

‘SCSI over IP’ (iSCSI), een van de drie standards om gegevensblokken via IP-netwerken te versturen, geniet een groeiende populariteit. De ondersteuning voor dit protocol groeit. Er zijn inmiddels enkele iSCSI-implementaties in zowel hard- als software op de markt. Bram Dons bespreekt een -nu nog ‘proprietary’- softwaretoepassing voor storage via IP: IPStor van FalconStor.

IPStor presteert goed volgens populaire ‘standaard’

FalconStors opslagmethode krijgt op termijn plaats in een heterogene infrastructuur

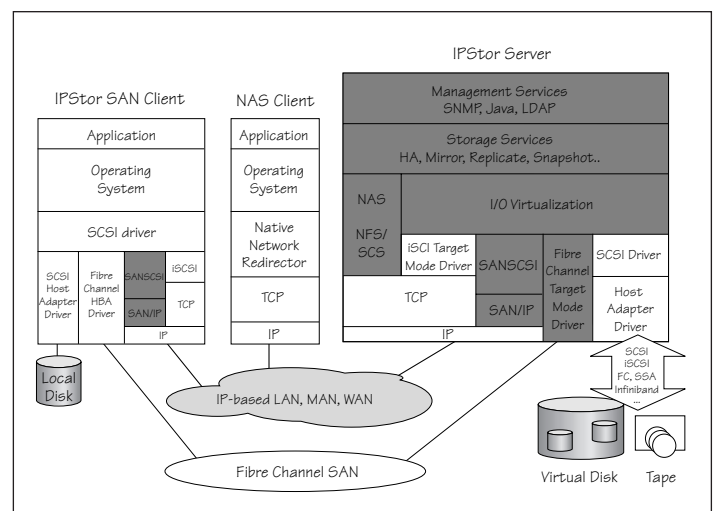
Bij de iSCSI-methode (zie kader *Nieuwe toepassing IP-netwerken* onder het kopje *Drie standards*) worden de SCSI-bloks (blokken) in TCP/IP-pakketjes verpakt, waarna ze met behulp van het TCP/IP-protocol via LAN of WAN worden verzonden.

In hardware zijn vorig jaar al twee iSCSI-routers op de markt verschenen die een koppeling mogelijk maken tussen een FC- en IP-netwerk: Cisco's Nuspeed 5420 en CNT's Ultranet Edge Storage router; de laatste ondersteunt ook FCIP.

Wat is IPStor?

IPStor is een softwarepakket waarmee gebruikers opslagnetwerken via standaard ethernet kunnen bouwen. In tegenstelling tot andere methoden van storage via IP, maakt FalconStor uitsluitend gebruik van software. Dit op het Linux-besturingssysteem gebaseerd pakket draait op elk off-the-shelf Intel-systeem en vereist geen speciale hardware. De IPStor-server is aan de ene kant via een of meer ethernet-verbindingen op het LAN aangesloten en aan de andere kant via een FC- of SCSI-bus verbonden met de opslagssystemen. De IPStor-clients -de eigenlijke applicatieservers- zijn via standaard ethernet (lieft GbE) met de IPStor-server verbonden, maar dat is ook via FC mogelijk.

De applicatieservers ‘zien’ de IPStor-server als één -weliswaar virtueel- standaard blokdata-device. Op de applicatieserver draait een speciale software-driver, die de SCSI block calls afvangt en rechtstreeks naar de NIC redigeert. Bij aankomst van de IP-pakketjes op de IPStor-server



Afbeelding 1: Netwerkkarchitectuur van IPStor.

Nieuwe toepassing IP-netwerken

Tot nu toe worden IP-netwerken hoofdzakelijk op applicatieniveau gebruikt voor het overbrengen van bestanden, webpagina's en andere vormen van gegevenstransport. Op *bestandniveau* zijn bepaalde bestandsystemen (NFS, CIFS) wel direct via IP-netwerken te bereiken; sinds enkele jaren wordt dit veel in praktijk gebracht door de populaire systemen van network attached storage (NAS) - de zogenaamde appliances. Maar al heel lang bestaat de wens opslagsystemen ook op *blokniveau* over langere afstanden met servers te kunnen verbinden. Voor de 'middellange' afstand is daarvoor fibre channel (FC) het aangewezen middel. FC is tot over een afstand tot tien kilometer te gebruiken. Wie verder weg wil, kan bijvoorbeeld ATM inzetten of een FCIP-router, die via een IP-netwerk de koppeling tussen SAN's aanbrengt. Een nieuwe ontwikkeling in lange-afstandoverdracht is de uitwisseling via IP-netwerken van gegevens op blokniveau tussen servers en opslagsystemen. Deze technologie van storage via IP ontkoppelt opslagsystemen van de serversystemen en maakt deze via IP-netwerken voor andere systemen over (wereldwijde) afstanden bereikbaar; de directe koppeling wordt in SAN-terminologie direct attached storage (DAS) genoemd. Fysieke ontkoppeling van opslagsystemen opent voor applicaties en opslagsystemen de deur voor allerlei nieuwe toepassingen. Op beheerniveau moeten we dan bijvoorbeeld denken aan remote mirroring, replicatie, snapshots en andere vormen van backup, bij opslagsystemen aan opslagvirtualisatie en op systeemniveau aan clustering en load balancing.

Drie standaards

Er is een aantal storage via IP-methoden om gegevensblokken via IP-netwerken te versturen. Drie standaards zijn in de maak: fibre channel via IP (FCIP), iFCP en SCSI via IP (iSCSI). Voor die laatste bestaat ook een aantal 'proprietary' varianten: SolP van de Nishan Systems en FalconStors SAN/IP-protocol.

Bij het binnenkort geratificeerde iSCSI-protocol worden SCSI-blokken in IP-pakketjes verpakt, bij het FCIP-protocol FC-frames. SAN-IP-standaards als iFCP en FCIP bieden alleen SAN-naar-FC verbindingen via IP-netwerken. Het belangrijkste verschil tussen FCIP en iSCSI is dat FCIP alleen van het low-level IP-protocol voor het transport gebruikt maakt, maar de eigenlijke verbindingen overlaat aan de hoger gelegen SAN-protocollen. FCIP maakt dan ook geen gebruik van de TCP/IP-routingprotocollen om het intelligente *multi-hop* routingwerk te doen. Bij de iSCSI-methode worden de SCSI-blokken in TCP/IP-pakketjes verpakt, waarna ze met behulp van het TCP/IP-protocol via LAN of WAN worden verzonden; dit in tegenstelling tot FCIP, dat voor de middelste en bovenste lagen de FC-protocollen blijft gebruiken.

worden de block calls door de serversoftware overgebracht naar de betreffende fysieke opslag-devices, die via een SCSI- of FC-bus met elkaar zijn verbonden.

SAN/IP, een gemodificeerd protocol

IPStor maakt gebruik van een 'proprietary' SAN/IP-protocol, dat het SCSI-verkeer in IP-pakketjes verpakt; ondersteuning van iSCSI is aangekondigd. De van de virtuele adapters van de client afkomstige requests worden verpakt in SAN/IP-pakketjes. De IPStor-server ontvangt deze pakketjes en pakt ze uit, zodat de oorspronkelijke SCSI-commando's weer tevoorschijn komen. FalconStors SAN/IP handelt het gehele proces af met minimale overhead, zodat de SCSI-devices op maximale snelheid kunnen werken, zelfs over een IP-netwerk. Het voordeel van het verpakken van opslaggegevens in SAN/IP-pakketjes

is dat gegevens over *trunked* adapters -een IP-verbinding gebundeld over meerdere NIC's- kunnen worden verstuurd. Zo kan de potentiële doorvoer naar een of meer opslagdevices eenvoudig worden verdubbeld. Dit is niet mogelijk bij busgebaseerde interfaces, omdat alle gegevens via dezelfde bus moeten worden verzonden en niet over meerdere bus-systemen zijn te verdelen.

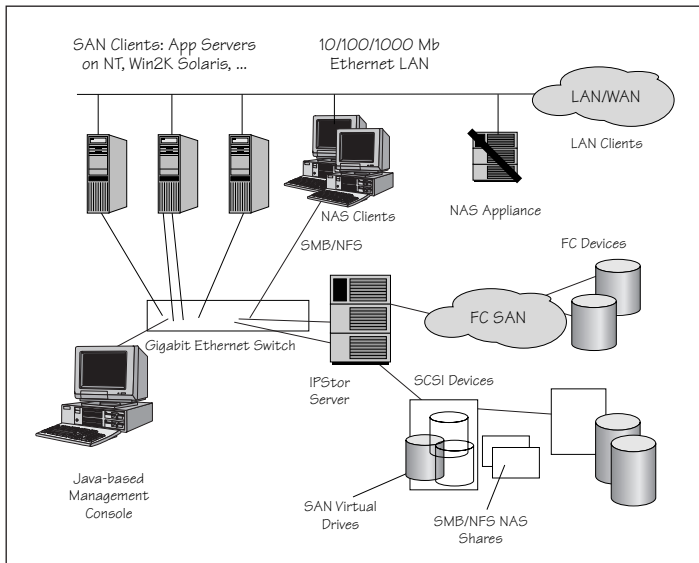
De vraag is hoe deze IPStor-toepassing presteert. Onafhankelijke, geverifieerde testen hebben aangetoond dat de gegevensoverdracht tussen host en opslag-device via een 1 GbE-netwerk 115 MBsec bedraagt. De theoretische limiet voor een 1 GbE-netwerk is 125 MBs; IPStor zal maximaal de beschikbare bandbreedte van een 1 GbE-netwerk benutten. IPStor bereikt deze resultaten doordat het UDP-deel van het TCP/IP-protocol-stack een 're-engineering' heeft ondergaan. SAN/IP is de naam van dit gemodificeerde protocol. De SAN/IP-technologie is overigens complementair aan de toekomstige iSCSI-standaards. Zonder veel aanpassingen zal IPStor ook de nieuwe iSCSI-standaard kunnen ondersteunen.

Architectuur IPStor-omgeving

De IPStor-omgeving bestaat uit de volgende hoofdcomponenten: IPStor Server, IPStor SAN client, IPStor native NAS client en IPStor Console. IPStor Server vormt de schakel tussen de opslagsystemen en SAN- en NAS-clients. IPStor Console, het beheersysteem, kan op één en op elke client worden geïnstalleerd. De componenten zijn in hetzelfde netwerksegment met elkaar verbonden, Storage Network. De IPStor-server bestaat uit een 'dedicated' netwerkopslagserver, waarop een standaard Linux-'appliance' draait. In feite is het een standaard pc-server, met daarop een Linux OS dat 'on top' de IPStor-applicatie laat draaien. Het serversysteem is via FC en/of direct via de SCSI-bus met opslagsystemen te verbinden. De taak van de server is het regelen van de communicatie tussen de SAN/NAS-clients en de achter de IPStor-server gelegen opslagsystemen. Om de beschikbaarheid te vergroten wordt zowel active-active als active/passive failover clustering ondersteund. Mirroring, replication en snapshot-technieken ondersteunen het concept van een snelle en flexibele disaster recovery.

De SAN-client heeft via een speciale meegeleverde softwaredriver toegang tot de server. Deze driver is als een virtuele SCSI-adapter geïmplementeerd, waardoor het betreffende besturingssysteem (Windows NT/2000, Linux of Solaris) transparant toegang heeft tot de virtuele

(IM)



Afbeelding 2: IPStor-omgeving.

opslagbronnen. Applicaties zien geen echte virtuele adapters, maar daadwerkelijk bestaande SCSI-devices. Opslagssystemen zien er voor het OS dan ook uit als lokaal verbonden devices, die zich in werkelijkheid als SCSI-devices 'achter' IPStor Server bevinden. Op de clients is geen hardware-SCSI-adapter nodig; de SAN-clientsoftware en een aparte NIC voor de verbinding met het IP-netwerk volstaan.

Virtualisatie

Virtualisatie is de nieuwe trend in 'storageland', waarvan veel definities in omloop zijn. Simpel gezegd is opslagvirtualisatie de logische presentatie van meerdere fysieke opslagdevices (JBOD's, RAID's of tapedevices) in één 'virtueel' device. Voor de hostserver ziet het virtuele device eruit als een lokaal verbonden device. Een van de sterkste punten van de virtualisatie-architectuur is dat devices aan bestaande of nieuwe storage pools (SAN of NAS) kunnen worden toegevoegd. Daarna zijn van de extra opslagruimte virtuele volumes te creëren. De meeste oplossingen ter zake zijn bijna uitsluitend gericht op de presentatie van meerdere fysieke devices aan hostservers als logische *virtual volumes*.

Nog twee aspecten ontbreken bij het bieden van een degelijke toepassing: connectiviteit en applicatieverbreding. Tot voor kort werd virtualisatie gezien als een techniek die alleen kon worden toegepast in de FC-SAN-omgeving. In de hedendaagse IT-omgeving ontstaat steeds meer vraag naar dynamisch adresseerbare opslagssystemen; niet alleen binnen geïntegreerde SAN-omgevingen, maar ook voor die gekoppeld aan NAS-omgevingen. Op dit punt onderscheiden de meeste producten op de markt zich van elkaar. De meeste bieden alleen virtualisatie tussen SAN's, maar niet voor de NAS-omgeving. Op dit moment wordt nog voorbijgegaan aan een van de belangrijkste aspecten van storagevirtualisatie: de beschikbaarheid van opslagssystemen binnen én buiten de SAN-omgeving.

High availability

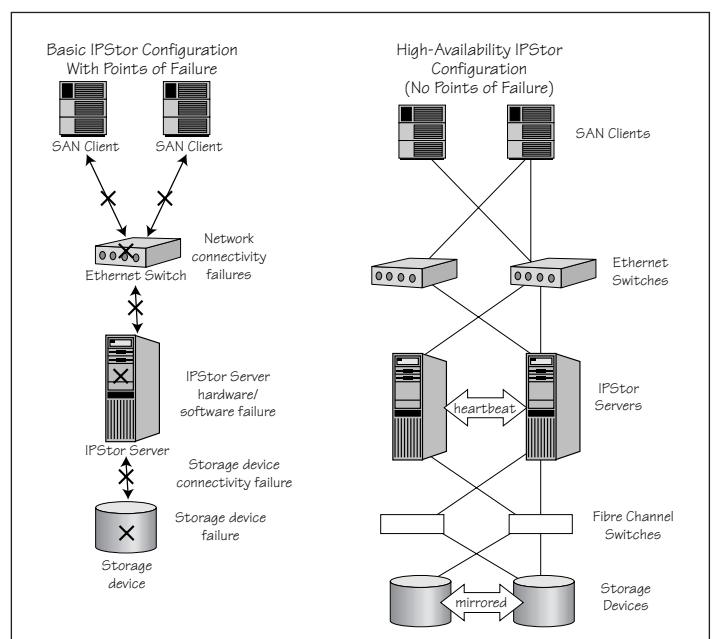
Het zal de lezer niet ontgaan zijn dat de IPStor-server een *single point of failure* vormt binnen de IPStor-architectuur. Immers, bij uitval van

de server heeft geen enkele client meer toegang tot de SAN/NAS-opslagbronnen. FalconStor heeft over deze tekortkoming natuurlijk nagedacht en is voor IPStor gekomen met de veel toegepaste failover-oplossing. In een dergelijke configuratie werken IPStors primaire en secundaire server onafhankelijk van elkaar, elk met de daaraan toegekende clients. Iedere server is verantwoordelijk voor zijn eigen clients; alleen in een failover-situatie worden clients van de andere server overgenomen. *Auto recovery* zorgt ervoor dat na fouthterstel de besturing naar de primaire server wordt teruggegeven. Failover gaat alleen niet op voor NAS-bronnen. Binnen de failover-configuratie vormt de gedeelde schijf de laatste single point of failure. IPStors feature voor mirroring biedt bovendien bescherming tegen uitval van deze gedeelde schijf.

IPStor NAS

IPStor bestaat in feite uit twee softwarestacks, voor de afhandeling van respectievelijk gegevens op blokniveau binnen SAN's en die op bestandsniveau binnen IP-netwerken (NAS). Binnen de NAS-omgeving hebben gebruikers via de standaard netwerkkinterfaces (NFS en CIFS) toegang tot gedeelde gegevens en opslagssystemen. De meeste NAS-systemen zijn uitgevoerd als aparte hardwareboxen, waarin een netwerkkinterfacé, netwerkbesturingssysteem en software voor het alloceren van opslag zijn opgenomen, de zogenaamde appliances. Zo'n NAS-box wordt rechtstreeks op het IP-netwerk aangesloten, waarna gebruikers via *file shares* gedeeld toegang hebben tot bestanden.

Bij een groeiend aantal gebruikers is steeds minder vrije opslagruimte beschikbaar. Door een NAS-box bij te plaatsen lijkt het capaciteitsprobleem in eerste instantie vrij eenvoudig op te lossen. Toch levert een uitdijende NAS-architectuur op termijn wel een probleem op. Met de NAS-box wordt weer een extra stukje hardware bijgeplaatst, net zoals dat met de direct aan de server verbonden opslagssystemen (DAS) het geval was. En dat brengt een eigen backup, toewijzing van opslag en een beheersysteem met zich mee. Maar in de NAS-toepassing



Afbeelding 3: IPStor failover.

IPStor Version 2

De recente IPStor versie 2 biedt verschillende verbeteringen ten opzichte van de vorige versie 1.0. Versie 1.0 ondersteunt FC ook in Initiator-mode, dat toegang bood tot FC-gebaseerde opslagdevices die deel uitmaken van de 'gevirtualiseerde' opslagpool. De compatibiliteit van IPStor als iSCSI-initiator is al met succes getest met IBM's TotalStorage 200i. Ondersteuning van iSCSI-target biedt IPStor de mogelijkheid iedere hostcomputer die is uitgerust met een of meer iSCSI-HBA's of een standaard NIC met een iSCSI-emulator, toegang te geven tot IPStors opslagpool. IPStor Server Software is nu ook beschikbaar op Solaris SPARC (32- en 64-bit versie). Linux kernel v2.4 wordt ondersteund voor zowel IPStor-server als -clients, met onder meer grotere geheugenruimte (tot 64 GB) en nieuwe bestandsvoorzieningen. IPStor ondersteunt Oracle Snapshot Agents voor point-in-time copying, zero-impact backup en remote replicatie.

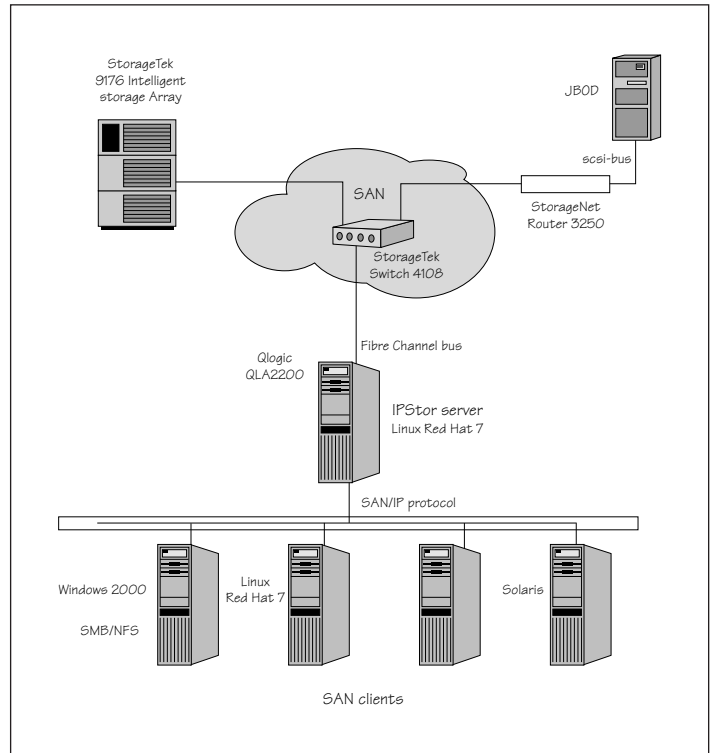
van IPStor betekent toevoeging van een opslagsysteem een extra opslagdevice binnen het *bestaande* opslagnetwerk. NAS gebruikt dezelfde opslagdevices als SAN, waardoor de noodzaak van aparte backup-devices komt te vervallen.

Een ander voordeel van IPStor NAS is de mogelijkheid snapshot en mirroring uniform toe te passen voor zowel SAN- als NAS-opslagbronnen en deze op basis van een eenduidige opslagbeheerpolicy te beheren. Traditionele NAS-boxes kunnen een dergelijke oplossing niet bieden, omdat ze in feite 'opslageilanden' binnen het LAN zijn. IPStor NAS echter maakt deel uit van een geïntegreerd opslagnetwerk. In tegenstelling tot bij IPStor SAN-clients is voor NAS geen speciale driversoftware aan de cliëntzijde nodig, omdat het OS in de regel zelf al de benodigde netwerkredirectors bezit voor de CIFS/SMB- en NFS-bestandsystemen.

Testopstelling IPStor

Onze testopstelling bestaat uit een IPStor-server (Pentium III 800 MHz en 256 MB RAM), waarop Linux Redhat 7.0 is geïnstalleerd. De clients zijn via een 100 MBs LAN met de Linux-server verbonden. De opslagkant bestaat uit een SAN met als schakelpunt een StorageTek 4100-switch, waarop een 9176 Storage-array (via een 3250 router) een JBOD is aangesloten. Voordat met de installatie van IPStor kan worden begonnen, moeten alle SAN-devices zijn geconfigureerd en vanuit de server bereikbaar zijn. Via een speciaal programma, IPStorcheck genaamd, is van tevoren te controleren of alle door IPStor gebruikte OS-componenten aanwezig zijn en de juiste revisie hebben.

(IM)



Afbeelding 4: IPStor-testopstelling.

De installatie van IPStor op server en clients is tamelijk simpel, is binnen enkele minuten gebeurd en verloopt zonder problemen. We installeren de IPStor-client en -console op een Windows 2000-systeem. Na afloop daarvan zien we dat een 'pseudo'-SCSI-adapter is aangemaakt, onder de naam FalconStor IPStor SCSI Host Adapter. Voor de installatie van IPStor Console is Java 2 Runtime Environment (JRE) nodig.

Configuratie IPStor SAN-bronnen

IPStor herkent nieuw geïnstalleerde devices direct bij het opstarten. Default worden alle vaste schijven voor *virtual* devices gereserveerd en alle tapedrives en library's als *direct* devices. SAN Resources zijn de eigenlijke opslagbronnen waartoe clients toegang hebben. We zagen al dat clients met IPStor geen rechtstreekse toegang hebben tot de achter de server gelegen Physical Resources. De fysieke opslagbronnen moeten eerst als SAN-bronnen worden gedefinieerd -de beheerder wijst ze toe aan clients- en ten slotte door de client zelf 'gemount', alsof ze een gewone lokale SCSI-device zijn. Bij de toekenning wordt voor de betreffende client een virtuele adapter gedefinieerd en de resource gekoppeld aan een virtuele SCSI ID van deze adapter. Deze methode bootst de configuratie van een echt SCSI-opslagdevice en -adapter na, en besturingssysteem en applicaties gedragen zich daarnaar.

De vanuit fysieke schijven gecreëerde virtuele devices bestaan uit verzamelingen van *storage blocks*. Deze gegevensblokken kunnen zich op een of meer fysieke schijven bevinden. Dit maakt het mogelijk virtuele devices samen te stellen uit een deel van een grotere fysieke schijf of een samenstelling van meerdere schijven. Virtuele devices bieden de mogelijkheid tot schijfuitbreiding, waarbij extra blokken aan het eind van bestaande virtuele devices kunnen worden toegevoegd. Dit alles is mogelijk zonder de noodzaak bestaande gegevens op de schijf te wissen.

Met name deze mogelijkheid is een van voordelen van virtualisatie, waardoor voorheen niet meer te gebruiken opslagruimte -als gevolg van het bekende 'LUN-probleem'- nu weer ter beschikking staat van de systeembeheerder. De blokken mogen afkomstig zijn van iedere vrije opslagruimte, op welke schijf dan ook. Een virtueel device kan dus worden samengesteld uit een of meer fysieke devices. Daardoor zijn zeer grote virtuele devices te creëren. Heeft een opslagdevice meer opslagruimte nodig, dan kan men die eenvoudig toevoegen aan het virtuele device. Uiteraard moet de clientapplicatie, afhankelijk van type bestandstelsel en OS, daarna de partitie en het bestandstelsel op het virtuele device nog zelf aanpassen. Default zijn alleen vaste schijven bruikbaar als virtuele devices, tapedrives en library's alleen als direct devices; sommige backup-programma's en devices verlangen een directe SCSI ID-adressering met vaste adressen.

Configuratie NAS

Er zijn twee typen NAS-clients: Windows-clients die van het CIFS-protocol en NFS-clients die NFS volgen. De configuratie van NAS doorloopt een aantal stappen:

- configuratie Windows clients;
- toevoeging NFS-clients;
- creatie NAS Resources;
- creatie share folder en toekenning daarvan aan een client;
- mapping/mounting van de 'share'.

Bij de configuratie van Windows-clients valt te kiezen uit drie toegangsmethoden: Domain-, Server- en Share-mode (default). Dit hangt

IPStor bereikt goede resultaten door 're-engineering' van het UDP-deel van het TCP/IP-protocol-stack

van al of niet gebruik van combinaties van een PDC/Domain-controller. Voor de NFS-client hoeft alleen de *fully qualified domain name* of een 'hard' IP-adres worden ingevuld; via het IP-subnetwerk en -netmasker ook alle systemen binnen een IP-subnetwerk.

Evenals bij SAN Resources, zijn bij NAS alle beschikbare fysieke opslagbronnen als NAS Resources te definiëren. Na de aanmaak van een NAS-bron volgt toekenning van een Windows- of NFS-client aan de betreffende share. In de New Share-wizard kan de beheerder voor Windows- of NFS-clients aangeven welke clients toegang hebben en de daarbij geldende privileges toekennen. Bij Windows-clients moet daarna met het bekende Map Network Drive-tool de share voor iedere client worden aangemaakt, bij NFS creëert men eerst een 'mount' directory en mount daarin vervolgens de share.

Testresultaten en toekomstverwachting iSCSI

De belangrijkste vraag bij de toepassing van storage via IP is hoe deze technologie zich verhoudt met een opslagsysteem binnen een DAS en SAN als het gaat om snelheid.

We testen met Intels IOMeter vanaf een Windows 2000 IPStor-client de prestaties van een virtuele SAN Resource. Het blijkt dat via een 100



Afbeelding 5: Console van IPStor.

MBs LAN een doorvoersnelheid van 3 MBs word gehaald. Dat steekt schril af bij de 25 MBs snelheid van een lokale schijf. Het 100 MBs LAN vormt hier duidelijk de flessenhals. Voor storage via IP is minimaal een 1 Gigabit-netwerk nodig.

Hoewel het iSCSI-protocol binnenkort geratificeerd wordt, verwacht de Gartner Group grootschalige inzet van storage via IP pas in 2003. De meeste SAN/NAS-analisten zijn het erover eens dat iSCSI en FCIP in de toekomst zullen samengaan en als gemeenschappelijke standaard gaan dienen voor netwerken van storage via IP. Steeds minder analisten geloven dat TCP/IP geen dominante rol gaat vervullen bij de bouw van opslagnetwerken over lange afstand. Storage via IP heeft weliswaar bepaalde voordelen boven SAN, maar zal de meer betrouwbare en snellere FC-gebaseerde netwerken zeker niet vervangen. De meeste marktonderzoekers gaan ervan uit dat SAN's en netwerken voor storage via IP op termijn beide toepassing zullen vinden in een heterogene opslaginfrastructuur.

Bram Dons

Bram Dons is onafhankelijk IT-adviseur. E-mail: b.dons@chello.nl

Informatie op Internet:

www.falconstor.com
www.nishansystems.com
www.cnt.com
www.cisco.com/warp/public/cc/pd/rt/5420/index.shtml
www.scsita.org/
www.t10.org/
www.ietf.org
www.snia.org