TECHNISCHE UNIVERSITÄT MÜNCHEN
INSTITUT FÜR INFORMATIK

Diplomarbeit

Einsetzbarkeit verschiedener Routing-Mechanismen in einer bestehenden ATM-basierten Netzwerkumgebung

Florian Wolters

Aufgabensteller:	Prof. Dr. Heinz-Gerd Hegering
Betreuer:	Norbert Wienold (TUM) Andreas Falkner (CompuNet)
Abgabedatum:	15. Mai 1999

Inhaltsverzeichnis

1. Einleitung 9

2. Analyse des bestehenden Netzes 10
2.1. Aufstellung der Anforderungen 10
2.2. Darstellung der Realisierungsvorschläge 12

3. Begriffsdefinitionen 19
3.1. Repeater 19
3.2. Hub 19
3.3. Bridge 19
3.4. Marketingbegriff Switch 20
3.5. Router 20
3.6. Layer-3-Switch 20

4. Beschreibung der verwendeten Technologien 21
4.1. ATM 21
4.1.1. ATM - Die Technologie 21
4.1.1.1. Grundlagen 21
4.1.1.2. B-ISDN-Schichtenmodell / ATM 21
4.1.1.2.1. Beschreibunng der Ebenen 22
4.1.1.2.2. Beschreibung der Schichten 22
4.1.1.3. ATM in LAN- und WAN-Welt 25
4.1.1.4. Multiplexing 27
4.1.1.5. Vermittlungsarten 28
4.1.1.6. Verbindungsorientierte und verbindungslose Kommunikation 28
4.1.2. Leistungsgrenzen von Ethernet 29
4.1.3. Leistungsgrenzen von Token-Ring und FDDI 31
4.1.4. Aufhebung der Leistungsgrenzen durch ATM 31
4.1.5. ATM-Standardisierung 32
4.1.6. PNNI 33
4.2. LAN Emulation 33
4.3. Virtual LAN-Technologie 37

5. Beschreibung traditioneller und neuer Routing-Mechanismen 39
5.1. Traditionelles Routing 40
5.1.1. Interne Gateway Protokolle 40
5.1.1.1. Link State Routing am Beispiel von OSPF 40
5.1.1.2. Distance Vector Routing am Beispiel von RIP 47
5.1.2. Externe Gateway Protokolle am Beispiel von BGP 52
5.2. Wirespeed Routing - Layer-3-Switching 52
5.3. ATM-Routing - PNNI und IISP 53
5.3.1. Private Network to Network Interface (PNNI) 53
5.3.2. Interim Inter-switch Signaling Protocol (IISP) 55
5.4. Shortcut Routing 56
5.4.1. IP Switching 56
5.4.2. Tag Switching 61
5.4.3. Net-Flow-Feature-Technik 65
5.4.4. Multi-Protocol over ATM (MPOA) 67

6. Einsetzbarkeit der Routing-Mechanismen 72
6.1. Darstellung der momentanen Topologie 72
6.2. Beispiele für Kommunikation 75
6.2.1. Kommunikation auf Etagenebene 75
6.2.2. Kommunikation im Gebäude 76
6.2.3. Kommunikation zwischen Gebäuden 77
6.3. Bewertung der traditionellen Routing-Mechanismen 78
6.3.1. Gegenüberstellung OSPF - RIP 78
6.3.2. Border Gateway Protokoll (BGP) 79
6.4. Bewertung von Layer-3-Switching 80
6.5. Bewertung der Shortcut-Mechanismen 80
6.5.1. IP-Switching und NFF-Technik 81
6.5.2. Tag-Switching 81
6.5.3. Multi-Protocol over ATM (MPOA) 82

7. Schlußwort 82

Literaturverzeichnis 83

1. Einleitung

Wenn Unternehmen fusionieren ergeben sich für alle Bereiche der beteiligten Konzerne eine Vielzahl von logistischen, terminlichen und organisatorischen Herausforderungen.

Eine in diesem Zusammenhang elementare Problematik tritt durch die Zusammenführung der komplexen, meist heterogenen Netzstrukturen auf. Definierte Anforderungen bezüglich verschiedener Qualitäts-Parameter wie Verfügbarkeit, Geschwindigkeit und Durchsatz, sowie der Skalierbarkeit und Fehlertoleranz des Netzes erfordern eine gut durchdachte Planung.

Vor diese Aufgabe sah sich die HypoVereinsbank AG Anfang des Jahres 1998 gestellt. Im Zuge der Fusion der ehemaligen Bayerischen Vereinsbank mit der Bayerischen Hypotheken- und Wechselbank mußte für die daraus entstandene Großbank eine neue IT-Infrastruktur mit hoher Performance und Ausfallsicherheit geplant und implementiert werden. Die Netztopologien der beiden Münchner Großbanken basierten Ende 1997 auf zwei unterschiedlichen Konzepten. Die Bayerische Vereinsbank setzte im Campus-Bereich und zur Anbindung der Gebäude am Hauptstandort Token-Ring ein, wobei zwischen den einzelnen Token-Ring Netzen traditionell geroutet wurde. Für die Anbindung der Client-PCs wurde neben Token-Ring auch Ethernet eingesetzt. Die Hypobank setzte bereits ATM im Backbone-Bereich mit Token-Ring LAN-Emulation ein. Für die Anbindung der Client-PCs wurden, wie bei der Vereinsbank, Token-Ring und Ethernet eingesetzt.

Die Aufgabe bestand nun in der Beauftragung eines Unternehmens, welches die Konsolidierung der unterschiedlichen Netztopologien planen und durchführen sollte. Hierfür wurde seitens der HypoVereinsbank eine Ausschreibung gemacht. Nachdem verschiedene Angebote eingegangen und bewertet waren, wurde der Auftrag für die Netzplanung und -implementierung an die Firma CompuNet vergeben. Gleichzeitig wurde die Entscheidung getroffen, die aktiven Netzwerkkomponenten von einem Hersteller zu beziehen. Die Wahl fiel hier auf die Firma Cisco Systems. Das Projekt "HypoVereinsbank Campus Redesign" wurde Mitte 1998 gestartet.

Ziel der Diplomarbeit ist zunächst das, zum Zeitpunkt der Erstellung, bestehende Netz der HypoVereinsbank zu analysieren. Hierzu gehört die Aufstellung der Anforderungen, die Darstellung der Realisierungsvorschläge und die Beschreibung der verwendeten Technologien. Anhand der Topologie-Beschreibung wird deutlich werden, daß zwischen den Gebäuden und dem Campus-Core geroutet wird. Die Wahl des Routing-Mechanismus spielt also eine entscheidende Rolle um die Anforderungen effizient umzusetzen. Aufbauend auf diese Einführung werden die am Markt erhältlichen Routing-Mechanismen untersucht und beschrieben. Neben traditionellem Routing werden die Verfahren Layer-3-Switching und Shortcut-Routing diskutiert. Abschließend werden die Konzepte bewertet und die aktuelle Netztopologie anhand eines Anwendungsbeispiels beschrieben.

2. Analyse des bestehenden Netzes

2.1. Aufstellung der Anforderungen

Für die Realisierung des Netzzusammenschlusses der beiden Banken gab es kein konkretes Anforderungsprofil. Die Hypobank hatte zum Zeitpunkt des Zusammenschlusses bereits ATM im Backbone-Bereich eingesetzt, so daß ATM als Backbone-Technik der neu entstandenen Großbank feststand. Die ATM-Komponenten der Hypobank waren von der Herstellerfirma Bay Networks, welche dann auch zuerst kontaktiert wurde, um Realisierungsvorschläge für die Migration abzugeben. Die Anforderungen waren hierbei sehr allgemein und stützten sich hauptsächlich auf die bereits, bei der Hypobank, realisierten Dienste:

- Die Bandbreite, bzw. Übertragungsgeschwindigkeit sollte nun im kompletten Backbone-Bereich der Hypovereinsbank mindestens 155 Mbit/s betragen. Hierfür wurden die ATM-OC/3 Komponenten eingesetzt.
- Die eingesetzten Komponenten sollten maximal skalierbar sein, so daß sie bei Neuentwicklungen entsprechend angepasst werden konnten (z.B. durch Ersetzung der ATM-OC/3 Komponenten durch ATM-OC/12 Komponenten was zu einer Erhöhung der Übertragungsgeschwindigkeit auf 622 Mbit/s führt). Auch die Topologie sollte so gewählt werden, daß die Netzgröße beliebig skalierbar bleibt.
- Die eingesetzten Komponenten sollten redundant ausgelegt sein, so daß zeitkritische Anwendungen, (wie z.B. Finanztransaktionen) bei einem Ausfall von Komponenten, weiterlaufen.
- Multimediafähigkeit war eine weitere wichtige Anforderung an das Netz, da bereits abzusehen war, daß Videoconferencing, Techniken wie "Voice over IP" oder auch das interne Fernseh- und Informationssystem VIA in das Netzwerk integriert werden sollten.
- Da die Komponenten mit den neuen Techniken nicht einfach aufgestellt werden und dann funktionieren, und sich eine Bank keine Ausfallzeiten leisten kann, war die Integration der vorhandenen Komponenten ein weiterer wichtiger Punkt für die schrittweise Migration.

Der Realisierungsvorschlag von Bay Networks wurde dann als Vorlage für Alternativvorschläge verwendet. Anzustreben war eine hierarchische ATM Peer Group mit Design auf Basis der verabschiedeten Spezifikationen des ATM Forums. Mittels einer PNNI Peer Group sollten autarke, also eigenständige und von anderen Systemen unabhängige, ATM Netze für die HypoVereinsbank realisiert werden. Mit der steigenden Anzahl der teilnehmenden Switche innerhalb einer Peer Group, steigt gleichzeitig die Anzahl der zu verarbeitenden Topologieinformationen. Um die Prozessorauslastung durch das PNNI-Routing auf ein vertretbares Maß zu beschränken, sollte die Größe einer PNNI Peer Group abhängig von der Leistungsfähigkeit eines jeden eingesetzten ATM Switchingsystems gewählt werden.

Ein weiterer wichtiger Aspekt in Bezug auf die Betriebssicherheit beim Zusammenschluß der Teilnetze beider Banken war die logische Trennung auf OSI-Schicht 3. Damit wurde gewährleistet, daß Störungen, wie z.B. Broadcast-Storms, sich nicht in das jeweils andere Teilnetz ausweiten können.
Darüber hinaus wurde dadurch eine zentrale Kontrollinstanz für die Steuerung und Pflege der Kommunikationsbeziehungen zwischen den beiden Häusern geschaffen.

Dieses durfte jedoch nicht zu Lasten des Datendurchsatzes gehen, wie es typische Routersysteme mit sich bringen.

Das in Kapitel 6.1. beschriebene Design bietet durch die "2-Hop-Strategie", welche noch detailliert erklärt wird, eine gute Skalierbarkeit hinsichtlich der Anwenderzahl (> 15.000) und der benötigten Bandbreiten.

Für die Realisierung wurden die verfügbaren Systeme der Hersteller Bay und Cisco in Betracht gezogen.

Aus folgenden Gründen fiel die Entscheidung auf den Hersteller Cisco:

- Der Einsatz von hierarchischen PNNI Strukturen war zum Zeitpunkt der Erstellung des Netzdesigns mit dem Bay Centillion 100 nicht realisierbar, da eine entsprechende Software nicht verfügbar war. Selbst wenn diese verfügbar gewesen wäre, hätte sich ein Einsatz dieser Systeme als leistungsfähige ATM Core Switche nicht empfohlen, da auch der Hersteller Bay Networks den Centillion 100 als Edge-Device bzw. Backbone Switch für kleine ATM Netze positioniert.

- Als leistungsfähige Alternative für den Bay Centillion 100 bot sich das System Centillion 1600 von Bay Networks als ATM Core Switch an. Der Einsatz dieses Systems erschien jedoch als zu riskant, da es sich um ein vollkommen neues Produkt handelte und somit eine Integration in die betrachtete unternehmenskritische Produktivumgebung noch nicht zu vertreten war.

- Im Gegensatz dazu konnte durch den Einsatz der Cisco Catalyst 5500 Serie bzw. LS-1010-Technologie bereits der Grundstein für eine zukunftsweisende hierarchische PNNI Struktur des Netzes gelegt werden. Diese Systeme haben sich in einer Vielzahl von Installationen als robuste und zuverlässige ATM Backbone Systeme erwiesen.

- Die Anforderungen bezüglich der kontrollierten Trennung der Netze auf OSI-Schicht 3 werden durch den Einsatz des Route Switch Moduls (RSM) und der Netflow Feature Card (NFFC) umgesetzt. Mittels dieser Netflow-Architektur können die Vorzüge des Routings in Bezug auf Funktionalitäten wie Zugangskontrolle und Accounting mit Switching-Performanz bei hohen Durchsatzraten von über 3 Millionen Paketen pro Sekunde verbunden werden.

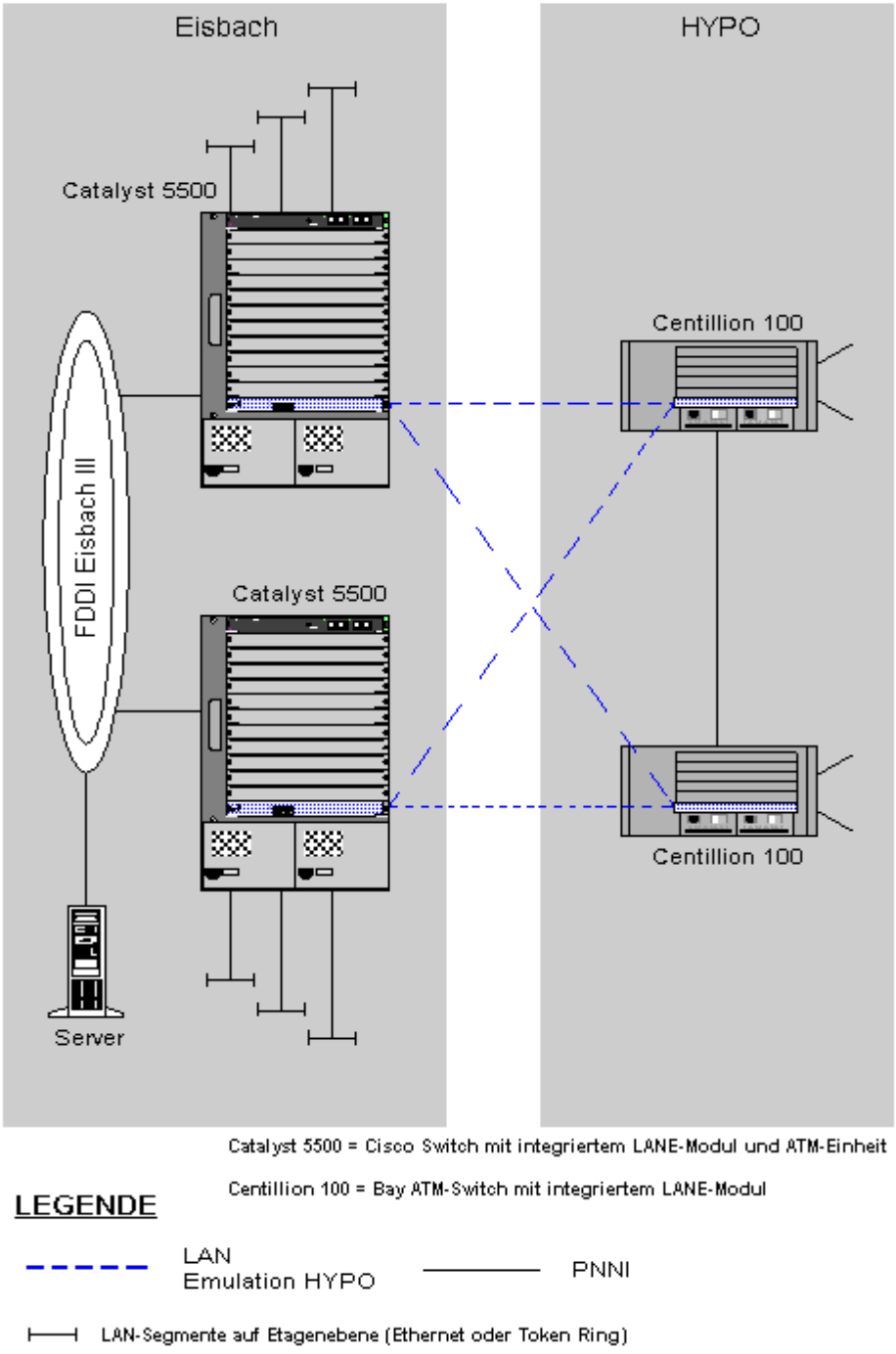
2.2. Darstellung der Realisierungsvorschläge

Die grundsätzliche zentrale Einheit von Banken, sind die Gebäude und Kommunikationseinheiten in denen die Finanztransaktionen ablaufen. Dieses Herzstück wird bei der HypoVereinsbank als "Treasury Netz" bezeichnet. Aufbauend auf der Zusammenlegung der beiden Treasury Netze bzw. dem Umzug der Händler und den damit verbundenen technischen und terminlichen Vorgaben wurden zwei, in mehrere Schritte unterteilbare, Lösungskonzepte erstellt. Die meisten Anwendungen, welche im Treasury-Bereich laufen, sind zeitkritisch, was zur Folge hat, daß eine hohe Ausfallsicherheit garantiert sein muß. Vor allem aus diesem Grund wurden viele Einheiten in der Netztopologie redundant ausgelegt.

Variante I

Bei diesem Lösungsvorschlag wird der Zusammenschluß durch Anbindung der bestehenden Cisco Catalyst 5500 Switche (Definition eines Switches, siehe Kapitel 3.4.) aus den Vereinsbank-Gebäuden "Eisbach 3" und "Eisbach 4" an die Bay ATM-Switche Centillion 100 am Standort der Hypobank gelöst. In den "Eisbach"-Gebäuden befinden sich die Treasury-Netze der ehemaligen Vereinsbank.

Variante I



Diese Variante stellt eine einfache, kostengünstige und schnell zu implementierende Lösung dar.

Um die Anbindung an das vorhandene ATM Netz der Hypo Bank zu realisieren, wird in die zwei bestehenden Cisco Catalyst 5500 Switchsysteme jeweils ein LANE Modul (zweifach 155 Mbit/s ATM Uplink für Single Mode Fiber) integriert. Über diese Verbindungen, die aus Sicht des Vereinsbank Treasury Netzes redundant ausgelegt sind, wird sowohl den Anwendern in der Vereinsbank der Zugriff auf bestehende

ELAN´s auf Basis von LAN Emulation in der Hypobank, als auch umgekehrt ermöglicht. Somit stehen beiden Häusern die Resourcen der jeweiligen anderen Seite zur Verfügung.

Ein wesentlicher Nachteil dieser Lösung ist die völlig transparente Zusammenschaltung beider Netze auf Layer 2-Ebene. Bei etwaigen Problemen auf Seiten der Vereinsbank oder der Hypobank besteht immer die Gefahr, daß sich Broadcasts ungehindert in beiden Netzen verbreiten und somit das ganze System belasten.

Ein weiterer Problempunkt dieser Variante 1 ergibt sich aus topologischen Gründen. Da zum Zeitpunkt der Planung ein Redesign des Campus-Backbones am Tucherpark bereits feststand, wollte man die Weichen für das zukünftige Netzkonzept früh genug stellen um eine mehrmalige Umstrukturierung der Netztopologie zu vermeiden.

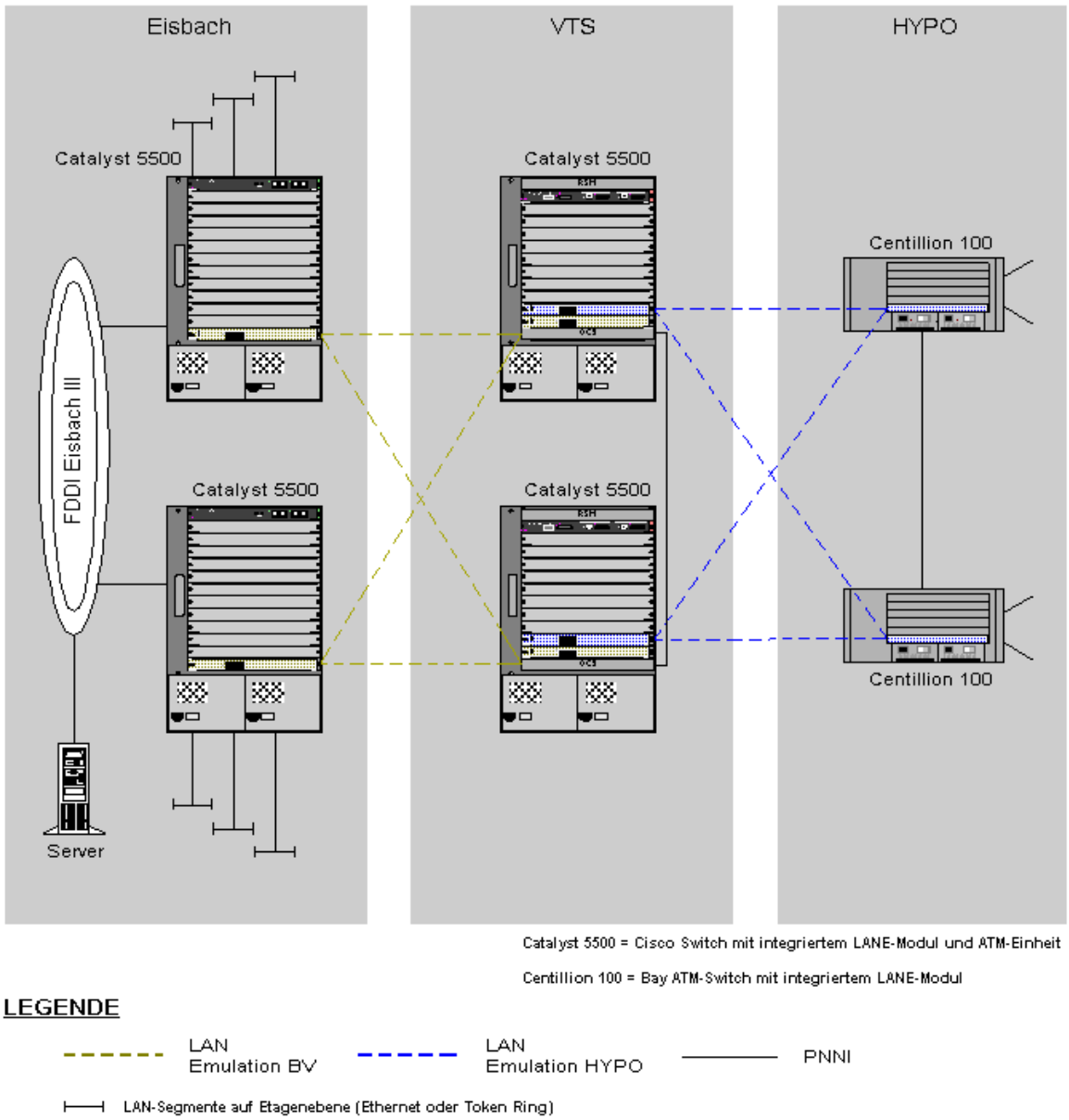
Bereits vor der Planung wurde das Vereinsbank-Gebäude "VTS", von HypoVereinsbank-Seite in Absprache mit den implementierenden Firmen, als der zentrale Datenkommunikationspunkt festgelegt. Die Gebäudeeinheiten, welche noch hinzugefügt werden sollten, sollten über das "VTS"-Gebäude direkt an das Campus-Netz angeschlossen werden. Hier spielt die 2-Hop-Strategie eine wichtige Rolle (siehe Kapitel 6.1.), durch die jede Gebäudeinheit mit maximal 2 Routing-Hops erreicht werden kann. Es empfahl sich also auch die Zusammenführung der beiden Banken-Netze dort zu realisieren, um eine zentrale Schalteinheit zu schaffen. Dies bedingte die Implementierung eines zentralen ATM Knotenpunktes im "VTS".

Variante II

Phase 1:

Die oben genannten Gründe führten zur Planung und Durchführung der zweiten vorgeschlagenen Variante. Die zweite Variante ist in drei Phasen aufgeteilt. In der ersten Phase geht es darum, im Vereinsbank-Gebäude "VTS" einen ATM-Backbone mit ATM-Uplink zum Vereinsbank Treasurynetz aufzubauen. Die Kopplung der beiden ATM Netze der Vereinsbank und Hypobank geschieht über ATM-Switche mit Layer-3 Funktionalität, also mit integrierten Routing-Einheiten.

Variante II
Phase 1



Durch den Aufbau eines zentralen ATM Knotenpunktes im VTS, der Ankopplung von "Eisbach 3" und "Eisbach 4" an den Backbone im VTS und den Aufbau einer Verbindung zum Hypo Netz über einen Switch mit Layer-3 Funktionalität wird der Grundbaustein für einen vollwertig skalierbaren ATM Campus gelegt. Als zentrale Switch Systeme werden zwei Cisco Catalyst 5500 mit je einem Route Switch Modul und einer Supervisor Engine 3, mit integrierter Net-Flow Feature Card (siehe 5.4.3. Net-Flow-Feature Technik), eingesetzt.

Ferner werden in jedem Catalyst ATM Switching Prozessoren, die den ATM Backbone mit Hilfe der OC 3-Module (155 Mbit/s) aufbauen, eingesetzt. Durch diese Bestückung werden die Switchsysteme sowohl als ATM Edge Devices, als auch als ATM Switch betrieben.

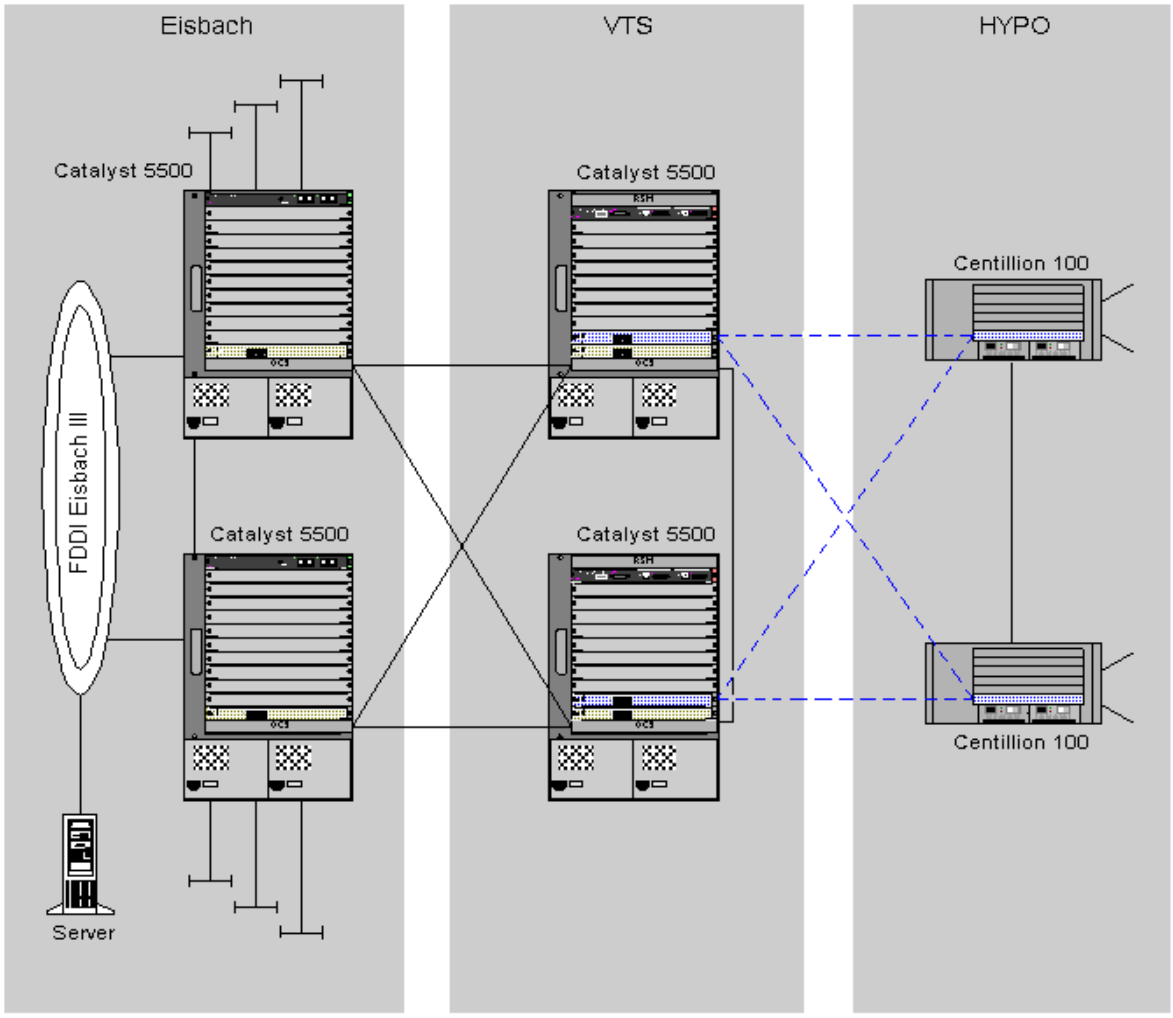
Da es sich hier um Komponenten handelt, die bereits über die entsprechenden hierarchischen PNNI Funktionalitäten gemäß ATM Forum verfügen, wird dadurch der Aufbau eines modularen, skalierbaren und für die Zukunft ausgerichteten ATM Netzes gewährleistet. (Da diese Funktionalität in den bereits installierten Centillion 100 Switchen noch nicht verfügbar war, wurde die Kopplung der beiden ATM Netze vorerst durch die Route Switch Module und die Netflow Feature Card realisiert.)

Desweiteren wird durch den Einsatz des Route Switch Modules in der ersten Phase der Kopplung eine zentrale Kontrollinstanz, mit welcher die Kommunikationsbeziehungen zwischen den beiden Häusern aktiv geregelt werden kann, geschaffen. Im Gegensatz zu Variante I, bei der eine sofortige und vollständige Transparenz beider Netze mit allen ihren Konsequenzen erreicht wird, stellt dieser Lösungsansatz dem Netzbetreiber ein Instrument zur Verfügung, mit dem er in die Lage versetzt wird, je nach Bedarf und Notwendigkeit die benötigten Applikationen nach und nach, durch Setzen von IP-Filtern, seinen Anwender zu Verfügung zu stellen.

Phase 2:

Mit Beginn der zweiten Phase werden die "Eisbach"-Gebäude, in denen sich die Treasury Netze befinden, in den ATM Backbone integriert. Wie schon beschrieben, werden hier die Finanztransaktionen durchgeführt. Die Server aus dem FDDI-Ring werden ins ATM-Netz migriert. Das FDDI-Netz wird somit abgelöst.

Variante II
Phase 2



Durch den Ausbau der Catalyst 5500 Systeme im Vereinsbank-Gebäude "Eisbach 3" mittels ATM-Switch-Prozessor (ASP, siehe Kapitel 6.1.) und ATM OC3 Modulen wird die Migration der in den FDDI Ringen befindlichen Server nach ATM ermöglicht.

Daraus resultiert die Auflösung der FDDI Ringe im "Eisbach"-Gebäude und am Campus.
Gleichzeitig wird hierdurch eine Integration in die hierarchische PNNI Struktur des Backbone und eine Vereinheitlichung der Topologiestruktur ermöglicht.
Durch die Vollvermaschung der Systeme über verschiedene Kabelwege wird gleichzeitig eine sehr hohe Redundanz erreicht.

Phase 3:

In dieser letzten Phase werden die beiden ATM Campi über hierarchisches PNNI zusammengeführt, so daß ein ATM Netz mit mehreren PNNI Peer Groups (Areas) entsteht. Dies führt zu einer für die Zukunft offenen einheitlichen Topologie beider Unternehmen, welche eine optimale Skalierung und Nutzung der ATM Switching Einheiten ermöglicht, und auch großen ATM-Netzerweiterungen, z.B. durch weitere Netzintegrationen von Bankentöchtern, standhält (siehe Kapitel 6.1. Darstellung der momentanen Topologie und Kapitel 6.2. Beispiele für Kommunikation). Durch die geplante Topologie ist es ohne Probleme möglich, mehrere Nebenstellen an den HypoVereinsbank-Campus anzuschliessen. Dies wurde zum Zeitpunkt der Erstellung der Diplomarbeit bereits bei 15 Gebäuden durchgeführt (z.B. Gebäude in der Hohenlindnerstraße in München mit ca. 1.000 Benutzern)

3. Begriffsdefinitionen

Wenn Hersteller von Netzwerkprodukten eine neue Technik auf den Markt bringen, kommt es vor allem in letzter Zeit öfters zu einer Sprachverwirrung, wenn es um die korrekte Beschreibung, bzw. Benennung der Technik geht. Begriffe wie "Layer-3-Switching" oder "Switch" kann man durchaus als "Techno-Jargon" bezeichnen, da sie im OSI-Referenzmodell nicht vorkommen. Tatsächlich handelt es sich in den meisten Fällen um Marketingbegriffe. Da die Produkte verkauft werden müssen und die Herstellerfirmen sich im harten Wettbewerb behaupten müssen, entwickeln die verantwortlichen Vertriebsmenschen einige Kreativität bei der Vermarktung ihrer Produkte, wobei der Name natürlich eine wichtige Rolle spielt. Die Beschreibung einer Technik durch entsprechende Benennung hat in diesem Fall nicht mehr oberste Priorität. Das beste Beispiel hierfür ist die Bezeichnung "Layer-3-Switch": Ein Switch (siehe Begriffsdefinition Switch) arbeitet auf OSI-Schicht 2, auf Layer-3 wird traditionell geroutet. Hier kann man nur vermuten, daß es sich um eine Kombination der Techniken handelt. Hinzu kommt die Tatsache, daß es für diese Technik auf dem Markt mehr als fünf verschiedene Bezeichnungen gibt (detaillierte Beschreibung in Kapitel 3.7. und 5.2.). Diese Sprachverwirrung erschwert natürlich die objektive Untersuchung und Beurteilung der Techniken, bezüglich Nutzen und Relevanz, durch den Kunden. Im folgenden werden einige Begriffe, welche auch anschließend in der Diplomarbeit verwendet werden, durch ihre Funktion beschrieben.

3.1. Repeater

Je länger die Kabel in einem Netzwerk sind, desto schlechter ist die Signalqualität. Repeater sind Signalverstärker. Sie verbinden jeweils zwei Kabeleinheiten miteineinander und passen bei der Signalübertragung die Signalstärke und das Signal-Timing an, so daß auch weiter entfernte Stationen die Daten korrekt empfangen können. Repeater sind auf OSI-Schicht 1, der Bitübertragungsschicht (Physical Layer), einzuordnen.

3.2. Hub

Übersetzt man den Begriff Hub ins Deutsche so erhält man Begriffe wie "Zentrum", "Radnabe", "Mittelpunkt". Den Begriff Hub kann man nicht mit einer speziellen Funktion verbinden, da es sich hierbei lediglich um einen Sternkoppler handelt, welcher verschiedene physikalische Einheiten in einem Punkt vereint. Die Bezeichnung Hub trifft also auf jedes Gerät zu, das über eine Anzahl von Ports verfügt. Normalerweise kann man davon ausgehen, daß, wenn die Rede von einem Hub ist, ein Repeater-Hub gemeint ist. Hierbei handelt es sich um einen Repeater mit mehreren Ports. Der Hub nimmt ein Datenpaket an einem Port an und leitet es an alle Ausgangsports weiter. Gleichzeitig werden alle notwendigen Korrekturen an Signalstärke und Signal-Timing vorgenommen.

3.3. Bridge

Repeater haben den Nachteil, daß sich alle Stationen im Netz, durch die bloße Weiterleitung der Daten, weiterhin die verfügbare Bandbreite (z.B. 10 Mbit/s bei Ethernet) teilen. Die Bridge geht hier einen Schritt weiter. Sie verbindet zwei

Netzwerke miteinander, arbeitet allerdings nach einem anderen Verfahren bei der Weiterleitung der Daten als der Repeater. Die Bridge untersucht jedes Paket, das sie erhält, nach der jeweiligen Absender-MAC-Adresse und merkt sich somit, welche Stationen sich in welchem Segment befinden. Bei jedem ankommenden Paket wird anschließend Quell- und Zielsegment verglichen. Nur wenn die beiden Segmente nicht identisch sind, wird das entsprechende Paket weitergeleitet, ansonsten verworfen. Bridges sind auf OSI-Schicht 2, der Sicherungsschicht (Data Link Layer), einzuordnen.

3.4. Marketingbegriff Switch

"Switch" ist, wie schon erwähnt, ein Marketingbegriff, welcher im OSI-Referenzmodell nicht definiert ist. Ist von einem Switch die Rede, so kann man in der Regel davon ausgehen, daß eine Multiport-Bridge gemeint ist. Alle an dem Switch angeschlossenen Netzwerksegmente bilden ein gemeinsames LAN, wobei zwischen den einzelnen Segmenten "gebridget" wird.
Der Switch empfängt an einem Port ein Datenpaket und entscheidet anhand der Empfänger-Adresse, an welchen Port er es weiterleitet. Gleichzeitig können auch Pakete auf anderen Ports empfangen und weitergegeben werden.

3.5. Router

Ein Router analysiert die Datenpakete nicht anhand ihrer MAC-Adressen, sondern anhand ihrer Protokollinformationen (z.B. von IP) auf OSI-Schicht 3, der Netzwerk- bzw. Vermittlungsschicht (Network Layer). Die Funktionen eines Routers werden detailliert in den Kapiteln 5.1. und 5.2. erklärt.

3.6. Layer-3-Switch

Der Begriff "Layer-3-Switch" ist auch definitiv ein Marketingbegriff, welcher nicht im OSI-Referenzmodell erscheint. Für die Funktion des Layer-3-Switching hat fast jeder Hersteller seine eigene Bezeichnung, sei es Gigabit Routing, Multilayer Switching, Switching Router, Routing Switch, Wirespeed Routing, usw. ([41]) . Die Implementierungen haben allerdings, abgesehen von hertellerspezifischen proprietären Ergänzungen, alle das gleiche Ziel: herkömmliches Routing schneller zu machen. Man kann einen Layer-3-Switch also durchaus als schnellen Router bezeichnen. Traditionelle Router nutzen Mikroprozessoren um die Berechnungen durchzuführen. Hier liegt der Hauptunterschied zum Layer-3-Switch, dessen Hardware aus anwendungsspezifischen Schaltkreisen, sogenannten ASICs (application specific integrated circuit) besteht. Eine detaillierte Beschreibung von Layer-3-Switching kann in Kapitel 5.2. nachgelesen werden.

4. Beschreibung der verwendeten Technologien

4.1. ATM

4.1.1. ATM - Die Technologie

4.1.1.1. Grundlagen

ATM gehört zu der Familie der zellenvermittelnden Datenübertragungs-Systeme (Cell Relay). Im Gegensatz zu paketvermittelnden Systemen (z.B. X.25) ist die Länge der Cell-Relay Datenpakete fest. Die Zellen haben eine Länge von 53 Bytes. Die ersten fünf der 53 Bytes sind für den Zellen-Header reserviert, der unter anderem Kanal- und Pfad-Adressierungen enthält. Alle Netzknoten sind über eine oder mehrere ATM-Schalteinheiten (ATM-Switches) miteinander verbunden. Die ATM-Switches vermitteln die Zellen an ihren jeweiligen Zielort. ATM repräsentiert die Implementation von Cell-Relay, die für die Spezifikation des B-ISDN-Standards ausgewählt wurde. ATM ist also ein Teil der ITU-Spezifikation für B-ISDN.

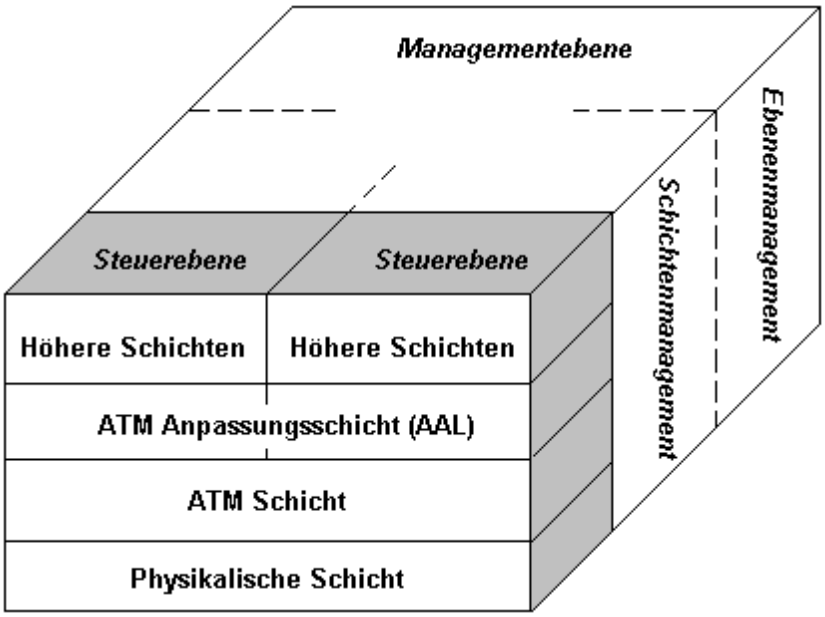
Die Netzteilnehmer im ATM-Netz teilen sich nicht ein gemeinsames Übertragungsmedium, wie das bei den lokalen Netzen der Fall ist, sondern sie schicken ihre ATM-Zellen an die nächste ATM-Schalteinheit, ohne Abhängigkeit von Zugriffsalgorithmen (media access control). Die insgesamt zur Verfügung stehende Übertragungsbandbreite wird von der ATM-Schalteinheit nach Bedarf verteilt (hierbei kann auch einzelnen Benutzern dediziert eine feste Übertragungsbandbreite zugewiesen werden).

Die Festlegung der Zellenlänge auf 53 Bytes ist das Resultat eines Kompromisses zwischen den Anforderungen von analoger Sprach- und digitaler Datenübertragung.
Für die Übertragung rein analoger Signale ist eine möglichst geringe Zellenlänge von Vorteil. In den europäischen Länder hat der reine Datenverkehr noch nicht die Bedeutung wie in den USA. Diese Tatsache führte zu dem Vorschlag der europäischen Gremien, die Zellenlänge der ATM-Zellen auf 32 Bytes festzulegen. Bei der Datenübertragung in Nordamerika spielt allerdings die Übertragung über Weitverkehrsnetze eine wesentlich wichtigere Rolle als in Europa. Nordamerika versuchte also eine Zellenlänge von 64 Bytes durchzusetzen. Größere Datenmengen werden aufgrund des geringeren Overheads günstiger in langen Datenpaketen übertragen. Die Diskussion führte schließlich zum Kompromiß von 53 Bytes. Die Zellenlänge von 53 Bytes ist für die Übertragung von analogen Signalen noch optimal geeignet. Auch für die Datenübertragung sind 53 Bytes bei einem Overhead-Anteil von 9,4 % akzeptabel um reine Datenströme ökonomisch zu übertragen. Die ATM-Architektur bietet eine Reihe von Vorteilen im Vergleich mit herkömmlichen Übertragungsmethoden.

4.1.1.2. B-ISDN-Schichtenmodell / ATM

B-ISDN (Broadband Integrated Systems Digital Network) ist der Versuch der ITU (ursprünglich CCITT) eine weltweit einheitlich aufgebaute Struktur für Hochgeschwindigkeitsnetze zu schaffen. Das universelle Netzwerk ist in der Lage die heute bestehenden Sprach-, Daten- und Videoübertragungen zu integrieren. Als

Datentransportmechanismus wird ATM eingesetzt. Die Standardisierung wurde von der CCITT 1990 begonnen. Die B-ISDN-Netzwerkarchitektur besteht aus vier Kommunikationsschichten, der physikalischen Schicht, der ATM-Schicht, der ATM-Adaptions-Schicht (AAL) und der Anwendungsschicht, welche über die Benutzer-, Steuer- und Managementebene verbunden sind (siehe Abbildung).



B-ISDN Protokoll-Referenzmodell ([3], S.84)

4.1.1.2.1. Beschreibung der Ebenen

Innerhalb der **Benutzerebene** (User Plane) findet der Informationsfluß über alle Schichten hinweg statt. Hierbei werden Mechanismen zur Fehlerkorrektur und Datenflußüberwachung eingesetzt.

Die **Steuerebene** (Control Plane) ist für den Auf- und Abbau von Verbindungen zuständig. Der VPI/VCI-Header von ATM-Zellen, welcher zu einer eindeutigen Zuordnung von Zellen zu einer Verbindung notwendig ist, wird durch das Signalisierungsverfahren der Steuerebene generiert und zugeordnet.

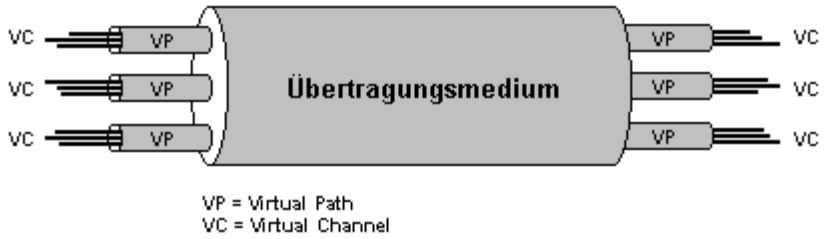
In der Abbildung ist zu sehen, daß die **Managementebene** zwei Funktionen beinhaltet: Ebenenmanagement (Plane Management) und Schichtenmanagement (Layer Management). Das Ebenenmanagement ist zuständig für die Koordination der zwei beschriebenen Ebenen mit der Managementebene. Das Schichtenmanagement ist, unter anderem, verantwortlich für Meta-Signalisierung (eigener Informationskanal zur Steuerung der Signalisierung) und den OAM-Informationsfluß (OAM = Operation and Maintenance). OAM-Daten werden zur Überwachung der Leistungsfähigkeit des ATM-Netzes und zum Fehlermanagement eingesetzt. Hierfür existieren spezielle OAM-Zellen.

4.1.1.2.2. Beschreibung der Schichten

Die **physikalische Schicht** wird in die zwei Bereiche Übertragungsanpassung (Transmission Convergence TC) und physikalisches Medium aufgeteilt, wobei die

Übertragungsanpassung die Aufgabe besitzt, die Zellen der ATM-Schicht in die Übertragungsrahmen des jeweiligen Transportmediums einzubetten und das Header-Error-Control-Byte (HEC) zu generieren. Das HEC-Byte ist eine Prüfsumme über die Headerinformationen der Zelle und wird im fünften Byte abgespeichert. Als physikalisches Medium wird meistens, vor allem wenn die Länge der Übertragungsstrecken über den lokalen Bereich hinausgehen, auf Glasfaserkabel zurückgegriffen.

Die **ATM-Schicht** hat die Aufgabe, die von der übergeordneten Anpassungsschicht gelieferten Daten an den Bestimmungsort zu transportieren. Die Daten werden in ATM-Zellen übertragen, welche durch den, im Zellenheader befindlichen, VPI/VCI-Identifikator (VPI/VCI = Virtual Path Identifier / Virtual Channel Identifier) einer Verbindung eindeutig zugeordnet werden. Zellenströme werden in zwei logische Hierarchien, ATM-Kanal und ATM-Pfad, unterteilt. Physikalische Übertragungsmedien können mehrere virtuelle Verbindungspfade übertragen, wobei jeder Verbindungspfad wiederum mehrere Verbindungskanäle beinhalten kann (siehe Abbildung).



ATM-Kanäle und ATM-Pfade

Die Aufgabe der ATM-Schicht ist es nun, für die Daten, welche sie von der übergeordneten AAL-Schicht erhält, dazugehörige ATM-Header zu erzeugen. Für die Kommunikation zwischen zwei Netzwerkkomponenten (ATM-Switches) werden NNI-Header und für die Kommunikation zwischen Benutzer und Netzwerkkomponente UNI-Header generiert (siehe Abbildung). Das GFC-Feld dient zur lokalen Flußkontrolle und ist deshalb nicht im NNI-Header enthalten.

Zellenheader für ATM-UNI-Zellen (UNI = User Network Interface)

1	2	3	4	5	6	7	8	Bit / Byte
Datenflußkontrolle (GFC)				Pfadidentifikation (VPI)				1
Pfadidentifikation (VPI)				Kanalidentifikation (VCI)				2
Kanalidentifikation (VCI)								3
Kanalidentifikation (VCI)				Payload Type (PT)			CLP	4
Headerprüfsumme (HEC)								5

Payload Type (PT) = Nutzlastidentifikator
Cell Loss Priority (CLP) = Zellenverlust-Priorität

Zellenheader für ATM-NNI-Zellen (NNI = Network Network Interface)

1	2	3	4	5	6	7	8	Bit / Byte
Pfadidentifikation (VPI)								1
Pfadidentifikation (VPI)				Kanalidentifikation (VCI)				2
Kanalidentifikation (VCI)								3
Kanalidentifikation (VCI)				Payload Type (PT)			CLP	4
Headerprüfsumme (HEC)								5

Payload Type (PT) = Nutzlastidentifikator
Cell Loss Priority (CLP) = Zellenverlust-Priorität

Vergleich UNI-Header mit NNI-Header ([3], S.87)

Die **ATM-Anpassungsschicht** (AAL = ATM-Adaptation Layer) ist in zwei Teilschichten aufgeteilt: die Konvergenzteilschicht (Convergence Sublayer CS) und die Segmentierungs- und Reassemblierungsteilschicht (Segmentation and Reassembly Sublayer SAR). Sie hat die Aufgabe, die Informationseinheiten der höheren Schichten auf die Zellenstruktur der ATM-Schicht abzubilden. Die SAR-Teilschicht zerteilt die Daten in 48 Bytes lange Fragmente (Segmentation), beziehungsweise setzt aus den ATM-Zellen wieder die ursprünglichen Datenströme zusammen. Die Konvergenzteilschicht beinhaltet Funktionen, welche den reibungslosen Ablauf der Datenkommunikation überwachen, wie beispielsweise:

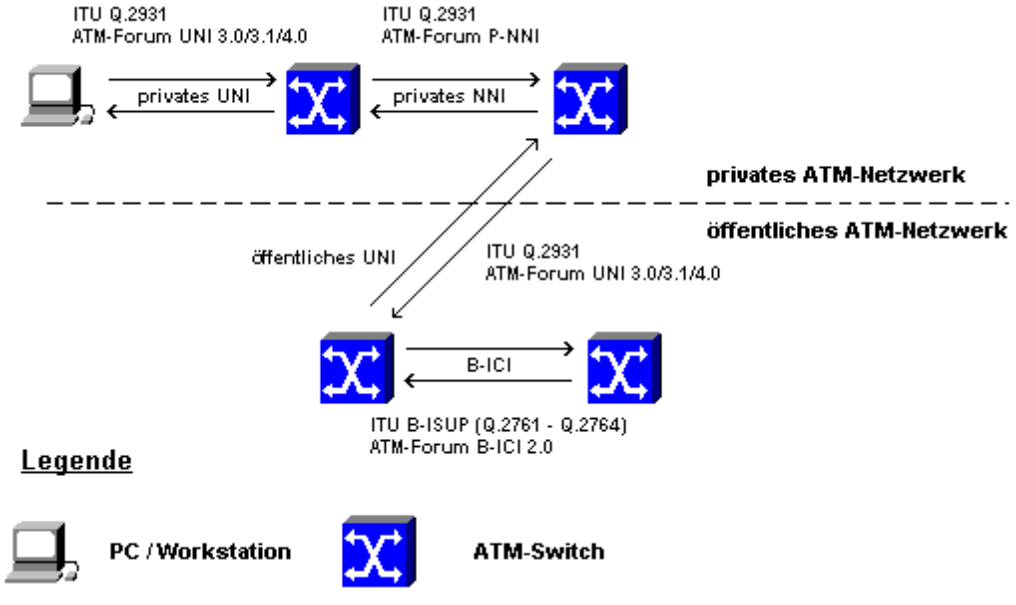
- Überprüfung des AAL-Protokollheaders auf Bitfehler
- Handhabung von fehlerhaften AAL-Protokoll-Headern
- Handhabung bei Überlauf des Pufferspeichers
- Handhabung bei Verlust der Synchronisation, bzw. Takt.

Die Funktionen der ATM-Anpassungsschicht (AAL) sind an die Charakteristik der übergeordneten Anwendungen gekoppelt. Hierfür wurden vier verschiedene Service-Klassen für die AAL-Schicht definiert: AAL1, AAL2, AAL3/4 und AAL5. Die Definition der verschiedenen Service-Klassen basiert auf den folgenden drei Parametern:

- Echtzeitanforderung
- Bitrate (konstant / variabel)
- Verbindungskonzept (verbindungsorientiert / nicht verbindungsorientiert)

In die ATM-Anpassungsschicht ist auch die Anpassungsschicht für Signalisierung integriert (SAAL), welche die Aufgabe hat, die Signalisierungsprotokolle der höheren Ebenen auf die ATM-Ebene umzusetzen. Signalisierung ist die Prozedur, nach der Verbindungen auf- und abgebaut werden. Hierbei werden auch Parameter wie beispielsweise der AAL-Typ, Bitrate, maximale Zellverlustrate oder die maximal zulässige Verzögerung festgelegt. Die ITU hat für die Signalisierung in B-ISDN-Netzen die Spezifikation Q.2931 entwickelt, welche für den Verbindungsaufbau über öffentliche und private UNI-Schnittstellen sowie private NNI-Schnittstellen gedacht ist. Weiterhin wurde von der ITU das B-ISUP-Protokoll in Q.2761 - Q.2764 spezifiziert, welches als Vermittlungsprotokoll für die B-ICI-Schnittstelle (B-ICI = Broadband-Inter-Carrier-Interface) in öffentlichen ATM-Weitverkehrsnetzen dient (siehe Abbildung). Die Signalisierungsspezifikationen des ATM-Forums, welche sich den Vorgaben der ITU anlehnen, sind folgende:

- UNI 3.0 / UNI 3.1 / UNI 4.0 Signalisierung für die UNI-Schnittstelle (UNI = User Network Interface)
- P-NNI Signalisierung für die private NNI-Schnittstelle (NNI = Network Network Interface)
- B-ICI 1.0 / B-ICI 2.0 Signalisierung für die öffentliche NNI-Schnittstelle



ATM-Signalisierung ([3], S.212)

4.1.1.3. ATM in LAN- und WAN-Welt

Datenkommunikation kann man in die zwei völlig unterschiedlichen Bereiche "Datenübertragung in Lokalen Netzwerken" und "Datenübertragung in Weitverkehrsnetzen" unterteilen. Die Datentransportmechanismen unterscheiden sich, in den einzelnen Bereichen, aus übertragungstechnischer Sicht wesentlich

voneinander. Die Datenkommunikation in Weitverkehrsnetzwerken basiert auf verbindungsorientierter Kommunikation. Hierbei wird eine dedizierte Verbindung zur Partnerstation aufgebaut, bevor die Datenübertragung beginnen kann. In lokalen Netzen basiert die Kommunikation hauptsächlich auf verbindungslosen Broadcastsendungen. Alle Stationen im lokalen Netz empfangen die gesendeten Daten, wobei die Empfängerstation die für sie bestimmten Daten herausfiltert. Aufgrund der geringen Datenverlustrate werden keine Empfangsbestätigungen verschickt.

Diese Trennung der Datenkommunikationsmethoden wird durch ATM aufgehoben. ATM entwickelt sich zu einer universellen Plattform für Datenübertragung. Die Technologie, die ursprünglich für die Übertragung von Weitverkehrsdaten konzipiert war, wird mittlerweile auch erfolgreich in lokalen Netzen, als Konkurrent zu Gigabit-Ethernet, eingesetzt. In ATM-LAN- und ATM-WAN-Netzwerken erfolgt die Übertragung identisch nach dem bereits beschriebenen Zellenvermittlungsschema. Zellen lassen sich, wie schon erwähnt, wesentlich effizienter und schneller als Datenpakete variabler Länge in Vermittlungseinheiten verarbeiten. Die feste Zellenlänge ermöglicht eine massiv parallele Architektur von Zellen-Vermittlungseinheiten. Zellen, die zur selben Zeit an den Eingangsports einer ATM-Vermittlungseinheit eingehen, können im Takt gleichzeitig an die gewünschten Ausgangsports vermittelt werden. ATM ist aufgrund seiner Architektur und der daraus resultierenden Übertragungseigenschaften die einzige standardisierte Übertragungstechnik, die in der Lage ist, alle heute bestehenden Datendienste wie Sprach-, Daten- und Videoübertragung in effizienter Weise zu vermitteln.

Der immer höhere Bandbreitenbedarf von immer mehr miteinander vernetzten und kommunizierenden Benutzern steigert kontinuierlich den Bedarf an Hochgeschwindigkeitsnetzwerken. Vor zwanzig Jahren dominierten noch Terminal-Host-Anwendungen die Netzwerke. Die Applikationen liefen auf dem Zentralrechner und konnten über die ASCII-basierenden Terminals bedient werden, wobei jedes Terminal über eine eigene Datenleitung an den zentralen Hostcomputer angeschlossen war. Aufgrund der zeichenbasierenden Anzeige war der daraus resultierende Bandbreitenbedarf gering. Dieser Bandbreitenbedarf stieg zum erstenmal sprunghaft an als man anfing Personal-Computer, ausgestattet mit eigenen Prozessoren und Speicherplatz, einzusetzen und diese miteinander zu vernetzen. Dieser Bedarf erhöhte sich weiter deutlich durch die Einführung von graphischen Benutzeroberflächen (Microsoft-Windows, X-Windows). Die Einführung und Nutzung von Videoapplikationen (Videokonferenzen,etc.) auf breiter Basis steigerte in den letzten Jahren den Bedarf an Übertragungsbandbreite nochmals enorm, wie auch die Anforderungen an das Echtzeitverhalten. Anwendungen beanspruchen mittlerweile dedizierte Bandbreiten im Bereich von mehreren Megabit pro Sekunde.

Um den Durchsatz bei traditionellen Netzwerktechnologien wie Ethernet und Token-Ring zu erhöhen, muß zur Anpassung an die immer höheren Netzlasten der einzig variable Parameter - die Größe der Netzwerksegmente - minimiert werden. Obwohl man die Segmentgröße bis auf einen PC reduzieren kann (Switched LANs, jedem PC wird dediziert die verfügbare Bandbreite zur Verfügung gestellt), stößt man bei den traditionellen LAN-Technologien auf Leistungsgrenzen, die die Übertragung von Echtzeitdaten nicht gewährleisten (siehe Leistungsgrenzen von Ethernet / Token-

Ring, FDDI). Übergangsweise finden derzeit Brückentechnologien wie die 100-Mbit/s-Varianten l00Base-T und FDDI noch ihre Einsatzgebiete. Sie sind allerdings nicht in der Lage den exponentiell ansteigenden Bandbreitenbedarf mittel- bis langfristig zu bewältigen. Echtzeitorientierte Multimedia-Applikationen finden immer mehr Verbreitung, so daß diese Topologien in naher Zukunft in den Einsatzbereichen wieder verschwinden wird, in denen spezielles Echtzeitverhalten garantiert sein muß.

4.1.1.4. Multiplexing

ATM funktioniert nach dem Prinzip des asynchronen Zeitmultiplexing. Was ist nun Multiplexing? Durch Multiplexing ist es möglich, mehrere voneinander unabhängige Datenströme über ein und dasselbe physikalische Medium zu übertragen. Dieses Konzept wurde erstmals von den Telefongesellschaften in den dreißiger Jahren realisiert. Bei den ersten Versuchen wurden zunächst 12 und später 60 Sprachkanäle mit Hilfe eines Frequenz-Multiplex-Verfahrens über ein Koaxialkabel übertragen. Vor der Übertragung wurden dabei die Sprachkanäle auf eine bestimmte Übertragungsfrequenz aufmoduliert und nach der gleichzeitigen Übertragung des gesamten Frequenzspektrums durch Demodulation wieder zurückgewonnen. Die enormen Einsparungen, die diese Technik mit sich brachte, liegen auf der Hand. Anstatt für jede Verbindung ein Kabel einzusetzen konnten nun mehrere Gespräche über eine Leitung vermittelt werden. Die Idee, beim Einsatz von Multiplex-Verfahren , ist also, die tatsächlich möglichen Ressourcen eines Übertragunsmediums effizient aufzuteilen, so daß mehrere Datenquellen dieses Medium parallel nutzen können. Heute gibt es eine Vielzahl von Multiplexverfahren wie Wellenlängen-, Code-, Frequenz- oder Zeit-Multiplexing. Zeit-Multiplex-Verfahren (Time-Division-Multiplexing TDM) können in synchrones und asynchrones Multiplexing unterteilt werden.

Bei synchronem Zeit-Multiplexing werden Übertragungsrahmen, welche aus einer bestimmten Anzahl von Zeitschlitzen fester Größe bestehen, definiert. Jedem Benutzer wird ein bestimmter Zeitschlitz innerhalb des Übertragungsrahmens zugeordnet. In diesem Zeitintervall kann der Benutzer senden und empfangen. Somit ist über die Position des Zeitschlitzes innerhalb des Übertragungsrahmens der Übertragungskanal identifizierbar. Synchrones Zeit-Multiplexing wird deshalb manchmal auch als Positions-Multiplexing bezeichnet. Ein Übertragungskanal befindet sich, in Bezug auf den Übertragungsrahmen, immer an derselben Stelle - deshalb die Bezeichnung "synchrones" Zeitmultiplexing.

Der Unterschied bei asynchronem Zeit-Multiplexing besteht darin, daß die Daten in Zeitschlitzen variabler Länge übertragen werden. Damit die Sendekanäle eindeutig identifizierbar sind, ordnet die Vermittlungseinheit den jeweiligen Daten in den Zeitschlitzen entsprechende Identifikationsnummern pro Kanal zu (Channel Identifiers). Dieses Verfahren gibt dem asynchronen Zeit-Multiplexing auch den Namen Adressen-Multiplexing (Label Multiplexing). Werden bei der Übertragung Datenpakete variabler Länge eingesetzt, spricht man von Paket-Vermittlung (Packet-Switching). Haben diese hingegen eine feste Länge, so werden sie als Zellen bezeichnet, und man spricht von Zellen-Vermittlung (Cell Switching).

4.1.1.5. Vermittlungsarten

Aber was versteht man nun eigentlich konkret unter Vermittlung ? Unter Vermittlung versteht man die Art und Weise, mit der ein Übertragungspfad zwischen Sender und Empfänger in einem Kommunikationsnetzwerk ermittelt wird. Hier gibt es wiederum zwei Methoden: Leitungsvermittlung (Circuit-Switched) und Paketvermittlung (Packet-Switched). Bei der Leitungsvermittlung wird zwischen Sender und Empfänger eine physikalische Verbindung aufgebaut, wie bei einem Telefongespräch. Dies geschieht über eine oder mehrere Vermittlungseinheiten. Der Übertragungspfad muß feststehen bevor Daten zwischen den Teilnehmern übertragen werden können. Ist die Verbindung etabliert, so verzögert nur noch die endliche Ausbreitungsgeschwindigkeit der elektromagnetischen Signale den Ablauf der Datenübertragung (ca. 6 ms / 1000 km).

Wird die zu übertragende Information in Datenpakete umgewandelt und über das Netzwerk übertragen, so spricht man von paketorientierter Vermittlung. Pakete können hierbei in Vermittlungseinheiten zwischengespeichert werden. Es besteht keine feste physikalische Verbindung zwischen Sender und Empfänger. Das Verfahren kann man sich ähnlich wie das Versenden mehrerer Briefe an eine Person vorstellen. Auch wenn man die Briefe an hintereinander folgenden Tagen einwirft ist nicht garantiert, daß sie in der entsprechenden zeitlichen Reihenfolge beim Empfänger ankommen. Die Pakete müssen also zusätzlich eine Nummerierung, die sogenannte Sequenznummer, beinhalten.

Wird eine feste Verbindung über Leitungsvermittlung hergestellt, so wird der Verbindung eine bestimmte Bandbreite fest zugeordnet, z.B. 64 kbit/s bei ISDN-Gesprächen bzw. -Datentransfer. Bei Paketvermittlung wird die Bandbreite bedarfsorientiert angefordert. Wird z.B. an einem Tag wesentlich mehr Post für eine Region geliefert als sonst, so wird an diesem Tag für diese Region ein zusätzlicher Post-LKW mit Briefträgern zur verfügung gestellt. Im Fall der Leitungsvermittlung ist somit die von einer Verbindung nicht benötigte Bandbreite verloren (theoretisch könnte man über die ISDN-Leitung parallel bis zu 8 Gespräche führen). Der limitierende Faktor in paketvermittelten Netzen ist die Rechen- und Speicherkapazität der vermittelnden Einheiten. Ungenutzte Übertragungskapazitäten stehen somit in paketvermittelten Netzen allen sendewilligen Stationen zur Verfügung. Aufgrund des stark variierenden Bandbreitenbedarfes in der Datenkommunikation werden Computernetzwerke deshalb hauptsächlich paketvermittelt betrieben.
Ein Problem der Paketvermittlung ist allerdings, daß Vermittlungseinheiten bei Speicherüberlastung Pakete "wegwerfen". In diesem Fall sind die beiden Endstationen für die reibungslose Kommunikation ohne Datenverlust, bzw. Übertragung von Duplikaten zuständig. Dieser Punkt führt zur Frage: Mit welcher Wahrscheinlichkeit kommen meine Daten korrekt beim Empfänger an, wie sicher ist also die Verbindung ?

4.1.1.6. Verbindungsorientierte und verbindungslose Kommunikation

Jede Schicht des OSI-Modelles kann den darüberliegenden Schichten zwei Arten von Verbindungsdiensten anbieten: verbindungsorientierte Kommunikation und verbindungslose Kommunikation.

Bei verbindungsorientierter Kommunikation wird zunächst eine virtuelle Verbindung zwischen den Kommunikationspartnern etabliert (mittels Handshake-Verfahren). Über diese Verbindung werden die Daten transferiert. Übertragungsfehler (Überlastung des Pufferspeichers, Empfang fehlerhafter Pakete, etc.) werden durch Einsatz von Sende- bzw. Empfangsbestätigungen (positive acknowledgement bei Erhalt der Daten, bzw. negative acknowledgement bei fehlerhaften oder nicht eingegangenen Datenpaketen) korrigiert. Die Qualität der Datenleitungen im Weitverkehrsbereich wächst zunehmend. Diese Tatsache führt dazu, daß die verbindungsorientierten Schicht 2-Übertragungsverfahren ihre hohe Bedeutung im Weitverkehrsbereich verlieren. Die Fehlerkontrolle geschieht mehr und mehr in höheren OSI-Schichten, auch in der Anwendungsschicht.

Bei verbindungsloser Kommunikation werden die Datenpakete mit der kompletten Zieladresse ausgestattet und ohne Empfangsbestätigung unabhängig voneinander versendet. Die Übertragungsreihenfolge der Pakete wird hierbei nicht notwendigerweise eingehalten. Für die korrekte Zusammensetzung der Pakete ist der Empfänger zuständig. Vorteile der verbindungslosen Kommunikation sind geringerer Verwaltungs-Overhead in den Datenpaketen und die gesparte Zeit, die für einen Verbindungsaufbau nötig wäre.

4.1.2. Leistungsgrenzen von Ethernet-Netzwerken

Die Übertragungsmethode, die auch heute noch standardmäßig in Ethernet-Netzwerken eingesetzt wird ist das Half-Duplex-Verfahren (mittlerweile werden aber immer mehr Ethernet-Netzwerke mit Full-Duplex betrieben). Zur Senderechtsvergabe wird das non-deterministische CSMA/CD-Verfahren eingesetzt. Die Bezeichnung "non-deterministisch" (nicht vorhersagbar) weist hier auf eine wichtige Eigenschaft von Ethernet-Netzwerken hin: Es gibt keine festgelegte garantierte Maximalzeit, nach der eine sendewillige Station den Versand der Daten starten kann. Diese Tatsache ist natürlich ein absolutes K.O.-Kriterium in Bezug auf Multimediaübertragung mit Echtzeitverhalten.

In Halb-Duplex-Netzen können die Stationen nicht gleichzeitig senden und empfangen. Sendet eine Station, so müssen sich alle anderen Stationen im Empfangsmodus befinden. Senden zwei Stationen gleichzeitig, so kollidieren die beiden Datenpakete und werden dabei zerstört. Nach einer Kollision müssen beide daran beteiligten Stationen eine durch einen Zufallsgenerator bestimmte Zeit warten, bevor sie einen erneuten Sendeversuch starten dürfen. Nimmt also die Anzahl der Netzteilnehmer zu, so wächst auch die Wahrscheinlichkeit von Kollisionen. Von entscheidender Bedeutung ist auch die Länge des Netzsegmentes. Mit zunehmender Länge wird nämlich der zentrale Parameter des CSMA/CD-Algorithmus - die sogenannte Slotzeit - größer. Die Slotzeit ist gleich zweimal der Signallaufzeit zwischen den beiden am weitesten voneinander entfernten Stationen. Diese Zeit benötigt eine Station im schlimmsten Fall, um zu erkennen, daß eine Kollision von Datenpaketen aufgetreten ist.

Ist die Netzlast gering, so funktioniert der CSMA/CD-Algorithmus einwandfrei gut. Ab einer Netzlast von 40%-50% nimmt die Ausnutzung der zur Verfügung stehenden Bandbreite von 10 Mbit/s allerdings dramatisch ab ([3], S.55). Die Ausnutzung der

(theoretisch) vorhandenen Übertragungsbandbreite von 10 Mbit/s ist um so effizienter, je größer die übertragenen Datenpakete sind. Dies begründet sich dadurch, daß bei der Übertragung größerer Datenpakete weniger Kollisionen auftreten.

Aus den genannten Gründen wird ersichtlich, daß Ethernet für die Übertragung von Multimedia-Applikationen denkbar schlecht geeignet ist. In der Beschreibung von ATM wurde zum Beispiel erwähnt, daß für die Übertragung von analogen Signalen viele kurze Pakete besser geeignet sind als wenige lange (das genaue Gegenteil gilt für die Leistungsfähigkeit von Ethernet-Netzen). Ein Beispiel dafür ist die Übertragung von analoger Sprachinformation durch Abtastung von analogen Sprachsignalen mit einer Frequenz von 8 kHz ([3], S.56) :

"Jeder Abtastwert wird mit 8 Bits kodiert, was in einem resultierenden Bitstrom von 64 kbit/s resultiert. Alle 125 µs wird dabei ein Meßwert übertragen. Man spricht von einer Abtastrate von 125 µs. Wird jeder 8 Bit große Abtastwert nun innerhalb eines 64 Bytes großen Datenpaketes übertragen, so wird bei derselben Übertragungsrate von 64 kbit/s lediglich eine Abtastrate von 8 ms erzielt. Um mit 64 Byte-Paketen dieselbe Abtastrate zu erzielen, ist eine Bandbreite von 4 Mbit/s notwendig, mit 512 Byte Paketen gar 32 Mbit/s. Man sieht, je größer die kleinste mögliche Dateneinheit ist, desto ineffizienter wird die Übertragung analoger Signale. Jedes Datenpaket ist zwar lediglich zu einem Bruchteil mit Daten belegt, aufgrund der Anforderungen an die maximal zulässige Übertragungsverzögerung kann aber nicht mit der Versendung eines Datenpaketes gewartet werden, bis es mit Abtastwerten aufgefüllt ist."

Die Anforderungen von Multimedia-Applikationen sind vor allem

- vorhersagbare (deterministische), minimale Übertragungsverzögerung (Jitter)
- auch bei kleinen Paketgrößen muß eine effiziente Übertragung garantiert sein
- die Möglichkeit der Priorisierung von Datenströmen
- hohe Bandbreiten, die auch dediziert von Anwendungen angefordert werden können.

Diese Anforderungen werden von Ethernet-Netzwerken in keinem Punkt erfüllt. Die stark variierenden Übertragungsverzögerungen machen eine Echtzeitübertragung unmöglich. Hinzukommt die Tatsache, daß aufgrund des non-deteministischen Charakters keine Minimal-Bandbreiten garantiert werden können. Die maximale Effizienz bei steigender Paketgröße ist ein weiteres Ausschlußkriterium bezüglich der Übertragung von Multimedia-Daten.

Natürlich sehen die Produktionsfirmen von Ethernet-Komponenten sowie die Organisation IEEE den Marktentwicklungen nicht tatenlos zu. Ethernet soll multimediafähig werden. Erste Ansätze liefert Gigabit-Ethernet im Full-Duplex Betrieb. Protokolle wie RSVP sollen Anwendungen dedizierte Bandbreite und Priorisierung garantieren. Ob sich diese Entwicklung durchsetzt werden die nächsten Jahre zeigen. Die Entwicklung, ob sich eine Technik im Netzwerkbereich durchsetzt ist vor allem durch zwei Kriterien beeinflusst: Standards und Marktpreis. Vor allem beim Preis besitzt Gigabit-Ethernet aufgrund der einfachen Struktur und Implementierung wesentliche Vorteile gegenüber ATM.

4.1.3. Leistungsgrenzen von Token-Ring- und FDDI-Netzwerken

Im Unterschied zu Ethernet-Netzwerken basieren Token-Ring und FDDI auf deterministischen Übertragungsverfahren. Es existiert eine definierte Maximalzeit, nach der eine Station spätestens senden darf. Innerhalb des Netzes rotiert eine spezielle Bitsequenz (Token), die zur Steuerung des Senderechtes benutzt wird, zwischen den Stationen. Erhält eine Station den Token, so ist sie zum Versenden von Daten berechtigt. Hierbei werden die Daten mit dem Token gesendet. Nach einer definierten maximalen Zeit (Token-Holding-Time) muß der Token an die nachfolgenden Stationen weitergegeben werden. Die Datenpakete rotieren nun durch das Netz, wobei jede Station untersucht ob die Daten an sie gerichtet sind. Die tatsächliche Empfängerstation kopiert die Daten und setzt die beiden Headerbits "Adresse erkannt" und "Paket kopiert" auf eins. Die modifizierten Daten rotieren nun weiter durch das Netz, bis sie wieder zum Sender gelangen. Die sendende Station wertet die Headerbits aus und nimmt die Daten bei korrekter Übertragung vom Ring.

Da ausschließlich jene Station, die den Token besitzt, senden darf, kommt es in Token-Ring- sowie FDDI-Netzwerken zu keinen Kollisionen. Aus diesem Grund nutzen diese Netzwerke die theoretisch verfügbare Bandbreite wesentlich besser aus als Ethernet-Netzwerke. Erst bei einer Netzlast von mehr als 80% verdoppelt sich die Zeit, die eine Station warten muß, bis sie in Besitz des Tokens kommt.

Das Grundkonzept von Token-Ring und FDDI passt also wesentlich besser zu den heutigen Anforderungen an moderne Netzwerke als dies bei Ethernet-Netzwerken der Fall ist. Trotzdem besitzen auch diese deterministischen Verfahren Schwachstellen in Bezug auf Multimedia-Übertragung. Wie schon erwähnt wurde, bestimmt die Token-Holding-Time die maximale Zeit, die eine Station zum Senden zur Verfügung steht, bevor sie den Token weitergeben muß. Der Standardwert für die Token-Holding-Time ist bei Token-Ring 10 ms und bei FDDI 4 ms. Befinden sich nun beispielsweise 25 PCs im Ring, so beträgt die maximale Zeit, die eine Station bis zum Erhalt der Sendeberechtigung warten muß bei Token-Ring 250 ms und bei FDDI 100 ms. Diese Werte sind um ein Vielfaches höher als der für den Betrieb von Multimedia-Aplikationen höchstzulässige Wert von etwa 10 ms ! ([3], S.58)

Dieses Beispiel zeigt, daß sogar ein Hochgeschwindigkeitsnetz wie FDDI nicht in der Lage ist, im Regelbetrieb Multimedia-Anwendungen zu übertragen. Auch für FDDI werden Anpassungen in Richtung Echtzeitfähigkeit realisiert. Allerdings wird die Technik FDDI in Zukunft, aufgrund der Konkurrenten ATM und Gigabit-Ethernet, immer weniger eingesetzt werden.

4.1.4. Aufhebung der Leistungsgrenzen durch ATM

Die Beschreibung der Leistungsgrenzen von Ethernet, Token-Ring und FDDI zeigt das Hauptproblem auf: Alle Stationen des Netzwerkes teilen sich das Übertragungsmedium; es handelt sich in diesen Fällen um ein sogenanntes "Shared-Medium". Nur jeweils eine Station ist sendeberechtigt, die anderen Stationen sind während der Sendung blockiert. Es stellt sich nun die Frage, mit welcher Technologie die bereits beschriebenen Anforderungen an ein multimedia-, bzw. echtzeitfähiges Netzwerk zu bewältigen sind. Zunächst muß garantiert sein, daß

theoretisch alle Stationen auf die verfügbare Bandbreite parallel zugreifen können. Die Bandbreite wird dann entsprechend aufgeteilt, je nachdem wieviel Stationen tatsächlich senden möchten. Weiterhin muß es eine garantierte maximale Verzögerungszeit geben, damit die Übertragung von Echtzeitdaten realisierbar ist. Zum Beispiel ist die Übertragung von Audiodaten zeitkritisch. Treten hierbei Verzögerungen während der Übertragung auf, so daß Datenpakete nicht innerhalb des akzeptierbaren Zeitrahmens beim Empfänger eintreffen, so sind diese aus Anwendungssicht nutzlos und werden verworfen. Die Tonqualität leidet in diesem Fall unter dem Teilverlust der Audiodaten, was relativ schnell zur Überschreitung der Toleranzschwelle des Benutzers für die entsprechende Anwendung führt.

ATM hebt die beschriebenen Leistungsgrenzen auf. Die angeforderte Bandbreite wird einzelnen Stationen, bzw. Anwendungen dediziert zugewiesen und steht den beiden kommunizierenden Stationen während der ganzen Übertragung zur Verfügung. Da es sich bei ATM-Verbindungen, im Gegensatz zu den besprochenen LAN-Technologien, um Punkt-zu-Punkt Verbindungen handelt, gibt es außer der Signallauf-, Puffer- und Vermittlungszeit keine Verzögerungen. Zusätzlich bietet ATM mittlerweile Durchsatzraten im Bereich von 622 Mbit/s bis 2,5 Gbit/s, wobei diese noch keine Obergrenzen darstellen. Diese Eigenschaften machen ATM zu einer ausgezeichneten Lösung wenn man ein schnelles Netz mit dedizierter Bandbreitenzuordnung und Echtzeitfähigkeit wünscht. Ein Nachteil ist der hohe Preis den die aktiven Komponenten kosten. Durch die dedizierte Bandbreitenzuordnung muß die ATM-Schalteinheit (ATM-Switch) selbst die Fähigkeit besitzen enorme Vermittlungsgeschwindigkeiten realisieren zu können. Will man beispielsweise 100 Endstationen parallel dediziert 10 Mbit/s pro Station zuweisen, so muß der ATM-Switch eine interne Vermittlungskapazität von 1 Gbit/s besitzen. Bei der HypoVereinsbank wird die ATM-Technologie bereits mit Erfolg eingesetzt.

Die zur Verfügung stehenden Übertragungsbandbreiten der bekannten Weitverkehrstechnologien (Frame-Relay, X.25 oder ISDN) sind für moderne Applikationen um Größenordnungen zu gering. ISDN ist zum Beispiel für die Übertragung von Sprache und Daten konzipiert. Dies garantiert unter anderem verbindungsorientierte Kommunikationspfade mit konstanter, kontrollierter Übertragungsverzögerung. Die Anforderungen für anspruchsvolle Multimedia-Anwendungen übersteigen allerdings deutlich die Kapazitäten einer einzelnen ISDN-Verbindung. Hinzu kommt, daß die Architektur der heutigen Weitverkehrsleitungen auf Dienste mit konstanten Bitraten ausgelegt ist. Bei Videoappliaktionen treten aber ständig variierende Datenraten auf. Bei der Verwendung herkömmlicher Techniken kann man sich nur durch die Bereitstellung von Leitungen mit festen Bandbreiten, die auch im Spitzenlastbereich ausreichend Kapazität besitzen, behelfen. Das Prinzip des asynchronen Zeitmultiplexing versetzt ATM in die Lage, auch stark schwankende Verkehrsprofile effizient übertragen zu können.

4.1.5. ATM-Standardisierung

Damit eine Technologie unter Komponenten-Herstellern und Nutzern breite Akzeptanz findet, bedarf es der Definition von Standards, wenn möglich weltweiter Standards. Die Standardisierung von ATM wird durch zwei Institutionen geregelt:
die ITU (International Telecommunications Union, vorher CCITT) und das ATM-Forum.

ITU (International Telecommunications Union)

Die International Telecommunication Union (ITU) wurde im Jahr 1932 unter dem Namen CCITT (International Telegraph and Telephone Consultative Commitee) gegründet. 1993 wurde das Gremium in ITU umbenannt. Es besteht mittlerweile aus mehr als 160 Mitgliedsländern (siehe http://www7.itu.int/GlobalDirectory/). Die Mitgliedschaft beschränkt sich auf die jeweiligen nationalen Organisationen für den Daten- und Telekommunikationsbereich.

Die I-Serie, der Standard für ISDN (Integrated Services Digital Network), wurde 1984 von der ITU geschaffen. Entwürfe für die Breitbanddienste B-ISDN wurden dieser Serie 1990 hinzugefügt. Die Beschlußfassung erfolgte 1991. Die ITU-Standards können auf der Internetseite der ITU (http://www.itu.int) angesehen und in elektronischer Form (z.B. im PDF-Format) übertragen werden.

Das ATM-Forum

Das ATM-Forum wurde im September 1991 gegründet. Gründer waren die vier Firmen CISCO-Systems, NET/Adaptive, Northern Telekom und US-Sprint. Seitdem ist die Anzahl der stimmberechtigten Mitglieder auf über siebenhundert gestiegen. Ziel des ATM-Forums ist es, die Standardisierung von ATM nicht mehr alleine der ITU zu überlassen, eigene Ideen mit einzubringen und die Zeitdauer vom Vorschlag bis zur Standardisierung einer Idee zu reduzieren. Durch enge Zusammenarbeit mit den offiziellen Standardisierungsgremien sollen Vorschläge der Industrie in die Gestaltung der Standards einfließen. Des weiteren sollen für die Bereiche, in denen noch keine Standards definiert sind, Industriestandards, mit Hilfe von Spezifikationen des ATM-Forums, geschaffen werden. Diese Quasistandards helfen den Unternehmen die Produkte schneller zu entwickeln, was durchaus Sinn macht, da die Prozeduren von der Idee bis zur Verabschiedung des Standards in den öffentlichen Entscheidungsgremien meist recht langwierig sind.

4.1.6 PNNI (Private Network-to-Network-Interface)

Eine wichtige Spezifikation des ATM-Forums ist PNNI (Private Network-to-Network-Interface). Innerhalb eines ATM-Netzwerkes regelt PNNI, mittels dynamischen Routings, die Kommunikation zwischen den ATM-Switching-Systemen. Hierbei werden zwei Protokolle eingesetzt: ein Topologie-Protokoll und ein Signalisierungsprotokoll. Das Topologieprotokoll hat die Aufgabe Informationen über die Netzwerktopologie an die einzelnen Netzwerkteilnehmer weiterzugeben. Das Signalisierungsprotokoll dient zum Verbindungsaufbau zwischen den ATM-Schalteinheiten. Eine detaillierte Beschreibung von PNNI erfolgt in Kapitel 5.3.1. .

4.2 LAN-Emulation

LAN-Emulation (LANE) ist eine Spezifikation des ATM-Forums die dem Zwecke dient, ATM-Netzwerke effizient in bestehende herkömmliche LAN-Strukturen zu integrieren. Die LAN-Emulations-Komponenten führen eine vollständige Emulation der LAN-MAC-Schicht durch. Die Emulation ermöglicht weiterhin den Einsatz bisheriger Anwendungen, ohne daß diese angepasst werden müssen. Die erste

Spezifikation für eine LAN-MAC-Schicht-Emulation auf ATM-Netzwerken wurde vom ATM-Forum im Januar 1995 unter dem Titel "LAN Emulation over ATM Version 1.0" fertiggestellt ([15]).

Die unterschiedliche Funktionsweise von ATM im Vergleich zu Ethernet-, Token-Ring- oder FDDI-Netzwerken macht es schwer ATM-Netze an herkömmliche LAN-Topologien anzukoppeln. Die unterschiedlichen Übertragungsgeschwindigkeiten oder inkompatiblen Paket- bzw. Zellenformate sind allerdings nicht das zentrale Problem. Solche Differenzen kann man durch Brücken relativ problemlos lösen. Dies geschieht bereits heute zwischen bestehenden Ethernet-, TokenRing und FDDI-LANs relativ problemlos. Der wesentliche Unterschied zwischen ATM und bestehenden LAN-Technologien besteht darin, daß ATM-Netzwerke auf verbindungsorientierten Kommunikationsbeziehungen basieren. Alle traditionellen LAN-MAC-Protokolle arbeiten dagegen nicht verbindungsorientiert. Die Datenpakete werden von letzteren über das gemeinsame Übertragungsmedium in der Hoffnung versendet, daß sie die Zielstation auch unversehrt erreichen. Auf Empfangsbestätigungen durch den Empfänger wird verzichtet. Gehen Datenpakete verloren, so müssen die Daten auf höherer Protokollebene erneut angefordert werden. Alle Stationen kommunizieren über ein und dasselbe Medium. Jedes Datenpaket erreicht auch jede andere angeschlossene Netzwerkstation desselben Empfänger-Netzsegmentes. Dem einzelnen Netzwerkteilnehmer wird, lediglich durch die Filterung nach entsprechenden Zieladressen durch die Netzwerk-Interface-Karte (NIC), der Eindruck einer exklusiven Kommunikationsbeziehung vermittelt. Wird als Zieladresse eine Broadcastadresse angegeben, so wird das betreffende Datenpaket von allen Stationen auch ausgewertet. Mit Hilfe der Broadcast-Technik können somit alle Stationen eines LAN-Segementes durch die Versendung eines einzigen Paketes angesprochen werden. Eine ganze Reihe von Funktionen werden von LAN-Protokollen unter Zuhilfenahme dieser Broadcastfunktionalität realisiert. Das auf Broadcasts basierende ARP-Protokoll dient zum Beispiel der Zuordnung von MAC-Adressen zu vorgegebenen IP-Adressen.

Diese Broadcastmechanismen stehen allerdings im Gegensatz zum Funktionsprinzip von ATM. Integrations-Lösungen lassen sich nur mit entsprechend hohem Aufwand realisieren. ATM arbeitet verbindungsorientiert und stellt für jedes Kommunkationspaar einen logischen Verbindungspfad zur Verfügung. Die für die Verbindung minimale zur Verfügung stehende Bandbreite, bzw. die maximale Übertragungsverzögerung wird während der Verbindungsaufbauphase vereinbart. Die Datenpakete werden zur Zielstation geleitet und von den anderen Stationen nicht wahrgenommen. Eine Broadcastsendung an n Stationen muß daher durch den Aufbau von n Verbindungspfaden zu den Empfängerstationen realisiert werden.

Die LAN-Emulation des ATM-Forums wird durch vier Dienstmodule realisiert:

- LE-Client (LEC),
- LE-Server (LES),
- LE-Emulation-Configuration-Server (LECS),
- Broadcast und Unknown-Server (BUS).

LEC (LAN Emulation Client)

Für die Durchführung aller notwendigen Steuerfunktionen sowie die eigentliche Datenübertragung über das ATM-Interface ist die LAN-Emulation-Client-Software zuständig. Den Anwendungen wird eine Standard-LAN-MAC-Schnittstelle zur Verfügung gestellt. LE-Clients bestehen aus den Komponenten

- Systemhardware (PC, Workstation, Brücke, etc.),
- Standard-LAN-Software (TCP/IP, Brücken/Routersoftware, IPX/SPX, etc.),
- LE-Client-Software und dem
- ATM-Interface.

LES (LAN Emulation Server)

Die Steuerung des emulierten LANs wird durch den LAN-Emulation Server durchgeführt. Er übernimmt die Registrierung von LE-Clients und die Zuordnung von MAC- zu ATM-Adressen aller registrierten Stationen. Jeder an einem emulierten LAN teilnehmende LE-Client muß seine LAN-MAC-Adresse, sowie die korrespondierende ATM-Adresse beim LAN-Emulation Server anmelden. Auch Routeninformationen werden an den LES weitergegeben. Beim Versenden eines LAN-Datenpakets wird zunächst in der Adresstabelle des LES nach der ATM-Adresse der Zielstation gesucht. Ist diese dort nicht vorhanden, so muß die Adressenzuordnung mit Hilfe eines Broadcasts vom BUS-Server durchgeführt werden.

LAN-Emulation-Configuration-Server (LECS)

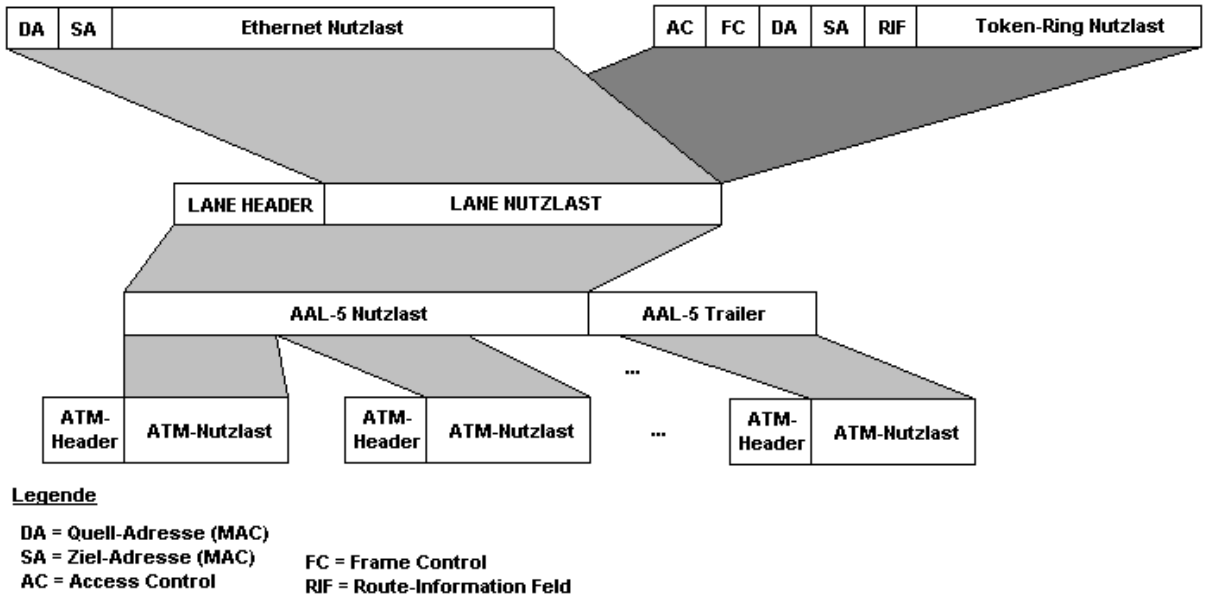
Der LAN-Emulation-Configuration-Server verwaltet die Zugehörigkeit von LECs zu unterschiedlichen emulierten LANs. Die Konfigurationsinformationen der definierten LAN-Emulationen stehen in einer Konfigurationsdatenbank. LE-Clients können über den LE-Configuration-Server auf diese Datenbank zugreifen und somit den entsprechenden LAN-Emulationen (auch Mitgliedschaft in mehreren parallel möglich) zugeordnet werden. Jedes Endgerät im ATM-Netz muß die ATM-Adresse des LECS kennen. Daher wird die Adresse beim Konfigurieren eines Endgeräts vom Administrator eingetragen. Ein gravierender Nachteil des LECS ist, daß er einen single point of failure darstellt, deshalb existiert auch mindestens ein Backup-LECS auf den die LE-Clients zugreifen können falls der Primary-LECS ausfällt.

Broadcast und Unknown Server (BUS)

Das Broadcast- und Unknown-Server-Modul dient zur Vermittlung aller Broadcast- und Multicast-Datenpakete von LE-Clients. Dazu gehören

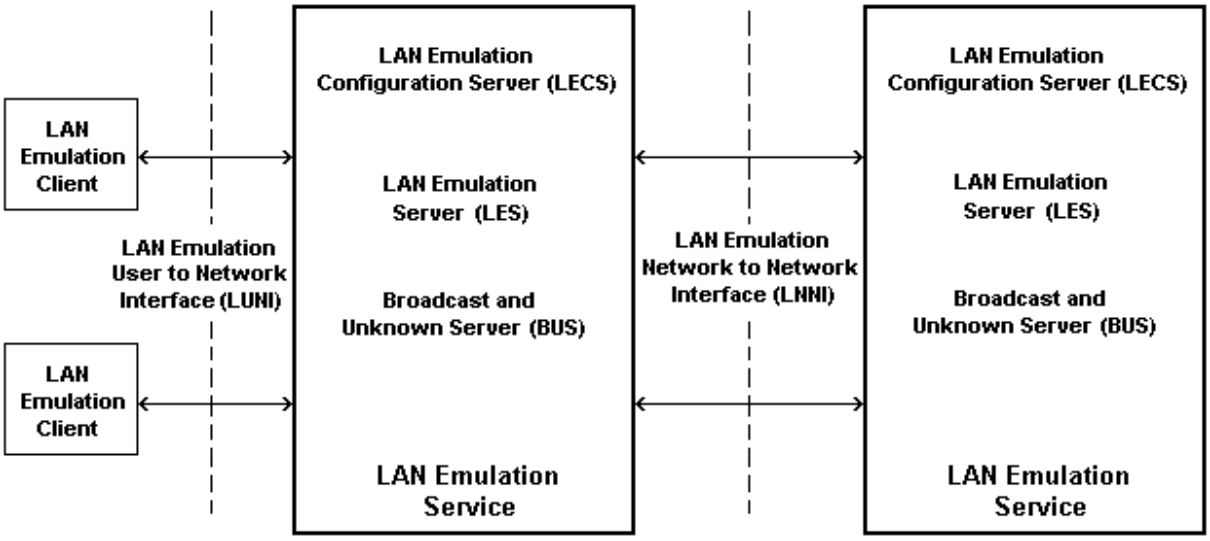
- alle Datenpakete mit Broadcast- bzw. Multicastadressen,
- Datenpakete mit einer MAC-Adresse, für die dem LE-Client die korrespondierende ATM-Adresse nicht bekannt ist, oder die vom LES nicht aufgelöst werden konnte,
- Explorer-Datenpakete, die zur Ermittlung der optimalen Routen dienen.

Die folgende Abbildung zeigt die Verkapselung der Frames der LAN-MAC-Schicht, welche in LANE-Rahmen verkapselt werden. Die ATM-Anpassungsschicht segmentiert anschließend die LANE-Rahmen in ATM-Zellen.



LANE-Kapselung ([3], S.249)

Wie schon beschrieben, funktioniert der LAN-Emulationsdienst nach dem Client/Server-Prinzip. Der LANE-Client (LEC) greift hierbei auf die Dienste des LANE-Servers (LES), des LANE-Konfigurations-Servers (LECS) und des Broadcast-and-Unknown-Servers zurück. Die Client/Server-Kommunikation wird vom ATM-Forum mit der Spezifikation L-UNI (LANE - User Network Interface) beschrieben. L-UNI regelt die Abläufe an der Schnittstelle zwischen Client und Server. Die Server/Server-Kommunikation ist in der Spezifikation L-NNI (LANE - Network Network Interface) festgelegt. Durch Einsatz von L-NNI ist es möglich, mehrere LEC-, LES- und BUS-Einheiten zu verbinden.



LUNI und LNNI ([15], S.21)

Soll ein neuer LAN-Emulation-Client einem emulierten LAN hinzugefügt werden, so baut der LEC zunächst eine Verbindung zum LE-Configuration-Server auf. Der LE-Client wird im emulierten LAN registriert und kann außerdem, unter Benutzung des LE-Konfigurationsprotokolles, verschiedene Parameter (Name des emulierten LANs,

maximale Rahmengröße, etc.) abfragen. Anschließend baut der LEC die Verbindung zum LAN-Emulation-Server auf. Der LES teilt dem LAN-Emulations-Client eine eindeutige LAN-Emulations-Client-ID (LECID) zu. Nun ist der LAN-Emulation-Client im Besitz aller für die Teilnahme am LAN-Emulationsdienst notwendigen Informationen. Abschließend wird ein Kommunikationspfad zum Broadcast-Unknown-Server eröffnet. Der LAN-Emulation-Client ist danach betriebsbereit für den LAN-Emulationsdienst.

LAN-Emulation unterstützt in der ersten Version noch keine Quality of Service-Zuordnung. Die Unterstützung ist ab Version zwei fester Bestandteil der Spezifikation. Eine detaillierte Beschreibung des LAN-Emulations-Dienstes in der Praxis kann in Kapitel 6.2.2. und 6.2.3. nachgelesen werden.

4.3. Virtual LAN-Technologie

Mit zunehmender Größe und Komplexität der Netzwerke wächst auch das Bedürfnis nach Strukturiertheit und Übersichtlichkeit. Hier hilft das VLAN-Konzept weiter. Der Begriff VLAN steht für Virtual Local Area Network.
Durch den Einsatz von VLANs ist es möglich, die vorhandene physikalische Struktur logisch zu trennen. Man möchte Einheiten, die logisch zusammengehören, in eigene Teilnetze gliedern, um den netzbelastenden Broadcastverkehr zu minimieren. In diesem Umfeld spielen auch Sicherheitskriterien eine wichtige Rolle. So soll der Netzverkehr aus sicherheitsrelevanten Bereichen nicht von anderen mitgehört werden
können. In großen Firmen ist die Zugehörigkeit eines Arbeitsplatzes zu einem logischen Teilbereich nicht unbedingt durch die örtliche Struktur gegeben. Das bedeutet, daß zum Beispiel die Mitarbeiter eines Forschungsprojekts über den Campus verteilt sind. Dies ist auch bei der HypoVereinsbank der Fall. Bei Änderungen von Teilnetzen sollte der anstehende Konfigurationsaufwand minimal sein. Bei der Bildung neuer Arbeitsgruppen sollten die Administratoren in der Lage sein, die Teilnehmer idealerweise per Knopfdruck einem neuen VLAN zuzuordnen. Es gibt sehr unterschiedliche Ausbaustufen von VLAN's, die jeweils eigene Vor- und Nachteile haben. Im folgenden Teil werden die verschiedenen Typen von VLAN's vorgestellt.

Der Begriff VLAN ist sehr weit auslegbar. VLAN ist nicht gleich VLAN. Auch gibt es keine Lösung, die alle Anforderungen erfüllt. Im folgenden Teil werden die einzelnen Ausbaustufen von VLAN's vorgestellt.

Layer-1-VLAN's: Die Zugehörigkeit eines Endgeräts oder Segments von Endgeräten zu einem VLAN wird durch den Port des Switches festgelegt, an den es angeschlossen ist. Vorteile sind neben der problemlosen Integration die geringen Kosten und die Unterstützung aller Layer-3 Protokolle. VLAN's auf Schicht 1 bringen allerdings große Sicherheitslücken mit sich. Zum Beispiel wird der Sender von Daten nicht untersucht. Hinzu kommt, daß das System völlig unflexibel bei Umzügen von Mitarbeitern ist, und statistisch gesehen zieht jeder Mitarbeiter einer Firma mindestens einmal im Jahr um. Der Administrator muß in diesen Fällen jedesmal die Portdefinitionen ändern.

Layer-2-VLAN's: Die Zugehörigkeit eines Endgeräts zu einem VLAN wird durch die

MAC-Adresse seiner Netzwerkkarte definiert. Es werden wiederum alle Layer-3 Protokolle unterstützt. Weiterhin fällt der administrative Aufwand bei Umzügen weg, da die virtuellen Segmente eben anhand der MAC-Adresse gebildet werden. Der Administrator muß lediglich dann neu konfigurieren, wenn neue Netzwerkkarten dem Netz hinzugefügt, bzw. alte Karten ausgetauscht werden. Ein weiterer Vorteil ist die Tatsache daß, aufgrund der hardwarenahen Zuordnung, sichere Strukturen aufgebaut werden können.

Layer-3-VLAN's: Die Zugehörigkeit eines Endgeräts zu einem VLAN wird durch die zugewiesene Netzwerkadresse (z.B. IP-Adresse) festgelegt. Somit ist das System wiederum unabhängig von der eingesetzten Hardware. Bei Umzügen oder sonstigen Änderungen konfiguriert sich das System weitgehend automatisch um. Konfigurationsaufwand entsteht erst, wenn Endgeräte einem anderen VLAN zugewiesen werden. Als Nachteil ist zu erwähnen, daß das System protokollabhängig ist: Es lassen sich zum Beispiel Netzwerkadressen per Software, also vom Nutzer selbst, ändern. Weiterhin müssen für Multiprotokollnetze (z.B. Einsatz von IP, IPX, etc.) verschiedene Systeme parallel eingesetzt werden.

5. Beschreibung traditioneller und neuer Routing-Mechanismen

Die Zahl der IP-Netze ist in den letzten Jahren, sowohl im Lokal- wie auch im Weitverkehrsbereich, sprunghaft angestiegen. Dieser Trend wird durch die steigende Begeisterung der Nutzer für das Internet begünstigt. In den letzten zehn Jahren hat sich die Zahl der Endrechner fast jährlich verdoppelt. Die Zahl der Web-Server stieg zweimal so schnell. Zur explosionsartigen Vermehrung der Hosts und der Nutzer im Internet kommt hinzu, daß immer mehr Dienste ins Internet verlagert werden, die vorher über andere Wege Anbieter und Nutzer fanden. Beispiele hierfür sind Internet-Telefonie, Audio- und Videoübertragung. Die hohe Kostenersparnis gegenüber traditionellen Lösungen ist hierbei der Motor für Motivation und Forschung. Der Bedarf an multimedialen Anwendungen, und somit gleichzeitig der Bedarf an Bandbreite und Echtzeitverhalten, wird sich in Zukunft noch deutlich erhöhen. Hierbei stossen allerdings die herkömmlichen IP-Router an ihre Leistungsgrenzen. Die Obergrenze der vermittelbaren Pakete liegt in der Sekunde bei ungefähr 1 Million, was zur Folge hat, daß Router bei wachsenden Datenströmen zum Flaschenhals werden.

An dieser Stelle waren die Herstellerfirmen der Komponenten, wie auch die Standardisierungsgremien gefordert Verbesserungen, bzw. alternative Lösungen zu erarbeiten. Als Ergebnis gab es mehrere Ansätze mit dem gleichen Ziel. Die bisher eingesetzte langsame software-gesteuerte Vermittlung der Daten (Forwarding), sollte in Switches integriert werden, welche für einen Großteil der Daten einen möglichst schnellen Weg zum Ziel "schalten" sollten. Diese Verfahren werden allgemein als "Shortcut-Routing" bezeichnet und im Kapitel 5.4. behandelt. In Kapitel 5.2. wird auch darauf eingegangen, was unter den meist verwirrenden Marketingbegriffen wie eben "Layer 3- Switching" tatsächlich zu verstehen ist. Ein Router arbeitet nämlich auf Ebene 3 des ISO/OSI-Referenzmodells, ein sogenannter "Switch" auf Ebene 2 (MAC-Ebene).

Router sind dafür konzipiert, verbindungslose Netzwerkdienste zur Verfügung zu stellen. Sie speichern die Frames in ihren Eingangspuffern, setzen sie zu IP-Datagrammen zusammen und stellen aufgrund der IP-Zieladresse fest, über welchen Router-Port das gewünschte Endgerät erreicht werden kann. IP-Adresse und Port-Nummer sind in Routingtabellen gespeichert, welche durch den Einsatz von Routingprotokollen erstellt werden (siehe "Traditionelles Routing"). Nachdem der Pfad bestimmt wurde, werden die IP-Datagramme wieder in MAC-Frames zerlegt, mit der neuen MAC-Zieladresse versehen und über den betreffenden Ausgangsport in Richtung Zielgerät gesendet. Dieser Vorgang wiederholt sich bei allen Transit-Routern, bis schließlich das Netz erreicht ist, in dem sich das Zielgerät befindet.
Jeder Router auf der Strecke setzt also die Mac-Frames zusammen, untersucht die IP-Adresse jedes Paketes, entscheidet über die Weiterleitung der Daten, zerlegt die Pakete wieder in Mac-Frames und sendet diese dann über den entsprechenden Ausgangsport. Das dieser Vorgang einiges an Zeit beansprucht liegt auf der Hand.

Man kann sich vorstellen, daß dieser Prozeß eine Menge Zeit in Anspruch nimmt. Hinzu kommt der Einsatz von Routingprotokollen (wie OSPF oder RIP) und die Durchführung von Protokollkonvertierungen (z.B. von Ethernet zu X.25 und zurück). Die Rechenleistung der Router stößt ab einer gewissen Bealstung an Grenzen, was

zu Verzögerungszeiten in der Größenordnung von 200 Millisekunden pro Router führen kann. Je mehr Router also auf dem Weg zur Zieldomäne liegen, um so höher ist die Verzögerung. Um die erwähnten Protokollkonvertierungen durchführen zu können benötigt man Multiprotokollrouter mit erheblicher Intelligenz, also hoher Rechenleistung und entsprechender Software. Diese Ausstattung macht den einzelnen Router entsprechend teuer.

ATM soll an dieser Stelle mit, speziell für IP-Übertragung entwickelten, Verfahren (siehe "Shortcut Routing") weiterhelfen. Diese Mechanismen besitzen einige Gemeinsamkeiten. Einem Datenstrom wird eine Marke (Label oder Tag) zugewiesen, auf die basierend ein Shortcut-Pfad im ATM-Netz eingerichtet wird. Die Weiterleitung der IP-Datenpakete wird beschleunigt, da sie zunächst in die kleinen ATM-Zellen fester Länge unterteilt und diese Zellen dann per Hardware-Entscheidung weitergeleitet werden. Dies ist möglich, weil der ATM-Switch nur den kleinen Adreßkopf der Zelle analysieren muß anstatt des komplexen IP-Datenpakets mit variabler Länge.

5.1. Traditionelles Routing

Das Internet (bzw. ein sogenanntes Internetwork allgemein) besteht aus selbständigen Netzen, welche zu einem großen Netzverbund zusammengeschlossen sind. Rechner, die sich in einem Teilnetz befinden, können immer nur mit den Rechnern des gleichen Teilnetzes direkt kommunizieren. Router sind für die Weiterleitung der Daten zu anderen Teilnetzen zuständig. Der Sender erfährt hierbei nicht welchen Weg die Datenpakete, bis zum Erreichen des Empfängers, nehmen. Damit die Daten korrekt weitergeleitet werden können, muß der Router Informationen über das Netz besitzen, sogenannte Topologieinformationen. Hierzu gehört vor allem die Information über das Vorhandensein und die Erreichbarkeit von Teilnetzen. Um diese wichtigen Informationen sammeln und verarbeiten zu können werden Routing-Protokolle wie OSPF,RIP oder BGP eingesetzt. Wichtig ist hier den Unterschied, zwischen Routing-Protokollen (OSPF, RIP) und "gerouteten" Protokollen wie IP, zu kennen. Die Router sammeln Topologieinformationen mit Hilfe eines oder mehrerer Routing-Protokolle um dann, aufbauend auf diesen Informationen, eine Entscheidung für das Weiterleiten von "routbaren" Protokollen wie IP zu fällen. Die derzeit am meisten eingesetzten Routingprotokolle sind OSPF, RIP und BGP. Ihre Funktion und möglichen Einsatzbereiche werden im folgenden besprochen.

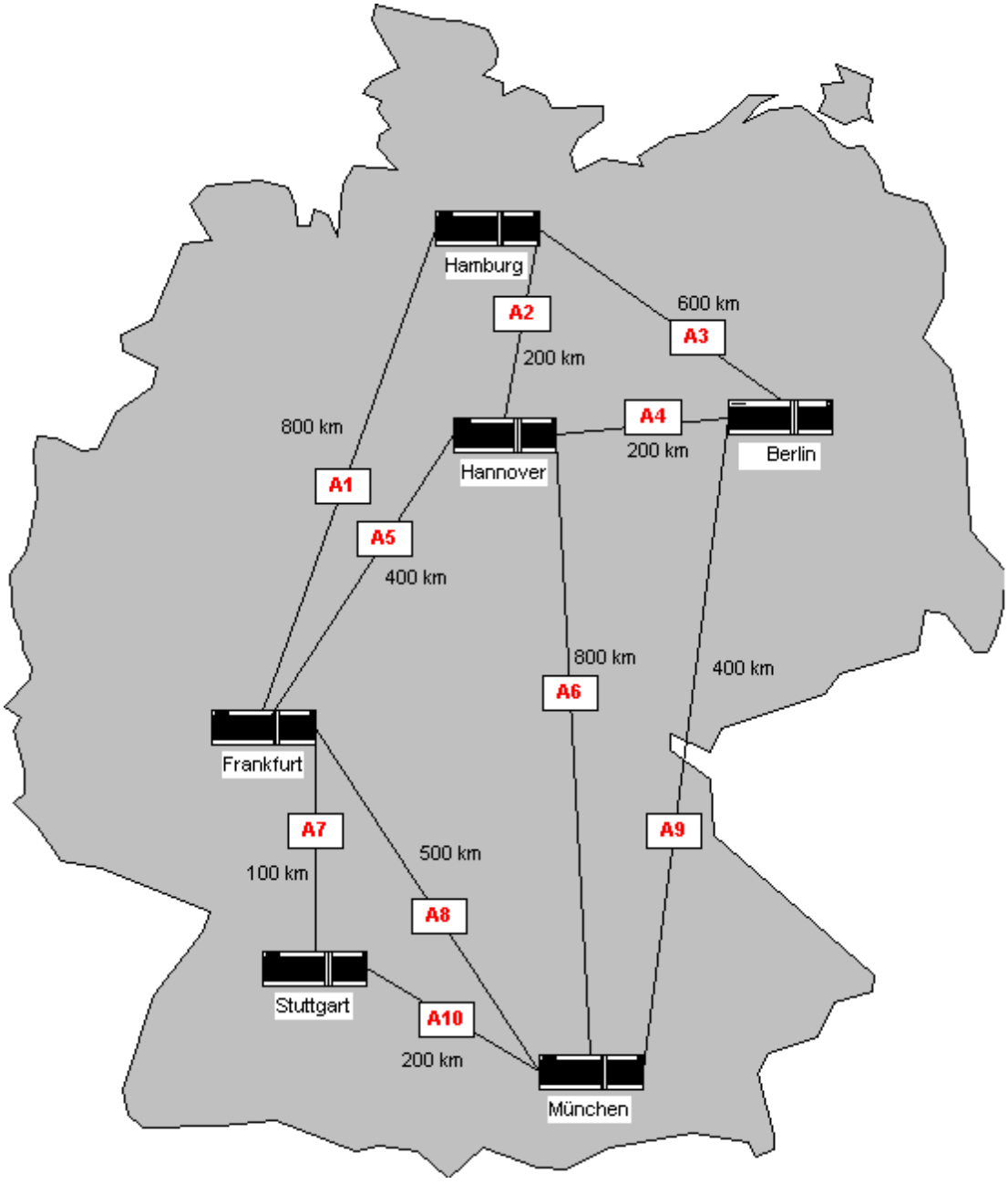
5.1.1. Interne Gateway Protokolle

5.1.1.1. Link State Routing am Beispiel von OSPF

OSPF (Open Shortest Path First) ist ein, in IP-Netzen, häufig eingesetztes Routing Protokoll, welches zur Klasse der Link-State-Routing Protokolle gehört. Die OSPF-Spezifikation kann in RFC 1583 nachgelesen werden ([31]). Bei Einsatz von OSPF haben alle Router eine komplette Sicht des Netzes.

Routingprotokolle haben die Aufgabe Router mit Topologieinformationen zu versorgen, so daß Routingentscheidungen getroffen werden können. Das Prinzip bei Link-State Protokollen ist, den Routern einen gesamten Überblick über das Netzwerk

zu verschaffen. Den Ablauf zur Erstellung dieses Gesamtüberblicks kann man in drei Schritte aufteilen. Jeder Router sammelt zunächst Informationen über seinen unmittelbaren Nachbarn. Diese Nachbarschaftsbeziehungen werden im nächsten Schritt allen anderen Routern bekanntgegeben. Zuletzt analysiert nun jeder einzelne Router die erhaltenen Informationen und berechnet sich daraus die Sicht auf das gesamte Netzwerk und die Pfade, die er für die Routingentscheidungen heranzieht. Als Beispiel zur allgemeinen Beschreibung der Funktionalität von Link-State-Routingprotokollen dient das Deutschlandbild mit Routerverbindungen zwischen verschiedenen Städten (siehe Abbildung).



OSPF-Beispiel: Deutschland-Router

Die Strecken sind jeweils zweifach gekennzeichnet. Die eine Marke (A1-A10) kennzeichnet das Teilnetz (z.B. IP-Subnetz) zwischen zwei Routern, und die andere Marke beschreibt die "Kosten", die für das Passieren dieses Teilnetzes festgelegt wurden. Die "Kosten" repräsentieren in diesem Beispiel die Entfernungen.

Die erste Aktion die nun von den einzelnen Routern durchgeführt wird, ist das Kennenlernen der direkten Nachbarn. Dies geschieht durch das Versenden von "hello packets" auf allen angrenzenden Teilnetzen des jeweiligen Routers. Der Router in München sendet also seine "hello packets" in die Teilnetze A10, A8, A6 und A9. Nachdem alle Pakete versendet und empfangen wurden, kennt jeder Router seine direkten Nachbarn und die jeweiligen "Kosten", bzw. Entfernungen. Jeder Router hält diese Daten in seiner Nachbar-Liste (Neighbour-List). Zum Beispiel besitzt der Router in München nach der ersten Aktion folgende Nachbar-Liste:

Nachbar	Teilnetz	Kosten
Stuttgart	A10	200 km
Frankfurt	A8	500 km
Hannover	A6	800 km
Berlin	A9	400 km

Es handelt sich hierbei, wie erwähnt, um ein Beispiel, um die Funktionalität zu erklären. Tatsächlich führen Router in diesen Tabellen beispielsweise auch Informationen über die eigenen Schnittstellen, über die der Router in die Teilnetze eingebunden ist (z.B. Interface "S0" als serielle Schnittstelle oder "Eth0" als Ethernet-Schnittstelle, etc.). Auf diese Detailinformationen wurde, zur Unterstützung der Übersichtlichkeit, verzichtet.

Jeder Router kennt nun seine Nachbarsysteme. Im nächsten Schritt werden die Nachbarbeziehungen im ganzen Netz bekannt gemacht. Dies geschieht, indem jeder Router seine Nachbar-Liste verschickt. Die somit verschickten Nachrichten werden auch LSA (Link State Advertisement) genannt. In dem Beispiel entspricht das LSA des Münchener Routers der obenstehenden Nachbar-Liste. Die LSA-Pakete werden im ganzen Netz durch "Flooding" verbreitet. Grundsätzlich versteht man unter "Flooding", daß jedes vom Router empfangene Paket auf allen Interfaces (abgesehen von der Eingangsschnittstelle) weitergesendet wird. Handelt es sich also um eine neue, noch nicht erhaltene, LSA-Nachricht so sendet der die Nachricht auf allen Verbindungen weiter. Ist das LSA dem Router bereits bekannt, wird es ohne weitere Behandlung verworfen. Im folgenden kurzen Beispiel soll das Flooding anhand der Verbreitung der LSA-Nachrichten des Münchener Routers dargestellt werden. Es gibt neben dem beschriebenen Ablauf noch mehrere Möglichkeiten, wie sich die Pakete im Netz verbreiten könnten. Dies ist abhängig von der Verarbeitungsgeschwindigkeit der einzelnen Router.

Schritt 1: Der Münchener Router schickt seine LSA-Pakete nach Stuttgart, Frankfurt, Hannover und Berlin.

Schritt 2: Stuttgart schickt das LSA weiter an Frankfurt. Frankfurt verwirft die Nachricht. Hannover schickt die Nachricht weiter an Berlin und Hamburg. Die Nachricht wird in Berlin verworfen. Frankfurt schickt die Nachricht nach Hannover

und Hamburg. Beide Nachrichten werden verworfen. Berlin schickt die Nachricht nach Hamburg. Dort wird sie verworfen.

Bereits nach zwei Schritten kennt jeder Router im Netz die Nachbarbeziehungen des Münchener Routers. Analog werden die LSA-Pakete der anderen Router verbreitet. Nachdem das Flooding beendet ist, kennen alle System die entfernten Nachbarschaftsbeziehungen genau so gut wie die lokalen. Jeder Router besitzt jetzt folgende Informationen (link-state Datenbank):

Router	Verbindung 1	Verbindung 2	Verbindung 3	Verbindung 4
München	Stuttgart, A10, 200km	Frankfurt, A8, 500km	Hannover, A6, 800km	Berlin, A9, 400km
Stuttgart	München, A10, 200km	Frankfurt, A7, 100km	-	-
Frankfurt	Stuttgart, A7, 100km	München, A8, 500km	Hamburg, A1, 800km	Hannover, A5, 400km
Hannover	Frankfurt, A5, 400km	Hamburg, A2, 200km	Berlin, A4, 200km	München, A6, 800km
Hamburg	Frankfurt, A1, 800km	Hannover, A2, 200km	Berlin, A3, 600km	-
Berlin	Hamburg, A3, 600km	Hannover, A4, 200km	München, A9, 400km	-

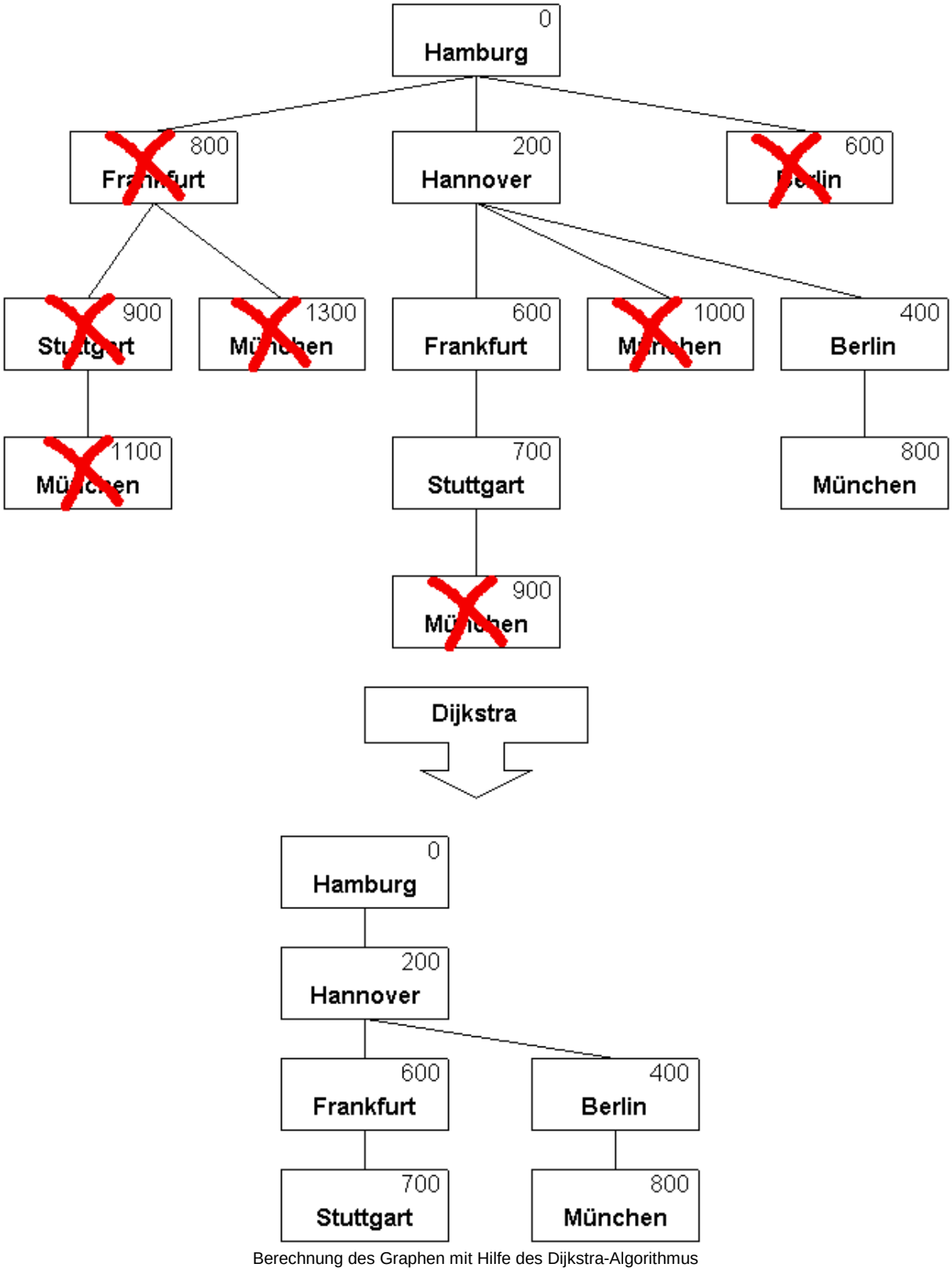
Schritt 3: Nun werden im letzten Schritt die Routen berechnet. Hierfür setzen die meisten Implementationen den "Shortest Path First"-Algorithmus von Dijkstra ein. Mit Hilfe dieses Algorithmus wird ein gerichteter Netzwerk-Graph erstellt, in dem zu jedem Netzwerkknoten genau ein Pfad existiert. Die Kriterien für die Festlegung des Pfades basieren auf den Kosten. Der Router, der die Berechnung durchführt, setzt sich jeweils selbst als Wurzel des Baumes ein. Jeder Router führt die Auswertung selbständig für sich und in drei Schritten aus:

1. Zunächst wird die Wurzel des Baumes eingesetzt, dies ist der Router selbst. Die Kosten betragen für diesen Router null.

2. Der Router durchsucht die link-state Datenbank um die Nachbarn der Router herauszufinden, die zuletzt in den Baum eingefügt wurden (im ersten Schritt der berechnende Router selbst). Für jeden Nachbarn werden die Kosten ausgewertet (Entfernung). Falls die jeweilige Entfernung kleiner als jede andere Route zu diesem Nachbarn ist, wird die Route temporär dem Baum hinzugefügt.

3. Nun muß im Baum nachgesehen werden, ob es eingetragene Routen gibt die, aus Kostensicht, ungünstiger sind, als die temporär eingebundenen Routen. Gibt es so eine Route, so wird sie entfernt. Anschließend wird, vorausgesetzt es gibt noch temporäre Einträge, Schritt 2 wiederholt. Gibt es keine temporären Einträge mehr so ist die Berechnung beendet.

Die folgende Abbildung und die zugehörige Beschreibung zeigen, wie der Router in Hamburg mit Hilfe der genannten drei Schritte zu seinen Topologieinformationen und dem zugehörigen Graphen gelangt.



Berechnung des Graphen mit Hilfe des Dijkstra-Algorithmus

Schritt 1: Der Router in Hamburg setzt sich selbst als Wurzel und seine drei Nachbarn (Frankfurt, Hannover und Berlin) und deren Kosten in den Graphen ein.

Schritt 2: Die Nachbarn von Frankfurt (Stuttgart und München) werden in den Baum

eingefügt. Die Verbindung nach Hannover wird nicht eingefügt, da bereits eine Strecke nach Hannover mit geringeren Kosten existiert.

Schritt 3: Der Nachbar von Stuttgart (München) wird eingefügt. Das System überprüft nach dem Einfügen, ob es andere Strecken zu den eingefügten Nachbarn gibt, die höhere Kosten erzeugen. In diesem Fall wird München als Nachbar von Frankfurt (1300 km von Hamburg aus) aus dem Baum entfernt.

Schritt 4: Die Nachbarn von Hannover (Frankfurt, München und Berlin) werden hinzugefügt. Da die Routen von Hamburg nach Frankfurt und von Hamburg nach Berlin, jeweils über Hannover, günstiger als die direkten Verbindungen sind, werden die letzteren (und ihre Unterbäume) gestrichen.

Schritt 5: Stuttgart wird als Nachbar von Frankfurt eingehängt. München wird nicht aufgenommen, da es bereits eine günstigere Verbindung über Hannover gibt.

Schritt 6: München wird als Nachbar von Stuttgart eingehängt. Diese Verbindung ist günstiger als die über Hannover. Somit wird München als Nachbar von Hannover gestrichen.

Schritt 7: München wird als Nachbar von Berlin eingetragen. Diese Verbindung ist mit 800km die günstigere als die über Stuttgart. Somit wird München als Nachbar von Stuttgart gestrichen. Der Graph des Hamburger Routers ist nun vollständig.

Die Berechnung der Routen führt jeder Router eigenständig durch. Nach der Auswertung weiss er dann zu jedem Ziel den günstigsten Pfad. Alle Router haben nun, vorausgesetzt es finden keine Veränderungen mehr statt, die komplette Sicht über das Netz. Es treten allerdings durchaus ständige Änderungen statt. Router werden neu gestartet oder fallen aus. Auch die Verbindungsstrecken werden teilweise unterbrochen. Für diese Fälle gibt es Mechanismen geben, die die Kommunikation aufrechterhalten. Die beschriebenen "hello packets" werden nicht einmalig sondern laufend verschickt. Empfängt ein Router keine "hello packets" mehr von einem seiner Nachbarn, so entfernt er den ehemaligen Nachbarn aus der "Neighbour-List". Ähnlich geschieht dieser Vorgang bei neuen Nachbarn. Diese werden in die Liste eingefügt. Die Nachbar-Listen (LSA's) werden laufend durch das "Flooding" im Netz verteilt, so daß Änderungen im Netz sehr schnell erkannt und, mit Hilfe der Dijkstra-Berechnung, übernommen werden. Nachteil dieses Verfahrens ist die entsprechend hohe Netzlast, die OSPF mittels dieses Verfahrens erzeugt. Sie steigt exponentiell zu der Anzahl der hinzukommenden Router.

Die Erzeugung dieser hohen Netzlast schließt aus, daß OSPF in so großen Netzen wie dem Internet eingesetzt wird. Die Dijkstra-Berechnung würde, unter Berücksichtigung aller Nachbarlisten der im Internet verbundenen Router, selbst auf einem Großrechner mehrere Stunden dauern. Und danach wären die Routen schon wieder veraltet. Um auch in größeren Netzen Einsatz zu finden, baut OSPF auf verschiedenen Hierarchiestufen auf (ähnlich wie bei der Beschreibung von PNNI).
Um einen Einblick in die Hierarchieordnung von OSPF zu geben , werden die zwei Aufteilungen "Autonomes System (AS)" und "Area" im folgenden beschrieben.

Das Internet selbst besteht aus zahlreichen autonomen Systemen. Unter einem

autonomen System versteht man ein geschlossenes System, welches einer "administrativen Einheit" untersteht, einem System, bestehend aus Rechnern und aktiven Netzkomponenten. Die Verwaltungseinheit, beispielsweise ein Provider oder das Leibniz-Rechen-Zentrum, ist für das Management aller Geräte innerhalb des AS verantwortlich. OSPF wird nur innerhalb des autonomen Systems eingesetzt. Router die "am Rand" des autonomen Systems sitzen, also den Verkehr aus dem AS nach außen und umgekehrt vermitteln, nennt man "Boundary Router". Für das Routing zwischen den autonomen Systemen werden sogenannte "Exterior Gateway" Protokolle wie BGP (siehe 3.1.2.) eingesetzt. Der "Boundary Router" setzt sowohl ein "Interior Gateway Protokoll" wie OSPF, für die internen Routing-Tabellen, wie auch ein "Exterior Gateway Protokoll" wie BGP, für die Informationen über die Verbindungen zu den einzelnen autonomen Systemen, ein. Diese Topologie-Informationen setzt er in einen Zusammenhang und kann somit als Schnittstelle fungieren. Intern gibt er die externen Topologie-Informationen mit Hilfe von "external LSA"-Paketen weiter.

Durch den Einsatz von autonomen Systemen wird der Einflußbereich von OSPF eingegrenzt. Ressourcen wie Rechenzeit oder Bandbreite werden eingespart. Manche autonome Systeme sind allerdings trotzdem zu groß um mit Hilfe von OSPF effizient verwaltet werden zu können. In solchen Fällen können im Netz weitere Hierarchiestufen eingefügt werden, sogenannte Areas. Areas sind wiederum eine beliebige Zusammensetzung von aktiven Netzkomponenten und Rechnern. Die Systeme müssen miteinander verbunden sein. Zwischen den Areas läuft die Kommunikation über sogenannte "Area Border Router", wobei jeder dieser Router genau eine Area vertritt. Innerhalb der Area wird ganz normal OSPF eingesetzt. Die Boundary Router versenden an alle anderen Boundary Router (durch Einsatz von Flooding) LSA's, die die Informationen der eigenen Area zusammenfassen.

Bisher wurde nur die Punkt-zu-Punkt Übertragung berücksichtigt, also die sogenannte Unicast-Sendung. Um Netzressourcen zu sparen, vor allem im Hinblick auf die anspruchsvollen Multimedia-Anwendungen, werden Multicasts (ein Netzknoten sendet an viele andere Netzknoten) oder auch Broadcasts (ein Netzknoten sendet an alle anderen Netzknoten) eingesetzt. Die Vermittlungseinheiten (Multicast-Router) sind in der Lage die Multicast-Übertragung als solche zu erkennen und von den Paketen entsprechend viele Kopien über die Empfänger-Schnittstellen zu senden. Es stellt also nicht der Sender sondern die Multicast-Router die entsprechende Anzahl an Paketen zur Verfügung. In der Abbildung sind zwei lokale Netze über Router verbunden.

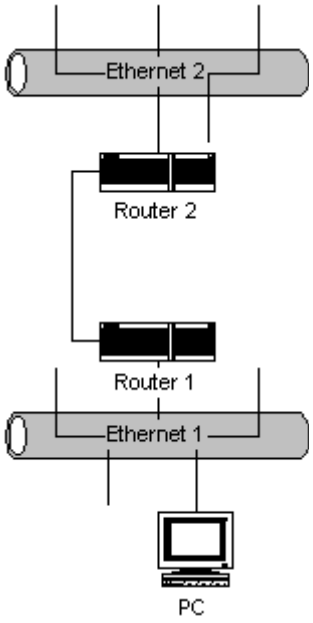


Abbildung Multicast-Übertragung

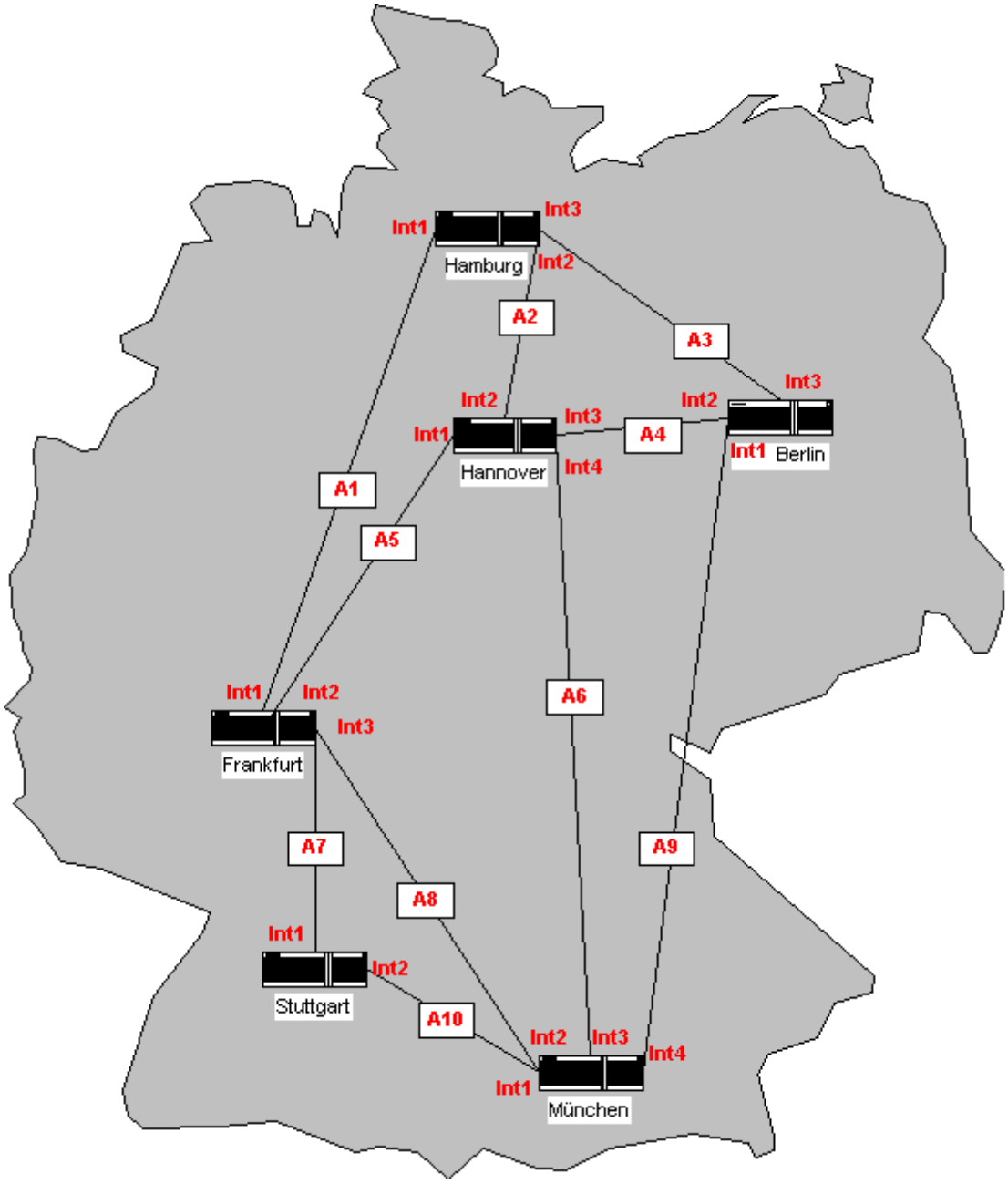
Will nun der PC aus Ethernet 1 eine Sendung an 30 PCs aus Ethernet 2 machen, so muß er bei einer Unicast-Sendung diesen Vorgang dreißig mal wiederholen. Bei einer Multicast-Sendung erkennen die Router den Vorgang. Router 1 schickt das Paket an Router 2, welcher dann entsprechend viele Kopien herstellt. Auf der Leitung zwischen den Routern wird somit nur 1/30 der ursprünglichen Bandbreite gebraucht. Natürlich müssen beide Router ein Multicast-Routing Protokoll beherrschen, wie zum Beispiel Multicast-OSPF (RFC 1584, RFC 1585). Die Übertragung wird dann mit Hilfe von Multicast-Adressen durchgeführt, welche jeweils eine Gruppe von Endsystemen repräsentieren. Die Endsysteme müssen sich in solchen Gruppen selbständig ein- und austragen. Weitere Details können in den angegebenen RFCs nachgelesen werden.

5.1.1.2 Distance Vector Routing am Beispiel von RIP

OSPF bietet schnelles und effizientes Routing. Allerdings ist das Verfahren sehr komplex. Die Router benötigen hohe Rechenkapazität, sind somit teurer, und die Netzlast ist entsprechend hoch. Wird nicht die volle Leistungsfähigkeit von OSPF benötigt, z.B. in kleineren Netzen, so empfiehlt sich der Einsatz des Routingprotokolls RIP (Routing Information Protokoll), welches in RFC 1058 erstmals beschrieben ist (RFC 1723 beschreibt ein RIP-Update, in RFC 2080 stehen Informationen über RIP der nächsten Generation "RIPnG for Ipv6").

RIP funktioniert nach dem "distance-vector"-Verfahren, welches in der Funktionsweise deutlich von OSPF abweicht. OSPF sammelt Informationen über seine Nachbarn und verteilt diese dann per "LSA-Flooding" im Netz. Die Router berechnen, basierend auf diesen Informationen, den gerichteten Dijkstra-Graphen, welcher die Pfade zu den einzelnen Teilnetzen repräsentiert. Bei RIP werden alle Informationen über das Netz, die der Router besitzt (nicht nur die über die Nachbarn) weitergegeben. Hier findet kein "Flooding" statt. Die einzelnen Router geben alle ihre Informationen in regelmäßigen Abständen lediglich an ihre direkten Nachbarn weiter.

Das Grundprinzip des distance-vector-Routing wird durch das folgende Beispiel, dem wieder die bereits bei OSPF verwendete Deutschlandkarte zugrunde liegt, erklärt.



RIP-Beispiel: Deutschland-Router

Die einzelnen Router besitzen Informationen über ihre Nachbarn da sie mit ihnen eine Verbindung und damit ein IP-Netz teilen. Ausgehend von diesen Startinformationen beginnt z.B. Berlin damit seine Nachbarn über seine Routing-Tabellen zu informieren. Die Mitteilung beinhaltet Informationen über IP-Netze von deren Existenz der Router weiß, sowie den Kosten um diese Netze zu erreichen. RIP setzt hier, wie die meisten distance-vector Routingprotokolle die Anzahl der benötigten Hops als Kosten ein. Die "Hop"-Zahl gibt die Anzahl der Router wieder (inklusive eigener Router), welche die Datenpakete passieren müssen um in das

Zielnetz zu gelangen. Für Berlin sind alle Ziele in A3, A4 und A9 einen hop entfernt. Für diese Ziele wird bei Kosten 1 übertragen. Parallel zu Berlin übermitteln alle anderen Städte ihren Informationsstand. Somit erhält Berlin Routing-Informationen aus Hamburg, Hannover und München. Durch diese Informationen ist Berlin bereits in der Lage über seinen Nachbarn hinauszublicken. Hannover ist beispielsweise 1 hop von Berlin entfernt und somit sind alle Nachbarn (bis auf Berlin) von Hannover 2 hops entfernt. Durch das regelmäßige Übertragen von Routerinformationen kann schließlich jeder Router eine Tabelle aufbauen die jedes Zielsystem im Netz enthält. Diese Tabelle enthält neben dem Zielnetz die Anzahl der Hops bis zum Ziel und das Interface des Routers, über welches er die Pakete schicken muß (Die Schnittstellen sind in der Zeichnung mit Int n beschriftet). Das Interface erfährt ein Router dadurch, daß er sich merkt woher er die Routinginformation bekommen hat. Berlin erfährt durch eine Nachricht von Hannover, daß A6 2 hops entfernt ist. Dadurch wird als Interface auf dem Weg zu A6 Int2 eingetragen. Für Berlin würde die vollständige Routingtabelle beispielsweise folgendermaßen aussehen:

Ziel	Kosten in hops	Interface
A3	1	Int3
A4	1	Int2
A9	1	Int1
A1	2	Int3
A5	2	Int2
A2	2	Int3
A6	2	Int2
A8	2	Int1
A10	2	Int1
A7	3	Int1

Für manche Teilnetze gibt es redundante Wege (z.B. für A2 sowohl über Hamburg wie auch Hannover von Berlin aus 2 hops), wenn man nur den Hop-Count als Kosten berücksichtigt. Da der Router in diesem Fall nur anhand der Hop-Zahl entscheidet, darf es nur einen Eintrag pro Teilnetz geben. Sobald der Router also einen Eintrag gemacht hat, ändert er ihn nur, wenn er über einen Weg mit kleinerer Hop-Zahl erfährt.

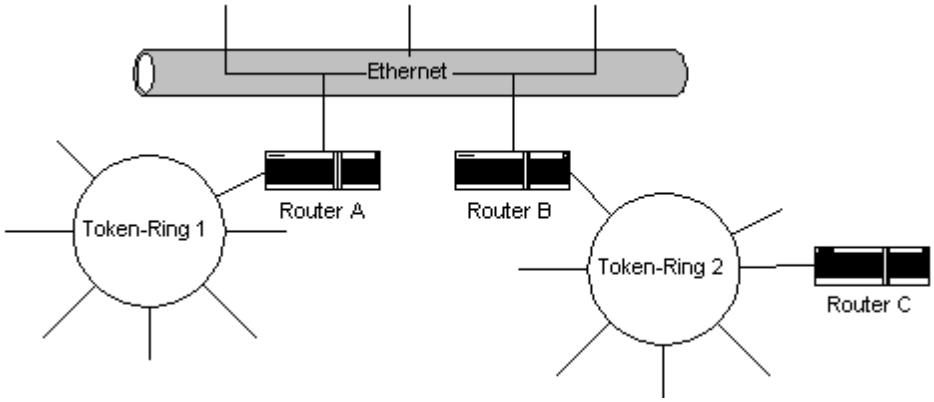
Nach dem Austausch aller RIP-Messages kennen die Router die komplette Netztopologie. Die Daten werden aktuell gehalten, indem regelmäßig RIP-Update-Pakete versendet werden. Diese Pakete teilen dem Nachbarn die aktuellen Informationen mit, so daß dieser seine Einträge anpassen kann. Für das Senden von Update-Paketen werden Timer festgelegt. Standardmäßig werden diese Informationen alle 30 Sekunden verschickt. Angenommen der Router in Stuttgart fällt aus und sendet somit keine Update-Pakete mehr. Die Nachbar-Router streichen alle Teilnetze die sie über Stuttgart bisher erreichen konnten aus ihrer Routing-Tabelle. Für das Herausnehmen von Informationen gibt es auch einen Timer, der standardmäßig auf 180 Sekunden, also 3 Minuten eingestellt ist. Der Router in München streicht also zum Beispiel den A7-Eintrag über Interface Int1 nach 3 Minuten aus seiner Routing-Tabelle. Spätestens 30 Sekunden später erhält der Router in Berlin eine Update-Nachricht aus München, woraufhin er auch den A7-Eintrag über Int1 streicht. Nach mehreren Minuten wird Berlin durch mehrere

Update-Nachrichten einen Alternativ-Weg finden, so daß die neue Routing-Tabelle wie folgt aussieht:

Ziel	Kosten in hops	Interface
A3	1	Int3
A4	1	Int2
A9	1	Int1
A1	2	Int3
A5	2	Int2
A2	2	Int3
A6	2	Int2
A8	2	Int1
A10	2	Int1
A7	3	Int2

Dieser Aktualisierungsvorgang dauert allerdings einige Minuten. Diese Tatsache zeigt eine Schwäche von RIP auf: die hohe Zeitdauer (time-to-convergence) die verstreicht, bis eine Netzänderung tatsächlich von allen Routern wahrgenommen wurde. Bereits im kleinen Beispielnetz werden 3 Runden an Updates benötigt damit jeder Router eine komplette Tabelle aufbauen kann. Angenommen die Updates werden in 30-Sekunden Intervallen abgeschickt, dann dauert dies 90 Sekunden. Um diese Schwäche teilweise zu umgehen gibt es die Möglichkeit von sogenannten "Triggered Updates". Sobald ein Router Änderungen in der Netztopologie feststellt sendet er sofort ein routing-update, ohne auf seinen nächsten, festgelegten Sendezeitpunkt zu warten. Nur die Informationen die sich geändert haben, werden übertragen. Die gesamten Informationen werden beim nächsten periodischen
Update gesendet. Es wurde bereits beschrieben, daß ein System bei Ausfall aus den Routing-Listen gestrichen wird. Tatsächlich wird das System vorerst in der Tabelle gehalten und die Kosten werden auf unendlich (meistens eine Maximalzahl wie z.B. 16, siehe weiter unten für mehr Informationen) gesetzt. Die getriggerten Updates verbreiten sich schnell im ganzen Netz, da alle Router, nach Erkennung von Änderungen in der Netztopologie, sofort ihre Update-Pakete senden. Die Reaktionszeit des Netzes auf Topologieänderungen wird somit deutlich verbessert.
Dieses Verfahren funktioniert allerdings nur schnell bei Ausfällen von Netzen, bzw. Verbindungsleitungen. Fällt ein Router aus, so wird dies, wie schon erwähnt, erst nach 3 Minuten von den Nachbar-Routern realisiert und weitergegeben.

Das bekannteste Problem von RIP ist "counting to infinity". Anhand der Abbildung wird dieses Problem aufgezeigt.



Beispiel: Counting to infinity

Alle drei Router besitzen bereits die aktuellen Routingtabellen in denen die drei Netze enthalten sind, welche auch auf der Abbildung zu sehen sind. Nach einiger Zeit fällt Router A aus, so daß Token Ring 1 nicht mehr erreichbar ist. Router B und Router C wissen zunächst nichts von diesem Ausfall. Nach maximal 30 Sekunden bemerkt Router B, daß von Router A keine Update-Pakete mehr gesendet werden. Router B startet in diesem Moment den Timer (180 Sekunden) um anschließend Router A aus der Routing-Tabelle zu entfernen, falls bis dahin keine Nachricht von Router A geschickt wurde. Parallel zu diesem Vorgang sendet Router B seine periodischen Update-Pakete, in denen Token-Ring 1 enthalten ist, an Router C. Router C bekommt nichts von den Veränderungen und dem Timer-Countdown mit.
Nach Ablauf des Timers entfernt Router B die Route zu Token-Ring 1 aus seiner Datenbank. Bevor er allerdings diese Aktualisierung an Router C weitergeben kann, erhält Router B ein Update-Paket von Router C, in dem Token-Ring 1 noch enthalten ist. Die Route zum Token Ring 1 ist zwar unterbrochen, der Router C kann es jedoch nicht wissen da er regelmäßige Updates vom Router B bekommen hat, in denen Token Ring 1 enthalten war. Router B geht jetzt davon aus, daß er Token-Ring 1 über Router C erreichen kann und trägt die neue Route in seiner Datenbank ein.
Im nächsten Schritt sendet Router B das nächste Update mit der erhöhten Entfernung zu Token Ring 1 an Router C. Dieser erhöht daraufhin seine eingetragene Entfernung zu Token Ring 1, da er nach wie vor annimmt ihn über Router B erreichen zu können. Dieser Ablauf, welcher als "Rooting-Loop" bezeichnet wird, setzt sich wie beschrieben fort. Die Entfernung wird so lange erhöht bis ein Wert erreicht wird der 'unendlich' darstellt. Erst durch Erreichen dieses Wertes erkennen beide Router daß der Token Ring nicht mehr erreichbar ist. Die Auswahl dieses Wertes wird von zwei Überlegungen beeinflußt. Er darf einerseits nicht zu klein sein, da er eine Obergrenze für die Größe des Netzwerks darstellt. Andererseits darf er nicht zu groß sein, da sonst "counting to infinity" entsprechend lange Zeit und Netzwerkressourcen in Anspruch nimmt. In RIP ist der Wert 16 als "unendlich" definiert. Ausgehend von diesem Wert darf kein System mehr als 15 hops von einem anderen entfernt sein. Mit "Split Horizon" stellt RIP ein weiteres Verfahren zur Verfügung um die "counting to infinity"-Problematik zu umgehen. Der Vorgang ist schnell erklärt: ein Router empfängt ein RIP-Update auf einem seiner Interfaces mit einer neuen Route. Nach dem Ablauf seines Timers, sendet auch er über das Interface seine aktuellen Daten als RIP-Update. Hierbei wird allerdings die neue Route nicht übertragen. Dieser Pfad wird nur allen anderen Verbindungen mitgeteilt.

5.1.2. Externe Gateway Protokolle am Beispiel von BGP

Das Border Gateway Protokoll (BGP) ist ein sogenanntes "exterior gateway protocol". Wie schon erwähnt, dient es zur Kommunikation zwischen Routern, welche als Vertreter ihres autonomen Systems ausgewählt wurden. In RFC 1267 wird BGP Version 3 beschrieben. RFC 1268 beschreibt die Verwendung von BGP im Internet. BGP-Router tauschen Netz-Erreichbarkeitsinformationen mit anderen BGP-Routern aus. Diese Daten beinhalten jeweils komplette Pfade von autonomen Systemen zu ihren Zielnetzen und dienen als Grundlage zur Berechnung der einzelnen Routing-Tabellen.

BGP unterscheidet zwischen zwei Arten von IP-Verkehr, nämlich den sogenannten "local traffic" und den "transit traffic". Von "local traffic" spricht man, wenn das Paket im betreffenden autonomen System entweder seinen Ursprung oder seinen Ziel hat. Jeder andere Verkehr wird als transit traffic bezeichnet. Eine der Hauptaufgaben von BGP im Internet ist die Reduzierung des transit traffic. Es kann zwischen verschiedenen autonomen Systemen unterschieden werden:

- Ein stub AS besitzt eine einzige Verbindung zu einem anderen AS. Der Verkehr ist immer lokal.
- Ein multihomed AS besitzt zu mehreren anderen AS eine Verbindung. Auch hier ist der Verkehr rein lokal.
- Ein transit AS hat Verbindungen zu mehr als einem anderen AS. Das autonome System ist allerdings in der Lage Lokal- und Transitverkehr zu transportieren. Für den Transitverkehr werden verschiedene Policy-Mechanismen berücksichtigt.

Die Topologie des Internet ist als Verbindung von transit-, multihomed- und stub-AS zu sehen. Ein wichtiges Kriterium für den Einsatz von BGP ist die Möglichkeit der Verwendung des policy-based routing. Der AS-Systemadministrator kann hierbei Einschränkungen, bzw. Prioritäten für den Ablauf des Routing-Vorganges festlegen. So kann beispielsweise eine Einschränkung für die Pfadwahl festgelegt werden (z.B. um die Vermittlung von inländischen Daten nicht über ausländische Router zuzulassen). Im Gegensatz zu den oben vorgestellten Protokollen RIP und OSPF, setzt BGP als Transportprotokoll TCP ein. Zwischen den kommunizierenden Routern wird also vor der Sendung der Update-Pakete eineVerbindung aufgebaut. Um den Ausfall von Verbindungen oder Hosts zu erkennen versenden die BGP-Router regelmäßig sogenannte "Keep-Alive"-Pakete. Das Prinzip von BGP entspricht dem distance-vector-Routing.

5.2. Wirespeed Routing - Layer-3-Switching

Das sogenannte "Layer-3-Switching" ist, wie schon in der kurzen Einführung in Kapitel 3.6. erwähnt, eine Weiterentwicklung des herkömmlichen Routings, welche sich im wesentlichen auf die Erhöhung des Datendurchsatzes beschränkt. Während herkömmliche Router mittlerweile ungefähr eine Million Pakete in der Sekunde (pps = Packets per second) an Datendurchsatz schaffen, schafft ein Layer-3-Switch mehrere Millionen Pakete in der Sekunde un kann somit als "Hochleistungs-Router" bezeichnet werden. Die drei Hauptaufgaben eines Routers sind:

- Paketweiterleitung
- Berechnen der Routen und Verteilen der Routinginformationen im Netz
- Bereitstellung von intelligenten Netzwerkdiensten wie beispielsweise Mechanismen zur Authentifizierung

Die Realisierung der Paketweiterleitung, also die physische Implementierung, unterscheidet sich bei herkömmlichen Routern von der Realisierung bei Layer-3-Switches. Während die Berechnungen in traditionellen Routern mit Hilfe von Mikroprozessoren durchgeführt werden, ist diese Funktion im Layer-3-Switch mit Hilfe von anwendungsspezifischen Schaltkreisen, sogenannten ASICs (application specific integrated circuit) realisiert. ASICs sind zwar teurer, aber dafür um ein vielfaches schneller als die Mikroprozessoren. Die Paketweiterleitung läuft wie folgt ab:

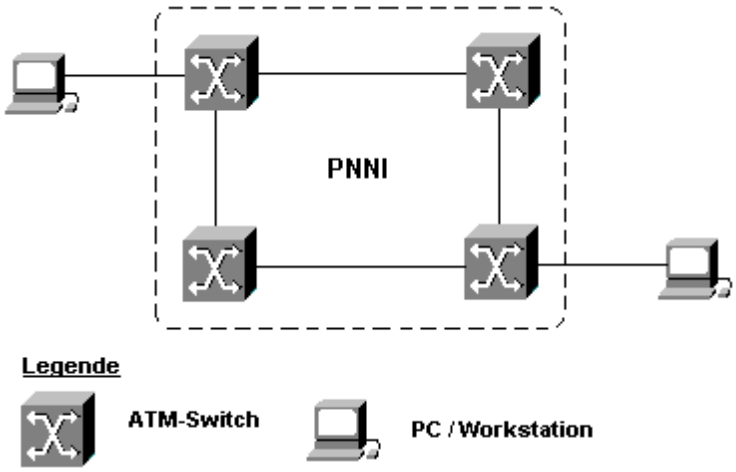
- Untersuchen des eingehenden Paketes, Feststellung der IP-Adresse des Zieles
- Vergleich des IP-Zielnetzes mit den Einträgen in den Routing-Tabellen
- Das Time-To-Live-Feld des IP-Paketes wird um eine Einheit verkleinert
- Ersetzen der Quell-MAC-Adresse durch die MAC-Adresse des Routers
- Berechnung der neuen Frame-Check-Sequence (FCS)
- Weiterleiten des Paketes zur Ausgangsschnittstelle

Dieser Ablauf wird auch "paket-by-packet"-Vermittlung genannt, und wird immer gleich, sowohl von traditionellen Routern wie auch von Layer-3-Switches, ausgeführt. Um es an dieser Stelle nochmals zu verdeutlichen: Der Layer-3-Switch beinhaltet in seinem Namen die Bezeichnung "Switch" aus Marketing-Gründen. Der Schicht-3-Mechanismus des Routings und der Paketanalyse wird nicht auf Schicht-2 umgangen, was man eventuell, aufgrund der Bezeichnung, erwarten würde. Mit der Bezeichnung "Switch" soll lediglich die erhöhte Geschwindigkeit assoziiert werden.

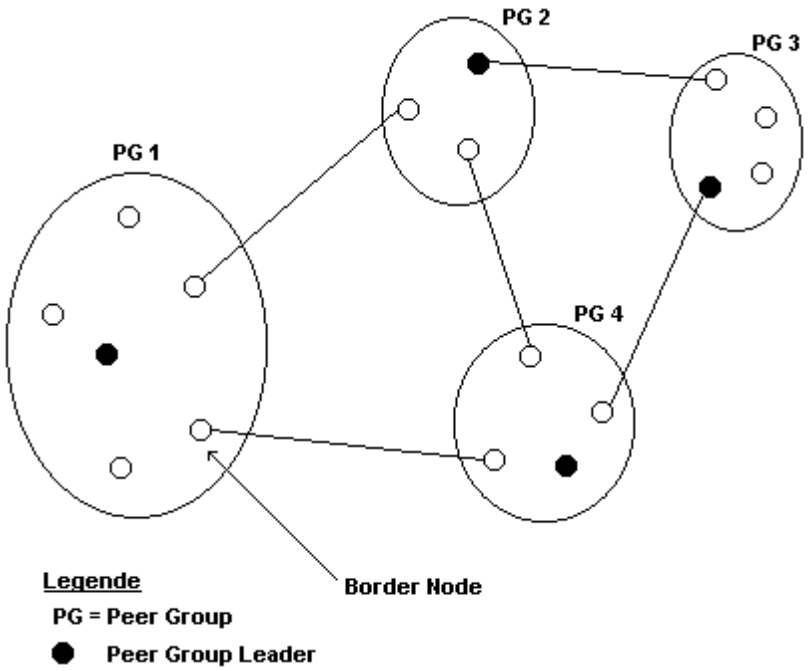
5.3. ATM Routing - PNNI und IISP

Damit eine Vermittlungseinheit eines ATM-Netzes Daten weiterleiten kann, benötigt sie Topologieinformationen. Die ATM-Spezifikationen stellt hierfür zwei Routing-Protokolle zur Verfügung: PNNI (Private Network Network Interface), für dynamisches Routing, und IISP (Interim Inter-Switch Signaling Protocol), für statisches Routing. Weiterhin beinhalten beide Protokolle Signalisierungsverfahren für den Verbindungsaufbau.

5.3.1. Private Network Network Interface (PNNI)



PNNI ist ein dynamisches Routing-Protokoll für ATM-Netzwerke welches, wie OSPF, seine Routing-Tabellen nach dem Link-State Prinzip erzeugt. Da die Verteilung der Topologie-Informationen Bandbreite beansprucht, bietet PNNI die Möglichkeit Hierarchie-Stufen zu definieren. Einzelne ATM-Switching-Systeme werden zu sogenannten, in der Einführung bereits erwähnten, Peer-Groups (PG) zusammengefaßt. In jeder Peer-Group existiert ein "Peer-Group-Leader", der die Gruppe nach außenhin repräsentiert.



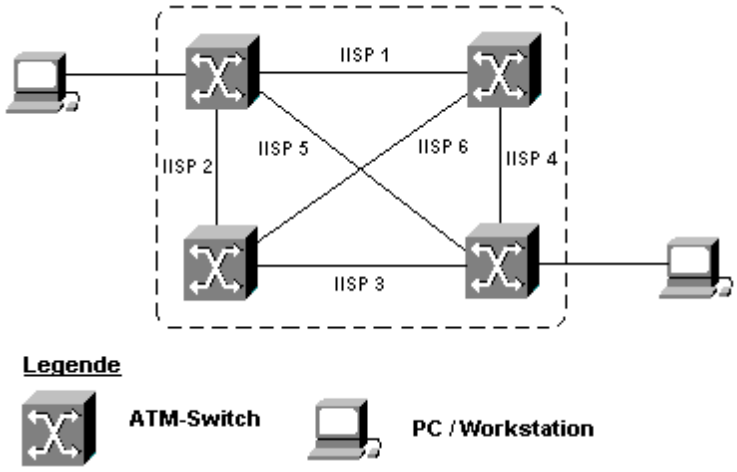
Jeder Netzknoten tauscht mit seinen Nachbarn innerhalb der Peer-Group Statusinformationen aus, auch PTSE (PNNI Topology State Elements) genannt. Diese Daten werden in sogenannten "Hello-Packets" übertragen und beinhalten Informationen über die Identität des Senders und den Status seiner Verbindungen zu den Nachbarn. ATM-Schalteinheiten, die über die eigene Peer-Group hinausgehende Verbindungen verfügen, werden "Boarder-Nodes" genannt. Die "Peer-Group-Leader" tauschen untereinander verkürzte Topologieinformationen über die Netzknoten der jeweils eigenen Peer-Group aus, und geben dann diese

Informationen innerhalb der eigenen Peer-Group weiter. Jede Schalteinheit besitzt somit eine Übersicht über das gesamte Netzwerk. Um den Speicher des ATM-Switches nicht zu überlasten, werden die Topologieinformationen mit zunehmender Entfernung des Zielknotens unschärfer, also weniger detailliert.

Für die Datenübertragung zwischen zwei Endsystemen wird, im Gegensatz zum IP-Routing, zunächst eine Verbindung aufgebaut (Connection Setup). Hierfür existieren bei der PNNI-Signalisierung verschiedene Methoden wie Source-Routing und Alternate-Routing (detaillierte Beschreibung mit Beispiel in [42] Seite 39 ff.). Während die Verbindung aufgebaut wird, werden QoS-Parameter mit den einzelnen Vermittlungseinheiten (ATM-Switches) vereinbart. Nur wenn jede Vermittlungseinheit die Einhaltung der Parameter garantieren kann, wird die Verbindung aufrechterhalten und die Übertragung kann beginnen.

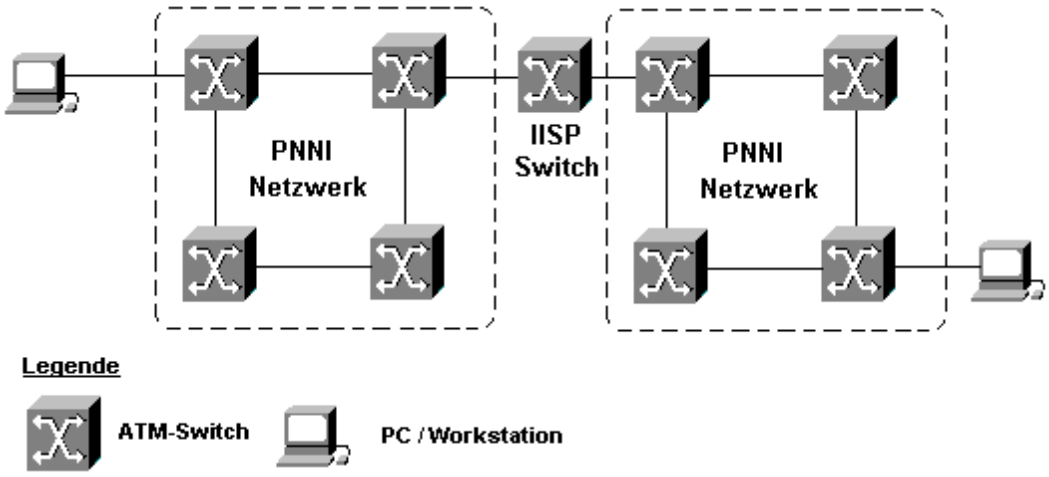
5.3.2. Interim Inter-switch Signalling Protocol (IISP)

IISP ist ein statisches Routing-Protokoll. Die Routen werden manuell vorgegeben und müssen bei jeder Topologieänderung entsprechend angepasst werden. Da sich der Administrationsaufwand mit steigender Anzahl der ATM-Switches und Topolgieänderungen deutlich erhöht, empfiehlt sich ein Einsatz von IISP nur in kleineren Netzen.



Interim Inter-switch Signalling Protocol (IISP)

IISP läßt sich mit PNNI kombinieren, wie beispielsweise in der folgenden Abbildung.



5.4. Shortcut Routing

5.4.1. IP-Switching

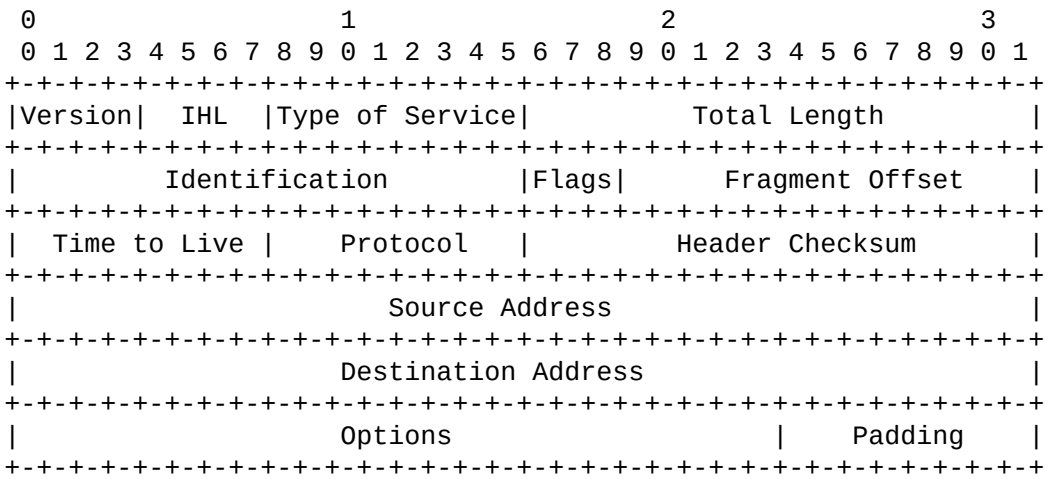
Wie schon erwähnt sind die meisten Ansätze für Shortcut-Routing Vorschläge, bzw. Initiativen von Herstellerfirmen. Grundsätzlich kann man alle, im Kapitel Shortcut-Routing, vorgestellten Verfahren als IP-Switching bezeichnen. Das unter dem Namen "IP-Switching" vorgestellte Konzept stammt von der Firma Ipsilon Networks, welche erstmals diese Bezeichnung für die spezielle Technik wählte. Man muß also bei den Namen unterscheiden, ob man grundsätzlich über eine Technik berichtet oder, ob man herstellerspezifische Informationen weitergibt. Ipsilon Networks bietet mit "IP-Switching" eine proprietäre Lösung, welche bereits zu einiger Verbreitung gelangt ist. Firmen wie IBM, 3Com und Compaq haben die Ideen aufgegriffen und Kooperationsprojekte gestartet.

Ein IP-Switch besteht aus zwei Komponenten: einem herkömmlichen ATM-Switch und einem sogenannten IP-Switch-Controller. An der ATM-Switch-Hardware werden keine Änderungen durchgeführt. Allerdings wird die ATM-Steuerungssoftware, welche unter anderem für die Signalisierung und das PNNI-Routing zuständig ist, durch das sogenannte "General Switch Management Protokoll" (GSMP, beschrieben in RFC 1987) ersetzt. Mit Hilfe des Kontrollprotokolles GSMP hat der IP-Switch-Controller Zugriff auf die Hardware des ATM-Switches. Der IP-Switch-Controller selbst ist meist ein gut ausgestatteter schneller Rechner (z.B. Pentium Pro III), auf dem die bereits beschriebene Standard IP-Router-Software läuft. Hinzu kommen einige Erweiterungen für den Umgang mit dem ATM-Switch:

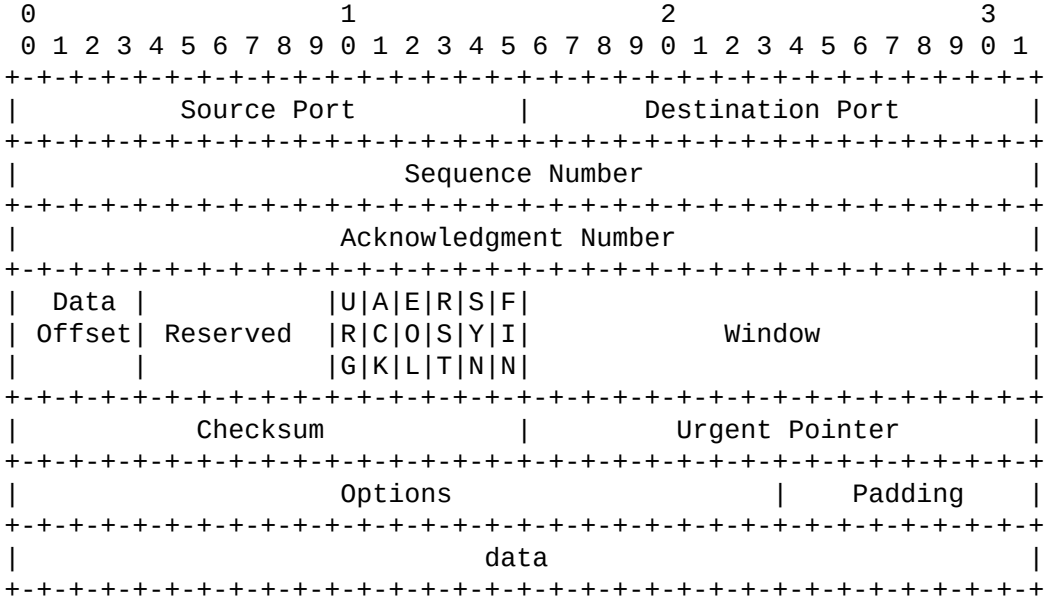
- das Ipsilon Flow Management Protocol (IFMP, beschrieben in RFC 1953) für die Abbildung von IP-Datenflüssen (flows) auf ATM-Verbindungen
- eine Fluß-Klassifizierung
- GSMP zur Kommunikation mit dem Switch.

Der IP-Switch ist in zwei Komponenten aufgeteilt. Diese Realisierung hat den Vorteil, daß Hardware und Software getrennt sind. Die einzige Bedingung an den ATM-Switch ist die GSMP-Fähigkeit. Zu Beginn baut der IP-Switch-Controller eine Default-ATM-Verbindung zu seinen Nachbarn auf. IP-Pakete werden über diesen Kanal in

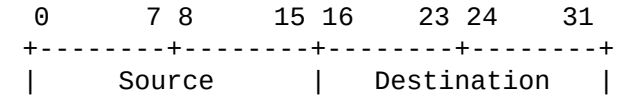
gewohnter Weise (hop-by-hop) weitergeleitet. Um nun die Vorteile des IP-Switchings zu nutzen, müssen die einzelnen IP-Flows auf ATM-Verbindungen abgebildet werden. Hierfür muß eine Flußklassifizierung stattfinden. Ein Fluß ist eine länger andauernde Datenübertragung gleichen Typs, also beispielsweise ein FTP-Filetransfer oder eine Multimedia-Sitzung. SNMP-Pakete oder DNS-Abfragen werden nicht als Fluß gewertet und in traditioneller Weise, nach dem Store-and-Forward-Prinzip, übertragen. Die Klassifizierung von IP-Paketen erfolgt anhand der Headerinformationen, die die Routing-Entscheidung bestimmen.

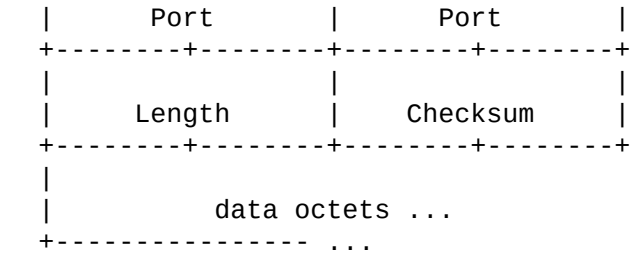


IP-Header, beschrieben in RFC 791



TCP Header Format, beschrieben in RFC 761





UDP Header Format, beschrieben in RFC 768

Hierfür werden folgende IP-Header-Felder untersucht:

- Type of Service (Diensttyp)
- Time-To-Live
- Protokoll (welches Transportprotokoll wird verwendet, z.B. TCP oder UDP)
- Quell-Adresse
- Ziel-Adresse

Folgende Header-Informationen werden dem Transportprotokoll entnommen:

- Quell-Port
- Ziel-Port

Zwei Pakete gehören zum selben Fluß, wenn bestimmte Felder übereinstimmen. Es lassen sich hierbei zwei Flußtypen festlegen: Port-Pair und Host-Pair. Port-Pair ist ein Fluß zwischen gleichen Quell- und Ziel-Ports (TCP/UDP) auf gleichen Quell- und Ziel-Adressen (IP). Hinzu kommt, daß auch der Diensttyp, das Transportprotokoll und das TTL-Feld übereinstimmen müssen. Host-Pair ist ein Fluß mit gleichen Quell- und Ziel-IP-Adressen mit gleichem TTL-Feld.

Wenn ein Paket vom ATM-Switch über den Default-Kanal empfangen wird, leitet er es an den IP-Switch-Controller, welcher dann eine Fluß-Klassifikation durchführt. Das Paket wird in gewohnter Weise weitergeleitet. Die Klassifizierung legt fest, ob es sich um einen Fluß handelt und, wenn ja, ob zukünftige Pakete desselben Flusses direkt von der ATM Hardware weitergeleitet (geswitcht) werden sollen. Die Fluß-Klassifizierung wird von jedem IP-Switch eigenständig durchgeführt. Es handelt sich also hierbei um eine lokale Entscheidung, die auf andere Switches keinen Einfluß hat. Dadurch bleibt der verbindungslose Charakter von IP erhalten und das Netz flexibel, z.B. bei Ausfall einer Station. Das Verhalten des IP-Switch-Controllers kann durch den Einsatz von Richtlinien (Policies) festgelegt werden. So kann der Netzverwalter beispielsweise bestimmen, daß FTP-Sessions geswitcht, DNS-Anfragen aber geforwardet werden. Weiterhin kann er festlegen nach wievielen Paketen (innerhalb einer gewissen Zeit) ein Fluß erkannt wird.

Mit Hilfe der folgenden Abbildung wird das Prinzip von Ipsilons "IP-Switching" erklärt.

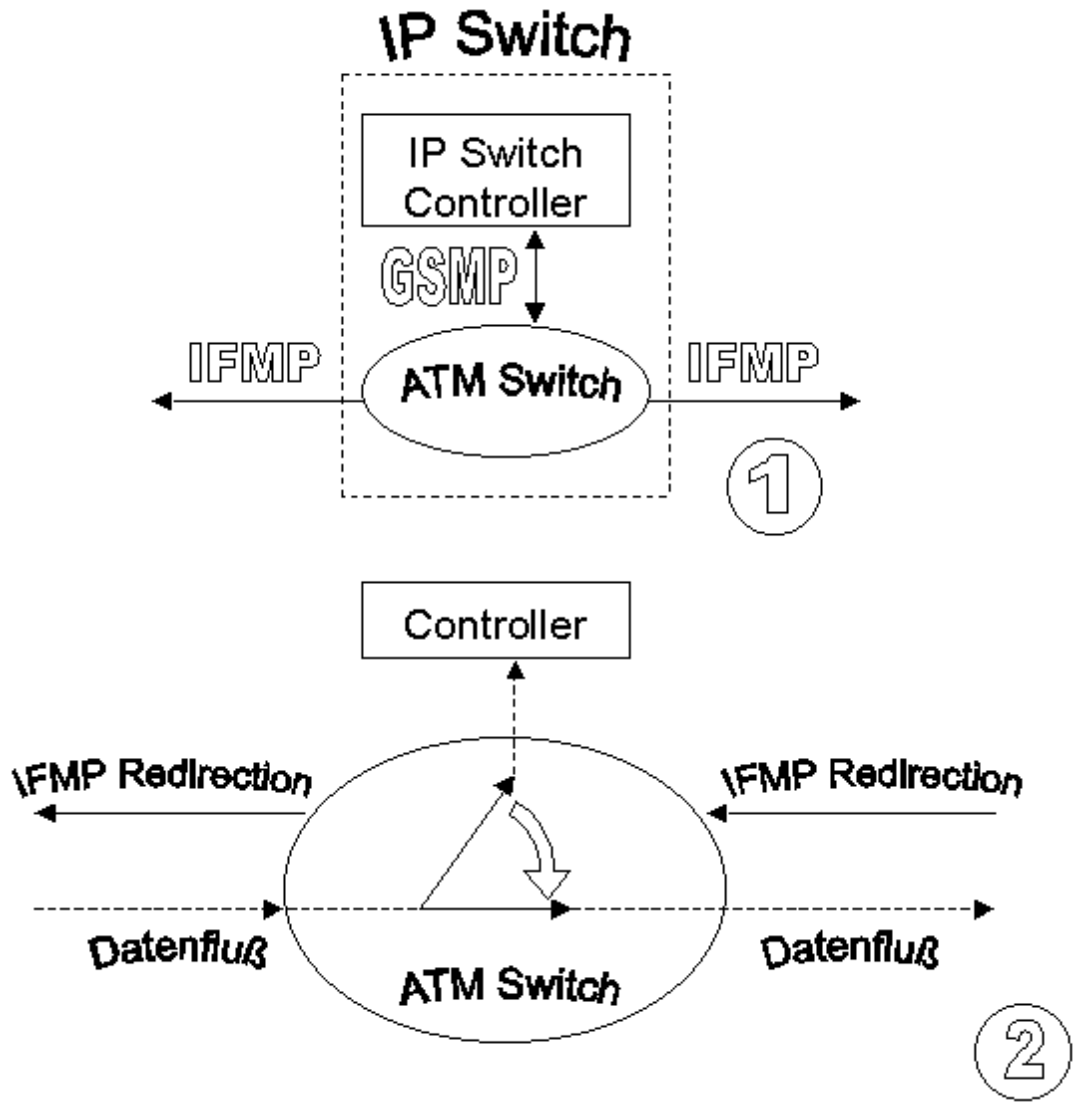


Abbildung: IP-Switching-Konzept

Alle in einem IP-Switching-Netz verbundenen Geräte kommunizieren über IFMP. Hierzu gehören die IP-Switches, direkt verbundene Hosts und IFMP-fähige Edge-Router (Router am Rande eines IP-Switching-Netzes). Bei ATM-Verbindungen benutzt IFMP den Default-Kanal (VPI 0, VCI 15). Das Ziel von IFMP ist, dem Sender der Daten ein Label (Wert für den VCI) zu übermitteln, der einem bestimmten IP-Fluß zugeordnet wurde. Der Sender versieht dann die Daten des entsprechenden Flusses mit dem Label, so daß der Empfänger diese direkt weiterleiten kann.
Pakete eines Flusses, für den noch keine Label-Zuordnung stattgefunden hat, werden über den Default-Kanal des IP-Switch-Controllers geforwardet und, wie bereits beschrieben, klassifiziert. Im folgenden wird das Einrichten eines geswitchten Datenflusses erklärt.

Nachdem der IP-Switch-Controller entschieden hat, einen Fluß zu switchen, wählt er ein freies Label (VCI=x) aus dem Label-Bereich für den Eingangsport (Port i) des ATM-Switches, an dem der Fluß ankommt. Weiterhin wählt der Controller ein freies Label (VCI=y) seines Kontroll-Ports (Port c) aus, der ihn mit dem ATM-Switch verbindet. Der IP-Switch-Controller legt nun über GSMP fest, daß der ATM-Switch

auf Port i und Label VCI=x empfangene Pakete zum Controller über Port c und VCI=y weiterleiten soll. Zu diesem Zeitpunkt findet noch kein Verkehr über die neu festgelegten Pfade statt. Um den ursprünglichen Fluß umzuleiten, nimmt der IP-Switch Controller über IFMP Kontakt zum direkten Nachbarn auf, von dem die Daten geschickt wurden. Dieser Nachbar erhält eine IFMP-Redirect-Botschaft, die das freie Label (VCI=x), einen Fluß-Identifikator und die Lebensdauer (TTL, Time To Live) dieser Zuordnung enthält. Der Fluß-Identifikator beinhaltet Informationen über die Felder, die ein Paket als zum Fluß zugehörig markieren. Nach Erhalt und Auswertung der Redirect-Botschaft sendet der Nachbar alle Pakete, bei denen die Felder mit den im Flußidentifikator übermittelten Werte übereinstimmen, auf dem virtuellen Kanal VCI=x. Diese Zuordnung gilt solange bis die im TTL-Feld der Redirect-Botschaft vereinbarte Zeitdauer abgelaufen ist. Danach muß eine neue Vereinbarung getroffen, bzw. die alte Vereinbarung erneuert werden. Die Switches befinden sich also in einem sogenannten "soft state"-Zustand. Dieses Prinzip hat den Vorteil, daß der Sender nicht unbegrenzt Daten weiterleitet, auch wenn der ursprüngliche Empfänger ausgefallen ist. Ist Verbindung etabliert, so kommen die Pakete des Flusses am Switch-Controller-Port c mit Label y an. Die ankommenden Pakete werden noch immer untersucht und weitergeleitet. Dieser Vorgang geschieht zwar schneller, da die Routing-Entscheidung im Cache unter dem Label-Index y abrufbar ist, doch zeigt sich der wahre Geschwindigkeitsvorteil erst, wenn der nachfolgende Knoten eine IFMP-Redirect-Nachricht zu diesem IP-Switch schickt. Der Vorgang der Label-Vergabe läuft unabhängig auf den einzelnen IP-Switch-Controllern ab. Somit ist der Ablauf des Verschickens der jeweiligen IFMP-Nachrichten zeitlich unabhängig voneinander ausführbar. Der IP-Switch-Controller ist nun in der Lage einen sogenannten "Shortcut" zu etablieren (siehe Abbildung 2). Der Fluß, welcher am Eingangsport i mit Label x ankommt, kann sofort zum Ausgangsport j mit Label z weitergeleitet werden, welcher dem nächsten Knoten für diesen Fluß zugeordnet wurde.

Mit dem Erhalt einer Redirect-Botschaft wechselt ein IP-Switch die Kapselung der Daten. Standardmäßig werden IP-Pakete gemäß RFC 1483 ([24], siehe auch [3] Seite 246) gemäß den Regeln der LLC-Enkapsulierung in AAL5-PDUs verpackt. Für Pakete eines mit IFMP umgeleiteten Flusses werden all die IP-Header Felder entfernt, die im Flußidentifikator übergeben wurden, also unter anderem die IP-Adressen von Quelle und Ziel. Der komprimierte Header wird in AAL-5 gekapselt und über den virtuellen Pfad versendet. Die entfernten Felder speichert der IP-Switch, der die Redirect-Botschaft ausgesandt hat. Die Pakete können so am Ziel eines Datenflusses wieder in die alten IP-Datagramme zusammengesetzt werden. Durch den Wegfall der einzelnen Header-Felder ist es nun auch möglich fluß-basierte Firewalls einzurichten. Dies hat den Vorteil, daß nicht jedes Paket einzeln untersucht werden muß.

IP-Multicast wird von Ipsilons Technik ohne Modifikationen des IP-Switches unterstützt. Ein IP-Switch, welcher mehrere Nachbarn innerhalb eines Multicast-Flusses mit Daten versorgt, erhält in diesem Fall mehrere Redirect-Nachrichten. Jeder Nachbar, der sich in der Multicast-Gruppe flußabwärts befindet schickt eine solche Nachricht. ATM unterstützt Multicast von sich aus. Der ATM-Switch kopiert Pakete , um sie an den entsprechenden Ports auszugeben, an denen eine Redirect-Nachricht ankam. Der Switch kann alternativ eine Kopie des Flusses an den Controller leiten, um ihn auf herkömmliche Art zu forwarden. Dies geschieht für die

Nachbarn welche sich zwar in der Multicast-Gruppe befinden, aber noch keine Redirect-Anfrage gestellt haben.

Ein weiteres wichtiges Einsatz-Kriterium ist die Unterstützung von Quality of Service (QoS). Ipsilons Konzept unterstützt hierbei sowohl policy-based als auch contract-based Quality of Service. Jeder IP-Switch kann aufgrund seiner Richtlinien (Policies), entscheiden, wann ein Paket zu einem Fluß gehört und welchen Weg die Daten nehmen sollen. Weiterhin können minimale Bandbreiten wie auch maximale Verzögerungszeiten definiert werden. Diese Vorgaben (Policies) werden vom Netzadministrator festgelegt. Bei der Flußklassifizierung werden auch die Felder beachtet, aus denen eine Dienstgüteanforderung abgeleitet werden kann. Jeder IP-Switch weiß über die Möglichkeiten, bzw. Implementierungen der darunterliegenden ATM-Hardware, und bildet die QoS-Realisierung entsprechend auf diese ab. Vertragsbasierte Dienstgüteanforderungen können mittels RSVP (Resource Reservation Protocol) vereinbart werden. Der Kontrakt wird normalerweise vom Empfänger der Daten initiiert. Jeder Switch auf dem Pfad zum Sender kann die Anforderungen annehmen und flußaufwärts weiterleiten oder ablehnen. Die Reservierung von Ressourcen kann entweder in der ATM-Switch-Hardware oder im IP-Switch-Controller durchgeführt werden. Ist der Vertrag von allen Komponenten auf dem Pfad angenommen, kann die Übertragung beginnen.

5.4.2. Tag-Switching

Tag-Switching ist ein Lösungsansatz der Firma Cisco Systems, beschrieben in RFC 2105. Auch bezüglich Tag-Switching arbeiten mehrere Firmen in einer IETF-Arbeitsgruppe zusammen um das Verfahren zu standardisieren. Ciscos Tag-Switch beinhaltet zwei Komponenten: eine Forwarding- und eine Kontrollkomponente. Die Beziehungen sind in der Abbildung verdeutlicht.

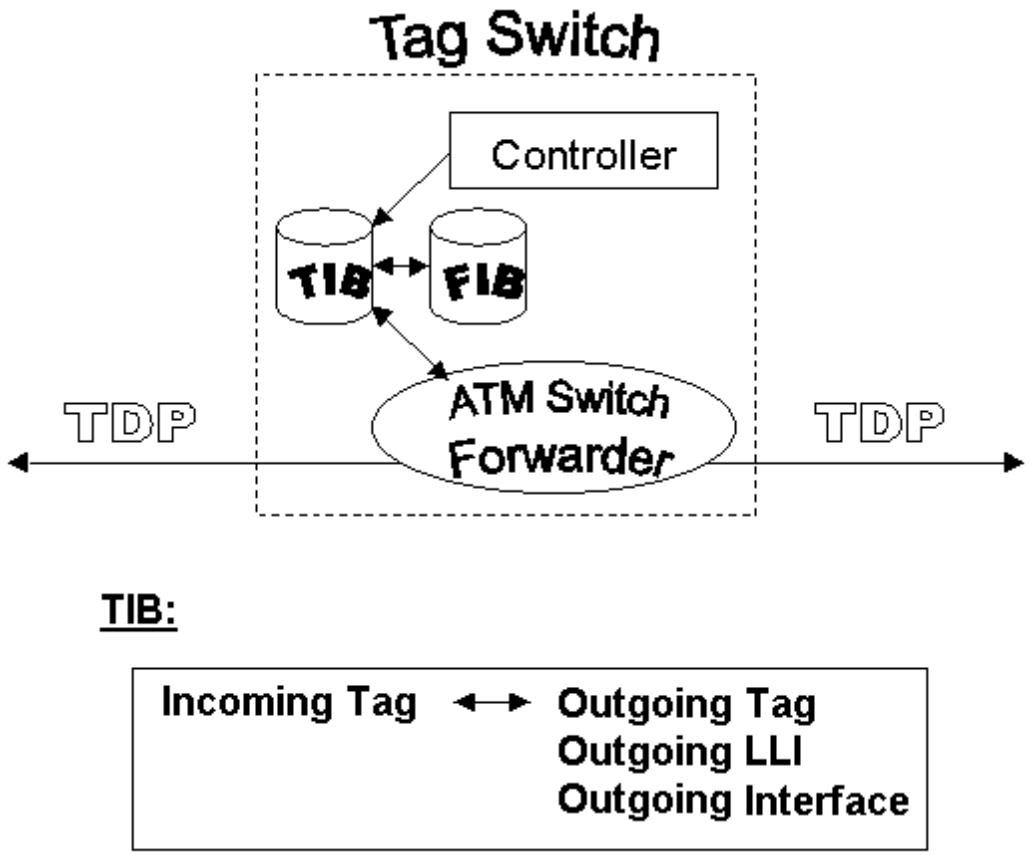


Abbildung: Tag-Switching-Konzept

Die Datenpakete beinhalten Labels, welche bei Ciscos Verfahren als "Tags" bezeichnet werden. Die Forwarding-Komponente entscheidet anhand des Labels und dem zugehörigen Pfad in der Tag Information Base (TIB) über die Weitersendung des Paketes. Die Control-Komponente ist dafür verantwortlich, korrekte Tag-Forwarding Informationen zu erzeugen, diese in die TIB und FIB (Forwarding Information Base) zu schreiben und innerhalb einer Gruppe von untereinander verbundenen Tag-Switches zu verteilen. Das grundlegende Prinzip das bei Tag-Switching Anwendung findet wird als Label-Swapping bezeichnet. Das Internet Protocol in der Version 6 (Ipv6) beinhaltet hierfür unter anderem ein eigenes Flow-Label-Feld. Empfangene Pakete, welche bereits ein Tag beinhalten, werden gesondert vom Tag-Switch untersucht. Falls für das Tag noch kein Eintrag in der TIB vorhanden ist, so wird dieses Tag als Index neu angelegt. Jeder Eintrag in der TIB besteht aus einem "Incoming-Tag" und einem oder mehreren Untereinträgen folgender Form:

- Outgoing-Tag
- Outgoing-Interface
- Outgoing-Link-Level-Information

Der Switch vergleicht das Tag des eingehenden Pakets mit den Einträgen in seiner TIB. Wird ein Eintrag gefunden so ersetzt der Switch das Tag im Paket mit dem Outgoing-Tag des Eintrags. Weiterhin wird die Link-Level-Information des Paketes, wie beispielsweise die MAC-Adresse, durch den Outgoing-LLI-Eintrag ersetzt. Das Paket wird danach an die eingetragene Schnittstelle "Outgoing-Interface" weitergeleitet. Ist das Tag nicht in der Datenbank registriert, so werden die Daten des Netzwerkschicht-Protokolls ausgewertet und traditionell geroutet.

Tag-Switching unterstützt sowohl Unicast- wie auch Multicast-Routingfunktionen. Unicast-Verbindungen haben einen Eintrag in der TIB pro Incoming-Tag. Für Multicast-Verbindungen werden mehrere Einträge angelegt. Das Label selbst kann auf verschiedene Weisen übertragen werden, zum Beispiel als Teil des Ebene-2-Headers, vorausgesetzt diese Option ist, wie bei ATM, gegeben. Alternativ kann das Tag auch zwischen den Headern von Ebene 2 und Ebene 3 übertragen werden. IPv6 bietet eine weitere Option mit Hilfe des Flow-Label-Feldes. Ein wichtiger Punkt ist hierbei, daß Tag-Switching im Vergleich zu IP-Switching unabhängig vom Ebene-3-Protokoll konzipiert wurde. Diese Tatsache ermöglicht den breiteren Einsatz des Verfahrens, besonders in den großen meist heterogenen Netzumgebungen.

Die Kontrollkomponente erzeugt die Tag-Bindungen und verteilt diese Informationen an die anderen Tag-Switches. Sie kann mehrere Standard-Routing-Module beinhalten. Neue Routing-Funktionen lassen sich somit leicht hinzufügen, indem ein neues Modul eingefügt wird.
Das "Tag-Switching"-Verfahren ist topologieorientiert. Besonders wichtig ist hierbei die Bindung zwischen einem Tag und einer Route. Solange ein Tag-Switch nicht auf diese Information zurückgreifen kann, kann er kein schnelles Forwarding durchführen. Die Möglichkeit der Tag-Zuweisung lässt sich flexibel skalieren. Einerseits kann ein Tag mehreren Routen zugeordnet werden. Es besteht aber auch die Möglichkeit ein Tag für einzelne Datenflüsse zu vergeben, zum Beispiel um für spezielle Anwendungen eine zugesicherte Datenrate zu realisieren.
Tags können auf drei verschiedene Verfahren zugewiesen werden:

- Downstream-Tag-Allocation
- Downstream-Tag-Allocation On Demand
- Upstream-Tag-Allocation

Bei dem Vefahren "Downstream-Tag-Allocation" wird für jede Route in der Forwarding Information Base (FIB) des Tag-Switches ein Tag generiert. Die Routen in der FIB werden traditionell durch Routingprotokolle wie OSPF oder BGP erzeugt. In der Tag Information Base wird ein neuer Eintrag erstellt, in dem das generierte Tag in das "Incoming-Tag" Feld eingesetzt wird. Anschließend gibt der Tag-Switch diese Information, über den Neueintrag und dessen Route, an die benachbarten Switches weiter. Die Übertragung der Informationen geschieht mit Hilfe des Tag Distribution Protocols (TDP) oder alternativ durch Integration der Daten in Routing-Update-Pakete (Huckepack) der bestehenden Routingprotokolle. Erhält nun ein Tag-Switch eine Tag-Bindungsinformation und diese kam vom nächsten Hop auf der entsprechenden Route, so plaziert der Switch das zugewiesene Tag im Outgoing-Tag-Feld seines TIB-Eintrags, welcher mit der Route verknüpft ist.

Bei "Downstream-Tag-Allocation On Demand" erstellt der Tag-Switch für jede FIB-Route eine TDP-Anfrage, welche jeweils an den "Next Hop Switch" auf der Route gesendet wird. Hierbei wird für die Route das Tag-Binding erfragt. Der angefragte Switch sendet anschließend seinen Tag für die Route zurück. Sollte der Switch noch keinen TIB-Eintrag für die angefragte Route besitzen so erzeugt er einen neuen Eintrag wobei er ein Tag generiert und das Incoming-Tag-Feld auf das gerade geschaffene Tag setzt. Diese TIB-Information, also die Bindung zwischen Route und Incoming Tag, wird dann als Antwort auf die entsprechende TDP-Anfrage zurückgeschickt. Der anfragende Switch setzt nun in seinem TIB-Eintrag, für die entsprechende Route, das empfangene Label in das Outgoing-Tag-Feld ein. Sollte noch kein Eintrag existieren so erstellt der Switch einen neuen Eintrag mit generiertem Incoming-Tag und dem empfangenen Tag als Outgoing-Tag.

Bei "Upstream-Tag-Allocation" erstellt der Tag-Switch für jede Route in seiner FIB, ein Tag und einen TIB-Eintrag, wobei das Outgoing-Tag auf das neue Tag gesetzt wird. Diese Bindung zwischen Route und Outgoing-Tag wird über TDP an den nächsten Hop gesendet. Dieser wiederum plaziert das Tag im Incoming-Tag-Feld seines TIB-Eintrages für die Route. Sollte ein solcher Eintrag noch nicht existieren, so wird er generiert.

Bei allen drei Zuweisungsverfahren ist der Tag-Switch dann in der Lage die Pakete schnell weiterzuleiten, wenn sowohl Incoming- als auch Outgoing-Tag eines TIB-Eintrags festgelegt wurden.

Tag-Switching ist vom Prinzip her eng verwandt mit dem Forwarding-Mechanismus von ATM-Zellen. Beide Verfahren treffen die Entscheidung über den einzuschlagenden Weg anhand eines kleinen, fest definierten Labels. ATM benutzt hierzu das VPI/VCI-Paar (Virtual Path Identifier / Virtual Channel Identifier), welches die Verbindung eindeutig festlegt. Die Tag-Switching Technologie lässt sich somit leicht auf einen ATM-Switch übertragen. Die Tag-Information kann im VCI-Feld des ATM-Headers mitgeführt werden. Im ATM-Switch müssen Layer-3-Routing-Protokolle implementiert sein, da sich der Tag-Switch die Information über seine Routing-Nachbarn mit gängigen Routing-Protokollen wie OSPF und BGP beschafft. Hinzu kommt die angesprochene Kontrollkomponente, die auch in den ATM-Switch integriert sein muß. Aus der Sicht eines benachbarten Routers verhält sich ein Tag-Switch wie ein gängiger Router. Diese Tatsache ist beabsichtigt. Bei der Migration zu einer neuen Technologie ist es eine Bedingung, daß sie in die alte Technologie aus Sicherheitsgründen integrierbar ist. Somit kann man bei Ausfällen auf die alte Techik zurückgreifen. Die vollständige Migration findet erst nach voller Funktionsfähigkeit des neuen Systems statt.

Auch Tag-Switching unterstützt das QoS-Prinzip. Die Pakete werden in verschiedene Klassen eingeteilt. Die Klassifikation geschieht anhand der Header der Daten auf der Netzwerkschicht oder auch auf höheren Schichten. Für einzelne Klassen werden dann spezielle Tags vergeben. Anhand dieses Tags und den dazugehörigen, vorher festgelegten, Richtlinien ist der Tag-Switch in der Lage die Pakete, entsprechend den Anforderungen, weiterzuleiten. Beispielsweise kann per RSVP eine bestimmte Dienstgüte angefordert werden. Für diese Anforderung wird ein Tag generiert, welches repräsentativ für diese RSVP-Session steht.

5.4.3. Net-Flow-Feature-Technik

Cisco setzt mit seiner proprietären Net-Flow-Feature-Technik (NFFT) bereits ein Verfahren ein, um die hohen Anforderungen an multimedia- und sicherheitsorientierte Netze zu bewältigen. Das Prinzip besitzt gewisse Ähnlichkeiten mit dem Ansatz von Ipsilon. Wie der Name schon sagt, wird hierbei auch eine Flußklassifizierung durchgeführt. Ein Fluß ist, wie schon erwähnt, eine unidirektionale Sequenz von Paketen zwischen gegebenen Quell- und Zielendpunkten. Die charakteristischen Eigenschaften eines speziellen Flusses können sehr detailliert und anhand mehrerer verschiedener Kriterien, welche bereits im Kapitel "IP-Switching" erwähnt wurden, festgelegt werden. NFFT untersucht das jeweils erste Paket eines Flusses um dann einen Eintrag in den Net-Flow-Cache zu schreiben. Die Klassifizierung ist hierbei von folgenden Kriterien abhängig: Quell- und Ziel-IP-Adressen, auf Transportschicht vergebene Portnummern für Anwendungen, IP-Protokolltyp, Type of Service, Interface-ID für das eingegangene Paket. Nach der Klassifitierung wird ein Job implementiert, welcher für nachfolgende Pakete des entsprechenden Flusses die schnelle Weiterleitung (Switching), die Dienstkontrolle (QoS) und die Datenanalyse (Accounting, Access-Listen, etc.) übernimmt. Dadurch wird der integrierte Router entlastet, damit er für seine eigentlichen Aufgaben, das Routing und das Austauschen von Routinginformationen, genügend Rechenkapazität zur Verfügung hat. Die Net-Flow-Technik bietet neben dem schnellen Forwarding weitere wichtige Funktionen:

- Accounting / Billing: durch die Analyse der Flußdaten (IP-Adressen, Zeitstempel, Type-of-Service-Felder, etc.) können flexible und detaillierte Messungen bezüglich der Beanspruchung von Ressourcen durchgeführt werden. Service Provider können diese Informationen nutzen um ihre bisher statischen Nutzungsgebühren flexibler zu gestalten, z.B. in Abhängigkeit der QoS-Anforderungen, der tatsächlich genutzten Bandbreite oder der Tageszeit. Großunternehmen können durch diesen Mechanismus Nutzungskosten einzelnen Abteilungen und auch Personen zuordnen.
- Netzwerkplanung und Performance-Analyse: wichtige Daten werden an Netzmanagement-Anwendungen weitergegeben. Die Ergebnisse der Datenanalyse liefern die Grundlage für effiziente Netzwerkplanung (Festlegung der Kommunikationspfade zwischen Abteilungen und deren Rechte, Routing-Policies, Durchführung von Hardware- bzw. Softwareupgrades). Die Übertragungsgeschwindigkeiten werden somit, wie auch die Bandbreitenkapazitäten und die Verlässlichkeit des Systems, maximiert. Gleichzeitig werden die Kosten für Netzwerkadministration minimiert.
- Echtzeit Network Monitoring: diese Funktion ist wichtig für die schnelle Fehlererkennung und die anschließende Fehlersuche, welche dann im Idealfall schnell zu einer Problemlösung führt. Die Verkehrsdaten werden für das Netzwerkmanagementsystem entsprechend aufbereitet, so daß sie dann visualisiert werden können. Je nachdem was gerade untersucht wird, können die Daten aus Verkehrssicht oder Anwendungssicht dargestellt werden.
- Application Monitoring and Profiling: dem Netzwerkadministrator wird hierdurch die Möglichkeit gegeben detaillierte, zeitorientierte Informationen über die Nutzung von Anwendungen zu erhalten und auszuwerten.
- User Monitoring and Profiling: die Untersuchungen können auch benutzerorientiert durchgeführt werden. Das Verhalten einzelner Benutzer und ihrer Beanspruchung

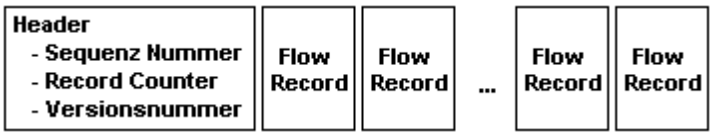
von Netzwerkressourcen kann detailliert protokolliert werden. Die Auswertung führt dann zur effizienten Planung und Zuweisung von Benutzerrechten. Bandbreiten- und Anwendungsressourcen können flexibel vergeben werden. Die potentielle Verletzung von Sicherheits- und Verhaltensrichtlinien (security, policies) kann früh erkannt und behoben werden. Die Anwendung des User Monitoring and Profiling ist allerdings, aus Sicht des Datenschutzes, ein heikles Thema.

- Data Warehouse and Mining: dem Administrator wird ermöglicht eine Datenbank zu organisieren, welche dann entsprechende Einträge über einen längeren Zeitraum beinhaltet. Hauptsächlich geht es darum festzustellen, **wer**, **wann**, **was**, **wo** und **wie lange** gemacht hat.

Die Net-Flow-Technik bewirkt die Performancesteigerung durch intelligentes Flow-Cache-Management. Vor allem bei hoher Auslastung der Router durch Verarbeitung von vielen parallelen Flüssen kurzer Dauer schafft die Net-Flow-Technik deutliche Geschwindigkeitssteigerungen. Die Flow-Cache-Management-Software beinhaltet mehrere intelligente anspruchsvolle Algorithmen, welche für folgende Punkte zuständig sind:

- Feststellung, ob ein eingehendes Paket einer bereits vorhandenen Flußdefinition zugeordnet werden kann.
- Erstellung von neuen Flow-Cache-Einträgen für nicht bekannte Flüsse
- Dynamische Aktualisierung von Flow-Cache-Einträgen
- Löschen von alten Flow-Cache-Einträgen bei Überschreitung eines definierten Zeitlimits (Vorgabewert: 30 Minuten). Alternativ werden Einträge auch gelöscht, wenn die Speicherkapazität des Flow-Cache überschritten wurde oder eine gewisse Zeit lang keine Pakete eines speziellen Flusses mehr weitergeleitet wurden.

Die einzelnen Komponenten kommunizieren, wie bei Ipsilons IP-Switching, über Flow-Datagramme, welche im Sekundentakt versendet werden (siehe Bild).



Flow-Datagramm Struktur ([43], S. 4)

Die einzelnen Flow-Records haben folgende Struktur:

Source IP address
Destination IP address
Source TCP/UDP application port
Destination TCP/UDP application port
Next Hop Router IP address
Input Physical Interface Index
Output Physical Interface Index
Packet count for this flow
Byte count for this flow
Start of flow timestamp
End of flow timestamp
IP Protocol (z.B. TCP=6, UDP=16)
Type of Service (ToS) byte
TCP Flags
Source AS Number
Destination AS Number
Source subnet mask
Destination subnet mask
Flags
Shortcut router IP address

Flow Record Struktur

Gelöschte Einträge werden nicht tatsächlich gelöscht. Sie werden aus dem lokalen Flow-Cache entfernt und als sogenannter "Net-Flow-Export" (auch ein Net-Flow Datagramm) an eine zentrale Einheit, den Net-Flow-Collector geschickt. Diese Einheit speichert die Daten in seine Datenbank und hält sie abrufbereit.

Die Net-Flow-Feature-Technik bietet sich, durch die erwähnten Möglichkeiten der Datenanalyse, als Instanz (z.B. von RSVP) zur Realisierung von QoS an. QoS-Einheiten müssen allerdings gesondert im entsprechenden Switch etabliert sein.

5.4.4. Multi-Protocol over ATM (MPOA)

MPOA (Multiprotocol over ATM) ist eine Spezifikation des ATM-Forums, welche auch Konzepte der IETF beinhaltet. Das Ziel des Gremiums war die Schaffung eines Standards für die Integration von ATM in bestehende Multiprotokollnetze. Die Eigenschaften von MPOA sind unter anderem das effiziente Routing von Ebene-3-Protokollen wie IP, die Ergänzung und Zusammenführung bestehender Standards sowie die Bereitstellung von Dienstegarantien (QoS) für entsprechende Anwendungen. Das Prinzip von MPOA beruht auf einem Client/Server-Modell. Alle Dienste, wie z.B. das Routing, werden zentral verwaltet und den nutzenden Komponenten bereitgestellt. Für die einzelnen Funktionen werden sogenannte Functional Groups (FG) gebildet. Die Clients speichern Serverinformationen in ihren lokalen Cache um die Serverlast zu reduzieren. Clients können in zwei Gruppen eingeteilt werden: ATM-attached Hosts und Edge Devices. ATM-attached Hosts, z.B.

Workstations, sind ausgestattet mit ATM-Karten, in denen die entsprechenden MPOA-Treiber integriert wurden. Sie werden in einzelne "ATM attached Hosts Functional Groups" (AHFG) zusammengefasst. Edge Devices, z.B. Ethernet-Switches mit ATM-Interface und integrierten MPOA-Treibern (LAN-Emulation, etc.). Die Edge Devices werden in den "Edge Device Functional Groups" (EDFG) zusammengefasst.

Für MPOA-Server wurden fünf verschiedene Funktionsgruppen festgelegt:

- Der "Configuration Server" ist für die Anfangskonfiguration der Komponenten zuständig. Er versendet beispielsweise die ATM-Adressen von Servern oder Informationen über die verwendeten Ebene-3-Protokolle und die entsprechenden Paketgrößen (MTU).
- Die "Coordination Functional Group" übernimmt die Verwaltung von Subnetzen.
- Solange ein Client noch keine direkte ATM-Verbindung zu seinem Ziel aufgebaut hat, nutzt er die "Default Forwarder Functional Group". Er schickt seine Daten, z.B. IP-Pakete, zu den Default Forwardern um so die Wartezeit bis zum Aufbau einer direkten ATM-Verbindung zu umgehen. Bei längeren Übertragungen werden dann direkte Verbindungen zum jeweiligen Zielsystem aufgebaut.
- Die "Route Server Functional Group" ist für den Austausch von Routinginformationen zwischen Routeservern verschiedener Subnetze, sowie zwischen Routeservern und herkömmlichen Routern, zuständig.
- Die "Remote Forwarder Functional Group" funktioniert ähnlich wie die "Default Forwarder Functional Group". Sie ist für den Verkehr zwischen den einzelnen Subnetzen zuständig.

MPOA beinhaltet bestehende Spezifikationen des ATM-Forums und der Internet Engineering Task Force (IETF).
Vom ATM-Forum sind dies LAN-Emulation (LANE) und Private Network-to-Network Interface (PNNI). LAN-Emulation ermöglicht, wie bereits beschrieben, die Kopplung zwischen bestehenden Netzen, wie Ethernet, und ATM. PNNI ist zuständig für das Routing auf ATM-Ebene. MPOA erweitert PNNI um die Fähigkeit IP-Routing durchzuführen.
Von der IETF übernimmt MPOA Teile der Spezifikationen von RFC 1483, NHRP (Next Hop Resolution Protocol) , RSVP (Ressource Reservation Protocol) und MARS (Multicast Address Resolution Server). In RFC 1483 ist die Verkapselung von Ebene-3-Protokollen, wie beispielsweise IP, über ATM beschrieben. MPOA verwendet die beschriebene LLC-Verkapselung für den Austausch von Kontroll- und Nutzdaten. NHRP ermöglicht die Benutzung von Abkürzungen (Shortcuts) beim Versenden von IP-Paketen über ATM-, X.25- und Frame-Relay-Netze. Somit können logische Subnetze effizient verbunden werden. RSVP ermöglicht die Realisierung von Dienstegarantien (QoS). MARS regelt das Versenden von IP-Multicast-Paketen über ATM. Hierbei ist eine direkte Zustellung der Pakete vom Sender zum Empfänger, wie auch die zentrale Verwaltung über einen Multicast-Server möglich.

Für die Datenübertragung mit MPOA gibt es zwei Varianten:

- Default Flows erfolgen über ATM-Verbindungen zwischen Clients, Default- und Remote- Forwardern. Diese Verbindungen bleiben ständig aufgebaut und haben den Vorteil, daß Pakete ohne Verzögerungen durch das Netz laufen. Für kurze

Nachrichten wäre der Aufbau einer Verbindung, aufgrund der Verzögerung, ineffizient.
- Shortcuts (Abkürzungen) sind direkte ATM-Verbindungen zwischen Sender und Empfänger. Diese Verbindungen werden nur bei längeren Datenflüssen eingerichtet. Die Route-Server, welche über NHRP kommunizieren (seihe weiter unten), sind für das Routing und die Ermittlung der ATM-Adressen für die Endpunkte zuständig.

MPOA ist nach dem Client/Server-Prinzip entworfen, somit existieren zwei logische Komponenten: MPOA Clients (MPC) und deren MPOA Server (MPS). Sie kommunizieren miteinander über einen Satz von Protokollen. Als Transitdomänen kommen emulierte LANs (ELAN), Logische IP-Subnetze (LIS) oder Lokale Adreßgruppen (LAG) vor. Mittlerweile spricht man bei LIS und LAG von sogenannten ION-Subnets (Internetworking Over NBMA Networks, NBMA = Non Broadcast Multiple Access). MPOA stellt außerdem die Zusammenarbeit mit herkömmlichen Router-basierenden Netzen sicher, da in den MPS auch Routing-Protokolle wie beispielsweise OSPF implementiert sind. Für die Kommunikation in MPOA-Netzen hat das nachfolgend beschriebene "Next Hop Resolution Protocol" einen essentiellen Charakter.

Das Next Hop Resolution Protocol (NHRP), beschrieben in RFC 2332, wurde von der IETF entwickelt und ist bereits Standard. NHRP ist kein Routing-Protokoll sondern ein Address-Resolution-Protokoll, welches allerdings auf die Informationen der Standard-Routing-Protokolle, wie OSPF oder BGP, zurückgreift. NHRP bietet auch die Unterstützung anderer Layer-3-Protokolle als IP. Das folgende Szenario bezieht sich aber auf IP-Routing über ATM-Netze, da dieser Fall die meiste Verbreitung findet. ATM ist ein NBMA (Non-Broadcast, Multiple Access)-Netz. NBMA bedeutet, daß in dem Netz kein Broadcast-Verkehr möglich ist. Ein physisches NBMA-Netzwerk kann in logische Subnetze (Logical IP-Subnet, LIS) unterteilt werden. Die Kommunikation zwischen zwei Stationen, welche in zwei verschiedenen LIS liegen, führt über einen Router, obwohl beide Hosts physisch durch dasselbe Netz verbunden sind. NHRP ermöglicht nun dem Quell-Host die NBMA-Adresse (z.B. MAC-Adresse) des geeigneten "Next Hop" herauszufinden. Ist der Ziel-Host direkt mit dem NBMA-Netz verbunden, so ist der "Next Hop" die Zielstation selbst. Andernfalls liefert NHRP den Router zurück, der am nächsten zur Zielstation liegt.
Das Prinzip von NHRP und die Etablierung von Shortcuts werden mit Hilfe der folgenden Abbildung erklärt.

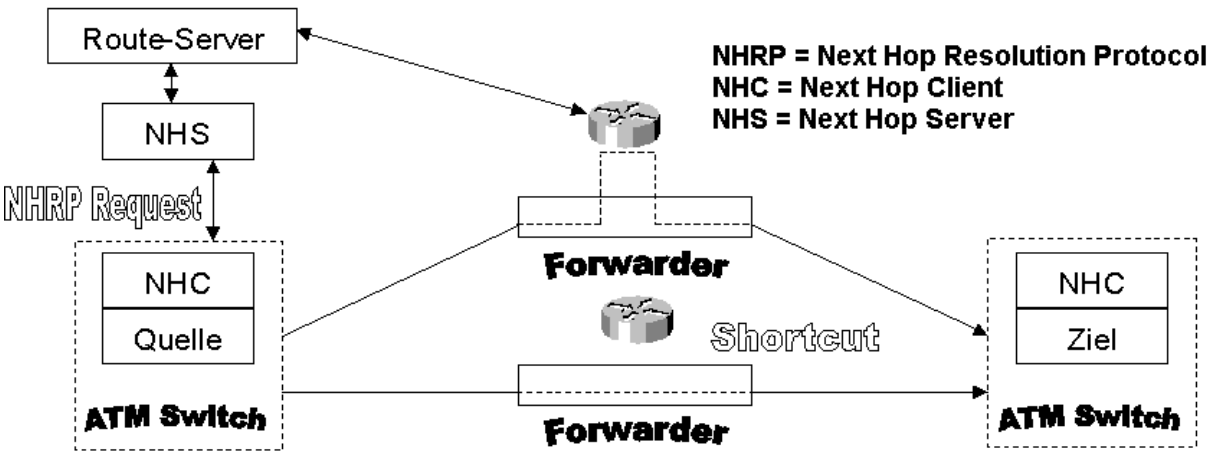


Abbildung: MPOA-Konzept

Der Next Hop Client (NHC) ist in einen ATM-Switch integriert. Er ist mit einem oder mehreren Next Hop Servern (NHS) direkt über ATM verbunden. Die Next Hop Client-Software verfügt über einen Adress-Cache mit IP/ATM-Adresszuordnungen.
Der Next Hop Server (NHS) arbeitet immer zusammen mit einem Router und unterhält ebenfalls einen Adress-Cache mit IP/ATM-Adresszuordnungen, welche er den Clients zur Verfügung stellt. Er kennt die ATM-Adressen der Clients, denen er als Server dient.
Ein IP-Datagramm soll nun von der Quellstation A zu der Zielstation B geschickt werden. Die Quellstation A betrachtet zunächst den NHC-Cache, ob bereits eine Adreßauflösung für die Zielstation B gespeichert ist. Ist dies der Fall, so kann bereits an dieser Stelle eine direkte ATM-Verbindung (Switched Virtual Circuit, SVC) aufgebaut werden. Sollte sich kein Eintrag im NHC-Cache befinden, so schickt die Quellstation einen NHRP Resolution Request an die, ihr bekannten, Next Hop Server. Der NHRP Resolution Request beinhaltet die IP- und ATM-Adresse der Quellstation, sowie die IP-Adresse der Zielstation. Die Anfrage kann als "authoritative" (zuverlässig) gekennzeichnet. In diesem Fall muß ein Next Hop Server antworten, der als Server für die Zielstation festgelegt wurde. Erhält nun ein NHS einen NHRP Resolution Request mit der Kennzeichnung "authoritative", so prüft er, ob er als Server für die Zielstation fungiert. Ist dies nicht der Fall, so leitet er die Anfrage an einen anderen NHS weiter. Ist der NHS als Server für die Zielstation festgelegt, so generiert er anhand der IP/ATM-Adreßzuordnung, welche ihm bekannt ist, einen positiven NHRP Resolution Reply. In diesem Fall wird die Antwort auch mit "authoritative" gekennzeichnet und an die Quellstation gesendet. Sollte die Zielstation nicht im gleichen IP-Subnetz liegen, so wird die IP-Adresse des Routers zurückgeliefert, über den die kürzeste Verbindung zur Zielstation hergestellt werden kann. Diese Entscheidung wird anhand der Routing-Tabelle des Route-Servers erstellt, welche mit Standard-Routingprotokollen wie OSPF erstellt wurde. Ein Transit-NHS (ein NHS, der einen Request nur weiterleitet) nimmt die Informationen eines Reply-Pakets in seinen Cache auf. Somit können nachfolgende Anfragen, welche nicht als "authoritative" gekennzeichnet sind beantwortet werden.
Sollte kein NHS im ATM-Netz auf einen NHRP Resolution Request antworten können, so wird ein negatives NHRP Resolution Reply an die Quellstation zurückgegeben. Dies kann passieren, wenn keine Informationen über die Zielstation (und das zugehörige IP-Subnetz) existieren oder, bei Ausfall einzelner NHS, welche zu dem Zeitpunkt die einzige Antwort auf eine "authoritative"-Anfrage geben könnten. Nach Erhalt einer positiven NHRP Resolution Reply, kann eine direkte ATM-Verbindung als Switched Virtual Circuit (SVC) aufgebaut werden.

Neben den bereits genannten, gibt es noch folgende weitere NHRP-Pakete:

- NHRP Purge Request: wird benötigt um gecachte Informationen auf dem NHS zu löschen
- NHRP Purge Reply: hiermit werden gelöschte Informationen bestätigt
- NHRP Error Indication: der Sender einer NHRP-Nachricht wird informiert, daß ein Fehler aufgetreten ist.

Die Bedeutung von Multicast-Verkehr wird in Zukunft noch weiter stark zunehmen. Multicast ist für viele Multimedia-Anwendungen die ideale Verteiltechnik. Außerdem

werden bisherige Protokolle, welche via Broadcast ihre Daten verteilt haben, zur Reduzierung der Netzlast Multicast-fähig gemacht (z.B. das Network Time Protocol).
Die Unterstützung von IP-Multicast ist also eine wichtige Forderung auch an MPOA, welche durch zwei verschiedene Varianten realisiert wird. Die erste mögliche Variante ist die Realisierung eines Multicast-Servers (MCS) im Default Forwarder, an den alle Sender einer Multicast-Gruppe die Pakete weiterleiten. Der MCS versendet dann die Daten an alle Empänger der Gruppe über eine Punkt-zu-Mehrpunkt-Verbindung. Alternativ kann jeder Sender eine eigene Punkt-zu-Mehrpunkt-Verbindung zu den Empfängern der Multicast-Gruppe aufbauen. Hierbei werden zwar wesentlich mehr Verbindungen geschaltet, die Anfälligkeit des Systems bei Ausfall des MCS und die sehr hohe Bandbreitenbelastung rund um den MCS wird aber umgangen.

6. Einsetzbarkeit der Routing-Mechanismen

6.1. Darstellung der momentanen Topologie

Aus der folgenden Abbildung wird ersichtlich, daß das VTS-Gebäude den zentralen Kommunikationspunkt darstellt, indem alle Gebäude direkt am VTS-Gebäude angeschlossen sind.

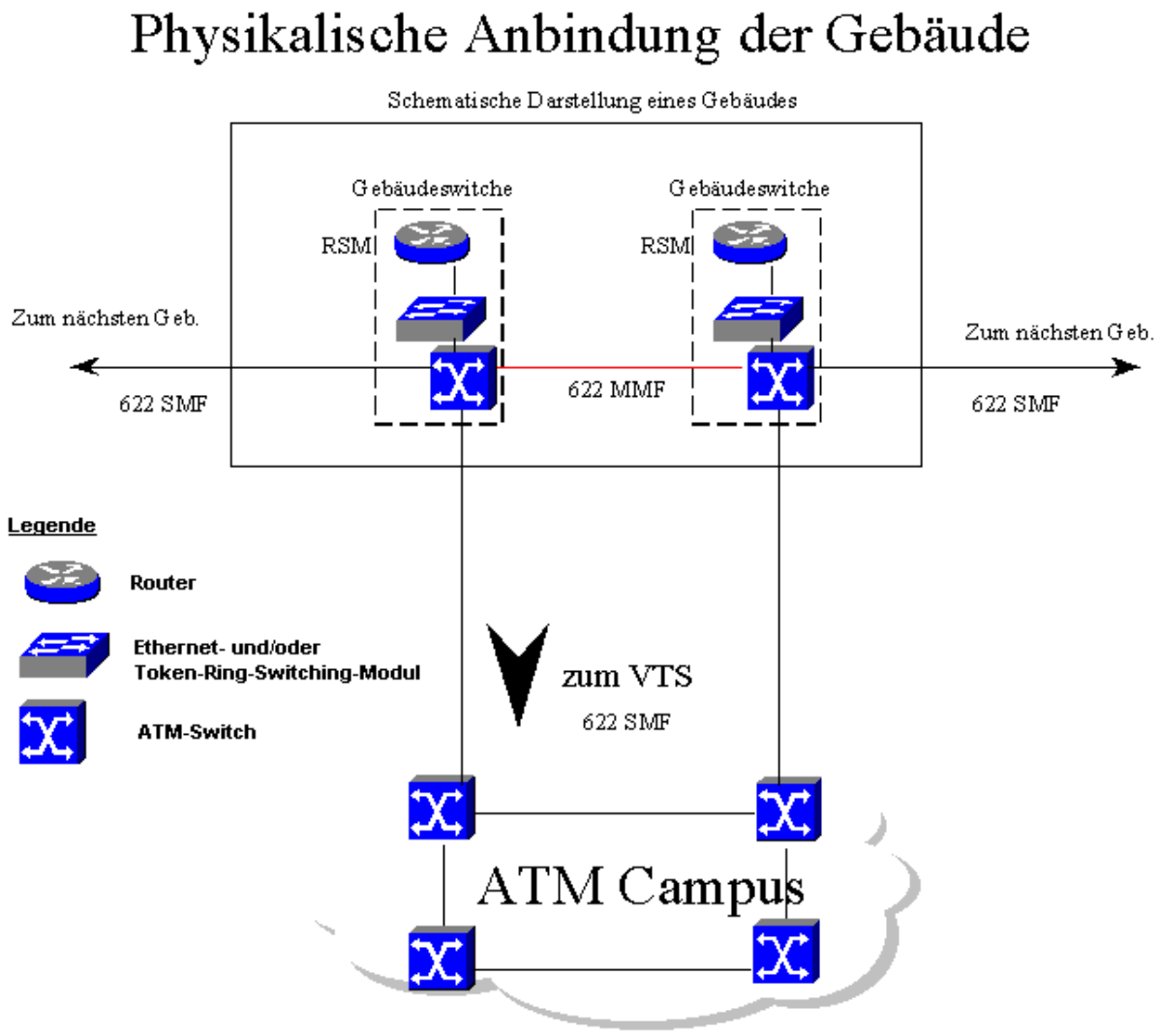


Abbildung: Anbindung der Gebäude (1)

Die Gebäude selbst sind alle identisch geplant: In den unteren Etagen werden zwei Gebäudeswitche mit integriertem ATM-Switch-Prozessor (ASP) installiert. Auf jeder Etage werden Switche mit Ethernet, bzw. Token-Ring Modulen installiert. Diese Etagenswitche besitzen ein LAN-Emulations-Modul mit ATM-Ports über das der Switch redundant mit den Gebäudeswitchen verbunden ist. Die Gebäudeswitche selbst sind über 622 Mbit/s Leitungen mit dem HypoVereinsbank Campus verbunden. Weiterhin bestehen auch ATM-Verbindungen zwischen den Nebenstellengebäuden. Somit kann auch eine Kommunikation bei Ausfall, bzw.

Überlast, des Campus-Backbone zwischen den Gebäuden stattfinden. Intern wird die Vernetzung auch als "Mercedes-Stern" bezeichnet. Bei der folgenden Abbildung fällt vor allem die "2-Hop-Strategie" auf: Die Kommunikation zwischen zwei Gebäuden läuft über maximal zwei Router.

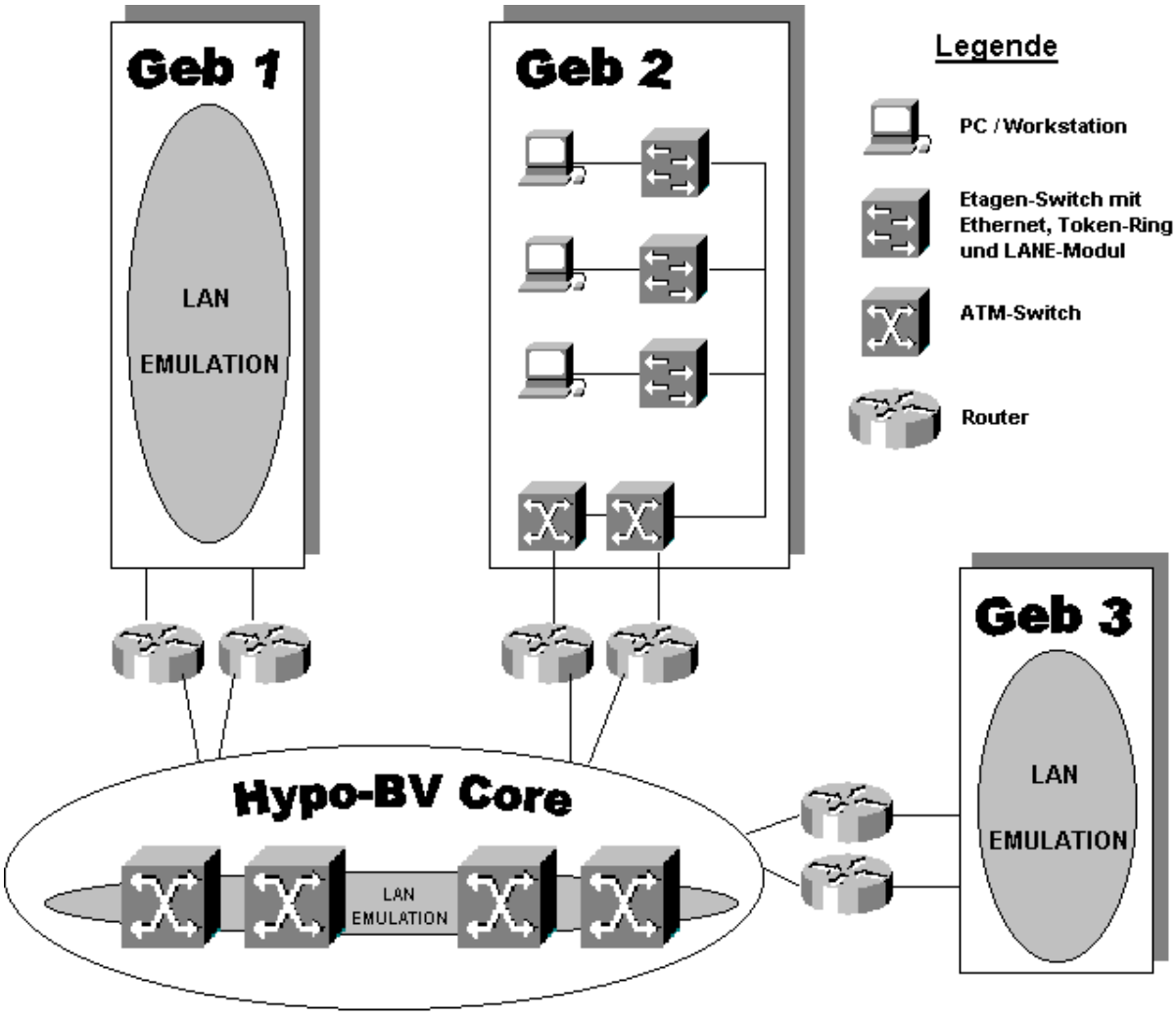
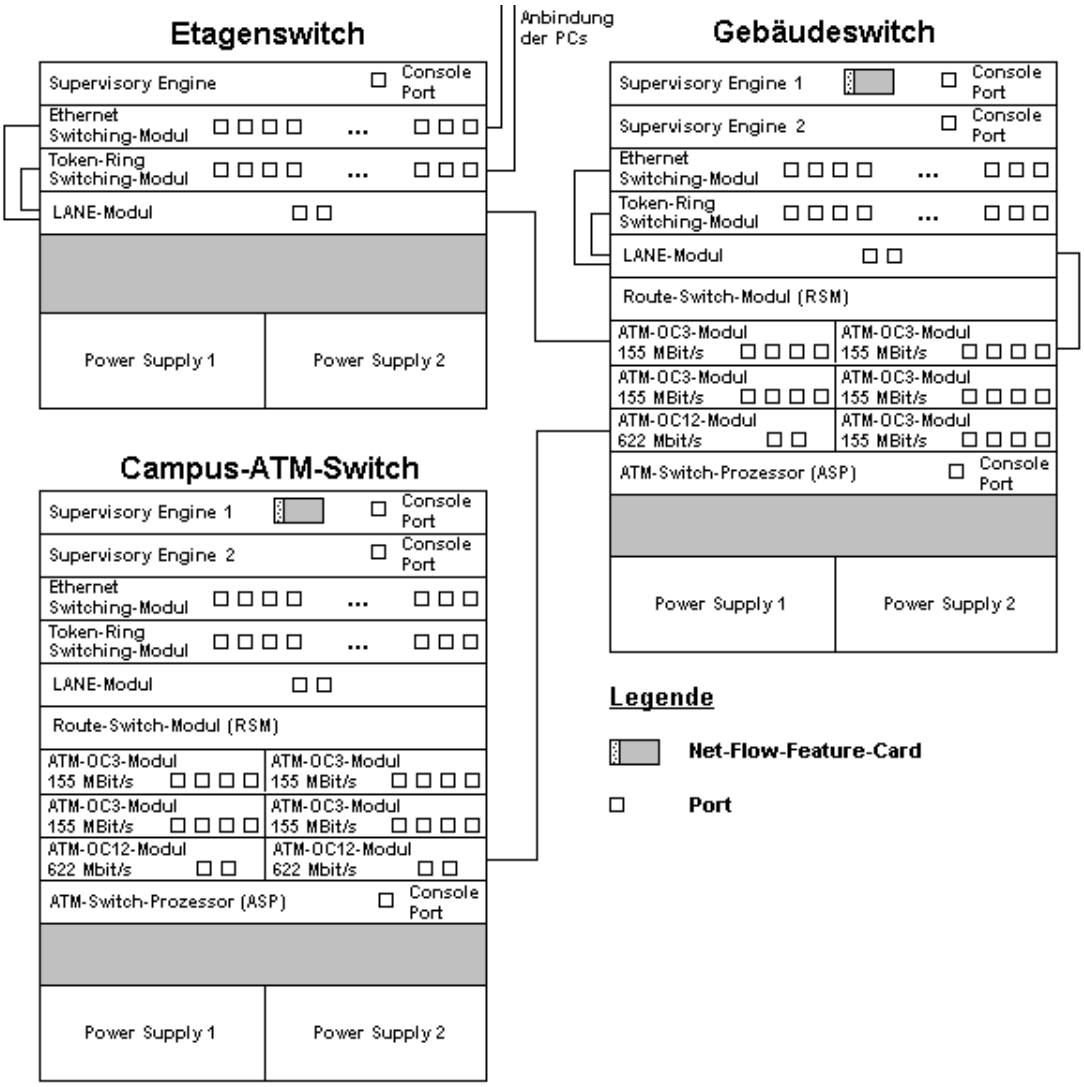


Abbildung: Anbindung der Gebäude (2), 2-Hop-Strategie

Die folgende Abbildung zeigt die verschiedene Bestückung der eingesetzten Cisco Catalyst 5500 Switche.



Etagen-Switch, Gebäude-Switch, Campus-ATM-Switch

Der Catalyst 5500 ist das größte Modell der Cisco-Produktreihe an Switches. Er besitzt 13 Einschubmöglichkeiten für diverse Module wie beispielsweise für Ethernet-, Token-Ring-, Routing-, ATM-Switch-Prozessor- oder LANE-Module. Diese Module werden alle über einen internen Bus verbunden, die sogenannte "Backplane". Die Steuerung der internen Kommunikation, sowie die Managebarkeit der Komponenten über das Netzwerk oder einen integrierten "Console Port" wird durch die sogenannte "Supervisory Engine", einem Einschubmodul, realisiert. Die "Supervisory Engine" muß vorhanden und in Einschubfach 1 eingelegt sein. Der Etagenswitch besitzt hauptsächlich Ethernet-Switching-Module, an deren Ports die PCs der Etage direkt angeschlossen sind. Neben Ethernet-Modulen werden auch Token-Ring-Switching-Module eingesetzt. Allerdings wird Token-Ring aus Kosten-, Geschwindigkeits- und Managementgründen wo es geht durch Ethernet ersetzt. Die Ethernet- und Token-Ring-Module sind direkt über den internen Bus mit dem LAN-Emulationsmodul verbunden. Dieses ist wiederum redundant über zwei Verbindungen an jeweils einen der ATM-Ports (155 Mbit/s) der beiden Gebäudeswitche angeschlossen. Zum Zeitpunkt der Erstellung dieser Diplomarbeit wurden für den Einsatz der LAN-Emulation für Ethernet und Token-Ring noch zwei LANE-Module installiert. Diese Maßnahme war nötig, da die LAN-Emulation nicht stabil lief, solange Ethernet und Token-Ring parallel emuliert wurden. Der Gebäudeswitch besitzt auch Ethernet und

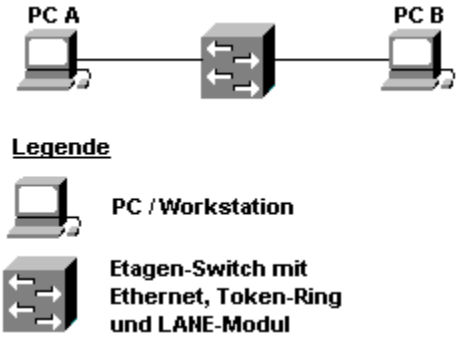
Token-Ring-Module, sowie ein LANE-Modul, da er parallel zu seiner Funktion auch die Funktion eines Etagenswitches übernimmt. Der ATM-Switch-Prozessor (ASP) ist auf dem Einschubmodul 13 integriert und für die Abwicklung aller ATM-Dienste zuständig. Da die einzelnen Gebäude in verschiedenen IP-Subnetzen liegen, muß in den Switch eine IP-Routing-Einheit integriert sein. Diese Aufgabe übernimmt das Route-Switch-Modul (RSM, detailliertere Beschreibung in Kapitel 7.3.). Alle Gebäudeswitche sind redundant über jeweils zwei 622 Mbit/s-Leitungen mit den ATM-Campus-Switchen verbunden.

6.2. Beispiele für Kommunikation

In den folgenden Beispielen werden ein paar Beispielszenarien skizziert, um einen Einblick in die Kommunikationsabläufe zu geben. Es versucht jeweils ein PC A mit einem PC B Kontakt aufzunehmen (z.B. könnte PC B ein Web Server sein). PC A

6.2.1. Kommunikation auf Etagenebene

Im ersten Beispiel wird die Kommunikation auf Etagenebene beschrieben. PC A möchte mit PC B Kontakt aufnehmen. Die PCs besitzen eine IP-Adresse im gleichen Subnetz und sind an das Ethernet-Modul des Etagenswitches angebunden. PC A weiß seine eigene IP-Adresse und die eigene MAC-Adresse, wie auch die IP-Adresse von PC B.



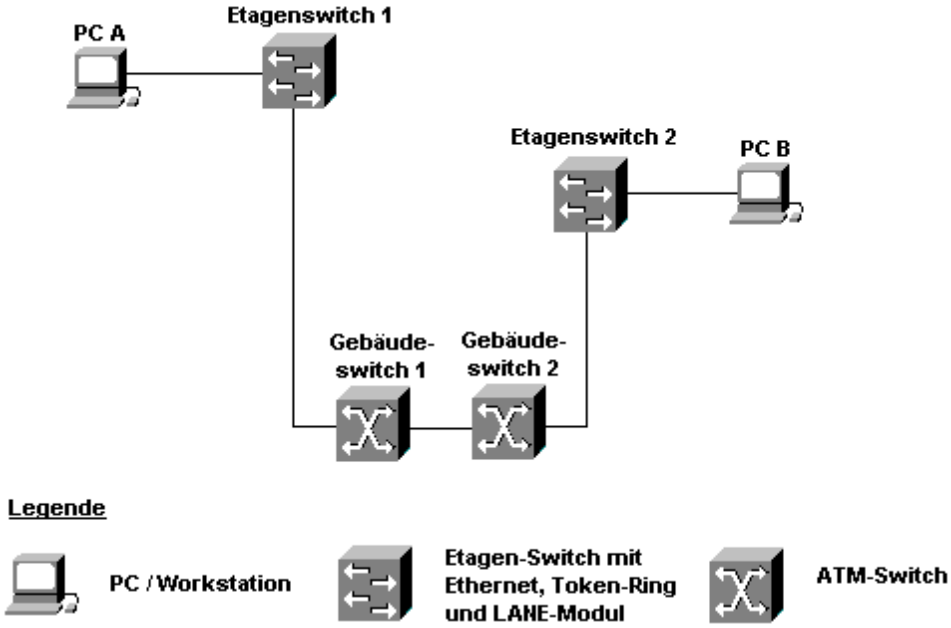
Kommunikation auf Etagenebene

PC A muß, um die Kommunikation beginnen zu können, die MAC-Adresse von PC B's Netzwerkkarte kennen. Wenn PC A bereits vorher schon mal Kontakt zu PC B hatte, so ist es wahrscheinlich, daß PC A die MAC-Adresse von PC B im lokalen ARP-Cache gespeichert hat (ARP = Address Resolution Protocol, dient zur Auflösung von IP- zu MAC-Adressen). Ist dies nicht der Fall, so muß PC A die MAC-Adresse des Kommunikationspartners mittels ARP-Request herausfinden. Der ARP-Request, also eine Anfrage für eine Adressauflösung, beinhaltet die IP- und MAC-Adresse von PC A, sowie die IP-Adresse des gewünschten Kommunikations-partners. Die Nachricht wird per Broadcast verschickt. PC B erhält somit neben allen anderen PCs auch die Anfrage, erkennt die eigene IP-Adresse und schickt an die angegebene MAC-Adresse seine eigene MAC-Adresse als ARP-Reply. PC A ist nun im Besitz der benötigten Informationen und verschickt die Daten, welche mit der MAC-Adresse von PC B versehen sind, über sein Ethernet-Interface (die Netzwerkkarte), welches direkt an den Etagenswitch angeschlossen ist. Der Etagenswitch kennt die MAC-Adresse von PC B und leitet die Daten direkt an den Port, an den PC B angeschlossen ist. Sollte der Etagenswitch die MAC-Adresse von

PC B noch nicht gelernt haben, so sendet er die Daten über alle Ethernet-Ports. Die Empfänger-PCs entscheiden anhand der Ziel-MAC-Adresse, ob die Daten für sie bestimmt sind, oder nicht.

6.2.2. Kommunikation im Gebäude

Im zweiten Beispiel befinden sich die zwei PCs immer noch im gleichen Subnetz, allerdings in zwei verschiedenen Etagen. Die Kommunikation muß also mindestens über einen der Gebäudeswitche laufen, da die Etagenswitche nicht untereinander, sondern über das LAN-Emulationsmodul an die Gebäudeswitche gekoppelt sind. PC A ist wiederum die eigene IP- und MAC-Adresse, sowie die IP-Adresse von PC B bekannt. Vereinfachend wird bei diesem Beispiel angenommen, daß auch die MAC-Adresse von PC B bekannt ist. Sollte sie nicht bekannt sein, so führt PC A wiederum einen ARP-Request aus, welcher dann durch den Broadcast-and-Unknown-Server (BUS) der LAN-Emulation realisiert wird.



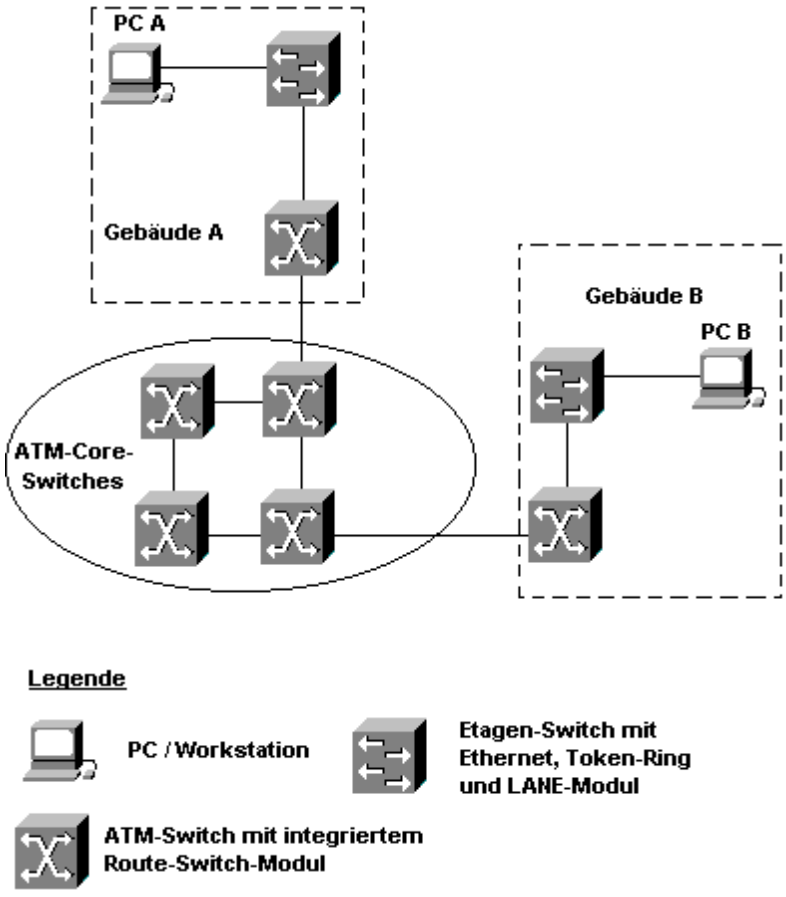
Kommunikation im Gebäude

PC A beginnt also mit der Datenübertragung (MAC-Adresse von PC B ist bekannt und als Zieladresse angegeben). Beim Eintreffen der Daten beim Etagenswitch prüft der dort installierte LANE-Client (LEC), ob er für die Ziel-MAC-Adresse einen lokalen Eintrag im Cache mit korrespondierender ATM-Adresse findet (logische ATM-Adressen sind von der Form 00.3217.1346.2435.2436.2524.1454.1352.2422.5452.23). Ist dies nicht der Fall, so fragt er die entsprechend zugehörige ATM-Adresse vom LANE-Server (LES) ab. An dieser Stelle wird angenommen, daß der LES die MAC-Adresse kennt, ansonsten müßte wiederum der BUS die Adresszuordnung klären. Der LEC ist nun über die korrespondierende ATM-Adresse (in diesem Fall die ATM-Adresse des LANE-Modules von Etagenswitch 2) informiert. Im nächsten Schritt initiiert der LEC von Etagenswitch 1 einen Verbindungsaufbau zu der entsprechenden Einheit. Gebäudeswitch 1 kennt die ATM-Adresse des Empfänger-Moduls anhand seiner Routing-Tabellen und initiiert die Signalisierung. Nachdem ein Pfad geschaltet wurde, werden die Daten an das LANE-Modul von Etagenswitch 2 gesendet,

welches die Daten, entkapselt, an das Ethernet-Modul und somit an PC B weitergibt.

6.2.3. Kommunikation zwischen Gebäuden

Im letzten Beispiel befinden sich die zwei PCs in verschiedenen Subnetzen und auch in verschiedenen Gebäuden. Da sich die Kommunikationspartner in zwei verschiedenen IP-Netzen befinden muß zur Datenkommunikation mindestens eine Routing-Einheit zwischengeschaltet sein. Diese Routing-Einheit ist in Form des Route-Switch-Moduls in den Gebäudeswitch integriert. Bei der Kommunikation zwischen verschiedenen IP-Netzen ist der sendenden Einheit nicht mehr die MAC-Adresse des Empfängers, sondern nur die MAC-Adresse des Routers, der die Daten weiterleitet, bekannt. Dieser Vermittlungsrouter verwirft, bei Empfang der Daten, die Empfangs-MAC-Adresse (seine eigene), untersucht den IP-Header, trifft eine Routing-Entscheidung über welches Interface er die Daten weiter schickt, und versieht dann das Paket mit einer neuen Empfänger-MAC-Adresse, entweder der des tatsächlichen Empfängers oder der einer weiteren Vermittlungseinheit.



Kommunikation zwischen Gebäuden

PC A möchte nun mit PC B kommunizieren und weiß bereits die MAC-Adresse des, in den Gebäude-ATM-Switch integrierten, Routers. PC A schickt also die Daten, welche die IP-Adresse von PC B und die MAC-Adresse des Routers im Gebäude-Switch enthalten, über die eigene Netzwerkkarte an den Etagenswitch. Beim Eintreffen der Daten überprüft der dort installierte LANE-Client (LEC) wiederum den lokalen Cache auf eine entsprechende ATM-Adresse (in diesem Fall für das LANE-Modul des Gebäude-Switches). Ist keine lokale Zuordnung gespeichert, so fragt er

wieder die entsprechend zugehörige ATM-Adresse vom LANE-Server (LES) ab. Der LEC kennt nun die ATM-Adresse des LANE-Modules des Gebäude-Switches. Im nächsten Schritt initiiert der LEC Verbindungsaufbau, welcher wie in Beispiel 2 vom Gebäudeswitch realisiert wird. Das LANE-Modul des Gebäudeswitches erhält die Daten und leitet sie an den internen Router weiter. Dieser untersucht die IP-Pakete nach dem Zielnetz und stellt fest, daß er dieses über den integrierten Router in Gebäude B erreicht, dessen IP- und MAC-Adresse er im lokalen ARP-Cache vorliegen hat (ansonsten müsste der Router wiederum einen ARP-Request verschicken). Die Daten versieht der Router nun mit der neuen Ziel-MAC-Adresse des integrierten Routers im Gebäudeswitch von Gebäude B. Der LEC überprüft wiederum die Zugehörigkeit der MAC-Adresse zu einer ATM-Adresse und initiiert den Verbindungsaufbau zum Gebäudeswitch des Gebäudes B. Der Router im Gebäudeswitch erhält die Daten vom LANE-Modul, erkennt das Gebäudenetz als Zielnetz und tauscht die Ziel-MAC-Adresse der Daten gegen die MAC-Adresse von PC B. Anschließend werden die Pakete an das LANE-Modul des Etagenswitches geschickt, an dem PC B angeschlossen ist. Das LANE-Modul entkapselt die Daten und leitet sie über das Ethernet-Modul an den PC B.

Der beschriebene Ablauf ist Teil der 2-Hop-Strategie: jeder Punkt des Netzes muß über maximal 2 Router erreichbar sein. Verbindungsorientierte Pfade sind jeweils vom LANE-Modul des jeweiligen Etagenswitches zum entsprechenden Gebäudeswitch, und zwischen den Gebäuden geschaltet. An dieser Stelle setzt die schon beschriebene Net-Flow-Feature-Technik an. Die Gebäudeswitch sind alle mit der Net-Flow-Feature-Card (NFFC) ausgerüstet, womit sie in der Lage sind eine Flußklassifizierung durchzuführen. Das Schalten des Shortcuts wird, wie beschrieben, immer vom letzten Knoten initiiert, also in diesem Fall dem Gebäudeswitch des Gebäudes B. Für den Fluß wird ein Label generiert, welches der Net-Flow-Einheit von Gebäudeswitch in Gebäude A mitgeteilt wird. Somit können die beiden Gebäudeswitches einen Shortcut schalten und die beiden Etagenswitches sind über einen direkten Pfad verbunden.

6.3. Bewertung der traditionellen Routing-Mechanismen

6.3.1. Gegenüberstellung RIP - OSPF

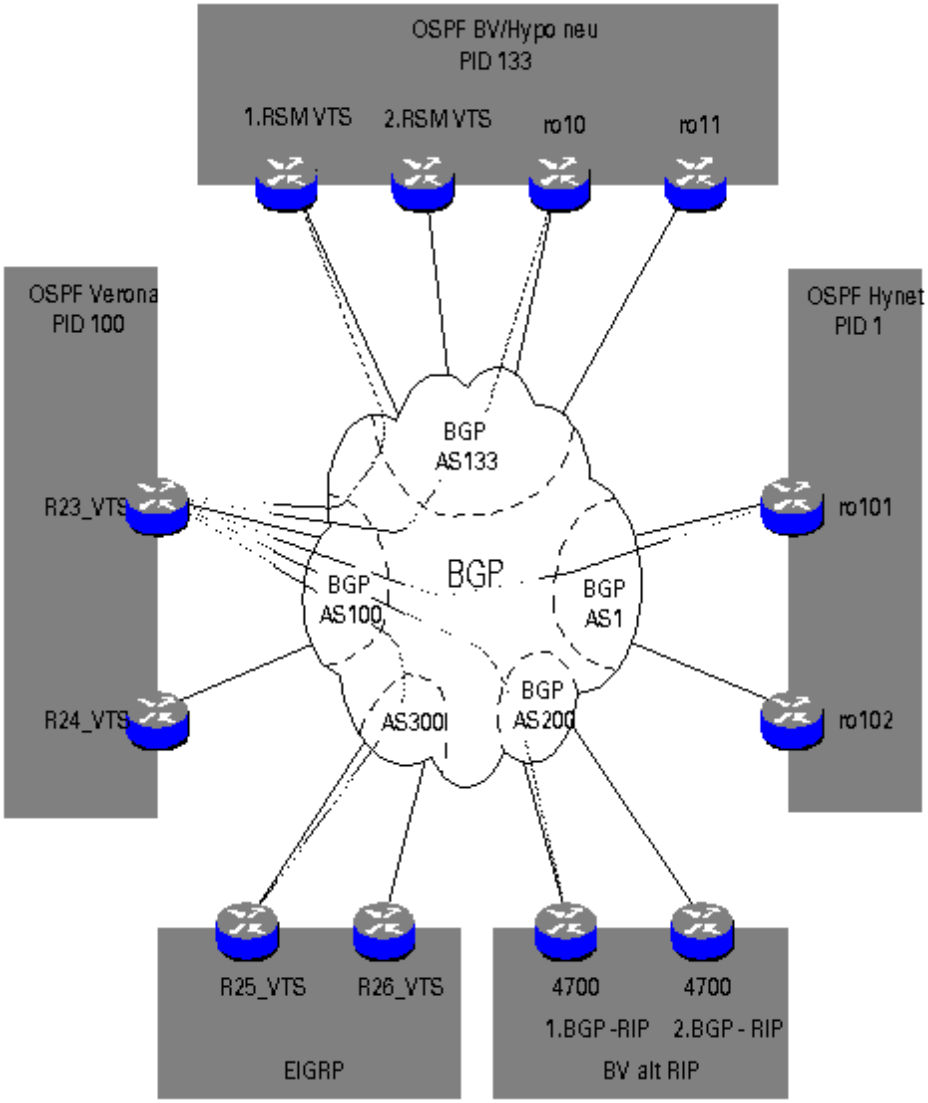
Die beiden beschriebenen Routing-Verfahren können nochmals grob skizziert wie folgt dargestellt werden. Angenommen jemand möchte mit seinem Auto von München nach Frankfurt fahren und weiss die Route nicht. Nun hat er zwei Möglichkeiten. Er könnte einfach losfahren und an den entsprechenden Autobahnkreuzen der Beschilderung folgen. Dieses Weiterleitungs-Verfahren entspricht dem bei RIP. Alternativ könnte er vor der Fahrt beim ADAC anrufen und sich die, zu diesem Zeitpunkt, beste Route durchgeben lassen. Die Route wird anhand des Verkehrsaufkommens (Kosten) bestimmt. Beim Erreichen des jeweils nächsten Autobahnkreuzes ruft der Fahrer wieder beim ADAC an um sich erneut die beste Route, abhängig vom Verkehr und seinem Standpunkt, durchsagen zu lassen. Dieses Weiterleitungs-Verfahren entspricht dem bei OSPF.

OSPF ist im Vergleich zu RIP wesentlich komplexer und umfangreicher. Solange das Netzwerk nicht zu groß wird, ist RIP durchaus effektiv einsetzbar. Der limitierende Faktor ist hierbei der Wert der "Unendlichkeit", welcher auf 16 festgelegt ist. Die Zahl

der hops, die nötig sind um ein Teilnetz über veschiedene Routerverbindungen zu erreichen, darf also nicht größer als 15 sein. Ein weiterer Nachteil von RIP ist die hohe Reaktionszeit (3 Minuten) auf Netzänderungen bei Ausfall eines Routers. Bei OSPF gibt es keine "counting to infinity"-Probleme und somit auch keine maximale Netzgröße. Der erhöhte Verkehr im Netz, welcher durch OSPF-Pakete entsteht, kann durch den beschriebenen Einsatz von Hierarchiestufen reguliert werden. Zusammenfassend kann gesagt werden, daß RIP in kleineren Netzen, die nicht auf die Vorteile von OSPF zurückgreifen müssen, effizient eingesetzt werden kann. Ab einer gewissen Größe empfiehlt sich dann der Gebrauch von OSPF.

6.3.2. Border Gateway Protokoll (BGP)

Das, für den Einsatz in den Routing-Einheiten der Gebäudeswitche vorgesehene, Routing-Protokoll war und ist OSPF. Allerdings gab es zum Zeitpunkt der Erstellung der Diplomarbeit Stabilitätsprobleme bei den ersten Einsatzversuchen. Diese Probleme sind darauf zurückzuführen, daß in den einzelnen Gebäuden teilweise noch alte Netze existieren, welche in das Gesamtnetz eingebunden sind und in denen noch Routing-Protokolle wie RIP oder EIGRP eingesetzt werden. OSPF hat somit die Aufgabe die lokalen Routinginformationen, aus den Gebäuden, für die anderen Netzteilnehmer sichtbar zu machen. Hierfür gibt es Funktionen zur Zusammenführung, bzw. zum Austausch der Routing-Tabellen innerhalb eines Gebäude-Routers (route consolidation). Diese Funktion war zwar implementiert, es kam aber, wie gesagt, zu häufigen Abstürzen des Systems, da die integrierte Software fehlerhaft, bzw. nicht stabil lief. Als Aushilfe wurde auf BGP zurückgegriffen, das seine Stabilität im Internet bewiesen hat. Die alten Netze wurden in einzelne autonome Systeme aufgeteilt (siehe Abbildung) und die Kommunikation läuft seit dem Einsatz von BGP stabil.



BGP-Anbindung einiger Gebäude ([45])

6.4. Bewertung von Layer-3-Switching

Die Funktionsweise von Layer-3-Switching wurde bereits erläutert. Die erhöhte Durchsatzgeschwindigkeit wird im HypoVereinsbank-Netz bereits durch den Einsatz der Route-Switch-Module genutzt und geschätzt.

6.5. Bewertung der Shortcut-Mechanismen

IP-Switching von Ipsilon Networks und Tag-Switching von Cisco Systems waren zum Zeitpunkt der Erstellung dieser Arbeit noch kein Internet-Standard. IP-Switching und die Net-Flow-Feature-Technik eignen sich gut für einen Einsatz in Campus-Netzen, da hier die Anzahl der Vermittlungseinheiten überschaubar ist. Tag-Switching eignet sich eher für den Einsatz in Weitverkehrsnetzen und im Internet. Alle drei Ansätze verbindet die Einfachheit der Protokolle. Somit wird die schnelle Verarbeitung der Daten gewährleistet. Hinzu kommen die geringen Kosten für die eingesetzte Hardware. Bestehende Netze können leicht migriert werden.
Der Vorschlag des ATM-Forums, MPOA, läßt viel Spielraum bei der tatsächlichen Umsetzung in Produkte. MPOA kann somit an die jeweiligen Anforderungen gut

angepasst werden. Damit geht man dann allerdings die Gefahr ein, daß verschiedene Implementierungen nicht miteinander operieren. Die Komplexität eines MPOA-Netzes ist sehr hoch, da versucht wird, möglichst viele Techniken miteinander zu vereinen. Hierbei stellt sich die Frage der Stabilität und der Interoperabilität dieser Netze. Aufgrund der genannten Probleme ist man sich heute schon sicher, daß sich MPOA in der Praxis in der beschriebenen Form nicht gegen die Konkurrenz-Techniken wird durchsetzen können. Neben den beschriebenen Neuentwicklungen gibt es natürlich Weiterentwicklungen, von bestehenden Techniken, wie Gigabit-Ethernet oder Gigabit-Routing, welche bereits eingesetzt werden, und somit auch den Einsatz von MPOA unwahrscheinlich machen.

6.5.1. IP-Switching und NFF-Technik

Die beiden Techniken IP-Switching und die NFF-Technik sind vor allem für Netze mittlerer Größe geeignet, also firmeninterne LANs und ATM-Campusnetze. Dies hängt damit zusammen, daß die Technik in beiden Fällen auf der Analyse von individuellen Datenströmen beruht. Es müssen also im Switch genügend Ressourcen zur Verfügung stehen, damit jedem Datenstrom ein eigenes Label, bei ATM ein VPI/VCI-Paar, zugewiesen werden kann. Als allgemeiner Vorteil des Shortcut-Routings kann die deutlich höhere Performance (höhere Geschwindigkeit beim Weiterleiten der Daten), durch das Ersetzen der langsamen und teuren Weiterleitung über die Netzwerkschicht durch schnelles und preiswertes Label-Swapping (Shortcut-Routing), genannt werden.
Obwohl die beiden Techniken bereits in der Praxis erfolgreich eingesetzt werden, gibt es doch ein paar Nachteile: Da die ersten Pakete noch über langsames Store-and-Forward ihr Ziel erreichen überholen die nachfolgenden Pakete, welche geswitcht werden, ihre Vorgänger. Dieses Problem tritt aber bereits bei Netzen auf, in denen standardmäßig geroutet wird und wird durch Sequenznummernüberprüfung gelöst.
Kritischer ist der Fall, wenn die ersten Switche auf dem Pfad zum Ziel bereits einen Shortcut geschaltet haben, somit ein Label generiert haben, und die nachfolgenden noch nicht. Erhält nämlich ein Switch ein Paket mit einem Label, das er nicht kennt so verwirft er das Paket. Es muß in diesem Fall neu angefordert werden. Die Implementierungen berücksichtigen diesen Fall bereits. Shortcut-Pfade werden grundsätzlich rückwärts vom Ziel zur Quelle aufgebaut. Der Prozess der Shortcut-Schaltung beginnt also mit dem Eintreffen des ersten Paketes. Natürlich muß auch als allgemeiner Nachteil der erhöhte Overhead durch Label-Vergabe erwähnt werden. Durch die schnelle Weiterleitung nach der Shortcut-Etablierung fällt dieser Faktor allerdings nicht ins Gewicht. Voraussetzung für den Einsatz der Techniken ist, daß alle Komponenten die Techniken integriert haben.

6.5.2. Tag-Switching

Mit Tag-Switching versucht Cisco Systems vor allem die Leistung von Routern in Weitverkehrsnetzen zu steigern. Die Technik soll also hauptsächlich im Internet oder im Backbone von Internet Service Providern (ISPs) eingesetzt werden. Der Grundgedanke ist, den komplexen Vorgang bei der Entscheidung zur Weiterleitung der Pakete zu vereinfachen. Für IP-Zieladressen werden Tags generiert, die allen integrierten Switches weitergegeben werden. Dadurch lassen sich individuelle IP-Datenströme, die ein gemeinsames Ziel haben, anhand der Tags zusammenfassen.

Somit wird eine preisgünstige Möglichkeit geschaffen, Switching und Routing zu kombinieren. Wie bei IP-Switching und der NFF-Technik muß in jedem teilnehmenden Vermittlungssystem eine Komponente integriert sein, die die Vermittlungstechnik behandeln kann. Für den Einsatz in Campusnetzen eignet sich Tag-Switching nicht gut, da die sehr individuellen Dienstgüteanforderungen nicht umgesetzt werden können: Das Tag-Switching-Verfahren ist topologieorientiert. Die Tags werden auf IP-Zielnetz-Basis vergeben und können, wie beschrieben, anhand der Type-of-Service Felder in Klassen eingeteilt werden. Allerdings kann die QoS-Realisierung niemals die feine Granularität einer flußorientierten Technik wie IP-Switching oder der NFF-Technik erreichen.

6.5.3. Multi-Protocol over ATM (MPOA)

MPOA beinhaltet, wie schon beschrieben mehrere Techniken des ATM-Forums und der IETF. Die Vorteile der Technik liegen unter anderem in der effizienten Kommunikation zwischen einzelnen Subnetzen, welche durch die Realisierung von Shortcut-Routen erreicht wird. Das zentrale Routing mit hilfe des virtuellen Routers schafft eine höhere Skalierbarkeit und eine leichte Managebarkeit des Systems (weniger Router die überwacht werden müssen). Hinzu kommt, daß Edge-Devices durch den zentralen Ansatz entlastet werden.
Neben den vielen Vorzügen, die MPOA bietet, gibt es aber auch einige Kritikpunkte Ein genereller Einwand kommt insbesondere von Seiten der IETF und richtet sich gegen den Einsatz von ATM als zugrundeliegende Technik. Die Komplexität der bei ATM eingesetzten Protokolle und die damit verbundenen Implementierungs- und Wartungskosten, sowie die mangelnde Stabilität sind die Hauptkritikpunkte. Stellt man dem MPOA-Modell ein klassisches IP-Netz mit Gigabit-Ethernet im lokalen Bereich und mit Standleitungen im WAN gegenüber, so schneidet Ethernet im Kosten/Leistungs-Vergleich wesentlich günstiger ab. Besonders da Standleitungen und Gigabit-Ethernet auch hier noch Raum für Wachstum bieten. Auch von anderen Seiten wird das MPOA-Modell wegen der bereits angesprochenen Komplexität und geringer Stabilität kritisch beurteilt und wird wahrscheinlich deshalb in der beschriebenen Form keinen Einsatz finden.

7. Schlußwort

Abschließend kann bemerkt werden, daß die "2-Hop-Strategie" in Kombination mit der Net-Flow-Feature-Technik und dem schnellen Routing mit Hilfe des Route-Switch-Moduls eine gut durchdachte Strategie ist, die sich im konkreten Fall in punkto Ausfallsicherheit und Durchsatz sehr gut bewährt hat.

Literaturverzeichnis

[1] "Eine Bank, ein Netz", Christian Lössner, CompuNet, 1998

[2] "Designing High-Performance Campus Intranets with Multilayer Switching", Cisco Systems, 1998

[3] "ATM-Netzwerke", 3. Auflage, Othmar Kyas, Datacom Verlag, 1996

[4] "ATM - Die Technik des B-ISDN", Hauptseminar WS 97/98 am Lehrstuhl für Rechnerkommunikation von Prof. Jessen, TU München

[5] "Kurze Wege durchs Netz - MPOA und IP-Switching", Andreas Gillhuber, CT 11/97 S.334

[6] "MPOA - Funktionsvielfalt in ATM-Netzen", Lothar Zier, iX 9/96 S.126

[7] "IP über ATM mit Classical IP oder LAN-Emulation", Lothar Klein, iX 10/95 S.170

[8] "Gigabit Ethernet versus ATM - Kampf und Koexistenz", Kai-Oliver Detken, iX 7/98 S.101

[9] "Sicherheitsprobleme in IP-über-ATM-Netzen", Iouri Pakhomenko, Eva Pless, iX 3/98 S.118

[10] "Gigabit Ethernet - jetzt wird´s auf der Glasfaser eng", Wolfgang Kemmler, ix 2/98 S.84

[11] "Weiter Weg - ATM Year in San Jose", Dr.Sabine Cianciolo, iX 8/98 S.20

[12] "Multi Protocol over ATM", ATM Forum, 1997

[13] "Multi Protocol over ATM - a brief description", ATM Forum, 1996

[14] "MPOA goes under the microscope at user panel discussion", ATM Forum, 1998

[15] "LAN-Emulation over ATM Version 1.0", ATM Forum, 1995

[16] "Computer Netzwerke", Tanenbaum, Prentice-Hall Verlag, 1997

[17] "Wire-Speed IP-Routing", Extreme Networks, 1998

[18] "Auf Sand gebaut - IP Switching und Tag Switching bald im Abseits ?", Dr.Sabine Cianciolo, CT 1/98 S.27

[19] "Netzdesign", Cisco Systems, 1998

[20] "Network Implementation", Cisco Systems, 1998

[21] "Cisco Campus ATM Solutions", Cisco Systems, 1998

[22] "Introduction to Cisco Router Configuration", Cisco Systems, 1998

[23] "Advanced Cisco Router Configuration", Cisco Systems, 1998

[24] "Multiprotocol Encapsulation over AAL 5", RFC 1483

[25] "Classical IP and ARP over ATM", RFC 1577

[26] "Default MTU use over AAL 5", RFC 1626

[27] "An Experimental Encapsulation of IP Datagrams on Top of ATM", RFC 1926

[28] "IP over ATM: A Framework Document", RFC 1932

[29] "Routende Switches - Multimedia im Netz", Rolf Manfred Schmidt, Networks & Communications 10/97 S.137

[30] "Routing mit Mehrwert: Layer-4-Switching", Mike Lange, Networks & Communications 9/98 S.82

[31] "The OSPF Specification", RFC 1131

[32] "OSPF Version 2", RFC 1247, RFC 1583

[33] "Ipsilon Flow Management Protocol Specification", RFC 1953

[34] "Ipsilon General Switch Management Protocol Specification", RFC 1987

[35] "Cisco Systems Tag Switching Architecture Overview", RFC 2105

[36] "Internet Protocol", RFC 791

[37] "Transmission Control Protocol", RFC 761

[38] "User Datagram Protocol", RFC 768

[39] "NBMA Next Hop Resolution Protocol (NHRP)", RFC 2332

[40] "NHRP Protocol Applicability Statement", RFC 2333

[41] "Layer 3 Switching Demystified", Cisco Systems Whitepaper, März 1999, "http://www.cisco.com/warp/public/cc/sol/mkt/ent/cmps/l3c85_wp.htm"

[42] "Private Network-Network Interface Specification Version 1.0 (PNNI 1.0)", ATM-Forum März 1996

[43] "NetFlow Services and Applications", Cisco System, 1998

[44] "Internet Protokolle", Hauptseminar SS 98 am Lehrstuhl für Rechnerkommunikation von Professor Jessen, TU München

[45] "Einsatz von BGP bei der HypoVereinsbank", Christian Lössner, Roman Svoboda, CompuNet, 1999