Management-Architektur zur Unterstützung von Gruppenkommunikation in Differentiated-Services-Netzen

Roland Bless, Klaus Wehrle
Institut für Telematik, Universität Karlsruhe (TH)
Zirkel 2, D-76128 Karlsruhe, Germany
Tel.: +49 721 608-{6396,6414}, Fax: +49 721 388097
E-Mail: {bless,wehrle}@telematik.informatik.uni-karlsruhe.de

Zusammenfassung. Die derzeitigen Anstrengungen, das Internet mit neuartigen Mechanismen zur Unterstützung von Dienstgüte (Quality-of-Service) auszustatten, konzentrieren sich auf die „Differentiated-Services"-Architektur. Am Institut für Telematik wurden Dienste und Mechanismen dieser Architektur bereits implementiert und evaluiert. Dabei hat sich gezeigt, daß die angebotenen Dienstgüten skalierbar erbracht werden können, wenn das Management in den Differentiated-Services-Domänen (DS-Domänen) bestimmte Voraussetzungen erfüllt. Vor allem der auf „Expedited Forwarding" basierende Dienst (Premium Service) [JaNP99] bietet die garantierte Übertragung einer bestimmten Datenrate bei minimaler Verzögerung.

Der erste Teil dieses Beitrags zeigt, daß sich die Garantien des Premium-Service dazu nutzen lassen, eine zuverlässige und skalierbare Gruppenkommunikation (Reliable and Scalable Multicast) auf der Basis von IP-Multicast und einem einfachen Transportprotokoll zu erreichen. Die eigentlichen Weiterleitungsmechanismen zur Erbringung der Dienstgarantien wurden bereits für den Unicast-Fall untersucht [BlWe99] und wiesen bezüglich der Gruppenkommunikation keine Unterschiede zur Unicast-Kommunikation auf. Für Gruppenkommunikationsszenarien werden somit die gleichen Garantien geboten.

Es ergeben sich aber verwaltungsbedingte Probleme, wie etwa beim Beitritt neuer Empfänger zu einer Gruppe oder die Unterstützung heterogener Multicast-Gruppen. Mit diesen Problemen beschäftigt sich der zweite Teil des Beitrags, wobei mit dem Konzept des *Differentiated Services Domain Manager (DSDM)* ein Lösungsvorschlag präsentiert wird.

1 Einleitung und Motivation

Im Internet steigt der Bedarf an Kommunikationsdiensten, welche über mehr Dienstqualität als der bisherige Best-Effort-Dienst verfügen. Viele fortgeschrittene Anwendungen benötigen von der Netzwerkschicht gewisse Garantien, wie bestimmte maximale Verzögerungszeiten, geringe Paketverlustraten oder einen zugesicherten Mindestdurchsatz. Die derzeit eingesetzten IP-Mechanismen können solche Garantien nicht bieten, erst recht nicht, wenn zusätzlich Gruppenkommunikation gefordert wird.

Eine zuverlässige und skalierbare Multicast-Kommunikation durch zusätzliche Transportprotokollmechanismen zu ermöglichen, ist das Ziel von zahlreichen Anstrengungen, wie beispielsweise dem *Local Group Concept (LGC)* [Hofm98]. Wie die meisten Transportschicht-orientierten Konzepte setzt auch LGC Übertragungswiederholungen vom Sender oder einem naheliegenden Gruppenmitglied ein, um auftretende Paketverluste auszugleichen. Die erneute Übertragung eines Pakets bedeutet jedoch eine zusätzliche Verzögerung, weil erst einmal der Verlust des Pakets bemerkt werden muß (nach einem Timeout) und anschließend das Paket erneut angefordert werden muß. Die korrekte Auslieferung beim Empfänger beansprucht in

diesem Fall mindestens die dreifache Übertragungszeit. Diese setzt sich aus der Verzögerung beim ersten Senden und der Übertragungswiederholung zusammen, sowie der Zeit, welche für die Anforderung der erneuten Übertragung benötigt wird.

Interaktive Anwendungen tolerieren solche Verzögerungen im allgemeinen nicht, vor allem weil diese linear zur Entfernung vom Sender steigen. Einige Ansätze versuchen diese Zeit zu verringern, indem sie die Pakete nicht direkt beim Sender anfordern, sondern möglichst bei naheliegenden Rechnern. Jedoch ist nicht sichergestellt, daß die Pakete lokal vorhanden sind, weshalb die erneute Anforderung insgesamt noch länger dauern kann als die sofortige Anforderung beim Sender, weil erst eine lokale Paketwiederholung angestrebt wird, und falls diese fehlschlägt, eine globale.

Zusammenfassend läßt sich sagen, daß zuverlässige und skalierbare Multicast-Kommunikation, welche zusätzlich noch geringe Verzögerungen benötigt, auf der Grundlage von unzuverlässigem IP-Multicast bisher nicht zu erreichen ist.

Die Differentiated-Services-Architektur, welche derzeit von der Internet Engineering Task Force entwickelt wird [BBCD+98], stellt einen neueren Ansatz dar, um Dienstgüte skalierbar im Internet bereitzustellen. Der Einsatz der bisher definierten Weiterleitungsmechanismen, die zur Realisierung von Ende-zu-Ende Diensten verwendet werden, wurde im Zusammenhang mit Multicast-Kommunikation bisher noch kaum betrachtet. Aufgrund der Einfachheit der Architektur ergeben sich jedoch einige Probleme, die in Abschnitt 3.2 beschrieben sind. Der nächste Abschnitt gibt einen kurzen Überblick über die Konzepte der Differentiated-Services.

2 Differentiated Services

Der Differentiated-Services-Ansatz stellt eine Architektur zur Verfügung, welche durch geringe Modifikationen an den Routern im Internet die Verwendung verschiedener Klassen von Dienstgüten erlaubt. Sie betrachtet, im Gegensatz zur Integrated-Services-Architektur [BrCS94], nicht jeden Datenstrom einzeln, sondern behandelt die Ströme eines Dienstes aggregiert, um so eine bessere Skalierbarkeit – vor allem im Netzinnern – zu erreichen. Weiterhin wird die unvermeidliche Komplexität an die Netzgrenzen verlagert. So wird lediglich im ersten Router (*First-Hop-Router*) und nicht in jedem Zwischensystem festgestellt, zu welchem Dienst und zu welcher Reservierung ein Paket gehört. Dieser unterscheidet (als einziger Router) einzelne Datenströme und ermittelt das jeweils zugehörige *Weiterleitungsverhalten (Per Hop Behavior – PHB)* auf dessen Basis der Dienst realisiert wird. Gleichzeitig überprüft er die Konformität eines Datenstroms mit einem zuvor vereinbarten Verkehrsvertrag, paßt ihn eventuell an und vermerkt dann die Kennung für das Weiterleitungsverhalten (*Codepoint*) im IP-Paketkopf. Hierfür wurde durch RFC 2474 [BBBN98] das bisher als TOS-Byte bezeichnete Feld in DS-Byte umbenannt.

Die nachfolgenden Router einer Differentiated-Services-Domäne (*Interior Router*) führen keine aufwendige Klassifikation mehr durch. Sie betrachten lediglich den Codepoint im DS-Feld und behandeln die Pakete eines Stroms nur noch anhand des dadurch bezeichneten zugehörigen Per-Hop-Behaviors. Durch diese Vereinfachung wird eine deutliche Reduktion der Komplexität der Router im Netzinnern erreicht, welche keine einzelnen Datenströme mehr unterscheiden, sondern nur noch Dienstklassen, deren Anzahl auf 64 beschränkt ist. Sie benötigen vor allem keinerlei Reservierungsinformationen mehr.

An den Netzgrenzen findet in den sogenannten *Border-Routern* eine erneute Überprüfung und Anpassung der Verkehrsströme statt. Es wird aber nicht jeder Strom einzeln, sondern wiederum alle Ströme eines Dienstes aggregiert betrachtet.

2.1 Premium Service

Die Differentiated-Services-Architektur bietet mit dem auf *Expedited-Forwarding (EF)* basierenden *Premium-Service* [JaNP99] einen Dienst mit sehr niedrigen Paketverlustraten und minimalen Verzögerungszeiten an. Im Rahmen des *UNIQuE*-Projekts [UNIQ99, KIDS99] wurde der Premium-Service implementiert und evaluiert [BlWe99]. Es hat sich dabei gezeigt, daß der Premium-Service skalierbar ist und eine harte Garantie der reservierten Bandbreite bei minimalen Verzögerungszeiten und praktisch keinen Paketverlusten gewährt. Dieser Dienst eignet sich daher als Basis zur Realisierung des *RMPS-Konzepts (Reliable Multicast based on Premium Service)*. Es erlaubt eine zuverlässige und skalierbare Gruppenkommunikation und wird im nächsten Abschnitt noch genauer vorgestellt.

3 RMPS – Dienstgüte für skalierbare Gruppenkommunikation

In [BlWe99] wurden bereits umfangreiche Untersuchungen zu Premium-Service vorgestellt. Dabei wurden keinerlei Paketverluste beobachtet, obwohl diese – zumindest theoretisch – aufgrund von Aggregationseffekten nicht gänzlich auszuschließen sind. Die gemessenen Verzögerungszeiten der Premium-Pakete waren minimal.

Zur Untersuchung der Eigenschaften von Premium-Service in Gruppenkommunikationsszenarien wurde die Differentiated-Services-Implementierung KIDS (*Karlsruhe Implementation of Differentiated Services*) anschließend um die Multicast-Funktionalität erweitert. Anwendungen könnten daher auf komplexe zusätzliche Mechanismen in der Transportebene verzichten, wenn die durch den Premium-Service angebotene Zuverlässigkeit ausreicht. Zur Überprüfung der Zuverlässigkeit des „Multicast-Premium-Services“ wurden einige Versuche aus [BlWe99] mit Multicast-Paketströmen wiederholt.

3.1 Evaluierung

Wie Abbildung 1 zeigt, arbeiten die Weiterleitungsmechanismen der Differentiated-Services-Architektur in Schicht 2 der Zwischensysteme, und zwar auf der (Ausgangs-)Warteschlange eines Netzwerkadapters. Weil die Paketreplikation schon vor den eigentlichen Weiterleitungsmechanismen in der IP-Schicht stattfindet, ist die Erbringung der Dienstgüte unabhängig von der Tatsache, ob es sich um Unicast- oder Multicast-Pakete handelt. Vor der Replizierung der Pakete findet in First-Hop-Routern noch die Klassifikation und Verkehrskontrolle statt.

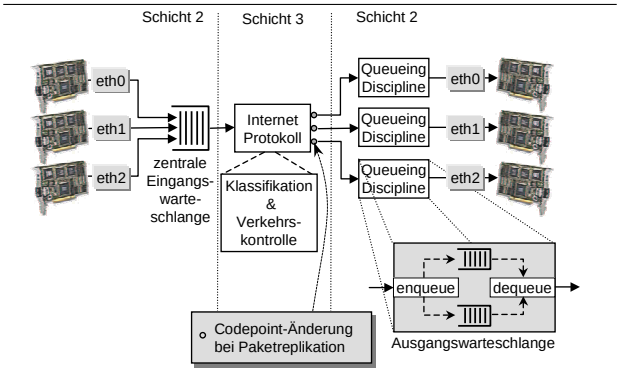

Abbildung 1 Architektur eines Differentiated-Services-Routers

Abbildung 3 zeigt die Ergebnisse einiger Messungen, die von den Autoren durchgeführt wurden. Dazu wurde das in Abbildung 2 skizzierte Testnetz, bestehend aus einem Multicast-Router, zwei Sendern und zwei Empfängern aufgebaut.

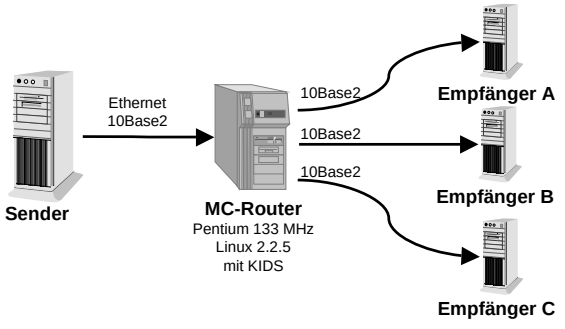

Abbildung 2 Struktur des Testbeds

Der Software-Router arbeitete mit der um die Multicast-Fähigkeit erweiterten KIDS.

In Übereinstimmung mit den Versuchen aus [BlWe99] wurden UDP-Datenströme mit den Raten 100 kbps, 500 kbps und 1 Mbps erzeugt, welche der Sender an die Multicast-Gruppe 233.1.1.1 schickte. In den verschiedenen Durchläufen waren jeweils ein bis drei Empfänger an der Gruppe beteiligt. Abbildung 3 zeigt die Paketzwischenankunftszeit (zeitl. Abstand zweier aufeinanderfolgender Pakete), welche bei den jeweiligen Empfängern (A und B) gemessen wurden. Dadurch, daß der Router als First-Hop-Router konfiguriert war, wird beim Senden der Pakete eine Glättung des Verkehrs (Traffic Shaping) durchgeführt. In der Abbildung ist dies an der fast gleichmäßigen Verteilung der Zwischenankunftszeit zu erkennen. Die geringen Schwankungen, d.h. der Raten-Jitter von durchschnittlich 120 μs ergibt sich durch die Ungenauigkeit des Timers im Router, welche bei 244 μs liegt.

Bei den Versuchen unter voller Last (jeder Link wurde mit 10 Mbps Best-Effort belastet) wurde ein Jitter von maximal 852 μs gemessen, welches der zeitlichen Länge eines Pakets auf dem Medium (Paketzeit) entspricht. Die Ursache dieses Jitters wurde bereits in [Wehr99] und [BlWe99] ermittelt und darin begründet, daß das zu sendende Premium-Service-Paket warten muß, bis das sich in Übertragung befindliche vorige Paket vollständig auf dem Medium ausgegeben ist. Die gemessenen 852 μs entsprechen genau der zeitlichen Länge eines UDP-Pakets mit 1000 byte Nutzdaten auf einem 10Base2-Segment. Es sei noch erwähnt, daß während den Messungen Kollisionen auf dem Ethernet-Segment vermieden wurden, weil diese die Ergebnisse verfälscht hätten.

Die Ergebnisse der Versuche in Abbildung 3 zeigen, daß sich die in [Wehr99, BlWe99] aufgezeigten Eigenschaften des Premium-Service (bzw. Expedited Forwarding) auch auf die Gruppenkommunikation übertragen lassen. Die Ergebnisse zeigen vor allem auch, daß sich durch die Paketreplikation keinerlei zusätzliche Verzögerungen ergeben. Premium-Service bietet somit auch für die Gruppenkommunikation eine harte Garantie der reservierten Bandbreite bei minimalen Verzögerungszeiten und praktisch keinen Paketverlusten.

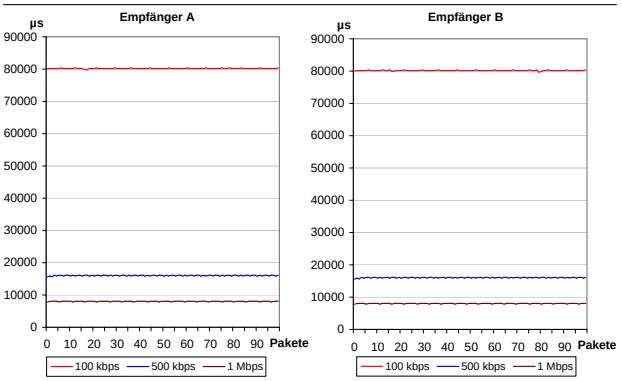

Abbildung 3 Zwischenankunftszeit der Pakete bei Multicast-Premium-Services

Für ein korrektes Funktionieren des Premium-Service sind jedoch einige Bedingungen seitens des Managements vorausgesetzt. Unter anderem ist eine strikte Zugangskontrolle notwendig, welche sicherstellt, daß nur so viel EF-Verkehr in eine Domäne eintritt, wie vereinbart ist. Auf diese Punkte wird in Abschnitt 4 genauer eingegangen.

3.2 Probleme von IP-Multicast in DS-Domänen

Bei der Unterstützung von IP-Multicast in Differentiated-Services-Domänen ergeben sich einige grundlegende Probleme, die im folgenden beschrieben werden.

3.2.1 Empfänger-initiierte Reservierungen

Differentiated-Services sind eher Sender-orientiert, d.h. Pakete erfahren den durch den sogenannten *„Codepoint“* [BBBN98] im Paketkopf angegebenen Dienst „stromabwärts“, wobei der Codepoint zu Anfang durch den First-Hop-Router oder den Sender selbst gesetzt wird. Die Bezeichnung „Sender-orientiert“ bezieht sich in diesem Zusammenhang daher nicht auf die eigentliche Reservierung des Dienstes, sondern auf

die Art und Weise, wie der Dienst dem Paketstrom zugewiesen wird. Die Reservierung kann, wie im Verlauf dieses Beitrags noch gezeigt wird, durchaus auf Veranlassung des Empfängers erfolgen. Jedoch wird das gewünschte Weiterleitungsverhalten immer im First-Hop-Router festgelegt und den Paketen zugeordnet.

Dazu wird im First-Hop-Router des Senders ein *Verkehrsprofil* gespeichert. Es beinhaltet Beschreibungsdaten der Charakteristik des Verkehrsstroms, welche in einem *Dienstvertrag (Service Level Agreement – SLA)* mit dem *Internet Service Provider (ISP)* vereinbart wurden. Durch weitere bilaterale Dienstverträge zwischen den ISPs auf dem Weg zum Ziel werden die Ressourcen Ende-zu-Ende reserviert und die Garantien können Ende-zu-Ende erbracht werden. Möchte ein Empfänger einen Dienst nutzen, so muß auf seine Anforderung hin im First-Hop-Router des Senders ein entsprechendes Profil etabliert und die Bereitstellung der notwendigen Ressourcen entlang des Weges durch weitere SLAs gesichert werden.

Gruppenkommunikationsszenarien zeichnen sich durch dynamische Empfängermengen aus, in welchen ein Empfänger einer Gruppe beitritt, ohne den Sender darüber zu informieren. Er weiß also nicht, welche Empfänger derzeit der Gruppe angehören. D.h. es muß ein Mechanismus bereitgestellt werden, damit die Verkehrsprofile vom Sender ausgehend entlang des Pfades eingerichtet bzw. aktualisiert werden. Die vorausgehende Zugangskontrolle kann allerdings erst durchgeführt werden, wenn der Empfänger seinen Beitrittswunsch zur Gruppe äußert, d.h. es macht für die meisten Fälle wenig Sinn, statische Profile vorab einzurichten. Überdies ist in vielen Fällen eine Abrechnung zu Lasten der Empfänger erwünscht.

3.2.2 Das Neglected Reservation Subtree-Problem (NRS-Problem)

Die inneren Router von DS-Domänen sind sehr einfach aufgebaut, weshalb sie über keinerlei weitergehende Klassifikationsmechanismen bzw. Verkehrsprofile verfügen und folglich auch keine Überwachung des Verkehrs durchführen. Bei der Vervielfältigung der Pakete durch IP-Multicast wird auch der Codepoint im IP-Paketkopf kopiert, so daß alle Paketreplikate den gleichen Dienst erfahren. Der Router kann allerdings nicht prüfen, ob für das replizierte Paket auch tatsächlich eine entsprechende Reservierung auf der Ausgangsleitung vorliegt. Ihm fehlen dazu die entsprechenden Informationen der Verkehrsprofile.

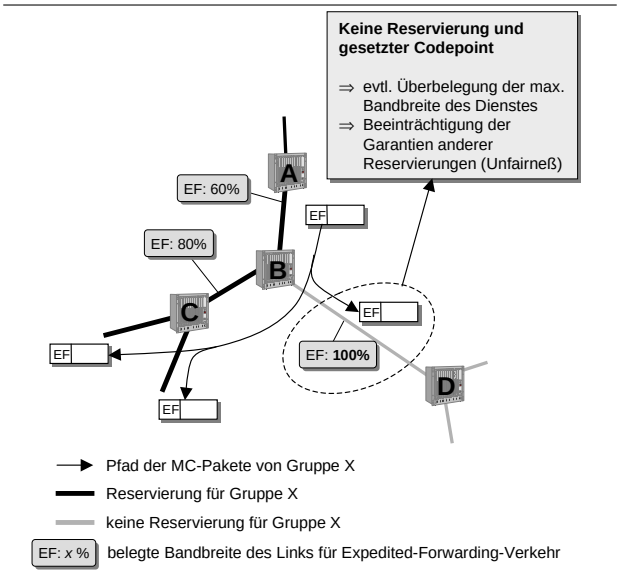

Abbildung 4 Das Neglected Reservation Subtree-Problem

Bleibt beim Hinzufügen von Empfängern eine solche implizite Dienstnutzung durch einfaches Erweitern des Multicast-Baumes unberücksichtigt, kann die Dienstqualität anderer Nutzer eines Mehrwertdienstes auf der betroffenen Ausgangsleitung aufgrund einer Überbelegung der Ressourcen beeinträchtigt werden (siehe Abb. 4).

Diese Gefährdung bestehender Dienstverträge durch eine unterlassene Reservierung – im weiteren als *Neglected Reservation Subtree-Problem (NRS-Problem)* bezeichnet – muß auf jeden Fall vermieden werden. Das Management einer DS-Domäne muß deshalb dafür Sorge tragen, daß innerhalb einer Gruppe, die differenzierte Dienste nutzt, nur dann Pakete auf den Pfaden des

Multicast-Baumes der Gruppe gesendet werden, wenn dafür auch entsprechende Ressourcen reserviert sind bzw. bestehende Dienstverträge dadurch nicht verletzt werden. In einer homogenen Gruppe, d.h. alle Teilnehmer erhalten den gleichen Dienst, wird dies am einfachsten dadurch erreicht, daß nur Teilnehmer zu einer Gruppe hinzugenommen werden, wenn für sie zuvor eine Reservierung durchgeführt wurde, bzw. die Reservierung eines anderen Gruppenmitglieds mitbenutzt werden kann. Auf die Probleme bzw. Fairneß bei der Tarifierung wird im folgenden nicht eingegangen.

3.2.3 Unterstützung heterogener Multicast-Gruppen

Heterogene Multicast-Gruppen enthalten einen oder mehrere Empfänger, die einen anderen Dienst bzw. Dienstqualität als diejenige, die der Sender vorgegeben hat, erhalten möchten. So sollten z.B. Empfänger, die lediglich einen Best-Effort-Dienst wünschen, auch an einer Gruppenkommunikation teilnehmen können, die sonst einen höherwertigen Dienst wie beispielsweise Premium-Service verwendet. Die Betrachtungen beschäftigen sich in diesem Beitrag nur mit unterschiedlichen Diensten, nicht aber mit unterschiedlich geforderten Dienstqualitätsparametern innerhalb eines Dienstes.

In heterogenen Gruppen müssen folglich einige der inneren Router einer DS-Domäne bei der Paketreplikation den Wert des Codepoints ändern und die neu erzeugten Pakete dem zugehörigen Dienst entsprechend weiterleiten. Dies erfordert in der Routing-Tabelle allerdings eine gesonderte Kennzeichnung, die beim Hinzufügen des neuen Multicast-Teilbaums angezeigt werden muß. Somit ist die Erweiterung des verwendeten Multicast-Routing-Protokolls oder zumindest der Router-Konfigurationsfunktionalität notwendig. Weiterhin lassen sich so keine unterschiedlichen Dienstqualitäten in Abhängigkeit der Port-Zieladresse erreichen, weil der Eintrag in der Routing-Tabelle sich lediglich auf Senderadresse, Gruppenadresse und Ausgangsschnittstelle bezieht (bei Verwendung von Dense-Mode-Routing-Protokollen).

3.2.4 Dynamischer Senderwechsel

Innerhalb einer IP-Multicast-Gruppe kann *jeder* Teilnehmer prinzipiell als Sender agieren. Für jeden durch einen Sender implizierten Multicast-Baum müssen die Ressourcen separat reserviert werden, falls gleichzeitiges Senden mit einem Dienst möglich sein soll. Dies ist zum einen darin begründet, daß DS-Dienste Simplex-Charakter besitzen, und zum anderen dadurch, daß die für einen Dienst reservierten Ressourcen innerhalb eines Multicast-Baumes nicht für das Verkehrsaufkommen von mehreren gleichzeitig sendenden Teilnehmern ausreicht (NRS-Problem). First-Hop-Router sollten deshalb immer in Abhängigkeit der Sender- und Gruppenadresse klassifizieren. Hierdurch kann überprüft werden, für welche Sender eine Reservierung vorliegt. Sendet ein Teilnehmer der Multicast-Gruppe, ohne daß in seinem First-Hop-Router ein entsprechendes Profil vorhanden ist, d.h. es liegt keine Reservierung vor, so wird im DS-Feld der Codepoint von Best-Effort-Service eingetragen und diese Pakete erfahren in allen Domänen keinen mehrwertigen Dienst. Gibt es hingegen eine Reservierung und somit auch ein Verkehrsprofil, so erfahren die Pakete in allen Domänen diesen Dienst.

Gleiches gilt für halbduplex Verkehr zur gemeinsamen Nutzung der reservierten Ressourcen durch mehrere Sender, weil durch das Netzwerk nicht sichergestellt werden kann, daß zu jedem Zeitpunkt nur genau ein Sender Pakete an die Gruppe schickt.

Will ein Empfänger mit einem anderen Dienst an die Gruppe senden, müßte jeder Teilnehmer erneut eine explizite Reservierung veranlassen. Im speziellen Fall des Best-Effort-Dienstes ist es für die Empfänger aber trotzdem ohne zusätzliche Mechanismen wie bisher möglich, Pakete an die Gruppe zu senden, weil diese aufgrund fehlender Profile im First-Hop-Router als Best-Effort-Pakete markiert und behandelt werden.

Zur Bereitstellung von Multicast-Kommunikation mit Differentiated-Services und zur Lösung der zuvor beschriebenen Probleme kann eine Management-Architektur eingesetzt werden, die außerdem weitere Vorteile bietet. Die Hauptkomponente dieser Architektur wird im folgenden Abschnitt vorgestellt.

4 Das Konzept der Differentiated-Services-Domain-Manager

Damit Ende-zu-Ende-Dienste auf der Basis von Differentiated-Services erbracht werden können, müssen für einige solcher Dienste (z.B. Premium-Service) Zugangskontrollen durchgeführt werden, um die geforderte Funktionalität und Leistung überhaupt zu Erzielen. Weiterhin müssen Verkehrsprofile in den Grenz-Routern installiert werden, so daß von ihnen das tatsächliche Verkehrsaufkommen eines Dienstes anhand der vereinbarten Menge im Verkehrsprofil überwacht werden kann. Um Dienste auf Anforderung bereitzustellen, sollte die Einrichtung und Aktivierung eines solchen Verkehrsprofils auch dynamisch erfolgen können.

Die statische Bereitstellung von Diensten durch die vom Netzwerkadministrator einer Domäne durchgeführte manuelle Konfiguration der Router (insbesondere die Einrichtung der Verkehrsprofile) mag zwar für einige wenige Fälle praktikabel sein, ist jedoch für eine größere Anzahl von Dienstnutzern zu aufwendig. Weiterhin wird bei einer weitgehend statischen Ressourcenreservierung für Dienste die Auslastung des Netzes nicht optimal sein, weil die Ressourcen vermutlich nicht immer voll genutzt werden.

Ein geeignetes Management kann die notwendige Funktionalität zur Erfüllung solcher Aufgaben bereitstellen. Gerade im Zusammenhang mit den in Gruppenkommunikationsszenarien häufig anzutreffenden dynamischen Empfängermengen müssen Zugangskontrollen bei Dienstanforderung des Empfängers durchgeführt und die entsprechenden Verkehrsprofile eingerichtet oder aktualisiert werden.

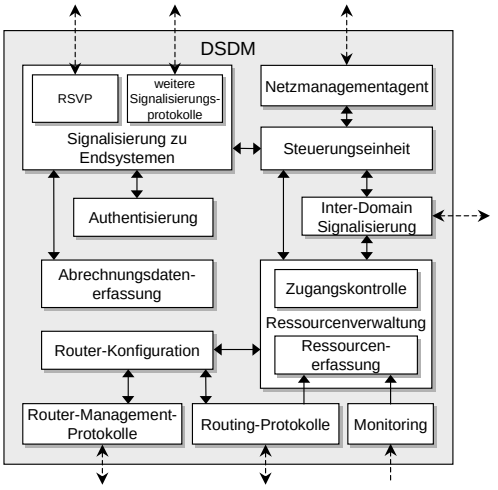

Abbildung 5 Logische Architektur eines DS-Domain-Managers

Um eine automatisierte Zugangskontrolle und Installation der Verkehrsprofile in den Grenz-Routern vorzunehmen, wurde in [JaNZ99] die Verwendung mehrerer sogenannter *Bandwidth-Broker* vorgeschlagen. Sie sollen sowohl Ressourcenanforderungen der Dienstteilnehmer entgegennehmen als auch die bilateralen Verhandlungen zwischen benachbarten Domänen durchführen. Während der Bandwidth-Broker aber hauptsächlich zur Verwaltung und Aushandlung der Ressourcen vorgesehen ist, verfolgt der hier beschriebene Ansatz ein integratives *Management der Dienste*. Teilnehmer erwarten schließlich einen Ende-zu-Ende Dienst und fordern diesen beim ISP mit den gewünschten Parametern an. Welche Weiterleitungsverhalten intern zur Realisierung des Dienstes verwendet werden, bleibt den jeweiligen ISPs überlassen, sofern das Weiterleitungsverhalten die geforderten Eigenschaften und Leistungsmerkmale aufweist, um den Dienst von Ende-zu-Ende zu erbringen.

Der *Differentiated-Services-Domain-Manager (DSDM)* ist eine logische Verwaltungseinheit einer DS-Domäne, welche für die folgenden Aufgaben verantwortlich ist (siehe Abbildung 5):

- *Zugangskontrolle* — Wie bereits zuvor erwähnt, benötigen einige der DS-Dienste

eine Zugangskontrolle, um überhaupt zu funktionieren. So muß beispielsweise beim Premium-Service garantiert werden, daß die Bandbreite des aggregierten Eingangsverkehrs kleiner oder höchstens gleich der für diesen Dienst zur Verfügung stehenden Bandbreite des Ausgangs ist [JaNP99]. Dadurch ist die Ausgangswarteschlange des Premium-Service nahezu immer leer. Weiterhin darf die aggregierte Bandbreite eines Dienstes innerhalb einer Domäne üblicherweise einen bestimmten Anteil (ca. 20-40%) der gesamten Bandbreite auf einer physikalischen Netzwerkteilstrecke nicht überschreiten [Wehr99, BlWe99], um die anderen Dienste nicht zu stark zu benachteiligen.

Vor Benutzung eines Dienstes muß dieser daher vom Teilnehmer angefordert und die Erfüllbarkeit bezüglich der Ressourcen entlang des Pfades vom Sender zum Empfänger geprüft werden. So prüft sukzessive jeder DSDM in jeder Domäne die Verfügbarkeit der Ressourcen und gibt die Anfrage der nächsten Nachbardomäne entlang des Pfades zum Empfänger weiter. Prinzipiell muß dies für jeden Datenstrom durchgeführt werden, allerdings können die gleichzeitig gestellten Anforderungen mehrerer Datenströme des selben Dienstes mit gleichem Zielnetzwerk zusammengefaßt werden. Außer der Ressourcenverfügbarkeit spielt noch die Einhaltung der zugrundeliegenden Politik des Domänenverwalters eine Rolle. So wird beispielsweise Verkehr, der aus bestimmten anderen Domänen stammt, aus (wirtschafts-)politischen Gründen nicht durch die eigene Domäne weitergeleitet. Dies geschieht häufig, um die eigenen Ressourcen vor Mißbrauch durch andere ISPs zu schützen. Die Zugangskontrolle umfaßt daher auch solche politischen Aspekte.

- *Ressourcenverwaltung* — Für die Zugangskontrolle ist eine konsistente Sicht der Ressourcen notwendig, die bei einer vollständig verteilten Ressourcenvergabe nur mit extremem Aufwand zu realisieren wäre. Der DSDM als zentrale Vergabestelle für Netzwerkressourcen innerhalb der Domäne verfügt aber zwangsläufig über eine konsistente Sicht. In Verbindung mit Routing-Protokollen und dem Netzwerkmanagement kann der DSDM die Topologie und den Zustand des Netzwerks erfassen. Weiterhin benötigt er Angaben von der Netzwerkadministration über die Ressourcenkontingentierung (z.B. maximale Bandbreite der physikalischen Verbindungen, Zuteilung der Bandbreite zu einzelnen Diensten). Vergibt der DSDM beispielsweise Bandbreite für einen Dienst, müssen die inneren Router nicht über diese Zuteilung informiert oder gar konfiguriert werden, wodurch eine deutliche Entlastung der Komponenten im Kernnetz von Verwaltungsaufgaben stattfindet.

- *Konfiguration der Router* — In Grenz-Router müssen u.a. Verkehrsprofile eingebracht werden, anhand derer die Konformität des Verkehrs überprüft und notfalls wiederhergestellt wird (z.B. durch Verwerfen oder Verkehrsformung).

- *Signalisierung mit Dienstnutzern* — Empfänger müssen die Möglichkeit haben, Ressourcen für einen Dienst anzufordern, d.h. die gewünschten Verkehrsprofile müssen der Domänen-Administration mitgeteilt werden. In einem voll-dynamischen Szenario geschieht dies unmittelbar vor der Dienstnutzung und kann mittels eines speziellen Reservierungsprotokolls oder z.B. aus Kompatibilitätsgründen mittels RSVP [BBHJ$^+$97] erfolgen. Die Pfad-Nachrichten von RSVP sind jedoch insbesondere in Unicast-Kommunikationsszenarien nicht notwendig.

- *Inter-Domänen Signalisierung* — Dienstanforderungen müssen im Rahmen der Zugangskontrolle auch von und zu Nachbar-Domänen übertragen werden können (siehe Abbildung 6). Die Kommunikation zwischen DSDMs erfolgt quasi direkt, denn physikalisch zwischengeschaltete Router nehmen

keine Verarbeitung dieser Nachrichten vor, sondern leiten sie nur weiter.

- *Authentisierung* — Im Zusammenhang mit der Signalisierung muß für eine Authentisierung gesorgt werden, damit eine Dienstnutzung auch nachweisbar abgerechnet werden kann. Zwischen Nachbardomänen können „vereinfachte“ Verfahren verwendet werden, da ihre Beziehungen zueinander eher von statischer Natur sind. So sollte beispielsweise eine Integritätssicherung der zwischen Domänen ausgetauschten Signalisierungsnachrichten erfolgen, durch die implizit die Authentizität einer jeden Nachricht nachgewiesen wird.

- *Dienstabrechnung* — Da sämtliche Dienstnutzungen dem DSDM bekannt sind, lassen sich die abrechnungsrelevanten Kommunikationsdaten dort zentral sammeln.

- *Mobilitätsmanagement* — Sollen auch mobile Teilnehmer unterstützt werden, sind zusätzliche Mechanismen, wie beispielsweise Reservierungen für Dienstnutzungen im voraus zur Vorbereitung eines Teilnehmerwechsels zwischen zwei Domänen (Handover), vorzusehen.

- *Netzmanagement* — Da über den DSDM die Konfiguration der gesamten DS-Domäne gesteuert wird, müssen die Kernfunktionen des traditionellen Netzmanagements für den Netzwerkadministrator bereitgestellt werden: Konfigurationsmanagement, Fehlermanagement, Abrechnungsmanagement, Sicherheitsmanagement und Leistungsmanagement.

Weil viele der Aufgaben andernfalls größtenteils durch die Grenz-Router bewerkstelligt werden müßten, entlastet der DSDM diese von den Managementaufgaben. Das Konzept des DSDM geht deutlich über die Funktionalität der in [JaNZ99] beschriebenen Bandwidth-Broker hinaus, weil es das *Management der Dienste* und nicht nur der Ressourcen bzw. Profile umfaßt.

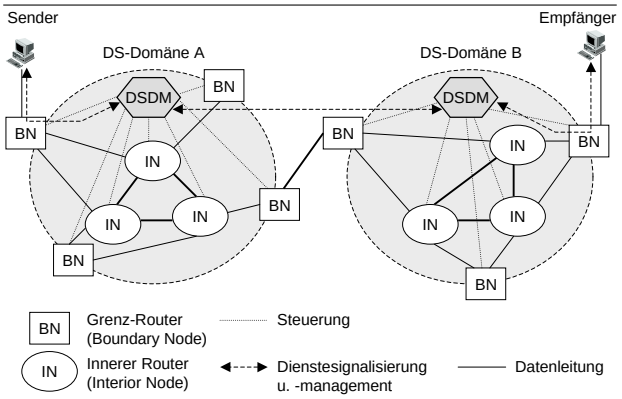

Abbildung 6 Dienstaushandlung über Differentiated-Services-Domain-Manager

Obwohl ein Ausfall eines DSDM die bereits etablierten Dienste nicht beeinträchtigt, kann er aus Gründen der Fehlertoleranz und Leistungssteigerung physikalisch durch mehrere Einheiten realisiert werden. Der DSDM operiert nicht im Datenpfad, d.h. er wird lediglich beim Verbindungsaufbau bzw. -abbau eingesetzt, so daß seine Leistung die Anzahl der vermittelten Verbindungen pro Zeiteinheit und die Dauer eines Verbindungsaufbaus beeinflußt.

5 Lösungsansätze

Mit Hilfe des DSDM ergeben sich folgende Lösungen für die im Abschnitt 3.2 angeführten Multicast-Probleme.

5.1 Beseitigung des NRS-Problems

Wenn ein Teilnetz durch eine „Graft“/„Join“-Nachricht einen Beitrittswunsch zur Multicast-Gruppe übermittelt (Vorgang ① in den Abbildungen 7 und 8), so passiert dieser eventuell zunächst die Grenz-Router und wird anschließend von den inneren Routern der DS-Domänen bearbeitet.

Derjenige Router, welcher den Anknüpfungspunkt an den bereits bestehenden Multicast-

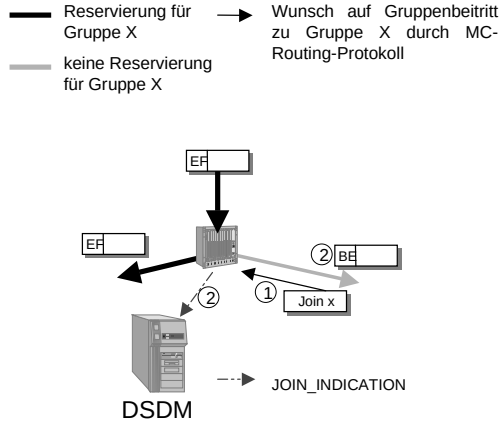

Abbildung 7 Bearbeitung eines Beitrittswunsches – ohne Reservierung

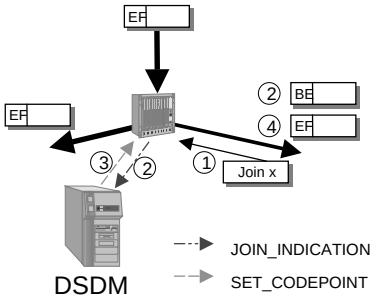

Abbildung 8 Gruppenbeitritt bei bestehender Reservierung

Baum darstellt, leitet die Pakete der Gruppe zum neuen Teilbaum aber vorerst ohne Dienstqualität weiter, d.h. als Best-Effort- oder gegebenenfalls als Lower-Than-Best-Effort-Verkehr, um eine Benachteiligung der übrigen Datenströme in der Best-Effort-Dienstklasse zu vermeiden. Dazu müssen die Codepoints der eintreffenden Pakete für die entsprechende Ausgangsschnittstelle auf Best-Effort umgesetzt werden, wenn sie zuvor einen höherwertigen Dienst erfahren haben. Die Umsetzung auf einen anderen Codepoint in Abhängigkeit von der Ausgangsschnittstelle wird in der Routing-Tabelle zusätzlich beim entsprechenden Eintrag der Gruppenadresse vermerkt. Das Umsetzen der Codepoints kann deshalb innerhalb eines Routers für jede Gruppe und für jede Ausgangsschnittstelle getrennt erfolgen. Wie in Abbildung 1 dargestellt wird, findet die Paketreplikation auf Schicht 3 statt, während die einzelnen Weiterleitungsmechanismen der DS-Dienste nach der Replikation auf Schicht 2 arbeiten. Somit muß lediglich das Multicast-Routing dahingehend erweitert werden, daß bei der Paketreplikation zuvor noch der Codepoint umgesetzt wird. Innerhalb der Routing-Tabelle existiert üblicherweise bereits für jede durch den Router weitergeleitete Multicast-Gruppe ein Eintrag, in welchem jede Schnittstelle aufgeführt ist, auf welcher Pakete für die Gruppe ausgegeben werden. Diese Einträge müssen für die Unterstützung von Differentiated-Services lediglich um ein Byte ergänzt werden, welches den zu setzenden Codepoint enthält. Es entsteht also ein vernachlässigbarer zusätzlicher Aufwand an Bearbeitungszeit und Speicherplatz. Die Kontrolle über den Inhalt dieses Codepoint-Feldes besitzt der DSDM.

Wird durch eine Join-Nachricht eine Schnittstelle für eine Multicast-Gruppe aktiviert, so wird der Eintrag standardmäßig zunächst auf (Lower-Than-)Best-Effort gesetzt (② in den Abbildungen). Dadurch wird erstens verhindert, daß das NRS-Problem auftritt, und zweitens wird den Empfängern die Möglichkeit gegeben, erst einmal in die Gruppe „reinzuhören“, ohne daß sie irgendwelche Dienstqualitäten in Anspruch nehmen. Es muß jedoch dafür gesorgt werden, daß die neu markierten Pakete keine Unfairneß in der neuen Dienstklasse Best-Effort hervorrufen, weil sie mit einer höheren Qualität in diese Dienstklasse gelangen und somit den übrigen Verkehr in dieser Klasse verdrängen bzw. benachteiligen. Deshalb könnten Pakete in diesem Fall in eine „Lower-Than-Best-Effort“-Dienstklasse eingeordnet werden, die eine solche Benachteiligung vermeidet.

Bei Ankunft einer Join-Nachricht wird an den DSDM eine `JOIN_INDICATION`-Nachricht gesendet, welche ihn über das Join-Ereignis informiert (ebenfalls ② in den Abbildungen). Der DSDM überprüft daraufhin, ob für diese Gruppe und die betroffene Ausgangsschnittstelle des Routers bereits eine Reservierung vorliegt. Es muß sich dabei um den gleichen Dienst handeln und außerdem darf die reservierte Bandbreite auch

nicht geringer sein, als im bestehenden Multicast-Baum. Wäre dies der Fall, so würde auf dem neuen Teilbaum der gleiche Dienst erbracht, wie auf dem Rest des Baumes, obwohl nur ein Teil der Bandbreite reserviert wurde. Es tritt also wieder das NRS-Problem auf. Um dies zu verhindern, beläßt der Router in diesem Fall einfach den voreingestellten Eintrag im Codepoint-Feld der Routing-Tabelle des Routers, wodurch alle Pakete als Best-Effort umdeklariert werden.

Wenn nun eine ausreichende Reservierung für den neuen Teilbaum vorliegt, können die Einträge in der Routing-Tabelle des Routers dahingehend verändert werden, daß der ursprüngliche Codepoint der ankommenden Pakete beibehalten wird. Der DSDM sendet dazu eine `SET_CODEPOINT`-Nachricht an den Router (③ in Abbildung 8), woraufhin er die Codepoints der eintreffenden Pakete nicht mehr verändert. In dem neuen Teilbaum werden somit die gleichen Garantien erbracht, wie im „alten“ Multicast-Baum.

5.2 Empfänger-initierte Reservierungen

Bei der Benutzung des DSDM können Empfänger-initiierte Reservierungsprotokolle eingesetzt werden, unter anderem auch RSVP.

Wird RSVP verwendet, ergibt sich der folgende Verlauf einer Reservierung:

1. Der Empfänger tritt der Multicast-Gruppe bei. Das Multicast-Routing-Protokoll seines First-Hop-Routers sendet Join-Nachrichten (bzw. entsprechende Nachrichten in Abhängigkeit vom eingesetzten Routing-Protokoll) bis der Anschluß an den Multicast-Baum erreicht ist.

2. Innerhalb von DS-Domänen werden bei neuen Teilbäumen einer Multicast-Gruppe die Codepoint-Felder in den Routing-Tabellen auf den voreingestellten Wert „(Lower-Than-)Best-Effort“ gesetzt. Somit erhält das neue Mitglied PATH-Nachrichten des Senders.

3. Wenn ein Empfänger reservieren möchte, sendet er stromaufwärts, d.h. in Richtung Sender, periodisch RESV-Nachrichten. Diese werden an den Grenzen der DS-Domänen von den Grenzroutern abgefangen und an den jeweiligen DSDM weitergeleitet.

4. Der DSDM überprüft die RESV-Nachrichten und reserviert gegebenenfalls die gewünschte Bandbreite. Ist dies der Fall, sendet er eine `SET_CODEPOINT`-Nachricht an den betroffenen Router, welcher den Codepoint-Eintrag in der Routing-Tabelle entsprechend ändert. Alle Pakete erfahren im folgenden den reservierten Dienst.

5. Der DSDM sendet die RESV-Nachricht stromaufwärts in die nächste Domäne. Innerhalb seiner eigenen Domäne wird sie transparent weitergereicht, ohne weiter bearbeitet zu werden.

5.3 Unterstützung dynamischer Senderwechsel

Wenn in einer Gruppe gleichzeitig mehrere Teilnehmer senden und dabei auch gleichzeitig Dienstgüte in Anspruch nehmen wollen, muß für jeden Sender – und damit für jeden durch ihn implizierten Multicast-Baum – eine eigene Reservierung vorgenommen werden. Dies kann problemlos über die vom DSDM bereitgestellten Management-Mechanismen realisiert werden.

Getrennte Reservierungen für jeden Sender sind auch dann unumgänglich, wenn eine Anwendung mit Halbduplex-Charakter vorliegt, d.h. wenn zu einem Zeitpunkt immer nur ein Sender Daten verschickt. Bei Halbduplex-Anwendungen könnte man nun gemeinsam benutzte Teilstrecken nur einfach reservieren, weil immer nur einer sendet, und somit z.B. nur die einfache Bandbreite benötigt wird. Es kann jedoch von der IP-Schicht nicht überprüft werden, ob zu jedem Zeitpunkt tatsächlich nur ein Sender aktiv ist. Würden mehrere Sender gleichzeitig Daten schicken, so würde auf gemeinsam benutzten Teilstrecken

zu viel Bandbreite in Anspruch genommen und es würde das NRS-Problem auftreten.

5.4 Unterstützung der Heterogenität

Innerhalb einer Gruppe sollte nur ein Dienst mit höherer Qualität benutzt werden. Wahlweise kann man sich als Empfänger noch ohne Dienstgüte mit Best-Effort an die Gruppe anschließen. Die Unterstützung von mehreren Dienstkategorien gleichzeitig innerhalb einer Gruppe ist zwar theoretisch möglich, aber es muß sichergestellt werden, daß nur von qualitativ höherwertigen Diensten zu niederwertigen konvertiert wird (hier ist ebenfalls das Problem der Unfairneß zu berücksichtigen). So würde es wenig Erfolg versprechen, wenn der Sender einen EF-Codepoint setzt, ein Teilbaum dann hauptsächlich Assured-Forwarding verwendet, bis auf einen Empfänger innerhalb des Teilbaums, der wiederum EF fordert. Weiterhin sind die Relationen zwischen den Weiterleitungsverhalten bzw. Diensten nicht eindeutig festzulegen und werden mit zunehmender Anzahl komplexer.

5.5 Weitere Vorteile

Schließlich seien noch einige weitere Vorteile des DSDM-Konzepts erwähnt:

- Sämtliche Reservierungen von Ende-zu-Ende können schneller bearbeitet werden, weil sie nur noch durch ein System pro Domäne behandelt werden, anstatt durch jeden beteiligten Router. Somit verringern sich die Bearbeitungskosten für jede Domäne auf ein Zwischensystem.

  Ein weiterer Vorteil ist, daß in den Routern auch keinerlei Reservierungsinformationen mehr gespeichert werden müssen. Die Authentisierung des Dienstnutzers ist auch nur noch einmal pro Domäne notwendig.

- Beim Ausfall eines DSDM bleiben bereits bestehende Reservierungen erhalten, weil die Verkehrsprofile in den Grenz-Routern bzw. First-Hop-Routern weiterhin aktiviert bleiben. Falls ein DSDM einmal ausfällt, bleiben die Profile weiterhin aktiv und die betroffenen Verkehrsströme erfahren weiterhin die reservierte Dienstgüte.

- Die gleichzeitige Verwendung mehrerer Signalisierungs- bzw. Reservierungsprotokolle ist möglich, weil die Ressourcenverwaltung des DSDM eine einheitliche Schnittstelle zur Verfügung stellt, über welche gezielt Ressourcen einer Teilstrecke reserviert werden können. Diese Schnittstelle verwaltet nur die Ressourcen der lokalen Domäne und ist unabhängig von der Art und Weise der Reservierung. Es können somit mehrere Reservierungsprotokolle parallel benutzt werden, ohne daß sich Inkonsistenzen ergeben.

- Durch die Unabhängigkeit vom Reservierungsprotokoll können sowohl sender- und empfängerbasierte als auch davon unabhängig statische und dynamische Reservierungen unterstützt werden. Mit diesen Eigenschaften bietet der DSDM alle Freiheitsgrade in bezug auf die möglichen Reservierungsprotokolle. So wird es möglich, für die Empfänger-orientierte Multicast-Kommunikation ein anderes Protokoll einzusetzen als für die Peer-to-Peer-Unicast-Kommunikation. Es kann für jede Kommunikationsform ein spezialisiertes Protokoll eingesetzt werden.

6 Zusammenfassung und Ausblick

Es wurden einige Probleme im Zusammenhang von Differentiated-Services mit Gruppenkommunikation identifiziert und beschrieben. Zur Lösung der Probleme lassen sich verschiedene Management-Mechanismen einsetzen, welche die Skalierbarkeit und Leistung der Differentiated-Services-Architektur weiterhin erhalten. Das beschriebene Konzept des Differentiated-Services-Domain-Managers stellt eine sehr flexible Lösung der Management-Aufgaben von DS-Domänen

dar. Vor allem ermöglicht es überhaupt erst den Einsatz von Differentiated-Services mit IP-Multicast, wodurch nun auch eine skalierbare und zuverlässige Gruppenkommunikation mit hohen Dienstgüteanforderungen möglich wird.

In weiteren Arbeiten am Institut für Telematik werden die Untersuchungen hinsichtlich der Leistungsfähigkeit von Differentiated-Services-Diensten weiter ausgebaut. Hierfür wurde ein leistungsfähiges Testnetz (das UNIQuE-Net [UNIQ99]) aufgebaut. Es besteht aus mehreren Linux-Software-Routern, welche mit 100 Mbps untereinander vernetzt sind.

Weiterhin wird das Konzept des DSDM detailliert entworfen und evaluiert. Hierfür sind sowohl Simulationen als auch reale Implementierungen des DSDM geplant.

Ein weiteres Vorhaben innerhalb des UNIQuE-Projekts sieht vor, neue Dienste für die Differentiated-Services-Architektur zu entwerfen und zu implementieren. Zudem werden geeignete Modifikationen am Transportprotokoll TCP durchgeführt, welche die Probleme bei der Ausnutzung der gesamten Bandbreite einer Reservierung [Wehr99] beheben sollen.

Literatur

[BBBN98] Fred Baker, David Black, Steven Blake und Kathleen Nichols. Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers. RFC 2474, Dezember 1998.

[BBCD+98] Steven Blake, David Black, Mark Carlson, Elwyn Davies, Zheng Wang und Walter Weiss. An Architecture for Differentiated Services. RFC 2475, Dezember 1998.

[BBHJ+97] Robert Braden, Steve Berson, Shai Herzog, Sugih Jamin und Lixia Zhang. Resource ReSerVation Protocol (RSVP) – Version 1. RFC 2205, September 1997.

[BlWe99] Roland Bless und Klaus Wehrle. Evaluation of Differentiated Services using an Implementation under Linux. In *Proceedings of the 7th IFIP Workshop on Quality of Service, London, Juni 1999*. IEEE, 1999.

[BrCS94] Robert Braden, David Clark und Scott Shenker. Integrated Services in the Internet Architecture: an Overview. RFC 1633, Juni 1994.

[Hofm98] Markus Hofmann. *Skalierbare Multicast-Kommunikation in Weitverkehrsnetzen*. Dissertation, Universität Karlsruhe (TH), Februar 1998.

[JaNP99] Van Jacobson, Kathleen Nichols und Kedarnath Poduri. An Expedited Forwarding PHB. Internet draft – draft-ietf-diffserv-phb-ef-02.txt, Februar 1999.

[JaNZ99] Van Jacobson, Kathleen Nichols und Lixia Zhang. A Two-bit Differentiated Services Architecture for the Internet. Internet draft – draft-nichols-diff-svc-arch-01.txt, April 1999.

[KIDS99] Karlsruhe Implementation of Differentiated Services (KIDS) homepage. http://www.telematik.informatik.uni-karlsruhe.de/forschung/diffserv/KIDS/, April 1999.

[UNIQ99] UNIQuE project homepage. http://www.telematik.informatik.uni-karlsruhe.de/forschung/UNIQuE/, April 1999.

[Wehr99] Klaus Wehrle. Implementierung und Evaluierung neuartiger Dienste für das Internet der nächsten Generation. Diplomarbeit, Universität Karlsruhe (TH), Institut für Telematik, Januar 1999.