Multicast-Kommunikation im Internet

T. Braun[1]

Gruppenkommunikation wird im Internet durch das Multicast-Konzept unterstützt. Zur Realisierung von Multicast sind Erweiterungen an verschiedenen Internet-Protokollen vorgenommen worden sowie neue Protokolle entwickelt worden. Im Mittelpunkt dieses Beitrags steht die IP-Multicast-Architektur, d. h. insbesondere die Multicast-Erweiterungen des IP-Protokolls und der dazugehörenden Multicast-Routing-Protokolle. Das IP-Multicast-Modell hat signifikante Auswirkungen auf die darunter liegenden Netztechnologien und die auf der IP-Schicht aufsetzenden Transportprotokolle und Anwendungen.

Schlüsselwörter: Gruppenkommunikation; Internet; Multicast; Routing

Multicast communication in the Internet. The Internet supports group communications by its multicast concept. Several Internet protocol extensions and new protocols have been developed in order to realize multicast in the Internet. This paper focuses on the IP multicast architecture, in particular on the IP multicast extensions and the corresponding multicast routing protocols. The IP multicast model has significant impacts on the underlying network technologies and on the transport protocols and applications on top of IP. These impacts are also discussed in the paper.

Keywords: group communication; Internet; multicast; routing

1. Einleitung

Gruppenkommunikation gewinnt zunehmend in multimedialen Anwendungsszenarienen an Bedeutung, z. B. für Audio/Video-Konferenzen, Computer Supported Cooperative Work oder verteilte Spiele. Neue Multimedia-Anwendungen erlauben oft gleichzeitig die Daten-, Video- und Audiokommunikation zwischen mehreren Teilnehmern. Populäre Beispiele sind die sogenannten MBone Tools [6, 9, 24], welche für Audio/Video-Konferenzen und verteilte Whiteboard-Anwendungen eingesetzt werden können. Ein noch größeres Wachstumspotenzial wird verteilten, interaktiven Spielen eingeräumt [31].

Um diese Anwendungen zu ermöglichen, muss das Internet als zukünftige globale Kommunikationsplattform Gruppenkommunikation unterstützen [30]. Die Gruppenkommunikation basiert im Internet auf dem Konzept des IP Multicast. Für die aktuelle IP-Version 4 wurden seit Anfang der 90er Jahre nachträglich Funktionen entwickelt, die effiziente Multicast-Kommunikation ermöglichen. Multicast-Funktionalität ist daher nicht in allen IPv4-Implementierungen enthalten, wird aber in allen Implementierungen der neuen, derzeit im Internet eingeführten IP-Version 6 vorhanden sein [29]. Ein wichtiger Aspekt bei der Entwicklung von IP-Multicast-Mechanismen ist die Skalierbarkeit, d. h. die Multicast-Mechanismen müssen für große Benutzer- und Gruppenzahlen (bis in den Bereich von Millionen) geeignet sein.

[1] Prof. Dr. Torsten Braun, Institut für Informatik und Angewandte Mathematik, Universität Bern, Neubrückstraße 10, CH-3012 Bern, Schweiz. (E-Mail: braun@iam.unibe.ch)

2. IP Multicast

Die Grundlage für Multicast-Kommunikation über das Internet sind Erweiterungen des IP-Protokolls sowie der dazugehörigen Routing- und anderen Kontrollprotokolle. Dieses Kapitel beschreibt diese Erweiterungen bzw. neuen Protokolle.

2.1 IP-Multicast-Modell

IP unterstützte ursprünglich nur eine 1:1-Kommunikation, indem ein mit einer Unicast-Quelladresse versehenes IP-Paket an einen durch eine Unicast-Zieladresse identifizierbaren Empfänger geschickt wurde. IP Multicast erweitert den IP-Unicast-Dienst um die Möglichkeit, IP-Pakete an eine Gruppe von Empfängern anstatt an einen einzelnen Empfänger zu senden. Ein Sender sendet dabei ein Multicast-Paket an eine durch eine IP-Multicast-Adresse gekennzeichnete Gruppe, wodurch eine $1:n$-Kommunikationsbezie-

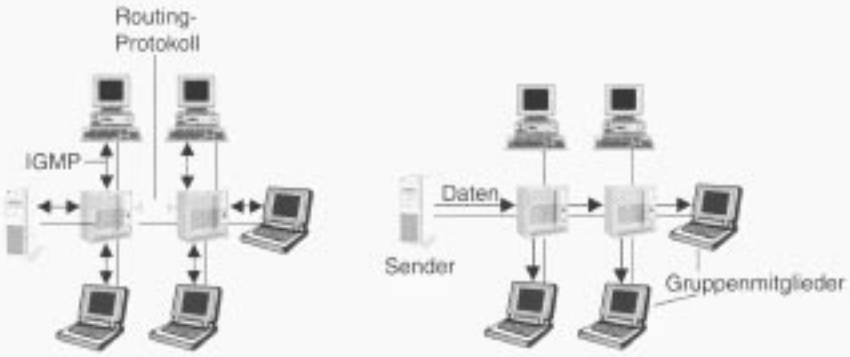

Abb. 1. IP-Multicast-Modell

hung verwirklicht wird. Die im IP-Paket enthaltene IP-Multicast-Zieladresse wird auf eine Multicast-Adresse auf Netzebene abgebildet, falls das Netz (z. B. Ethernet, ATM usw., vgl. Kap. 3) Multicast unterstützt.

Die wesentlichen sich aus Multicast ergebenden Vorteile bestehen darin, dass ein Multicast-Paket für n Empfänger vom Sender nur einmal gesendet werden muss, und dass es jeden zwischen Sender und Empfängern liegenden Übertragungsabschnitt nur einmal passieren muss. Im Unicast-Fall müsste ein Sender das Paket bei n Empfängern n-mal senden, was zu einer deutlich grösseren Belastung des Senders und einem erhöhten Bandbreitenbedarf auf dem Netz führt.

Ein Sender weiß bei der IP-Multicast-Kommunikation nicht, wie viele und welche Empfänger die an eine Multicast-Gruppe gesendeten Pakete empfangen wollen. Ein Sender sendet ein Multicast-Paket einfach unter Angabe der Gruppenadresse in der IP-Zieladresse auf das angeschlossene Subnetz. Die an diesem Subnetz angeschlossenen Gruppenmitglieder empfangen das Multicast-Paket direkt. Gleichzeitig erhalten alle an das Subnetz angeschlossenen IP-Router das Paket und leiten dieses über einen Übertragungsabschnitt weiter, sofern weitere Gruppenmitglieder über diesen erreichbar sind.

Ein Router muss deshalb wissen, auf welchen Wegen weitere Gruppenmitglieder erreichbar sind und auf welchen Übertragungsabschnitten er empfangene Multicast-Pakete weiterleiten muss. Diese Information lernt ein Router, indem er mit anderen Routern über Multicast-Routing-Protokolle wie z. B. DVMRP (Distance Vector Multicast Routing Protocol) oder MOSPF (Multicast Open Shortest Path First; vgl. Kap. 2.3) die Multicast-Topologie-Informationen austauscht. Schließlich müssen sämtliche Router, die direkt mit Endsystemen verbunden sind, feststellen, ob innerhalb der angeschlossenen Subnetze Mitglieder einer Multicast-Gruppe existieren. Hierzu zeigen Endsysteme durch das Senden von IGMP-Nachrichten (IGMP: Internet Group Management Protocol; Kap. 2.2) dem Router an, dass sie die an eine bestimmte Multicast-Gruppe gesendeten Nachrichten empfangen wollen (Abb. 1).

2.2 Internet Group Management Protocol

Im obigen Beispiel ist es notwendig, dass die Router genau wissen, ob sich in ihren Subnetzen Mitglieder einer Multicast-Gruppe befinden; z. B. muss der rechte stromabwärts liegende Router dem linken stromaufwärts liegenden Router (über Multicast-Routing-Protokolle) mitteilen, ob er selbst Gruppenmitglieder bedienen muss. Nur in diesem Fall ist es notwendig, dass der stromaufwärts liegende Router dem stromabwärts liegenden Router Multicast-Pakete weiterleitet. Hat der stromabwärts liegende Router keine Gruppenmitglieder zu bedienen, so ist das Weiterleiten der Multicast-Pakete durch den stromaufwärts liegenden Router nicht erforderlich.

Das Anzeigen, ob ein Endsystem die Pakete einer Multicast-Gruppe empfangen will, erfolgt mit Hilfe des Internet Group Management Protokolls (IGMP) [11]. An der IGMP-Dienstschnittstelle werden mit

- JoinHostGroup (group address, interface) und
- LeaveHostGroup (group address, interface)

zwei spezielle Operationen zum Beitreten und Verlassen einer Multicast-Gruppe angeboten. Des Weiteren muss das darüber liegende Protokoll auch die Paketlebenszeit zum Senden von Paketen spezifizieren können.

IGMP setzt über IP auf und unterscheidet Query- (Anfragen) und Response-Pakete (Antworten). Die Multicast-Router versenden dabei Query-Pakete an eine permanente Gruppe (all hosts group), mit denen alle Endsysteme erreicht werden. Mit diesen Query-Paketen wird periodisch angefragt, ob die jeweiligen Endsysteme zu irgendwelchen Multicast-Gruppen gehören. Das Endsystem antwortet auf die Anfrage mit einem Response-Paket, falls weiterhin die Mitgliedschaft in einer Multicast-Gruppe aufrecht erhalten werden soll (Abb. 2). Die Antwort erfolgt erst nach einem gewissen Zeitintervall. Sollte während dieses Intervalls ein anderes Endsystem im selben Subnetz auf eine Anfrage antworten, so verzichten die anderen Gruppenmitglieder in diesem Subnetz auf das Senden einer Antwort, da durch die erste Antwort gewährleistet wird, dass die Multicast-Pakete weiterhin zu diesem

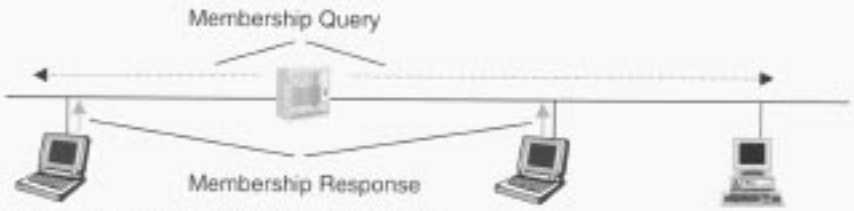

Abb. 2. Internet Group Management Protocol

Subnetz weitergeleitet werden. Da die Anfragen nur relativ selten (z. B. im Minutenabstand) gesendet werden, besitzt ein Endsystem die Möglichkeit, auch ohne Anfrage eine Antwort zu senden, um beispielsweise den Eintritt in eine Gruppe anzuzeigen. Es reicht dabei aus, dass mindestens ein Gruppenmitglied innerhalb eines Subnetzes dies signalisiert, so dass der Router danach alle Multicast-Pakete an diese Gruppe über das Subnetz verteilt.

Eine Zugangskontrolle zu Multicast-Gruppen gibt es nicht, d. h., jedes beliebige Endsystem kann einer Multicast-Gruppe beitreten und empfängt dann in der Folgezeit auch alle für die Gruppe bestimmten Pakete. Geschlossene Benutzergruppen können daher nur mit Verschlüsselungsmethoden realisiert werden, so dass nur Empfänger, die einen Schlüssel für die Multicast-Gruppe besitzen, ein Multicast-Paket an diese Gruppe dekodieren können.

IGMP wird gegenwärtig weiterentwickelt. IGMPv2 [22] enthält spezielle Nachrichten, um das Verlassen einer Gruppe anzuzeigen. Damit kann ein Router schneller erkennen, dass das letzte Endsystem eines IP-Subnetzes eine Multicast-Gruppe verlassen hat. In diesem Fall kann das unnötige Weiterleiten von Multicast-Paketen in ein Subnetz ohne Empfänger vermieden werden. Hierbei setzt ein Endsystem ein Flag, falls es auf das letzte Membership Query mit einer Response-Nachricht geantwortet hat. Falls nun ein Endsystem eine Multicast-Gruppe verlassen will, erfolgt keine spezielle Aktion, falls das Flag nicht gesetzt ist, während bei gesetztem Flag das Endsystem eine Leave-Group-Nachricht sendet. In diesem Fall fragt der Router durch ein gruppenspezifisches Membership-Query-Paket, ob noch weitere Gruppenmitglieder im Subnetz vorhanden sind. Falls kein Endsystem darauf antwortet, kann der Router das Weiterleiten der Multicast-Pakete sofort unterdrücken und auch die stromaufwärts liegenden Router davon in Kenntnis setzen, dass er die Multicast-Pakete für diese Gruppe nicht mehr weiterleiten soll. Ohne dieses Verfahren hätte der Router nur nach dem Ausbleiben der Endsystem-Reaktionen auf das periodisch gesendete Membership-Query-Paket feststellen können, dass keine weiteren Gruppenmitglieder mehr existieren. In der Zwischenzeit (Intervalle bis in den Minutenbereich) wären dann viele unnötige Pakete über den Router in das Subnetz weitergeleitet worden.

Eine weitere Optimierung besteht in der Einführung von Source-Filtering-Funktionen. Bei IGMPv1 und IGMPv2 zeigt das Gruppenmitglied durch die IGMP-Nachrichten an, dass es alle Multicast-Pakete an die spezifizierte Multicast-Gruppe, d. h. von jedem beliebigen Sender, empfangen will. Bei IGMPv3 [2] kann zusätzlich zur Gruppe auch eine Sendermenge angegeben werden, so dass nur Pakete, die von diesen angegebenen Sendern gesendet wurden, an das entsprechende Subnetz des Empfängers weitergeleitet werden. Diese Informationen können dann durch die entsprechenden Multicast-Routing-Protokolle weiterverarbeitet werden, d. h. auch die Router signalisieren sich untereinander, dass an sie nur Pakete von bestimmten Sendern weitergeleitet werden müssen. Insgesamt reduziert dieser Mechanismus die benötigte Bandbreite weiter. Als neues Dientsprimitiv wurde IPMulticastListen (socket, interface, multicast-address, filter-mode, source-list) definiert, wobei als Filter-Modi exclude oder include möglich sind. Damit kann dieses eine Primitiv die IGMPv1-Primitive JoinHostGroup und LeaveHostGroup ersetzen.

2.3 Multicast Routing

Mit Hilfe von Multicast-Routing-Protokollen tauschen Router Informationen über die Existenz von Gruppenmitgliedern in Teilen des Internets aus. In Abb. 1 muss beispielsweise der stromabwärts liegende Router dem stromaufwärts liegenden Router mitteilen, dass dieser Multicast-Pakete für die an ihn direkt angeschlossenen Empfänger weiterleiten soll.

Wie bei Unicast-Routing-Protokollen können Multicast-Routing-Protokolle in Distanz/Vektor-Verfahren und Link-State-Routing-Verfahren eingeteilt werden. Für Distanz/Vektor-Multicast-Routing existieren verschiedene Varianten. Beim Reverse Path Forwarding (RPF) [7] wird ein Paket auf allen Links mit Ausnahme des Links, an dem das Paket ankam, weitergeleitet, falls das Paket auf dem Link empfangen wurde, welcher den kürzesten Weg zurück zur Quelle darstellt. Ein Problem dieses Verfahrens ist, dass, falls ein Link zwei Router besitzt, die ein Paket jeweils empfangen haben, dieses von jedem Router weitergeleitet wird, d. h. doppelt über den Link übertragen wird. Beispielsweise senden beim RPF-Verfahren Router x und Router y in Abb. 3 das gleiche Paket auf Link a.

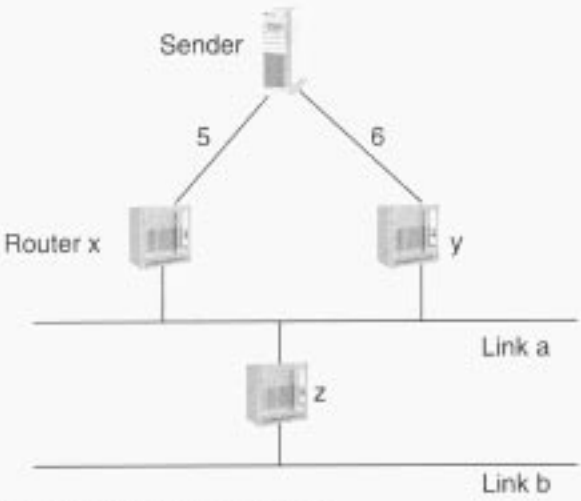

Abb. 3. Reverse Path Forwarding

Dieses Problem wird beim Reverse Path Broadcast-Verfahren (RPB) dadurch gelöst, dass zwischen Router und Link eine Eltern-Kind-Beziehung ermittelt wird. Sind beispielsweise wieder zwei Router eines Multicast-Baums an einen Link angeschlossen, so ist derjenige Router ein Elternteil des Links, der die kürzeste Entfernung zur Quelle, d. h. zum Sender, besitzt. In Abb. 3 ist Router *z* ein Kind von Router *x*, da Router *x* eine kürzere Entfernung zur Quelle besitzt als Router *y*. Bei Entfernungsgleichheit gibt die kleinere Adresse den Ausschlag. Das Verfahren erfordert, dass die Router die entsprechenden Informationen, wie die Entfernung zur Quelle über einen Link, mit den benachbarten Routern austauschen.

Die beiden Verfahren RPF und RPB implementieren ein Broadcast-Verfahren, d. h., um ein wirkliches Multicast zu erreichen, ist es notwendig, Pakete nur dann weiterzuleiten, wenn mit dem Weiterleiten Mitglieder einer Multicast-Gruppe erreicht werden. Mit dem Truncated RPB- (TRPB-)Verfahren gelingt dies, indem Pakete nur auf Links weitergeleitet werden, welche kein sogenanntes Blatt-Link ohne Gruppenmitglieder darstellen. Blatt-Links sind Links, die von keinem Router benutzt werden, um die Quelle zu erreichen (Link *b* in Abb. 3). Nachbar-Router lernen dabei durch spezielle Nachrichten von anderen Routern, ob ein Link ein Blatt-Link ist oder nicht. Router *z* sendet z. B. eine Nachricht „Link *a* ist der nächste Weg zur Quelle *S*" auf Link *a*. Die Router *x* und *y* erkennen dann, dass Link *a* kein Blatt-Link darstellt. Pakete werden nur dann weitergeleitet, wenn an einem Link Gruppenmitglieder vorhanden sind oder der Link benötigt wird, um einen anderen Router in der Multicast-Gruppe zu erreichen.

Das Reverse Path Multicast- (RPM-)Verfahren verwendet eine sogenannte Pruning- Technik, um anzuzeigen, ob Pakete über einen Link weitergeleitet werden müssen oder nicht. Zunächst werden dabei die Pakete nach dem TRPB-Verfahren gesendet. Erhält ein Router ein Paket, wobei alle Links des Routers Blatt-Links sind, an die keine Gruppenmitglieder angeschlossen sind, so sendet er an den nächsten Router in Richtung Quelle die Non-Membership-Report-Nachricht für die spezielle Gruppe und die spezielle Quelle. Um eine Wiederaufnahme eines Gruppenmitglieds zu erreichen, werden ebenfalls Kontrollnachrichten gesendet, die das Blockieren von Links wieder aufheben. Eine alternative Möglichkeit besteht in der automatischen, periodischen Wiederaufnahme von Links in den Baum.

Eine bereits seit längerem im Internet Verwendung findende Multicast-Version der Distanz-Vektor-Routing-Technik ist DVMRP (Distance Vector Multicast Routing Protocol) [10]. Bei diesem Protokoll werden Informationen ausgetauscht, welche die Mitgliedschaft von Knoten in Gruppen und die Routing-Kosten zwischen den Knoten enthält. Für jede Gruppe und jeden darin enthaltenen Sender wird dabei ein Routing-Baum entsprechend dem TRPB-Verfahren erstellt. Empfangene IP-Multicast-Pakete werden auf allen Links weitergesendet, welche zu den Zweigen des Multicast-Baums der Gruppe gehören. DVMRP setzt auf IGMP-Paketformaten auf und definiert zusätzliche Nachrichten zum Austausch von Multicast-Routing-Informationen.

Eine Alternative zur Distanz-Vektor-Technik besteht im Shortest Path First (SPF) oder Link-State Routing. Bei dieser Technik überprüft jeder Router periodisch, ob die Verbindungen zu seinen Nachbar-Routern noch gültig sind. Die Routing-Informationen müssen innerhalb einer Domäne per Broadcast an alle anderen Router verteilt werden. Jeder Router unterhält eine Tabelle, in der die gesamte, vollständige Netztopologie einer Routing-Domäne gespeichert ist. Da jeder Router eine komplette Sicht des Netzzustands besitzt, kann er die benötigten Routing-Informationen selbständig berechnen. Hierzu wird in der Regel Dijkstras Shortest Path Tree- (SPT-)Algorithmus verwendet. In der Internet-Welt wird beispielsweise das Protokoll Open SPF (OSPF) [12] eingesetzt. Bei MOSPF, der Multicast-Erweiterung von OSPF [13, 14], berechnet jeder Multicast-Router innerhalb der Routing-Domäne für jede Gruppe und für jeden Sender dieser Gruppe einen Baum. Dies ist gerade für eine große Anzahl von großen Gruppen etwas ineffizient.

Die beiden Ansätze MOSPF und DVMRP berechnen jeweils die kürzesten Pfade zwischen einem Sender und den verschiedenen Empfängern. Sie haben daher speziell dann Nachteile, wenn sich die Gruppen nicht auf wenige Regionen eines Netzes konzentrieren, sondern sehr weit verteilt sind, da dann relativ viele Netzwerkressourcen beansprucht werden. Aus diesem Grund wurden neue Architekturen entwickelt, die besonders im Fall von weit verstreuten und weniger dicht konzentrierten Multicast-Gruppen ein gutes Ver-

halten hinsichtlich Leistungsfähigkeit und Bandbreitenbedarf besitzen.

Der CBT-Ansatz (Core Based Tree) [1, 19, 20] benutzt lediglich einen Baum für die gesamte Gruppe. Alle Sender dieser Gruppe senden dann die Dateneinheiten über diesen Baum, so dass sich der gesamte Verkehr innerhalb der Gruppe auf den einen gemeinsamen Baum konzentriert. Dadurch werden zwangsläufig nicht alle Pakete auf dem kürzest möglichen Pfad übertragen. In bestimmten Situationen ist die entstehende zusätzliche Verzögerung nicht besonders gravierend (typischerweise eine Erhöhung um bis 100 %) und kann zugunsten eines effizienteren und vereinfachten Routings manchmal toleriert werden. Speziell bei vielen Sendern mit relativ geringer Datenrate erscheint z. B. ein CBT-Ansatz vorteilhaft, während sich ein SPT-Routing für wenige, mit hoher Datenrate sendende Quellen als sinnvoller erweist. Es bietet sich daher an, einen flexiblen Ansatz zu wählen und in Abhängigkeit des gewünschten Einsatzszenarios die eine oder die andere Technik zu wählen.

Genau diese Flexibilität ist in PIM (Protocol Independent Multicast) möglich. Eine Multicast-Gruppe kann entweder Bäume nach dem SPT-Verfahren oder aber gemeinsam benutzte Bäume auswählen. In PIM sind zwei verschiedene Arbeitsmodi vorgesehen:

(1) PIM-Dense Mode (PIM-DM) [4]
(2) PIM-Sparse Mode (PIM-SM) [23]

PIM-DM ist ein DVMRP-ähnliches Verfahren, basiert jedoch auf Reverse Path Multicasting, d. h., ein Router leitet eine Dateneinheit von einer für ihn bislang unbekannten Multicast-Gruppe auf allen Links weiter mit Ausnahme des Links, auf dem das Paket empfangen wurde. Erst nach einer gewissen Zeit stellen die Router innerhalb des Multicast-Baums auf Grund der (Pruning-)Kontrollnachrichten fest, ob auf gewissen Zweigen die Pakete einer Gruppe gesendet werden müssen oder nicht, falls sich keine Gruppenmitglieder mehr in diesem Zweig befinden. Spezielle Join-Pakete werden dadurch nicht benötigt, jedoch wird durch das initiale Weiterleiten an alle Links ein gewisser Overhead erzeugt. Dieser Overhead wird in Kauf genommen, weil man durch dieses vereinfachte Verfahren kein spezielles Routing-Protokoll benötigt und auf einem beliebigen Unicast-Routing-Protokoll aufsetzen kann.

PIM-SM unterscheidet sich von PIM-DM darin, dass Router mit dahinter liegenden Gruppenmitgliedern explizit Join-Pakete senden müssen, um das Beitreten zu einer Gruppe zu veranlassen, während in PIM-DM zunächst die Mitgliedschaft von dahinter liegenden Endsystemen angenommen wird und diese erst nach den entsprechenden Kontrollnachrichten ausscheiden. Die angesprochenen Join-Pakete werden von den Gruppenmitgliedern (sowohl Sender als auch Empfänger) an den nächsten, innerhalb einer Gruppe ausgezeichneten Router, den sogenannten Rendezvous-Punkt, gesendet. Diese Rendezvous-Punkte sind Gruppen-spezifisch und bilden eine Art Backbone für die entsprechende Gruppe. Zwischen den Rendezvous-Punkten und den Gruppenmitgliedern werden dann Pfade etabliert und in den Multicast-Baum übernommen. Nicht mehr benötigte Pfade werden durch spezielle Kontrollnachrichten aufgelöst. In Abb. 4 sendet der nächste PIM-Router (Router A) des einzugliedernden Empfängers ein PIM-Join-Paket an den nächsten Rendezvous-Punkt, wodurch ein Zweig zwischen dem Rendezvous-Punkt und dem Empfänger erzeugt und in den Multicast-Baum aufgenommen wird. Ähnlich wird ein Sender aufgenommen, falls dieser zu senden beginnt. PIM-Router D empfängt das Datenpaket, kapselt es in ein spezielles PIM-Register-Paket ein und sendet dieses über weitere PIM-Router an einen Rendezvous-Punkt. Der Rendezvous-Punkt sendet eine PIM-Join-Nachricht an den PIM-Router D.

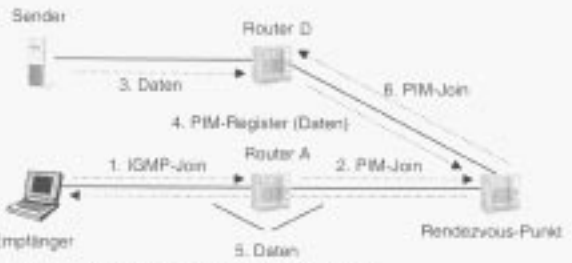

Abb. 4. Protocol Independent Multicast

2.4 Das Multicast Backbone im Internet

Der erste Multicast-Router wurde 1992 im Internet in Betrieb genommen, die erste Anwendung war im selben Jahr die Übertragung einer Konferenz der Internet Engineering Task Force (IETF) mit entsprechenden Audio/Video-Konferenzanwendungen. Die damals noch wenigen Multicast-fähigen Router wurden über nicht-Multicast-fähige Router verbunden, wobei zwischen zwei Multicast-Routern sogenannte Tunnel konfiguriert wurden. Am Anfang eines Tunnels werden dabei die Multicast-Pakete in Unicast-Pakete verpackt und mit der neuen Zieladresse des Multicast-Routers am Ende des Tunnels versehen. Dieser Vorgang wird auch als IP-in-IP-Encapsulation (Einkapseln) bezeichnet. Ein Tunnel bewirkt das gleiche als wenn die beiden Router über eine direkte Punkt-zu-Punkt-Verbindung miteinander verbunden wären. Der Multicast-Router am Ende des Tunnels stellt jeweils das ursprüngliche, eingekapselte Multicast-Paket wieder her (Abb. 5). Sind zwei Multicast-Router über ein gemeinsames Netz direkt miteinander verbunden, so ist ein Aufsetzen eines Tunnels nicht erforderlich, die beiden Multicast-Router erkennen sich gegenseitig automatisch. Das Netz der Multicast-Router im Internet wird auch Multicast-Backbone (MBone) genannt.

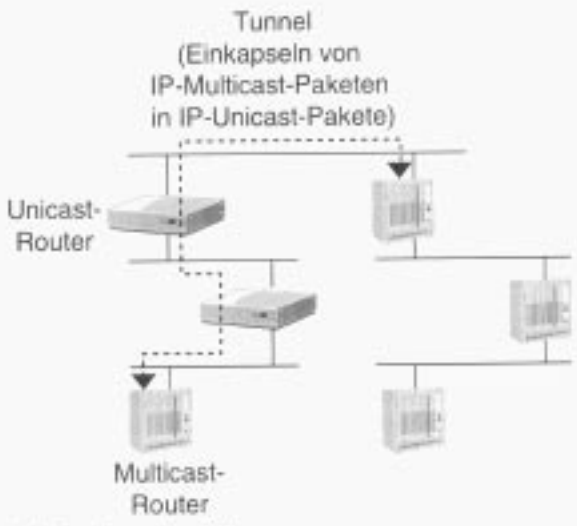

Abb. 5. Tunnel im MBone

Wichtige Voraussetzung zum Anschluss eines Endsystems an das MBone ist die Existenz eines Multicast-Routers am lokalen Netz des Endsystems. Der Multicast-Router ist dann entweder direkt mit dem nächsten Multicast-Router verbunden, oder es existiert ein Tunnel zum nächsten Multicast-Router. Ist kein Multicast-Router im Subnetz vorhanden, müssen von allen Endsystemen aus explizit Tunnel zu einem entfernten Multicast-Router eingerichtet werden.

3. Multicast über spezifische Netze

3.1 Multicast über IEEE LANs

IP Multicast benötigt natürlich eine entsprechende Unterstützung der darunter liegenden Netze. Relativ einfach ist die Abbildung von IP Multicast auf lokale Netze wie z. B. Ethernet oder Token Ring. Bei solchen Netzen gibt es bereits Multicast-Adressen auf MAC-Ebene, auf die eine IP-Adresse abgebildet werden kann. Eine 48-bit-Ethernet-Multicast-Zieladresse wird beispielsweise gebildet, indem das 24-bit-Präfix 01 00 5E sowie ein weiteres 0-Bit den 23 niederwertigsten Bits der 32-bit-langen IP-Multicast-Adresse vorangestellt werden. Das resultierende Ethernet-Paket wird dann über ein Ethernet auf Shared-Medium oder Switching-Basis gesendet. Bei einem gemeinsam benutzten Medium können die Multicast-Empfänger potenzielle Multicast-Pakete vom Medium kopieren und weiterverarbeiten. Bei einem auf Switching basierenden lokalen Netz müssen die Switches entweder einfach alle Multicast-Pakete auf den ausgehenden Links fluten oder sie müssen wissen, auf welchen der ausgehenden Links Gruppenmitglieder erreicht werden können. Hierzu wurde mit dem GARP (Generic Address Registration Protocol) Multicast Registration Protocol (GMRP) von IEEE ein Protokoll definiert, welches ähnlich wie IGMP anstatt auf der IP-Ebene den Endsystemen nun auf MAC-Ebene erlaubt, Gruppenmitgliedschaften anzukündigen. Diese GMRP-Nachrichten werden von den Switches aufgenommen und weiterverarbeitet, indem Ports, über die eine solche Nachricht empfangen wurde, in die Liste der Ausgangs-Ports für ein Multicast-Paket an eine bestimmte Gruppe aufgenommen wird.

3.2 Multicast über ATM

Die Abbildung von IP Multicast über ATM erfolgt am einfachsten, wenn über ATM LAN Emulation (LANE) eingesetzt wird [8]. In diesem Fall sorgt LANE für das Verteilen der Multicast-Nachrichten im emulierten LAN. Gegebenenfalls wird ein Multicast dann aber auf einen Broadcast-Mechanismus abgebildet, speziell dann, wenn die LANE-Implementierung nur Broadcast unterstützt.

Bei Classical IP über ATM [21] war Multicast-Unterstützung zunächst nicht im Standard enthalten. Deshalb wurde mit dem MARS-Ansatz eine Art Multicast-Erweiterung entwickelt. Der sogenannte Multicast Address Resolution Server (MARS) übernimmt dabei ähnliche Funktionen wie der ATM-ARP-Server im Unicast-Fall, d. h., er löst IP- Adressen nach ATM-Adressen auf [17, 18]. Sowohl bei ATM ARP als auch bei MARS sendet ein Client eine Adressauflösungsanfrage an einen Server (ATM-ARP-Server für Unicast oder MARS für Multicast). Der Server hat in beiden Fällen eine Tabelle, welche eine IP-Adresse auf ATM-Adressen abbildet. Während im Unicast-Fall eine IP-Unicast-Adresse auf eine einzige ATM-Adresse abgebildet wird, wird eine IP-Multicast-Adresse auf eine Liste von ATM-Adressen abgebildet. Die aufgelösten ATM-Adressen werden dann vom Server an den Client zurückgesendet, welcher dann damit Punkt-zu-Punkt- oder Punkt-zu-Multipunkt-ATM-Verbindungen aufbauen kann.

Der Aufbau der Abbildungstabelle erfolgt basierend auf der Registrierung der ATM-Endsysteme beim Server. Im Unicast-Fall registriert ein Endsystem die IP-Unicast-Adresse beim Initialisieren, die Multicast-Adressen werden erst bei Bedarf registriert, z. B. wenn eine Anwendung den Beitritt zu einer IP-Multicast-Gruppe der IP-Instanz mitgeteilt hat. Ändert sich die Zusammensetzung der Multicast-Gruppe, während eine Punkt-zu-Mehrpunkt-ATM-Verbindung existiert, so muss der MARS-Server alle ATM-Endsysteme über diese Veränderungen unterrichten. Die Registrierungsnachrichten werden dabei über bidirektionale ATM-SVCs, die MARS-Kontroll-VCs (Abb. 6), ausgetauscht, während die Aktualisierungsinformationen über einen Punkt-zu-Mehrpunkt-VC (MARS-Cluster-Kontroll-VC), ausgehend vom MARS-Server, an alle registrierten Endsysteme verteilt werden.

Die vom MARS-Server zurückgelieferten ATM-Adressen können Endsystem- oder Multicast-Server-

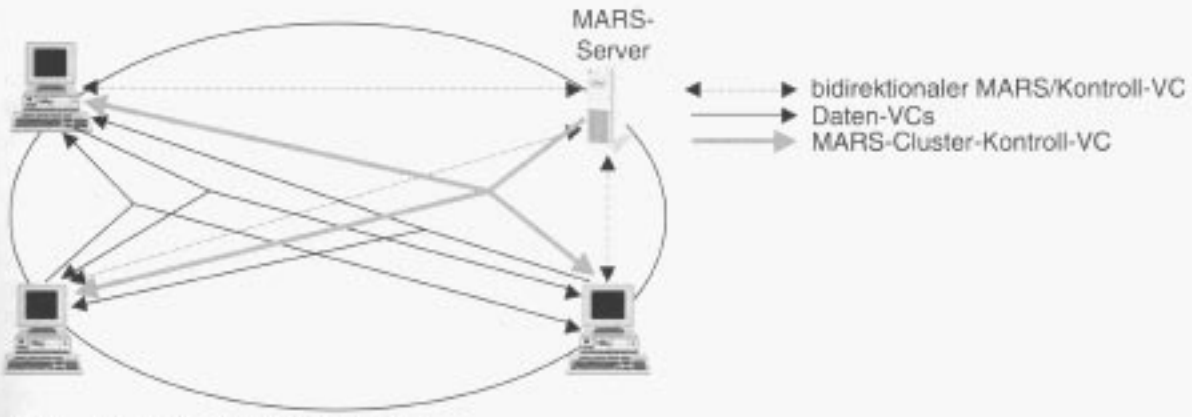

Abb. 6. Multicast Address Resolution Server

Adressen darstellen. Im ersten Fall etabliert dann jeder Sender zu jedem Endsystem einer IP-Multicast-Gruppe eine direkte Punkt-zu-Mehrpunkt-Verbindung. Man spricht dann auch von einem VC-Netz. Allerdings werden in einem Szenario mit n Sendern, die alle an m unterschiedliche Multicast-Gruppen Daten senden wollen, $n \cdot m$ Punkt-zu-Mehrpunkt-VCs erforderlich. Dies stellt dann ein Problem dar, wenn ATM-Switches verwendet werden, welche nur wenige Hundert Punkt-zu-Mehrpunkt-VCs unterstützen können. Um den Verbrauch von Punkt-zu-Mehrpunkt-VCs zu reduzieren, können nun sogenannte Multicast-Server etabliert werden, welche dann das Verteilen der Daten an eine Multicast-Gruppe übernehmen. In diesem Fall, man spricht auch von einem Multicast-Server-Szenario, werden die Punkt-zu-Mehrpunkt-VCs nur einmal pro Gruppe vom jeweils zuständigen Server aufgebaut. Alle n Sender senden dann aber über einen Punkt-zu-Punkt-VC (Multicast Send VC) die zu verteilenden Daten an den Server, welcher dann die Daten schließlich über den Punkt-zu-Mehrpunkt-VC (Multicast Forward VC) verteilt. Der Nachteil dieses Konzepts besteht dann aber in einem im Vergleich zum VC-Netz zusätzlichen System im Datenpfad, allerdings werden bei beliebig vielen Sendern nur m Punkt-zu-Mehrpunkt-VCs für die m Gruppen benötigt. Dies stellt aber keinen Nachteil dar, wenn der Multicast-Server in einen IP-Multicast-Router integriert ist und Daten von anderen Netzen über den Router im ATM-Netz verteilt werden. Ein weiterer Vorteil ist die Entlastung der Endsysteme von Signalisierungsaufgaben und Verbindungsverwaltung. Lediglich die Multicast-Server tauschen mit dem MARS-Server Kontrollnachrichten aus. Die VC-Netz-Konfiguration ist in Abb. 7 dargestellt, während eine Multicast-Server-Konfiguration in Abb. 8 zu sehen ist.

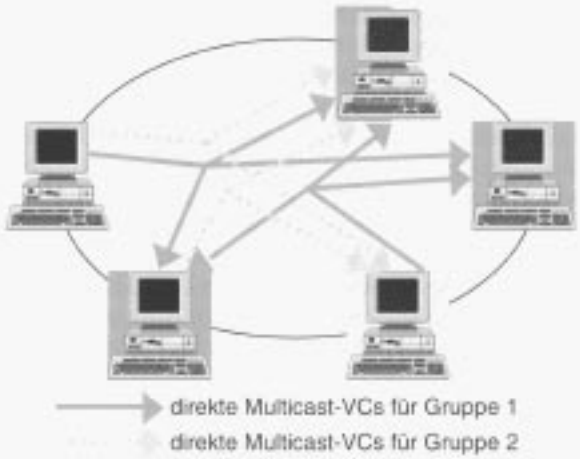

Abb. 7. MARS: VC-Netz

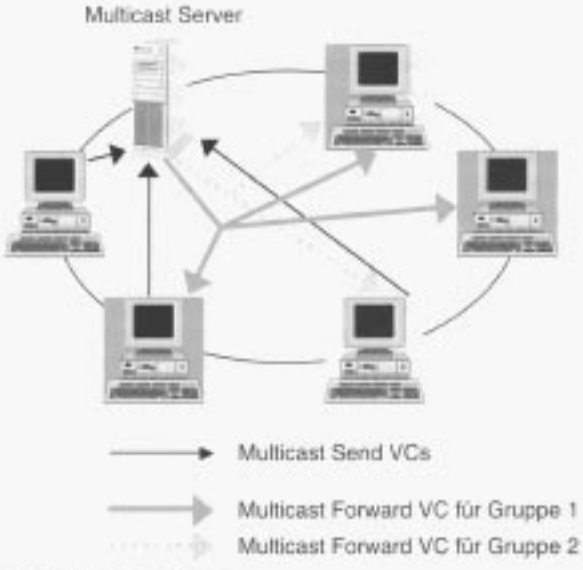

Abb. 8. Multicast Server

Im diskutieren Szenario sind die MARS-Erweiterungen nur dann notwendig, wenn die Endsysteme über ATM erreichbar sind. In Campus- und Unternehmensnetzen ist aber kaum ein Trend zu „ATM to the Desktop" festzustellen. Für den Netzanschluss von Arbeitsplatzrechnern werden eher Technologien wie Fast Ethernet zunehmend zum Einsatz kommen. Eine Ausnahme bildet vielleicht ADSL, wo zwischen dem DSL

Abb. 9. IP Multicast und ADSL

Access Multiplexer bis zu einem ADSL Gateway beim Teilnehmer ATM über ADSL verwendet wird (Abb. 9). Oft agieren solche Gateways als IP-Router und setzen ATM auf Ethernet um, so dass das Endsystem über Ethernet kommunizieren kann. Es ist dann die Aufgabe des Routers die über ATM empfangenen Multicast-Ströme via Ethernet an die Endsysteme weiterzuverteilen. Zwischen dem ADSL Gateway und dem IP Router kann dann eine ATM-Verbindung (Punkt-zu-Punkt-Link) aufgesetzt werden. Dieser Punkt-zu-Punkt-Link kann in diesem Fall durch das Aufsetzen eines Multicast-Tunnels zwischen den beiden Unicast-IP-Interfaces der beiden Router (IP Router und ADSL Gateway) überbrückt werden, so dass auf ein Multicast-fähiges Netz zwischen den IP Routern verzichtet werden kann. Hat der IP Router aber mehrere ADSL Gateways zu bedienen, muss er für jedes ADSL Gateway die Multicast-Daten duplizieren. Mit Hilfe von Punkt-zu-Mehrpunkt-ATM-Verbindungen und MARS könnte dieser Nachteil vermieden werden.

3.3 Multicast über Wave Divison Multiplexing-Netze

Als Backbone-Technologie für das Internet werden Wave Division Multiplexing- (WDM-)Netze zunehmend an Bedeutung gewinnen. WDM basiert auf der Mehrfachausnutzung von Glasfasern, d. h., über eine Glasfaser werden mehrere Wellenlängen gleichzeitig übertragen, wobei eine Wellenlänge als ein Kanal bezeichnet wird. Bei Unicast-Datenströmen müssen sich die WDM-Knoten, d. h. Sender und Empfänger auf einen Kanal einigen, über den dann die Daten ausgetauscht werden. Bei Multicast-Kommunikation soll wenn möglich die verfügbare Bandbreite wieder effizient ausgenutzt werden. Dies bedeutet, dass der Sender eine Wellenlänge zum Senden wählt und diese Wellenlänge mit entsprechenden optischen Einrichtungen an alle Multicast-Empfänger sendet. Die Empfänger müssen dann entsprechend benachrichtigt werden, damit sie die für den Multicast-Datenstrom verwendete Wellenlänge beachten. Diese Benachrichtigung und die Einstellung der Tuner können nicht vernachlässigbare Verzögerungen verursachen. Unter Umständen kann es daher geschickter sein, einen Multicast-Datenstrom an n Empfänger über n Unicast-Kanäle zu senden. Geeignete Scheduling-Strategien sind derzeit Gegenstand der Forschung [32].

4. Application Level Framing und zuverlässiger Multicast

Das IP Multicast-Konzept hat nicht nur Auswirkungen auf die darunter liegenden Schichten, sondern auch auf die darüber liegenden Transportprotokolle und Anwendungen. Dies liegt in erster Linie daran, dass IP Multicast einen unzuverlässigen Dienst anbietet, das TCP-Protokoll für Multicast aber nicht geeignet ist. TCP hat Mechanismen wie Staukontrolle und Übertragungswiederholung integriert, welche auf 1:1-Kommunikation (Unicast), aber nicht auf 1:n-Kommunikation (Multicast) abgestimmt sind. Daher kann TCP nicht über IP Multicast aufsetzen, sondern lediglich UDP, welches einen unzuverlässigen Dienst ohne Fluss- und Staukontrolle realisiert. Da UDP allerdings wenig zusätzliche Funktionalität im Vergleich zu IP anbietet, müssen von der Anwendung zusätzlich geforderte Funktionalitäten (z. B. zuverlässige Multicast-Kommunikation) oberhalb von UDP, d. h. in den Anwendungen selbst, implementiert werden.

Für solche zusätzlichen Funktionen wurden sogenannte Protokoll-Rahmenwerke entwickelt, die grundlegende Mechanismen und Paketformate bereitstellen, aber von den Anwendungen für ihre individuellen Anforderungen um spezielle Funktionen und Algorithmen erweitert werden müssen. Dieses Entwurfsprinzip wird auch Application Level Framing (ALF) [3] genannt. Für zuverlässige Multicast-Kommunikation wurde z. B. das Scalable Reliable Multicast- (SRM-) Protokoll entwickelt [5], für Audio/Video-Kommunikation hat sich das Real-Time Transport Protocol (RTP) etabliert. RTP standardisiert die Paketformate, in denen Audio/Video-Daten über UDP/IP transportiert werden [15, 16].

SRM, das Rahmenwerk für zuverlässige Multicast-Kommunikation, wird oft in verteilten Whiteboard-Anwendungen eingesetzt. SRM definiert Nachrichtenformate zur Anforderung und zur Übertragungswiederholung von Daten. Negative Quittungen werden auf Anwendungsebene beschrieben (z. B. „3. wb-Seite von Teilnehmer A") anstatt mit Sequenznummern, wie es in Transportprotokollen üblich ist. Eine Übertragungswiederholung kann von der Wichtigkeit der Daten innerhalb der Anwendung abhängen und wird unabhängig vom Senden neuer Daten durchgeführt. Beliebige Teilnehmer einer verteilten Anwendung

(also auch reine Empfänger) können Daten wiederholt senden, was zu einer Minimierung von Verzögerungen und zu einer Lastverteilung der Übertragungswiederholungen auf mehrere Instanzen führen kann. Übertragungswiederholungen werden auch zum Aktualisieren von der Gruppe später beigetretenen Teilnehmern verwendet. Problematisch ist bei SRM allerdings die Skalierbarkeit und die Tatsache, dass eine 100 %ige Zuverlässigkeit kaum möglich ist, da ein Sender nie die vollständige Kenntnis der Empfängermenge besitzt.

Zur Vermeidung dieser Probleme sind in den letzten Jahren einige vielversprechende Ansätze vorgestellt worden. Exemplarisch sei hierbei auf das Local Group Concept verwiesen [26]. Bei diesem Konzept übernehmen mehrere verteilte Controller die Überwachung der zuverlässigen Kommunikation. Dabei werden einem Controller mehrere untergeordnete Empfänger oder Controller einer niedrigeren Hierarchiestufe zugeordnet. Der übergeordnete Controller sorgt für die zuverlässige Verteilung der Multicast-Daten mit Hilfe von Übertragungswiederholungen. Gleichzeitig werden Quittungen der Empfänger eingesammelt und kumuliert, d. h., Quittungen werden an übergeordnete Controller bzw. den Sender kumuliert und nicht einzeln weitergegeben. Leider hat sich bislang keines der vorgeschlagenen Verfahren im Internet durchgesetzt. Eine neue IETF-Arbeitsgruppe „Reliable Multicast Transport" versucht, einen Standard zu entwickeln [25].

Ein weiteres offenes Problem ist das Thema Staukontrolle bei Multicast-Strömen, da UDP keine Staukontrollmechanismen unterstützt. Staukontrollmechanismen müssen dann auch wieder in die Anwendungen integriert werden. Aktueller Forschungsbedarf besteht hierzu aber auch hinsichtlich der Mechanismen, da Verfahren wie sie in TCP integriert sind, für Multicast ungeeignet sind. Beispielsweise ist es nicht sinnvoll, die Senderate zu reduzieren, wenn nur ein Teilbaum eines großen Multicast-Verteilbaums von Staus betroffen ist, andere Teilbäume aber nicht. Staus in einem Teilbaum würden dann zu einer Reduzierung der Datenrate und damit zu Qualitätseinbußen bei einer Videoübertragung über einen nicht von Staus betroffenen Teilbaum führen. Verfahren, die derzeit in Diskussion sind, basieren darauf, einen Datenstrom (z. B. Video) auf mehrere Teilströme zu verteilen. Diese Teilströme werden dann auf verschiedene Multicast-Gruppen abgebildet. Erkennen die Empfänger einen Stau, z. B. durch erhöhte Paketverluste, können sie eine oder mehrere Multicast-Gruppen verlassen, was dazu führt, dass diese Teilströme nicht mehr weiter über den belasteten Teilbaum übertragen werden. Die Arbeitsgruppe „Reliable Multicast Group Charter" [27] der Internet Research Task Force (IRTF) nimmt sich derzeit dieses Themas an.

5. Quality of Service

Ein derzeit sicherlich grosses Problem bei Netzen, die auf Internet-Technologie basieren, besteht darin, dass die Kommunikationsprotokolle keine Dienstgüten (Quality of Service, QoS), z. B. maximale Verzögerungszeiten oder Mindestbandbreiten, garantieren können. Speziell Audio/Video-Kommunikationsanwendungen haben hohe QoS-Anforderungen, z. B. Verzögerungen unter etwa 200 ms bei Audiokommunikation und Mindestbandbreiten von mehreren Mbit/s für hochqualitative Videokommunikation. Um diesen Mangel zu beheben, wurde in einer IETF-Arbeitsgruppe das Ressourcenreservierungsprotokoll Resource Reservation Setup Protocol (RSVP) [28] definiert, welches als eine Art Signalisierungsprotokoll zwischen IP-Endsystemen und Routern dient. Die bei einer Kommunikationsbeziehung beteiligten IP-Systeme können über RSVP QoS-Anforderungen für einzelne Anwendungen austauschen und die erforderlichen Reservierungen der Betriebsmittel (CPU-Zeit, Speicher) und Netzbandbreiten vornehmen. RSVP unterstützt Multicast durch das Empfänger-orientierte Prinzip, indem jeder Empfänger seine individuellen QoS-Bedürfnisse ausdrücken und dem nächsten stromaufwärts liegenden RSVP-Routern mitteilen kann. Dieser Router kann dann mehrere unterschiedliche QoS-Anforderungen diverser Empfänger mischen, indem das Maximum über mehrere Anforderungen ermittelt wird (z. B. die höchsten Bandbreiten, die niedrigste Verzögerung) und an den nächsten stromaufwärts liegenden Router weitergeleitet wird.

RSVP hat allerdings einige Skalierungsprobleme, weil Internet-Router, um das Konzept durchgängig unterstützen zu können, sich für jeden Datenstrom die QoS-Anforderungen merken müssten und durch geeignete Maßnahmen wie Bandbreitenreservierung oder CPU-Scheduling individuell unterstützen müssten. Aus diesem Grund wird derzeit ein besser skalierbares Konzept unter der Bezeichnung Differentiated Services entwickelt, wobei das Thema Multicast hierbei noch nicht intensiv bearbeitet wurde.

Schrifttum

[1] Ballardie, T., Francis, P., Crowcroft, J.: Core Based Tree (CBT), an architecture for inter-domain routing. ACM SIGCOMM'93, September 1993.

[2] Cain, B., Deering, S., Thyagarajan, A.: Internet Group Management Protocol, Version 3. Internet Draft, work in progress.

[3] Clark, D., Tenenhouse, D.: Architectural considerations for a new generation of protocols. ACM SIGCOMM '90, S. 200–208.

[4] Deering, S., Estrin, D., Farinacci, D., Jacobson, V., Helmy, A., Meyer, D., Wei, L.: Protocol Independent Multicast Version 2 dense mode specification. Internet Draft, work in progress.

[5] Floyd, S., Jacobson, V., McCanne, S.: A reliable multicast framework for light-weight sessions and Application Level Framing. ACM SIGCOMM'95, August 1995, S. 342–356.

[6] Frederick, R.: Experiences with real-time software video compression. 6th International Workshop on Packet Video, September 1994, S. F1.1-F1.4.

[7] Huitema, C.: Routing in the Internet. Prentice-Hall, 1995.

[8] ATM Forum: LANE v2.0 LUNI Interface, af-lane-0084.000. Juli, 1997.

[9] McCanne, S., Jacobson, V.: Vic: a flexible framework for Packet Video. ACM Multimedia '95, San Francisco, November 1995, S. 511–522.

[10] Waitzman, D., Partridge, C., Deering, S.: Distance Vector Multicast Routing Protocol. RFC 1075, November 1988.

[11] Deering, S.: Host extensions for IP multicasting. RFC 1112, August 1989.

[12] Moy, J.: OSPF Version 2. RFC1583. März 1994.

[13] Moy, J.: Multicast extensions to OSPF. RFC1584, März 1994.

[14] Moy, J.: MOSPF: analysis and experience. RFC1585, März 1994.

[15] Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V.: RTP: a transport protocol for real-time applications. RFC 1889, Januar 1996.

[16] Schulzrinne, H.: RTP profile for audio and video conferences with minimal control. RFC1890, Januar 1996.

[17] Armitage, G.: Support for multicast over UNI 3.0/3.1 based ATM Networks. RFC 2022, November 1996.

[18] Talpade, R., Ammar, M.: Multicast server architectures for MARS-based ATM multicasting. RFC 2149, Mai 1997.

[19] Ballardie, A.: Core Based Trees (CBT version 2) multicast routing – protocol specification. RFC 2189, September 1997.

[20] Ballardie, A.: Core Based Trees (CBT) multicast routing architecture. RFC 2201, September 1997.

[21] Laubach, M., Halpern, J.: Classical IP and ARP over ATM. RFC 2225, April 1998.

[22] Fenner, W.: Internet Group Management Protocol, Vers. 2. RFC 2236, November 1997.

[23] Estrin, D., Farinacci, D., Helmy, A., Thaler, D., Deering, S., Handley, M., Jacobson, V., Liu, C., Sharma, P., Wei, L.: Protocol Independent Multicast-Sparse Mode (PIM-SM): protocol specification. RFC 2362, Juni 1998.

[24] Jacobson, V., McCanne, S.: vat – LBNL Audio Conferencing Tool. *http://www-nrg.ee.lbl.gov/vat/.*

[25] IETF Arbeitsgruppe Reliable Multicast Transport (rmt): *http://www.ietf.org/ html.charters/rmt-charter.html.*

[26] Hofmann, M.: Skalierbare Multicast-Kommunikation in Weitverkehrsnetzen. Infix-Verlag. 1998.

[27] IRTF Arbeitsgruppe Reliable Multicast Group Charter: *http://www.ietf.org/charters/reliable-multicast.htm.*

[28] Braden, R., Zhang, L. Berson, S., Herzog, S., Jamin, S.: Resource ReSerVation Protocol (RSVP)-Version 1 functional specification. RFC 2205, September 1997.

[29] Braun, T.: IPng: Neue Internet-Dienste und virtuelle Netze. dpunkt.verlag. 1999

[30] Wittmann, R., Zitterbart, M.: Multicast: Protokolle und Anwendungen. dpunkt.verlag. 1999.

[31] Diot, C., Gautier, L.: A distributed architecture for multiplayer interactive applications on the Internet. IEEE Network Magazine vol 13 (Juli 1999), no. 4, S. 6–15.

[32] Ortiz, Z., Rouskas, G., Perros, H.: Scheduling of multicast traffic in tunable-receiver WDM networks with non-negligible tuning latencies. ACM SIGCOMM 97, S. 301–310.