

Routing-Strategien im Universitätsrechnernetz

Das Rechnernetz der Humboldt-Universität wächst nach wie vor schnell, wenn auch langsamer als im *SERVUZ-Projekt* geplant. Es kommen ständig neu verkabelte Institute und Datenleitungen hinzu.

Das Universitätsrechnernetz besteht aus einzelnen Segmenten in den Bereichen, in der Regel Ethernet (Koax oder S/UTP), und Datenleitungen, die diese verbinden. Die Datenleitungen, die bisher einen maximalen Durchsatz von 64 kbit/s hatten, werden seit Januar diesen Jahres durch Laser-Links mit einer Geschwindigkeit von 10 Mbit/s z.B. vom Hauptgebäude zur Burgstraße und Spandauer Straße und über das Charité-Hochhaus zu den umliegenden Fachbereichen u.a. in der Hessischen Straße ergänzt. In Kürze werden die ersten Glasfaserverbindungen in Betrieb genommen, die 100 Mbit/s (FDDI-Standard) und mehr erlauben. Dazu kommen im Rechenzentrum die Verbindungen in die *Außenwelt* über das *WiN* und die *2 Mbit-Standleitung zur Technischen Universität*. Das Wachstum des Netzes erfordert neue Überlegungen zu Struktur und Routing.

Netzweit werden zwei unterschiedliche Datenübertragungsprotokolle benutzt, das **Internet-Protokoll** (IP) und **Banyan VINES** sowie mit abnehmender Bedeutung auch **X.25**. In diesem Beitrag wird nur auf das Internet-Protokoll eingegangen. An den Knotenstellen des Netzes, d.h. zwischen den Segmenten und an den Enden von Datenleitungen, befinden sich spezielle Rechner, sogenannte **Router** (im Internet Sprachgebrauch auch *Gateways*), die entscheiden, ob Datenpakete, die von einem Rechner in einem Segment abgesendet wurden, in ein anderes Segment oder über eine Datenleitung weiterzuleiten sind. Diesen Prozeß bezeichnet man als **Routing**. Er ist für jedes Protokoll standardisiert. Die Internet-Standards heißen **RFCs** (Request For Comment). Der Router trifft seine Entscheidungen anhand der 32 Bit langen **Internet-Adressen** in jedem Datenpaket. Internet-Adressen sind im *RFC-950* definiert und werden durch die ersten zwei Bit in die Klassen A, B und C unterteilt. Wir haben Adressen der *Klasse B* mit folgendem Aufbau:

1 0	Netz-Nr.	Subnetz-Nr. / Rechner-Nr.
0 1 2 ... 15 16 31

Bei der Klasse B belegt die Netznummer 14+2 Bit. Die restlichen 16 Bit können beliebig auf Subnetznummer und Rechnernummer aufgeteilt werden. Auch Netze ohne Subnetze sind möglich. Internet-Adressen werden mit vier Dezimalzahlen für jedes Byte, getrennt durch Punkte, geschrieben. Z.B. hat die CONVEX im Rechenzentrum die Adresse 141.20.1.2. Dabei ist 141.20.0.0 unsere Netznummer. Die Aufteilung der restlichen 16 Bit auf Subnetz- und Rechnernummer wird durch **Subnetzmasken** festgelegt. Im Beispiel gilt die Subnetzmaske 255.255.255.0,

d.h. es sind jeweils 8 Bit Subnetz- und Rechnernummer (hier Subnetz 1, Rechner 2). Die Nummern 0 und 255 sind für *Broadcasts*, d.h. Nachrichten an alle, reserviert. Folglich können 254 Rechner in einem Subnetz adressiert werden. In der Regel wird man kaum so viele Rechner in einem Segment haben. Das ist zwar mit Bridges und Repeatern realisierbar, aber vom Durchsatz her unzweckmäßig. Da außerdem beim Festhalten an der bisherigen Subnetzstruktur in absehbarer Zeit die Subnetznummern erschöpft wären, sind wir dazu übergegangen, in einigen Bereichen längere Subnetznummern zu verwenden. Dehnt man die Subnetznummer auf 10 Bit aus, lautet die Subnetzmaske 255.255.255.192. Betrachten wir die Adressen 141.20.20.3 und 141.20.20.67. Die Rechnernummer ist dann in beiden Fällen 3, die Subnetznummer 20.0 bzw. 20.64.

Zurück zum Routing. Der Router prüft bei jedem gelesenen Paket anhand von Netz- und Subnetznummer, ob die Adresse zum gleichen Segment gehört oder ob es in ein anderes angeschlossenes Segment transportiert werden muß. Gegebenenfalls wird das Paket auch an einen benachbarten Router geschickt, dazu jedoch später.

Neuere Internet-Standards lassen es zu, die Subnetzmaske für jedes Segment gesondert (*variabel*) zu definieren. Das bringt für den einzelnen Rechner keine Probleme, da er in der Regel nur einen Netzanschluß und damit *eine* Internetadresse und *eine* Subnetzmaske hat, anhand derer er entscheidet, ob Pakete direkt an den Zielrechner oder an den nächsten Router (*Default Gateway*) zu schicken sind.

Router bauen für ihre Routing-Entscheidungen *Routing-Tabellen* auf (wie auch in vereinfachter Form jeder Rechner). Darin werden automatisch alle anliegenden Netze und Subnetze mit den entsprechenden Masken eingetragen. Um zu entscheiden, wohin mit Paketen für fremde Netze oder Subnetze, muß der Router Informationen über die weitere Netztopologie erhalten. Das kann *statisch*, d.h. manuell durch den Netzadministrator erfolgen. Im einfachsten Fall genügt dazu an der Peripherie des Netzes die Eintragung eines *Default Gateways*, also der Adresse eines Routers, der für alle unzustellbaren Pakete zuständig ist. Im zentralen Teil, dem sogenannten **Backbone**, das die Verbindung zwischen den verschiedenen Bereichen bildet, sieht es viel komplizierter aus. Hier gilt es, in unserem Netz z.Z. 15 Router vom Typ *3Com NetBuilder II*, ca. 15 *VINES-Server mit IP-Routing-Option* und weitere Router anderer Hersteller zu koordinieren. Das ist mit statischen Verfahren ein fast unlösbares Problem. Hinzu kommt, daß für einige Datenverbindungen aus Sicherheitsgründen **Back-up-Leitungen** bestehen. Wenn Laserverbindungen wegen schlechter Sicht ausfallen, soll auto-

matisch auf Reserveleitungen umgeschaltet werden. Dazu werden *dynamische* Verfahren benötigt, mit deren Hilfe Router die Netztopologie automatisch kennenlernen. Übrigens werden im Banyan VINES nur in Sonderfällen statische Eintragungen benutzt. Das einfachste dynamische Verfahren im LAN-Bereich des Internets ist das **RIP** (Routing Information Protokoll), das seinen Ursprung im BSD-UNIX hat. Hierbei versendet jeder Router per Broadcast (an alle) alle 30 Sekunden seine Routinginformationen. Andere Router lernen so *Routen*, d.h. Wege zu fremden Subnetzen innerhalb eines Netzes, und vergessen sie wieder, wenn sie eine Zeit lang nicht empfangen wurden. Dieses Verfahren hat mehrere Nachteile:

- Es werden viele zusätzliche Informationen im Netz transportiert (*Overhead*).
- Die Güte von gefundenen Wegen wird nur nach der Zahl der beteiligten Router (*Hops*) berechnet. Die Kapazität der Leitungen geht nicht in die Berechnung ein.
- Das Verfahren reagiert langsam (mehrere Minuten) auf Änderungen der Topologie, wie z.B. ausgefallene Verbindungen oder Router. Es können auch *Routing Loops* (Schleifen) oder andere Fehler auftreten.
- Es werden keine Subnetzmasken transportiert.

RIP wird bei uns in einzelnen Segmenten verwendet, wenn Router kein anderes dynamisches Protokoll verstehen und nicht statisch gearbeitet werden soll. VINES-Server beherrschen z.Z. kein dynamisches IP-Routing-Protokoll, so daß Routen immer statisch auf einem benachbarten Router eingetragen werden müssen. Allerdings werden VINES Server künftig nur noch einen kleinen Teil des IP-Routing im peripheren Bereich realisieren.

Im Backbone-Bereich haben wir uns für ein neueres dynamisches Routing-Protokoll entschieden, das die genannten Nachteile von RIP auf der Basis eines LinkState-Algorithmus vermeidet und inzwischen von vielen Router-Herstellern implementiert wurde, **OSPF** (Open Shortest Path First), das im RFC 1131 spezifiziert ist

Das Verfahren arbeitet im LAN-Bereich in einem sogenannten *autonomen System* (AS), das aus mehreren Netzen bestehen kann und über *AS Boundary Router* (ASBR) mit anderen autonomen Systemen kommuniziert. Zu unserem AS gehören außer dem Netz 141.20.0.0 auch das Netz der Charité 141.42.0.0.

Die Router verschicken innerhalb des Systems regelmäßig kurze Broadcasts mit ihrer Identifikation. Damit können sie von anderen Routern gefunden werden. Hat ein Router einen potentiellen Nachbarn

erkannt, tauscht er mit diesem Router Informationen aus und trägt sie in seine eigene *Netzdatenbasis* ein. In die Datenbasis gehen auch die evtl. variablen Subnetzmasken und die *Kosten* für eine Verbindung ein. Die Kosten im Sinne von OSPF sind ein Wert, der z.B. die Kapazität und den Mietpreis einer Datenleitung ausdrückt. Die Datenbasis wird normalerweise nur alle 30 Minuten aktualisiert oder *ereignisgesteuert*, d.h. wenn eine Verbindung ausfällt oder sich ein benachbarter Router nicht mehr meldet. Auf diese Weise ist der normale Datenaustausch zwischen den Routern gering. Trotzdem kann innerhalb weniger Sekunden auf Veränderungen der Netztopologie reagiert werden.

Gefundene Routen werden mit der Summe der Kosten aller auf dem Weg liegenden Verbindungen bewertet. So kann eine Route über drei Laser-Links günstiger sein, als über eine direkte aber langsame Datenleitung. Auch kann die Last, wenn mehrere Routen mit gleichen Kosten existieren, verteilt werden.

Mit OSPF kann das Netz auch in *Areas* unterteilt werden. Das sind Teilnetze, wie z.B. der Backbone-Bereich, oder relativ abgeschlossene Teilnetze in Fachbereichen. Dazu werden *Area IDs*, d.h. Bezeichnungen für die Areas vergeben, die wie IP-Adressen aussehen. Das Backbone erhält immer die Bezeichnung 0.0.0.0. Für den Bereich Invalidenstraße 110 ist 141.20.40.0 vorgesehen. Der Informationsfluß zwischen den Areas wird auf das Mindestmaß beschränkt. Die Area ID wird an Teilnetze (meist mehrere Segmente) vergeben, nicht an Router. So kann ein Router mit mehreren Areas verbunden sein, etwa dem Backbone und einem Fachbereichsnetz. Ein solcher Router wird *Area Border Router* (ABR) genannt und erhält mehr Informationen als ein Router, der nur einer Area (nicht dem Backbone) angehört.

Die Implementation von OSPF durch 3Com ist sehr einfach zu handhaben. OSPF wird für jedes Routerinterface mit IP-Routing eingeschaltet und gegebenenfalls die entsprechende Area ID eingetragen. Wird in besonderen Fällen außerdem mit RIP oder statischen Routen gearbeitet (s.o.), ist noch festzulegen, wie durch RIP gelernte oder statisch eingetragene Routen durch OSPF verbreitet werden sollen. Alles weitere regeln die Router selbständig. Mitunter ist es zweckmäßig, für Verbindungen andere als die standardmäßig angenommenen Kosten einzutragen.

Abschließend kann festgestellt werden, daß die dreimonatige Erprobung von OSPF die genannten Vorteile gegenüber anderen Verfahren anschaulich demonstriert hat. OSPF sollte also überall, wo es möglich ist, auch eingesetzt werden.

Klemens Müller
Tel.: 2093-2880