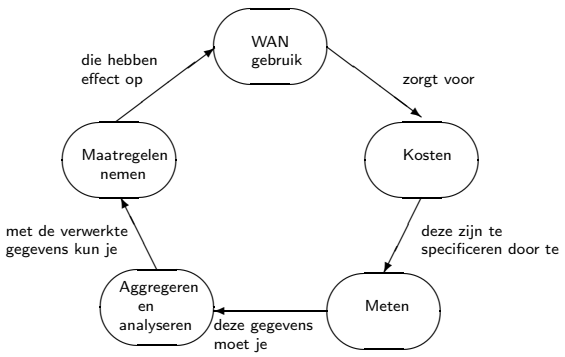
De enige manier die dan nog resteert om de gewenste gegevens te krijgen is zelf te gaan meten. Er kan dan gelijk zelf worden geregeld hoe gedetailleerd men de gegevens wil hebben, in welke vorm en wanneer. Dit wordt uitgewerkt in Hoofdstuk 4 en Hoofdstuk 5.

Wanneer de provider toch iets van gegevens kan leveren, dan is het meteen mogelijk om deze gegevens naast de zelf gemeten waarden te leggen, en met elkaar te vergelijken. Hiermee kan men zich enerzijds wapenen tegen foutieve facturen van de provider, en anderzijds het eigen meetproces controleren en calibreren.

Wanneer de diverse meetgegevens beschikbaar komen, zullen deze nog moeten worden bewerkt (geaggregeerd) totdat ze de gewenste vorm hebben. Wanneer er met de resulterende gegevens voor meer dan alleen toerekenen van kosten zal worden gebruikt, dienen ze nog verder te worden geanalyseerd. Op die manier kunnen er trends of gebruikspatronen uit worden gehaald.

2.3 Gebruik van stuurinformatie

Wanneer het management gedetailleerde kosten informatie heeft, dan kan het besluitvormingsproces beter worden doorlopen. Er kan bijvoorbeeld worden gekeken of de gemaakte kosten wel overeenkomen met de verwachte kosten behorende bij de werkzaamheden. Andere beslissingen kunnen zijn omtrent het al dan niet nodig zijn om bandbreedte bij te kopen of af te stoten, het verplaatsen van werkzaamheden enzovoorts. Nadat er maatregelen zijn getroffen kan er uiteraard weer worden gemeten en gedetailleerde kosten informatie worden gegenereerd. Dit alles is een continue proces van meten, aggregeren, analyseren, en maatregelen nemen.
De volgende figuur geeft dit proces schematisch weer:



Figuur 2.1: Het continue proces van informatie verzamelen, verwerken en gebruiken.

2.4 Praktijkvoorbeelden van kosten

Om een globale indruk te krijgen van wat een aansluiting op en het gebruik van een WAN kost, zijn een aantal offertes bekeken en is een een aantal netwerkbeheerders de vraag voorgelegd wat de WAN verbindingen zelf kosten.

Accountants PwC: Men wist slechts één voorbeeld van kosten van WAN koppelingen te noemen, en dit zijn de rekeningen van UUnet voor de aansluiting op Internet. Hierop staat alleen vermeld hoeveel tijd er verbinding is geweest. Naast het vaste maandtarief dient er bij UUnet ook per minuut betaald te worden. Naast deze kosten dient er ook nog te worden betaald aan de provider van de verbinding naar UUnet, meestal de KPN, wat normaal gesproken ook per tijdseenheid wordt afgerekend.
Op het gebied van LAN zijn er volgens de accountants geen kosten als zodanig. Die worden volgens hun alleen gemaakt bij de aanschaf van apparatuur.

Offerte transportonderneming (1999): Bij een grote transportonderneming bestond de wens om wereldwijd alle kantoren met elkaar verbinden. De oplossing die hiervoor werd aangedragen bestond uit een vast bedrag per aansluiting voor het gebruik van de backbone (ongeacht hoeveelheid gebruik) van de provider, en een uurtarief voor het inbellen op die backbone.
De bedragen waar het hier dan om gaat zijn gemiddeld zo'n 120 duizend gulden per jaar voor een grote vestiging en zo'n 45 duizend gulden voor een kleine vestiging aan exploitatie kosten. Totaal komt dit dan op een 5,5 miljoen gulden, op een omzet van net onder de 2 miljard gulden (is dus 0,0275%)

PwC - GTS: De Nederlandse kantoren van PwC zijn met elkaar verbonden met een diversiteit aan soorten verbindingen. De kantoren in Amsterdam hadden een aansluiting op CityRing van KPN van 3x10Mbps, tezamen kosten deze zo'n 18000 gulden per maand. Inmiddels liggen er eigen glasvezel kabels (met aanzienlijk meer capaciteit dan 10Mbps) tussen de kantoren welke iets van 1 miljoen gulden hebben gekost. De maandelijkse kosten zitten hier in de vergoeding voor de stukken grond waar de kabel in ligt.
Op een aantal andere locaties is gekozen voor straalverbindingen, die aan apparatuur ergens rond de 100 duizend gulden hebben gekost voor een 10-100Mbps verbinding (afhankelijk van de afstand). De enige terugkomende kosten hier zijn de kosten voor de zendmachtiging.

TU Delft: De TU Delft heeft capaciteit te over. De gebouwen op het campus terrein zelf worden door glasvezel kabels verbonden met een snelheid van 155 of 622Mbps. De verbinding naar Surfnet is momenteel een ATM verbinding van 155Mbps welke volgend jaar wordt vervangen door een Gigabit verbinding. De kosten van deze verbinding liggen in de orde van grote van 1 tot 2 miljoen gulden.
Het probleem bij deze verbinding is echter omgekeerd, namelijk dat men de verbinding niet vol krijgt. De beste belasting ooit was slechts 80Mb.
De koppelingen tussen de gebouwen van de TU zijn glasvezel verbindingen. Deze liggen echter allemaal op eigen terrein en zijn ook eigendom van de TU. Dit zijn eenmalige kosten geweest en in die zin niet van belang voor dit onderzoek.
De benodigde capaciteit van de verbindingen naar de diverse faculteiten wordt bepaald door een vaste rekensom met als enige variabele het aantal studenten.

Surfnet: Een aansluiting bij Surfnet bestaat uit kosten voor de fysieke aansluiting en kosten per aangesloten gebruiker. Een aansluiting van 2 Mbit kost bijvoorbeeld 6400 gulden per maand. Hierbij komen dan nog de kosten van de fysieke verbinding naar Surfnet toe. Indien hier dan bijvoorbeeld 100 gebruikers achter zouden zitten, dan kost dat 5550 per maand (bovenop de 6400). Totaal zou dit dan uitkomen op 143400 gulden per jaar, plus de kosten van de fysieke verbinding.

KPN: De prijs van een Digistream verbinding van 2 Mbit tussen een tweetal punten (als voorbeeld 2 locaties ca. 11km van elkaar verwijderd, omgeving Rotterdam) bestaat uit een eenmalige component van 10000 gulden en een maandelijkse component van 3890 gulden De tarieven van KPN zijn gebaseerd op afstand tussen de aansluitpunten, en niet van het verbruik; de connectie is een dedicated connectie.
De prijs van een Frame Relay (FR) aansluiting, bestaat uit diverse componenten, maar is eigenlijk alleen afhankelijk van de gewenste minimum snelheid. Een nota voor een FR aansluiting bestaat uit een vast bedrag per aansluiting en een vast bedrag per PVC.

2.5 Samenvatting

In dit hoofdstuk is er aangegeven waarom er stuurinformatie nodig is, en in grote lijnen uitgezet wat er moet worden gedaan om deze stuurinformatie te verkrijgen. Uit deze grote lijnen volgt het stappenplan, dat er als volgt uit ziet:

1. **Identificeren van de gewenste informatie**
   Het management zal moeten bepalen welke informatie zij willen hebben, welke vorm en/of eenheid, en gedurende welke periode. Bijvoorbeeld per eenheid het verbruik en het bijbehorende kostenplaatje, gespecificeerd per week.

2. **Bepalen van de benodigde meetgegevens**
   Wanneer bekend is welke eindgegevens het management wil, volgt hieruit welke meetgegevens moeten worden verzameld.

3. **Bepalen waar deze gegevens vandaan moeten worden gehaald**
   Nu bekend is welke meetgevens nodig zijn, moet worden bepaald waar deze vandaan moeten worden gehaald. Dat wil zeggen, van welke locaties in het netwerk.

4. **Vaststellen hoe deze gegevens moeten worden verzameld**
   Per locatie moet er worden bekeken hoe de gewenste gegevens kunnen worden verzameld. Dus van welke tools (software, hardware) en/of faciliteiten (bijv. logs) moet gebruik worden gemaakt.

5. **Verzamelen van de gegevens**
   Met de tools en/of faciliteiten kan men nu aan de slag gaan, en meetgegevens verzamelen.

6. **Gegevensaggregatie**
   Nu de ruwe meetgegevens beschikbaar zijn dienen ze nog in de door het management gewenste vorm te worden gezet.

7. **Gegevensanalyse**
   Op de ruwe of geagreggeerde gegevens kunnen er vervolgens diverse analyses op worden losgelaten.

8. **Conclusies en maatregelen**
   Uit de analyses kunnen bepaalde conclusies worden getrokken, die vervolgens kunnen worden gebruikt om maatregelen te nemen.

In het volgende hoofdstuk wordt er kort ingegaan op een aantal aspecten van netwerken en de technieken die alles laten functioneren. In de hoofdstukken die daarop volgen wordt het hierboven genoemde stappenplan uitgewerkt. Stappen 1 en 2 worden in Hoofdstuk 4 uitgewerkt, stappen 3, 4 en 5 in Hoofdstuk 5, en stappen 6 en 7 en 8 in Hoofdstuk 6.

Hoofdstuk 3

Communicatienetwerken

Om twee of meer machines met elkaar te laten communiceren is er een verbinding tussen die machines nodig. Alle verbindingen bij elkaar wordt een netwerk genoemd. Een verbinding alleen is niet genoeg; de machines moeten ook nog met elkaar kunnen communiceren.

Deze communicatie gebeurt met een veelvoud aan protocollen. Die protocollen werken in een soort hiërarchisch lagen model; elke laag gebruikt een protocol van een daaronder gelegen laag om een bepaalde functionaliteit te bieden. Die functionaliteit is dan meestal het transport van gegevens, aangevuld met een systeem dat transport problemen oplost.

In dit hoofdstuk wordt een korte uitleg gegeven over hoe dit alles met elkaar samenwerkt, en worden een aantal veelgebruikte protocollen toegelicht. Voor een uitgebreide uitleg verwijs ik de lezer naar [32]

3.1 Het lagen model

Om het geheel van protocollen en hun onderlinge relatie zichtbaar te maken en gestructureerd te kunnen aanpakken, is er enkele tientallen jaren geleden al een lagen model ontwikkeld waarin de relaties van de verschillende protocollen zichtbaar kan worden gemaakt. Dit lagen model, het ISO/OSI model, deelt netwerkcommunicatie in zeven lagen in waarvan de belangrijkste zijn:

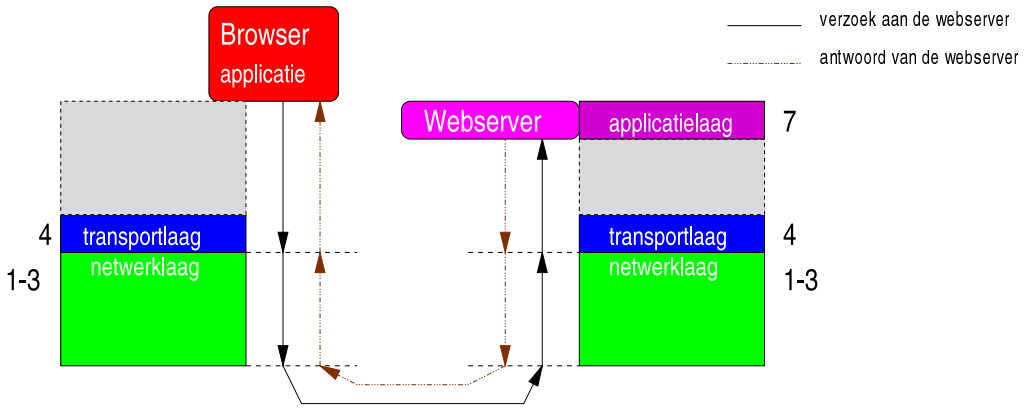
1-3 de netwerklaag: verzorgt de logische en fysieke adressering en routering van data

4 de transportlaag: verzorgt de foutafhandeling en kwaliteit van de verbindingen

7 de applicatielaag: het protocol voor de meeste eind applicaties (web, ftp, email, enz.).

Bovenop de applicatielaag werkt de uiteindelijke applicatie, zoals een webbrowser of een email programma. De webbrowser is dan de applicatie en de webserver implementeert de applicatielaag waarmee de applicatie communiceert. Zie ook Figuur 3.1 voor een grafische weergave van de onderlinge relaties.

De netwerklaag en de transportlaag zijn voor het meten interessant omdat daarin de adressering van de verzonden data wordt geregeld.
De applicatielaag is voor het meten eveneens interessant, omdat daar uit te halen is wat er nu eigenlijk wordt verstuurd aan data en voor welke toepassing dit is.



Figuur 3.1: Het lagen model. Verzoek van de browser aan de server en zijn respons.

3.2 Specifieke protocollen

3.2.1 Algemeen

Ruwweg bekeken bestaan er 3 typen netwerken:

LAN Local Area Network
Meestal het netwerk binnen één of zeer dicht bij elkaar gelegen gebouwen.

MAN Metropolitan Area Network
Een netwerk dat verspreid is over verschillende locaties binnen een grote stad. De koppelingen hiertussen bestaan uit bijvoorbeeld korte afstand radio verbindingen of speciaal hiervoor aangelegde of ingehuurde kabels. Een voorbeeld hiervan is CityRing van KPN Telecom. De bandbreedte is vaak hetzelfde als in een LAN.

WAN Wide Area Network
Een netwerk dat verspreid is over grote tot zeer grote afstanden. Denk hierbij aan verbindingen tussen steden, landen, of zelfs continenten. De verbindingen worden gemaakt door middel van diverse media, zoals satelliet of glasvezelkabels van een backbone provider.

De protocollen die over deze 3 typen netwerken worden gebruikt variëren ook. Dit komt door de verschillende omstandigheden en afstanden waarbij het netwerk wordt ingezet. Bij een LAN kan er worden aangenomen dat de tijd die nodig is om tussen twee eindpunten (machines) data te versturen behoorlijk klein is, doordat de maximale lengte kabel hiertussen meestal niet meer dan enkele kilometers lang is. Door deze korte tijd is het goed mogelijk om snel een robuuste fout detectie en -correctie uit te voeren. Bij een WAN is die tijd veel groter en wordt dat een heel stuk lastiger, en zijn er andere methodes nodig om een betrouwbare communicatie te kunnen garanderen.

3.2.2 IP, TCP en UDP

I(nternet) P(rotocol) Het IP protocol is een protocol van de netwerklaag (laag 3). Dit protocol wordt momenteel gebruikt als basis protocol voor de meeste Internet programmatuur. Maar het wordt ook steeds meer gebruikt als basis protocol op een LAN. Een machine die IP gebruikt wordt geacht een uniek nummer, een IP adres, te hebben. Zo'n nummer wordt, net als een regulier postadres, gebruikt voor het afleveren van een stukje data op een ander adres. Het verzend adres kan dan weer worden gebruikt om naar te antwoorden. Een IP adres bestaat uit vier getallen kleiner dan 256 welke meestal door een centrale instantie wordt toegewezen aan een machine. Het is mogelijk dat een machine

meer dan één IP adres gebruikt, en dat er meer machines zijn met hetzelfde adres (maar dat kan veel problemen geven).

T(ransmission) C(ontrol) P(rotocol) en U(ser) D(atagram) P(rotocol) Waar het IP protocol alleen zorgt voor adressering en routering, daar zorgen TCP en UDP ervoor dat de data daadwerkelijk in orde aankomt. TCP is een zogenaamd sessie-protocol. Dwz. dat TCP een sessie onderhoudt met de ontvanger; hierbij blijft er continu een verbinding tussen de twee machines. Zelfs als er geen data teversturen is blijft deze open. Dit is belangrijk om een continue stroom van data te kunnen garanderen. UDP werkt weer heel anders. De gedachte achter UDP is dat alle data moet aankomen binnen een bepaalde tijd (time-out). Gebeurt dit niet dan wordt de niet-aangekomen data opnieuw verstuurt.

Poorten Zodat er op één machine meer dan één dienst of applicatie kan werken bestaat er naast de adressering op machine niveau (het IP adres) ook een adressering op diensten niveau. Dit zijn de zogenaamde poortnummers. Deze adressering gebeurt in het TCP of UDP protocol. Deze adressering is een beetje te vergelijken met een flatgebouw waar je met de lift naar een bepaalde verdieping gaat (het IP adres), en op die verdieping bij een specifiek huisnummer binnengaat (het poortnummer) omdat alleen in dat huis een bepaalde dienst wordt verleend.

3.2.3 IPX en SPX

I(nternetwork) P(acket) (e)X(change) is een protocol dat tot het eind van vorige eeuw voornamelijk door Novell gebruikt werkt in netwerken met een Netware server. Voor wat betreft functionaliteit is het te vergelijken met het IP protocol.

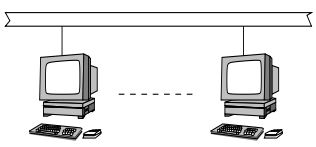
Een IPX adres bestaat uit het hardware adres van een machine, gecombineerd met het netwerk adres dat door een server wordt bepaald. Een IPX adres is dus in principe altijd uniek en behorend bij één specifieke machine.

S(tructured) P(acket) (e)X(change) vervult via IPX dezelfde rol als TCP en UDP bij IP.

3.3 Netwerk apparatuur

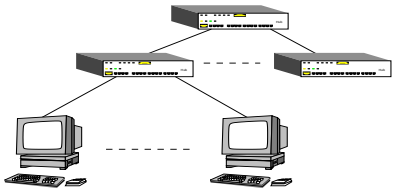
Wanneer er slechts 2 machines aan elkaar gekoppeld hoeven te worden, is een (directe) kabel tussen die twee voldoende. Zodra er echter meer aan elkaar gekoppeld worden is er een bepaalde combinatie van kabels en hulp apparatuur nodig.

In de meest eenvoudige situatie zitten alle machines verbonden aan dezelfde kabel. In die situatie gaat al het communicatieverkeer van alle machines over dezelfde kabel. Indien een machine wil gaan communiceren, zal hij eerst moeten wachten tot er niemand anders meer communiceert en het dus even "stil" is op de kabel. Dit alles is te vergelijken met de televisiekabel waarover Internet wordt gebruikt.
Deze manier van aansluiten wordt meestal aangeduid als een **bus-structuur**.



Figuur 3.2: Netwerk met Bus-structuur.

Een bus-structuur heeft echter een aantal belangrijke nadelen. Wanneer de bus op enig punt onderbroken wordt, kan geen van de machines die aan die bus aangesloten zijn nog communiceren. Dat is uiteraard een erg groot risico. Naast dat probleem speelt er ook nog een veiligheids aspect mee. Doordat de communicatie van iedere machine op zo'n bus kan worden afgeluisterd op elke andere machine op die bus, is beveiliging van gegevens erg lastig.



Figuur 3.3: Netwerk met hubs of switches.

In een betere situatie hebben alle machines hun eigen stukje kabel, en praten ze allemaal via een centraal punt. Dit centrale punt vangt de gesprekken van iedereen op en stuurt ze één voor één naar alle aangesloten kabels toe. Elke aangesloten machine is dus nog steeds in staat om de communicatie van andere aangesloten machines af te luisteren, maar machines kunnen wel gaan communiceren wanneer ze dat willen.

Zo'n centraal punt heet een **netwerkhub**. In een wat groter netwerk kunnen er hiervan meerdere staan (met iets van 2 tot 32 aansluitingen per hub), meestal hiërarchisch aan elkaar gekoppeld. Hubs werken op de datalinklaag (laag 2) en zijn daarmee onafhankelijk van de gebruikte netwerkprotocollen van de netwerklaag (laag 3) en hoger (zoals IP en IPX).

In een nog betere situatie worden er geen hubs meer gebruikt, maar **switches**. Switches zijn eigenlijk te beschouwen als intelligente hubs. Ze zijn intelligent in dat ze zelf onthouden welke machines er op welke aangesloten kabel zitten. Daardoor wordt de communicatie tussen machines alleen gerouteerd naar de juiste kabel, en worden de machines op de andere kabels niet lastig gevallen met communicatie die niet voor hun bedoeld is; ze kunnen hierdoor ook geen communicatie van andere machines meer afluisteren. Switches werken op de datalinklaag (laag 2) of op de netwerklaag (laag 3).

Wanneer er tussen twee stukken netwerk een verbinding met een ander soort medium zit, dan is er aan beide kanten hiervan een apparaat nodig dat dataverkeer van het ene naar het andere medium kan overzetten. Zo'n apparaat heet een **bridge**, en werkt op de datalinklaag (laag 2). Wanneer er meer van het apparaat wordt verwacht dan simpelweg overzetten, dan wordt er een aangepast apparaat ingezet. Dit is een **router** (vergelijk met de stap van hub naar switch), en deze werkt op de netwerklaag (laag 3). De intelligentie zit hem dan meestal in het opbouwen en afbreken van de verbinding wanneer dit nodig is, en wederom in een stuk routeer techniek. Een bridge of router heeft vaak maar twee aansluitingen, één voor elk medium.

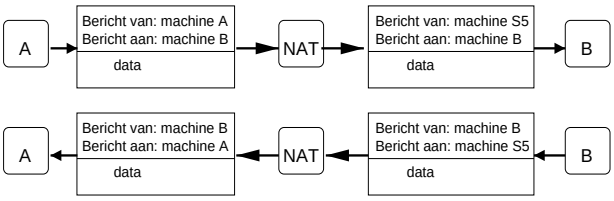
3.4 Servers en diensten

In een netwerk bevinden zich normaal gesproken diverse machines die diensten aanbieden aan gebruikers of andere machines. Hierbij dient te worden gedacht aan bestandsopslag, printen, email ophalen of versturen, enzovoorts. Een aantal diensten kunnen problemen geven bij het verwerken van de meetgegevens. Hieronder zullen deze diensten kort worden uitgelegd. De oorzaak van de problemen zullen kort worden aangegeven. Een diepere behandeling van deze problemen gebeurt in sectie 5.6.4.

3.4.1 Network Address Translation / Masquerading

Het is soms wenselijk of noodzakelijk om een machine aan het ene einde van een connectie te laten denken dat hij communiceert met een andere machine dan dat hij daadwerkelijk mee communiceert. Als bijvoorbeeld machine A met machine C wil communiceren, en deze communicatie verloopt via machine B, dan kan N(etwork) A(ddress) T(ranslation) er voor zorgen dat C denkt dat hij met B aan het communiceren is. Machine B verandert in de berichten van A naar C het afzender adres, en zorgt dat de antwoorden van C weer bij A terecht komen. Zie ook Figuur 3.4.

Wanneer bijvoorbeeld een LAN aan het Internet gekoppeld wordt, dan zou eigenlijk iedere machine die met het Internet wil communiceren een uniek IP adres moeten hebben. Voor IP adressen op het Internet moet er echter betaald worden, zodat het al snel een hele dure zaak kan worden bij een groot aantal lokale machines. Dit probleem is op te lossen



Figuur 3.4: Werking van NAT.

door NAT te gebruiken, waardoor er in het gunstige geval zelfs maar één IP adres nodig is voor een heel LAN, en wel voor de machine die alle communicatie naar het Internet routeert.

Naast het kostenaspect kan hier ook een beveiligingsaspect de reden zijn voor het gebruiken van NAT. In bovenstaand voorbeeld zal machine C nooit van het bestaan van machine A af weten. Hij communiceert immers alleen met machine B.

Het probleem voor het meten is dat netwerkverkeer kan worden gezien, dat als bron of doel adres machine B heeft, terwijl dit eigenlijk machine A moet zijn.

3.4.2 Proxies

Proxies doen, heel grof bekeken, eigenlijk ook een vorm van Network Adress Translation zij het in een iets andere vorm. Een proxy is te beschouwen als een tussenpersoon. In plaats van een verzoek rechtstreeks aan een eindbestemming te doen, wordt het verzoek aan de proxy gegeven die vervolgens zelf het verzoek aan de eindbestemming doet en het antwoord terug geeft aan degene die het verzoek had gedaan. De redenen om dit op die manier te doen kunnen er diversen zijn; uit beveiligingsredenen, uit efficiëntie redenen (zie verderop), uit beheersredenen, enzovoorts.

Dit is te vergelijken met bellen naar het buitenland via de bedrijfscentrale. Vaak is het niet toegestaan om van een willekeurig telefoontoestel naar het buitenland te bellen; dit moet dan worden aangevraagd bij de telefoniste die vervolgens het nummer draait en het gesprek doorschakelt.

Efficiëntie redenen kunnen ook leiden tot het inzetten van een proxy. In zo'n geval spreken we van een **caching proxy** (bufferende proxy). Doordat alle verzoeken voor bepaalde zaken van meerdere machines of gebruikers via één punt lopen (de proxy) bestaat de mogelijkheid om antwoorden te gaan bufferen. Bij het eerste verzoek wordt het antwoord in de buffer gezet, zodat bij volgende identieke verzoeken het antwoord uit de buffer kan worden gegeven in plaats van dat deze opnieuw moet worden opgevraagd. Aan de interne kant (het LAN) van de proxy heeft dit een snelheidswinst tot gevolg. Aan de externe kant (het WAN) kan dit een kostenvoordeel opleveren, doordat er minder communicatie nodig is.

Het probleem voor het meten is hier zowel het adresserings probleem (zoals bij NAT) als het probleem dat één gebruiker zorgt voor WAN verkeer terwijl andere gebruikers gebruik kunnen maken van het gebufferde antwoord en zelf geen WAN verkeer meer veroorzaken. Indirect hebben zij dit natuurlijk wel gedaan.

3.4.3 DHCP / BOOTP

Om diverse redenen kan het wenselijk of noodzakelijk zijn dat machines niet vast geconfigureerd worden voor een bepaald IP adres. Om deze machines toch te voorzien van een IP adres bestaan er een aantal protocollen. De bekendste hiervan zijn BOOTP en DHCP, waarvan de laatste zal worden besproken.
DHCP staat voor Dynamic Host Configuration Protocol. Dit protocol werkt als volgt:

1. Een machine (Host) wil gaan communiceren maar heeft nog geen IP adres.

2. Hij stuurt hiervoor een algemene aanvraag het netwerk op.

3. Een DHCP server vangt deze aanvraag op en kijkt of hij een configuratie heeft voor de betreffende machine

4. Indien deze server een configuratie heeft, stuurt hij alle configuratie gegevens terug naar de aanvrager.

5. De aanvrager ontvangt de configuratie en stelt zichzelf hiermee in. Hij heeft nu een IP adres en mogelijk wat extra gegevens, en kan nu gaan communiceren.

Dit klinkt allemaal erg eenvoudig, maar er zijn wel een aantal valkuilen. Op de DHCP server is het mogelijk om een vast IP adres te reserveren voor een specifieke machine. Het is echter ook mogelijk om een poule van adressen te maken. Een aanvrager krijgt dan een vrij adres uit deze poule toegewezen. Het is dus zeer goed mogelijk dat één machine elke dag (of zelfs vaker) een ander adres toegewezen krijgt.

Een ander punt dat hier mee te maken heeft, is dat een toegewezen adres slechts tijdelijk wordt toegekend; een soort tijdelijk huurcontract. Na een bepaalde tijd (meestal enkele uren) moet de machine een aanvraag doen om het huurcontract te verlengen. Het kan hierdoor dus voorkomen dat een machine iedere keer een ander adres krijgt. Overigens is het ook zo dat een machine tegen de server kan vertellen dat hij het toegwezen IP adres niet meer nodig heeft.

Gelukkig hebben de meeste DHCP servers een logging functionaliteit waarin staat welke machine op welk tijdstip welk IP adres heeft gekregen, zijn huurcontract heeft verlengd en wanneer deze het IP adres heeft vrijgegeven.

Problematisch voor het meten is dat hier dus niet zomaar een gebruiker aan één IP adres gekoppeld kan worden voor de verwerking van de meetgegevens.

3.4.4 Firewall

Een **firewall** werkt als een soort politie agent op het netwerkverkeer, normaal gesproken op de enige verbindingen tussen twee netwerken. Hij laat dan alleen specifiek verkeer door, en al het andere blokkeert hij.

Wanneer de firewall tussen het laatste meetpunt en de router is geplaatst, kan het dus gebeuren dat een stuk netwerkverkeer wordt geteld maar nooit het WAN op komt. Op dat ogenblik kloppen de meetgegevens niet meer.

3.4.5 Authenticatie / autorisatie en logging

In de hiervoor beschreven gedeelten zijn een aantal diensten besproken die allemaal voornamelijk met machines te maken hebben. Achter de meeste machines (werkstations) zullen gebruikers zitten die deze machines bedienen. Indien een gebruiker diensten in het netwerk wil gebruiken zal hij zich in het algemeen eerst dienen aan te melden op het netwerk. Afhankelijk van de gebruikersnaam en soms zelfs zijn locatie verkijgt de gebruiker op deze manier rechten en toegang tot bepaalde diensten. Bij de meeste netwerken en servers zal dit aanmelden in een logbestand worden bijgehouden. Achteraf kan er dan in dat log worden bekeken wanneer en op welke machine een gebruiker zich heeft aangemeld en tevens wanneer deze zich heeft afgemeld.

Het is mogelijk dat gebruikers geen vaste werkplek of werkstation hebben, en dus iedere dag (of vaker) zich op een ander werkstation op het netwerk zullen aanmelden. Bij het verwerken van de meetgegevens zal dus altijd rekening moeten worden gehouden wanneer en waar een gebruiker zich heeft aangemeld, en eventueel heeft afgemeld. Gebeurt dit niet dan kan er netwerkverkeer aan de verkeerde gebruikers toe worden gerekend.

3.5 Kostenstructuur van Wide Area Networks

Waar bij telefonie de kosten steeds meer naar betalen naar daadwerkelijk gebruik gaan (per seconden, in plaats van per tien seconden of zelfs per minuut), ontwikkelt de kostenstructuur voor dataverkeer over langere afstanden zich steeds meer naar een vast bedrag onafhankelijk van het gebruik ervan. Dit heeft vermoedelijk alles te maken met het feit dat voor dataverkeer een complete lijn naar de backbone (permanent) wordt gereserveerd en dus niet meer wordt gedeeld met meerdere afnemers. Het is voor de netwerk provider (NSP, zie pagina 70) dan niet meer interessant hoeveel dataverkeer er nou over dat lijntje gaat; de lijn is toch al apart gezet voor de afnemer. Bovendien hoeft de NSP nu ook geen gedetailleerde gegevens bij te gaan houden en te factureren. Dit alles lijdt er toe dat de NSP slechts een vast bedrag in rekening brengt.

Het gebruik van een backbone, zoals met Frame Relay, neigt ook steeds meer naar een vast bedrag onafhankelijk van het gebruik. Dit heeft er waarschijnlijk alles mee te maken dat de provider van deze backbone alleen maar geïnteresseerd is om zoveel mogelijk aansluitingen op die backbone te realiseren.

Hoofdstuk 4

Identificatie van de informatie- en gegevensbehoefte

Voordat er daadwerkelijk iets kan worden gedaan moet er als eerste worden vastgesteld wat het management wil bereiken en de informatie die daar voor nodig is. Met het vaststellen van deze informatiebehoefte wordt er ook bepaald welke gegevensbehoefte er is. In dit hoofdstuk zullen deze behoeften worden toegelicht.

4.1 Identificeren van de informatiebehoefte

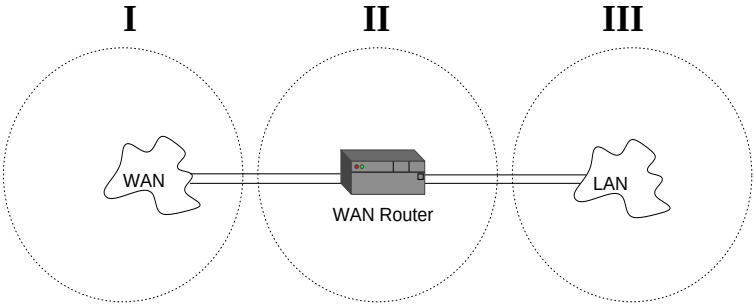
Als begin van het gehele proces moet er eerst duidelijkheid worden geschept in de behoefte van het management. Voor het gebied van Network Control, waar het in dit onderzoek om gaat, zijn de mogelijke informatiebehoeften van het management globaal op te delen in 3 categoriëen:

I - Prijs/prestatie verhouding

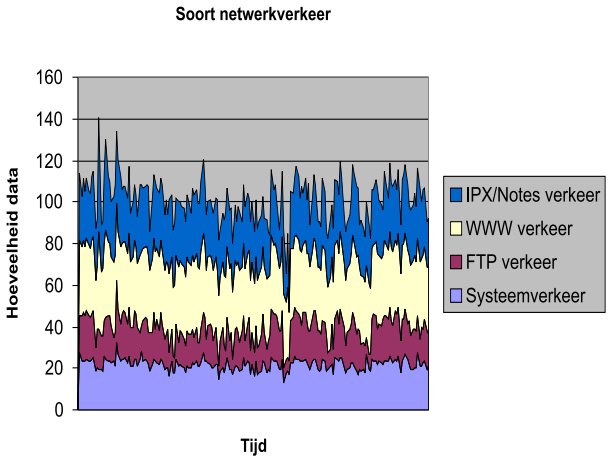
II - Functioneel gebruik van het netwerk

III - Wie maakt gebruik van het netwerk

Elk van deze categoriëen bevat diverse onderdelen waar het management in geïnteresseerd kan zijn. In Figuur 4.1 is zichtbaar gemaakt op welk gebied van het netwerk de drie categoriëen zich richten. Hieronder worden deze verder uitgewerkt tot concrete informatiebehoeften.



Figuur 4.1: De gegevenscategoriëen.



Figuur 4.2: Netwerkgebruik onderverdeeld per type netwerkverkeer.

4.1.1 Prijs/prestatieverhouding

Deze categorie richt zich voornamelijk op het gedeelte dat zich "buiten de deur" bevindt. De aandacht ligt daardoor hoofdzakelijk bij de zogenaamde Service Level Agreements (**SLA**).
Wanneer er een contract is afgesloten voor de levering van een WAN verbinding, staan hier normaal gesproken afspraken in met betrekking tot de prestaties van deze verbinding en wat hiervoor betaalt dient te worden. Bij prestaties moet worden gedacht aan bandbreedte- (hoeveelheid dataverkeer per tijdseenheid) en beschikbaarheidsgaranties. Afhankelijk van het soort verbinding kunnen er ook afspraken zijn over minimale - en maximale **bandbreedte** die wordt geleverd. Evenals afspraken over bijvoorbeeld de tijd dat, indien mogelijk, de minimale bandbreedte achter elkaar mag worden overschreden.

In deze categorie valt daarmee de vraag van het management of de geleverde prestaties overeenkomen met de afspraken. Ook de vraag of de prestaties voor een concurrerende prijs worden geleverd valt hieronder, maar die vraag valt buiten het onderzoeksgebied doordat de "prijs" van een WAN verbinding vaak is samengesteld uit diverse onderdelen (verbinding, apparatuur, support, boetes, enzovoorts). Het vergelijken van die prijzen is daardoor bijna een apart vak en zal daardoor niet hier worden behandeld.

4.1.2 Functioneel gebruik van het netwerk

Functioneel gebruik van het netwerk is hoofdzakelijk gericht op de vraag waar het netwerk (het WAN) voor wordt gebruikt en of dit in overeenstemming is met het toegestane gebruik. Om deze vraag te beantwoorden moet men weten welke applicaties in combinatie met het WAN worden gebruikt, en in welke mate ten opzichte van elkaar (zie de voorbeeldgrafiek in Figuur 4.2).

Met deze gegevens kan er bijvoorbeeld worden gecontroleerd of bepaalde applicaties elkaar niet "in de weg zitten". Een andere controle die hiermee kan worden uitgevoerd is of de verhoudingen tussen het gebruik van de verschillende applicaties overeenkomen met wat verwacht kan worden als gevolg van bepaalde werkzaamheden.

4.1.3 Wie gebruikt het netwerk

Het management kan ook een beter inzicht willen verkrijgen in het gebruik van het netwerk (het WAN), maar dan op een meer gedetailleerd niveau dan in de vorige categorie vereist is. Deze categorie richt zich dan ook hoofdzakelijk op het gedeelte dat zich "binnenshuis"

bevindt (het LAN).
De gedetailleerdere gegevens kunnen voor een grote diversiteit aan controles en maatregelen worden gebruikt. Hierbij kan worden gedacht aan

- het vaststellen van benodigde capaciteit
- het vaststellen van het gebruik of het controleren hiervan aan de werkzaamheden per gebruiker, of per bedrijfseenheid.
- het opdelen van de totale kosten naar gebruik (in verband met doorbelasten)
- het vaststellen van een budget per bedrijfseenheid

Er zijn uiteraard nog meer controles te verzinnen, echter deze zullen in mindere of meerdere maten een afgeleide zijn van die hierboven genoemde controles.

4.2 Identificeren van de gegevensbehoefte

Wanneer vast is gesteld wat de informatiebehoefte is, moet er worden bepaald welke gegevens er nodig zijn om aan deze informatiebehoefte te kunnen voldoen. De gegevensbehoeften van de drie categoriëen lopen enigszins in elkaar over, maar voor de duidelijkheid zullen ze hieronder toch per categorie worden beschreven.

4.2.1 Prijs/prestatieverhouding

Om op de vragen uit deze categorie een antwoord te kunnen geven zijn er uit het netwerk zelf bijna geen gegevens nodig. Om te kunnen controleren welke prestaties er worden geleverd kunnen er een aantal gegevens worden verzameld:

Aanwezigheid Om de aanwezigheid van de verbinding te controleren zal er iets continu moeten controleren of er communicatie via het WAN mogelijk is, en hierover rapporteren. Dit zal iets van ja/nee waarden opleveren.

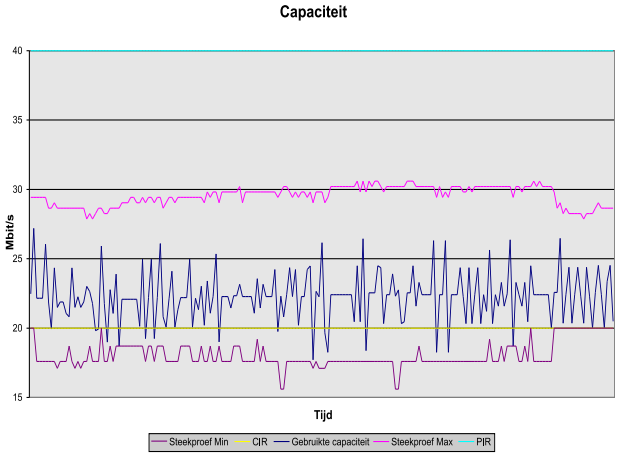
Bandbreedte Controle of de capaciteit aan een bepaald minimum of maximum voldoet kan wederom verkregen worden door iets dat hier periodiek tests voor uitvoert en hierover rapporteert. Ook het daadwerkelijke gebruik van de capaciteit kan worden opgenomen in de meetgegevens.

Als totaalbeeld voor deze categorie zal er dus iets van een grafiek zoals in Figuur 4.3 moeten komen met daarin de gebruikte bandbreedte en de resultaten van periodieke tests op aanwezigheid en bandbreedte.

4.2.2 Functioneel gebruik van het netwerk

Om aan de informatiebehoeften in deze categorie te kunnen voldoen moeten er gegevens worden verzameld over wat voor soort communicatie er via het WAN gebeurt. Hoe deze soorten kunnen worden onderscheiden is sterk afhankelijk van welke protocollen er worden gebruikt (zie ook sectie 3.2). Bij het bekendste protocol, IP, gebeurt dit in zijn eenvoudigste vorm door naar het zogenaamde poortnummer te kijken. Elke dienst gebruikt in principe een uniek poortnummer (uitzonderingen even buiten beschouwing latend) waaraan het soort communicatie kan worden herkend. Voor een zeer diepgaand onderzoek kan er ook nog eens *in* de verzonden data worden gekeken om daarmee eveneens vast te stellen om wat voor soort communicatie het gaat. Dit kan echter een zeer tijdrovende en ingewikkelde klus zijn zodat dit alleen bij zeer goede redenen moet worden gedaan.

Om aan de informatiebehoeften te kunnen voldoen moeten er dus statistieken per soort communicatie worden verzameld; bij het gebruik van IP zijn dat dan statistieken per poortnummer.



Piek gebruikte capaciteit:
30.7 Mbit/s
Gemiddelde gebruikte capaciteit:
23.5 Mbit/s (117.6% van contract)
Gemiddelde maximum capaciteit:
29.3 Mbit/s (73.3% van contract)

Figuur 4.3: Capaciteit van de verbinding.

4.2.3 Wie gebruikt het netwerk

Om informatie omtrent het gebruik van het netwerk te kunnen geven, moeten er gegevens worden verzameld over wie het WAN gebruikt en hoeveel die het WAN gebruiken. Om dit voor elkaar te krijgen dienen er statistieken te worden verzameld per machine of groep machines uit het lokale netwerk. In elk stukje communicatie dat over het netwerk gaat zijn bron- en bestemmingsadres, alsmede de hoeveelheid data vastgelegd. Ook het type communicatie staat bij de meest gebruikte protocollen daarin en kan ook worden vastgelegd om te kunnen bekijken waarvoor men het netwerk gebruikt. Deze statistieken kunnen “on the fly” worden gemaakt, dat wil zeggen dat iedere keer dat er netwerkverkeer langs het meetpunt komt de statistiek worden aangepast. Het is ook mogelijk om gewoon bij te houden wat er langs komt, en vervolgens pas achteraf de gegevens te aggregeren tot de hiervoor genoemde statistieken.

Deze gegevens moeten dus voor elke communicatie die via het WAN verloopt worden bewaard om uiteindelijk aan de informatiebehoefte te kunnen voldoen.

4.3 Overige attentiepunten

Twee andere belangrijke punten die moeten worden bepaald zijn over welk **tijdsinterval** er meetgegevens moeten worden verzameld en in **welke vorm** deze gepresenteerd moeten worden.

Het tijdsinterval is in de meeste gevallen cruciaal om zinvolle vergelijkingen of analyses te maken. Wanneer er bijvoorbeeld een controle op een ontvangen nota moet plaatsvinden heeft het uiteraard alleen zin om die vergelijking te maken als de meetgegevens uit dezelfde periode zijn. Ook voor bijvoorbeeld trendanalyse of capaciteitsbepaling moet goed worden bekeken in welk tijdsinterval de metingen plaasts dienen te vinden. Vlak voor een vakantieperiode of feestdag zal bijvoorbeeld meer gebruik worden gemaakt van het World Wide Web zodat zo’n periode niet gebruikt moet worden als maatstaf voor benodigde capaciteit.

Inzicht in de bedrijfsactiviteiten en eventuele trends zijn dus van groot belang bij het bepalen van het tijdsinterval en de analyse van de gegevens achteraf.

De vorm waarin de verwerkte en geanalyseerde gegevens uiteindelijk gepresenteerd moeten worden zijn van belang voor de verwerking en moeten dus vooraf worden bepaald. Is het niet nodig dat er in de te presenteren gegevens veel detail aanwezig is, dan kan

dit enorm uitmaken voor de verwerking van de gegevens; er kan dan mogelijk een flinke reductie in de hoeveelheid gegevens plaatsvinden.

Hoofdstuk 5

Gegevens verzamelen

In het vorige hoofdstuk is aangegeven dat de informatiebehoefte, en daarmee de gegevensbehoefte, aanzienlijk kan variëren. Wanneer er slechts grove informatie gewenst is, kan er wellicht worden volstaan met slechts één lokatie om gegevens te verzamelen. Die ene lokatie is dan meestal de router die netwerkdata naar het WAN stuurt. In de meeste gevallen zijn er echter meer gegevens gewenst dan er via één lokatie valt te verzamelen. Er dienen dan in het netwerk zelf gegevens te worden verzameld.

In die situaties zal er zeer gestructureerd moeten worden bepaald waar en hoe er gegevens moeten worden verzameld. Gebeurt dit niet, dan is het mogelijk dat de verzamelde gegevens een (totaal) verkeerd beeld geven van de werkelijkheid. Er zijn namelijk diverse omstandigheden waardoor meetgegevens verstoord kunnen worden of gewoonweg niet compleet zijn.

Om het risico van inaccurate gegevens te vermijden dient het gehele netwerk te worden doorgelicht. Hierbij dient dan te worden bekeken wat er aanwezig is aan machines en welke functies of diensten er worden uitgevoerd. Er zijn diensten die de meetgegevens kunnen verstoren; hier dient op de juiste wijze mee om te worden gegaan.

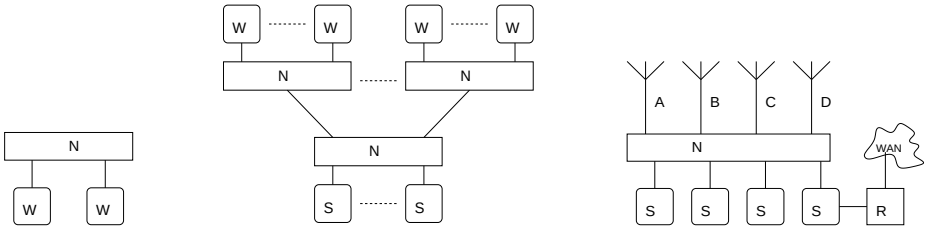
Het doorlichten van het netwerk kan het beste gebeuren door een overzicht te maken van hoe het netwerk in elkaar zit, welke diensten er aanwezig zijn, en welke routes bepaald applicatie gebruik over het netwerk maakt. Een overzicht is te maken door de volgende stappen te doorlopen:

1. Er moet een overzicht worden gemaakt van hoe alle apparatuur aan elkaar is gekoppeld. Dit vormt het overzicht van het fysieke netwerk (een soort wegenkaart).
2. In het **fysieke overzicht** kunnen nu controle punten worden aangewezen.
3. Bij elk aangesloten apparaat aan het netwerk moet worden bekeken welke diensten deze levert aan het netwerk.
4. Nu kan er een extra overzicht worden gemaakt met daarin de routes die bepaald applicatie gebruik maakt over het netwerk. Dit overzicht is het **logische overzicht**.
5. Uit het logische overzicht kunnen nu de lokaties worden bepaald waar gegevens verzameld kunnen worden.

Nadat deze stappen zijn doorlopen moet er bij elke gevonden lokatie worden bekeken hoe de gegevens kunnen worden verzameld. Dit kan voor bepaalde diensten of lokaties erg lastig zijn, zoals zal worden aangetoond in sectie 5.6.4

5.1 Fysiek overzicht

In een **fysiek overzicht** dient zichtbaar te worden gemaakt welke (fysieke) connecties er zijn tussen de machines in het netwerk. Dit houdt bijvoorbeeld in dat een netwerk met



(a) Simpel netwerk

(b) Klein vertakt netwerk

(c) Groter netwerk met WAN

W = werkstation, S = server, N = netwerkapparaat
R = router, A-D = takken met werkstations

Figuur 5.1: Voorbeelden van fysieke overzichten.

slechts 2 machines en een hub of switch het plaatje van Figuur 5.1(a) op levert; de beide machines zijn allebei rechtstreeks verbonden met de hub of switch en zijn daarom ook in de figuur verbonden hiermee. Bij een groter netwerk met meer machines wordt het overzicht op dezelfde manier gemaakt. In Figuur 5.1 staan een drietal voorbeelden van drie verschillende netwerken.

In veel situaties zal er al een overzicht aanwezig, zijn in de vorm van tekeningen die gebruikt zijn bij de aanleg van het netwerk. Zo'n overzicht kan worden gebruikt, maar uiteraard dient men er zich van zeker te stellen dat het overzicht nog compleet en correct is. Voor het gebruiksgemak van zo'n tekening kan het nuttig zijn om aan de hand hiervan een nieuw overzicht te maken door hem te simplificeren. In de (bouw)tekeningen staan namelijk normaal gesproken allerlei zaken getekend met betrekking tot de fysieke lokatie in het gebouw, muren, etc. Deze zaken zijn niet interessant voor het fysieke overzicht; daar is alleen van belang hoe alle machines met elkaar verbonden zijn en niet waar ze zich echt bevinden.

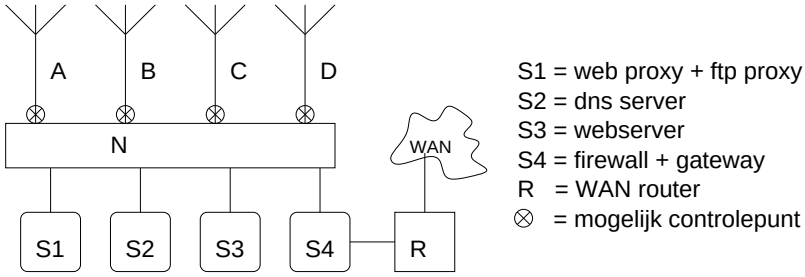
5.2 Controlepunten

Afhankelijk van de opbouw van het netwerk is het mogelijk dat zich in het fysieke overzicht een soort van "eilandjes" met machines bevinden, welke met slechts één verbinding aan de rest van het netwerk verbonden zijn. Via die verbinding verloopt dan alle communicatie met de rest van het netwerk en daarmee ook het WAN. Dit is daardoor een uitstekende plek om een meetlokatie in te richten, in ieder geval om controlegegevens te verzamelen. Zo'n lokatie is een **controlepunt**. Deze controlegegevens kunnen na het verwerken van de gewone gegevens worden gebruikt om het meetproces te kunnen controleren en calibreren.

Die verbinding zal in de meeste gevallen op in ieder geval één kant op een hub, router, of soortgelijk apparaat aangesloten zijn. Indien dit apparaat daar geschikt voor is, dan kan deze globale meetgegevens voor controlemogelijkheden verzamelen. Als het apparaat er niet geschikt voor is, dan kan hier ook een apart apparaat voor worden bijgeplaatst, een zogenaamd RMON device (Remote MONitoring, zie sectie 5.6.1 en Bijlage C).

5.3 Diensten

Een machine in het netwerk kan **diensten** verzorgen voor andere machines of gebruikers in het netwerk. Zo'n dienst kan van alles zijn, variëerend van klok (een tijdserver) tot



Figuur 5.2: Controle punten en diensten in het fysieke overzicht.

bestandsopslag (een fileserver) of WAN toegang (een router of gateway). Het is essentiëel om alle diensten in kaart te brengen, om later de juiste meetlokaties te kunnen bepalen. Communicatie met een dienst verloopt in grote lijnen als volgt:

1. de klant (een machine in het netwerk) wil dat een server iets voor hem uitvoert (de dienst),
2. de klant stuurt zijn verzoek naar de server,
3. de server ontvangt het verzoek en gaat hiermee aan de slag;
   Bij sommige diensten zal de server nu zelf tijdelijk klant worden, doordat hij zelf een verzoek gaat versturen aan een andere server,
4. de server stuurt zijn antwoord terug naar de klant.

Machines kunnen meer dan één dienst leveren, en een dienst kan door meer dan één machine geleverd worden. Dit is iets om goed in de gaten te houden zodat het overzicht compleet wordt.

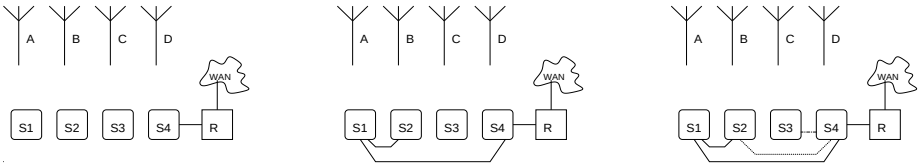
5.4 Logisch overzicht

Elke applicatie die op een machine wordt gebruikt genereert bepaalde **communicatiestromen** over het netwerk. Voor elke applicatie kunnen en zullen deze communicatiestromen anders zijn. Bijna elke applicatie veroorzaakt ook meer dan één communicatiestroom. Neem als voorbeeld een eenvoudige telnet sessie: deze veroorzaakt eerst DNS verkeer, en vervolgens het echte telnet verkeer. Die stromen kunnen ook zonder probleem gericht zijn op verschillende machines.

Omdat bepaalde communicatiestromen, dankzij de diensten in het netwerk waar zij op gericht zijn, weer andere communicatiestromen teweeg kunnen brengen, is het van belang om deze stromen goed in kaart te brengen. Deze stromen zijn immers hetgene wat moet worden gemeten. Het overzicht waar deze stromen op worden weergegeven wordt het **logische overzicht** van het netwerk genoemd.

Om dit overzicht te maken, kan er worden gestart met het fysieke overzicht. In dit overzicht zijn in de vorige sectie reeds alle diensten aangegeven. Per dienst moet er nu worden bekeken of deze zelf weer netwerkverkeer kan veroorzaken. Dit dient kritisch te worden bekeken, omdat veel diensten in eerste instantie zelfstandig lijken te zijn, maar dit toch niet blijken te zijn. Wanneer een dienst niet zelfstandig is, moet er worden bekeken waar deze dienst zelf weer netwerkverkeer naar toe kan genereren.

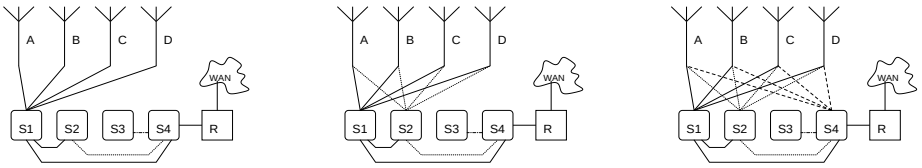
Neem als voorbeeld een **fileserver**. Wanneer deze bestanden van zijn lokale opslag levert is er niets aan de hand. Het is echter ook zeer goed mogelijk dat deze bestanden levert die niet lokaal op de fileserver staan opgeslagen, maar rechtstreeks of gerepliceerd van een andere server afkomen. Een verzoek van een gebruiker voor een bestand dat, voor



(a) Fysiek overzicht zonder netwerkapparatuur

(b) Verkeer vanuit dienst van S1

(c) Verkeer vanuit dienst van S2 en S3



(d) Verkeer uit het netwerk naar S1

(e) Verkeer uit het netwerk naar S2

(f) Verkeer uit het netwerk naar S4

Figuur 5.3: Opbouw van een logisch overzicht.

wat hem betreft, lokaal op een fileserver staat, kan dan een verzoek door die fileserver aan een andere server voor dat bestand als gevolg hebben. De fileserver dienst is dan niet zelfstandig, omdat deze zelf ook weer netwerkverkeer genereert.

Wanneer nu blijkt dat de betreffende dienst zelf geen direct of indirect verkeer naar het WAN genereerd, of vanuit het WAN te benaderen is, dan kan deze dienst ter vereenvoudiging van het overzicht hieruit weg worden gelaten.

Voorbeeld In Figuur 5.3 wordt het voorbeeld uit Figuur 5.1(c) uitgewerkt tot een logisch overzicht. Dit netwerk bestaat uit diverse groepen werkstations die verder geen diensten leveren in het netwerk (takken A, B, C en D). Daarnaast zijn er 4 servers aanwezig:

S1 - een proxy server voor web en ftp verkeer

S2 - een DNS server

S3 - een webserver, bereikbaar via S4. Vanaf het WAN of vanuit het netwerk.

S4 - een firewall en de gateway naar het WAN via router **R**

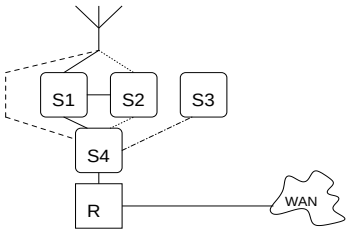
Nu kan stap voor stap het logische overzicht worden gebouwd:

a Als eerste verwijderen we nu uit Figuur 5.1(c) alle netwerkapparaten (in dit geval de switch of hub)

b De dienst die op S1 draait (de proxy) kan zelf zorgen voor DNS verkeer (naar S2) en voor web en ftp verkeer naar het WAN (via S4)

c De dienst die op S2 draait (de DNS) kan zelf zorgen voor DNS verkeer naar het WAN (via S4). S3 is afgezonderd, en kan alleen maar communiceren via S4, en dit zal webverkeer zijn.

d Vanuit het hele netwerk kan er gecommuniceerd worden met S1, dus alle takken met werkstations worden verbonden met S1



Figuur 5.4: Vereenvoudiging van een logisch overzicht.

e Vanuit het hele netwerk kan er gecommuniceerd worden met S2, dus alle takken met werkstations worden verbonden met S2

f Vanuit het hele netwerk kan er gecommuniceerd worden met S4, dus alle takken met werkstations worden verbonden met S4

Omdat er hier in de takken met werkstations verder geen onderscheid in hoeft te worden gemaakt, kunnen deze eventueel tot 1 tak worden vereenvoudigd. De tekening kan hierna opnieuw worden georganiseerd zodat deze er meer als een soort **stroomdiagram** uit gaat zien (zie Figuur 5.4).

Het logische overzicht staat nu in een vorm die handelbaar is, en waarin de **lokatie bepaling** kan plaatsvinden.

5.5 Lokatie bepaling

Om er zeker van te zijn dat de te verkrijgen gegevens uit het netwerk compleet zijn, moeten zowel de directe - als de indirecte stromen (die via een dienst lopen) naar het WAN worden afgevangen.

Ter verduidelijking gaan we hier weer verder met het hiervoor behandelde voorbeeld (Figuur 5.4). In dit voorbeeld zijn er 4 directe stromen te herkennen, namelijk direct vanaf de werkstations (de onderbroken lijn), en de drie stromen vanuit de 3 servers (S1, S2, S3). De **indirecte stromen** zijn die stromen die niet rechtstreeks vanaf een werkstation of server naar het WAN gaan (via S4). Hier zijn dat de stromen vanaf de werkstations naar S1 en S2, en de stroom tussen S1 en S2.

Alle stromen naar het WAN moeten via S4, dus op S4 kunnen het beste de directe stromen worden gemeten. Op S1 en S2 bevinden zich de diensten die de indirecte stromen genereren. Er zal daarom op een geschikte manier het gebruik van die diensten moeten worden gemeten, wat er meestal toe leidt dat op S1 en S2 een meetpunt wordt ingericht.

De manier waarop daadwerkelijk meetgegevens verkregen kunnen worden, wordt toegelicht in de volgende secties.

5.6 Hoe kunnen de gegevens worden verkregen

Nu bepaald is op welke lokaties en/of diensten er gegevens verzameld moeten worden, komt de vraag hoe deze gegevens dan echt kunnen worden verzameld. Door de grote verscheidenheid aan systemen die gebruikt kunnen worden is het onmogelijk om een complete lijst te geven met wat te doen bij welk systeem. We zullen het hier dus beperkt moeten houden tot algemene en een beperkt aantal specifieke systemen en manieren.

Er bestaan vele tools welke kunnen meten aan communicatiestromen. Deze tools kunnen een stuk programmatuur zijn welke op een server of werkstation werkt, of een speciaal apparaat welke in het netwerk moet worden geplaatst. Hoewel niet helemaal correct,

zullen we deze hier verder respectievelijk software tools en hardware tools noemen. Aan deze tools dienen een aantal strenge eisen te worden gesteld:

- De tools dienen minimaal de gegevens te kunnen leveren zoals deze in stap 2 van het stappenplan (zie sectie 4.2) geïdentificeerd zijn.
- Enige aggregatie bijvoorbeeld per machine is niet onwenselijk ter beperking van de hoeveelheid gegevens, maar niet meer. Aggregatie wordt pas later (zelf) gedaan in stap 6 van het stappenplan (zie sectie 6.1).
- Indien de tool de gegevens in een eigen formaat bewaart, dan moet er een export of conversie mogelijkheid zijn. Hiermee moeten de gegevens dan worden omgezet naar een formaat zoals deze bij de verwerking gebruikt kan worden.

Het grote nadeel bij bijna alle tools is dat ze slechts een beperkte hoeveelheid typen communicatiestromen kunnen onderscheiden en benoemen. De typen die ze niet kunnen onderscheiden of niet zijn benoemd worden onder één noemer overig geplaatst. Hierdoor kan het voorkomen dat bepaalde diensten niet kunnen worden onderscheiden, en daarmee de meetgegevens niet compleet of gedetailleerd genoeg zijn. Onbekende communicatiestromen kunnen daardoor niet ontdekt worden.

Sommige tools zijn hierin nog wel enigszins in aan te passen, maar dan moet wel elke voor de tool onbekende dienst apart geconfigureerd worden. Kan dit niet of worden er diensten over geslagen, dan zullen bijvoorbeeld stromen voor IP-poortnummer 8080 (meestal een proxy) niet apart worden vermeld. Indien er dus een dienst via die poort bereikbaar is, dan zal het gebruik daarvan nooit apart vermeld kunnen worden. Van de onderzochte tools zijn Net-acct (zie pagina 62) en de Packetshaper van Packeteer de enigen die hier geen last van hebben en alles apart vermelden, of de dienst nu benoemd is of niet.

Naast het meten aan communicatiestromen zal er vaak ook op een dienst moeten worden gemeten. Deze dienst zal hiervoor zijn activiteiten ergens moeten vastleggen, meestal gebeurt dit in een log bestand. Deze log bestanden hebben een geheel andere vorm dan de gegevens die de hierboven bedoelde tools afleveren. Ook hier zal iets op moeten worden gevonden.

5.6.1 Hardware tools

RMON Er is voor het monitoren een standaard gedefinieerd welke voornamelijk door de grotere netwerkapparaten wordt gebruikt om netwerken te bewaken. Deze standaard, RMON (Remote MONitoring) , en de uitbreidingen en varianten hierop (RMON2 en SMON) maakt het mogelijk om gegevens bij te houden betreffende verschillende grootheden, waaronder:

- netwerk statistieken
- machine statistieken per machine – Statistieken per ontdekte machine op het netwerk.
- machine statistieken voor paren van machines – Statistieken voor communicatie tussen tweetallen machines

Zoals uit de naam en de omschrijvingen hierboven blijkt is de standaard voornamelijk gericht op statistieken en niet op details. In de praktijk wordt RMON voornamelijk gebruikt om de “gezondheid” van een stuk netwerk te bewaken. Dat houdt in het in de gaten houden van belasting van de apparatuur (en daarmee ook het netwerk) en de eventuele boosdoeners (overbelasten van het netwerk) in het netwerk identificeren. Bovendien schrijft de standaard voor dat, wegens de beperkte opslagruimte in de meeste netwerkapparatuur, oude gegevens zonder waarschuwing verwijderd mogen worden.

Een ander algemeen bekend probleem is dat de apparatuur die de RMON standaard implementeert niet krachtig genoeg is om *én* veel te monitoren *én* zijn primaire functie uit te voeren. Hiermee wordt de inzetbaarheid beperkt.

De gegevens van een RMON probe (het apparaat of de software die de meetgegevens verzamelt) worden normaal gesproken verzameld in een groot netwerk management pakket zoals HP OpenView, CiscoWorks of 3Com Transcend Network Supervisor. Wanneer deze de gegevens vaak genoeg van de probe ophaalt wordt het risico van het verdwijnen van gegevens op de probe verminderd, maar nog niet opgelost.

Door deze beperkingen is de standaard daarmee wel geschikt om een ruwe analyse van het gebruik van het netwerk te kunnen doen (de tweede categorie uit Hoofdstuk 4), maar voor analyses die detail vereisen is RMON niet geschikt. Er zijn ook bijna geen diensten die gedetailleerde gebruiksgegevens leveren via RMON, waarmee deze standaard ook af valt voor het meten aan de indirecte stromen uit sectie 5.4 en sectie 5.5.

Andere hardware tools Er bestaan ook zogenaamde **packetshapers**. Deze apparaten zijn eigenlijk kleine computers, en zijn bedoeld om bandbreedte te beheren. Dit doen ze door bepaalde typen verkeer voorrang te verlenen boven andere typen door middel van diverse mechanismen. De meeste van deze apparaten kunnen ook zeer goed gegevens verzamelen over wat er allemaal aan netwerkverkeer de packetshaper passeert. Het type gegevens dat deze apparaten verzameld komt minimaal overeen met wat binnen de RMON standaarden gedefinieerd is, maar met in ieder geval het essentiële verschil dat ze de gegevens veel langer kunnen bewaren. De PacketShaper van Packeteer bijvoorbeeld bewaart de gegevens minimaal 2 maanden.

Een ander voordeel van deze apparaten is dat ze ook al voorbereid zijn om een diversiteit aan grafieken uit de opgeslagen gegevens kunnen genereren. Voor een eerste (snelle) analyse van het netwerk, althans voor de directe stromen, zijn deze apparaten dus zeer geschikt.

5.6.2 Software tools

Software tools die kunnen meten aan het netwerk zijn er legio. Bij sommigen is de hoofdfunctionaliteit het meten, bij anderen is dit slechts een (kleine) deelfunctionaliteit. Helaas zijn er bijna geen pakketten die gedetailleerd genoeg kunnen meten voor wanneer het grootste detail van de gegevens vereist is. Een uitzondering hierop is het pakket Net-Acct (zie pagina 62), alhoewel deze alleen IP verkeer kan meten en bijvoorbeeld niet IPX. Een ander probleem dat bij de meeste pakketten voorkomt is dat de meetgegevens niet in een direct bruikbaar formaat geleverd kunnen worden. Sommige pakketten geven alleen rapporten af, anderen weer alleen algemene gegevens of grafieken. Die pakketten zijn daardoor onbruikbaar. In Bijlage A.2 staat een kort overzicht van een aantal bekeken tools met commentaar over wat goed en slecht aan die tools is.

5.6.3 Meten op de diensten zelf

Veel diensten, maar lang niet allemaal, houden alle of een gedeelte van hun activiteiten bij in **log bestanden**. De meldingen die hier in worden geschreven kunnen variëren van eenvoudige berichten als “Dienst gestart” tot complete debug gegevens of volledige gedetaileerde gebruiksgegevens. Een dienst die zelf als meetpunt moet gaan dienen kan alleen als zodanig gebruikt worden als deze een log bestand gebruikt. Daarin moet deze dan wel genoeg gegevens stoppen zodat die gebruikt kunnen worden om hiervan meetgegevens te maken. In Hoofdstuk B staan onder andere een aantal voorbeelden van log bestanden van diensten, welke gebruikt zouden kunnen worden als meetgegevens.

De meeste log bestanden staat echter weer zoveel tekst in dat het al snel onmogelijk wordt om hier handmatig de gegevens uit te halen. Voor de log bestanden van de meest

populaire diensten zijn er vaak al door allerlei mensen kleine tooltjes gemaakt (meestal Perl-scripts of kleine programma's) om de gegevens er uit te halen en in een ander formaat te zetten. Dit soort tools, vaak gratis van het Internet af te halen, kunnen dan worden gebruikt, eventueel met een aanpassing, om de log bestanden om te zetten in bestanden met bruikbare meetgegevens.

Zelfs wanneer er nog geen tool voor is gemaakt is het vaak maar een kleine moeite om zelf iets te maken, al dan niet door een bestaande tool aan te passen.

Is een dienst niet in staat om "gebruiksgegevens" te produceren, al dan niet in een log bestand, dan wordt het lastig om het gebruik exact vast te stellen. De enige methode die dan nog overblijft is om vlak voor en vlak na de dienst het netwerkverkeer te gaan meten. Uit die gegevens valt dan misschien het indirecte verkeer te herleiden, afhankelijk van of er nog andere diensten op de zelfde machine actief zijn. Dit kan helaas lastig en onnauwkeurig worden. In dit soort situaties kan het wellicht beter zijn om, wanneer het om kosten doorbereken gaat, een **tarief per request** in te stellen. Gaat het niet om doorberekenen maar alleen om het gebruik vast te stellen, dan kan er wellicht worden beslist om deze dienst apart te beschouwen.

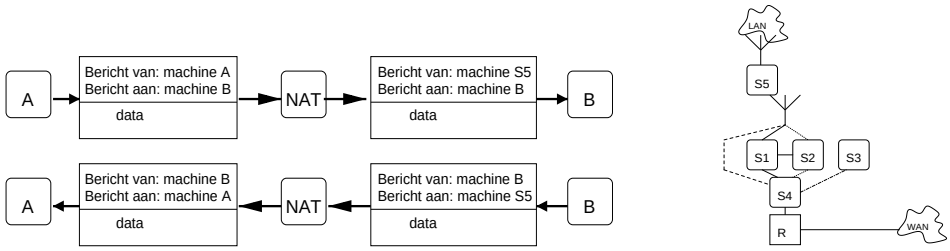
5.6.4 Obstakels

In sectie 5.4 werd reeds vermeld dat er diensten in het netwerk aanwezig kunnen zijn die zelf WAN verkeer kunnen genereren naar aanleiding van verzoeken uit het netwerk. Dit soort diensten maken het noodzakelijk om bij te houden naar aanleiding van welk verzoek (en van wie) dat verkeer is gegenereerd. Wordt dit niet gedaan dan geven de verzamelde meetgegevens aan dat de machine waar die dienst draait veel verkeer zelf heeft veroorzaakt, terwijl dit natuurlijk in dienst van anderen was.
In de volgende gedeelten worden een aantal veel voorkomende typen diensten besproken en wordt uitgelegd waar de moeilijkheden zitten.

Network Address Translation / Masquerading

Bij N(etwork) A(ddress) T(ranslation), ook wel aangeduid met Masquerading, wordt van communicatiestromen die langskomen het afzenderadres veranderd in een ander adres (normaal gesproken het adres van de machine die deze verandering uitvoert). Bij de terugkerende stromen wordt het omgekeerde gedaan, namelijk het bestemmingsadres wordt weer terugveranderd in de originele afzender (zie ook Figuur 5.5(a) voor een voorbeeld van **NAT in werking**). Deze verandering van adressen wordt gedaan om de hoeveelheid benodigde externe adressen te beperken of uit beveiligingsoogpunt om interne adressen niet naar buiten kenbaar te maken. NAT is iets dat meestal wordt uitgevoerd door de WAN router of een machine die daar vlak voor staat (machine S4 in de figuren in dit hoofdstuk). Indien dit daadwerkelijk gebeurt op zo'n plek, en er worden verder geen andere diensten die indirect verkeer veroorzaken op die machine, dan geeft NAT geen problemen. Immers, het meten gebeurt dan toch al op of voor die machine zodat de originele adressen nog bekend zijn. Gebeurt het meten achter de machine die NAT doet, dan lijken alle communicatiestromen van de NAT machine af te komen.

Wanneer NAT ergens in het netwerk gebeurt en niet op een lokatie zoals hierboven aangegeven, dan wordt het probleem een stuk groter. Indien de betreffende dienst (NAT) een log bijhoudt met wat en hoeveel hij verandert, dan kunnen de communicatiestromen die de NAT machine verlaten nog wel uit elkaar worden gehaald met behulp van de logs. Wanneer echter die communicatiestromen ook nog naar een andere dienst gaan (die **indirect verkeer** veroorzaken) dan wordt het bijna ondoenlijk om hier nog individueel gebruik er uit te lichten. Er moet dan immers uit de logs van beide diensten per tijdseenheid worden gehaald welke communicatiestromen door NAT veranderd zijn en vervolgens



(a) Werking van NAT

(b) NAT dienst in een netwerk

Verzoek en antwoord bij NAT

S5 = server met NAT dienst

Figuur 5.5: Network Address Translation

naar de betreffende dienst zijn gegaan. In een netwerk met meer dan een tiental machines en druk verkeer is dat al niet meer te doen.

De conclusie die hier uit getrokken kan worden is dat **NAT *binnen* een netwerk** niet gewenst is en normaal gesproken ook niet nodig is. Is die NAT dienst wel essentieel dan gaat het vermoedelijk om een netwerk met conflicterende adressen; een hernummering van die adressen is dan vereist, ook wegens een duidelijkere inrichting van het netwerk. Is zo'n hernummering niet haalbaar wegens de grootte van dat netwerk of andere redenen, dan moet er worden bedacht of het niet mogelijk is om voor dat netwerk apart meetgegevens te verzamelen. Het netwerk dat achter de NAT dienst zit moet dan voor het netwerk verder als één machine worden beschouwd. NAT aan de rand van een netwerk, zoals S4 in de voorgaande figuren, is daarentegen geen probleem en meestal *wel* gewenst.

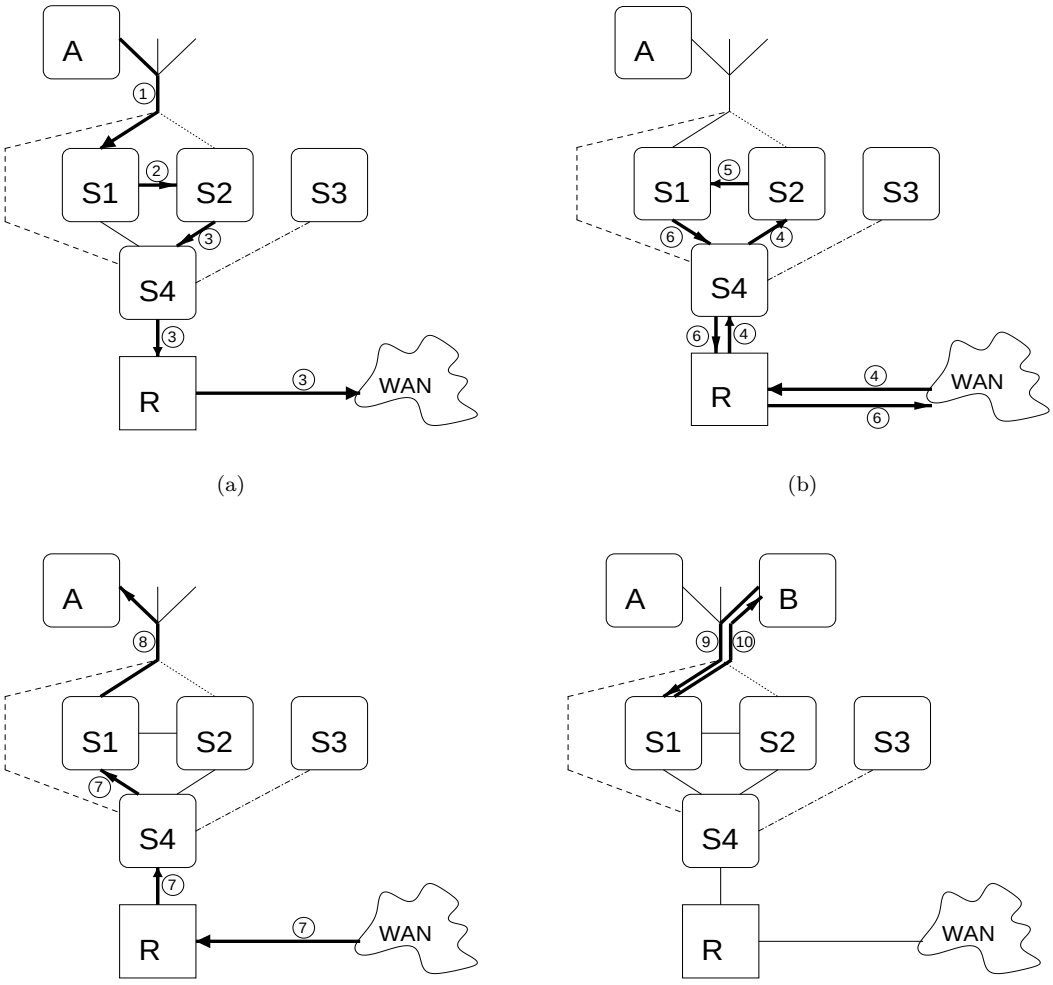
In Figuur 5.5(b) is zo'n situatie geschetst. Hierin is S5 de server waarop de NAT dienst draait voor het LAN dat zich daarachter bevindt. Dat netwerk noemen we hier verder netwerk B. Indien een machine in netwerk B nu gebruik maakt van bijvoorbeeld de proxy dienst op S1 in het andere netwerk (netwerk A), dan zal voor wat betreft S1 het verzoek komen vanaf S5. Indien dan uit de logs of andere meetgegevens van S4 moet worden gehaald wie dat verzoek heeft gedaan, dan moeten tijd, datum en grootte bekend zijn en deze met de logs of andere meetgevens van S5 worden gecombineerd om te achterhalen wie echt het verzoek heeft gedaan. In een klein netwerk met weinig machines en weinig netwerkverkeer zal dit nog wel lukken. In een druk netwerk met veel machines is dit al niet meer haalbaar, al is het maar wegens de vele verzoeken per seconde. Logbestanden of andere meetapparatuur gaan in hun tijdsregistratie normaal gesproken niet verder dan seconden nauwkeurig en dat is dan al te weinig. In de situatie zoals hier voorgesteld moet netwerk B voor wat betreft de metingen in netwerk A worden gezien als 1 machine (S5). Binnen netwerk B kan dan een zelfde meetsysteem worden ingevoerd als in netwerk A.

Wanneer er ook aan billing wordt gedaan, dan kan er niet meer aan worden ontkomen om een tarief per request in te stellen en niet per hoeveelheid data. Het is hier nu immers niet meer mogelijk om een request naar een gebruiker terug te leiden.

Proxies

Een proxy heeft als primaire taak het ophalen van gegevens van een andere server of dienst, naar aanleiding van een ontvangen verzoek hiervoor. Het antwoordt geeft hij, al dan niet aangepast, door aan degene die het verzoek heeft gedaan. Per definitie genereert een proxy dus indirect netwerkverkeer, al dan niet via het WAN.

Om te kunnen bepalen wie (welke machine en / of gebruiker) voor welk stuk indirect netwerkverkeer verantwoordelijk is, is het esssentieel dat de proxy zelf ergens vastlegt wie



(a) (b)

(c) (d) Verzoek voor de gebufferde pagina

Figuur 5.6: Communicatiestromen van een proxy-request.

een verzoek instuurt, voor wat, en hoeveel netwerkverkeer het verzoek (inclusief antwoord) veroorzaakt.

Bij **indirect verkeer** moet er bij een proxy bijvoorbeeld worden gedacht aan DNS requests. Een machine wil bijvoorbeeld gaan communiceren met een interne of externe machine, en gebruikt hiervoor een naam in plaats van een adres. De machine zal dan eerst aan een DNS server vragen of deze een adres heeft bij die naam. Wanneer het antwoord wordt ontvangen zal de echte connectie pas worden opgezet.

Wanneer de proxy ook als buffer (cache) fungeert komt er nog een extra probleem bij. Het probleem met **gebufferde gegevens** is dat de eerste gebruiker die de gegevens opvraagt het echte WAN verkeer veroorzaakt, terwijl requests van andere gebruikers voor diezelfde gegevens uit de buffer worden gehaald en geen extra WAN verkeer veroorzaken.

De proxy moet in dat geval dus ook vastleggen of hij zijn antwoord uit de buffer geeft of "nieuw" heeft opgehaald. Zie verderop voor meer over buffers. Zo'n bufferende proxy wordt ook wel een **caching proxy** genoemd.

Proxy Voorbeeld Een voorbeeld maakt alles een stuk duidelijker. Hieronder staan de stappen beschreven die worden uitgevoerd wanneer een machine A de webpagina www.domain.com opvraagt via de proxy, gevolgd door een verzoek van machine B voor diezelfde webpagina.

Stap	Source	Data	Destination
1	A	get http://www.domain.com	Proxy
2	Proxy	req. ip for www.domain.com	DNS
3	DNS	req. ip for www.domain.com	Internet
4	Internet	ip=1.2.3.4	DNS
5	DNS	ip=1.2.3.4	Proxy
6	Proxy	get http://1.2.3.4	Internet
7	Internet	webpagina 1.2.3.4	Proxy
8	Proxy	webpagina 1.2.3.4	A

Vervolgens vraagt machine B dezelfde pagina op via de proxy.

Stap	Source	Data	Destination
9	B	get http://www.domain.com	Proxy
10	Proxy	webpagina 1.2.3.4	B

In Figuur 5.6 staan de communicatiestromen ingetekend in het logische overzicht van sectie 5.4.

Stappen 2 t/m 7 worden dus voor B helemaal niet meer uitgevoerd doordat de proxy reeds het ip adres weet van www.domain.com en ook de webpagina reeds in zijn buffer heeft. De hoeveelheid WAN verkeer gegenereerd door stappen 2 t/m 7 zal dus moeten worden verdeeld over zowel A als B. Overigens gebruikt dit voorbeeld het Internet, maar soortgelijke situaties bestaan ook buiten het Internet om binnen bedrijfsnetwerken.

Om bij het bovenstaande voorbeeld volledige meetgegevens te kunnen krijgen, dienen de volgende acties te worden gemeten:

1. de communicatie tussen A en de proxy
2. de communicatie tussen B en de proxy
3. de communicatie tussen de proxy en de DNS server
4. de communicatie tussen DNS en het Internet
5. de communicatie tussen de proxy en het Internet

Transacties 1 en 2 kunnen worden gemeten door op of vlak voor de proxy het verkeer bestemd voor de proxy dienst te meten, of beter nog, door de gegevens uit de logs van de proxy te halen. De transactie tussen de proxy en de DNS zullen meestal niet uit logs

gehaald kunnen worden en moeten dus uit aparte meetgegevens worden gehaald. Dit maakt alles er niet makkelijker op.

Om dat soort gegevens allemaal met elkaar te combineren zal er erg veel werk moeten worden verricht, waardoor een ietwat andere aanpak beter zal zijn.

In plaats van elk request exact in rekening te brengen (wat niet zal lukken) is het misschien beter haalbaar om een vast **tarief per request** vast te stellen. Dit kan bijvoorbeeld worden gedaan door te meten hoe groot een request (plus antwoord) gemiddeld is en dit te combineren met het elders bepaalde tarief per eenheid data. Overigens is dit slechts een van de mogelijkheden en zal een specifieke situatie uit moeten wijzen wat daar de meest geschikte methode is.

User - IP mapping Een ander probleem wat het een en ander erg lastig kan maken is het zoeken van de juiste gebruiker bij een bepaald (IP) adres. Er zijn verscheidene manieren waarop een machine aan zijn IP adres kan komen. Dit zijn: op de machine zelf ingesteld, dynamisch toegekend, of volledig dynamisch toegekend.

Wanneer het **adres op de machine zelf ingesteld** is, dan zijn er nog een aantal problemen te erkennen. Als eerste dient de machine afdoende beveiligd te zijn tegen het ongewenst veranderen van het adres. Indien een ongewenste verandering gebeurt, dan zal bij de aggregatie van de gegevens het netwerkverkeer van die machine worden aangezien voor het verkeer van een andere machine, mogelijk iemand die bij een andere bedrijfseenheid werkt.

Een ander probleem onstaat wanneer er meer dan 1 gebruiker van een machine gebruik maakt. Indien men zich altijd netjes individueel aanmeldt bij het netwerk dan valt het netwerkverkeer nog te combineren met een gebruiker met behulp van de gebruikersauthenticatie log bestanden. Wanneer dit niet gebeurt dan zal hier een compromis moeten worden gesloten en voor die machine de gegevens worden geaggregeerd naar meerdere gebruikers.

Een machine kan ook met behulp van DHCP, BOOTP, RAS, e.d. een **dynamisch toegewezen IP adres** krijgen. Deze diensten kunnen zo ingesteld worden dat een bepaalde machine altijd hetzelfde adres krijgt. Bij deze methode kan de instelling van de dienst erbij worden gehaald om de mapping van IP adres naar machine te maken.

De hierboven genoemde diensten kunnen (daarnaast) ook zo worden ingesteld dat een machine een **volledig dynamisch toegewezen IP adres** krijgt. Er wordt dan uit een groep adressen geput waaruit een adres aan een machine wordt toegewezen. Doordat de adressen uit een groep komen kan een machine zelfs ieder (op de dienst ingesteld) interval een ander adres krijgen.

Uiteraard dient bij alle methodes van IP adressen instellen of toewijzen de machine afdoende beveiligd te zijn tegen het ongewenst veranderen van de instellingen, anders is er niet meer met zekerheid te zeggen wie bij welk IP adres hoort.

Deze diensten zijn meestal wel voorzien van een log faciliteit zodat de relevante activiteiten (het toewijzen of vrijgeven van een adres) goed bijgehouden worden (zie sectie B.2.2 voor een voorbeeld van zo'n log bestand). Op die manier kan in ieder geval de mapping van adres naar machine gemaakt worden. De mapping van user naar machine moet dan wel nog gemaakt worden. Niet alle netwerken (lees: servers) houden op dezelfde manier bij welke gebruiker vanaf welke machine zich aanmeldt. Sommige typen houden dit bij als een combinatie van machine adres en gebruikersnaam, anderen houden het weer bij als combinatie van protocol specifiek adres en gebruikersnaam. In het eerste geval is het dan nu mogelijk om een gebruiker te koppelen aan een IP adres met behulp van de gebruikersauthenticatie log bestanden. In het tweede geval is het direct mogelijk om een gebruiker aan een IP adres te koppelen, alleen veranderd het probleem dan. Er zal dan moeten worden gecontroleerd of er niet meerdere gebruikers tegelijkertijd aangemeld zijn

geweest vanaf hetzelfde adres. De meeste meetgegevens zullen echter uitgaan van een protocol specifiek adres zodat het gehele probleem eigenlijk een **fraude detectie** probleem is geworden. Er moet dan immers gecontroleerd worden of er geen onmogelijke situaties tevoorschijn komen uit de log bestanden.

Firewalls zijn niet echt als een dienst te beschouwen zoals de voorgaande voorbeelden dat wel waren. Ze kunnen echter net zo zeer voor problemen zorgen. Een firewall is er om ongewenste communicatie tegen te houden. Dit kan zowel om inkomende als uitgaande communicatie gaan. Wanneer er verkeer wordt tegengehouden dan zal dit normaal gesproken gebeuren zonder de originele verzender hiervan op de hoogte te stellen.

Een firewall wordt meestal ingericht op het knooppunt met een groter netwerk, zoals een verbinding met een WAN. Dit gebeurt dan vaak op de router of de machine direct daar voor (zoals machine S4 in de voorbeelden in dit hoofdstuk). Hier komt dan ook meteen het probleem tevoorschijn. Wordt het netwerkverkeer zowel voor de firewall (aan de LAN zijde) als tussen deze firewall en de router gemeten (wat in de meeste gevallen helemaal niet nodig is), en blokkeert de firewall bepaalde netwerkstromen, dan zullen de meetgegevens van beide lokaties niet meer overeenkomen. Ze kunnen dan ook alleen nog maar met elkaar in relatie gebracht worden als de firewall extra informatie levert.

Andersom bestaat ook een probleem. Als er alleen voor de firewall gemeten wordt en de firewall blokkeert bepaalde netwerkstromen, dan komt de werkelijk verzonden hoeveelheid gegevens niet meer overeen met de hoeveelheid die de eigen metingen aangeven. Ook hier geldt dat de firewall extra informatie moet leveren om weer alles compleet te krijgen.

Deze extra informatie moet dan bestaan uit gegevens die aangeven dat er netwerkstromen zijn geblokkeerd en nog belangrijker welke dat dan zijn inclusief verzender, bestemming en hoeveelheid. Die gegevens zullen bijna altijd in een log bestand worden opgeslagen. In sectie 5.6.3 werd reeds het een en ander over dit soort log bestanden verteld.

Packetshapers introduceren een soortgelijk probleem als firewalls. Een packetshaper is bedoeld om bepaalde typen netwerkverkeer voorrang te geven boven anderen. Een manier waarop dit wordt bereikt is door aan de verzendende machine te vragen of hij opnieuw wil verzenden, eventueel met bijvoorbeeld een kleinere hoeveelheid data tegelijk. Op die manier kan het daardoor voorkomen dat een machine meerdere malen dezelfde data verstuurt terwijl deze toch maar één maal het WAN op worden gestuurd. Dit is dus het omgekeerde probleem van de firewalls waar data *niet* verzonden werd, terwijl bij een packetshaper soms data *dubbel* verzonden wordt.

In tegenstelling tot firewalls leggen de packetshapers dit soort activiteiten niet vast; dit is ook niet echt verwonderlijk aangezien zo'n logbestand dan een buitengewoon gigantische omvang zal krijgen. Dit betekent echter wel dat hier niets aan te doen valt voor de meetgegevens.

5.7 Verzamelen en opslaan van de gegevens

Wanneer alle programmatuur en hardware hun werk hebben gedaan gedurende een bepaald tijdsinterval, dan resulteert dit in een heleboel gegevens. Lang niet al deze gegevens staan in de zelfde vorm. Sommige gegevens zijn lijsten met adres, type en aantal, anderen kunnen bijvoorbeeld log bestanden zijn. Om enigszins efficiënt met de gegevens te kunnen gaan werken (in het volgende hoofdstuk) zullen deze gegevens ergens moeten worden opgeslagen. Waarschijnlijk de beste methode is om alles in een database te plaatsen. Op die manier kunnen er direct relaties tussen de diverse soorten gegevens worden gelegd voor zover deze er zijn. Ook het omvormen en combineren van diverse gegevens kan op die manier redelijk makkelijk gebeuren. Als extra voordeel valt nog te noemen dat, met het in elkaar zetten van de juiste bewerkingen (queries), uiteindelijk het geheel van verwerken van de gegevens automatisch kan gebeuren.

De eerste stap die moet worden genomen voordat de gegevens in een database kunnen worden opgeslagen is uiteraard het ontwerpen van die database. Bepalend voor het ontwerp zijn in ieder geval welke soort gegevens er in zullen komen. Er zijn een aantal soorten gegevens:

1. lijsten met ruwe data (tijd, adres, type, hoeveelheid)

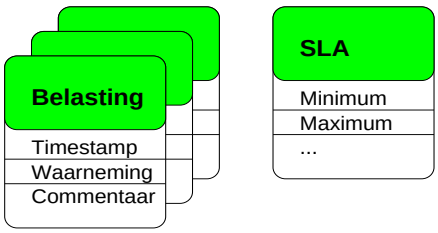
2. log bestanden

3. "externe" gegevens (van nota's etc.)

Wanneer er wordt teruggegrepen op de categoriëen uit Hoofdstuk 4 dan kan er een onderscheid worden gemaakt in de complexiteit van de meetgegevens. Deze complexiteit loopt eigenlijk gelijk met de categoriëen: *prijs/prestatie* levert relatief simpele en weinig gegevens op, *functioneel gebruik* levert redelijk eenvoudige gegevens op maar wel al meer, en *wie gebruikt het netwerk* levert mogelijk een enorme vracht aan (verschillende) gegevens op. Voor alle drie de categoriëen kan een ander database ontwerp voor worden gemaakt om op die manier de onderliggende verwerking niet onnodig complex te maken.

Prijs / prestatie

Wanneer het bij "prijs/prestatie" alleen gaat om controle op aanwezigheid en de beschikbare bandbreedte, dan zullen één of meer van de volgende gegevens worden verzameld:



Figuur 5.7: Tabel ontwerp

- Periodiek uitgevoerde waarneming of de verbinding aanwezig is. De resultaten bestaan uit een ja/nee waarde, en in het geval van een nee waarde de oorzaak hiervan (voorzover bekend uiteraard).

- Periodiek uitgevoerde test of waarneming (afhankelijk van het soort verbinding en de mogelijkheid om dit te meten) wat de bruikbare (minimum) capaciteit is.

- Belastingsgegevens van de verbinding. Dat wil zeggen periodieke uitgevoerde waarneming van de **doorvoer snelheid** (de capaciteit).

- Periodiek uitgevoerde test of waarneming (afhankelijk van het soort verbinding en de mogelijkheid om dit te meten) wat de maximum capaciteit is.

- Gegevens over de capaciteit en beschikbaarheid zoals overeengekomen in de **SLA**.

Indien de gegevens van de SLA even buiten beschouwing worden gelaten, dan bestaan de anderen allemaal uit regels met een timestamp, een waarde en een eventueel commentaar. Opslag van deze gegevens kan dan eenvoudigweg gebeuren door voor ieder item een aparte tabel te maken met deze drie velden.

De gegevens uit de SLA zijn pas bij de analyse nodig waardoor het niet noodzakelijk is om deze in een database te stoppen. Voor compleetheid van de gegevens is het beter om dit wel te doen.

Functioneel gebruik

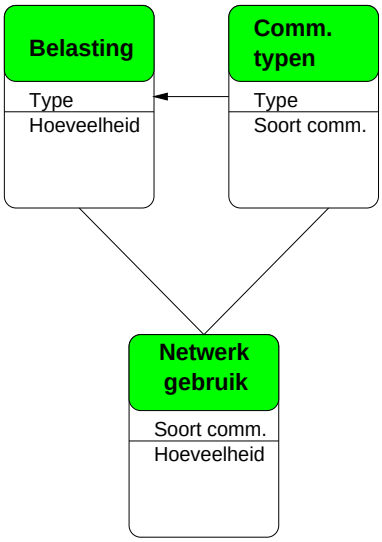
De verzamelde gegevens voor "functioneel gebruik van het netwerk" bestaan uit lijsten met hoeveelheden verzonden data, eventueel de snelheid, en het type van de communicatie (zoals WWW, Lotus Notes, telnet, enzovoorts). De gegevens komen per verbinding maar van één punt, ook hier is het dus niet nodig om een ingewikkelde structuur te maken.

Er hoeven daarom geen gegevens van verschillende meetpunten bij elkaar te worden gezocht. Wel zullen er verschillende **communicatietypen** bij elkaar moeten worden gezocht. Een applicatie kan namelijk voor meer dan één communicatietype aan verkeer zorgen (zie ook sectie 3.2.2). Bijvoorbeeld FTP verkeer zorgt normaliter voor twee typen verkeer, namelijk FTP (poort 21) en FTP-data (poort 20). Ook email verkeer kan voor meerdere typen verkeer zorgen (SMTP(25), POP3(110), auth(113), en nog meer). Welke typen bij elkaar horen moet (in principe eenmalig) bepaald worden aan de hand van standaarden en documentatie van de verschillende applicaties.

Is eenmaal bekend welke typen bij elkaar horen dan kunnen alle meetgegevens die volgens hun type bij elkaar horen onder één noemer (meestal de bijbehorende applicatie) worden gebracht. Het beste is het om dit te doen door middel van een query en niet door de originele data aan te passen.

Indien er communicatietypen overblijven waarvan niet gevonden kan worden waar deze bij horen, dan zal er gedurende een korte periode ook inhoudelijk in dat type netwerkverkeer moeten worden gekeken (in de applicatielaag, zie sectie 3.1). Hier kunnen zogenaamde **sniffers** voor worden gebruikt, welke vaak een heel eind komen in het ontleden van het netwerkverkeer (zie Bijlage A.2). Soms kan er door in de daadwerkelijk verzonden data te kijken worden achterhaald van welke applicatie dit is. Dit kan gebeuren door bijvoorbeeld herkenbare tekst in die data, of door de combinatie met ander netwerkverkeer dat daaromheen voorbij kwam. Bij email ontvangst met SMTP komt er bijvoorbeeld eerst de SMTP connectie, dan een AUTH(enticatie) connectie de andere kant op, waarna de echte email pas wordt verstuurd in de originele richting via de initiële SMTP connectie.

De bekende types kunnen samen met hun naam of omschrijving weer in een aparte tabel worden geplaatst welke als referentie voor de hierboven genoemde query kan functioneren.

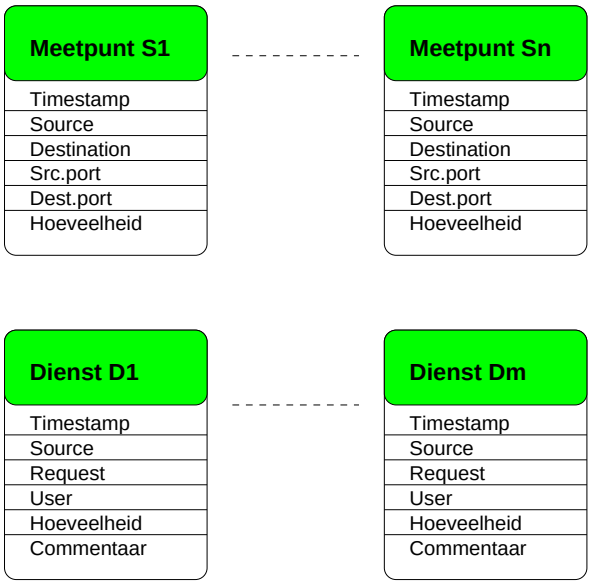


Figuur 5.8: Tabel ontwerp

Wie gebruikt het netwerk

Bij deze categorie worden de meeste gegevens verzameld, vaak in een veelvoud aan vormen. Er zijn grofweg twee vormen: (1) log bestanden en (2) transactielijsten op protocol niveau. Binnen die twee vormen kan ook weer een veelvoud aan verschillende vormen bestaan door de variatie in de informatie die ze bevatten. Zeer belangrijk wordt het hier dus om uit alle verschillende vormen dezelfde gegevens te kunnen halen. Door enigszins abstract naar beide vormen te kijken, evenals naar de vereisten welke in dit hoofdstuk aangegeven zijn, kan er voor beide vormen worden aangegeven welke gegevens hieraan te ontrekken zijn.

Voor de log bestanden zijn dit:



Figuur 5.9: Tabel ontwerp

Timestamp	-	Het tijdstip van het request aan de dienst
Source	-	Het bron adres van waar het request komt
Request	-	Het request zelf in welke vorm dan ook
User	-	De gebruiker die het request heeft gedaan (indien bekend)
Hoeveelheid	-	De hoeveelheid data van het antwoord (indien bekend)
Commentaar	-	Extra opmerkingen (zoals cache-hits of cache-misses)

Voor de transactielijsten zijn dit:

Timestamp	-	Het tijdstip van de transactie
Source	-	Het bron adres van waar de transactie komt
Destination	-	Het adres waar de transactie heen gaat
Source Port	-	Het poortnummer of soort communicatie van de source
Destination Port	-	Het poortnummer of soort communicatie bij de destination
Hoeveelheid	-	De hoeveelheid data van het antwoord

Met dit ontwerp wordt het dan nu zaak om de verschillende meetgegevens in de juiste tabellen onder te brengen. Het duidelijkst voor de verwerking zal zijn om voor elke dienst of meetpunt een aparte tabel te gebruiken. Bij het verdere verwerken van de gegevens kan dan, analoog aan het logische overzicht van het netwerk, elke indirecte route worden omgezet in een directe route. Wat hier nog niet kan worden gedaan is het onderbrengen van verschillende typen communicatieverkeer onder één noemer zoals bij de vorige categorie (functioneel gebruik) is gedaan. Doordat er nog veel gegevens gecombineerd zullen moeten worden kan dat pas veel later worden gedaan.

Hoofdstuk 6

Verwerking van de gegevens

Nadat de informatie- en gegevensbehoeften zijn vastgesteld en de benodigde gegevens zijn verzameld resulteert dit in de meeste gevallen in een enorme hoeveelheid meetgegevens. Deze gegevens zijn meestal ongeschikt om onveranderd aan het management door te spelen; de gegevens zullen eerst verwerkt moeten worden tot presenteerbare of tot een enigszins handelbare hoeveelheid gegevens. Zodra dat is gebeurd wordt het ook mogelijk om analyses uit te voeren op die gegevens en daarop gebaseerde conclusies te trekken.

Evenals in Hoofdstuk 4 maken we ook hier weer het onderscheid in drie categoriëen. Hier is dat onderscheid van evengroot belang, omdat de aard van de verzamelde gegevens nogal verschillen en een eigen aanpak vereisen.

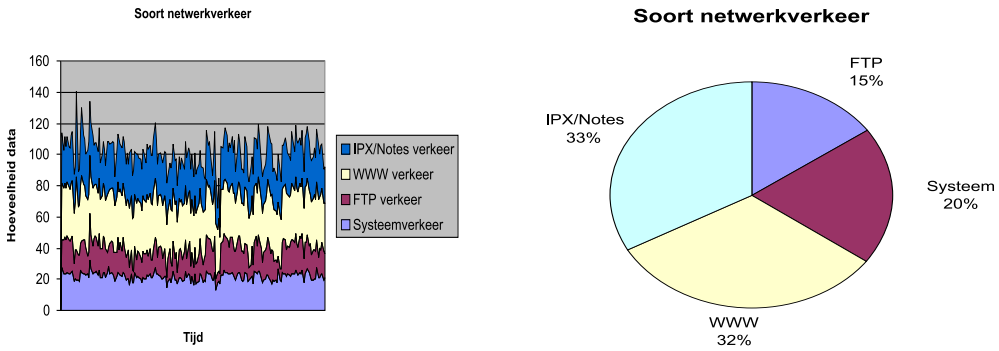
6.1 Aggregatie

6.1.1 Prijs/prestatieverhouding

Wanneer het alleen gaat om controle op aanwezigheid en de beschikbare bandbreedte, dan zouden één of meer van de volgende gegevens nu verzameld moeten zijn:

- Periodiek uitgevoerde waarneming of de verbinding aanwezig is. De resultaten bestaan uit een ja/nee waarde, en in het geval van een nee waarde de oorzaak hiervan (voorzover bekend uiteraard).
- Periodiek uitgevoerde test of waarneming (afhankelijk van het soort verbinding en de mogelijkheid om dit te meten) wat de bruikbare (minimum) capaciteit is.
- Belastingsgegevens van de verbinding. Dat wil zeggen periodieke uitgevoerde waarneming van de **doorvoer snelheid** (de capaciteit).
- Periodiek uitgevoerde test of waarneming (afhankelijk van het soort verbinding en de mogelijkheid om dit te meten) wat de maximum capaciteit is.
- Gegevens over de capaciteit en beschikbaarheid zoals overeengekomen in de SLA.

Aangezien deze gegevens allemaal betrekking hebben op dezelfde verbinding en tijd lenen zij zich uitstekend om in één grafiek te verwerken. De MRTG tool (zie Bijlage A.2) is hiervoor gemaakt om dit allemaal continu te doen en netjes te presenteren. De tool is er weliswaar op gericht om dit allemaal in real-time te doen, maar hij is aan te passen om dit ook met gegevens achteraf te kunnen doen. Wordt zo'n tool niet gebruikt, maar bijvoorbeeld een spreadsheet of database, dan is het vooral belangrijk dat alle gegevens voorzien zijn van de bijbehorende "timestamp". Wanneer dit niet juist gebeurt dan laat de grafiek een vertekend beeld zien.



(a) Vb. 1 (b) Vb. 2

Figuur 6.1: Netwerkgebruik onderverdeeld per type netwerkverkeer.

Voor de latere analyse zullen vooral de “negatieve” gegevens van belang zijn, omdat die de problemen aangeven die opgelost moeten worden. Met “negatieve” gegevens worden bijvoorbeeld de meetpunten bedoeld waar de minimum beschikbare capaciteit zich onder het afgesproken minimum bevindt. Over dit soort meetpunten kan een stukje statistiek worden gemaakt. Daarin kan dan bijvoorbeeld worden opgenomen:

- hoeveel procent van de tijd de verbinding niet aanwezig was
- de oorzaken van complete afwezigheid van de verbinding
- hoeveel procent van de tijd de capaciteit onder het minimum lag
- hoeveel procent van de tijd de gebruikte capaciteit onder het minimum of boven het maximum zat

Wanneer dit soort statistieken naast de grafiek worden gemaakt, dan is feitelijk de analyse ook al grotendeels gebeurt. Immers, de berekende percentages kunnen naast de referentie percentages (SLA, nota, andere meetgegevens, enz.) worden gelegd, waarmee de zogenaamde “SOLL” en “IST” positie naast elkaar staan. Zie ook Figuur 4.2.

6.1.2 Functioneel gebruik van het netwerk

In het vorige hoofdstuk is reeds aangegeven dat bij het opslaan van de meetgegevens al een aggregatiestap kan worden gemaakt, door diverse meetgegevens die bij elkaar horen als zijnde van één applicatie, reeds onder één noemer te stoppen. Dit is eigenlijk al de belangrijkste verwerkings stap die kan worden genomen met deze gegevens. De resulterende gegevens behoeven nu eigenlijk alleen nog maar zodanig gevormd te worden dat zij geschikt zijn om een grafiek mee te maken.

Na de genoemde aggregatiestap uit het vorige hoofdstuk resulteert dit in één grote lijst met type communicatie en hoeveelheid welke gedurende de meetperiode zijn verstuurd. Met deze gegevens kan er nu of een eenvoudige tabel worden gemaakt met daarin deze gegevens, of een grafiek (meestal een cirkeldiagram). Van belang voor de latere analyse zullen vooral de onderlinge verhoudingen zijn en wellicht ook de hoeveelheden. Het is dus verstandig om dit soort cijfers bij de tabel of grafiek te vermelden.

Zijn de gegevens telkens verzameld over kleinere aansluitende tijdsintervallen, dan zullen ze voorzien moeten zijn van een timestamp. Met dit extra gegeven wordt het mogelijk om bepaald applicatiegebruik in verloop van tijd in beeld te brengen. Hiertoe kunnen tijd en hoeveelheid per applicatie tegen elkaar worden uitgezet in een grafiek. Dit

kan ook worden gedaan voor meerdere applicaties in één grafiek. Op die manier kunnen er mogelijk verbanden tussen bepaald applicatiegebruik worden gedetecteerd. Vooral voor de communicatietypen die niet geïdentificeerd zijn kunnen dit soort grafieken een mogelijke hint vormen waar zij bij horen (wanneer bepaalde vormen terugkeren bij andere typen verkeer).

In Figuur 6.1 staan een aantal voorbeelden van dit soort grafieken. Ook in sectie 7.5 zijn enkele voorbeelden te zien.

6.1.3 Wie gebruikt het netwerk

Deze categorie levert de meeste meetgegevens op, maar ook de meeste problemen en werk. In zijn meest uitgebreide vorm zijn er van het WAN en van diverse lokaties in het LAN meetgegevens verzameld. Zoals in het vorige hoofdstuk reeds aangetoond komen hier problemen tevoorschijn als gevolg van het indirecte verkeer.

Bij de aggregatie van de meetgegevens moeten de gegevens van de verschillende diensten en andere meetpunten samengevoegd worden tot één lijst met gegevens. Hierbij moeten de indirecte stromen worden verwerkt tot directe stromen, zodat al het netwerkverkeer aan specifieke gebruikers kan worden toegerekend. Een ander (eenmalig) probleem dat moet worden overwonnen zijn de grote verschillen in vorm van de meetgegevens. Zo zullen sommige meetgegevens in een eenvoudig verwerkbaar formaat staan (zoals die van de Network Accounting daemon, zie Bijlage A.2) en andere gegevens staan weer in log bestanden. In sectie 5.7 is reeds aangegeven welke gegevens uit de diverse vormen gehaald moeten worden.

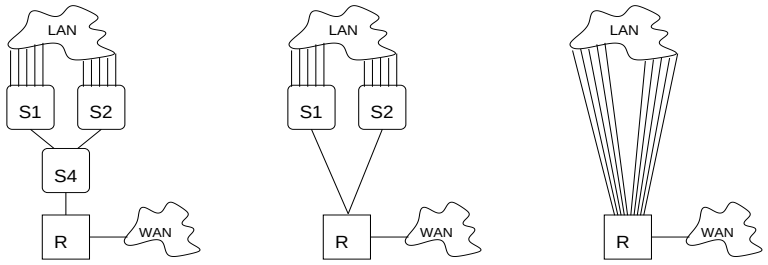
Welk detailniveau de gegevens moeten behouden is vastgesteld bij het bepalen van de informatie- en gegevensbehoefte in Hoofdstuk 4 en wordt hier de leidraad voor de verwerking van de gegevens. Hoe minder detail er hoeft te worden behouden hoe meer meetgegevens onder één noemer kunnen worden gestopt. Voor het verwerken van de gegevens geldt wel dat er beter pas aan het einde van het verwerkingsproces de gegevens bij elkaar kunnen worden gestopt

Reeds in het begin van het gehele verwerkingsproces is het al mogelijk om de hoeveelheid meetgegevens op een eenvoudige manier terug te brengen. Het kan zo zijn dat er bij de verzamelde gegevens naast de reguliere meetgegevens (lijsten met adressen en hoeveelheden) ook log bestanden zijn. Als er in zo'n log bestand voldoende gegevens aanwezig zijn om het gebruik van die dienst (in hoeveelheden) er uit te halen zijn, dan kunnen mogelijk een flink aantal van de reguliere meetgegevens worden weggelaten. Deze zijn dan immers uit de log bestanden te halen met mogelijk nog meer informatie. Wordt dit niet gedaan in zo'n situatie dan worden er zaken dubbel geteld zodat de eindresultaten niet meer accuraat zijn. Zie ook de voorbeelden van meetgegevens in Bijlage B.2 waar de verschillen in detail goed naar voren komen. De weg te laten gegevens kunnen uiteraard nog wel dienen als controlemiddel.

Als er **indirecte stromen** zijn, en er ook binnen die stromen meetgegevens zijn verzameld, dan zullen deze indirecte stromen eerst verwerkt moeten worden tot directe stromen. Wanneer dat is gebeurt dan staan alle gegevens in vorm afzender-ontvanger-type-hoeveelheid. En met *die* gegevens kan er dan eventueel gegroepeerd worden.

Het verwerken van de indirecte stromen kan in principe twee richtingen op gebeuren, van smal (WAN) naar breed (LAN) of andersom.

Aggregeren van WAN naar LAN Bij deze manier wordt begonnen met de meetgegevens van het meetpunt dat zich het dichtse bij de aansluiting naar het WAN bevindt. Van elke netwerkstroom die van een dienst uit het netwerk vandaan komt moet dan worden uitgezocht door welke machine deze veroorzaakt is. Dit kan worden gedaan door deze data uit de (log)gegevens van die dienst te halen.



(a) Stap 1 (b) Stap 2 (c) Stap 3

Figuur 6.2: Verwerking van de indirecte stromen van WAN naar LAN.

Aggregeren van LAN naar WAN Hierbij worden netwerkstromen gevolgd vanuit de machine die ze initiëren. Wanneer een stroom naar een dienst gaat, moet er uit de meetgegevens van die dienst (de logs) worden gehaald hoeveel netwerkverkeer heeft geresulteerd uit die stroom. Dat resulterende netwerkverkeer kan dan worden toegerekend aan de initiërende machine.

Beide manieren hebben zo hun eigen voor- en nadelen, het zal praktisch gezien meestal handiger zijn om van WAN naar LAN te werken. Wanneer er dan bijvoorbeeld een stuk LAN verscholen zit achter één adres (met behulp van NAT, zie sectie 3.4.1) en dit LAN als één geheel mag worden gezien dan kan er worden gestopt met verwerken bij dat ene adres. Daarnaast is het aantal vergelijkingen (van het ene adres met het andere) minder bij verwerking van WAN naar LAN dan andersom.

Neem als voorbeeld Figuur 6.2. In deze figuur zou S4 bijvoorbeeld een DNS server kunnen zijn die door de diensten op S1 en S2 (bijvoorbeeld een proxy en een mailserver) wordt gebruikt. Achter S1 en S2 bevinden zich de werkstations. We krijgen dan meetgegevens van de werkstations naar S1 en S2, van S1 en S2 naar S4, en van S4 naar R. Wanneer er nu meetgegevens op deze drie punten zijn verzameld evenals op de router R dan zou het verwerken als volgt moeten gebeuren:

1. De stromen tussen R en S4 moeten uit elkaar worden gerafeld. Dit gebeurt met de meetgegevens van S4. Deze geven aan hoeveel en wanneer S1 en S2 gebruik hebben gemaakt van S4, die vervolgens via R met het WAN heeft gecommuniceerd. De meetgegevens van de stromen tussen R en S4 worden met die informatie herschreven zodat er nu meetgegevens ontstaan voor de stromen tussen S1/S2 en R (via S4).

2. Nu kunnen met behulp van de meetgegevens van S1 en S2 de stromen tussen S1/S2 en R worden verwerkt tot stromen vanuit het LAN (de werkstations) naar R.

Het matchen van gebruiksgegevens uit logs met de “reguliere” meetgegevens hoeft in principe geen groot probleem te zijn, zolang er geen of nauwelijks tijd verstrijkt tussen het ontvangen van het request door een dienst en het doorsturen van dat request richting WAN of een andere dienst. Wanneer alle stappen gevolgd worden, wordt deze bewering duidelijker.

1. Een client doet een request bij een dienst. Dit geeft een separate logregel dat het request is gedaan, of nog een regel die pas verderop wordt gecompleteerd.
   Deze regel bevat minimaal een timestamp en een source van het request.

2. Vervolgens stuurt de dienst (indien deze nog geen antwoord heeft) het request door. Op dit request krijgt hij een antwoord. In de meetgegevens van een meetpunt

waarlangs dit request komt, zullen er gegevens worden vastgelegd zowel wegens het doorgestuurde request als het antwoord. Deze combinatie (dit heet een sessie, zie ook sectie 3.2.2) is in principe zonder moeite bij elkaar te vinden in die gegevens. Hierdoor is zowel de timestamp van het request en van het antwoord bekend, evenals de hoeveelheid data in het antwoord.

3. De dienst stuurt nu het antwoord door aan de client en maakt hiervoor een logregel aan, of completeert de initiële logregel uit stap 1.

Dit zijn dan de bijbehorende meetgegevens:

Log Dienst 1	Time.1a	Mach.A	Req.1			
Meetgeg. P1	Time.1b	Mach.1	Mach.2 / WAN	Hoeveelheid	Src Poort	Dest. poort
Meetgeg. P1	Time.2a	Mach.2	Mach.1 / WAN	Hoeveelheid	Src Poort	Dest. poort
Log Dienst 1	Time.2b	Mach.A	Reply op Req.1	Hoeveelheid		

In theorie zouden timestamp 1a en 1b, evenals 2a en 2b aan elkaar gelijk zijn. In de praktijk zal hier een minimaal verschil in zitten. Dit verschil is in principe echter redelijk constant en erg klein. Het zal daardoor geen probleem zijn om de juiste gegevens bij elkaar te vinden (ook automatisch niet).
Alle vier de gegevens bij elkaar gestopt geeft dit dan:
Time.1a — Mach.A — Soort request of communicatietype — Hoeveelheid

De gegevens staan nu echter nog in een formaat waarbij een machine als veroorzaker wordt aangegeven. Wanneer er nu ook authenticatie gegevens bekend zijn en deze ook in de database zijn ingevoerd, dan kunnen die gegevens gecombineerd worden, zodat de meetgegevens nu aan een gebruiker worden toegewezen.

Na het verwerken van de indirecte stromen resulteert dit in nog altijd een heleboel meetgegevens. Deze staan nu echter wel in een formaat dat direct aan geeft welke machine wat voor soort netwerkverkeer heeft veroorzaakt en hoeveel data hiermee is gemoeid. Deze meetgegevens zijn nu eigenlijk een soort transactierecords.

De gegevens kunnen nu naar wens geaggregeerd worden naar de eenheden waarin het management deze wil zien. Hierbij valt te denken aan aggregeren per:

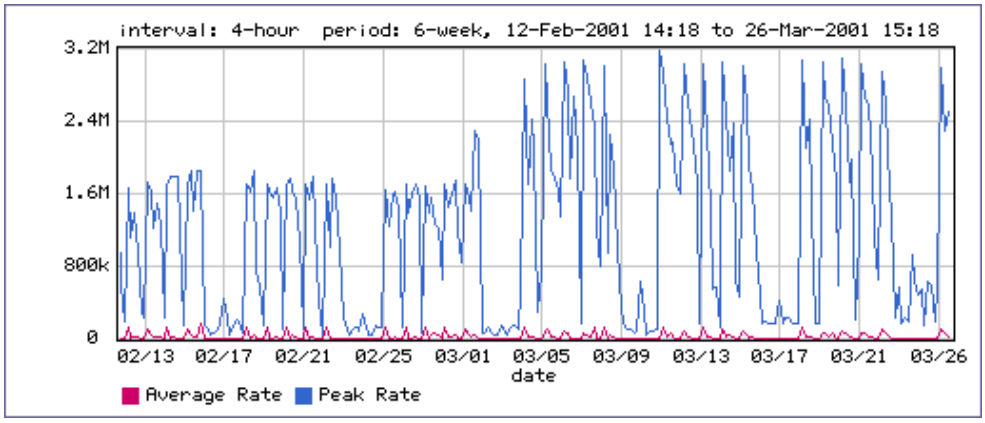
- gebruiker, gebruikersgroep, afdeling, of een andere bedrijfseenheid
- type gebruikte applicatie
- tijdseenheid
- of enige combinatie van bovenstaande

De gegevens kunnen nu echter ook compleet genoeg zijn om te gebruiken voor het doorbelasten van het netwerk gebruik. Er is immers een combinatie gemaakt tussen gebruikers en hoeveelheden, wat de twee meest essentiële gegevens zijn voor het doorbelasten.

Voor het visueel maken van de gegevens kan er wederom voor worden gekozen om grafieken te maken van het gebruik. Dezelfde grafieken als in de vorige sectie kunnen ook hier hun dienst doen. Doordat er nu echter meer gegevens beschikbaar zijn, kunnen deze grafieken nu bijvoorbeeld ook per afdeling of zelfs per machine (of gebruiker) worden gemaakt. Bij opmerkelijke pieken in grafieken kan er dan voor worden gekozen om deze meteen uit te diepen naar de grootste veroorzakers hiervan.

6.2 Analyse

Voor het analyseren van de verzamelde gegevens zullen in eerste instantie de diverse soorten grafieken worden gebruikt, omdat deze veel meer zeggen dan de individuele getallen kunnen. Voor de analyse van deze grafieken gelden de gebruikelijke vuistregels:



Figuur 6.3: Plotselinge belastings verandering.

Pieken en dalen die uit de globale lijn "springen" moeten verder worden bekeken

Vreemde patronen in de lijn van de grafiek zullen verklaard moeten worden

Scheve verhoudingen in taartgrafieken zullen eveneens verklaard moeten worden

Nadat de analyses zijn uitgevoerd moet er kritisch naar de uitkomsten hiervan worden gekeken en worden bepaald of er maatregelen getroffen kunnen worden.

6.2.1 Pieken en dalen

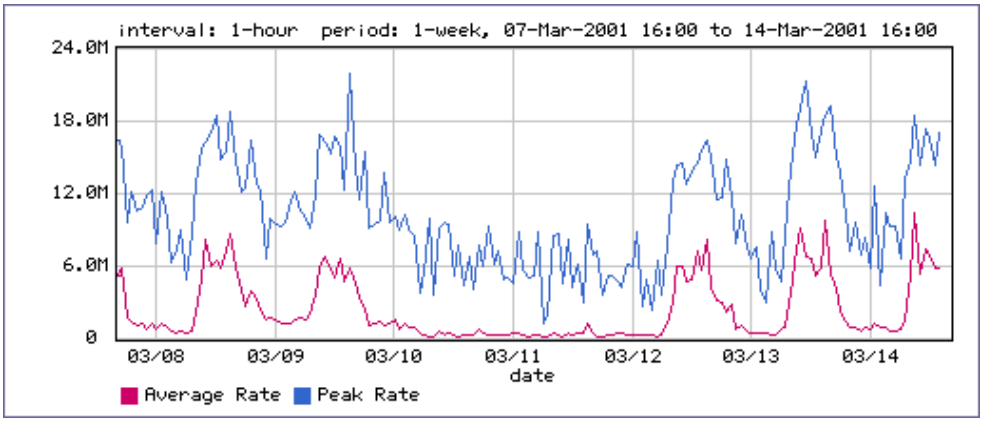
Pieken en dalen in grafieken zullen, zeker als ze buitensporig groot zijn of meer dan incidenteel zijn, verder moeten worden uitgezocht naar wat de oorzaak hiervan is. Wanneer er geen verdere meetgegevens hierover bekend zijn, dan zal er in de organisatie zelf moeten worden gezocht. Een eenmalig hoge piek is meestal niet de moeite waard om uit te zoeken. Ook een plotselinge continue verhoging of daling van het gebruik zal verklaard moeten worden.

Onverwacht (langdurig) laag gebruik van een bepaald communicatietype is vaak ook iets wat moet worden uitgezocht. Dit kan duiden op een verkeerd meetproces, een "achterdeur" ergens (zoals een niet bekende Internet aansluiting), of iets anders. Zie bijvoorbeeld Figuur 6.3 waar zichtbaar is dat vanaf het midden van de grafiek de belasting ineens continu ruim de helft hoger is. Zulke verandering zullen moeten worden verklaard.

Bij PricewaterhouseCoopers (zie sectie 7.5) was de belasting van de lijn naar een van de vestigingen enkele weken lang gedurende elke nacht enorm hoog ten opzichte van normaal 's nachts en zelfs overdag. De precieze oorzaak was uit de meetgegevens van de Packetshaper niet vast te stellen. Navraag bij de automatiseringsafdeling leverde echter op dat er als noodvoorziening 's nachts van diverse servers een backup werd gemaakt via die lijn naar de backupvoorziening op een andere lokatie.

6.2.2 Vreemde patronen

Bij applicaties die de hele dag door worden gebruikt zal dit een gebruiksgrafiek opleveren waar vaak een vast patroon per dag in te ontdekken valt. Dit patroon zal dan vaak van dag tot dag en/of iedere week zich herhalen. Zo'n patroon is vaak zelfs nog te verklaren binnen een dag, zoals ochtend, lunchtijd en middag. Wanneer er binnen zo'n vast patroon wat afwijkende vormen in zitten, en dit niet eenmalig is, dan zal ook dit kunnen worden onderzocht. Hetzelfde geldt voor patronen die niet kloppen met de verwachting, zoals applicatiegebruik in het weekend terwijl er niemand aanwezig hoort te zijn.



Figuur 6.4: Uitgaand netwerk gebruik over een week.

In Figuur 6.4 is het netwerkgebruik gedurende een week te zien, dus het totale uitgaande gebruik en niet applicatie specifiek. In het midden, het lage gedeelte, is het weekend. Daar omheen zijn duidelijk de inviduele dagen zichtbaar, evenals de lunchperiode (de ingezakte gedeeltes van de grote pieken).

6.2.3 Scheve verhoudingen

Wanneer de verhoudingen tussen applicatiegebruik zichtbaar worden gemaakt dan dient er kritisch naar deze verhoudingen te worden gekeken. Uit deze verhoudingen zou bijvoorbeeld oneigenlijk veel niet gewenst of niet essentieel applicatie gebruik kunnen blijken. Ook wanneer een bepaalde applicatie een onwaarschijnlijk klein deel in het geheel heeft kan dit leiden tot verder onderzoek.

Wanneer er vooraf is onderzocht welke applicaties er worden gebruikt binnen de organisatie en mogelijk ook hoeveel en door wie, dan heeft dit een zekere verwachting gecreëerd omtrent de verhoudingen hiertussen. Deze verhoudingen zijn nu te vergelijken met de gevonden verhoudingen. Wanneer hier grote verschillen in zitten dan zullen die uiteraard verklaard moeten worden. Iets waar men zich bij het onderzoek vooraf snel in zal kunnen vergissen is de hoeveelheid verkeer die een bepaalde applicatie teweeg brengt.

Bij een vestiging van PricewaterhouseCoopers lag bij een bepaalde vestiging de verhouding tussen een bedrijfskritieke applicatie (Lotus Notes) en WWW verkeer (HTTP) ten opzichte van het geheel op respectievelijk 63 procent en 7 procent. Bij andere vestigingen lagen deze op bijvoorbeeld 44 en 24 procent of op 37 en 21 procent. De conclusie die hier uit moest worden getrokken was dat die vestiging blijkbaar een eigen Internet aansluiting had (wat niet de bedoeling was).

6.3 Doorbelasting van de kosten

Wanneer er een systeem is ingericht om het gebruik van het WAN in kaart te brengen, kunnen er een aantal acties worden ondernomen. Een van de acties die kan worden ondernomen is dit systeem op een regelmatige basis te gaan gebruiken om deze gegevens te verzamelen. Die gegevens kunnen vervolgens worden gebruikt om het gebruik van het WAN in rekening te gaan brengen bij de gebruikers of bedrijfseenheden. Op die manier worden de kosten eerlijk verdeeld over de veroorzakers hiervan, in plaats van via een overhead post of een (wellicht) arbitrair gekozen verdeelsleutel.

Wegens de vele obstakels (zie sectie 5.6.4) die aanwezig kunnen zijn is dit echter lang niet altijd eenvoudig. Er zullen vele afwegingen moeten worden gemaakt alvorens de kosten kunnen worden doorbelast, zoals:

- Hoe wordt er omgegaan met het gebruik van diensten in het netwerk die WAN verkeer kunnen genereren (de indirecte stromen)
- Wat wordt er gedaan met het gebruik van caching diensten
- Hoe worden de niet-toewijsbare gedeelten van het WAN gebruik alsnog toegewezen

Indien het te lastig of ingewikkeld wordt om bepaalde indirecte stromen direct te kunnen doorbelasten, dan moet er naar een alternatief worden gezocht. Het meest voor de hand liggende alternatief is waarschijnlijk het instellen van een tarief voor het gebruik van de dienst waar de indirecte stromen via lopen (zie ook pagina 30). Op die manier kunnen er toch kosten worden doorbelast zonder een ingewikkeld verwerkingsproces te hoeven doorlopen.

Bij caching diensten kan er aan een zelfde oplossing worden gedacht. Er zal dan wel periodiek moeten worden bekeken of de toegewezen kosten voor die dienst niet ver van de daadwerkelijke kosten afwijken. Wijken deze ver af (de toegewezen kosten zullen in principe altijd hoger uitkomen), dan geeft dat in ieder geval aan dat de cache uitstekend wordt gebruikt. Er kan worden beargumenteerd dat de "meeropbrengsten" kunnen worden gebruikt als compensatie voor de kosten om de caching dienst te onderhouden. De kosten voor onderhoud komen anders toch weer via een andere verrekening bij de gebruikers, en op die manier worden dan ook *die* kosten naar rato toegewezen. Maar dit argument is maar tot beperkte hoogte te gebruiken.

Of er nu oplossingen als hierboven beschreven worden gebruikt of niet, er zullen altijd gedeeltes van de meetgegevens overblijven die niet aan een specifieke machine/gebruiker toe te wijzen zijn. Dit is helemaal het geval wanneer ook meetgegevens van de router zelf worden meegenomen. Hierin zitten namelijk ook stukken netwerkverkeer die als "systeemverkeer" kunnen worden beschouwd. Dit systeemverkeer bevat zaken als tijdsynchronisatie, verkeer om te controleren of een systeem aanwezig is, en diverse andere. Er zullen ook typen netwerkverkeer zijn waarvan niet bekend is bij welke applicatie dit hoort. Wanneer dit zo'n klein gedeelte is dat het niet de moeite waard is om te gaan onderzoeken bij welke applicatie of gebruiker dit hoort, dan zal ook dit verkeer bij het systeemverkeer moeten worden gerekend.

Zodra er over kosten en doorbelasten hiervan wordt gepraat, zal er ook kritisch naar beveiliging van in ieder geval de werkstations moeten worden gekeken. Wanneer dat niet goed geregeld is, bestaat de mogelijkheid dat er instellingen worden gewijzigd waardoor bepaald netwerkverkeer op de rekening van anderen komt.

Tariefstelling Wordt er voor het gebruik van een dienst besloten om hier een tarief voor in te stellen, dan rijst de vraag hoe dan de **tariefstelling** moet zijn. Hier kan van alles voor worden verzonnen, maar de eenvoudigste manier is waarschijnlijk om een gemiddelde te nemen.

Over een korte periode kan er worden gemeten (of mogelijk uit de log bestanden worden gehaald) hoeveel verzoeken er naar de dienst toe komen, en hoeveel netwerkverkeer er van de dienst uit naar het WAN toegaat. Door nu eenvoudigweg deze twee op elkaar te delen komt hier een bepaalde gemiddelde hoeveelheid data per verzoek uit. Hiernaast kan er een prijs per verzonden data via het WAN worden berekend door de totale hoeveelheid te delen op de kosten. Door deze twee met elkaar te vermenigvuldigen komt er een tarief per request uit.

Zijn alle benodigde gegevens elke periode beschikbaar dan kan er zelfs voor worden gekozen om ieder periode de tarieven opnieuw vast te stellen. Na verloop van tijd zal blijken of er veel fluctuatie in de tarieven zit, zodat er dan voor kan worden gekozen om eventueel het tarief vast te zetten.

6.4 Kosten effectiviteit van dit proces

Het gehele proces vanaf de identificatie van de informatiebehoefte tot aan het verwerken en analyseren van de gegevens zal in veel gevallen veel tijd kosten. Vooral de eerste keer dat dit proces wordt uitgevoerd gaat er veel tijd zitten in het eerste gedeelte van het proces. Dan moet immers meestal van niets af worden bepaald wat men wil weten en welke gegevens hiervoor nodig zijn. Moet er vervolgens daadwerkelijk in het netwerk zelf worden gemeten, dan kosten de aanpassingen die moeten worden uitgevoerd vele uren tijd aan planning en uitvoering. Ook het tweede gedeelte, de verwerking en de analyse, zullen door hun nieuwigheid en onverwachte dingen nogal wat tijd kosten.

De potentiële besparingen zijn echter ook flink groot, zeker bij de eerste keer dat dit proces wordt uitgevoerd. Is alles eenmaal doorlopen, dan kan al het werk vervolgens veel vlotter verlopen. De aanpassingen zijn immers al gebeurd en de verwerking is nu ook geen onbekend iets meer. Hoe vaker het proces wordt doorlopen hoe meer het een soort monitoring proces wordt en minder tijd gaat kosten.

De vraag of het geheel kosten effectief is hangt eigenlijk voornamelijk af van de bedragen die aan de WAN verbindingen uitgegeven worden. Wanneer er slechts iets van 10.000 gulden per jaar aan een verbinding wordt uitgegeven, zal het weinig zin hebben om enkele duizenden guldens uit te geven aan manuren en andere zaken om vervolgens misschien tien procent kostenbesparing te kunnen realiseren. In dat soort gevallen zal het effectiever zijn om één keer per jaar prijzen van WAN providers te vergelijken of naar andere technologiën over te stappen (zoals momenteel bijvoorbeeld ADSL in plaats van een ISDN-2 verbinding).

Wanneer het over grotere totaalbedragen gaat dan zijn ook de potentiële besparingen of gevolgen van maatregelen groter. Er kan dan ook meer tijd worden gestopt in het verzorgen van een automatische gegevensverzameling en -verwerking, zodat het gehele proces meerdere keren en sneller kan worden doorlopen.

Wanneer er ook wordt gekeken naar *waar* het netwerk voor wordt gebruikt dan zitten de potentiële besparingen niet alleen meer in de kosten van de WAN verbindingen. Afhankelijk van welk (ongewenst) gebruik er wordt gemaakt van het WAN zijn er ook mogelijk besparingen te realiseren bij servers in het netwerk of bij de eindgebruikers zelf. Hier gaat dus het aantal gebruikers in het netwerk een rol spelen: meer gebruikers betekent meer netwerk- en WAN verkeer en meer potentiële besparingen.

Wordt er gekeken naar wie het netwerk voor wat gebruikt dan zal ook de stap naar het doorbelasten van de kosten niet zo groot meer zijn. Voor beide moet er initiëel veel werk worden verricht en zullen de besparingen lang niet altijd direct zichtbaar zijn. Die zullen vooral op langere termijn pas gerealiseerd worden. Het meeste effect zal dit alles hebben bij het management dat over betere stuurinformatie kan beschikken. Ook hier geldt daardoor dat de effecten bij kleine netwerken gering zullen zijn waardoor deze niet opwegen tegen de te maken kosten om die effecten te bereiken.

Concluderend moet er worden gezegt dat de grote van het netwerk bepalend is of het hele proces kosten effectief kan zijn. Waarschijnlijk zal, zodra het netwerk uit meer dan zo'n honderd gebruikers bestaan (die ook actief van het WAN gebruik maken) dat het de moeite waard kan zijn om het proces te doorlopen.

Hoofdstuk 7

Praktijk situaties

7.1 Erasmus Universiteit Rotterdam

De Erasmus Universiteit Rotterdam (EUR) heeft een zeer uitgebreid netwerk in bezit waarvan een aantal gedeeltes zich verspreid over de stad bevinden. Deze zijn echter allemaal aan elkaar gekoppeld middels WAN verbindingen. Naast deze verbindingen is er ook een verbinding met Surfnet, de backbone van alle Universiteiten in Nederland.

De beheerder (J.H. Bovenlander) van het fysieke netwerk verleende geen toestemming om zelf ergens in het netwerk metingen te verrichten. Hij wist echter redelijk goed te vertellen wat er over het netwerk gaat, doordat zij zelf een zeer beperkte statistiek kunnen opvragen over de verdeling van de verschillende typen netwerkverkeer. Hij wist echter direct te vertellen dat het grootste deel (zo'n 80%) van al het verkeer webverkeer (World Wide Web) is. Bovendien wordt de koppeling met Internet via Surfnet volledig gebruikt; dwz. hoeveel capaciteit er ook wordt aangeboden, deze zit altijd vol. Binnenkort wordt er overgestapt naar een Gigabit WAN verbinding naar Surfnet, waarmee het volledig irrelevant wordt of er wel efficiënt gebruik wordt gemaakt hiervan, of om een billingsysteem in te voeren. Momenteel is de financiële situatie ook dermate riant en de bezetting van het netwerk scheef genoeg (80% is al bekend) dat het voor de EUR niet interessant of kosteneffectief is om in de vorm van dit onderzoek iets aan Network (Cost) Control te doen.

7.2 TU Delft

Net als de EUR heeft ook de Technische Universiteit Delft (TUD) een enorm netwerk met zelfs nog meer WAN verbindingen als de EUR. De beheerder (P. de Nie) geeft ook hier aan niet bereid te zijn om het een en ander te herconfigureren om de gewenste gegevens naar boven te halen. De metingen die nu reeds worden uitgevoerd blijven beperkt tot belastings gegevens van de diverse routers en switches. Het bekijken van welk soort data er over het LAN gaat wordt alleen gedaan bij zeer ernstige problemen. Een demo van ophalen van wat belastings gegevens van 1 switch kostte bijna een halve minuut, ondanks de backbone van 155Mbps.

Het netwerk van de TUD bestaat volledig uit Cisco apparatuur. De monitoring software is een combinatie van HP OpenView en CiscoWorks.
Een overzicht van de netwerk apparatuur is ook met HP OpenView gemaakt met behulp van SNMP discoveries, gevolgd door een hoop handmatig werk om het gegenereerde plaatje te organiseren. Hierop bleek dat het netwerk bestaat uit enkele honderden switches en routers.

Onderverdeling van belastings gegevens op gebouw niveau gebeurt wel, door middel van

de gegevens van de toegangsrouter van dat gebouw, maar enkel en alleen om te controleren of de beschikbare capaciteit voldoende is. Verder wordt er niet gegaan. Capaciteit wordt toegewezen en kosten worden doorbelast op basis van het aantal studenten dat bij een bepaalde faculteit (lees: gebouw) hoort.

7.3 Een klein proefopstelling

In een zeer klein netwerk van slechts drie machines, waarvan er één een verbinding met Internet heeft is gedurende korte tijd proef gedraaid met de Network accounting daemon. Door nu diverse activiteiten op die machines uit te voeren, zoals WWW browsen, NFS en Samba gebruiken, en nog een aantal toepassingen, werden diverse meetgegevens gegenereerd.

Met een gekorte versie van die meetgegevens is vervolgens geprobeerd om deze te aggregeren naar begrijpbaarder formaat. De initiële opslag werd in een database gedaan, conform de suggesties in sectie 5.7. Omdat de meetgegevens naast LAN-naar-WAN ook lokale transacties bevatten was het nodig om deze te scheiden. Doordat er binnen SQL niet erg veel in één enkele query kan worden uitgevoerd waren er hier relatief veel queries nodig, wat de duidelijkheid niet ten goede kwa.

Dezelfde handelingen kunnen echter ook in een zeer simpel programmatje worden gedaan wat direct op de logfile kan werken. Met veel minder code kan nu hetzelfde resultaat worden bereikt.

Het is nu ook mogelijk om per werkstation een overzicht te maken van welke diensten intern gebruik is gemaakt.

Op deze manier kun je ruwe meetgegevens zoals:

Timestamp	Type	Src	SrcPort	Dst	DstPort	Bytes	I/F	User
970135052	1	198.1.2.204	0	198.1.2.200	0	104	eth0	unknown
970135052	6	198.1.2.200	1036	198.1.2.204	8080	399	eth0	unknown
970135052	6	198.1.2.204	8080	198.1.2.200	1036	14311	eth0	unknown
970135052	6	194.159.73.2	46706	212.238.52.39	25	4352	ppp0	unknown
970135052	6	212.238.52.39	25	194.159.73.2	46706	1017	ppp0	unknown
970135052	17	212.238.52.39	1025	207.69.194.186	53	71	ppp0	unknown
970135052	17	207.69.194.186	53	212.238.52.39	1025	259	ppp0	unknown
970135052	6	212.238.52.39	4923	194.159.73.2	113	44	ppp0	unknown
970135052	6	194.159.73.2	113	212.238.52.39	4923	40	ppp0	unknown
971269938	17	198.1.2.204	753	198.1.2.200	2049	320	eth0	unknown

verwerken tot meerdere tabellen, zoals lokaal-servers, lokaal-WAN, overzicht gebruikte services, enz. Dit zijn een aantal tabellen die gemaakt kunnen worden:

Gebruikte services:

Src	Dst	Port
198.1.2.200	198.1.2.204	0
198.1.2.200	198.1.2.204	123
198.1.2.200	198.1.2.204	513
198.1.2.200	198.1.2.204	741
198.1.2.200	198.1.2.204	745
198.1.2.200	198.1.2.204	753
198.1.2.200	198.1.2.204	757
198.1.2.200	198.1.2.204	761
198.1.2.200	198.1.2.204	1023
198.1.2.200	198.1.2.204	8080
198.1.2.205	198.1.2.204	0
198.1.2.205	198.1.2.204	53
198.1.2.205	198.1.2.204	119
198.1.2.205	198.1.2.204	8080

Gebruikte caching services:

Src	Dst	Port	Service
198.1.2.200	198.1.2.204	8080	Proxy
198.1.2.205	198.1.2.204	53	DNS
198.1.2.205	198.1.2.204	8080	Proxy

Lokaal naar Servers:

Source	Server	Bytes
198.1.1.200	198.1.1.255	1916
198.1.2.200	198.1.2.204	4445817
198.1.2.200	198.1.2.204	6347887
198.1.2.205	198.1.2.204	29031
198.1.2.205	198.1.2.204	250016
198.1.2.205	198.1.2.255	1770
198.1.2.205	224.0.0.2	148

Lokaal naar WAN:

Src	Dst	Bytes
198.1.2.200	130.115.1.1	1140

Deze en andere tabellen leveren nu eigenlijk al voldoende informatie om een eenvoudige vorm van gebruiksanalyse te doen, of zelfs om al iets aan doorbelasting van kosten te doen.

Hoewel dit slechts een zeer simpel voorbeeld is, is wel gebleken dat met weliswaar enige initiële moeite er behoorlijk snel toch zeer waardevolle informatie kan worden onttrokken.

7.4 KPN: billing op de vaste verbindingen

De KPN biedt aan zijn klanten diverse soorten verbindingen aan, waaronder ook vaste verbindingen met bijvoorbeeld het Internet. De kosten voor zo'n lijn bestaan uit een vast bedrag en een variabel bedrag welke wordt bepaald door de frequentie en duur dat de klant de afgesproken bandbreedte overschrijd.

Aan elke lijn zit (uiteraard) een router, zowel aan de kant van de klant als aan de kant van de KPN. Elke 5 minuten wordt er van elke router aan de KPN zijde (ca. 1300) gegevens opgevraagd met betrekking tot de huidige belasting. Deze gegevens gaan onder andere een database in. Op basis van deze gegevens worden er notas gemaakt. Bij de betreffende lijnen is er sprake van een CIR. Elk meetpunt dat boven de CIR ligt kost de klant extra geld (bovenop de vaste kosten voor de verbinding). Het zou dus het mooiste zijn voor de KPN om elke seconde te meten. Dit zou echter een te grote belasting voor de routers betekenen, en bovendien niet haalbaar zijn met 1300 routers.

Een voorbeeld van de belasting werd getoond van één van de bekendere Internet providers in Nederland. In de belastingsgrafiek waren bepaalde patronen goed zichtbaar. Op dag niveau was duidelijk de nacht herkenbaar (bijna geen belasting), oplopend naar het maximum rond lunchtijd. Tijdens de lunch was een kleine dip zichtbaar waarna de belasting hoog blijft tot het einde van de middag. Dit patroon was bijna elke werkdag hetzelfde, met uitzondering van de vrijdag waar de belasting eerder in de middag zakt (mensen die vroeg naar huis gaan). Op de zaterdag is de belasting overdag redelijk hoog met wat fluctuaties. De zondag is wat vreemder met een hoge belasting aan het begin van de dag tot vroeg in de middag, waarna de belasting flink zakt. Aan het begin van de avond is er vervolgens weer een piek te zien die echter kort duurt. Deze patronen van dag tot dag herhalen zich vervolgens netjes elke week weer. De maximum belasting zat tegen de 100Mbit aan.

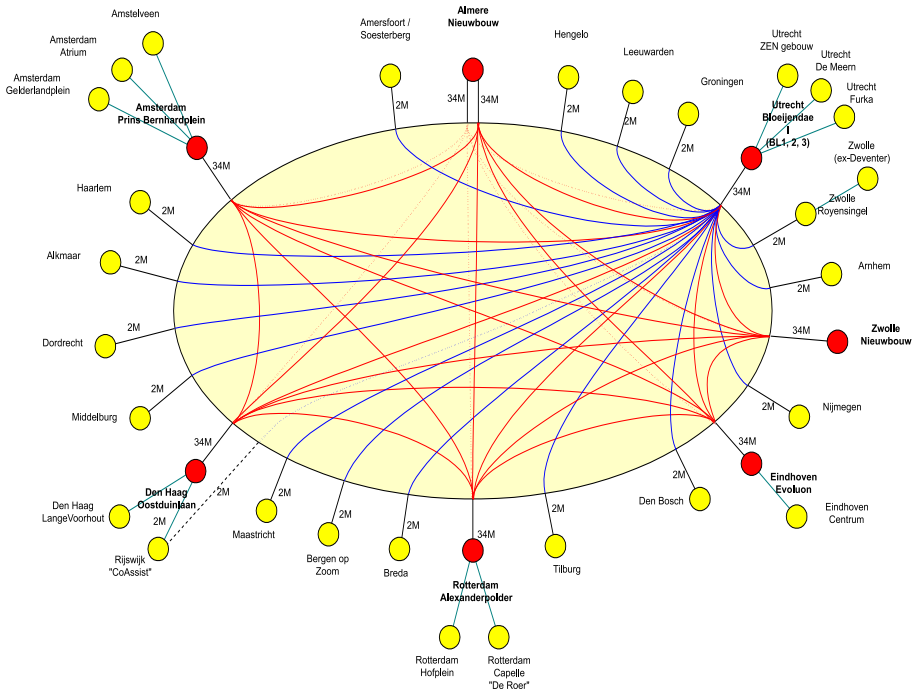
7.5 PricewaterhouseCoopers

PricewaterhouseCoopers (PwC) heeft in Nederland op veel lokaties vestigingen. Al deze vestigingen zijn gekoppeld met een landelijk netwerk, dat zelf weer gekoppeld is aan een wereldwijd netwerk. De vestigingen (ongeveer 35) zijn landelijk verbonden middels ATM of Frame Relay verbindingen. Op lokaal niveau (binnen een stad) is er één knooppunt dat aan het landelijke netwerk is verbonden. De andere vestigingen zijn verbonden met diverse soorten MAN verbindingen, variërend van huurlijnen van een telecom maatschappij tot in opdracht aangelegde kabels en straalverbindingen. De aansluitingen zijn ook niet allemaal bij dezelfde maatschappij afgenomen.

De snelheden van de diverse verbindingen zijn niet allemaal gelijk, omdat uiteraard niet iedere vestiging evenveel werknemers herbergt (bijvoorbeeld 23 in Haarlem tegenover bijna 2000 in Amsterdam) en daarmee ook andere hoeveelheden netwerkverkeer veroorzaken. Er is één vestiging (Utrecht) die als centraal punt voor het netwerk geldt, bijna alle aansluitingen hebben alleen een Virtual Circuit naar Utrecht toe. Die aansluitingen hebben allemaal een maximum snelheid van 2Mbps. De overige aansluitingen zijn "hoofdaansluitingen" welke allemaal met elkaar verbonden zijn, en hebben ook een veel grotere maximum snelheid (34Mbps) dan de anderen. De maximum snelheden zijn niet noodzakelijkerwijs de snelheid die gebruikt wordt. De hoofdaansluitingen zijn dubbel uitgevoerd, zodat er één aansluiting kan uitvallen zonder problemen te veroorzaken. De andere aansluitingen hebben deze voorziening niet.

Nota's

Op notas van providers staat de CIR en de overschrijding hiervan. De overschrijdingen worden niet gecontroleerd aan de eigen gegevens, maar de overschrijdingen zelf worden wel



Figuur 7.1: PwC WAN

in de gaten gehouden voor grote veranderingen. Deze zouden immers kunnen betekenen dat er mogelijk een hogere CIR moet worden afgesloten, of in ieder geval dat er moet worden bekeken waarom er zoveel verkeer is.

Packetshapers

Op alle verbindingen staan Packetshapers welke tot nu toe voornamelijk stonden te “leren” wat er allemaal gebeurt over de verbindingen. Er zijn wel instellingen gedaan, maar deze zijn allesbehalve definitief. De voorzetten hiervoor zijn gedaan door een extern bedrijf aan de hand van een matrix met activiteiten en hoeveelheden. Deze hoeveelheden zijn bepaald door een combinatie van ervaring, gevoel, enige observatie en natte vinger werk. De instellingen zullen tezijnertijd worden aangepast aan de echte situatie. Daarna zullen deze regelmatig moeten worden gereviewed en eventueel aangepast.

Met onbekende verkeerstypen wordt eigenlijk niets gedaan. Bekende typen wordt wel naar gekeken maar er wordt geen indruk gegeven dat er echt iets mee wordt gedaan. Sinds het installeren van de Packetshapers hebben deze wel staan “leren” welk soort verkeer er langs komt. Nu deze dit een tijdje hebben gedaan wordt er vanuit gegaan dat de meeste soorten nu zijn herkend, en wordt deze leer stand weer uitgeschakeld. Eventueel nieuwe soorten verkeer zullen dus onherkend blijven en op de grote hoop “overig” komen.

Op individueel gebruik wordt niet echt gekeken. Wel worden er top-talkers en top-listeners bijgehouden voor bepaalde protocollen (zoals HTTP) en daar wordt soms naar gekeken, maar dieper dan dat gaat men niet. Er wordt dus niet gekeken of de activiteiten van die toppers wel ”gepast” zijn, evenmin of er notoire veelverbruikers zijn.

Doorbelasting van het gebruik

Individueel/groeps gebruik wordt eveneens niets mee gedaan. IP adressen worden uit een pool uitgegeven per vestiging, maar niet per afdeling oid. Op die manier opdelen werkt momenteel dus niet. Wellicht dat de behoefte een keer komt maar nu is die er niet en wordt er dus niets aan gedaan. Bovendien voelt men er niets voor om dat uit te voeren:

ze zijn techneuten en geen administrateurs. Indien het geheel automatisch zou kunnen worden gedaan dan zou de interesse er wel zijn.

Enige vorm van activity tracking of mogelijke doorbelasting wordt geheel gezocht in de Notes sfeer. Daar wordt het een en ander bijgehouden. Eventuele doorbelasting van kosten zou dan kunnen gebeuren op basis van activiteiten in Notes. De meeste (netwerk) activiteiten gebeuren toch in Notes (en WWW). Momenteel gebeurt er dus niets voor wat betreft doorbelasting op basis van gebruik.

De verbindingen

De WAN en MAN verbindingen bestaan uit allerlei soorten. Op diverse locaties zijn er oplossingen anders dan een gehuurde lijn. Dit is meestal doordat het slechts een tijdelijke locatie (tijdelijk kan ook een aantal jaar zijn !) is, of doordat "standaard" oplossingen niet beschikbaar zijn. De "eigen kabels" die op diverse locaties liggen zijn ook gehuurde kabels, echter deze zijn dan in opdracht aangelegd. Dit kan echter wederom flink wat kosten. In Den Haag is voor een periode van ongeveer drie jaar een verbinding aangelegd (glasvezel, 1Gbit/s) welke initieel ongeveer 600.000 gulden heeft gekost. Daarnaast wordt er dan ook nog maandelijks hiervoor betaald. De straalverbindingen op een aantal lokaties zijn eveneens gehuurd.
Zo even grofweg gerekend wordt er maandelijks ruim 250.000 uitgegeven aan de diverse verbindingen, jaarlijks dus zo'n drie miljoen.

Een inzage in een uitgebracht advies met betrekking tot uitbreiding en vervanging van het toenmalige WAN bracht een aantal interessante zaken naar voren. Uit het advies bleek dat ondanks diverse boete het een aanzienlijke besparing opleverde om over te stappen naar een andere WAN leverancier. Het verschil in kosten over een periode van drie jaar, zelfs na aftrek van de boetes, was ruim één miljoen gulden, bijna twintig procent van het totaal. Ook bij een uitgezet migratie traject blijkt het verschil tussen de twee aanbieders nog zo'n 16 procent te zijn. In de overweging werden uiteraard wel meer dan alleen het financiële verschil meegenomen.

Traffic Matrix

De lijncapaciteiten voor het huidige WAN zijn voornamelijk bepaald aan de hand van de gegevens die de eerste Packetshaper (in Utrecht) in het begin heeft opgeleverd. De pieken in het verkeer bepaalden de maximum waarde voor een verbinding, en ongeveer het maximum van de gemiddelde doorvoer bepaalt de "gewone" capaciteit (bij Frame Relay dus respectievelijk de PIR en de CIR).

Een activiteit die in het kader van dit onderzoek kan worden uitgevoerd is het controleren en bijwerken van de getallen in de Traffic Matrix. Er zijn na enkele maanden genoeg meetgegevens via de inmiddels op elke verbinding geïnstalleerde Packetshapers te verkrijgen om de getallen een realistischere invulling te geven. Dit is te beschouwen als een vorm van capaciteitsbepaling (zie ook sectie 4.1.3).

Ter voorbereiding is er bekeken hoe het zat met het Notes replicatie verkeer van Den Haag naar Utrecht. Er schijnt gezegd te zijn dat er geen replicatie plaats vindt tussen Den Haag en Utrecht waardoor er ook geen waarde in de Traffic Matrix staat. De metingen van de Packetshaper wijzen echter iets geheel anders aan: gemiddeld ruim 80kbps met maximale pieken van 1.5Mbps !

In een eerdere analyse van Packetshaper gegevens viel het totale gebrek aan HTTP verkeer vanuit één vestiging op, terwijl er flink wat ander verkeer zoals Notes aanwezig was. De conclusie die moest worden getrokken was dat de betreffende vestiging blijkbaar *zelf* over een verbinding met Internet beschikt.

Met behulp van de gegevens van de Packetshaper zijn reeds meer interessante vondsten gedaan. Zo bleek dat ondanks vele pogingen om Internet via het EuroWAN tegen te gaan dit op een gegeven moment toch weer mogelijk was (en nog steeds mogelijk is).

De verwerking van de gegevens gaat met behulp van de volgende stappen:

1. Er worden van de Packetshapers belastings gegevens gehaald over een bepaalde periode

2. Van deze gegevens wordt er iedere keer een bepaald percentiel genomen (meestal 90 of 95)

3. De gegevens worden gedeeld door het aantal piek connecties; dit vormt de waargenomen waarde

4. Het aantal werknemers van de betreffende locatie wordt teruggebracht naar een aantal gelijktijdig actieve gebruikers.

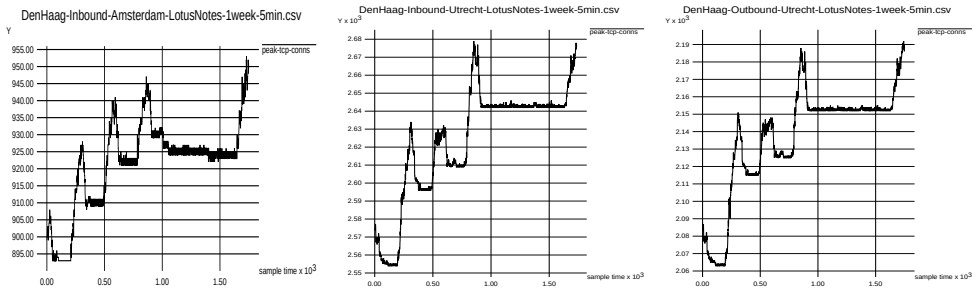
5. De gecorrigeerde belastingsgegevens uit stap 2 worden vervolgens gedeeld door het aantal gelijktijdig actieve gebruikers; dit vormt de berekende waarde.

6. De waargenomen en berekende waarden moeten enigszins bij elkaar in de buurt liggen. Is dit niet het geval dan kloppen blijkbaar de aannames bij stap 4 niet, en zal daar iets mee moeten worden geschoven tot de waarden wel bij elkaar in de buurt komen.

Toelichting hierop:
De 90 of 95 percentiel wordt genomen om de uitschieters (zowel naar boven als naar beneden) uit de metingen te verwijderen. Deze uitschieters zijn niet representatief voor het reguliere netwerkverkeer.
Wanneer er bij een vestiging een x aantal werknemers behoren betekent dit niet dat deze altijd aanwezig zijn, en zeker niet dat zij tegelijkertijd actief zijn. Een gedeelte van hen zal afwezig zijn doordat ze bijvoorbeeld bij klanten zijn, ziek zijn, of gewoon een vrije dag hebben. Niet geheel willekeurig is er gezegd dat van het aantal werknemers ongeveer 70 procent aanwezig zal zijn. Van deze aanwezige werknemers zal zeker niet iedereen tegelijkertijd bezig zijn met dezelfde activiteit (zoals hier Lotus Notes gebruik). Door wat te spelen met wat percentages en de metingen blijkt dit in de buurt van de 8 procent (van die 70 procent aanwezigen) te liggen.
Er wordt wat geschoven met percentages bij de berekeningen, omdat het doel hier is om getallen te vinden zodat het gebruik *binnen* de beschikbare bandbreedte blijft. Zodra er moet worden gekeken of de beschikbare bandbreedte voldoende is, zou men de percentages met rust kunnen laten. Bij de Packetshapers is dat echter veel makkelijker te zien, door naar de totale belasting van de verbinding te kijken. Zit deze te vaak over de CIR of zelfs tegen de PIR, dan is het duidelijk dat de bandbreedte moet worden verhoogd (maar pas nadat er is gecontroleerd of er geen overbodige communicatie tussen zit die de bandbreedte onnodig verbruikt).

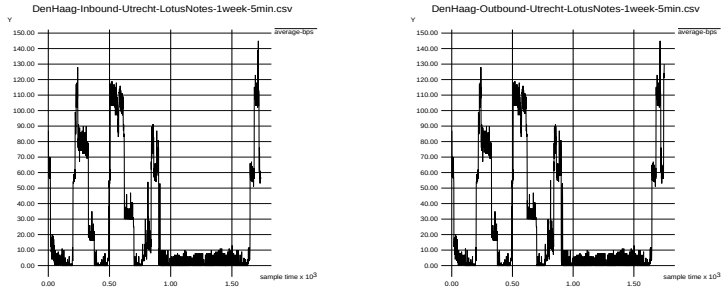
Bij een proef met het verwerken van de meetgegevens viel reeds op dat er een grote hoeveelheid connecties open stonden. Een gedeelde verklaring hiervoor bleek te zijn dat er indertijd meer soorten verkeer onder de betreffende klasse vielen. Toen een aantal soorten in een andere klasse werden ondergebracht stonden er blijkbaar op het moment van activeren een (groot) aantal sessies open. De Packetshaper heeft daardoor nooit enig teken gezien dat de betreffende sessies niet meer bestaan. Wat de Packetshaper betreft staan deze nog immer open.

Een tweede onregelmatigheid kwam naar voren bij het uitzetten van de peak-tcp-conns om te controleren of er nog andere onregelmatigheden in zaten. De resulterende grafiek (met meetgegevens van een week) liet een globale stijgende lijn zien (zie ook Figuur 7.2(a) en Figuur 7.2(b)). Blijkbaar blijven er dagelijks een (redelijk constante) hoeveelheid sessies open staan volgens de Packetshaper. Dit waren ongeveer zo'n 20 stuks per dag. Om een probleem uit de meetgegevens te verwijderen en daardoor een realistischer aantal sessies



(a) (b) (c)

Figuur 7.2: In- en outbound verloop voor peak-tcp-conns.



(a) (b)

Figuur 7.3: In- en outbound verloop voor average-bps.

per dag hieruit te verkrijgen, moet er bij elke dag worden bekeken wat het minimum was, en dit van de meetgegevens van die dag worden afgehaald.

Een van de dingen die we willen bepalen is het bandbreedte gebruik voor deze applicatie per gebruiker. Er hoeft niet naar de classes per lokatie te worden gekeken, het gaat hier immers om bandbreedte gebruik per applicatie en niet per lokatie waarmee gecommuniceerd wordt.

Ook hoeft er niet naar zowel outbound - als inbound verkeer te worden gekeken. Om alles niet nodeloos complex te maken, wordt de aanname gedaan dat inbound - en outbound verkeer niet schrikbarend veel van elkaar verschillen. Dit gaat uiteraard alleen op als uit de meetgegevens blijkt dat ze inderdaad niet veel verschillen. Bij de Packetshapers is dit te controleren aan de hand van de peak-tcp-conns, welke dan voor inbound en outbound ongeveer gelijk dienen te zijn. In figuren Figuur 7.2(b) en Figuur 7.2(c) is te zien dat Inbound en Outbound dezelfde patronen en waarden vertonen. Tussen figuren Figuur 7.3(a) en Figuur 7.3(b) is eveneens een gelijkenis te constateren, maar dan voor de gemiddelde doorvoer. De aanname blijkt dus te kloppen voor Notes verkeer, en er hoeft daardoor alleen maar naar Inbound of naar Outbound te worden gekeken.
Uit de uiteindelijke gegevens kan nu met een beetje inzicht een bandbreedte per gebruiker per sessie worden gedestilleerd.
Uit de berekende en de waargenomen getallen moet een acceptabel getal worden gevormd welke als maatstaf wordt gebruikt.

In beide type grafieken die hier getoond zijn (Figuur 7.2 en Figuur 7.3) valt ook heel netjes een dagelijks patroon te ontdekken. De grafieken zijn enigszins verschoven door

het tijdstip waarop de gegevens zijn onttrokken (Maandag middag). Het grote vlakke stuk zijn de zaterdag en zondag wanneer er nauwelijks iemand werkt. Naar rechts gaand is er een grote piek, welke de Maandag is. De linkerkant van de grafiek begint met de Dinsdag, waarna de pieken volgen van de Woensdag en de Donderdag. De laatste piek is van de Vrijdag welke opmerkelijk anders is als de andere doordeweekse dagen. Dit heeft als oorzaak dat er op vrijdag veel mensen weg blijven van kantoor (ATV-dag of anderszins). Deze patronen blijken zich iedere week weer te herhalen.

Om ook eventuele dag-specifieke zaken te verwijderen is er voor gekozen om het tijdsinterval te vergroten naar een week. Wanneer er nu rekening wordt gehouden met alles wat hiervoor is aangevoerd, dan komen de volgende getallen te voorschijn:

class: Inbound-LotusNotes				
	Berekend		Waargenomen	
	avg-bps	peak-bps	avg-bps	peak-bps
Almere	52	999	390	7395
Amsterdam	1055	9250	1674	14671
Eindhoven	18	298	103	1659
Rotterdam	3926	19358	4912	24216
Utrecht	1565	28985	1139	21101

Zoals redelijk goed opvalt hebben Almere en Eindhoven een verwaarloosbare hoeveelheid verkeer ten opzichte van de andere drie. Het is zeer waarschijnlijk dat deze twee vestigingen (nog) niet rechtstreeks met Den Haag repliceren. Het verkeer dat hier zichtbaar is voor die twee vestigingen zijn dan blijkbaar puur interactief verkeer. En dat zijn dan werknemers van Den Haag die in Almere of Eindhoven aan het werk zijn en hun email lezen, wat weer op een server in Den Haag staat.

Almere en Eindhoven kunnen daarmee buiten beschouwing worden gelaten. Wanneer we dan naar de gemiddelde doorvoer gaan kijken, dan is te zien dat 2000bps toch wel het minimum is dat er gegarandeerd moet worden. De meetgegevens geven weer dat er gemiddeld per sessie zo'n 2500bps doorvoer is geweest $(1055 + 3926 + 1565)/3)$, en dat is dan zonder dat er een beperking aan doorvoer is geweest. Deze snelheid is dus blijkbaar de gemiddelde werksnelheid; bij minder snelheid zal het worden ervaren als langzaam.

Door het grote aantal samples (meer dan 1700) en sessies, is het statistisch verantwoord om de 95-percentiel waarden van avg-bps en peak-bps te delen door de 95-percentiel waarde voor het aantal peak tcp connections. Delen we de avg en peak bps beide per sample door de bijbehorende peak-tcp-conns, dan blijft resultaat hier gelijk. De waargenomen avg waarde wordt dan $(1025 + 5021 + 1933)/3 = 2660$ tegenover 2575 in de bovenstaande tabel.

Gaan we naar de piek doorvoer kijken dan liggen de getallen een stuk hoger. Pieken zijn normaal bij veel soorten communicatie. De grootste uitschieters zijn reeds uit de metingen gehaald door met percentielen te werken. De pieken die overblijven zijn dan "normale" werkpieken. Deze zullen daarmee dus wel de bandbreedte beschikbaar moeten hebben om een normale werksnelheid te kunnen waarborgen. Uit de getallen blijkt dat 20kbps iets is dat zeker rekening mee moet worden gehouden, en waarschijnlijk is het dan zelfs beter om van 25kbps uit te gaan om wat meer pieken te kunnen herbergen.

Totaal bezien moet er worden gezegd dat 2.5kpbs het minimum is dat per sessie moet worden gegarandeerd, en 25kbps als maximum per sessie kan worden aangehouden. Uitgaande van de 70% van het aantal werknemers is aanwezig en 10% van die werknemers is simultaan bezig, dan komt dit totaal op een minimaal te reserveren bandbreedte van $44 \times 2500 = 110kbps$ en een maximum bandbreedte van $44 \times 25000 = 1.1Mbps$.

Deze 1.1Mbps is aanzienlijk hoger dan die 400kbps en is in ieder het gevolg van een foute inschatting van het volume. Een andere oorzaken zal zeker ook de groei in de bezetting van Den Haag zijn (van 577 naar 793, een stijging van 37%).

In de Traffic Matrix stond nu slechts 400kbps. Hoe is deze enorme toename te verklaren ? Als eerste was er in de oorspronkelijke matrix alleen voorzien in replicatie verkeer tussen Den Haag en Rotterdam, en niet met de andere vestigingen. Het aantal werknemers in Den Haag is met 37 procent gegroeid van 577 naar 793 waardoor de hoeveelheid mail verkeer eveneens is toegenomen. Daarnaast vertoont email gebruik in zijn algemeen al een stijgende lijn evenals de hoeveelheid data die wordt meegezonden (documenten, plaatjes, etc.), zodat ook dit voor een toename zorgt. Net als mail groeit ook de hoeveelheid databases in de loop der tijd, waardoor er eveneens meer gerepliceerd moet worden.

Een andere factor kan zijn dat het Notes gedeelte dat nu nog grotendeels via IPX liep (Notes Database Access) wordt verplaatst naar TCP/IP. Dit houdt in dat alle nieuwe machines (waaronder dus van nieuwe medewerkers) TCP/IP verkeer veroorzaken dat onder de LotusNotes klasse valt, terwijl dit bij de “oude” machines nog via IPX verloopt. Deze factoren bij elkaar hebben daardoor gezorgd voor de enorme toename in het LotusNotes verkeer.

Hoofdstuk 8

Conclusies en aanbevelingen

De doelstelling van deze scriptie, het samenstellen van een methodiek ter beheersing van gebruik en kosten van WAN verbindingen, heeft reeds in Hoofdstuk 2 vorm gekregen door het introduceren van een stappenplan. In de daarop volgende hoofdstukken werden al die stappen uitgewerkt tot zoveel mogelijk concrete handelingen.
Voor de duidelijkheid wordt het stappenplan hier nog eens herhaald:

1. Identificeren van de gewenste informatie
2. Bepalen van de benodigde meetgegevens
3. Bepalen waar deze gegevens vandaan moeten worden gehaald
4. Vaststellen hoe deze gegevens moeten worden verzameld
5. Verzamelen van de gegevens
6. Gegevens aggregatie
7. Gegevens analyse
8. Conclusies en maatregelen

De eerste stap is iets dat in eerste instantie door het management zelf zal moeten worden uitgevoerd, waarna de tweede stap door de "techneuten" verder kan worden uitgewerkt naar de concrete gegevens die nodig zijn.

Bij stappen drie, vier en vijf komt initieel het meeste werk kijken en eveneens de meeste problemen die moeten worden opgelost. Er werd een poging gedaan om de oorzaak van die problemen duidelijk te maken. Vervolgens werd voor een aantal concrete, maar veel voorkomende, toepassingen de mogelijke problemen specifiek toegelicht en oplossingen hiervoor aangedragen.

Hoe er daadwerkelijk meetgegevens verzameld moeten worden is ook aan bod gekomen en een aantal voorstellen voor te gebruiken tools zijn gedaan. In elke specifieke situatie zal er echter toch moeten worden bekeken wat daar de beste oplossing is. Het is bijna onmogelijk om één specifieke tool aan te wijzen, doordat elk netwerk net als de gegevensbehoefte weer totaal anders kan zijn.

Het uiteindelijke verwerken van de meetgegevens tot informatie waar een analyse op uit kan worden uitgevoerd is in Hoofdstuk 6 aan bod gekomen. Zonder concrete gegevens is het echter lastig om specifieke zaken naar voren te brengen. Er is een poging gedaan om duidelijk te maken wat voor soort moeilijkheden men kan tegenkomen en wat daar aan gedaan kan worden. Er zijn een aantal voorbeelden gegeven van hoe de gegevens gepresenteerd kunnen worden voor de analyse, en waar men bij een analyse naar dient te

kijken. Alle mogelijkheden aangeven is ook hier weer onmogelijk, voornamelijk omdat dit soort analyses geen “exact science” is en er veel gevoelsmatig zal worden omgegaan met de analyses. Het verdiend dan ook sterk de voorkeur om bij de analyses mensen te betrekken die veel inzicht hebben in zowel de werkzaamheden van de mensen in het netwerk, als het een en ander van de techniek van het netwerk weten.

Wanneer er goed met de verzamelde meetgegevens wordt omgegaan kan dit een bron van informatie zijn voor zowel het management als voor de mensen die het netwerk onderhouden. Zij kunnen dan uit dezelfde gegevens informatie halen die op hun wensen is toegespitst. Feitelijk is het geheel van aggregatie en analyse een stuk data-mining met alle voor- en nadelen die daar bij horen.

Bij het toelichten van de analyse in Hoofdstuk 6 en bij het bepalen van de informatiebehoefte in Hoofdstuk 4 werden er reeds diverse hints gegeven over conclusies die getrokken kunnen worden uit de analyses. De maatregelen die hiermee getroffen kunnen worden zullen per situatie verschillen en daar valt hier dan ook weinig concreets over te vermelden. Wel moet worden bedacht dat met het nemen van eventuele maatregelen men nog niet klaar is. Het effect van die maatregelen moet wel zichtbaar worden gemaakt, zodat het stappenplan weer doorlopen zal moeten worden. Besturen is immers een proces van informatie verzamelen, verwerken en gebruiken.

Het gehele onderzoek is grotendeels gevoed uit eigen ervaring met het inrichten en onderhouden van netwerken. Om diverse redenen bleek het onmogelijk het eigen stappenplan in zijn geheel aan de praktijk te toetsen. De meeste stappen hebben echter wel individueel een praktische toets gehad, onder meer in een stuk praktijk bij de automatiseringsafdeling van PricewaterhouseCoopers welke in het vorige hoofdstuk uitgebreid is beschreven. Was de mogelijkheid er geweest dan was het uiteraard mooi geweest om alles in zijn geheel uit te kunnen voeren, zeker wanneer er ook sprake zou zijn van kostendoorbelasting op basis van (individueel) gebruik waarbij echt alle aspecten uit dit onderzoek volledig worden gebruikt.

Vooral bij het uitwerken van de diverse problemen bij stappen drie, vier en vijf is al snel gebleken dat de hoeveelheid initieel werk enorm groeit naarmate men meer detail informatie wil hebben. Het zal duidelijk zijn dat het implementeren van dit stappenplan niet iets is dat eventjes snel kan worden gedaan. De toegevoegde waarde van de informatie die uit het proces kan vloeien kan echter wel ontzettend groot zijn. Dit behoeft zich initieel niet eens direct in een financiële vorm te zijn, maar meer in inzicht in het gebruik van het netwerk en van het werkgedrag van de mensen. Dat dit uitstekend als control(e) middel kan worden gebruikt behoeft hier geen verder betoog.

Alles bij elkaar is met dit onderzoek een methodiek gepresenteerd waarmee beheersing van gebruik en kosten van WAN verbindingen kan worden aangepakt. Met behulp van deze methodiek kan er een stuk bestuurlijke informatieverzorging worden ingericht die zich naar verloop van tijd steeds verder laat verbeteren en verfijnen. De praktijk heeft reeds bewezen dat uit deze informatie soms onverwachte zaken tevoorschijn kunnen komen die vervolgens wel uitgezocht zullen moeten worden. Dit soort situaties bewijzen daarmee ook meteen het nut van het toepassen van de methodiek.

Een eventueel vervolg onderzoek zou zich zowel op de technische kant van dit onderzoek kunnen richten, als op de bestuurlijke kant. Bij de technische kant zal de focus liggen bij zowel het verkrijgen van de gegevens uit allerlei bronnen, als bij de opslag en verwerking van die gegevens (stappen 3, 4, 5 en 6 uit het stappenplan). De bestuurlijke kant kan zich richten op het uitdiepen van de mogelijke informatie behoeften, en op maatregelen die getroffen kunnen worden naar aanleiding van de conclusies die getrokken zijn uit de verkregen informatie (stappen 1, 2, 7 en 8 uit het stappenplan).

Bijlage A

Tools

A.1 Networkmapping

Voor het maken van network overzichten bestaan diverse tools. Een aantal hiervan zijn bekeken. Gelet is hoofdzakelijk op hun geschiktheid voor het maken van een logisch overzicht zoals in dit onderzoek bedoeld wordt (zie sectie 5.4).

Cheops Ziet er wel aardig uit, maar bleek niet in staat om alle koppelingen automatisch te maken, of om deze achteraf handmatig te maken. Sommige koppelingen werden ook niet correct gemaakt.

tkIned Draait zowel onder Unix als onder WinNT. Detecteert alle routes correct, maar de initiële overzichten zijn onbruikbaar. Na wat schuiven en hulpmiddelen van het programma gebruikt te hebben, kan er iets overzichtelijks worden gemaakt. SNMP functionaliteit is aanwezig, maar mist (standaard) de mogelijkheid om meetgegevens te combineren met het gemaakte overzicht.

HP Openview Niet zelf getest. Openview gebruikt SNMP queries om apparatuur in het netwerk te kunnen ontdekken. Evenals bij tkIned zijn de initiële overzichten onbruikbaar en moet er veel handmatig werk worden verricht om ze leesbaar te maken.

Er zijn uiteraard nog vele anderen, zoals Visio Professional, CyberCop, Solarwinds, Npulse, NetRecon, en nog veel meer. Het mappen van het netwerk doen zij eigenlijk allemaal met behulp van SNMP.

A.2 Meet tools

Voor het meten in het netwerk van het gebruik zijn zeer veel tools beschikbaar. Ze zijn echter lang niet allemaal geschikt voor de verkrijgen en verwerken van meetgegevens zoals in Hoofdstuk 5 wordt aangegeven. Een kleine selectie tools zijn bekeken en beoordeeld op het detail van de meetgegevens en het gemak waarmee deze gegevens ook buiten de tool om kunnen worden verwerkt.

MRTG Een zeer mooi en uitgebreid freeware pakket is MRTG (Multi Router Traffic Grapher), welke op zowel UNIX als Windows NT kan draaien. Deze levert zijn resultaten als webpagina's met daarin grafieken opgebouwd uit gegevens die via SNMP zijn opgehaald. In de standaard configuratie levert hij echter alleen belastings statistieken. Het pakket is voornamelijk gericht op het omzetten van gegevens in grafieken.

Transcend Network Supervisor Dit pakket is voornamelijk gericht op het in de gaten houden van apparatuur en niet zozeer op meten.

Chevin CNAPro Dit pakket is ook voornamelijk bedoeld als management software zoals HP OpenView, maar kan zelf ook meten vanaf het werkstation. De gegevens die hieruit geexporteerd kunnen worden missen echter weer diepgang. Ze gaan namelijk niet verder dan laag 3 (dus: IPX, IP, enz.).

PacketShaper PacketShaper is een apparaat dat bedoeld is om netwerk verkeer te vormen ("shapen"), maar welke ook gebruikt kan worden om te meten. Hij kan heel mooi onderscheid maken tussen diverse typen verkeer. De meetgegevens kunnen via webpagina's worden bekeken en diverse rapporten worden opgevraagd, en kunnen zonder problemen worden opgehaald en in een spreadsheet of database worden ingelezen.

Twee andere bekende pakketten zijn CiscoWorks en HP OpenView.

Voor het meten op een werkstation zijn er vele pakketten beschikbaar, zowel commercieel als shareware / freeware. Alle bekeken pakketten bieden min of meer dezelfde functionaliteit en verschillen eigenlijk voornamelijk in gebruikersinterface en rapportage mogelijkheden. Van de bekeken tools is SnifferPro de enige die de gegevens netjes kon exporteren.

Voor het meten op één plek zijn er diverse software tools beschikbaar. Gedistribueerd meten (op desktops) met alleen software wordt al lastiger maar is uit te voeren. Het pakket NeTraMet kan dit bijvoorbeeld, maar implementeert een eigen structuur van gegevens hiervoor (welke wel via SNMP kunnen worden overgedragen).

Etherboy Laat op het scherm alleen een activiteiten cirkel zien, maar meet in de achtergrond behoorlijk gedetailleerd. Als output kun je aantal pakketten en bytes verzonden/ontvangen, gespecificeerd per host en type data (ipx, ip, udp, icmp, arp, http, enz.) De gegevens zijn te exporteren, maar komen in een wat ongemakkelijke vorm.

NetXray / SnifferPro Dit pakket is voornamelijk gericht op het analyseren van het netwerk. Meten doet hij echter ook goed, en de gegevens kunnen zelfs netjes naar csv-formaat geexporteerd worden. De uitvoer staat in een redelijk handelbaar formaat.

NeTraMet Ongetest. De documentatie geeft aan dat ongeveer dezelfde gegevens als die binnen RMON (2) zijn gedefiniërd en worden opgeslagen.

Net-Acct Deze meet alleen IP verkeer, maar laat wel al het verkeer gescheiden zien; hij aggregeert niets. De uitvoer kan daardoor behoorlijk groot worden, maar met wat handige tools is dit aanzienlijk terug te brengen, tot de totalen die gewenst zijn.

Bijlage B

Voorbeelden van meetgegevens

B.1 Transactie gegevens

B.1.1 Network Accounting daemon

De net-accounting daemon is een klein programma dat dat onder de diverse UNIX varianten werkt. Zijn output bestaat uit lijsten die er als volgt uit zien:

Timestamp	Type	Src	SrcPort	Dst	DstPort	Bytes	I/F
970135052	1	198.1.2.204	0	198.1.2.200	0	104	eth0
970135052	6	198.1.2.200	1036	198.1.2.204	8080	399	eth0
970135052	6	198.1.2.204	8080	198.1.2.200	1036	14311	eth0
970135052	6	194.159.73.2	46706	212.238.52.39	25	4352	ppp0
970135052	6	212.238.52.39	25	194.159.73.2	46706	1017	ppp0
970135052	17	212.238.52.39	1025	207.69.194.186	53	71	ppp0
970135052	17	207.69.194.186	53	212.238.52.39	1025	259	ppp0
970135052	6	212.238.52.39	4923	194.159.73.2	113	44	ppp0
970135052	6	194.159.73.2	113	212.238.52.39	4923	40	ppp0
971269938	17	198.1.2.204	753	198.1.2.200	2049	320	eth0

De timestamp geeft de tijd aan waarop de betreffende transactie plaats vondt. Het type geeft aan welk protocol er bovenop het IP protocol werd gebruikt voor de transactie. Src, Dst, SrcPort en DstPort geven aan van welk (source) adres en poort er werd gecommuniceerd met welk (destination) adres en poort.

De bovenstaande vorm is de enige vorm waarin het programma zijn gegevens aflevert. Deze gegevens zijn echter in de meeste gevallen genoeg om individueel gebruik vast te kunnen stellen.

B.2 Log bestanden

B.2.1 Webproxy servers

Squid is een zeer populaire caching proxy. Voor wat betreft de opbouw van de log bestanden is Squid erg flexibel. In zijn meest uitgebreid vorm levert hij genoeg gegevens om allerlei vormen van accounting toe te kunnen passen.

In onderstaand voorbeeld zijn de kolommen achtereenvolgens:

1. Timestamp
2. ??
3. Sourceadres van het request
4. Resultaat van het request en resultaatcode
5. Hoeveelheid data in het reply
6. Het request zelf
7. User (degene die het request maakte)
8. Doorverwijsactie en doorverwijzing
9. Type van de data

Log bestand voorbeeld:

1)	982707387.746	121	127.0.0.1	TCP_DENIED/403	996
2)	982707598.050	765	127.0.0.1	TCP_MISS/200	8721
3)	982707618.926	431	127.0.0.1	TCP_HIT/200	8721
4)	982707850.190	366	127.0.0.1	TCP_HIT/200	8722
5)	982707862.181	316	127.0.0.1	TCP_MISS/200	3931

(Vervolg van de regels)

1)	GET http://dirtyhill.rotjeknor.nl	-	NONE/-	-
2)	GET http://dirtyhill.rotjeknor.nl	-	DIRECT/dirtyhill.rotjeknor.nl	text/html
3)	GET http://dirtyhill.rotjeknor.nl	-	NONE/-	text/html
4)	GET http://dirtyhill.rotjeknor.nl	-	NONE/-	text/html
5)	GET http://www.rotjeknor.nl	-	DIRECT/www.rotjeknor.nl	text/html

De acties de gebeuren in de bovenstaande regels zijn:

1. Er werd een webpagina opgevraagd (GET ...), maar de machine waar het request vandaan kwam (127.0.0.1) had daar geen rechten toe (TCP_DENIED). Het antwoord (foutmelding 403) van de proxy was 996 bytes groot.

2. Er werd een webpagina opgevraagd (GET ...), maar de webpagina bevond zich niet in de cache (TCP_MISS). Het verzoek wordt doorgestuurd naar de betreffende webserver (DIRECT/...). Het antwoord van de proxy (de webpagina) was 8721 bytes groot en de webpagina was van het type text/html.

3. Er werd een webpagina opgevraagd (GET ...), de webpagina bevond zich wel in de cache (TCP_HIT). Er is geen verdere actie nodig (NONE/-). Het antwoord van de proxy (de webpagina) was 8721 bytes groot en de webpagina was van het type text/html.

4. De zelfde actie als bij 3 wordt nogmaals uitgevoerd

5. Er werd een webpagina opgevraagd (GET ...), maar de webpagina bevond zich niet in de cache (TCP_MISS). Het verzoek wordt doorgestuurd naar de betreffende webserver (DIRECT/...). Het antwoord van de proxy (de webpagina) was 3931 bytes groot en de webpagina was van het type text/html.

Apache (met mod_proxy) is een van de meest gebruikte webservers op het Internet. Reeds enige tijd bezit ook deze webserver de mogelijkheid om als (caching) proxy te fungeren. De log bestanden van Apache zijn een standaard geworden en kunnen ook een heleboel informatie geven. Wat echter ontbreekt is de cache hit/miss informatie die bijvoorbeeld Squid wel levert.

In onderstaand voorbeeld zijn de kolommen achtereenvolgens:

1. Sourceadres van het request
2. ??
3. User (degene die het request maakte)
4. Timestamp
5. Het request zelf
6. Resultaatcode van het request
7. Hoeveelheid data in het reply
8. Referentiepagina (de huidige webpagina op het moment dat de client het request deed)
9. De identificatiestring van de clientbrowser

Log bestand voorbeeld (de requests zijn hier iets aangepast om alles passend op papier te maken):

1)	127.0.0.1	-	edward	[10/Feb/2001:23:34:43 +0100]
2)	127.0.0.1	-	edward	[10/Feb/2001:23:34:44 +0100]
3)	127.0.0.1	-	edward	[10/Feb/2001:23:34:47 +0100]
4)	127.0.0.1	-	edward	[10/Feb/2001:23:34:47 +0100]
5)	127.0.0.1	-	edward	[10/Feb/2001:23:34:48 +0100]

(Vervolg van de regels)

1)	"GET http://deskpro/el_trigo.dir.plate-1_10.html HTTP/1.0"	200	351
2)	"GET http://deskpro/IMAGES/El-Trigo.1.10.jpg HTTP/1.0"	200	53907
3)	"GET http://deskpro/service-station.html HTTP/1.0"	200	1239
4)	"GET http://deskpro/IMAGES/El-Trigo.Service-Station.jpg HTTP/1.0"	200	57520
5)	"GET http://deskpro/el_trigo.dir.service-station.html HTTP/1.0"	200	401

(Vervolg van de regels)

1)	"http://deskpro/plate-1_10.html"	"Mozilla/4.7 [en] (X11; I; Linux 2.2.13 i686)"
2)	"http://deskpro/plate-1_10.html"	"Mozilla/4.7 [en] (X11; I; Linux 2.2.13 i686)"
3)	"http://deskpro/el_trigo.dir.plate-1_10.html"	"Mozilla/4.7 [en] (X11; I; Linux 2.2.13 i686)"
4)	"http://deskpro/service-station.html"	"Mozilla/4.7 [en] (X11; I; Linux 2.2.13 i686)"
5)	"http://deskpro/service-station.html"	"Mozilla/4.7 [en] (X11; I; Linux 2.2.13 i686)"

De acties de gebeuren in de bovenstaande regels zijn:

1. Vanaf een machine (127.0.0.1) werd door een geauthenticeerde gebruiker (edward) een webpagina opgevraagd (GET ...). Deze pagina werd gevonden (code 200) en aan de aanvrager gestuurd. De pagina was 351 bytes groot. Op het moment dat de browser dit verzoek verstuurde stond deze in de pagina plate-1_10.html, en het de browser die het verzoek deed identificeert zich als "Mozilla/4.7...".

2. Vanuit dezelfde pagina als bij het vorige request wordt nu ook een plaatje opgevraagd (53907 groot).

3. Vanaf de pagina die bij het eerste request is opgehaald wordt nu weer een andere pagina opgevraagd (service-station.html)

4. Vanaf de pagina die bij het vorige request is opgehaald wordt een plaatje opgevraagd

5. Vanaf de pagina die bij het derde request is opgehaald wordt een nieuwe webpagina opgevraagd.

Het is hier niet niet helemaal duidelijk aan te tonen zonder een hele massa logregels toe te voegen, maar wat dan zou opvallen is dat requests niet noodzakelijkerwijs in dezelfde volgorde worden beantwoord als dat ze verzonden zijn. In het bovenstaande voorbeeld (en wat daar nog bijhoort) wordt er iedere keer een pagina opgevraagd, waarna er automatisch weer een andere webpagina wordt opgevraagd evenals een plaatje. Deze constructie bestaat vele malen op de betreffende website maar de log bestanden laten zien dat de ene keer eerst het plaatje wordt teruggestuurd en dan pas de webpagina, en andere keren precies andersom (afgeleid uit de volgorde en timestamp in de log bestanden).

B.2.2 DHCP servers

ISC dhcpd is één van de vele (gratis) verkrijgbare DHCP servers. De server kan flink wat informatie loggen, waaronder de belangrijkste informatie voor het meten.

```
1)  Feb 6 14:07:11  whisper  dhcpd:  DHCPDISCOVER  from  00:10:4b:7e:8e:34
2)  Feb 6 14:07:11  whisper  dhcpd:  DHCPOFFER     on    198.1.2.205
3)  Feb 6 14:07:11  whisper  dhcpd:  DHCPREQUEST   for   198.1.2.205
4)  Feb 6 14:07:11  whisper  dhcpd:  DHCPACK      on    198.1.2.205
5)  Feb 6 14:07:20  whisper  dhcpd:  DHCPREQUEST   for   198.1.2.205
6)  Feb 6 14:07:20  whisper  dhcpd:  DHCPACK      on    198.1.2.205
7)  Feb 6 15:07:19  whisper  dhcpd:  DHCPREQUEST   for   198.1.2.205
8)  Feb 6 15:07:19  whisper  dhcpd:  DHCPACK      on    198.1.2.205
9)  Feb 6 16:07:19  whisper  dhcpd:  DHCPREQUEST   for   198.1.2.205
10) Feb 6 16:07:19  whisper  dhcpd:  DHCPACK      on    198.1.2.205
11) Feb 6 17:07:19  whisper  dhcpd:  DHCPREQUEST   for   198.1.2.205
12) Feb 6 17:07:19  whisper  dhcpd:  DHCPACK      on    198.1.2.205
```

(Vervolg)

```
1)
2)  to    00:10:4b:7e:8e:34  via eth0
3)  from  00:10:4b:7e:8e:34  via eth0
4)  to    00:10:4b:7e:8e:34  via eth0
5)  from  00:10:4b:7e:8e:34  via eth0
6)  to    00:10:4b:7e:8e:34  via eth0
7)  from  00:10:4b:7e:8e:34  via eth0
8)  to    00:10:4b:7e:8e:34  via eth0
9)  from  00:10:4b:7e:8e:34  via eth0
10) to    00:10:4b:7e:8e:34  via eth0
11) from  00:10:4b:7e:8e:34  via eth0
12) to    00:10:4b:7e:8e:34  via eth0
```

In bovenstaande regels is zichtbaar hoe een werkstation zonder IP adres op het netwerk kenbaar maakt dat hij een dhcp server zoekt (DHCPDISCOVER) die hem een IP adres kan geven. Dit bericht wordt door deze server opgevangen en hij stuurt een aanbieding terug (DHCPOFFER voor adres 198.1.2.205). Het werkstation beslist dat hij deze server wil gaan gebruiken en verstuurt het verzoek voor het aangeboden adres naar de server (DHCPREQUEST) terug. De server antwoord hierop (naar 198.1.2.205) met een bevestiging (DHCPACKnowledge).

In de volgende regels is te zien dat het werkstation elk uur opnieuw een verzoek instuurt voor hetzelfde adres. Dit doet hij omdat zijn "huurcontract" voor het adres elk uur afloopt. Bij elke regel is ook te zien met welk hardware adres er wordt gecommuniceerd door de server. Hier is dan dus uit de log bestanden te halen welke machine welk IP adres op welk tijdstip heeft gekregen.

B.2.3 Authenticatie logs

Samba is een stuk Unix software dat zich voor de andere machines in het netwerk voordoet als een Windows NT server. De server biedt een aantal diensten aan zoals printers en fileservices. Hier onder staan een aantal regels uit het log van zo'n Samba server.

```
2001/04/01 17:37:33 pwc (198.1.2.205) connect to service edward as user edward
2001/04/01 18:06:34 pwc (198.1.2.205) closed connection to service edward
2001/04/05 17:56:09 vmwhisper (198.1.2.211) connect to service edward as user edward
2001/04/05 17:57:37 vmwhisper (198.1.2.211) closed connection to service edward
2001/04/05 17:59:29 vmwhisper (198.1.2.208) connect to service edward as user edward
2001/04/05 18:00:19 vmwhisper (198.1.2.208) closed connection to service edward
2001/04/06 00:27:57 whisper98 (198.1.2.206) connect to service public as user nobody
2001/04/06 01:00:56 whisper98 (198.1.2.206) closed connection to service public
```

In dit voorbeeld zijn er een twee soorten regels te herkennen: een “connect” regel, en een “disconnect” of “close connection” regel. In de eerste regel hierboven zien we dat op 1 April 2001 om 17:37 vanaf de machine met IP adres 198.1.2.205 (welke zichzelf adverteert met de naam pwc), er een connectie wordt gemaakt met de service edward (een home directory van een gebruiker). Hiertoe wordt geautenticeerd als gebruiker edward.
De volgende regel laat zien dat deze connectie een half uur later weer wordt verbroken.
Op de volgende 4 regels wordt er twee maal ook een connectie gemaakt met service edward. Echter dit gebeurt van 2 verschillende IP adressen. De betreffende machines identificeren zich echter wel allebei met dezelfde naam.

De laatste twee regels laten zien dat wederom een andere machine een connectie maakt met de service public (publiek toegankelijke mappen en bestanden) als gebruiker nobody (de anonieme gebruiker).

UNIX houdt een redelijk simpel maar voldoende log bij met succesvolle authenticaties. Helaas zijn er services die er toch een eigen authenticatie log op na houden.

edward	ttyp1	Whisper	Thu Feb 1 23:02 - 23:38	(00:36)
edward	ftp	130.1.10.19	Thu Feb 1 23:03 - 23:04	(00:00)
edward	NCP002	002018583d7c	Thu Feb 1 23:06 - 23:11	(00:05)
edward	NCP001	002018583d7c	Thu Feb 1 23:13 - 23:17	(00:03)
edward	NCP001	002018583d7c	Thu Feb 1 23:18 - 23:22	(00:03)
reboot	system boot		Wed Feb 7 23:07	
edward	NCP002	002018583d7c	Wed Feb 7 23:27 - 23:35	(00:07)
edward	ttyp0	198.1.2.205	Thu Feb 8 01:19 - down	(00:06)
edward	ftp	198.1.2.205	Thu Feb 8 01:21 - 01:21	(00:00)
reboot	system boot		Fri Feb 9 22:47	
edward	tty2		Fri Feb 9 22:52 - 00:35	(01:42)
root	tty3		Fri Feb 9 22:53 - 00:35	(01:41)

In de bovenstaande regels valt achtereenvolgens af te lezen:

- Gebruiker edward heeft met telnet ingelogd vanaf een IP adres dat werd vertaald door de DNS server naar de machine naam Whisper.
- Gebruiker edward heeft via ftp ingelogd vanaf IP adres 130.1.10.19 dat niet kon worden vertaald door de DNS server.
- Gebruiker edward heeft 3 maal ingelogd op een NCP connectie (een connectie naar een Novell emulator) vanaf hardware adres 002018583d7c.
- Het UNIX systeem is opgestart
- Gebruiker edward heeft nog 3 maal via een externe connectie ingelogd
- Het UNIX systeem is opgestart
- Gebruiker edward is ingelogd op tty2, een login op het console van het UNIX systeem.
- Gebruiker root is ingelogd op tty3, een login op het console van het UNIX systeem.

Bijlage C

Woordenlijst

ASP Application Service Provider. Een provider die een applicatie als dienst aanbiedt. Dat wil zeggen dat de applicatie die je gebruikt op de apparatuur bij de provider draait.

ATM Asynchronous Transfer Mode

Backbone Het 'netwerk' dat als een soort ruggegraat andere netwerken met elkaar verbindt. Bijvoorbeeld tussen twee gebouwen of tussen twee landen.

Bandbreedte Hoeveelheid dataverkeer per tijdseenheid

BOOTP Boot Protocol. Een protocol dat meestal wordt gebruikt om een machine een IP-adres te geven, en eventueel om hem op te laten starten.

Bridge Ook wel bekend als selectieve repeater. Zie 3.3

CIR Committed Information Rate, de gegarandeerde bandbreedte waar een vast bedrag voor wordt betaald

DHCP Dynamic Host Configuration Protocol. Een protocol dat wordt gebruikt om een machine een IP-adres en eventuele extra parameters te geven.

DNS Domain Name Server. Een dienst die een hostnaam omzet in het bijbehorende IP-adres.

EIR Extended Information Rate, de extra bandbreedte die men mag gebruiken, indien beschikbaar

FR Frame Relay

FDDI Fiber Distributed Data Interface

Hub Centrale schakelkast. Zie 3.3

IEEE Institute of Electrical and Electronics Engineers

IETF Internet Engineering Task Force. Houdt zich bezig met standaarden rondom het Internet.

InterNet Een wereldwijd publiek netwerk.

IntraNet Het (bedrijfs)interne netwerk.

IP Internet Protocol

IPX Internetwork Packet eXchange

ISDN Integrated Services Digital Network

ISP Internet Service Provider. Een provider biedt als dienst de toegang tot het Internet.

MAN Metropolitan Area Network. Dit kan worden gezien als een WAN op kortere afstand; binnen een "metropool".

NSP Network Service Provider. De provider biedt als dienst een netwerk aan. Een aanbieder van een Frame Relay dienst is bijvoorbeeld een NSP.

LAN Local Area Network

PIR Peak Information Rate, de maximum bandbreedte die men mag gebruiken

Proxy Een "tussenpersoon". Een proxy voert iets uit in opdracht van iemand anders.

POTS Plain Old Telephone Service. De gewone telefoon.

RAS Remote Access Service

RFC Request For Comments

RMON device Remote Monitoring device.

SNMP Simple Network Management Protocol

SLA Service Level Agreement

SPX Structured Packet eXchange

TCP Transmission Control Protocol

UDP User Datagram Protocol

WAN Wide Area Network

Bijlage D

Vragenlijsten

D.1 Vragenlijst PwC - GTS

Korte vragenlijst en onderzoeks aanpak. Meeting van 9 Maart 2001

- Wat wordt er nu gedaan bij GTS op het gebied van Network Cost Control ?
  - Welke gegevens worden hiervoor gebruikt.
  - Wat voor analyses en met welke frequentie worden die uitgevoerd.
  - Welke criteria worden er gehanteerd bij het beslissen over aanleg of vervanging van een WAN verbinding.
  - Indien er niets wordt gedaan op dit gebied, waarom dan niet.
- Wat voor zaken zou GTS nog willen weten of onderzoeken en in hoeverre kan mijn onderzoek hierin voorzien ?
  - Welke gegevens zijn er nodig om in die behoefte te voorzien.
  - Kunnen we aan die gegevens komen met behulp van de Packetshaper of andere aanwezige apparatuur.
    - zoniet, wat moet er dan extra gebeuren en mag dat worden uitgevoerd
    - zoja, trek daar dan de gegevens uit,
    verwerk ze tot het gewenste formaat
    analyseer deze gegevens
    en trek eventuele conclusies

Gedachtes vooraf wat er gedaan kan worden:

- Prijs/prestatie
  - Packetsheper doorvoorgegevens of doorvoortest (bijvoorbeeld met ftp)
  - lange termijn doorvoer analyse
  - Is de gewenste capaciteit gelijk aan de huidige capaciteit ?
- WAN uitbreidingen ⇒ welke type verkeer en volume
  - Zijn de toppers hierin gewenst en acceptabel
  - Zijn deze toppers aan te passen door bijvoorbeeld proxy/cache/etc te verhuizen
- Wie gebruikt het netwerk ⇒ Type en hoeveelheid per gebruiker
  - Totaal per groep/gebruiker onderling vergelijken

Bijlage E

Definities

E.1 Informatie en besluitvorming

Decision making The process of defining problems, gathering information, generating alternatives, and choosing a course of action. (ref. [24])

Bestuurlijke Informatieverzorging Alle activiteiten met betrekking tot het systematisch verzamelen, vastleggen en verwerken van gegevens, gericht op het verstrekken van informatie ten behoeve van het besturen-in-engere-zin (kiezen uit alternatieve mogelijkheden), het doen functioneren en het beheersen van een huishouding en ten behoeve van de verantwoordingen die daarover moeten worden afgeleged. (ref [20])

Nemen van beslissingen De *kern* van het besturen bestaat uit het *nemen van beslissingen* met betrekkingen tot de doelstellingen, middelen en werkwijze van de betrokken huishouding, alsmede het nemen van beslsissingen met betrekking tot de te verrichten handelingen: welke handelingen moeten door wie, waarmee, wanneer en hoe worden verricht ? (ref [20], I.2)

E.2 Netwerken en techniek

X.25 X.25 network links are a robust connection between systems, with or without an intervening public network. Suitable for noisy communication environments or connecting to legacy communications equipment, X.25 (and its telephony-oriented cousin, BX.25) provides reliable worldwide communications through public common carrier data networks. (ref [36])

Frame Relay Frame Relay networks offer high transmission speeds and low cost, relative to traditional leased lines and X.25 networks. Suitable for high quality facilities, Frame Relay links are often used as carrier facilities for digital voice or IP connections.

Frame relay is a telecommunication service designed for cost-efficient data transmission for intermittent traffic between local area networks () and between end-points in a wide area network (wide area network). Frame relay puts data in a variable-size unit called a frame and leaves any necessary error correction (retransmission of data) up to the end-points, which speeds up overall data transmission. For most services, the network provides a permanent virtual circuit (Permanent Virtual Circuit), which means that the customer sees a continous, dedicated connection without having to pay for a full-time leased line, while the service provider figures out the route each frame travels to its destination and can charge based on usage. An enterprise can

select a level of service quality - prioritizing some frames and making others less important. Frame relay is offered by a number of service providers, including AT&T. Frame relay is provided on fractional T-1 or full T-carrier system carriers. Frame relay complements and provides a mid-range service between Integrated Services Digital Network, which offers bandwidth at 128 Kbps, and Asynchronous Transfer Mode (asynchronous transfer mode), which operates in somewhat similar fashion to frame relay but at speeds from 155.520 Mbps or 622.080 Mbps.

Frame relay is based on the older X.25 packet-switching technology which was designed for transmitting analog data such as voice conversations. Unlike X.25 which was designed for analog signals, frame relay is a fast packet technology technology, which means that the protocol does not attempt to correct errors. When an error is detected in a frame, it is simply "dropped." (thrown away). The end points are responsible for detecting and retransmitting dropped frames. (However, the incidence of error in digital networks is extraordinarily small relative to analog networks.)

Frame relay is often used to connect local area networks with major backbones as well as on public wide area networks and also in private network environments with leased lines over T-1 lines. . It requires a dedicated connection during the transmission period. It's not ideally suited for voice or video transmission, which requires a steady flow of transmissions. However, under certain circumstances, it is used for voice and video transmission.

Frame relay relays packets at the data link layer of the Open Systems Interconnection (OSI) model rather than at the Network layer. A frame can incorporate packets from different protocols such as Ethernet and X.25. It is variable in size and can be as large as a thousand bytes or more. (ref [36])

Fysiek overzicht Physical diagrams show every component in a network, including the physical connections among them. A physical diagram can help show what pieces are required and demonstrate the connections between components. (ref [15])

Logisch overzicht Logical diagrams show the relationships between network components without going into detail about their physical placement. The figures throughout this book are good examples of the sort of graphical representation that is needed. In general, the more complex the network, the more necessary a visual mapping. (ref [15])

Bibliografie

[1] 3Com. Remote monitoring (rmon). http://support.3com.com, unknown. Corebuilder 3500 Implementation Guide.

[2] Andy Bierman. Remote monitoring mib extensions for differentiated services. Request for comments, The Internet Society, June 2000. draft-ietf-rmonmib-dsmon-mib-02.txt.

[3] Chevin. Chevin cnapro. http://www.chevin.com, 2000.

[4] David W. Crawford. Pricing network usage: A market for bandwidth or market for communication ? *The Journal of Electronic Publishing*, 2(1), May 1996. ISSN 1080-2711.

[5] META Group. Reducing wan total cost of ownership. Technical report, META Group, http://www.metagroup.com, 1997.

[6] C.Mills D.Hirsch G.Ruth. Internet accounting: Background. Request for comments, unknown, November 1991.

[7] N.Brownlee C.Mills G.Ruth. Traffic flow measurement: Architecture. Request for comments, The Internet Society, October 1999.

[8] Steve Kaplan and Marc Mangus. *Citrix MetaFrame for Windows Terminal Services*, chapter 6: Designing your network for server-based computing. McGraw-Hill, http://www.pbg.mcgraw-hill.com/barnesandnoble/jun00/kaplan/chap06.html, 2000.

[9] Hari Balakrishkan Srinivasan Seshan Mark Stemm Randy H. Katz. Analyzing stability in wide-area network performance. In *Sigmetrics 1997*, http://www.seshan.org/papers/conference/sigmetrics97/html/sigmetrics97.html, 1997.

[10] Apogee Networks. Netcountant product brochure. http://www.apogeenet.com.

[11] Apogee networks. Netscope: Real-time network monitoring & performance analysis. http://www.apogeenet.com, 2000.

[12] NISS. Hei requirements for management of network charging. Technical report, Joint Information Systems Committee, http://www.niss.ac.uk, 2000. http://www.jisc.ac.uk.

[13] Npulse. Npulse. http://www.horsburgh.com/h_npulse.html, 2000.

[14] Thobias Oetiker. Multi router traffic grapher. http://www.ee.ethz.ch/stats/mrtg/, 1997.

[15] Oracle. Understanding sql*net. *http://info-it.umsystem.edu/oradocs/doc/net/doc/NWUS233/ch3.htm*, 2000.

[16] David Rogerson Barry Ladbrook Mark Donnely Nick Owen. Implementing cost based interconnect. Technical report, Ovum, http://www.ovum.com/reports/cost_based_interconnect.htm, 1999.

[17] Packeteer. Packeteer packetshaper. http://www.packeteer.com, 2000.

[18] Check Point. The intelligent queueing (iq) engine. Technical report, Check Point Software Technologies Ltd., http://www.checkpoint.com/products/floodgate-1/iqengine.html, 2000.

[19] Check Point. Key requirements for bandwidth management. Technical report, Check Point Software Technologies Ltd., http://www.checkpoint.com/products/floodgate-1/iqkeyreq.html, 2000.

[20] Prof. E.J. Joels R.A. Prof. R.W. Starreveld R.A., Prof. Drs. H.B. de Mare R.A. *Bestuurlijke Informatieverzorging, deel 1.* Samson Bedrijfsinformatie, 4e druk edition, 1994.

[21] Bruce Robertson. Traffic shaping: Assuring application performance. *http://www.networkcomputing.com/822/822colrobertson.html*, December 1997.

[22] Juergen Schoenwaelder. Cheops network user interface. http://wwwsnmp.cs.utwente.nl/šchoenw/scotty, 1999.

[23] Limor Schweizer. Meeting the ip network billing challenge. *TMCnet*, 2000. http://www.tmcnet.com.

[24] Hellriegel & Slocum. *Management*. International Thomson Publishing, 7th edition, 1996.

[25] Mark Spencer. Cheops network user interface. http://www.marko.net/cheops, 1999.

[26] SURFnet. Surfnet tarieven. http://www.surfnet.nl/publicaties/aansluiten/tarieven.html, 2000.

[27] S.Waldblusser. Remote network monitoring management information base. Request for comments, The Internet Society, June 1999. draft-ietf-rmonmib-rmon2hc-00.txt.

[28] S.Waldblusser. Application performance measurement mib. Request for comments, The Internet Society, July 2000. draft-ietf-rmonmib-apm-mib-01.txt.

[29] S.Waldblusser. Remote network monitoring management information base. Request for comments, The Internet Society, May 2000.

[30] R.Waterman B.Lahaye D.Romascanu S.Waldblusser. Remote network monitoring mib extensions for switched networks. Request for comments, The Internet Society, June 1999.

[31] Cisco Systems. Resource reservation protocol (rsvp). Technical report, Cisco Press, http://www.cisco.com/cpress/cc/td/cpress/fund/ith2nd/it2443.htm, 2000.

[32] A.S. Tanenbaum. *Computer Networks*. Prentice Hall, 2nd edition, 1995.

[33] unknown. Remote network monitoring (rmonmib) charter. http://www.ietf.org/html.charters/rmonmib-charter.html, 1999.

[34] unknown. Network management requirements. http://www.gate.net/ ksiles/netman/requirements.htm, unknown.

[35] Steven Waldbusser. Remote network monitoring management information base version 2. http://www.ietf.org, 1999.

[36] ZDnet. www.whatis.com. http://www.whatis.com, 2000.

Index