

IP-over-ATM: ATM/Hochgeschwindigkeitsnetze in der Internetprotokollumgebung

*Dipl.-Ing. Kai-Oliver Detken
Bremen Institut für Betriebstechnik und angewandte Arbeitswissenschaft (BIBA)
an der Universität Bremen
Hochschulring 20, D-28359 Bremen
Email: det@biba.uni-bremen.de
URL: <http://www.biba.uni-bremen.de>*



*Institute for International Research (IIR)
Konferenzbeitrag
„Strategische Aspekte von Hochgeschwindigkeitsnetzen – ATM und Gigabit-
Ethernet im Vergleich“
Düsseldorf, 7.-9. Oktober 1997*

1 Einleitung	2
2 TCP/IP- und ATM-Eigenschaften im Vergleich	2
3 Anpassungsmöglichkeiten von ATM auf TCP/IP-Netze	3
3.1 Classical-IP (RFC-1577)	3
3.2 LAN-Emulation (LANE)	5
3.3 Multiprotocol over ATM (MPOA)	7
3.3.1 Multicast Address Resolution Server (MARS)	9
3.3.2 Next-Hop Resolution Protocol (NHRP)	9
3.3.3 Resource Reservation Protocol (RSVP)	10
3.3.4 Private Network-to-Network-Interface (P-NNI)	11
3.4 Gegenüberstellung der IP-over-ATM-Verfahren	12
4 IP-Unterstützung von Gigabit-Ethernet und SDH/SONET	13
4.1 Gigabit-Ethernet (IEEE802.3z)	13
4.2 Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH)	14
5 Effektivität der TCP/IP-Protokolle über ATM	15
5.1 Protokoll-Overhead	15
5.2 Durchsatzverminderungen	16
5.3 Tuning der Endsysteme	17
6 Aussichten	18
7 Weiterführende Literaturliste	19

1 Einleitung

TCP/IP-Protokolle nehmen in heterogenen Netzen einen immer höheren Stellenwert ein und sind zu einem Quasi-standard expandiert. Das wachsende Internet ist ein Beispiel für die Verbreitung der TCP/IP-Protokollfamilie. Der Bandbreitenbedarf der Endbenutzer wächst aber zusätzlich mit den Anforderungen an die verwendeten Applikationen. Weiterhin sind Quality-of-Service (QoS) Merkmale notwendig, um bei isochronen Datenströmen die Dienstgüte sicherstellen zu können. Heterogene Netzstrukturen in denen die TCP/IP-Protokollfamilie vornehmlich eingesetzt wird, bieten diese QoS-Unterstützung bislang nicht. Der Asynchrone Transfer Modus (ATM) ist in der Lage diese Engpässe zu beseitigen. Aus diesem Grund hängt der Erfolg von ATM auch von der Anpassung an das Internet Protokoll (IP) ab.

Die Effizienz von TCP/IP-Protokollen wird aber für Hochgeschwindigkeitsnetze angezweifelt, da diese Protokollfamilie eigentlich für geringe Datenraten entwickelt wurde. Das ist auch nicht ganz unberechtigt – vor allem bei Echtzeitanwendungen. Neue Verfahren wie Multiprotokoll über ATM (MPOA) machen es durch die Zusammenarbeit zwischen der IETF und dem ATM-Forum möglich, daß unterschiedlichste Protokollarten aufeinander angepaßt werden können. Durch Integration unterschiedlicher Protokolle (P-NNI, RFC-1483, MARS, NHRP, RSVP) entsteht so ein virtuelles Schicht-3-Routing, das flexibel auf unterschiedliche Netzsituationen reagieren kann.

Dieser Bericht gibt einen Überblick über bestehende Ansätze bzw. Anpassungsmöglichkeiten von ATM auf IP, vergleicht diese Möglichkeiten miteinander und gibt über die Effizienz von IP-Protokollen über Hochgeschwindigkeitsnetze Auskunft. Bei gründlicherer Studie der vorgestellten Verfahren und Vergleiche wird auf das Buch *ATM in TCP/IP-Netzen – Grundlagen und Migration zu High Speed Networks* verwiesen, welches von mir demnächst über den Hüthig-Verlag erscheinen wird.

2 TCP/IP- und ATM-Eigenschaften im Vergleich

Um im lokalen Netzwerk oder Weitverkehrsbereich TCP/IP-Protokolle über ATM-Netzwerke übertragen zu können, mußten einige Spezifikationen und Anpassungen an ATM vorgenommen werden. Das liegt an den völlig unterschiedlichen Strukturen und Eigenschaften dieser beiden Übertragungsarten.

ATM ist aufgrund seiner virtuellen Verbindungen verbindungsorientiert aufgebaut, besitzt eine eigene Adressenstruktur und Routing-Funktionen. Dadurch besitzt ATM den Vorzug der universellen Skalierbarkeit gegenüber anderen Netzstrukturen. Kleine 53-Byte-Zellen transportiert ATM über diese virtuellen Verbindungen. Die ATM-Signalisierung übernimmt dabei den Auf- und Abbau der virtuellen Verbindungen (VCC/VPC). Das heißt, es können Verbindungen manuell über Permanent Virtual Circuits (PVC) oder automatisch über Switched Virtual Circuits (SVC) konfiguriert werden. Dafür wird ein spezieller Verkehrsvertrag ausgehandelt und der QoS festgelegt. Der QoS definiert dabei die Parameter wie Bandbreite, Verlustrate, Jitters usw., um eine garantierte Dienstgüte einhalten zu können. Allerdings sind bisher nur Punkt-zu-Punkt-Verbindungen realisiert. Für den Verbindungsaufbau wird die Signalisierung benötigt, die man in unterschiedliche Bereiche unterteilt:

- User Network Interface (UNIv3.0/v3.1/v4.0 für die Kommunikation zwischen Endgerät und Switch
- Interim Inter-Switch Signaling (IISP) und Private Network-to-Network-Interface (P-NNI) für die Kommunikation zwischen den ATM-Switches
- Broadband-ISDN User Part (B-ISUP)
- Broadband-Intercarrier-Interface (B-ICIV2.0)

Das Internet Protokoll (IP) ist hingegen verbindungslos realisiert. TCP/IP leitet Datenpakete auf Hop-by-Hop-Basis (Netzwerk-zu-Netzwerk) weiter, unabhängig vom darunterliegenden Netzwerk (Ethernet, Token Ring, FDDI im LAN-Bereich sowie X.25, ISDN, Festleitungen, SDH, ATM im WAN-Bereich). Fehlererkennung und -korrektur sowie Erhaltung der Sendereihenfolge und Duplikaterkennung, werden nicht durch IP angeboten und müssen durch Protokolle höherer Schichten (Transport Control Protocol – TCP) vorgenommen werden. TCP/IP ist inzwischen für fast alle Rechnerplattformen und Betriebssysteme verfügbar und hat sich aufgrund seiner Verbreitung durch das Internet als Quasistandard durchgesetzt. Weiterentwicklungen in den letzten Jahren haben das Protokoll immer weiter verbessert.

MAC-Protokolle (Media Access Control Protocols) für Shared-Medium-LANs arbeiten ebenfalls nicht verbindungsorientiert. Somit sind keine Quittungsmechanismen beim Empfang implementiert worden. Verlorene Datenpakete müssen durch höhere Protokollebenen wiederholt angefordert werden, wie das bei IP-Protokollen durch das TCP-Protokoll geschieht. LAN-Netze übertragen Daten über ein physikalisches Medium, wodurch die Information durch Zugriffsmechanismen für jede angeschlossene Station zur Verfügung steht. Broadcast-Funktionen sind nötig, da jede Station mit den anderen angeschlossenen Stationen kommunizieren muß. Dies geschieht durch spezielle Angabe der Zieladresse, wodurch sich eindeutige Kommunikationsbeziehungen zu anderen Netzteilnehmern auf-

bauen lassen.

Diese Eigenschaften liegen im starken Gegensatz zu den ATM-Mechanismen. Zwar lassen sich Broadcast-Nachrichten auch über viele virtuelle Verbindungspfade realisierbar: Dies wird durch Signalisierungsmechanismen mit Hilfe von Zuordnungstabellen ermöglicht. ATM unterstützt diese Möglichkeit aber noch nicht ausreichend. Zusätzlich zu den verschiedenen Eigenschaften treten Probleme im Bereich unterschiedlicher Übertragungsgeschwindigkeiten und inkompatibler Paket- bzw. Zellenformate auf. Diese Schwierigkeiten können jedoch heutzutage durch leistungsfähige Router überwunden werden, wie dieses bereits bei Übergängen von Ethernet/Token Ring zu FDDI-Netzwerken der Fall ist. Weiterhin ist die Adressenumsetzung sowie das Routing selbst als nicht unproblematisch anzusehen.

3 Anpassungsmöglichkeiten von ATM auf TCP/IP-Netze

Zwei Institutionen sind bei der Entwicklung von Anpassungsmöglichkeiten zwischen IP und ATM entscheidend. Die Internet Engineering Task Force (IETF) arbeitet in der Arbeitsgruppe Internetworking Over NBMA (ION) an Verfahren IP an NBMA (Non-Broadcast Multiple Access) Netze, wie ATM, X.25 und Frame Relay anzupassen. Weiterhin ist das ATM-Forum, ein Zusammenschluß aus Herstellern, Universitäten, Entwicklern, Diensteanbietern und Nutzern, vorhanden. Ziel dieses Forums ist die schnellere Spezifizierung von ATM-Standards mit Hilfe der Industrie, um die Arbeiten der International Telecommunication Union (ITU) zu beschleunigen. Herausgebildet hat sich besonders die LANE- und die MPOA-Arbeitsgruppe. Um die vermittlungstechnisch völlig unterschiedlichen Welten von IP und ATM miteinander zu verbinden, sind aus diesen Arbeitsgruppen mittlerweile die folgenden Verfahren hervorgegangen:

- Classical-IP (RFC-1577)
- LAN-Emulation (LANE)
- Multiprotocol over ATM (MPOA)

Classical-IP ist von der Internet Engineering Task Force (IETF) definiert worden und läßt ausschließlich Unicast-Verkehr zu. Das Ziel ist es dabei, Endsysteme mit ATM-Schnittstelle die IP-Nutzung zu ermöglichen. Classical-IP war die erste Spezifizierung, die die IP-over-ATM-Übertragung ermöglichte. Dabei sind alle Endsysteme eines Logical IP Subnet (LIS) über ATM erreichbar. Die Kommunikation kann man dabei über PVC- oder SVC-Verbindungen vornehmen. Die weitere Anbindung eines LIS an andere LIS oder traditionelle Router erfolgt über Router. Die Adreßauflösung der IP-Adressen zu den ATM-Adressen erfolgt mittels eines zentralen ATMARP-Servers.

Dagegen wurde die LAN-Emulation vom ATM-Forum definiert und hat die Anbindung von ATM an traditionelle lokale Netzwerke (Token Ring, Ethernet, Token Bus, FDDI etc.) zur Aufgabe. Hauptanwendungsgebiete sind deshalb die LAN-Kopplung über ATM sowie die Verbindung von LANs mit Endgeräten am ATM auf MAC-Ebene. Dadurch wird eine transparente Nutzung von ATM erlaubt. Emulierte LANs (ELANs) und traditionelle LANs können durch Ethernet-Switche miteinander verbunden werden. Dies geschieht durch Bridges, während unterschiedliche IP-Netze durch Router miteinander kommunizieren.

Die letzte Möglichkeit der IP-Anpassung stellt das MPOA-Verfahren dar, wobei man hier eigentlich nicht von einer reinen Anpassung ausgehen kann, da komplette Routing-Mechanismen realisiert wurden. Dieses Verfahren entwickelte wiederum das ATM-Forum, welches für die Verwendung von ATM in bestehende Datennetze spezifiziert wurde. Zusätzlich werden weitere Protokolle der IETF wie RSVP, MARS, NHRP, RFC-1483 und RFC-1577 mit einbezogen. Ziel ist es ATM problemlos in die TCP/IP-Umgebung integrieren zu können. Es lassen sich zusätzlich virtuelle Netze realisieren.

3.1 Classical-IP (RFC-1577)

Die IETF veröffentlichte 1993 die erste verfügbare Spezifikation zur Übertragung von LAN-Daten über ATM-Netzwerke. Dadurch wurde es erstmals möglich TCP/IP-Protokolle für ATM-Netzwerke zu nutzen. Das erste Dokument RFC-1483 definierte, wie unterschiedliche Protokolle in ATM-Zellen eingepackt und verschickt werden können. Dabei werden in dieser Spezifikation zwei grundsätzlich verschiedene Methoden zur Einkapselung von LAN-Datenpaketen beschrieben:

- Logical Link Control Encapsulation: Mehrfachprotokollarten können über eine einzelne Verbindung transportiert werden. Die eingekapselten Pakete lassen sich dabei durch einen LLC/SNAP-Header identifizieren. Eine weitere Auswirkung der LLC/SNAP-Einkapselung ist es, daß alle Verbindungen, die eine solche Einkapselung verwenden, an der LLC-Schicht innerhalb der Endsysteme enden, da hier das Paketmultiplexen stattfindet.
- VC Based Multiplexing: Bei der VC-Multiplexmethode wird nur ein einzelnes Protokoll über eine ATM-

Verbindung, inklusive mit dem an der Verbindungskonfiguration identifizierten Protokolls, übertragen. Folglich ist kein zusätzliches Multiplex- oder Pakettypenfeld erforderlich, obwohl dem eingekapselten Paket ein Blockfeld vorangestellt werden könnte. Die Art der Einkapselung, die von der LANE für die Datenpakete verwendet wird, ist eine Form des VC-Multiplexens.

Ein halbes Jahr später folgte die eigentliche Classical-IP-Spezifizierung (RFC-1577), welche die IP-Übertragung über ATM realisieren sollte. Hier werden die Datenpakete nicht nur eingepackt, sondern vollständig in ATM implementiert. Dabei behandelt RFC-1577 hauptsächlich die Adressenumsetzung im logischen IP-Subnetz. Weitere RFC-Dokumente folgten, wie RFC-1626 und RFC-1755. RFC-1626 beschreibt die maximal zulässigen Paketgrößen, während RFC-1755 die Verwendung der ATM-Signalisierung UNI 3.1 in IP-Netzen definiert. Classical-IP legt nun logische Subnetze auf ATM-Netzen mit Hilfe der Logical IP Subnet (LIS) fest. Ein LIS besteht dabei aus einer Reihe von Clients (Rechner und Router) sowie einem ATMARP-Server. Dieser Server dient ausschließlich zur Adressenauflösung. Alle Clients besitzen eine identische IP-Adresse und Subnetzmasken, womit sichergestellt ist, daß jedes andere Endsystem über Punkt-zu-Punkt-ATM-Verbindungen im gleichen Subnetz erreichen werden kann. Außerhalb liegende Systeme können über Routerverbindungen kommunizieren. Die Mitgliedschaft von Routern und Endsystemen in mehr als einem Netzsegment ist durch physikalische ATM-Schnittstellen oder mehrere virtuelle Schnittstellen möglich. Die Kommunikation der verschiedenen Subnetze erfolgt wie bisher über Hop-by-Hop-Routing. Das gilt ebenfalls, wenn beide LIS zum selben physikalischen Subnetz gehören, also prinzipiell direkt miteinander kommunizieren können.

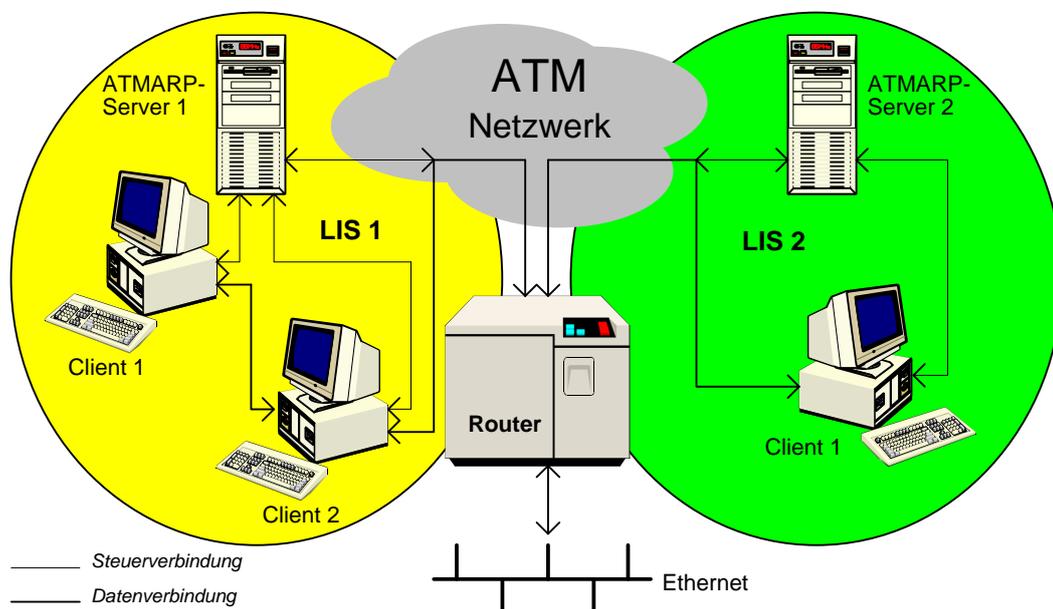


Abb. 1: Classical-IP Datenaustausch zwischen zwei logischen Subnetzen

Soll nun durch eine PVC- oder SVC-Schaltung eine ATM-Verbindung über einen ATMARP-Server aufgebaut werden, muß eine Registrierung der IP-Adresse im ATMARP-Server vorliegen. Nur so kann gewährleistet werden, daß dieser Server die IP/ATM-Adressenabbildungen für alle Endsysteme im ATM-Netz hat. Falls ein Endsystem IP-Pakete an eine IP-Adresse schicken will, die nicht in der ATMARP-Tabelle vorhanden sind, folgt eine Request-Mitteilung zum ATMARP-Server. Dieser gibt daraufhin eine ARP-Antwort, die die ATM-Adresse des Zielsystems beinhaltet. Anschließend kann der Rechner, der den Verbindungsaufbau eingeleitet hat, durch die UNI-Signalisierung eine SVC-Verbindung zum Ziel-endsystem aufbauen. Abb. 1 zeigt die Struktur von Classical-IP mit den LIS-Subnetzen, dem ATM-Netz dazwischen, dem ATMARP-Servern pro LIS und der Routerverbindung.

Classical-IP benötigt einen zusätzlichen Mechanismus, um IP-Pakete über ATM-Netzwerke zu übertragen. Dieser bedient sich der LLC/SNAP-Verpackung (IEEE 802.2) um die IP-Pakete einzukapseln. Das Logical Link Control (LLC) und seine Ergänzung Sub-Network Access Protocol (SNAP) versehen jedes versendete Paket mit einem Protokollkopf, der das betreffende Netzprotokoll kennzeichnet. Dabei legt der LLC-Header fest, daß es sich um ein weitergeleitetes Protokoll vom Router handelt. Der SNAP-Header besteht hingegen aus dem Organizational Unique Identifier (OUI) und dem Protocol Identifier (PID). Die Werte des OUI geben an, um welches Netzwerk es sich handelt, während PID die Protokollart angibt. Bei Ethernet-Anpassung steht an dieser Stelle beispielsweise der Wert 0x00-00-00. Dabei werden zwei Arten von IP-Paketen eingesetzt: normale IP-Struktur und ATM Address Resolution Protocol (ATMARP). Dadurch ist es möglich dieselbe ATM-Verbindung für verschiedene darüberliegende Netzprotokolle auszunutzen. Neben IP ist hier insbesondere ARP zu nennen. Die LLC/SNAP-Einkapselung muß stets von einem IP-System unterstützt werden. Prinzipiell kann eine Verbindung auch direkt zum IP als Nutzungs-

protokoll aufgebaut werden, so daß der Overhead durch LLC/SNAP entfällt. Dieses ist bei der LAN-Emulation bei Verwendung eines DIX-Ethernets der Fall. In diesem Zusammenhang lassen sich auch die Überlegungen zu Light-Weight-IP-Modellen wie TULIP (TCP and UDP over Lightweight IP) einordnen, bei denen der Overhead reduziert wird, indem vorgegebene Parameter bereits beim Verbindungsaufbau ausgetauscht werden. Abb. 2 zeigt die genaue Umsetzung eines IP-Pakets in ATM-Zellenpakete über die Anpassungsschicht AAL-Typ 5.

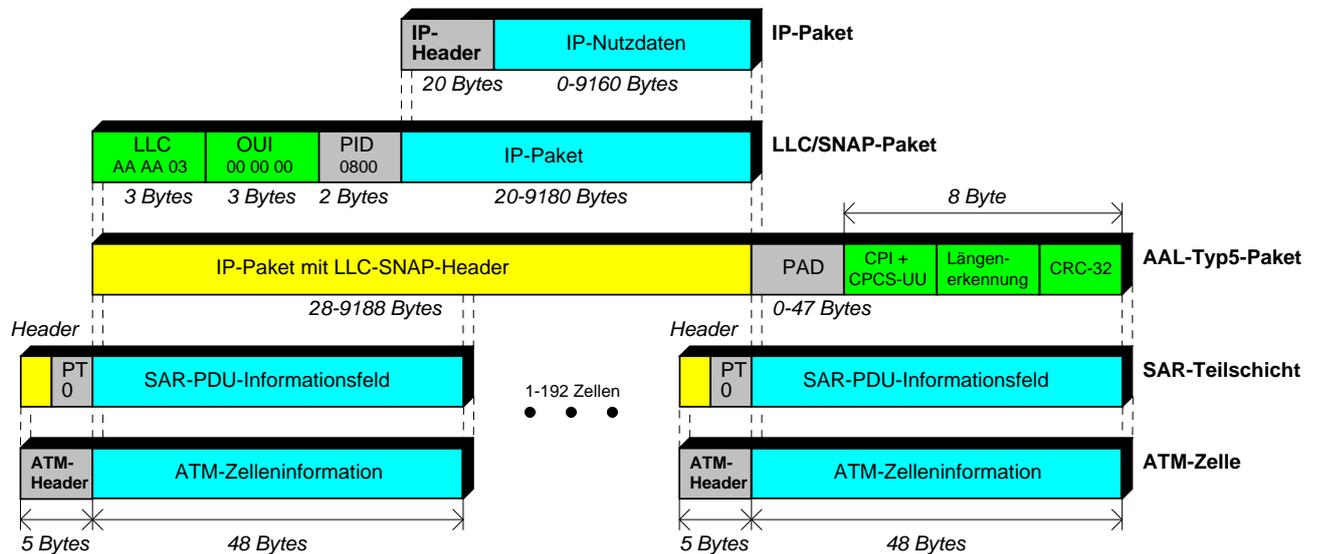


Abb. 2: Classical-IP Dateneinkapselung über die AAL-Typ5-Schicht von ATM

3.2 LAN-Emulation (LANE)

Die LAN-Emulation (LANE) ist eine Möglichkeit ATM-Produkte und -Netzwerke in vorhandene LAN-Topologien zu integrieren. Sie stellt dabei die universellere Methode der Netzanpassung dar, da sie die MAC-Schicht eines lokalen Netzwerkes vollständig emuliert. Dadurch wird bestehenden LAN-Anwendungen, ohne zusätzliche Einstellungen über ATM-Netzwerke hinweg, eine LAN-zu-LAN-Kommunikation vorgetäuscht. Im Januar 1995 wurde die erste LANE-Spezifizierung vom ATM-Forum vorgelegt, wodurch die Anbindung von lokalen Netzwerken über Protokolle wie IPX, AppleTalk sowie IP möglich wurde. Folgende Einsatzfälle werden dabei durch LANE abgedeckt:

- Verknüpfung traditioneller LANs über ein ATM-Netzwerk.
- Verbindung zwischen Endsystemen an traditionellen LANs mit denen an ATM-Netzwerken.
- Einsatz von LAN-Anwendungen zwischen Endsystemen, die direkt am ATM-Netzwerk angeschlossen sind.

Weiterhin hat LANE die Aufgabe vorhandene LAN-Treiber zu benutzen, da diese für traditionelle LANs auch in Zukunft weiter verwendet werden. Somit würde ein Endsystem in der Lage sein über eine standardisierte MAC-Schnittstelle, wie NDIS (Network Driver Interface Specification) und ODI (Open Data-Link Interface), auf ein bestimmtes LAN zuzugreifen, ohne eine Anpassung von Protokollen und Programmen vornehmen zu müssen. Aus diesem Grund setzt LANE unterhalb der Schicht 3 an und emuliert den MAC-Treiber eines herkömmlichen LANs. Die Verbindungssysteme zwischen den traditionellen und den emulierten LANs (ELAN) erfüllen dabei die Brückenfunktion: Transparent oder Source Routing.

Ein emuliertes LAN besteht aus vier wesentlichen Komponenten, die in Abb. 3 dargestellt sind:

- LAN Emulation Clients (LEC)
- LAN Emulation Server (LES)
- LAN Emulation Configuration Server (LECS)
- Broadcast and Unknown-Server (BUS)

Das bedeutet, daß LANE auf dem Client-Server-Prinzip basiert, wobei es einen LAN Emulation Service gibt, der die zentralen Dienste der LANE bereitstellt (Abb. 3). Jedes Endsystem, welches die LANE nutzen möchte, muß einen LAN Emulation Client (LEC) besitzen. Dieser ist meistens innerhalb entsprechender Treibersoftware auf einer ATM-Karte implementiert. Dabei besteht die Hauptaufgabe des LEC in der Bereitstellung der MAC-Schnittstelle für höhere Protokolle. Das heißt, er muß die MAC-Adressen auflösen, die Verbindungen abbauen und danach die Daten über ATM versenden. Für den LEC sind für diesen Fall zwei Adressenarten vorhanden: 6-Byte-

MAC-Adresse und 20-Byte-ATM-Adresse.

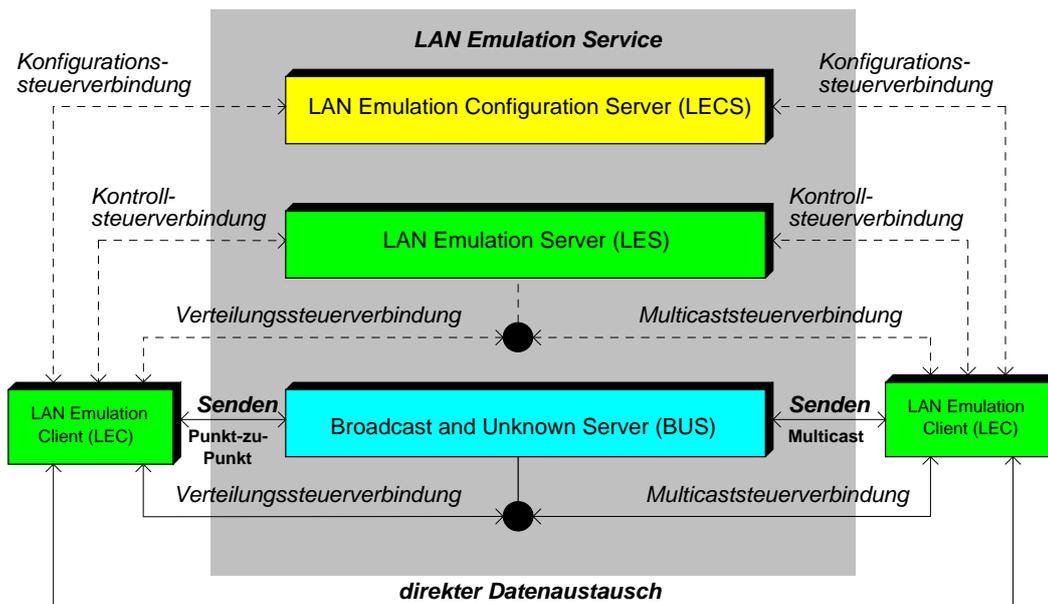


Abb. 3: LANE-Architektur

Abb. 3 zeigt die drei wesentlichen Komponenten des LAN Emulation Services (LES). Dabei dient der LECS zur automatischen Konfiguration der LECs und ermöglicht die einfache Konfiguration virtueller LANs von einer zentralen Stelle aus. Das heißt, es wird die Zugehörigkeit der LAN Emulation Server zu verschiedenen LANs zentral verwaltet. Die Konfigurationsinformationen werden dabei in einer Datenbank gespeichert und mit Hilfe der Clients einem emulierten LAN zugeordnet. Der LAN Emulation Server ist die sogenannte Schaltzentrale der LANE. Seine Aufgabe ist es, die LECs zu unterstützen, indem ATM-Adressen für vorgegebene MAC-Adressen ermittelt werden. Zusätzlich registriert er alle vorhandenen LECs. Der jeweilige LEC sendet dazu die eigene LAN-MAC-Adresse, die dazugehörige ATM-Adresse sowie Wegwahlinformationen an den LES. Dieser untersucht daraufhin die eigene Adressentabelle nach der ATM-Adresse. Findet er die ATM-Zieladresse, kann die Datenübertragung beginnen. Zur Ermittlung der ATM-Adresse wird das LAN Emulation Address Resolution Protocol (LE-ARP) verwendet, da ein LES nicht die Aufgabe einer Brücke übernehmen soll und deshalb nicht alle Adressen kennt. Wird die ATM-Zieladresse nicht gefunden, so wird die Adressenzuordnung durch eine Broadcast Message über den Broadcast and Unknown-Server (BUS) vorgenommen. Der BUS ist demnach für die Verteilung der Broadcast- und Multicast-Pakete zuständig. Dabei stellt er auch die Pakete zu, die noch keine ATM-Verbindung vom LEC zugeordnet bekommen haben. Zusätzlich sind Routermechanismen implementiert, die zur Ermittlung des optimalen Weges dienen. Die empfangenen Datenpakete werden vom BUS in der korrekten Reihenfolge an die angeschlossenen LECs weitergeleitet, um eine Überschneidung mit AAL-5-Datenpaketen von anderen Sendern vermeiden zu können. Die logischen Einheiten des LAN Emulation Service lassen sich getrennt sowie gemeinsam in bestehende Systeme integrieren. Als Zielsysteme kommen dabei Workstations, ATM-Switches und Router in Frage.

Bei den Ethernet-Paketen wird eine Einkapselung über LLC/SNAP der Spezifizierung RFC-1042 verwendet, die die Verpackung der Pakete über IEEE 802 abwickelt. Zusätzlich ist die Möglichkeit über DIX-Einkapselung der Spezifizierung RFC-894 vorhanden. Dabei können einige leichtere Unterschiede zwischen den traditionellen LAN-Paketen und den emulierten LAN-Paketen ausgemacht werden. Zum einen werden die redundanten Datenfelder nicht mehr berücksichtigt und zu anderen findet keine weitere Prüfsummenkontrolle mehr statt. Redundante Datenfelder können beispielsweise Synchronisationsdaten enthalten, die bei der LANE überflüssig wären. Die Prüfsumme muß auch nicht mehr ermittelt werden, da bereits in den AAL-5-Paketen eine Überprüfung stattfindet. Abb. 4 stellt die Paketstruktur übersichtlich für ein emuliertes Ethernet nach IEEE 802.3 dar. Bei DIX-Ethernet entfällt die LLC/SNAP-Teilschicht. Dabei werden die IP-Pakete zuerst in ein LLC/SNAP-Paket umgesetzt und anschließend im LANE-Paket mit einem Header und der Sender- und Empfangsadresse versehen. Danach hängt die AAL-5-Schicht einen entsprechenden Trailer an das LANE-Paket, bevor es in ATM-Zellen zerlegt wird. Am Empfänger sorgt sie dann automatisch für die Zusammensetzung des ursprünglichen Datenflusses. Bei einer Verwendung des DIX-Ethernets entfällt die LLC/SNAP-Schicht.

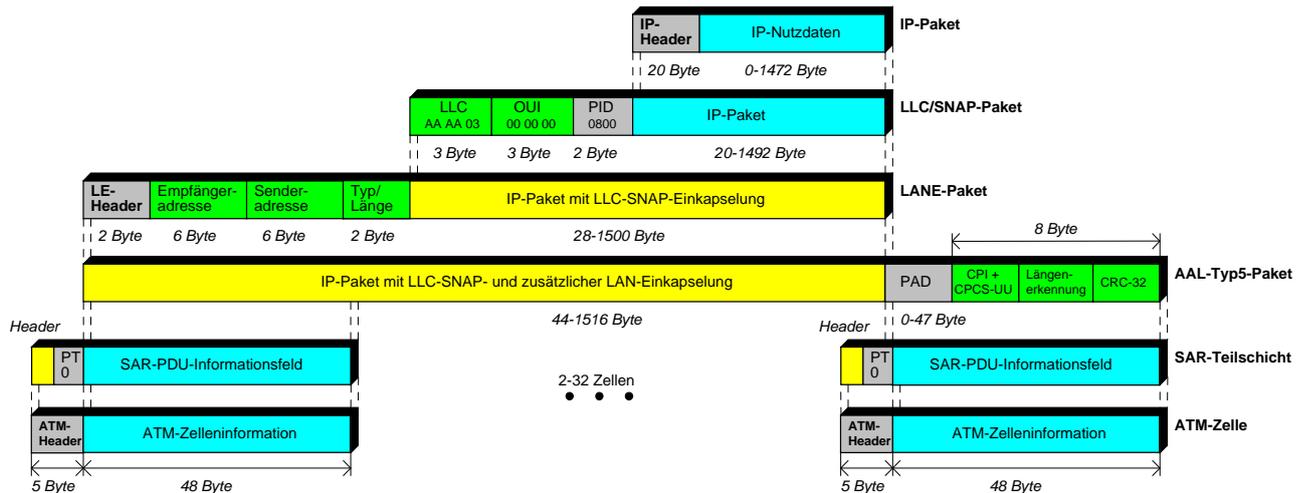


Abb. 4: LANE-Dateneinkapselung (IEEE802.3) über die AAL-Typ5-Schicht von ATM

3.3 Multiprotocol over ATM (MPOA)

Die vorhandenen Möglichkeiten von IP-over-ATM wie LANE und Classical-IP ermöglichen bisher nur den Aufbau einzelner Subnetze. Wenn mehrere Subnetze auftreten, sind Router für eine Kommunikation erforderlich. Dadurch kann die Effizienz erheblich beeinflusst werden, da Router den Datendurchsatz begrenzen und Verzögerungen durch Protokollbearbeitung hervorrufen. MPOA versucht diesen Nachteil durch neue Konstruktionen, die aus vielen Subnetzen bestehen, und sogenannten Shortcuts (direkte Verbindungen) zwischen unterschiedlichen IP-Subnetzen auszuschnitten. Da MPOA direkt auf der Netzwerkschicht 3 aufsetzt, wird die Ausnutzung von ATM-Eigenschaften ermöglicht. Durch die neue Idee eines virtuellen Routers und durch Integration bestehender Ansätze, versucht MPOA die Routerfunktionen Forwarding und Routing räumlich zu trennen. Bisherige Routerfunktionen ermöglichen den Informationsaustausch über die Wegewahl zum Empfänger (Routing) und der anschließenden Übertragung der Daten (Forwarding). Beim Ansatz des virtuellen Routers werden sogenannte Forwarder über ein standardisiertes Protokoll zentral von Router-Servern gesteuert. Die Routing-Funktionen werden dabei von den Router-Servern ausgeführt. Durch diese Trennung sind erhebliche Kosteneinsparungen durch Managementvereinfachungen über die Zentralisierung der Routerfunktionen möglich. Zusätzlich lassen sich einfache Forwarder mit einem geringeren technischen Aufwand realisieren. MPOA definiert zum virtuellen Router weiterhin das IP-Subnetz als Internetwork Address Sub-Group (IASG). Die IASG ist damit ein IP-Adressenbereich, der in einem IP-Routingprotokoll zu einer Route (Weg) zusammengefaßt ist. Wie beim traditionellen IP-Netz werden dadurch alle Adressen erfaßt, die man durch ein IP-Broadcast erreichen kann. Weitere Eigenschaften sind nicht mehr vorhanden. Somit sind nicht nur mehr Rechner lokaler Netze, sondern auch Rechner fremder Netze ohne Router erreichbar.

Der virtuelle MPOA-Router kann sich aus verschiedenen Spezifikationen zusammengesetzt kann. Das Private-Network Node Interface (P-NNI) und die LAN Emulation (LANE) sind dabei über das ATM-Forum definiert worden. Dabei stellt P-NNI ein Protokoll für das Routing auf ATM-Ebene dar, welches durch MPOA auf IP-Routing erweitert wird. Zusätzlich werden Spezifikationen der IETF verwendet, die die Paketeinkapselung, NHRP-, RSVP-Protokoll und MARS betreffen. Die Paketeinkapselung stellt eine Mehrfachprotokoll-Verpackung über eine ATM-Verbindung dar. Das NHRP-Protokoll erlaubt hingegen die Shortcut-Benutzung beim Versenden von IP-Paketen über ATM. Dadurch ist es möglich IP-Pakete direkt an andere Systeme in anderen Netzen weiterzuleiten, wodurch effiziente Methoden zur Verbindung virtueller Netze geschaffen wurden. RSVP ermöglicht die ATM-Eigenschaft der Bandbreitenreservierung ebenfalls über IP-Übertragung auszunutzen, während MARS das Versenden von IP-Multicast-Paketen übernimmt. Dabei können sowohl Multicast-Server (MCS) eingesetzt werden, als auch direkte Verbindungen zur Zieladresse. Ursprünglich waren die Ziele von MPOA darauf ausgelegt diese Protokolle komplett zu integrieren. Allerdings hat man sich, aufgrund der resultierenden Verzögerungen, darauf geeinigt LANE, RFC-1483 und NHRP im ersten Schritt zu implementieren. Später sollen weitere Protokolle (MARS, RSVP und P-NNI) folgen.

Die Grundarchitektur von MPOA besteht aus einer Client/Server-Umgebung und beschreibt die Funktionen von logischen Komponenten (Functional Groups – FG), die durch die Emulation eines virtuellen Routers über das ATM-Netzwerk miteinander verbunden sind.

Die MPOA-Architektur besteht dabei aus den folgenden Komponenten:

- Randgeräte (Edge Devices)

- ATM-Hosts
- Default Forwarder Functional Group (DFFG)
- Internet Address Sub-Group (IASG)

Bei MPOA sind zwei unterschiedliche Client-Arten vorhanden. Das sind zum einen die Randgeräte (Edge Devices) und zum anderen die ATM-Hosts, welche direkt mit einem ATM-Netzadapter an das ATM-Netz angeschlossen sind und das MPOA-Verfahren unterstützen. Sie sind dadurch in der Lage direkt mit anderen ATM-Hosts über ATM zu kommunizieren. Zusätzlich können traditionelle LANs über Edge Devices mit dem ATM-Netz kommunizieren. Diese Randgeräte vermitteln dazu Datenpakete zwischen herkömmlichen LAN-Segmenten und ATM-Schnittstellen. Zur Zielerkennung werden dabei Adressen der Netzwerkschicht oder MAC-Adressen verwendet. Deshalb kann man Edge Devices auch als Brücken oder Ethernet-Multilayer-Switches mit ATM-Anschluß und LANE bezeichnen. So ermöglichen sie Endgeräten MPOA-Dienste zu nutzen, die normalerweise nicht MPOA-tauglich wären.

Abb. 5 zeigt die Funktionsweise der MPOA-Architektur auf. Dabei sind für das MPOA-Verfahren fünf verschiedene Serverfunktionsgruppen definiert worden, die nicht alle in der Abbildung enthalten sind:

- Configuration Server Functional Group (CSFG)
- IASG Co-ordination Functional Group (ICFG)
- Default Forwarder Functional Group (DFFG)
- Route Server Functional Group (RSFG)
- Remote Forwarder Functional Group (RFFG)

Das MPOA-Verfahren definiert virtuelle LANs (VLAN) durch die Strukturen der Netzwerkschicht. Diese werden Internetwork Address Sub-Group (IASG) genannt. In der IASG Co-ordination Functional Group (ICFG) sind Funktionen vorhanden, die für die Verwaltung der IASG-Mitglieder nötig sind. Dieses beinhaltet, die Erfassung der ATM- und MAC-Adressen der ATM-Hosts wie der jeweiligen Randgeräte. Die Configuration Server Functional Group (CSFG) muß zuerst die fangskonfiguration der teilnehmenden Komponenten vornehmen und diesen ATM-Adressen zuteilen. Zusätzlich legt er die maximale Paketgröße (MTU) und die verwendeten Internet-

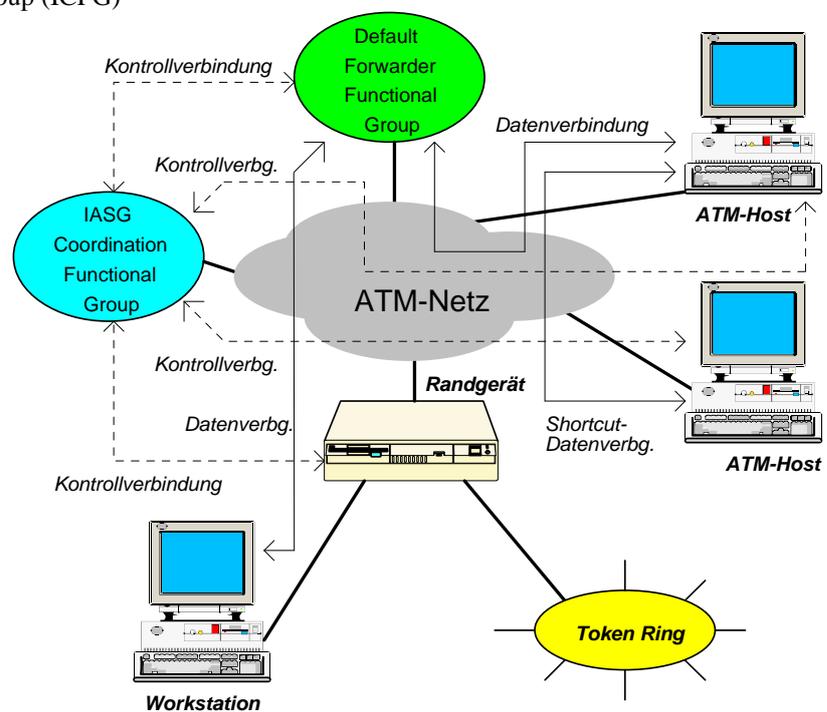


Abb. 5: MPOA-Architektur

Working-Protokolle fest. Eine weitere Kontrollverbindung ist von der ICFG zu der Default Forwarder Functional Group (DFFG) vorhanden, um die Anzahl der möglichen ICFG-ATM-Verbindungen vermindern zu können. Zusätzlich kann auch die Verzögerung beim Verbindungsaufbau über ATM verringert werden. Dieses geschieht, wenn IP-Pakete von einem Sender nicht direkt einer ATM-Verbindung zum Empfänger zugeordnet werden können. Der DFFG übernimmt zunächst die Pakete, wodurch es zu keiner Verbindungsaufbauverzögerung aufgrund fehlender Zieladresse kommen kann. Bei einer späteren Erkennung des Empfängers leitet dann der DFFG die IP-Pakete an den Empfänger weiter.

Zwei weitere Servergruppen, die nicht in der Abb. 5 aufgeführt sind, übernehmen die restlichen MPOA-Funktionen. Sie sind im Gegensatz zu den eben beschriebenen Servern für die Kommunikation zwischen den virtuellen Netzen zuständig. Die Remote Forwarder Functional Group (RFFG) realisiert den Forwarding-Teil eines virtuellen Routers und ist somit für die Verteilung von Multicast-Paketen zwischen MPOA-Clients, die keinen zusätzlichen Randgeräte benötigen, zuständig. Dabei kann man den Ansatz des Multicast Address Resolution Servers (MARS) verwenden, der mit seinen Funktionen im ICFG integriert ist. Wenn bestimmte Multicast-Gruppen senden oder empfangen wollen, müssen sich also die Clients des MPOA-Netzes an den MARS wenden. Die Route Server Functional Group (RSFG) realisiert hingegen den Routing-Teil eines virtuellen Routers und ist deshalb für die Kommunikation zwischen Wegen mit Router-Servern anderer IASGs zuständig. Auch die Router traditioneller Netze werden mit unterstützt. Beide Router-Server bieten somit eine Vielzahl von Funktionen an, mit denen sich Subnetze, die auf anderen

Netzwerkschichten definiert sind, auf ATM abbilden lassen. Die Routerserver können dabei ein eigenständiges Gerät darstellen sowie als Zusatzfunktionen in bestehende Router oder Switches integriert werden. Die Routerserver halten Adressentabellen für die Netzwerkschicht bereit, die MAC-Ebene und ATM bereit, um sie bei Anforderung an Randgeräte und ATM-Hosts übermitteln zu können. Dadurch sind überhaupt erst direkte virtuelle Verbindungen zwischen zwei beliebigen Endgeräten möglich geworden.

3.3.1 Multicast Address Resolution Server (MARS)

Das IETF hat in der RFC-1112 Spezifizierung die Unterstützung von IP-Multicasting durch eine Kombination zwischen einem Multicast-Server (MCS) und überlagernde Punkt-zu-Mehrpunkt-Verbindungen beschrieben. Diese Spezifikation befindet sich zwar in einer frühen Bearbeitungsphase, kann aber als ein Modell angesehen werden, welches Multicasting bei anderen Protokollen (NHRP und MPOA) unterstützen wird. Dabei wird davon ausgegangen, daß der Multicast Address Resolution Server (MARS) einer Gruppe von Netzknoten dient, die man auch als Cluster bezeichnen kann. Alle Endsysteme innerhalb des Clusters werden durch die ATM-Adresse des MARS konfiguriert. Dabei gibt es zwei Verfahren bei der Verteilung von Multicast-Paketen. Bei der ersten Möglichkeit werden die Pakete einer Multicast Group von den Sendern zum Multicast-Server transportiert. Der MCS ist dabei im DFFG integriert und leitet die empfangenen Pakete anschließend über Punkt-zu-Mehrpunktverbindungen an die jeweiligen Endsysteme der empfangenen Multicast Group weiter. Das hat den Vorteil, daß die Verbindungsanzahl immer auf einem erträglichen Niveau gehalten wird. Die zweite Möglichkeit verzichtet auf den zusätzlichen Server und stellt direkte Verbindungen zwischen jeder Sendestation und den Empfängern der Multicast Group her. Dadurch wird der Server-Engpaß vermieden, der bei einem zusätzlichen Multicast-Server auftreten könnte. Allerdings

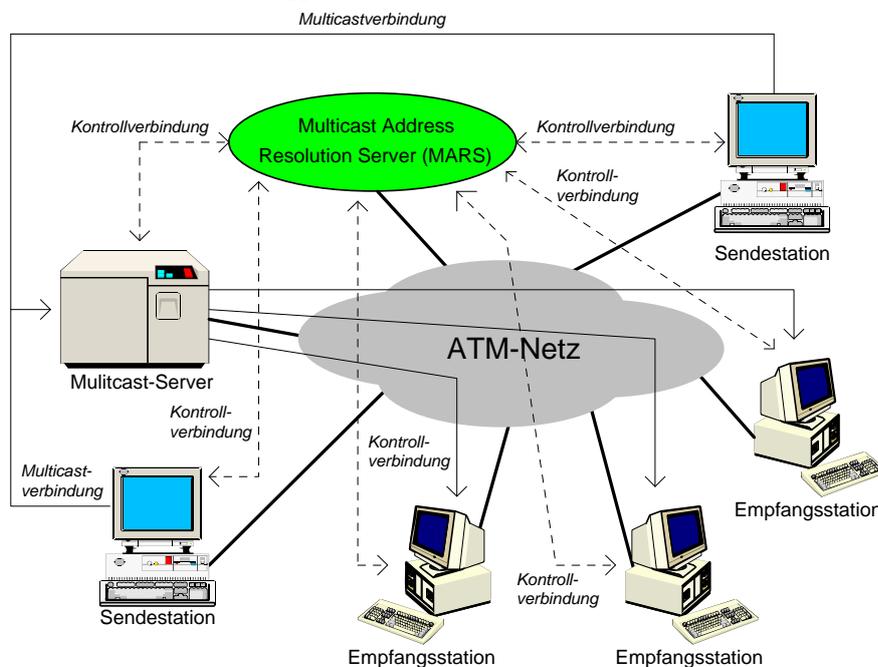


Abb. 6: Aufgaben des MARS-Servers bei Einsatz eines zusätzlichen Multicast-Servers

ATM-Netz zustande kommen.

3.3.2 Next-Hop Resolution Protocol (NHRP)

Bisher leiden die klassischen IP-over-ATM-Modelle an Begrenzungen, die ihnen von Host-Anforderungen auferlegt werden. Diese schließen Cut-Through-Routing aus und umgehen Router-Hops zwischen Knoten auf demselben ATM-Netz, die aber innerhalb zwei verschiedener Logical IP Subnetworks (LIS) vorhanden sind. Eine Arbeitsgruppe des IETF, welche sich Routing over Large Clouds (ROLC) nennt, arbeitet an Protokollen, die diese Begrenzungen aufheben sollen. Nach der Betrachtung von zahlreichen unterschiedlichen Ansätzen hat die Gruppe die Arbeit an dem Next-Hop Resolution Protocol (NHRP) abgeschlossen.

NHRP baut auf klassischen IP-Modellen auf und ersetzt die LIS-Begriffsvorstellung durch ein logisches Non-Broadcast Multi-Access (NBMA) Netzwerk. Das heißt, daß eine Netztechnologie (wie ATM, Frame Relay oder X.25), die den Zugriff vieler Geräte auf ein Netz ermöglicht, aber nicht die Verwendung von Broadcast-Mechanismen erlaubt, praktisch auf einem LAN angeordnet ist. Solch ein Netzwerk besteht aus einer Gruppe von

könnte dieses wiederum einen Anstieg von Serververbindungen bedeuten.

Die IETF-Spezifikation RFC-2022 beschreibt zusätzlich einen Mechanismus, wie man IP-Multicasting über ATM ausnutzen kann. MARS verwaltet analog den ATMARP-Servern die IP-Multicast-Adressen mit den jeweiligen ATM-Adressen. Dabei sind nicht nur einzelne ATM-Adressen vorhanden, sondern eine Liste von mehreren Empfängeradressen. Multicast-Pakete werden durch die zentrale MCS oder durch direkte Verteilung wie beschrieben vorgenommen. Abb. 6 verdeutlicht die Funktionsweise des MARS-Servers mit zusätzlichem Multicast-Server, wobei hierbei beachtet werden sollte, daß natürlich alle Verbindungen über das

Knoten, die in demselben NBMA-Netz (in diesem Fall ist ein ATM-Netz gemeint) vorhanden sind. Dabei findet die Kommunikation zwischen den unterschiedlichen Knoten auf dem direkten Wege statt. Ein einzelnes NBMA-Netz kann mehrfache administrative Domänen unterstützen, wobei jede über direkte Verbindungen mit der anderen kommunizieren kann. Dabei können auch bestimmte Verbindungen ausgeschlossen werden, beispielsweise um Firewalls errichten zu können. NHRP ist innerhalb jeder administrativen Region anwendbar, erlaubt aber direkte Verbindungen nur am Zutrittspunkt einer anderen administrativen Region. Statt eines ARP-Servers wurde beim NHRP die Vorstellung von einem NHRP-Server (NHS) verwirklicht. Jeder Next-Hop Server (NHS) verwaltet einen Next-Hop Resolution Cache (NHRC) mit IP-zu-ATM-Adressenangleichungen. Die Knoten sind mit der ATM-Adresse ihres NHS konfiguriert und melden anschließend ihre eigenen ATM- und IP-Adressen beim NHS mit Hilfe von Registrierungspaketen an, so daß das NHS seine Cache-Tabellen erweitern kann. Dafür hat man das ARP-Protokoll zu dem NBMA Address Resolution Protocol (NARP) erweitert. Nachteilig bleibt beim NHRP, daß Routing-Schleifen möglich sind und die Netzübergänge von NBMA-Netzen zu traditionellen LANs über Router abgewickelt werden müssen.

Letztendlich besitzt NHRP jedoch den großen Vorteil direkte Datenverbindungen schnell herzustellen. Einige Verbesserungen müssen dabei jedoch noch implementiert werden. Neben dem Routerproblem ist bisher auch nicht an eine Autokonfiguration gedacht worden. Weiterhin ist keine Unterstützung von Multicast/Broadcast-Funktionen vorhanden. Trotzdem wird sich NHRP weiterentwickeln und eine bedeutende Rolle im Zusammenhang der MPOA-Spezifikation spielen. Die MPOA-Spezifikation nimmt eine verbesserte Anpassung von NHRP an ATM vor, da bei ihr ausschließlich die Shortcut-Funktionen verwendet werden.

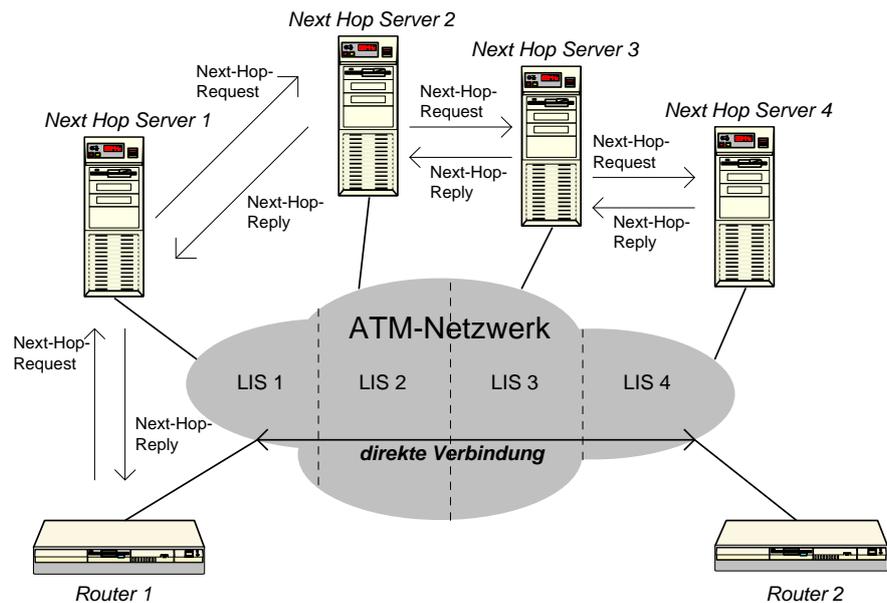


Abb. 7: NHRP-Funktionalität

Andere Funktionen, die NHRP noch nicht erfüllen kann, werden von anderen Protokollen aufgefangen.

3.3.3 Resource Reservation Protocol (RSVP)

ATM verwendet Signalisierungsprotokolle (Q.2931) um eine virtuelle Verbindung zwischen zwei Endpunkten aufbauen zu können. Zusätzlich werden in diesem Fall die Parameter des Quality-of-Service (QoS) festgelegt. Die Charakteristika haben dabei sehr viel von herkömmlichen Signalisierungsprotokollen, welche häufigen Senderwechsel handhaben und mit den Hard-State-Zuständen in den ATM-Switches die bestehende Verbindung aufrechterhalten müssen. Dabei kann bei der momentanen Standardisierung keine nachträgliche QoS-Änderung bei einer bestehenden Verbindung mehr vorgenommen werden. Das heißt, die Zuteilung bei der UNI-Signalisierung (UNI 3.1) ist statisch. Das bedeutet, daß zwar bei einer Unicast-Verbindung der QoS für Sender und Empfänger garantiert wird, aber bei Multicast-Verbindungen nur beim Sender eine Ressourcenzuteilung erfolgen kann.

Bei IP-Paketen war aufgrund der verbindungslosen Struktur bisher an gar keine Verbindungsgarantien gedacht worden. Um dieses zu ermöglichen, ist von der gleichlautenden Arbeitsgruppe der IETF das Resource Reservation Protocol (RSVP) vorgeschlagen worden. Dadurch wird es erstmals möglich Echtzeitsdienste in einer verbindungslosen Umgebung zu unterstützen. Auf der anderen Seite unterstützt ATM verbindungsorientierte Dienste. Dafür werden von ATM vor einem Verbindungsaufbau die notwendigen Ressourcen festgelegt.

Kategorie	RSVP-Protokoll	ATM-UNI-Signalisierungsprotokoll
Ausrichtung	Empfänger-basiert	Sender-basiert
Zustand	Soft-State (Aktualisierung/Zeitlimitüberschreitung)	Hard-State (explizites Löschen)
QoS-Konfigurationszeitpunkt	separat von der Routen-etablierung	gleichzeitig zur Routen-etablierung
QoS-Änderungen	dynamisches QoS	statisches QoS (Festlegung beim Konfigurationszeitpunkt)
Richtungsarten	unidirektionale Ressourcen-zuteilung	bidirektionale Zuteilung für Unicast; unidirektionale Zuteilung für Multicast
Heterogenität	Empfängerheterogenität	gleicher QoS für alle Empfänger

Tabelle 1: Gegenüberstellung der RSVP- und UNI-Eigenschaften

nachgefragt werden. Wenn dabei ein bestimmtes Zeitlimit überschritten wird, kommt es zum Verbindungsabbruch. Zusätzlich müssen die reservierten Ressourcen nachdrücklich gelöscht werden. Bei ATM hingegen bestehen die garantierten Ressourcen während der gesamten Verbindungsdauer (Hard State). Weiterhin erlaubt die Soft-State-Methode beim RSVP die Reservierung von QoS-Parametern für einen Datenfluß. Diese Parameter sind dynamisch und können jederzeit während der Verbindungsdauer verändert werden. Bei ATM ist dieses nicht möglich, da hier ein statischer QoS verwendet wird. RSVP ist ein Simplex-Protokoll, wodurch die Reservierung der Ressourcen nur in eine Richtung möglich ist. Bei ATM hingegen sind bidirektionale Punkt-zu-Punkt-Verbindungen vorhanden sowie unidirektionale Punkt-zu-Mehrpunktverbindungen. Die Ressourcen-Reservierung wird bei RSVP durch den Empfänger veranlaßt, während ATM die Reservierung von dem Endsystem durchführen läßt, welches eine Verbindung aufbauen möchte. Weiterhin unterstützt RSVP Multicast-Sitzungen, wodurch dynamisches Umschalten sowie Veränderung der Teilnehmeranzahl gewährleistet ist. ATM unterstützt dieses bisher noch nicht. Weitere Unterschiede lassen sich der Tabelle 1 entnehmen.

Somit kann man RSVP als ein Kontrollprotokoll bezeichnen, welches von Anwendungen innerhalb von IP-Endsystemen verwendet wird. IP-Endsysteme zeigen in einem IP-Netz den Netzknoten an, was für einen Datenstrom sie empfangen können. Diese Parameter beinhalten Bandbreite, Abweichungen, maximale Datenblöcke usw. Dazwischenliegende Systeme entlang eines Übertragungspfades vom Sender zum Empfänger interpretieren die RSVP-Kontrollpakete, um Zulassungskontrollen ausführen und die benötigten Ressourcen zuordnen zu können. Solche Systeme verwalten Soft-States über diese Datenströme zwischen den beteiligten Endsystemen. Durch die Eigenschaften des RSVP-Protokolls lassen sich die selben QoS-Spezifikationen hinsichtlich des Paketdatenstroms von ATM-Signalisierungen erreichen. Weiterhin baut RSVP im wesentlichen auf dem Multicast-Muster auf und leitet Datenströme über Punkt-zu-Mehrpunktverbindungen weiter. Neue Multicast-Protokolle, wie das Protocol Independent Multicast (PIM) werden zukünftig mit RSVP gekoppelt werden. Die Notwendigkeit einer verbesserten Schnittstelle zwischen RSVP und ATM hat man inzwischen beim ATM-Forum erkannt. Ergänzungen der ATM-Signalisierung hinsichtlich dynamischer Punkt-zu-Punkt/Mehrpunkt-Verbindungen sowie die Migration zum empfangenorientierten QoS sind in der Entwicklung.

3.3.4 Private Network-to-Network-Interface (P-NNI)

Das Private Network-to-Network-Interface (P-NNI) ist vom ATM-Forum definiert worden und regelt den Datenaustausch zwischen mehreren ATM-Endsystemen, die sich in einem privaten Netzwerk befinden. Für diese Aufgabe sind zwei verschiedene Protokolle festgelegt worden, die für den Verbindungsaufbau sowie den Informationsaustausch der jeweiligen Teilnehmer in einem ATM-Netzwerk zuständig sind. Dabei basiert das Signalisierungsprotokoll auf der UNI-Spezifikationen und wurde lediglich um Source-Routing- und Alternate Routing-Funktionen erweitert.

P-NNI liegt bisher erst in der Phase 1 vor und ging aus der Spezifizierung des Inter-Switch Signaling Protocols (IISP) hervor. IISP ist sehr einfach gehalten und soll später durch P-NNI vollständig abgelöst werden. Beide Protokollarten verhalten sich inkompatibel zueinander, da das IISP-Protokoll keine NNI-Signalisierung unterstützt. Weiterhin unterstützt IISP keinerlei QoS-Mechanismen, wodurch P-NNI klare Vorteile besitzt. In der Phase 1 von P-NNI sind allerdings nur begrenzte QoS-Möglichkeiten vorhanden. Deshalb wird das P-NNI-Protokoll daran gemessen werden, wie es die beiden oberen Ziele umsetzen kann. Momentan werden von dem P-NNI-Protokoll nur die UNIV3.0/3.1-Signalisierungen unterstützt. UNIV4.0 wird bisher nicht berücksichtigt, wodurch Gruppenadressierungen und ABR-Verbindungen nicht unterstützt werden. Diese Funktionalität wird erst in der Phase 2 implementiert werden.

Tabelle 1 zeigt die Unterschiede zwischen der UNI-Signalisierung von ATM und dem RSVP-Protokoll auf, die aufgrund der unterschiedlichen Verbindungsarten zustande kommen. Beispielsweise wird die Ressourcen-Reservierung in IP-Hosts und Routern durch den Soft-State nommen. Das bedeutet, daß die Reservierungen nicht permanent vorhanden sind, sondern in bestimmten Zeitintervallen immer

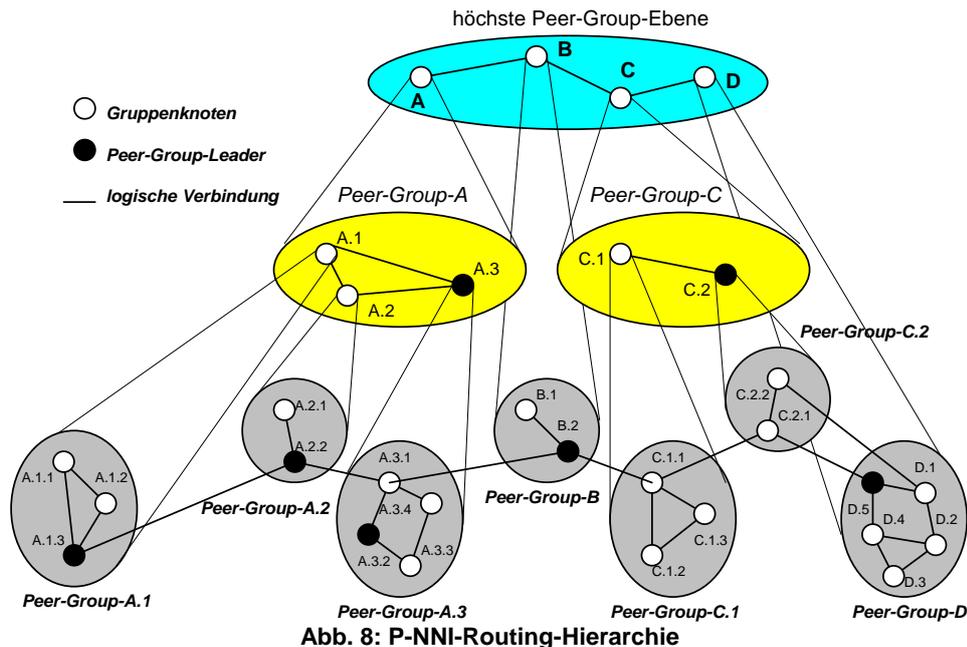


Abb. 8: P-NNI-Routing-Hierarchie

Da ATM eine verbindungsorientierte Werktechnologie ist, muß P-NNI für eine bestimmte Zeit eine virtuelle Verbindung aufbauen und aufrecht erhalten. Dadurch können durch ineffektives Routing virtuelle Pfad- und Kanalverbindungen zusätzlich belegt werden. Dieses soll das P-NNI-Routing verhindern. Weiterhin ermöglicht ATM dem Benutzer beim Aufbau den gewünschten Quality-of-Service (QoS) und die Bandbreite festzulegen. Dabei erfolgt der Aufbau einer

Verbindung in zwei Stufen: Selektion des ATM-Pfades und anschließende Verbindungsaufbau an jedem Punkt entlang des ATM-Pfades.

Die Pfadselektion muß dabei so durchgeführt werden, daß der gewünschte QoS und die Bandbreite an jedem Knoten der Verbindung eingehalten werden kann. Falls die geforderten Parameter nicht eingehalten werden können, kommt es zu einem sogenannten Crankback, d.h. ein neuer Pfad muß berechnet werden. Die Einbeziehung des QoS ist dabei immer noch ein aktuelles Forschungsthema. Es sind zwar Algorithmen bekannt, die in der Lage sind unter Einbeziehung mehrfach unabhängiger Verbindungsparameter eine QoS-sensitive Pfadauswahl treffen zu können, diese sind aber bisher schwer realisierbar. Aus diesem Grund ist es wichtig, daß P-NNI verschiedene Pfadauswahlalgorithmen für die QoS-Auswahl ermöglicht.

P-NNI-Signalisierung basiert als Untermenge auf der UNIV4.0-Signalisierung. Das heißt, es werden nicht alle Eigenschaften und Funktionen von P-NNI unterstützt. Die P-NNI-Signalisierung kann man als Nutzung der P-NNI-Routing-Informationen verstehen. D. h., die Wegwahlberechnung, die von der Erreichbarkeit, Konnektivität und Ressourceninformationen dynamisch durch das Routing verwaltet wird, kann abgeleitet werden. P-NNI ermöglicht somit ein Virtual Circuit Routing Protocol, welches für die Wegwahl der Signalisierungsanfragen durch das ATM-Netz verwendet wird. Dies ist auch die Strecke, auf der die ATM Verbindung eingerichtet wird und worüber anschließend die Daten transportiert werden. Damit ist ebenfalls das Routing der ATM-Verbindungen gemeint. Die Operation der Signalisierungsanfrage durch ein ATM-Netz hat dabei paradoxerweise mit dem verbindungslosen Paketrouting innerhalb vorhandener Netzschichtenprotokolle (wie bei IP) Ähnlichkeit. Das liegt daran, daß es vor einer Verbindung noch keine Signalisierungsanfrage geben kann. Solch ein VC-Wegplanungsprotokoll könnte daher einige Konzepte verwenden, die vielen der verbindungslosen Routing-Protokolle zugrunde liegen, welche in den letzten Jahren entwickelt wurden. Allerdings ist das P-NNI-Protokoll wesentlich komplexer, als bisherige Routing-Protokolle, weshalb es auch Verzögerungen bei der Standardisierung gab. Abb. 8 verdeutlicht die komplizierte P-NNI-Routing-Hierarchie.

3.4 Gegenüberstellung der IP-over-ATM-Verfahren

Der Hauptvorteil von Classical-IP ist die MTU-Größe von 9180 Byte, Ausnutzung von SVC-Verbindungen sowie die einfache Struktur. Die meisten UNIX-Systeme und Router mit ATM-Interface besitzen die Möglichkeit Classical-IP auszunutzen. Dadurch das Classical-IP bereits seit 1995 verfügbar ist, konnte man auch größtmögliche Interoperabilität verschiedener Hersteller erreichen. Leistungsfähige Endsysteme können bis an die Grenzen der theoretischen Nettoübertragungsrate die verfügbare Bandbreite ausnutzen. Nachteilig bleibt die fehlende Multicast/Broadcast-Unterstützung, keine redundanten ATMARP-Server sowie manuelle Konfiguration. Weiterhin müssen alle Systeme eines LIS an ATM angeschlossen sein, der Verbindungsaufbau kann größere Verzögerungen besitzen, direkte Verbindungen zu anderen Subnetzen ist nicht möglich und ausschließlich das IP-Protokoll wurde auf ATM angepaßt.

LANE hingegen ist die wesentlich komplexere Anpassung von IP an ATM, da die MAC-Schicht emuliert wird. LANE besitzt Autokonfiguration für verbesserte Managementfunktionen, verzögerungsfreier Transport der Daten-

pakete (auch ohne Verbindung zum Ziel), gute Integrationsmöglichkeiten von ATM in bestehende Netze über Ethernet-Switche mit ATM-Interface. SVC-Verbindungen und IP-Multicasting werden ebenfalls unterstützt. Durch Autokonfigurationsmöglichkeiten kann man LAN für kleine und mittlere Netze einsetzen. Nachteilig bleiben die kleinere MTU (1500 Byte) gegenüber Classical-IP und die wesentlich höhere Komplexität durch die verschiedenen Server und Multicast-Verbindungen. Weiterhin ist die LANE-Struktur noch nicht redundant ausgelegt, es fehlt die vollständige Standardisierung und der BUS sowie ältere Netzadapter begrenzen die Leistung des Gesamtsystems. Zusätzlich werden Multicast-Meldungen wie Broadcast-Meldungen verwendet und deshalb an alle angeschlossenen Endsysteme geschickt. Das größte Problem stellt aber weiterhin die fehlende QoS-Unterstützung bei LANE und Classical-IP dar.

Das MPOA-Verfahren ist ein völlig neuer Ansatz, der versucht die Datenübertragung über ATM noch effizienter zu gestalten. Verschiedene Protokolle wie RSVP, NHRP, MARS will man aus den IETF-Spezifikationen dafür integrieren oder zumindest auf MPOA anpassen. Dadurch werden sich zukünftig erstmals die Eigenschaften von ATM ausnutzen lassen. Weiterhin entstehen durch das MPOA-Verfahren virtuelle LANs, die völlig neue Möglichkeiten der Konfiguration und des Managements von komplexen Netzstrukturen ermöglichen. Beispielsweise lassen sich Router-Informationen durch Clients nutzen, so daß Verbindungen zu Clients anderer Netze sehr effizient ermöglicht werden. Router-Server erleichtern zusätzlich das Management großer Netzwerke. Reduzierung der Teilnehmeranzahl durch die logische Zugehörigkeit zu einem VLAN und Verringerung des IP-Adressenraums sind weiteren Vorteile einer strukturierten Netzkonfiguration. Bei Erweiterungen des Netzwerkes wird der Verwaltungsaufwand enorm verringert, da die Teilnehmer sich durch ihre logischen Adressen zuordnen lassen. Zusätzlich ermöglicht das MPOA-Verfahren den Aufbau eines virtuellen Routers mit einem einzigen Hop über das gesamte Netzwerk, wodurch sich hohe Datenübertragungsraten und geringere Verzögerungszeiten für ergeben. Dabei spielt die Zugehörigkeit zu bestimmten Subnetzen keine Bedeutung. Ein weiterer großer Vorteil ist die direkte Abbildung der Netzwerkschicht auf die ATM-Schichten, wodurch die Umsetzung von Subnetzstrukturen in VLANs unterstützt wird. Außerdem entfällt für den Verbindungsaufbau ein großer Teil des Overheads, da beispielsweise keine Broadcast-Meldungen versendet werden, um IP-ARPs aufzulösen. Nachteilig ist die fehlende Standardisierung und die hohe Komplexität, wodurch es zu Stabilitätsproblemen kommen kann. Weiterhin wird aufgrund der Entwicklungsverzögerung zuerst an der NHRP- und LANE-Integration gearbeitet. MARS und RSVP sollen folgen, wenn die Spezifikationen vollständig vorliegen. Somit bleibt es beim Hauptvorteil von MPOA Router-Engpässe zu beseitigen und virtuelle Netzstrukturen zu etablieren.

Classical-IP ist momentan die stabilste Variante für IP-Anpassung auf ATM. LANE wird sicherlich im Bereich kleiner und mittlerer traditioneller Netzwerke Verwendung finden, um die vorhandene höhere Bandbreite auszunutzen. MPOA hingegen stellt eine wirkliche Alternative und Implementierung von IP in ATM dar, die bei fertiggestellter Standardisierung im Bereich großer Netzwerke ihren Einsatz finden wird. Aufgrund ihrer unterschiedlichen Eigenschaften wird es alle drei Alternativen parallel im Netzbereich geben. Alle Verfahren werden kontinuierlich weiterentwickelt.

4 IP-Unterstützung von Gigabit-Ethernet und SDH/SONET

Das Internet-Protokoll (IP) ist ursprünglich für heterogene Strukturen entwickelt worden. Aus diesem Grund kann es beliebige Physical Layer zum Datentransport ausnutzen. Die Synchron Digital Hierarchy (SDH) hat sich im WAN-Bereich als Vermittlungstechnik durchgesetzt, so daß sogar ATM-Zellen über diese Technologie übertragen wird. Gigabit-Ethernet wird als mächtiger Konkurrent für ATM betrachtet. Im LAN-Bereich hat dies auch durchaus seine Berechtigung. Beide Hochgeschwindigkeitsnetze müssen aufgrund des Quasi-Standards IP in der Lage sein dieses Protokoll zu unterstützen. Die nächsten beiden Unterkapitel geben Aufschluß darüber.

4.1 Gigabit-Ethernet (IEEE802.3z)

Aufgrund der Abwärtskompatibilität und den kostengünstigen Netzadapterkarten haben sich Fast-Ethernet-Netzwerke sehr schnell am Markt etabliert. Aufgrund der aber immer noch bestehenden Mängel wurde bald über eine weitere Erhöhung der Bandbreite nachgedacht. Die Spezifikation für ein Gigabit-Ethernet wird deshalb gegenwärtig durch die IEEE 802.3z Gigabit Task Force erarbeitet. Das endgültige Ziel dieses Standards, der wahrscheinlich im Laufe von 1998 abgeschlossen sein wird, ist es Vollduplexübertragung mit 1000 MBit/s zu ermöglichen und dabei dieselben Datenpakete wie 10Base-T und 100Base-T zu verwenden. Die Empfehlungen für Protokoll Media Access Controller (MAC), Repeater und Physical Layer liegen bereits vor, wobei komplexere Themen wie, Flußkontrolle, Repeater-Architektur, Multimode/Monomode-Übertragung sowie Codierungsverfahren für Kupferleitungsübertragung noch diskutiert werden. Mehr als 60 Unternehmen haben sich parallel zu diesen Standardisierungsvorhaben in der Gigabit-Ethernet-Allianz (GEA) zusammengeschlossen, um die Standardisierung und Entwicklung von Gigabit-Ethernet-Produkten zu beschleunigen.

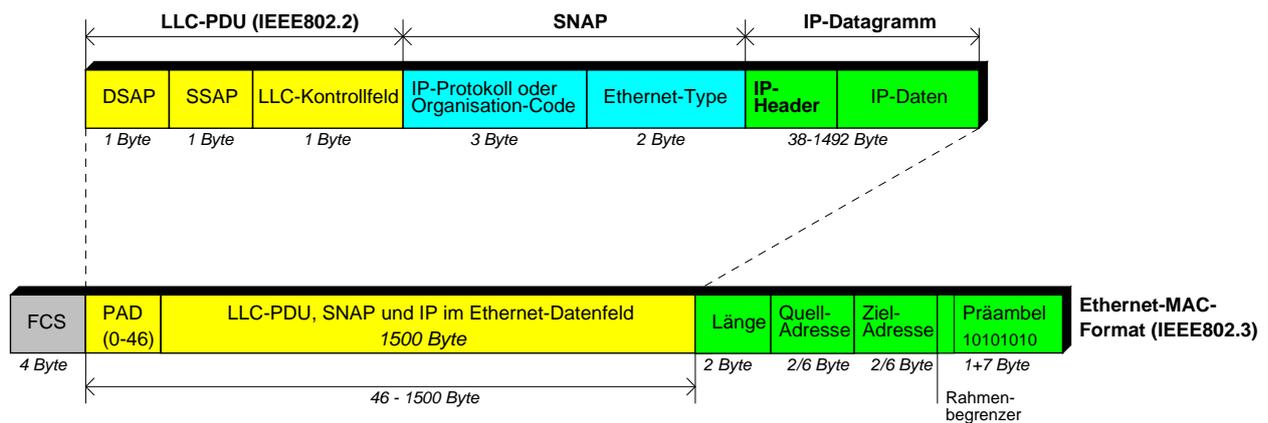


Abb. 9: IP-Paketeinkapselung bei Gigabit-Ethernet

Das Ethernet-Rahmenformat legt eine Länge von 64-1518 Byte einschließlich Header, Daten und Trailer fest. Dabei unterscheidet man zwischen zwei Formaten, da DEC, Intel und Xerox (DIX) Ethernet etwas anders definiert haben, als die IEEE mit der Spezifikation IEEE802.3 (Carrier Sense Multiple Access Bus with Collision Detection – CSMA/CD). Die IEEE verbesserte die DIX-Version und veröffentlichte IEEE802.3 im Jahr 1983. Unterschiede betreffen die Quell- und Zieladresse, die zusätzlich auch 2-Byte-Werte annehmen kann, das Address Resolution Protocol (ARP), welches die IP-Adresse (32 Bit) in die IEEE802.3-Adressen (48 Bit) umwandelt und das Längenfild, das die Dateneinheitlänge festlegt. Da sich der IEEE802.3-Standard durchgesetzt hat und Gigabit-Ethernet auch darauf aufbaut, wird dieser Rahmen in Abb. 9 dargestellt.

Das Datenfeld enthält die Informationen der oberen Schicht sowie zwei weitere IEEE-Protokollköpfe. Der Logical Link Control (LLC) Protokollkopf ist in IEEE802.3 und beinhaltet Destination Service Access Point (DSAP), Source Service Access Point (SSAP) und ein weiteres LLC-Kontrollfeld. Der LLC-Header ermöglicht dadurch die Adressierungsmöglichkeit für den gleichzeitigen Betrieb mehrerer Kommunikationsverbindungen auf einer Station. Das Kontrollfeld unterscheidet die verschiedenen LLC-Rahmen. Nach der LLC-PDU folgt das Sub-Network Access Protocol (SNAP), welches die Felder Protokoll-ID oder Organisation-Code sowie Ethernet-Typ enthält. Tabelle 2 vergleicht die verschiedenen Ethernet-Technologien miteinander und zeigt Gemeinsamkeiten auf.

Merkmale	Ethernet	Fast-Ethernet	Gigabit-Ethernet
Zeitschlitz	512 Bit	512 Bit	4096 Bit
interner Rahmenabstand	9,6 μ s	0,96 μ s	0,096 μ s
max. Paketlänge	1518 Byte	1518 Byte	1518 Byte
min. Paketlänge	64 Byte	64 Byte	64 Byte
Carrier Extension	nicht vorhanden	nicht vorhanden	448 Byte
Frame-Burst-Länge	12 000 Bit	12 000 Bit	12 000 Bit

Tabelle 2: Ethernet-Parameter

4.2 Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH)

SONET/SDH hingegen ist für Übertragungssysteme im WAN-Bereich entwickelt worden und wird von den Telekommunikationsanbietern weltweit verwendet. 1988 wurde diese Architektur zum SDH-Standard der ITU. Dieser Standard beschreibt ausschließlich die Übertragungsrahmen der OSI-Schicht-1 und unterscheidet zwei Varianten: die nordamerikanische ANSI-SONET und die europäische ETSI-SDH. Die erste SONET-Hierarchiestufe wurde dabei auf die speziellen amerikanischen Bedürfnisse auf eine Bitrate von 51,84 MBit/s ausgelegt. Die entsprechenden optischen und elektrischen Signale bei SONET werden als STS-1 (Synchronous Transport Signal, Ebene 1) bezeichnet. Im Gegensatz dazu basiert SDH, wie in den ITU-Empfehlung G.707 beschrieben, auf einer Bitrate von 155,520 MBit/s, also exakt der dreifachen SONET-Grundbitrate. Das zugehörige Signal wird hier als STM-1 (Synchronous Transport Module, Ebene 1) bezeichnet. Die Bitraten der nächsthöheren Hierarchiestufen sind ganze Vielfache der Bitrate von Stufe 1 und entstehen durch entsprechende Byte-Verschachtelung von Stufe 1 an aufwärts

SDH läßt den Transport verschiedener, durch unterschiedliche Bitraten und Strukturen gekennzeichnete Signale zu. Subsignalformate, die sogenannten Virtuellen Container (VC), die im STM transportiert und an Netzknoten unabhängig von ihrem Inhalt durchgeschaltet (Payload) werden, ermöglichen dieses. Ein VC besteht aus dem Path Overhead (POH) und dem eigentlichen Payload-Container (C). Der Inhalt des Containers besteht dabei entweder

aus kleineren synchronen VCs (Higher-Order-Container) oder aus Signalen der Plesiochronen Hierarchie (Low-Order-Container). Diese VCs besitzen Substrukturen, die von der zu transportierenden Nutzlast abhängig sind. Besteht die Nutzlast wiederum aus einzelnen VCs, so werden sie in Tributary Unit Group (TUG) unterteilt, welche ihrerseits wieder in Tributary Units (TU) aufgeteilt sind. Jeder TU stellt die Übertragungskapazität für einen VC-Kanal dar, während die TUG die Lage der TU im Higher-Order-VC bestimmt. Die TU, die sich unmittelbar unter der Hierarchie des STM-1-Rahmens befindet, wird Administrative Unit (AU) genannt, welche zusammengefaßt wiederum eine Administrative Group (AUG) ergeben. Der AU-Pointer referenziert den Beginn des VC-Containers. Einen Überblick über die Verschachtelungen von Bitratenuntersystemen in den STM-1-Rahmen zeigt Abb. 10.

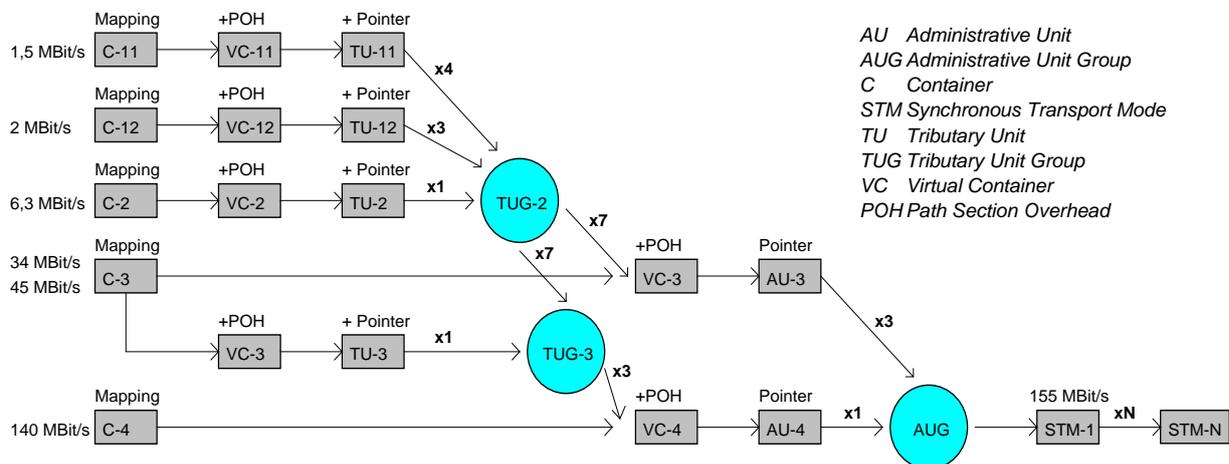


Abb. 10: SDH-Multiplexschema

Durch dieses Mapping können beliebige Protokollformate an die SDH-Rahmen angepaßt werden. CICSO wird als einer der ersten Hersteller ein entsprechendes Produkt anbieten. Dabei wird es sich um einen Router mit direkter SDH-Schnittstelle handeln, der bewußt ATM nicht mit einbezieht. Eine andere Möglichkeit speziell das Internet-Protokoll (IP) auf SDH-Rahmen anzupassen besteht über das Multiple Access Protocol over SONET/SDH (MAPOS). Dieses Protokoll definiert die Transportmethode HDLC-Rahmen über SONET/SDH zu übertragen (RFC-2171). Dies ermöglicht SONET/SDH-Rahmen gleichermaßen im LAN- wie auch im WAN-Bereich auszunutzen. Indem IPv4 auf das MAPOS-Protokoll angepaßt wird, erfolgt so eine IP-Übertragung durch Einkapselung in MAPOS-HDLC-Rahmen und ARP für das Mapping zwischen HDLC und IP (RFC-2176). Es handelt sich hierbei allerdings nicht um einen Standard, sondern ausschließlich um einen Vorschlag. Eine genauere Spezifikation müßte erfolgen, damit die IETF diesen Dokumenten größere Aufmerksamkeit widmen wird.

5 Effektivität der TCP/IP-Protokolle über ATM

Die Effektivität TCP/IP-Protokolle über ATM zu übertragen wird oft angezweifelt. Dies geschieht nicht unberechtigterweise, da TCP/IP für geringe Übertragungsraten entwickelt wurde. Zusätzlich rufen Echtzeitanwendungen, mittels IP-Protokolle über ATM-Netzwerke übertragen, weitere Probleme hervor. Dabei sind verschiedene Punkte relevant, die die Bandbreite oder Performance empfindlich begrenzen können. Ergebnisse von Messungen in verschiedenen ATM-Pilotprojekten zeigen, daß die Sende- und Empfangsleistung von TCP/IP-Paketen beispielsweise sehr schlecht ist, wenn Überlastungen des Netzes durch zu kleine ATM-Switch-Puffer und zu große TCP-Segmente sowie Fenstergrößen (Windows-Sizes) entstehen. Weiterhin wird oft der zusätzliche Overhead als kritisch gesehen sowie die Rechenleistung der Endgeräte.

5.1 Protokoll-Overhead

Eine Anwendung, die Daten über ein Netzwerk überträgt, kann nicht die volle Bandbreite des verwendeten Kommunikationsmediums ausnutzen. Der Overhead, welcher für den Transport der Daten benötigt wird, verbraucht immer einen gewissen Anteil der Bruttoübertragungsrate. Wenn noch zusätzliche Protokollstrukturen umgesetzt werden müssen, die auf höheren Schichten funktionieren, summieren sich die unterschiedlichen Overheads auf Kosten der Nettoübertragungsrate weiter auf. Das heißt, jede Protokollschicht fügt den übertragenden Daten einen Header und/oder Trailer hinzu. Zur Überprüfung der Protokollzuverlässigkeit werden ebenfalls noch weitere Felder, wie HEC, AAL-5-CRC und IP-Checksummen, angehängt. Für Multiplex-Funktionen sind ebenfalls zusätzliche Felder, wie ATM-VPI/VCI, IP-Quellen- und Empfängeradresse sowie UDP/TCP-Portnummern, vorhanden. Alle Felder zusammen besitzen dann noch Längeninformationen der Protocol Data Unit (PDU) und Verbindungszustände. Die Frage bei der gesamten Umsetzung der Protokoll- und Kontrollstrukturen ist, wie effektiv letztendlich noch

eine Datenübertragung bei kleiner werdender Nettobitrate ist.

Tabelle 3 zeigt die jeweiligen Schichten auf, die bei der IP-Übertragung über das ATM-Netzwerk verwendet werden. Dabei wurde von der Vielzahl der Physical Layer Interfaces die SDH-Schnittstelle mit 155-MBit/s-Übertragungsrates ausgewählt, da diese die meist verwendete WAN-Transportart darstellt. Die STM-1-Rahmenstruktur besitzt drei Overhead-Arten, die aus dem Regenerator Section Overhead, Multiplex Section Overhead sowie dem Path Layer Overhead bestehen. Der STM-1-Rahmen überträgt einen Rahmen der Größe 2430 Byte alle 125 µs mit einem Overhead von 90 Byte. Dadurch entsteht aufgrund der Bruttoübertragungsrates von 155,520 MBit/s ein Gesamt-Overhead von 5,760 MBit/s, wodurch die Nettoübertragungsrates noch 149,760 beträgt. Tabelle 3 vergleicht drei verschiedene MTU-Größen: 576 Byte (Standardgröße für das Internet), 9180 Byte (Festgelegte Standardgröße für IP-Übertragung über ATM) und 65527 Byte (Maximale IP-Paketgröße über die AAL-Typ-5-Schicht)

Hieran ist zu erkennen, daß bei einer MTU-Größe von 576 Byte der Overhead deutlich größer ist, als bei den anderen beiden Varianten. Da die LAN-Emulation nur einen maximalen MTU-Wert von 1500 Byte erreichen kann und ein zusätzlicher Header erforderlich ist, ist dieses Verfahren ineffizienter als beispielsweise Classical-IP mit 9180 Byte. Weiterhin verbessert eine Vergrößerung der MTU-Werte über 9180 Byte nicht mehr deutlich die Nettoübertragungsrates. Aus diesem Grund wurde 9180 Byte auch als Default-Maximum-Transmission-Unit (D-MTU) festgelegt. Durch diese Begrenzung werden die Übertragungsfehler geringer gehalten, da der Verlust von einzelnen ATM-Zellen den Verlust des gesamte AAL-Typ-5-Paketes verursachen würden. Die maximal mögliche Größe von 65527 Byte über die Anpassungsschicht 5 ist deshalb nur ein theoretischer Wert, der in der Praxis nicht angestrebt werden sollte.

Protokollschichten	OC-3c-Schnittstelle		
	576 Byte	9180 Byte	65527 Byte
Übertragungsgeschwindigkeit (PHY)	155,520 MBit/s		
ATM-Schicht	149,760 MBit/s		
Anpassungsschicht Typ 5	135,632 MBit/s		
Maximum Transport Unit (MTU)	576 Byte	9180 Byte	65527 Byte
LLC/SNAP-Einkapselung	126,937 MBit/s	135,220 MBit/s	135,563 MBit/s
IP-Protokoll	125,198 MBit/s	135,102 MBit/s	135,547 MBit/s
Transportschicht	120,851 MBit/s	134,808 MBit/s	135,506 MBit/s
Anwendungsschicht über TCP	116,504 MBit/s	134,513 MBit/s	135,464 MBit/s
Anwendungsschicht über UDP	119,112 MBit/s	134,690 MBit/s	135,489 MBit/s
Σ Overhead bei TCP/UDP	39,016 / 36,408 MBit/s (25,1 / 23 %)	21,007 / 20,830 MBit/s (13,5 / 13,4 %)	20,056 / 20,031 MBit/s (12,9 / 12,9 %)
(in Prozent)			

Tabelle 3: OC-3c-Interface-Overhead

5.2 Durchsatzverminderungen

Um die Datenrate bei TCP/IP-Verbindungen über ATM feststellen zu können, sind Messungen oder Simulationen nötig, die in europäischen ATM-Pilotprojekten gegenwärtig durchgeführt werden (z.B. EXPERT). Dabei haben sich die folgenden Faktoren als störend herausgestellt, wenn nicht die richtigen Einstellungen vorgenommen werden:

- Socket-Puffergrößen am Sender und Empfänger
- Netzwerk: Maximum Transport Unit (MTU)
- Protokoll: Maximum Segment Size (MSS)
- Sender: Verwendung des Nagle-Algorithmus
- Round-Trip-Time (RTT)
- Empfänger: verzögerte Quittungsmechanismen
- Sender: Silly Window Syndrome (SWS)
- Kopierstrategie an der Socketschnittstelle
- Netzverluste und Verlusterkennung

TCP ist ein bidirektionales Protokoll, welches sich verbindungsorientiert verhält. Beim Aufbau einer Verbindung über ATM beansprucht TCP deshalb einen virtuellen Kanal. Um zu verhindern, daß Datensegmente wegen zu kleiner Datenpuffer oder Überlastung des Empfängers nicht verarbeitet werden können, ist der Empfänger in der Lage den ankommenden Datenfluß vom Sender zu begrenzen. Dabei sind die Gründe, daß höhere Protokolle oder der Empfänger die TCP-Segmente nicht so schnell verarbeiten können, für eine Verminderung des Datenflusses ausschlaggebend. Dies liegt an nicht ausreichend dimensionierten Empfangspuffern. Sind die Empfangspuffer ausgelastet, werden neu empfangene Datensegmente verworfen. Dadurch wird eine Sendewiederholung nötig, die den

Durchsatz einer Übertragung empfindlich verringern kann. Aus diesem Grund wurde eine zusätzliche Funktion eingearbeitet, die Sliding-Window-Mechanismus genannt wird. Dieser funktioniert so, daß der Empfänger dem Sender die Datenmenge mitteilt, die er verarbeiten kann. Dabei wird der verfügbare Speicherbereich des Empfangspuffers zur Ermittlung der TCP-Window-Size verwendet. Bei jeder Verbindung wird deshalb eine Maximum Segment Size (MSS) für das Netzwerk festgelegt, um eine effektive Kommunikation gewährleisten zu können. Die MSS-Größe legt dabei die maximal zulässige TCP-Segmentgröße fest. Diese ist ebenfalls für die Durchsatzraten entscheidend, da zu kleine MSS-Werte das Netzwerk höher belasten würden. Normalerweise wird dabei die MSS durch das Abziehen des TCP/IP-Headers von der Maximum Transmission Unit (MTU) des Netzwerks gebildet. Das heißt, aufgrund der Festlegung auf eine MTU von 9180 Byte und einer TCP/IP-Header von 40 Byte, ergeben sich 9140 Byte für den MSS-Wert. Somit kann der Teilnehmer an einer Sitzung den Inhalt der Window Size ohne Bestätigung an andere Benutzer schicken. Die Fenstergröße wird anschließend immer weiter geschlossen und erst durch eine Quittierung von empfangenen Datenbytes wieder geöffnet. Durch diesen Mechanismus ist ein höherer Datendurchsatz möglich, da nicht nach jeder Datensendung auf eine Empfangsbestätigung gewartet werden muß.

Durch den Window-Mechanismus kann es allerdings auch zu Durchsatzverminderungen kommen, indem zu kleine Fenstergrößen den Datenverkehr zu groß werden lassen. Dadurch würde das Netzwerk einer zusätzlichen Belastung ausgesetzt, die teilweise eine Kommunikation unmöglich werden lassen. Dieses Phänomen wird als Silly Window Syndrome (SWS) bezeichnet. Um diesen Engpaß zu verhindern, muß jeweils am Empfänger und Sender ein sogenannter Umgehungsalgorithmus implementiert sein. Dieser Silly Window Avoidance-Algorithmus vergrößert das Empfangsfenster, wenn mindestens $\frac{1}{4}$ des Empfangspuffers frei ist oder ein maximal großes Segment empfangen werden kann.

Eine weitere Durchsatzratenverminderung kann der Nagle-Algorithmus hervorrufen, der eigentlich zur Verhinderung von kleinen Segmentgrößen entwickelt wurde. Bei dem Transport geringer TCP-Segmente würde der Overhead zu groß ausfallen, wodurch die effektive Bandbreite deutlich herabgesetzt wird. Deshalb verhindert dieser Algorithmus das Senden von Segmenten, die kleiner als die zugeordnete MSS der Verbindung sind, wenn irgendeine vorherige Datenübermittlung unbestätigt bleibt. Aus diesem Grund ist es eigentlich notwendig den Nagle-Algorithmus auszuschalten, wenn nur kleine Datenmengen transportiert werden oder eine Übertragung im Simplex-Modus, d.h. ohne Rückmeldung, erfolgt. Da bei Hochgeschwindigkeitsnetzen die D-MTU 9180 Byte beträgt, sind die kleinen Segmentgrößen kleiner als 9180 Byte. Für den Nagle-Algorithmus wären bereits unter normalen Umständen 9180 Byte ein kleiner Segmentwert, da ja 65535 Byte maximal zulässig sind. Aus diesem Grund kann es auch durch diesen Algorithmus zu Performance-Verlusten kommen.

Weiterhin sind auch nach einem erfolgten Verbindungsaufbau Quittierungsmeldungen (Acknowledgements) von dem Empfänger an den Sender nötig. Im Falle von verlorenen TCP-Segmenten wird eine Retransmission-Time-Out-Funktion am Sender gestartet, die die verlorenen Daten erneut überträgt. Zusätzlich werden alle 200 ms zyklisch Quittierungsmeldungen konstruiert, wenn Daten während der Übertragung keine Bestätigungen vom Empfänger enthalten. Das bedeutet, daß der effektive Datendurchsatz weiter durch die zusätzlichen Quittierungsmeldungen vermindert wird. Bei erfolgter Bestätigung des Empfängers werden nur alle 500 ms Acknowledgements verschickt (RFC-1122). Wenn diese Meldungen für längere Zeit ausbleiben, wird die Verbindung abgebrochen. Weiterhin hängt die Round-Trip-Time (RTT), d.h. die Zeit, die ein TCP-Segment vom Sender zum Empfänger und wieder zurück benötigt, von dem Time-Out-Wert ab. Wenn der Time-Out-Wert zu gering ist, kann das bei geringen RTT-Schwankungen eine erneute Übertragung bedeuten. Falls der Time-Out-Wert allerdings zu groß gewählt wird, kann es zu Segmentverlusten kommen, die ebenfalls den Datenfluß negativ beeinflussen würden. Die Abschätzung des richtigen RTT-Wertes stellt deshalb einen kritischen Bereich der TCP-Performance dar.

Bei UNIX-Systemen agiert weiterhin eine Socket-Schicht als Schnittstelle zwischen den Anwendungen und den Protokollen des UNIX-Kernels. Das heißt, es wird eine einheitliche Schnittstelle (Application Programming Interface - API) zur Programmierung von Netzwerkanwendungen zur Verfügung gestellt. In die Sockets implementiert sind dabei eine ganze Reihe von Kernel-Routinen, die zum Aufbau einer Verbindung zwischen zwei Rechnern und zur Datenübertragung benötigt werden. Zusätzlich kann das System dadurch verschiedene Verbindungsarten und Anwendungsprozesse erkennen. Für die Datenspeicherung sind für den Socket zwei Datenpuffer definiert worden. Der eine ist für Daten, die zum Kernel gelangen (Write) implementiert worden und der andere für Daten, die zu der Anwendungsschicht (Read) transportiert werden. Der Benutzer kann dabei eine maximale Byteanzahl festlegen. Der Anwendungsprozeß wird abgebrochen, falls die Socket-Schicht nicht in der Lage ist die gesamten Anwendungsdaten in den Write-Puffer zu schreiben. Dieses blockiert den Ablaufprozeß, welcher erst wieder startet, wenn Pufferspeicher wieder verfügbar ist.

5.3 Tuning der Endsysteme

Leistungsfähigere Rechner wie DEC, Alpha oder Ultra-SPARC können heute bereits Datenraten von bis zu 135 MBit/s auf einer STM-1/OC-3c-Strecke realisieren, womit fast die gesamte theoretisch mögliche Bandbreite ausge-

nutzt wird. Die restlichen 20 MBit/s sind zusätzlicher TCP/IP-Overhead und den ATM-Header mit dem SONET/SDH-Rahmen. Allerdings sind Endsysteme aufgrund der vorherrschenden LAN-Netzwerke für Hochgeschwindigkeitsnetze oft falsch konfiguriert. Um die Vorteile der erhöhten Bandbreite effizient ausnutzen zu können, müssen besonders die Endsysteme korrekt eingestellt werden. Dabei lassen sich folgende Hauptkriterien ableiten:

- **Maximum Transmission Unit (MTU):** Der MTU-Wert begrenzt die Rahmengröße eines Systems. Das heißt, angeschlossene Endsysteme müssen den gleichen MTU-Wert verwenden. Für Classical-IP ist 9180 Byte vorgeschrieben, während die LAN-Emulation nur 1500 Byte verwendet. Diese Werte sollten auf jeden Fall unverändert bleiben, da sie zwar die Performance beeinflussen, aber auch standardisiert sind.
- **Path Maximum Transmission Unit:** Dieser Wert legt die Minimumgröße des MTU-Wertes für alle Netzwerke fest, die sich zwischen zwei kommunizierenden Endsystemen befinden. Der Minimalwert definiert das größte IP-Paket, welches ohne Fragmentation gesendet und empfangen werden kann. RFC-1191 spezifiziert eine Methode, um diesen Wert festzustellen.
- **TCP-Sende- und Empfangspuffer:** Die Fenstergröße wird durch die Puffergröße bestimmt. Um die bestmögliche Performance zu erhalten, muß der gewählte Wert mindestens größer als der doppelte MTU-Wert sein. Der Maximalwert beim Betriebssystem Solaris 2.5 und Windows95/NT liegt bei 64 KByte.
- **Maximum Segment Size (MSS):** Der MSS-Wert begrenzt die TCP-Segmentgröße. Der MSS-Wert wird durch das Abziehen des TCP/IP-Headers von der Maximum Transmission Unit (MTU) des Netzwerks gebildet. Bei Remote Networks ist der Standardwert 536. Hingegen sollten 1436 Byte bei Hochgeschwindigkeitsnetzwerken eingestellt werden, um die optimale Leistungsfähigkeit zu erreichen. Dieser Wert wird allerdings bei geringeren Übertragungsraten zu Performance-Einbußen führen.
- **Window Scale Option:** Der Standard RFC-1323 sieht für eine weitere Verbesserung der Leistungsfähigkeit eine Option vor, um die Fenstergröße auf einen Wert von maximal 2^{30} Byte einzustellen. Diese Option gilt bidirektional und wird beim Verbindungsaufbau vereinbart. Enthalten ist ebenfalls eine Timestamp-Option, die von beiden Kommunikationspartnern unterstützt werden muß und den TCP-Header um 12 Byte erhöht. Durch diese Funktion wird der RTT-Wert durch den Sender erhöht, so daß es zu weniger kritischen Time-Outs kommen kann.

6 Aussichten

Die Arbeitsgruppe ION der IETF arbeitet an einer Erweiterung von Classical-IP. Hauptziel ist es, eine ATMARP-Serverliste einzuführen, welche es gestattet mehrere ATMARP-Server in einem LIS-Subnetz aufzubauen. Dadurch lassen sich redundante Strukturen realisieren. LANE Version 2.0 wird ebenfalls weiter durch das ATM-Forum vorangetrieben. Die Schnittstellen zwischen Client und Server (LAN-Emulation User Network Interface – L-UNI) und der Schnittstelle zwischen den Servern (LAN-Emulation Network Network Interface – L-NNI) werden darin abschließend spezifiziert. Die L-UNI soll kompatibel zur bisherigen Version bleiben und weitere Funktionen enthalten:

- ELANs können über ATM-Verbindungen gemultiplext werden.
- Verbesserte Unterstützung von Multicast-Paketen
- direkte Verbindungen über mehrere Subnetze sollen möglich sein
- QoS soll ebenfalls eingeschränkt unterstützt werden
- LECS wird zu einem allgemeinen Konfigurationsserver umfunktioniert

L-NNI wird in der Version 2.0 zum erstenmal vorliegen. Hauptziel ist es verteilte Server zu ermöglichen und die Kommunikation zwischen diesen zu verbessern. Dadurch können redundante ELANs aufgebaut werden, der BUS begrenzt das Gesamtsystem nicht weiter und unterschiedliche Hersteller sollten einsetzbar sein.

MPOA bietet langfristig gesehen die einzige Möglichkeit ATM-Vorzüge direkt auf andere Netze zu übertragen. Dabei nimmt die Idee des virtuellen Routers immer konkretere Strukturen an. Durch ihn wird es möglich Routerfunktionen auf das Netz zu verteilen, um dadurch Router-Engpässe zu vermeiden. Zentrale Routerserver steuern dann verteilte Edge Devices (Randgeräte), die die Schnittstelle zu den traditionellen LAN-Netzen darstellen. Ein entgegengesetzter Ansatz beinhaltet den Einsatz von Multilayer-Switches. Dieser Ansatz verlegt die Routing-Funktionalität in einzelne Switches, die in gemischter Ethernet/ATM-Umgebung das Routing (Schicht 3) und Ethernet-Switching (Schicht 2) übernehmen. Beachtet werden müssen auch Unterschiede zwischen ATM- (P-NNI) und IP-Routing. In diesem Bereich sind deshalb die Standardisierungsgremien bemüht für große NBMA-Netze Lösungen zu entwickeln.

IP-Switching ist ein weiterer Ansatz um eine Kombination zwischen ATM-Switch und IP-Router zu verwirklichen. Es sind viele Vorschläge vorhanden, die aber alle sehr unterschiedlich ausfallen. Allen gemeinsam ist der Routerbereich, der einem traditionellen Router ähnelt. Zusätzliche Funktion ist die Erkennung von IP-Paketflüssen. Das heißt, es werden die IP-Pakete nicht mehr einzeln eingelesen (Store-and-Forward-Prinzip), sondern im ATM-

Bereich geswitched. Dieser sehr jungen Technologie wird eine vielversprechende Zukunft bescheinigt. Sie wird sich allerdings erst durchsetzen, wenn sie mit LANE oder Classical-IP kooperiert und diese nicht als Konkurrenz betrachtet. Auch ist IP-Switching bisher eine proprietäre Lösung, die zwar auch bei anderen Firmen Unterstützung findet, aber erst durch Standardisierung eine echte Alternative zu bestehenden IP-over-ATM-Lösungen wird. Dabei bleibt abzuwarten, wie IP-Switching letztendlich in die ATM-Technik integriert wird.

Letztendlich sollte das MPOA-Verfahren konsequent weiter verfolgt werden, da hier die meisten positiven Ansätze zu verzeichnen sind. Aufgrund der Komplexität des Standards werden ATM-Forum und IETF aber noch einige Zeit für die endgültige Spezifikation benötigen.

7 Weiterführende Literaturliste

- [1] Alles, A.: *ATM Internetworking*; ATM Product Line Manager; Cisco Systems, Inc.; Engineering InterOp.; Las Vegas; März 1995
- [2] ATM-Forum-Spezifikation: *Private Network-Network Interface Specification*; Version 1 (PNNI 1.0); 1996
- [3] ATM-Forum-Spezifikation: *ATM User Network Interface (UNI) Signalling Specification*; Version 4.0; 1996
- [4] ATM-Forum-Spezifikation: *LAN Emulation Over ATM*; Version 1.0, 1995
- [5] ATM-Forum-Spezifikation: *MPOA Baseline*; Version 1.0, 1996
- [6] Borden, M.; Crwaley, E.; Davie, B.; Batsell, S.: *Integration of Real-time Services in an IP-ATM Network Architecture*; Network Working Group; Request for Comments 1821; Category: Informational; IETF 08/1995
- [7] Cavanaugh, J. D.: *Protocol Overhead in IP/ATM Networks*; Minnesota Supercomputer Center Inc.; ARPA; Contract Number F19-628-92-C-0072; USA 1994
- [8] Detken, K.-O.: *IP über ATM: Pakete huckepack*; Gateway 03/97; Verlag Heinz Heise; Hannover 1997
- [9] Detken, K.-O.: *Virtueller Router – Multiprotokolle über ATM, Teil 1*; Gateway 04/97; Verlag Heinz Heise; Hannover 1997
- [10] Detken, K.-O.: *Dienste mit Qualität – Multiprotokolle über ATM, Teil 2*; Gateway 05/97; Verlag Heinz Heise; Hannover 1997
- [11] Detken, K.-O.: *Kopflast als Leistungsbremse – Effizienz von IP über ATM*; Gateway 05/97; Verlag Heinz Heise; Hannover 1997
- [12] Heinanen, J.: *NBMA Address Resolution Protocol (NARP)*; Request for Comments 1735; Category: Standard Tracks; IETF 12/94
- [13] Katz, D.; Piscitello, D.: *NBMA Next Hop Resolution Protocol (NHRP)*; Routing over Large Clouds Working Group; Category: Internet Draft; IETF 05/1995
- [14] Laubach, M.: *Classical IP and ARP over ATM*; Request for Comments 1577; Network Working Group; Category: Standard Track; Hewlett-Packard Laboratories; IETF 01/94
- [15] Luciani, J.; Katz, D.; Piscitello, D.: *NBMA Next Hop Resolution Protocol (NHRP)*; Category: Internet Draft; IETF 10/96
- [16] Murakami, K.; Maruyama, M.: *MAPOS - Multiple Access Protocol over SONET/SDH*; Version 1; Category: Informational; IETF 06/97
- [17] Murakami, K.; Maruyama, M.: *IPv4 over MAPOS*; Version 1; Category: Informational; IETF 06/97