Seminar Breitband-Netztechnologien und Multimedia-Systeme SS96
Technische Universität Berlin
Fachbereich Elektrotechnik

IP over ATM

Jörn Duckrow, Matrikelnr. 143457
(ducki@cs.tu-berlin.de)

Betreuer: Dipl.-Ing. B. Rathke

Juli 1996

Abstrakt

Seit der Entstehung vor fast dreißig Jahren hat die Internet-Protokollsuite eine weltweite Verbreitung erreicht, die alle ursprünglichen Erwartungen übertrifft. Seit Anfang der 90er Jahre nimmt im Internet, anfänglich für Forschungszwecke gedacht, durch die fortschreitende Einführung von multimedialen Anwendungen und Online-Diensten der Anteil kommerzieller und privater Nutzer ständig zu. Durch diese Entwicklung wächst ständig der Bedarf an Übertragungsbandbreite. Die Engpässe bei der Übertragung existieren sowohl im Weitverkehrsbereich, durch den gebündelten IP-Verkehr vieler Nutzer, als auch im LAN-Bereich, durch neue immer bandbreitenintensivere Anwendungen. Als die Lösung wird die ATM-Technologie gesehen. ATM ermöglicht hohe Bandbreiten und eine gute Skalierbarkeit im WAN- und im LAN-Bereich.

In dieser Studie soll gezeigt werden wie IP-Datenverkehr in einem ATM-Netzwerk ermöglicht wird, welche Standards existieren, welche Ergebnisse bereits erzielt wurden und welche Probleme in der Zukunft noch gelöst werden müssen.

Inhaltsverzeichnis

1 Einleitung

Das Internet ist das derzeit größte Netzwerk der Welt. Das Internet-Protokoll (IP) war mit Mitte der 70er Jahre schon sehr früh verfügbar und ist im Vergleich zu anderen Protokollen recht einfach zu implementieren. Viele Betriebssysteme, enthalten schon standardmäßig eine IP-Implementierung. IP bietet einen verbindungslosen, ungesicherten und asynchronen Paketvermittlungsdienst. Die Länge der Pakete und derer Header sind variabel. Um nicht auf neue Netzwerk-Protokolle umsteigen zu müssen, ist eine Anpassung des IP-Stacks an derzeit modernste Netzwerkarchitekturen wie ATM notwendig.

Protokolle in Hochgeschwindigkeitsnetzwerken ("Gigabit-Netzwerke") müssen spezielle Anforderungen erfüllen. Die existierenden Gigabit-Netzwerke sind durch die Entstehung neuer Anwendungen, wie Ein- und Ausgabegeräte für Hochgeschwindigkeitsrechner, Multimedia-Konferenzen, Video on Demand, Virtual Reality Anwendungen für Simulationen, u.v.m., notwendig geworden. Der *Asynchronous Transfer Mode* (ATM) ist aus verschiedenen Gründen die zukunftsträchtigste Netzwerkarchitektur. Der Hauptgrund dafür sind wahrscheinlich die möglichen hohen Übertragungsgeschwindigkeiten von 155 Mbit/s bis 622 Mbit/s und mehr. Ein weiterer Vorteil von ATM ist die (derzeit noch nicht abgeschlossene) Standardisierung durch die International Telecommunication Union (ITU) und das ATM-Forum. Außerdem können mit einem einzigen Protokoll alle möglichen Telekommunikationsdienste (Übertragung von Daten, Sprache und Bewegtbildern) in lokalen und weltweiten Netzen realisiert werden.

Die Nutzung von IP über ATM bringt für beide Verfahren Vorteile. Zum einen kann ein allmählicher Übergang von konventionellen zu ATM-basierten Netzen erfolgen, was die Einführung von ATM fördert. Andererseits wird dadurch eine Beibehaltung der Internet-Protokollsuite unter dem Druck stetig steigender Bandbreitenanforderungen ermöglicht [8].

2 Integration von IP über ATM

Im Gegensatz zu IP ist ATM ein verbindungsorientiertes Verfahren zur schnellen Übertragung von kleinen Paketen (Zellen) fester Größe. Die Wegewahl (*Routing*) erfolgt beim Verbindungsaufbau über virtuelle Kanäle zwischen einzelnen Endpunkten. Außerdem wird bei ATM in der Regel auf jede Form von Flußkontrolle und auf Nutzdatenfehlerkontrolle verzichtet, da dies höhere Protokollschichten erledigen. Durch die zeitgemultiplexte Übertragung der Zellen läßt sich die Bandbreite gut skalieren und ermöglicht die Bildung von virtuellen Kanälen und Pfaden (*Virtual Channels/Paths*, VCs/VPs) und die Einführung von Dienstklassen bezüglich der Qualität bei der Datenübertagung (siehe Tabelle 1).

Dienstklasse	Klasse A	Klasse B	Klasse C	Klasse D
Synchronisation	erforderlich		nicht erforderlich	
Kommunikationsart	verbindungsorientiert			verbindungslos
Bitrate	konstant	variabel		
Anpassungsschicht	AAL 1	AAL 2	AAL 3/4 oder AAL 5	

Tabelle 1: Dienstklassen der ATM Anpassungsschicht

Um eine bestehende Anwendung transparent in einem ATM-Netz benutzen zu können, braucht man – gemäß dem OSI-Referenzmodell – eine ATM-Anpassungsschicht (*ATM Adaption Layer*, AAL), in der die Daten der höheren Protokollschichten, z.B. IP-Pakete, verpackt und an die ATM-Schicht weitergegeben werden (siehe Abb. 1). Zusätzlich kann den höheren Schichten weitere Funktionalitäten, wie beispielsweise Fehlererkennung und -behebung, Datenpaket-Segmentierung und Zusammenfügung (*Segmentation and Reassembly*, SAR), Flußkontrolle und Zeitanforderungen, angeboten werden. Die AAL besteht aus der Konvergenzunterschicht (*Convergence Sublayer*, CS) und der SAR-Unterschicht. Je nach Anwendung kann zwischen verschiedenen AAL-Typen gewählt werden (siehe Tabelle 1).

In der ATM-Schicht werden die Zell-Header generiert bzw. extrahiert, die Zellkopf-Fehlererkennung und -korrektur, das Multiplexen und Demultiplexen von Zellen unterschiedlicher Verbindungen auf einen einzigen Zellstrom und das Umsetzen von VCIs und VPIs in ATM-Switches vorgenommen.

Die physikalische Schicht besteht aus der übertragungsmediumabhängigen physikalischen Schicht (*Pysical Medium Dependant*, PMD) und der Übertragungsanpassungsteilschicht. Da ATM nicht an ein bestimmtes physikalisches Medium gebunden ist, kann man z.B. über E3 (34 Mbit/s), SDH-STM-1 (155 Mbit/s), TAXI (100 oder 140 Mbit/s) oder andere ATM-Zellen übertragen. Die PMD-Teilschicht umfaßt die korrekte Übertragung und den korrekten Empfang der Bits auf dem entsprechenden Medium (optisch oder elektrisch). In der Übertragungsanpassungsteilschicht erfolgt u.a. die Einpassung und Entnahme der Bits in und aus den Übertragungsrahmen und die Protokollkopf-Fehlersicherung (*Head Error Correction*, HEC).

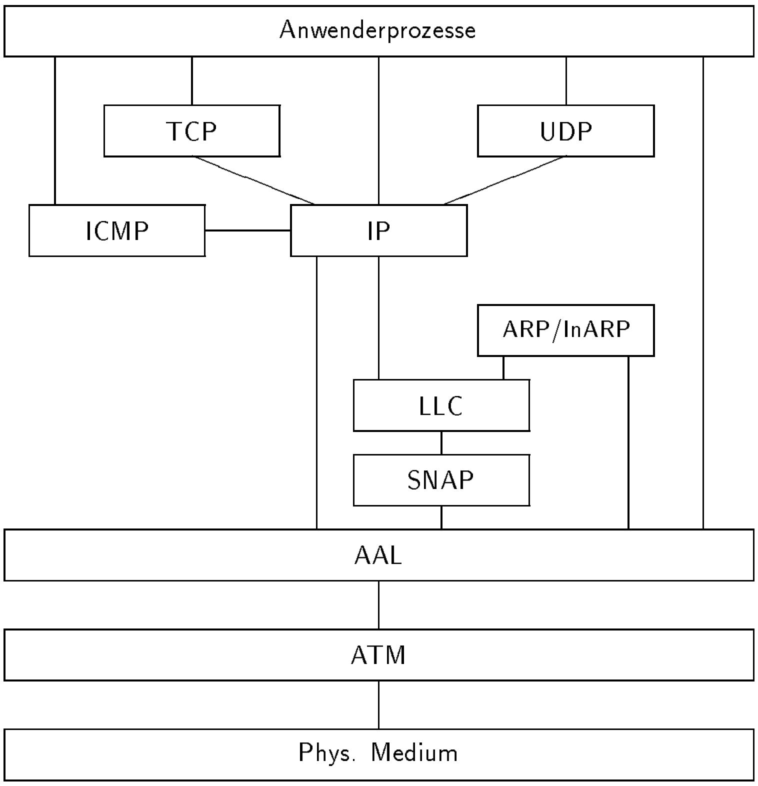

Abbildung 1: Internet-Protokollstack über ATM

3 Classical IP over ATM

Das Modell *"Classical IP and ARP over ATM"* wurde von der IETF im RFC1577 [2] spezifiziert. Ein ATM-Netz wird darin, analog zu einem bestehenden LAN, als ein logisches IP-Subnetz (LIS) definiert. Das Ziel der Spezifikation war es, eine Lösung für die Implementation zum transparenten Versenden von IP- und ATM-Adreßauflösungsprotokoll-Paketen über die ATM-Anpassungsschicht (AAL 5) zu schaffen. Alte LAN-Segmente, wie z.B. Ethernet-, Token Ring- und FDDI-Netzwerke sowie lokale Backbone-Verbindungen, lassen sich nach und nach durch ATM-Technologie ersetzen. Bei der Problematik *IP over ATM* existieren mehrere Problemfelder. Es handelt sich dabei im wesentlichen um *Protocoll Encapsulation*, d.h. die Kapselungsvarianten bei der Anpassung der IP-Schicht an die *Convergence Sublayer* von ATM, um die Wegewahl in LANs und WANs und um die Anpassung der Internet-Adressierung an Adressen im ATM-Netz *(Address Resolution)*.

3.1 Logical IP Subnetwork

In einem ATM-Netz sind, technisch bedingt, nur Punkt-zu-Punkt-Verbindungen möglich, die in einem LIS über virtuelle Kanäle realisiert werden. Ein LIS, wie in Abbildung 4 dargestellt, besteht in der Regel aus ein oder mehreren ATM-Rechnern, mindestens einem ATM-Switch und einem ATMARP-Server. Alle Rechner (Hosts, Router) eines LIS haben die gleiche IP-(Sub)Netznummer und Adreßmaske und sind über den Switch direkt an das ATM-Netz angeschlossen. Jeder ATM-Rechner in einem LIS muß die gleiche MTU-Größe, d.h. die gleiche Länge für die AAL-SDU, benutzen. In RFC1483 [1] wurde als Standardwert eine MTU-Größe von 9180 Bytes vorgeschlagen. Rechner außerhalb des LIS können nur über einen Router erreicht werden. Die Netzwerk-Administration (security, access, routing, filtering) erfolgt unverändert, da bisher noch keine (besseren) ATM-spezifischen Standards dafür existieren.

3.2 IP Paketkapselung über ATM

Bei der Übertragung von IP-Paketen müssen Vereinbarungen über die Kapselung der Daten getroffen werden. Dabei sind mehrere Lösungen möglich. Einige sind in den entsprechenden Internet-Standards RFC1483 [1], RFC1577 [2] und RFC1626 [3] spezifiziert worden. Da IP einen verbindungslosen, ungesicherten und asynchronen Paketvermittlungsdienst bietet, kommen nur die AAL-Typen 3/4 und 5 in Frage.

Es gibt zwei Methoden für die verbindungslose Übertragung über ein ATM-Netzwerk. Die erste Methode (*LLC Encapsulation*) erlaubt das Multiplexen von verschiedenen Protokollen über eine einzelne virtuelle Verbindung (*Virtual Circuit*, VC). Die Protokolle der übertragenen PDU werden dabei anhand eines IEEE 802.2 Logical Link Control (LLC) Header identifiziert. Diese Variante wird dort Einsatz finden, wo aus Kostengründen nicht beliebig viele VCs geöffnet werden können (z.B. in öffentlichen ATM-Netzwerken). Bei der zweiten Methode (*VC-basiertes Multiplexen*) werden Protokolle höherer Schichten implizit durch VCs gemultiplext, indem verschiedene virtuelle Kanäle verwendet werden. Diese Variante wird vorzugsweise dort Anwendung finden, wo das Öffnen von VCs schnell und ökonomisch ist [1, 6].

AAL 3/4

Historisch gesehen war AAL 3/4 der erste AAL-Typ, der im LAN- und Campus-Bereich von Hochschulen und anderen Forschungseinrichtungen eingesetzt wurde, um das stark verbreitete IP an ATM anzupassen. Zu dieser Zeit wurden fehlende Standards von der ITU-T oder dem ATM-Forum durch proprietäre Lösungen ersetzt. Die IP-Pakete (inkl. ARP/InARP und ICMP) werden ohne LLC/SNAP-Paketierung an die Konvergenzschicht (*Convergence Sublayer*, CS) weitergeleitet (siehe Abbildung 2). Mit Header und Trailer versehen wird die CS-PDU an die SAR-Schicht

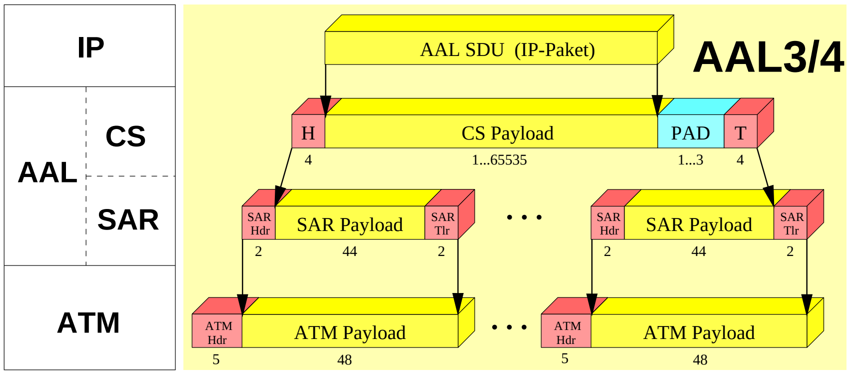

Abbildung 2: Segmentierung mit AAL 3/4

übergeben. Die SAR-PDUs (44 Oktette plus 4 Bytes Header und Trailer) werden in der ATM-Schicht mit dem Zellheader versehen. Der Trailer der SAR-PDU enthält eine 10 Bit lange CRC. Der zweifache Overhead einer AAL 3/4-PDU wirkt sich hier doppelt negativ aus. Zum einen entsteht ein höherer Rechenaufwand zur Verarbeitung der Protokollinformationen und zum anderen werden die ATM-Zellen durch die SAR-Paketierung schlechter ausgenutzt (nur zu ca. 83%).

AAL 5

Der AAL-Typ 5 entstand auf Druck der Computerindustrie, den Protokollaufwand von AAL 3/4 zu verringern. Die Vereinfachung besteht aus dem Verzicht einiger Funktionalitäten in der SAR-Schicht, wie Bitfehlererkennung und -behandlung sowie Multiplexing (siehe Abbildung 3). Das Multiplexen kann, falls gewünscht, von der CS-Schicht erbracht werden. Die Fehlerüberwachung erfolgt nun paketorientiert in der CS-Ebene und nicht mehr zellorientiert in der SAR-Ebene. Das ist deshalb sinnvoll, weil eine fehlerhafte ATM-Zelle die reassemblierte AAL 5-PDU sowieso unbrauchbar macht. Bei AAL 5 werden die IP-Pakete mit einem LLC/SNAP-Header gemäß IEEE 802.2 und IEEE 802.1a versehen, was die Nutzung von verschiedenen Protokollen über AAL 5 durch Multiplexen der Protokoll-ID ermöglicht [1]. Durch den nur noch in der CS-Schicht anfallenden Overhead beträgt die Zellausnutzung bei AAL 5 rund 90%.

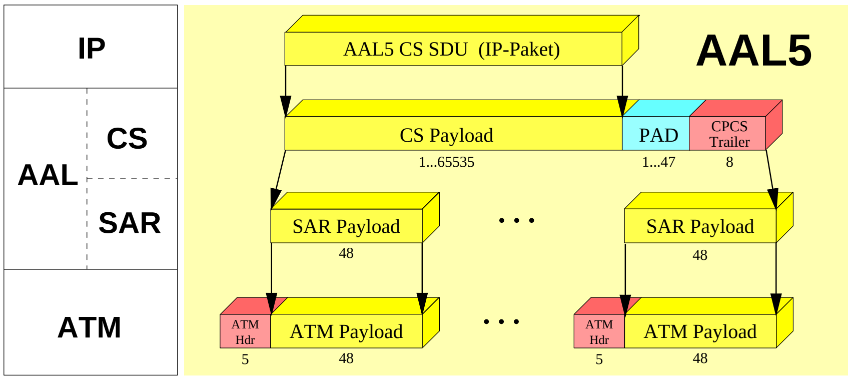

Abbildung 3: Segmentierung mit AAL 5

Im Laufe der Markteinführung von ATM erfuhr der AAL-Typ 5 aufgrund seiner Einfachheit und Effektivität eine größere Verbreitung als AAL 3/4. Durch die zunehmende Standardisierung von Diensten über AAL 5 wurden die proprietären Implementationen wieder verdrängt und durch Standards ersetzt.

3.3 Adressierung

In einem IP-Netz werden die Rechner auf der Vermittlungsschicht über eine 32 Bit lange IP-Adresse angesteuert. In der Sicherungsschicht wird in der Regel eine andere Adressierung benutzt – so auch bei ATM. In der Spezifikation UNI 3.0 des ATM-Forums wurde für private Netzwerke eine Adressierung mit *Network Service Access Points* (NSAP) gemäß ISO8348 oder Standard E.164 der ITU-T oder eine Mischung aus beiden festgelegt (siehe Tabelle 2). Zwischen zwei ATM-Stationen muß ein virtueller Kanal oder Pfad gefunden oder aufgebaut werden. Adressen in privaten Netzwerken beschreiben ein ATM-Endgerät eindeutig. In öffentlichen Netzen erfolgt die eindeutige Beschreibung eines Interfaces entweder durch die nach dem Standard E.164 definierte Adreßstruktur oder durch private Adressen.

ATM Nummer	**ATM Unteradresse**
ATM Forum NSAP	Null
E.164	Null
E.164	ATM Forum NSAP

Tabelle 2: Struktur für private ATM-Adressen

In einem ATM-Netz werden die IP-Adressen, analog zum ARP-Mechanismus nach RFC826, über einen ATMARP-Dienst innerhalb eines LIS aufgelöst. Im Gegensatz zu den alten IP-LANs ist eine broadcastbasierte Adreßumsetzung bei ATM nicht sinnvoll. Für AAL 3/4 wurde noch ein verteilter ARP-Mechanismus implementiert. Mit der Standardisierung des ARP-Dienstes über AAL 5 in RFC1577 wurde eine zentrale Lösung gewählt. Danach ist innerhalb eines LIS ein ATMARP-Server für die Umsetzung der IP-Adressen zuständig. Dafür müssen für alle ATM-Geräte gewisse Parameter existieren. Jede einzelne ATM-Station besitzt eine ATM-Hardware-Adresse `atm$ha` und kennt außerdem die ATMARP Request Adresse `atm$arp-req`, also die Adresse des ATMARP-Servers des LIS, an den die ARP-Anfragen abgeschickt werden. In einem LIS mit festen virtuellen Kanälen (*Permanent Virtual Circuits*, PVCs) wird der ATMARP-Server nicht benötigt, da alle Verbindungen schon manuell konfiguriert sind. Für die umgekehrte Adreßumsetzung wird InATMARP verwendet, das eine an ATM angepaßte Variante des im RFC1293 beschriebenen InARP ist.

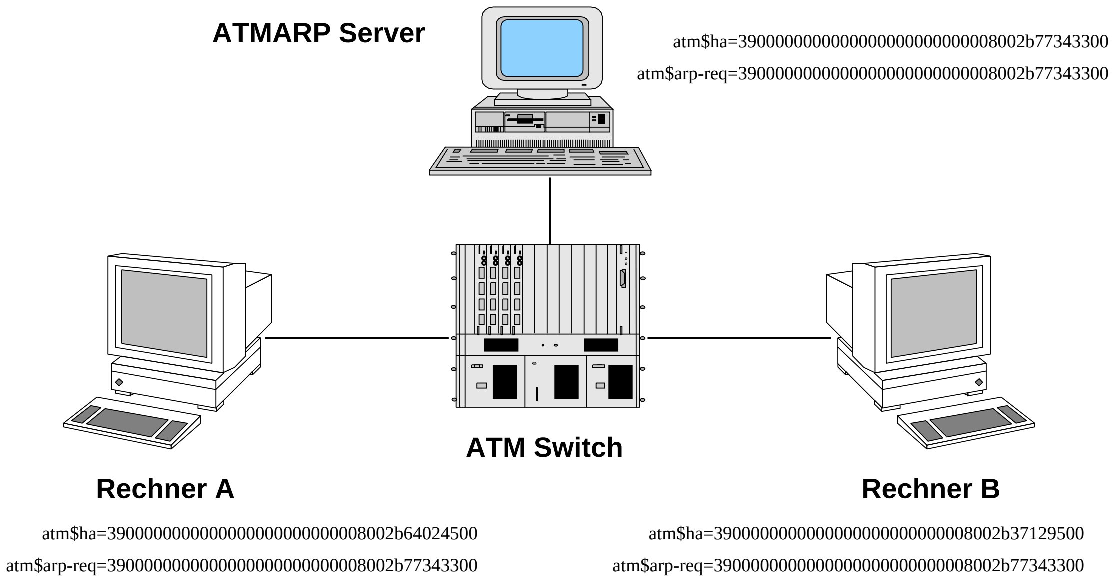

Abbildung 4: Beispiel einer Adressierung im LIS

Im ATMARP-Server werden die IP-Adressen und die ATM-Hardware-Adressen aller im LIS befindlichen Rechner tabellarisch verwaltet. Die Einträge der ATMARP-Tabelle sollten jeweils mindestens alle 20 Minuten aktualisiert werden [2]. Wenn IP-Daten übertragen werden sollen muß die ATM-Adresse des Zielrechners bekannt sein. Unbekannte ATM-Adressen werden vom Rechner beim ATMARP-Server erfragt. Der dabei verwendete Mechanismus zur Signalisierung ist in Kapitel 3.4 beschrieben. Die ermittelten Adressen haben eine maximale Gültigkeit von 15 Minuten und müssen danach per ATMARP-Request erneut ermittelt werden. Durch diesen Mechanismus ist ein LIS flexibel erweiterbar. Alle Rechner außerhalb des LIS sind nur über einen Router erreichbar. Ein Beispiel dafür ist in Abbildung 5 illustriert.

Der dargestellte Adressierungsmechanismus hat den Vorteil, daß er innerhalb eines LIS einfach zu implementieren ist, ohne auf kommende Broadcast- oder Multicastprinzipien zurückzugreifen. Allerdings ist das Adressierungsschema nach RFC1577 nicht skalierbar, d.h. für größere Netzwerke ungeeignet. Ein einziger ATMARP-Server würde dabei einen Flaschenhals bilden und außerdem bei Ausfall ein Betriebsrisiko für das gesamte LIS darstellen. Eine erste Lösung wäre ein verteilter ATMARP-Server. In sehr großen Netzwerken wird sich eine Client-Server-Architektur jedoch nicht mehr als sinnvoll erweisen.

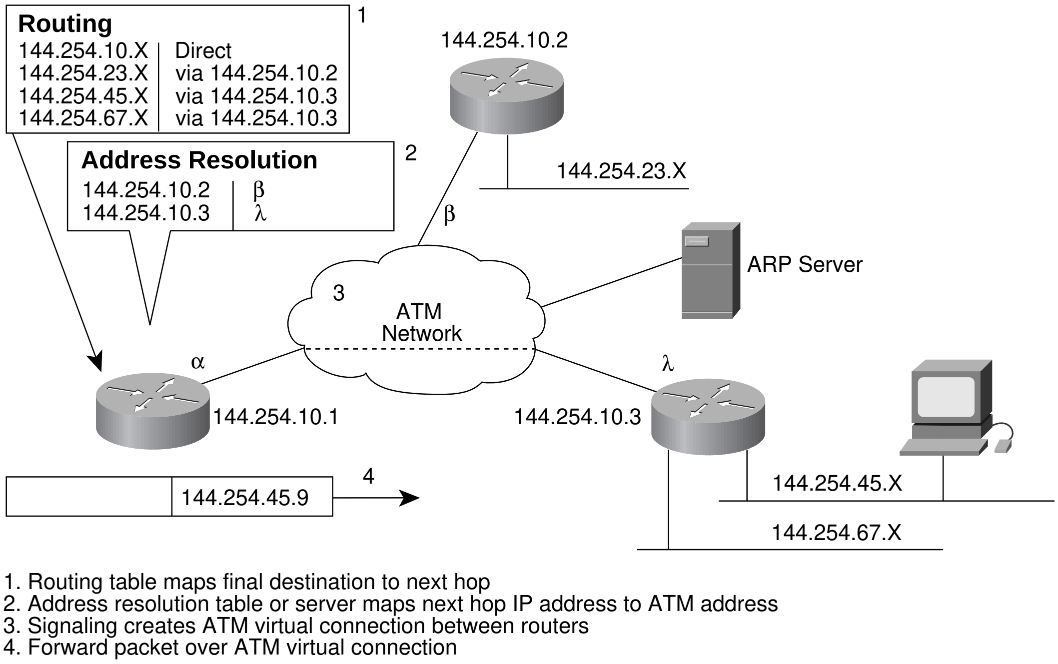

Abbildung 5: Routing über ATM nach dem Klassischen Modell [5]

3.4 Signalisierung

Die im folgenden beschriebenen Signalisierungsprozeduren sind den UNI-Standards 3.0 und 3.1 des ATM-Forums, die wiederum auf den Standards Q.93B und Q.2931 der ITU-T basieren, entnommen und zeigen nur die wichtigsten Signalisierungsabläufe. Mit UNI 4.0 soll ein endgültiger Standard für die Signalisierung in ATM-Netzen geschaffen werden. Zum Zeitpunkt der Erstellung dieser Arbeit war UNI 4.0 noch nicht verfügbar.

Auf- und Abbau von virtuellen Verbindungen

In einem ATM-Netzwerk mit dynamisch erzeugten virtuellen Verbindungen (*Switched Virtual Circuits*, SVCs) müssen diese explizit angefordert werden. Die Daten für die Signalisierung werden in der Regel über einen bekannten und extra dafür reservierten VC gesendet[1]. Soll eine Verbindung aufgebaut werden, schickt die Station eine SETUP-Nachricht mit allen notwendigen Informationen an die Zielstation (siehe Abbildung 6). Dem Sender kann optional von dem Netzwerk oder von der Empfängerstation mit einer CALL PROCEEDING-Nachricht signalisiert werden, daß der Verbindungsaufbau initiiert wird und keine weiteren SETUPs akzeptiert werden. Wird die Verbindung vom Empfänger angenommen, antwortet er mit einer CONNECT-Nachricht oder er lehnt mit einem RELEASE COMPLETE ab. Im letzten Schritt wird die Verbindung durch ein CONNECT ACKNOWLEDGE quittiert. Ein Wunsch zum Verbindungsabbau wird mit einer RELEASE-Nachricht signalisiert und mit RELEASE COMPLETE bestätigt.

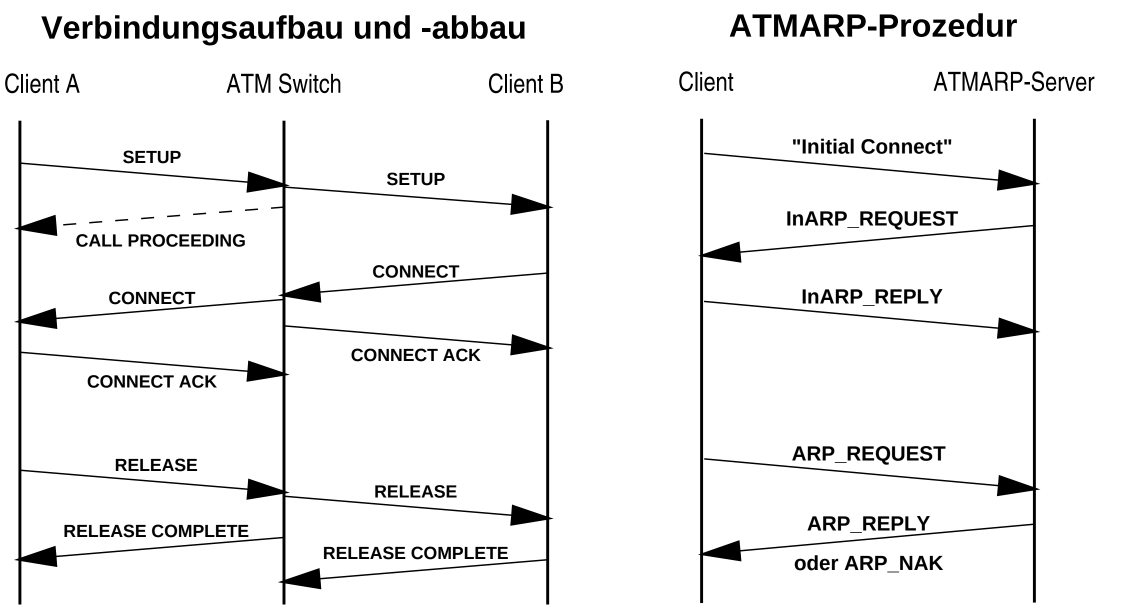

Abbildung 6: Signalisierung bei ATM

ATMARP und InATMARP nach RFC1577

Ein grundlegender Mechanismus für die IP-Kommunikation in ATM-Netzen ist der schon erwähnte ATMARP-Dienst zur Umsetzung der IP- in ATM-Adressen. Die verwendete Signalisierungsprozedur verläuft dabei folgendermaßen. Zuerst muß sich jeder Rechner, z.B. während des Bootvorganges, durch Senden der ATM-Hardware-Adresse beim ATMARP-Server anmelden (initial Connect in Abbildung 6). Der ATMARP-Server erfragt im nächsten Schritt per InATMARP-Request die IP-Adresse des jeweiligen ATM-Rechners, der dann mit einer InATMARP-Reply seine IP-Adresse zurückschickt. Der ATMARP-Server merkt sich das ATM-/IP-Adreßpaar zusammen mit einem Zeitstempel in seiner ATMARP-Tabelle. Sucht ein Klient nach der Adresse eines anderen Rechners, schickt er eine Anfrage (ARP_REQUEST) nach der zu einer IP-Adresse gehörenden ATM-Adresse. Der Server sendet dann entweder per ARP_REPLY die gesuchte ATM-Adresse oder generiert eine Negativantwort ARP_NAK. Die Erweiterung von ARP um die ARP_NAK-Antwort soll dem Klienten eine bessere Erkennung ermöglichen, einen Serverausfall von einer gescheiterten Adreßauflösung zu unterscheiden.

[1] Nach UNI 3.0 werden dabei folgende Standardwerte benutzt: VPI=0, VCI=5.

4 Performance von IP über ATM

4.1 Einführung

Allgemein gültige Aussagen über die Leistungsfähigkeit von Rechnern und Rechnernetzwerken sind möglich. Die Leistungsangaben verschiedener Netzkonfigurationen sind relativ zueinander nur bedingt aussagekräftig. Die Meßergebnisse sind von zu vielen Faktoren abhängig, wie z.B. die Leistung der Rechner, die verwendeten Betriebssysteme bei Rechnern und Routern, der zu verarbeitende Protokolloverhead, die Größe der verwendeten Warteschlangen (Puffer) in den Rechnern und Switches und die Art der Protokollimplementierungen (Hardware- oder Softwarelösungen). Neben den physikalischen Hardware-Eigenschaften spielen auch typische Arbeitsprofile der Benutzer einzelner Anwendungen eine Rolle. Mehrere Nutzer, die während eines Projektes ihre Programme regelmäßig neu kompilieren, erzeugen in einem NFS-basiertem Netzwerk eine wesentlich höhere Netzlast als eine Vielzahl von Nutzern, die Textverarbeitung betreiben und nur gelegentlich ihre Dokumente abspeichern. Aus diesen Beispielen wird deutlich, daß eine pauschale Aussage wie “155 Mbit/s” wenig aussagekräftig ist im Bezug auf die in der Praxis zu erwartende Leistungsfähigkeit – sie gibt eher einen groben Anhaltspunkt. Wie lassen sich also aussagefähige Ergebnisse ermitteln?

Leistungsbewertung

Für die Messungen zur Leistungsbewertung von IP über ATM sind eine Vielzahl von Parametern zu berücksichtigen. Die Datenübertragungsrate gibt an, wieviele Nutzdaten pro Zeit übertragen werden können. Sie kann auf allen Schichten des OSI-Schichtenmodells gemessen werden. Ein weiterer Leistungsfaktor ist die Fehlerrate. Sie gibt das Verhältnis der verlorenen oder fehlerhaften Daten zu der insgesamt übertragenen Datenmenge an. Die Fehlerrate kann unabhängig vom Medium durch bestimmte Fehlerbehebungsstrategien, wie Datenredundanz, verringert werden. Die Zugangsgeschwindigkeit (*Setup Time*) ist hauptsächlich bei Protokollen mit Mediumzugriff (MAC), z.B. Token- oder CSMA/CD-Protokollen, entscheidend. Sie gibt die Zeit an, um den Zugriff auf das Medium zu bekommen. Bei Protokollen, die Kanäle, schalten ist die Verbindungsaufbauzeit die entscheidende Meßgröße. *Network Delay* ist die Laufzeit der Daten im Netz. Die Netzwerkauslastung beschreibt das Verhältnis der gemessenen übermittelten zu der maximal übermittelbaren Datenmenge.

Für die Beurteilung der Leistungsfähigkeit eines Netzes ist die Geschwindigkeit eines sendenden, empfangenden oder routenden Rechners wichtig. Die Leistungsfähigkeit von Rechnern und Routern hängt von mehreren Kriterien ab und ist bei unterschiedlichen Rechnerarchitekturen besonders schwer vergleichbar. Die Geschwindigkeit, mit der Rechner, Router und Switches in der Lage sind, Daten zu kapseln, weiterzuleiten und zu entkapseln, sind wichtige Faktoren. Ein Kriterium für die Beurteilung der Leistungsfähigkeit eines Rechners oder Routers ist seine Rechenleistung in Millionen Instruktionen pro Sekunde (MIPS). Wichtige Parameter für den Router sind ferner der durch das Betriebssystem verursachte Overhead, CPU- und Bus-Geschwindigkeit sowie Speicher- und Puffergröße.

Theoretische Performance

Ein charakteristisches Merkmal von ATM ist die feste Paketgröße. Die sog. Zellen sind mit 53 Bytes vergleichsweise klein und beinhalten mit 5 Bytes einen relativ großen Anteil an Verwaltungsinformationen, der etwa $\frac{5}{53} = 9,43\%$ entspricht. Damit können nur maximal 90,57% der Gesamtbandbreite des Netzwerkes für die Übertragung von Nutzdaten verwendet werden. Zum Vergleich gehen bei Ethernet auf MAC-Ebene nur $\frac{26}{1500} = 1,73\%$ und bei FDDI sogar nur $\frac{28}{4500} = 0,62\%$ für Protokollinformationen verloren (jeweils inklusive Präambel und Frame Check Sequence).

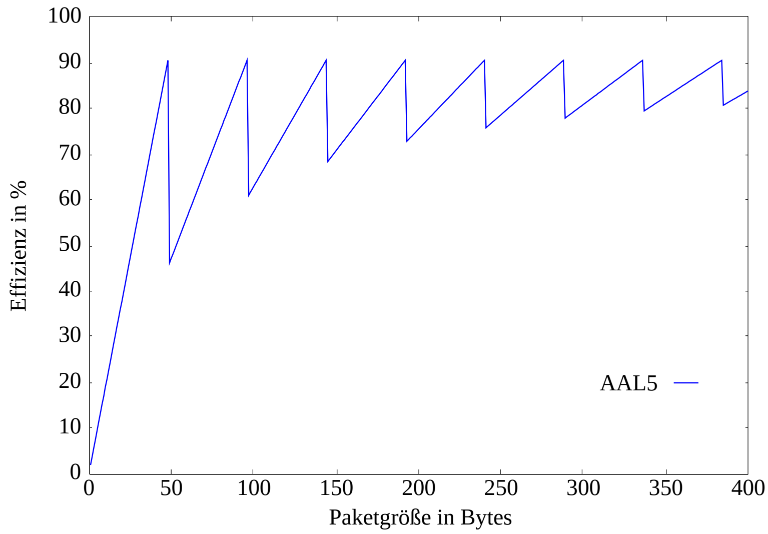

Abbildung 7: Theoretische Effizienz der Zellausnutzung bei kleinen Paketen

Zellen, die nicht voll ausgenutzt werden, enthalten bei AAL 5 je nach Paketlänge 0 bis 47 Füllbytes (*Padding*). Bei kleineren Paketen mit weniger als 512 Bytes liegt deshalb die Ausnutzung zwischen 90 und nur 50% (siehe Abbildung 7). Da kleine Pakete in LANs typisch sind, wirkt sich die schlechte Ausnutzung recht stark auf die Protokolle höherer Schichten aus. Mit `NNstat` wurde in [7] ermittelt, daß 75% der Pakete im PRZ-Netz kleiner als 320 Bytes sind (siehe Abbildung 10).

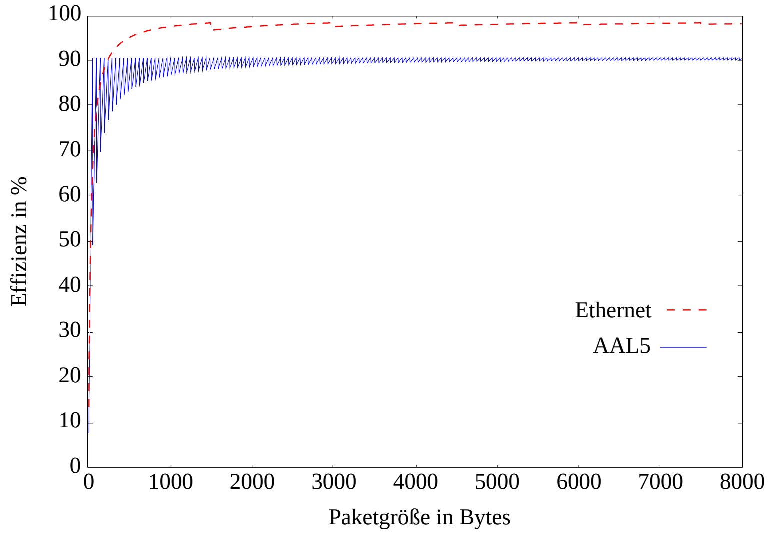

Abbildung 8: Theoretische Effizienz der Zellausnutzung bei größeren Paketen

Bei größeren Datenpaketen wirken sich die Füllbytes in der letzten Zelle kaum noch aus, so daß die Zellen fast optimal mit 90,57% ausgenutzt werden. Das vergleichsweise wesentlich langsamere Ethernet, dargestellt in Abbildung 8, nutzt seine zur Verfügung stehende Bandbreite deutlich besser aus.

Bei der Betrachtung der theoretischen Effizienz von Vermittlungs- und Transportprotokollen über AAL 5 in Abbildung 9 fällt auf, daß die Kurven noch steiler ansteigen und schneller zum Optimum konvergieren. Das kommt daher, daß schon bei einem nur 1 Byte großen Datenpaket für TCP 20 Bytes Protokollinformationen bzw. 8 Bytes für UDP sowie zusätzliche 20 Bytes für den

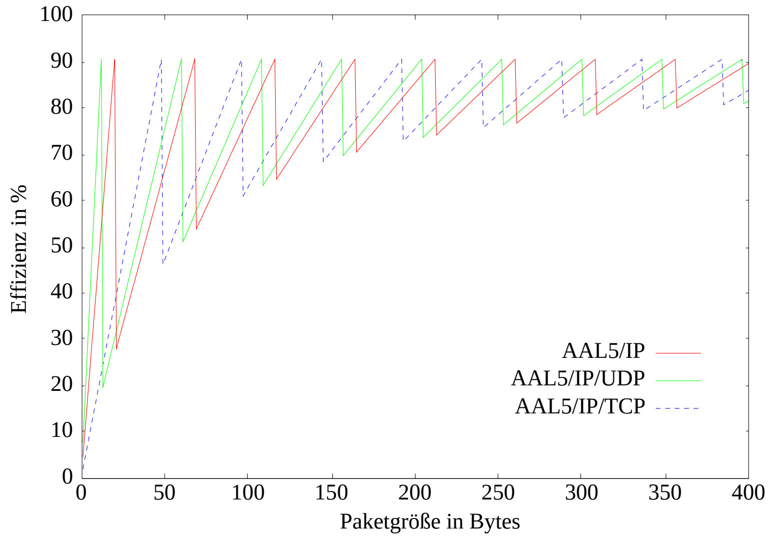

Abbildung 9: Theoretische Effizienz von IP, TCP und UDP über AAL 5

IP-Header anfallen. Bei der Kapselung gemäß RFC1483 fällt noch ein zusätzlicher LLC/SNAP-Header von 8 Bytes an. Bei einem 1 Byte langen TCP-Nutzdatenpaket wird eine 49 Bytes lange PDU übertragen, die dann auf *zwei ATM-Zellen* verteilt werden muß. Bisher kann ein einzelnes Byte lediglich unter Verwendung von (bis heute) noch nicht für ATM standardisierten Kapselungsverfahren mit komprimiertem IP-Header nach RFC1144 (z.B. *Serial Line IP*, SLIP) in einer einzigen ATM-Zelle übertragen werden!

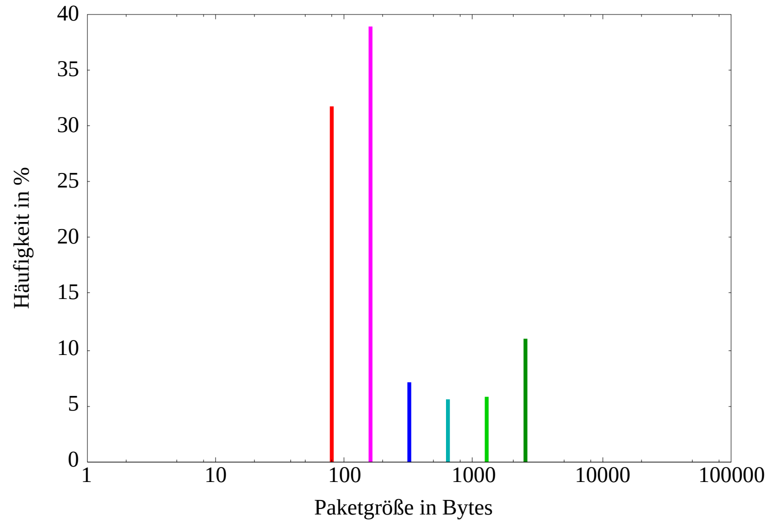

Abbildung 10: Häufigkeiten von TCP- und UDP-Paketgrößen im PRZ-Netz

Typischer IP-Datenverkehr

Um eine Aussage über einen typischen IP-Datenverkehr treffen zu können, wurde in [7] im LAN der Prozeßrechnerverbundzentrale der TU Berlin (PRZ) an einem Router über 3 Wochen der Paketverkehr gemessen. Der verwendete Router verbindet vier Class C Internet Subnetze (2×Ethernet, FDDI, ATM), die insgesamt rund 150 Workstations, PCs und X-Terminals sowie mehrere Server umfassen. Es wurden insgesamt 82.755.777 Pakete gemessen, wobei das kleinste 28 und das größte

Paket 17184 Bytes lang war. Die Verteilung der TCP- und UDP-Paketgrößen[2] ist in Abbildung 10 dargestellt.

Studien zu Weitverkehrsnetzen anderer Autoren [10, 11] zeigen, daß die Häufigkeit von kleinen Paketen im Weitverkehrsbereich noch größer ist. Cáceres hat den IP-Weitverkehr zweier Universitäten und von zwei Industrie-Forschungseinrichtungen untersucht und gab an, daß 84% der TCP-Datagramme kleiner als 256 Bytes und 97% der UDP-Datagramme kleiner als 150 Bytes waren, wobei der Verkehrsanteil von TCP 80% betrug. Die Erklärung dafür ist, daß die hier typischen Anwendungen wie `telnet` und `rlogin` beim interaktiven Arbeiten einzelne Befehle oder gar nur einzelne Zeichen übertragen.

4.2 Gemessene Leistung von *IP over ATM*

Die folgenden Meßergebnisse sind auch aus [7] entnommen und stellen typische Werte verschiedener häufig benutzter Hardwarekonfigurationen dar. Die dabei verwendeten Rechnerkonfigurationen stehen in Tabelle 3. Als Switches kamen ein Fore Systems ForeRunner ASX200 und ein Cisco LS100 zu Einsatz. Beide Switches ermöglichten Verbindungen über Anschlüsse für 100 Mbit/s TAXI und 155 Mbit/s SDH OC3c. Im Versuchsaufbau wurden immer zwei gleiche Rechner über einen Switch verbunden. Die verwendeten ATM-Treiber waren konform zu den RFCs 1483 und 1577. Zum Aufbau der virtuellen Verbindungen wurde bei der Hardware von Fore Systems das *Simple Protocol for ATM Network Signalling* (SPANS) und bei den HP-Rechnern eine Signalisierung nach UNI 3.0 benutzt.

HP750	64 MB RAM, HP-UX 9.01, EISA-Bus, HP J2802 A, 155 Mbit/s SDH OC3c
HP755	96 MB RAM, HP-UX 9.01, EISA-Bus, HP J2802 A, 155 Mbit/s SDH OC3c
SPARCstation 2	16 MB RAM, SunOS 4.1.3, SBUS, Fore Systems SBA100 ATM Adapter, 100 Mbit/s TAXI
SPARCstation 10	32 MB RAM, SunOS 4.1.3, SBUS, Fore Systems SBA200 ATM Adapter, 100 Mbit/s TAXI

Tabelle 3: Bei den Messungen in [7] verwendete Rechnerkonfigurationen

Für die Messungen wurde das Programm `ttcp` verwendet, das im wesentlichen zwei Kommunikationsendpunkte (Sockets) öffnet, Daten an die Netzwerkschnittstelle übergibt und auf der Empfängerseite entgegennimmt. Die Puffergröße für die Sockets wurde auf 32 KBytes gesetzt. Es wurden 40 MBytes Daten in jeweils unterschiedlichen Paketgrößen per TCP und UDP transportiert. Hier sind nur einige UDP/IP-Ergebnisse dargestellt, da der ungesicherte Dienst von UDP im Vergleich zu IP nur unwesentlich mehr Protokoll-Overhead und Bearbeitungsaufwand erzeugt und damit für *IP over ATM* aussagekräftig ist. In einer weiteren Messung wurden modifizierte Kernels benutzt, um bis zu 25000 Bytes lange Pakete übertragen zu können (siehe Abbildung 12).

Für alle Messungen wurden in der Regel drei Durchläufe pro Meßreihe gestartet. An allen Meßpunkten wurde die Standardabweichung σ nach

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2}$$

berechnet, um eine Abschätzung über die Meßgenauigkeit und Reproduzierbarkeit zu bekommen. In den Diagrammen ist jeweils zur Wahrung der Übersichtlichkeit nur die letzte Meßreihe dargestellt.

[2] Ein Balken steht in der Abbildung für ein ganzes Intervall. Die Meßintervalle waren kleiner oder gleich 9, 19, 79, 159, 319, 639, 1279, 2559, 5119, 10239, 20479 und größer als 20479.

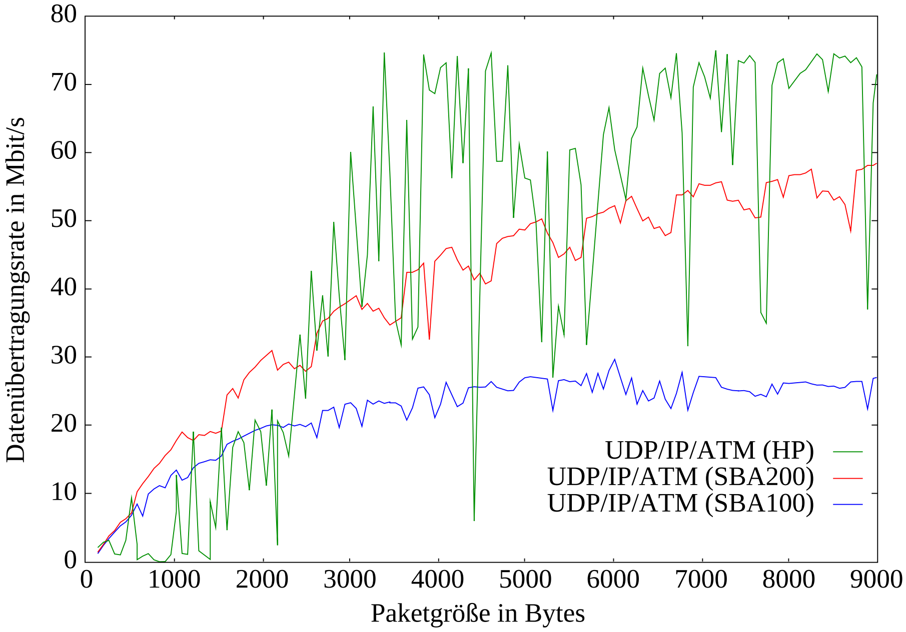

Abbildung 11: Datendurchsatz in Abhängigkeit von der Paketgröße

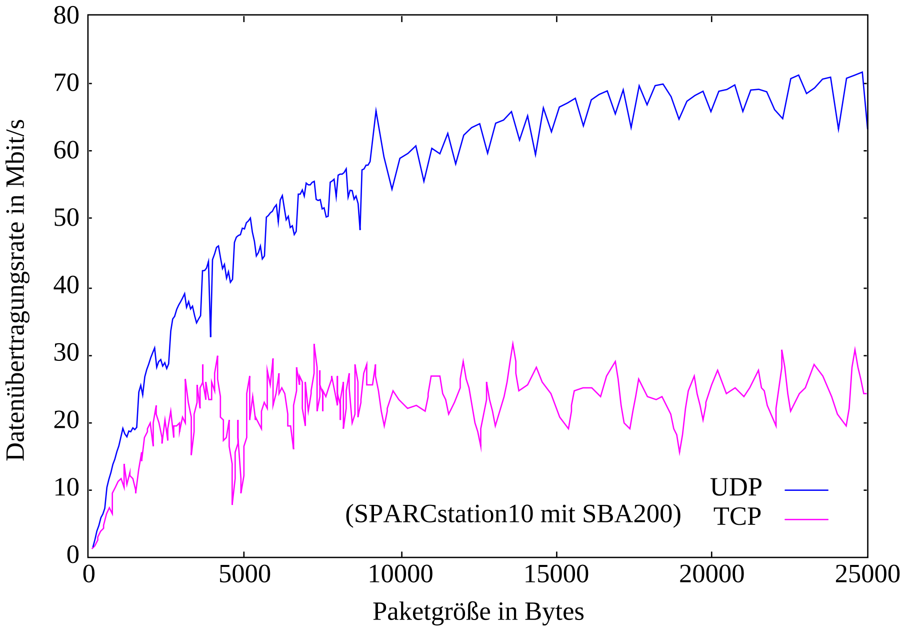

Abbildung 12: Datendurchsatz mit modifiziertem Kernel

Für weitere Tests wurden `netperf`, `ethertop`, `atmstat` und eigens entwickelte Tools verwendet. Die Messungen wurden unter *Zero Load Condition* durchgeführt, d.h. es liefen neben den nötigen Systemprozessen keine weiteren Anwendungen. In [7] wurden hauptsächlich der Datendurchsatz gemessen, aber auch Round Trip Timing (RTT), Netzwerkauslastung sowie Paket- und Bytefehlerraten.

4.3 Performanceprobleme bei *IP over ATM*

Neben den, auf dem Protokollstack zurückzuführenden, Performanceeinbußen gibt es noch weitere Quellen für die Leistungseinbrüche bei der Übertragungsrate: die Verzögerungszeiten im Netzwerk und die Übertragungsfehlerrate bzw. die Anzahl der Paketwiederholungen bei fehlendem Fehlerkorrekturverfahren. Bei ATM gibt es verschiedene Verzögerungszeiten (siehe Tabelle 4). Da die einzelnen Verzögerungen in der Praxis kaum meßbar sind, wird vielmehr die Gesamtverzögerung für den Hin- und Rückweg (*Round Trip Delay*) gemessen.

Übertragungsverzögerung	Zeit, die eine Nachricht zum Durchlaufen des Netzes benötigt.
Kapselungsverzögerung	Zeit, die benötigt wird, um die zu versendenden Daten in Pakete und Zellen umzuwandeln.
Vermittlungsverzögerung	Dauer des Umschaltens eines Kanals
Warteschlangenverzögerung	Wartezeit einer Zelle im Switch, bis alle vorher eingetroffenen Daten weitergesendet wurden.

Tabelle 4: Verzögerungszeiten in ATM-Netzen

Die Werte dieser Parameter sind stark von der verwendeten Technologie, wie z.B. die Art der Protokoll-Implementierung (in Software oder Hardware), abhängig. Die *Bitfehlerrate* der verwendeten Lichtwellenleiter ist mit 10^{-8} bis 10^{-10} sehr niedrig. Bei ATM aber kann sich die Fehlerrate noch erhöhen, wenn in den Switches wegen überlaufender Warteschlangen Zellen verworfen werden. Mit steigender Last kommt es im Switch oder auf den Endsystemen zu *Delay* und fallengelassenen Zellen, die sich wiederum empfindlich auf die höheren Protokollschichten auswirken.

Verluste in Switches

In [7] wurden während der Messungen mit Hilfe eines HP75000 Netzwerk-Analyzers im ForeRunner ASX200 Switch Zellverluste von durchschnittlich 0.012% festgestellt (siehe Abbildung 13). In der Transportschicht wurden Paketverluste zwischen 1 und 6% ermittelt. Im LAN-Bereich werden Werte für Paketverluste und Fehlerraten größer als 0.025% als schlecht bzw. als Indikator für Überlast oder Leitungsdefekt angesehen.

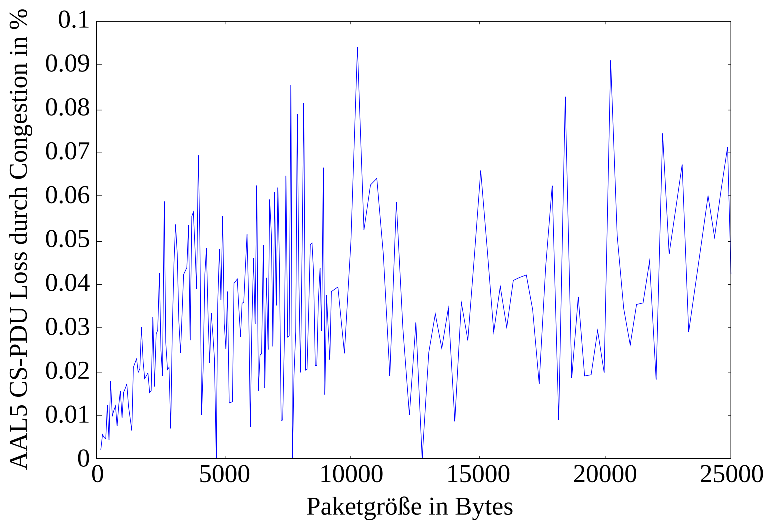

Abbildung 13: Verlust von AAL5-CS-PDUs wegen Network Congestion

Performanceeinbußen durch Zellverluste in Switches sind im allgemeinen die Folge von zu kurzen Warteschlangen oder von stark schwankenden Datenströmen (*bursty Traffic*). Problematisch sind auch Übergänge von schnellen Verbindungen (z.B. SDH-STM-1) auf "langsamere" Verbindungen (z.B. TAXI), in denen bei hoher Last die Warteschlangen relativ schnell überlaufen können. Durch den Verlust einer Zelle, wird eine ganze AAL5-PDU, die ein oder mehrere IP-Pakete enthält, unbrauchbar. Gerade hier könnte man den Switch entlasten, wenn beschädigte AAL5-PDUs erkannt und nicht mehr weitergeleitet würden (*Early* bzw. *Partial Packet Discard* [12]).

Verluste in Treiber und Betriebssystem

Neben Zellverlusten in ATM-Switches gehen die meisten Zellen in den ATM-Adapterkarten des Empfängers verloren (siehe Abbildung 14). Messungen ergaben, daß bis zu 8% und selten weniger als 1% der Zellen in der Treibersoftware der SBA200-Karte verloren gingen. Dieser große Zellverlust ist die Hauptursache für die hohen Paketverluste in den höheren Protokollschichten. Grund dafür sind laut Manual Page von `atmstat` mangelnder Pufferplatz oder Warteschlangenüberläufe im Gerätetreiber.

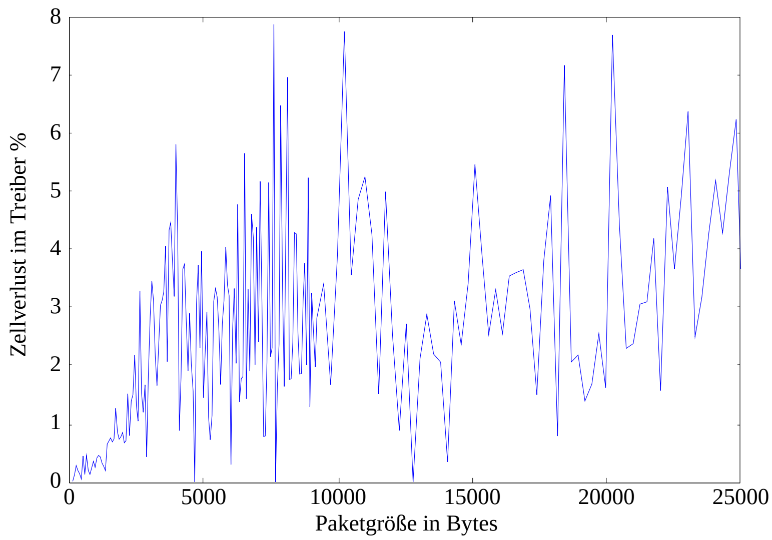

Abbildung 14: Zellverlust im Fore SBA200 Treiber

Ein weiteres Problem sind die Betriebssystempuffer. Für die, über eine Netzwerkschnittstelle zu versendenden Daten, wird in BSD basierten Systemen sogenannte *Mbufs* verwendet. Unter SunOS sind die Mbufs 128 Byte große Speicherbereiche mit 112 Bytes Nutzdatenanteil. Für Puffer größer als 512 Bytes werden 1024 Bytes große *Cluster Mbufs* verwendet.Über die Mbufs werden die für die Socketoperationen (z.B. `send()`) benötigten Daten in den Kernel kopiert und an entsprechende Module (z.B. `ip_input()`, `ip_output()`, ...) übergeben. Das Modul ergänzt die Protokollheader und reicht die Kette von Mbufs an den Netzwerktreiber weiter, der die Daten dann in den Speicher des Netzwerkadapters kopiert (siehe Abbildung 15). Die kleinen Mbufs komplizieren die Speicherverwaltung der Warteschlangen und sollten daher in zukünftigeren Betriebssystemen vergrößert werden.

Im Betriebssystem stellt das doppelte Kopieren der Daten zwischen den einzelnen Schichten bei Standard-BSD-Implementierungen eine weitere Verlustursache dar (*Double Copy Stack*, siehe Abbildung 15). Die bei HP entwickelte *Afterburner* Architektur umgeht diesen Performanceverlust mit einem *Single Copy Stack*. Die Daten werden dabei direkt vom User-Adreßraum in den Adreßraum des Netzwerkinterfaces kopiert.

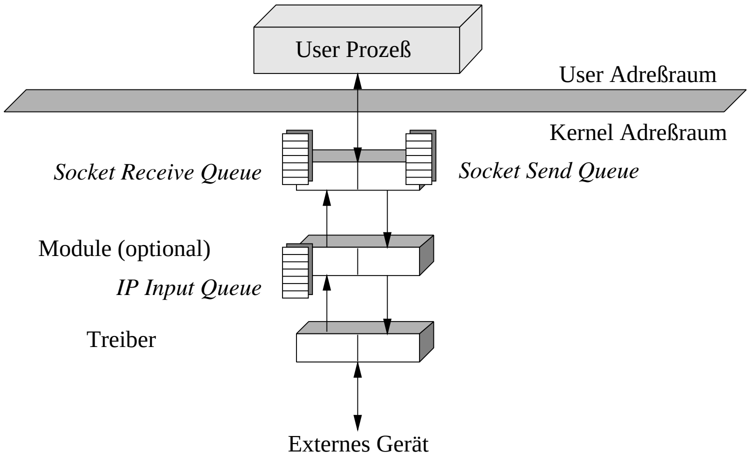

Abbildung 15: Kommunikationsdaten-Verarbeitung unter BSD 4.3-UNIX

4.4 *IP over ATM* in heterogenen Netz

In [7] wurden auch Messungen in heterogenen Netzwerken durchgeführt, wobei neben ATM auch Ethernet und FDDI zu Einsatz kamen. Die Performance liegt dabei immer deutlich unter der der langsamsten Architektur. Bei heterogenen Netzen ist eine Flußkontrolle absolut notwendig. Da auf ATM-Ebene keinerlei Flußkontrolle erfolgt, wurden im wesentlichen nur Messungen mit TCP durchgeführt. Als größter Flaschenhals erwiesen sich die Router durch die besonders bei größeren IP-Paketen nötige Fragmentierung wegen der unterschiedlichen MTU-Größen in den einzelnen Teilnetzen. Ein ATM-Netz konnte nur durch Koppelung mit *mehreren* langsameren Netzwerken ausgelastet werden. ATM bietet sich somit auch als Backbone-Architektur an.

5 Stand der Forschung

Da das Internet dem Benutzer keinerlei Garantie gewährt, ob und wann ein zu sendendes Paket wirklich übertragen wird, bietet es nur einen Best-Effort-Dienst. Mit der aktuellen Version 4 des Internet-Protokolls können die unterhalb von IP gebotenen ATM-Dienstgüten nicht genutzt werden. Eine auf IP basierende Anwendung kann daher über ATM weder einen Verkehrskontrakt mit einem ATM-Netz aushandeln, noch die Einhaltung eines durch Dritte vorgegebenen Kontrakts gewährleisten. Derzeit werden einige Vorschläge für bessere Policing-Algorithmen auf ATM-Ebene diskutiert, die höhere Protokolle unterstützen [13, 14]. Letztendlich scheint doch am besten zu sein, das Internet-Protokoll um QoS-Parameter (*Quality of Service*) zu erweitern, so daß jede auf IP über ATM basierende Anwendung einen den Anforderungen entsprechenden Verkehrskontrakt aushandeln kann. Dieser Ansatz wird mit der Entwicklung von IP Version 6 (IPng) verfolgt [15].

Im Rahmen der ATM-Standardisierung sind immer noch einige Probleme offen bzw. in der Diskussionsphase [9], z.B. Unterstützung von Multicast und Broadcast mittels eines *Multicast Address Resolution Servers* (MARS), Verbesserung des ARP-Services, Verbindungsverwaltung für eine effektivere Nutzung der virtuellen Verbindungen [6, 13, 14] und verbessertes Routing durch das *Next Hop Resolution Protocol* (NHRP) [16].

6 Zusammenfassung

In dieser Arbeit wurde dargestellt, mit welchen Verfahren das Internet-Protokoll in ATM-Netzwerken verwendet werden kann. Es wurden die wichtigsten von der Arbeitsgruppe *IP over ATM* der IETF entwickelten Standards erwähnt. Der relativ große Protokolloverhead, die schlechte Skalierbarkeit und die Anfälligkeit des ATMARP-Servers wurden angesprochen.

In der theoretischen Betrachtung der Performance von *IP over ATM* wurden die Leistungsgrenzen aufgezeigt. Der Performanceeinbruch bei den häufig vorkommenden kleinen Paketen wurde gezeigt. Bei praktischen Messungen auf einer SPARCstation 10 mit einer physikalischen Datenübertragungsrate von 100 Mbit/s (TAXI) konnten für TCP ca. 30 Mbit/s und für UDP etwa 70 Mbit/s als maximale Datenübertragungsraten ermittelt werden. Bei Übergängen zwischen unterschiedlich leistungsfähigen Rechnern traten 50 bis 100% Performanceverluste auf. Als Ursache dafür konnten fallengelassene Zellen im Switch und vor allem im Treiber der Netzwerkadapterkarte beim Empfänger gefunden werden.

Gründe für eine relativ niedrige Performance sind der erwähnte Bandbreitenverlust durch den Zellkopf-Overhead, fallengelassene Zellen im Switch und vor allem im Treiber der Netzwerkadapterkarte durch Pufferüberläufe im Kernel beim Empfänger sowie der Paketbearbeitungsaufwand bei höheren Protokollen. Auf die für nicht auf Hochgeschwindigkeitsprotokolle ausgelegte Speicherverwaltung und auf die eventuell mangelnde Prozessorleistung wurde hingewiesen. Damit bestehende Anwendungen künftig besser von der unter ATM möglichen Leistung profitieren können, müssen einerseits die Applikationen selbst auf das Versenden größerer Datenpakete optimiert und andererseits die Speicherverwaltung in künftigen Betriebssystemen beschleunigt werden.

Der hier beschriebene Ansatz *IP over ATM* zur Integration des Internet-Protokolls in ATM-Netzwerken ist im Vergleich zu anderen, wie *LAN Emulation*, die derzeit leistungsfähigste Lösung. Die genannten Performance-Probleme in den Endgeräten sind in Zukunft noch zu lösen.

Literatur

[1] Juha Heinanen: *Multiprotocol Encapsulation over ATM Adaptation Layer 5.* RFC 1483, Telecom Finland, Juli 1993.

[2] M. Laubach: *Classical IP and ARP over ATM.* RFC 1577, Hewlett-Packard Laboratories, Januar 1994.

[3] R. Atkinson: *Default IP MTU for use over ATM AAL5.* RFC 1626, Naval Research Laboratory, Mai 1994.

[4] M. Perez, F. Liaw, D. Grossman, A. Mankin, A. Hoffman: *ATM signalling support for IP over ATM.* RFC 1755, USC/Information Sciences Institute, FORE Systems, Inc., Motorola Codex, Ascom Timeplex, Inc., Januar 1995.

[5] Anthony Alles: *ATM Internetworking.* Engineering InterOp, Las Vegas, March 1995

[6] S. Keshav et al: *An Empirical Evaluation of Virtual Circuit Holding Time Policies in IP-over-ATM Networks.* IEEE JSAC Vol. 13 No. 8, October 1995

[7] Andrè Zehl: *IP Internetworking und Performance über ATM.* Diplomarbeit, Technische Universität Berlin, Dezember 1994.

[8] Lutz Henckel, Falk Schulz: *DFN-ATM-Vorhaben - Studie zur Anpassung von TCP/IP an ATM-AAL.* GMD FOKUS, September 1995. Weitere Informationen sind über URL `http://www.fokus.gmd.de/nthp/dfn-atm/` erhältlich.

[9] R. Cole, D. Shur, C. Villamizar: *IP over ATM: A Framework Document.* RFC 1932, AT&T Bell Laboratories, ANS, April 1994.

[10] Ramón Càceres: *Efficiency of Asynchronous Transfer Mode Networks in Transporting Wide-Area Data Traffic.* Computer Science Division, University of California, Berkeley, 1991

[11] Roy Campbell, Andrew Schmidt: *Internet Protocoll Traffic Analysis with Applications for ATM Switch Design.* Department of Computer Science, University of Illinois at Urbana-Champaign, 1992.

[12] Allyn Romanow, Sally Floyd: *Dynamics of TCP Traffic over ATM Networks.* IEEE JSAC, Vol. 13, No. 4, Mai 1995.

[13] Yoon Chang, Nada Golmie, David Su: *A Rate Based Control Switch Design for ABR Service in an ATM Network.* National Institute of Standards and Technology, Gaithersburg, Maryland, Juni 1995

[14] Li, Siu, Tzeng, Ikeda, Suzuki: *TCP Performance over ABR and UBR Services in ATM,* Beitrag zur International Pheonix Conference on Computers and Communications (IPCCC), März 1996.

[15] Peter Schulter: *A Framework for IPv6 over ATM.* Internet Draft, 1996. elektronisch verfügbar als `draft-schulter-ipv6atm-framework-02.txt`

[16] D. Katz, D. Piscitello: *NBMA Next Hop Resolution Protocol (NHRP).* Internet Draft, 1995.

Abkürzungen

AAL	ATM Adaption Layer
ABR	Available Bit Rate
ARP	Address Resolution Protocol
ATM	Asynchronous Transfer Mode
ATMARP	ATM Address Resolution Protocol
BOM	Begin of Message
CBR	Constant Bit Rate
COM	Continuation of Message
CPCS	Common Part Convergence Sublayer
CRC	Cyclic Redundancy Check
CS	Convergence Sublayer
EOM	End of Message
FCS	Frame Check Sequence
HEC	Header Error Control
ICMP	Internet Control Message Protocol
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
InARP	Inverse Address Resolution Protocol
IP	Internet Protocol
LAN	Local Area Network
LIS	Logical IP Subnet
LLC	Logical Link Control
MAC	Medium Access Control
MAN	Metropolitan Area Network
MIPS	Million Operations Per Second
MTU	Maximum Transmission Unit
NSAP	Network Layer Service Access Point
PDU	Protocol Data Unit
PVC	Permanent Virtual Circuit
QoS	Quality of Service
RFC	Request For Comments
SAR	Segmentation and Reassembly
SDH	Synchronous Digital Hierarchy
SDU	Service Data Unit
SMDS	Switched Multi-Megabit Data Service
SNAP	Subnetwork Access Point
SPANS	Simple Protocol for ATM Network Signalling
SSCS	Service Specific Convergence Sublayer
SVC	Switched Virtual Circuit
TAXI	Transparent Asynchronous transmitter/receiver Interface
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UNI	User Network Interface
VC	Virtual Channel
VCI	Virtual Channel Identifier
VP	Virtual Path
VPI	Virtual Path Identifier
WAN	Wide Area Network