

# **Ressourcen Management und Zugangskontrolle eines Vorabreservierungssystems**

Sabine Kühn

sabine.kuehn@dante.org.uk

Diese Dissertation entstand im Rahmen eines DFG-Projektes an der TU-Dresden. Ausgehend von der allgemeinen Zielsetzung des Projektes, die eine komplexe Betrachtung der Thematik "Vorabreservierungen in verteilten Systemen" darstellt, werden zunächst grundlegende gemeinschaftliche Schwerpunktthemen dargestellt und diskutiert. Darauf aufbauend befaßt sich diese Arbeit insbesondere mit den speziellen Teilgebieten der Zugangskontrolle und des Ressourcen-Management eines solchen Systems.

## **1 Einleitung**

Ohne Zweifel unterliegt das Internet einem enormen Entwicklungsprozeß, der sowohl durch eine rapide wachsende Anzahl aktiver Nutzer als auch durch eine zunehmende private bzw. kommerzielle Nutzung gekennzeichnet ist. Es ist abzusehen, daß mit der steigenden Anzahl der Internet-Teilnehmer der Datenverkehr und somit auch der Bandbreitenbedarf sich um ein Vielfaches erhöhen wird. So wird z. B. bei einem unveränderten Wachstum das zu übertragende Datenvolumen im Gigabit-Wissenschaftsnetz des Deutschen Forschungsnetzes (DFN) im Jahr 2004 voraussichtlich 6 PetaBytes (= 6.000 TeraBytes) pro Monat betragen.

Dabei wird deutlich, daß in zunehmenden Maße neben den "klassischen" Diensten, wie Filetransfer, Terminal Emulation, E-Mail oder World Wide Web neue Klassen von Anwendungen (Multimedia Anwendungen) und damit neue Dienste an Bedeutung gewinnen.

Zum einen liegt der Grund dafür in der drastischen Steigerung der Leistungsfähigkeit der Endsysteme. Somit können Anwendungen mit Benutzerschnittstellen verwirklicht werden, die den audiovisuellen Fähigkeiten von Menschen immer besser gerecht werden. Neue Möglichkeiten der Interaktion zwischen Computern und ihren Benutzern werden damit eröffnet.

Zum anderen wird dies begünstigt durch die Digitalisierung der Netze (z. B. Digital Video Broadcast - DVB, Universal Mobile Telecommunications System - UMTS, Integrated Services Digital Network - ISDN) und der damit einhergehenden Erweiterung der Funktionalität der Endgeräte wie Telefon, Fernsehgerät und Computer (Endgerätekonvergenz). Telefonie über Datennetze ("Voice over IP"), Teledienste über Mobilfunknetze oder SetTop-Boxen in Kabelnetzen sind nur einige Beispiele, die diese fortschreitende Konver-

genz verdeutlichen.

Durch den rasanten Fortschritt im Bereich der Telekommunikation und der Rechentechnik wird die Entwicklung vernetzter, multimedialer Anwendungen entscheidend vorangetrieben. Die Dienste, die dem Nutzer damit zur Verfügung stehen, umfassen dabei ein breites Spektrum. Insbesondere interaktive Dienste wie z. B. Videokonferenzen, Telematikanwendungen wie Fernunterricht und computergestützte Ausbildung, Telemedizin, Verkehrsleitsysteme und informationsgestützte Navigations- und Zielführungssysteme, CD-ROM-Datenbanken und On-line-Geschäfte mit multimedialen Informationen, Virtual Reality oder auch Verteildienste wie Audio- bzw. Video-on-Demand sind von starkem Interesse [HMB<sup>+</sup>00, ZS00, Ste99].

All diesen Anwendungen ist gemeinsam, daß sie gleichzeitig mehrere unterschiedliche Medienströme verarbeiten. Dabei wird unter einem Medienstrom ein unidirektionaler Strom digitaler Daten von einem Sender zu einem (oder mehreren) Empfänger(n) verstanden. Die digitalen Daten werden beim Sender erzeugt, mit Hilfe eines Kommunikationsdienstes zum Empfänger übertragen und dort weiterverarbeitet. Verteilte, multimediale Anwendungen stellen dabei höchste Anforderungen an das Zusammenwirken verschiedener Systeme und erfordern besondere Eigenschaften der zugrundeliegenden Kommunikationsnetze. Es muß garantiert werden können, daß ein Dienst den Echtzeitanforderungen der Datenströme bei der Übertragung gerecht wird. So sind zum Beispiel viele multimediale Anwendungen verzögerungsempfindlich und benötigen deshalb Zeitgarantien und Übertragungskapazitäten (Bandbreite), um eine gewisse Qualität bei der Wiedergabe der Audio- und Videodaten garantieren zu können.

Das heutige Internet ist für eine Dienstintegration, d. h. die gleichzeitige Übertragung von Audio, Video und klassischen Daten nur bedingt geeignet. Dies ist insbesondere darauf zurückzuführen, daß für die Übertragung die Datenströme nur ein Dienst (*best effort* Dienst) bereitgestellt wird. Bei diesem Dienst werden keinerlei Aussagen über die Zustellzeit der Pakete gemacht, was in der Praxis die unangenehme Auswirkung hat, daß einzelne Pakete teilweise erheblich verzögert werden können, die Datenpakete eines Datenstromes in ihrer logischen Reihenfolge vertauscht oder bei einer Überlastung des Netzwerkes verloren gehen können.

Um den Anforderungen bei der Übertragung von Anwendungsströmen mit Echtzeitanforderungen (z. B. Audio- und Videodaten) gerecht zu werden, sind sowohl zum einen höhere Übertragungsraten und zum anderen zusätzliche Mechanismen für eine differenzierte Behandlung von Datenströmen notwendig.

In diesem Zusammenhang ist seit einiger Zeit ein struktureller und technologischer Wandel des Internets zu verzeichnen. So wurden bereits in den 90er Jahren in den USA die ersten Gigabit-Netzwerke aufgebaut. Eine Reihe amerikanischer Universitäten gründete im Jahre 1997 die Non-Profit-OrganisationUCAID, mit dem Auftrag, ein neues Hochleistungsnetzwerk für Forschung und Wissenschaft auf Glasfaserbasis aufzubauen (Abilene Network). Auch in Deutschland wird das Wissenschaftsnetz des DFN (G-WIN) seit 1999 zu einem Gigabit-Netzwerk aufgerüstet [DV00].

Durch den Zusammenschluß dieser Netze soll ein *Next Generation Internet* (Internet 2) geschaffen werden, welches parallel zum herkömmlichen Internet läuft und über das Infor-

mationen mit hoher Geschwindigkeit und hoher Qualität durch breitbandige Glasfasernetze geleitet werden. Zunächst ist das Netz jedoch für einen kleinen Kreis von Wissenschafts- und Forschungseinrichtungen reserviert. Telemedizin, Videokonferenzen und virtueller Zugang zu Projekten werden die Hauptnutzungen sein.

Obwohl die in den Hochleistungsnetzen zum Einsatz kommenden Technologien wie zum Beispiel der Asynchrone Transfermodus (ATM), *Packet-over-Sonet* (POS) oder Gigabit-Ethernet hohe Bandbreiten zur Verfügung stellen, ist damit allein die Gewährleistung von Dienstgüte (*Quality of Service*, QoS) nicht gesichert. Um Netzressourcen flexibel und wirtschaftlich zu nutzen und gleichzeitig eine bestimmte Netz- und Dienstgüte garantieren zu können, ist ein sehr leistungsfähiges Verkehrsmanagement nötig. Diese Funktionalität wird jedoch z. Z. nur von ATM im vollen Umfang erbracht [Sie97].

Aufgrund der Heterogenität der Netze und des Umstandes, daß IP wegen seiner weiten Verbreitung, Einfachheit und Zuverlässigkeit weiterhin als Netzwerkprotokoll zum tragen kommt, wurden Architekturen entwickelt und untersucht, die eine differenzierte Behandlung von Datenströmen in heterogenen IP-Netzen ermöglichen. So werden bereits seit längerer Zeit durch die Internet Engineering Task Force (IETF) Mechanismen erarbeitet, mit denen Dienstgüte in IP-Netzen sichergestellt werden kann.

Das *Integrated Services* Modell (IntServ) erlaubt es, Dienstgüte für Anwendungen mit besonderen Anforderungen an Bandbreite, Verzögerungszeiten (*Delay*) und *Jitter* bereitzustellen [BCS94, BZB<sup>+</sup>97]. Im Mittelpunkt steht dabei das Signalisierungsprotokoll RSVP (Resource Reservation Protocol). Es ermöglicht eine empfangerbasierte Reservierung, indem in bestimmten Zeitabständen RSVP-Nachrichten ausgesandt werden. Jeder Netzwerkknoten muß Informationen zu den IP-Datenströmen, die ihn durchlaufen, verarbeiten und speichern. Dieses explizite Reservieren von Ressourcen wird bereits erfolgreich in Firmen- und Zugangsnetzen eingesetzt, um QoS bereitzustellen. Im Backbone-Bereich ergeben sich allerdings Probleme mit der Skalierbarkeit, weil dort Netzknotten unter Umständen von sehr vielen Datenströmen durchlaufen werden [MBB<sup>+</sup>97]. Daher befassen sich die aktuellen Arbeiten innerhalb der Intserv Working Group mit Techniken, die RSVP besser skalierbar machen sollen [BIFD01]. Außerdem wurde eine Methode vorgestellt, die das Datenvolumen verringert. Das Verfahren nutzt dazu die Ähnlichkeit der periodisch verschickten RSVP-Nachrichten. Ein weiterer Ansatz reduziert den Overhead bei der Verarbeitung von RSVP-Refresh-Nachrichten [BGS<sup>+</sup>01].

Die Arbeitsgruppe *Differentiated Services* (DiffServ) beschäftigt sich mit einer Architektur [BBC<sup>+</sup>98], die Dienstgüte unterstützt, aber die bei RSVP möglichen Skalierungsprobleme vermeidet. Im Unterschied zu IntServ werden nicht die einzelnen IP-Datenströme berücksichtigt, sondern Verkehrsflüsse mit ähnlichen QoS-Anforderungen zu einem "Behavior Aggregate" (BA) zusammengefaßt. Das "Diffserv-Code-Point"-Oktett im IP-Header [NBBB98] zeigt an, wie mit einem bestimmten IP-Paket verfahren werden soll. Eine wichtige Frage beim Diffserv-Modell ist, wie dynamische QoS-Anforderungen unterstützt werden können. Ein Lösungsansatz sind "Bandbreiten-Broker" (Bandwidth Broker). Sie sammeln die Anforderungen und vergeben dann die entsprechenden Ressourcen je nach Verfügbarkeit und Berechtigung.

Die oben genannten diensteintegrierende Netzwerke (ATM, IntServ, DiffServ) ermögli-

chen es, für die Übertragung ad hoc Ressourcen zu reservieren und Datenströmen zuzuordnen. Aus diesem Grund werden diese Systeme im weiteren auch als Reservierungssysteme bezeichnet. Da in diesen Reservierungssystemen die Reservierung mit dem Verbindungsaufbau (ATM) bzw. unmittelbar vor dem Senden der Daten (IntServ) erfolgt, wird diese Form der Reservierung im weiteren als **Sofortreservierung** bzw. als **immediate reservation** (ImRe) bezeichnet.

### 1.1 Problemstellung

Trotz technologischer Entwicklungen wie Gigabitnetzwerke (z. B. Gigabit-Ethernet) und Switching Technologien (z. B. ATM), werden aufgrund des drastischen Nutzerzuwachses und der zunehmenden kommerziellen Nutzung insbesondere Ressourcen wie Verbindungsbandbreite und Verarbeitungskapazität der Router aus Kostengründen auch in Zukunft begrenzt sein. Aus diesem Grund stellen Reservierungsmechanismen eine notwendige Grundlage zur Einhaltung einer garantierten Dienstgüte dar.

Neben dem Reservierungsprotokoll zur Aushandlung der Dienstgüteparameter regelt die Zugangskontrolle, als ein weiterer Bestandteil des Reservierungssystems, die Zuteilung der Ressourcen für die jeweiligen Verbindungen. Es ist bekannt, daß ohne weitere Kontroll- bzw. Steuerungsmechanismen breitbandige Verbindungen eine wesentlich höhere Blockierungswahrscheinlichkeit als Verbindungen mit geringen Bandbreitenanforderungen erfahren können [RVM96]. Neben Mechanismen, die eine explizite Zuteilung und Nutzung eines Teils der Ressource einzelnen Klassen von Verbindungen erlauben, stellt die Priorisierung eine weitere Möglichkeit dar, auf das Blockierungsverhalten dedizierten Anwendungsströmen Einfluß zu nehmen. Die Reservierung von Ressourcen im Voraus, die im folgenden auch als **Vorabreservierung** bzw. als **resource reservation in advance** (ReRA) bezeichnet wird, kann als eine Form der Priorisierung angesehen werden. So erfolgt entsprechend der zeitlichen Reihenfolge des Eintreffens der Reservierungsanforderungen eine Priorisierung der Anforderungen.

Vorabreservierungen haben einige Vorteile gegenüber anderen Prioritäts-Mechanismen. Sie sind aus Sicht der Nutzer einfach zu verstehen und anzuwenden, da diese aus dem täglichen Umgang mit Reservierungen gewisse Erfahrungen besitzen (Hotel- und Flugbuchung, Platzkartenreservierung). Zudem wird die Steuerung und Verantwortlichkeit vom Netzwerk in die Hände der Nutzer gegeben. Anstatt durch eine bestimmte Zugangskontrolle eingeschränkt zu werden, kann der Nutzer für wichtige Reservierungen weiter im Voraus buchen. Andererseits ermöglicht das Wissen über zukünftige Reservierungen eine bessere Planung bei der Vergabe der Netzwerkressourcen.

Anwendungen für Vorabreservierungen sind in der Regel dadurch gekennzeichnet, daß schon eine gewisse Zeit vor Inanspruchnahme der Reservierung Aussagen über die zu erwartende Datenstromcharakteristik und die gewünschte Dienstgüte getroffen werden können. Ein Einsatz von Vorabreservierungen bietet sich vor allem für Anwendungen und Ereignisse an, bei denen aufgrund hoher Anforderungen an die Dienstgüte oder einer starken Belastung des Netzwerkes eine Absage der Ressourcenreservierung zum Übertragungs-

zeitpunkt wahrscheinlich ist bzw. für die eine garantierte Zusage der Ressourcen gefordert wird.

Nachfolgend sollen drei mögliche Einsatzgebiete dies noch einmal verdeutlichen:

- Videokonferenzsysteme, Distance Learning, Telecommuting, Virtual Rooms

Nutzer können von ihrer gewohnten Arbeitsumgebung aus Konferenzen organisieren und daran teilnehmen, ohne den sonst üblichen Zeit- und Kostenaufwand für die Reise zum Tagungsort. Wenn mehrere Teilnehmer an einer Konferenz partizipieren wollen, ist es notwendig, einen gemeinsamen Termin zu finden. Da gerade die Audio- und Videoströme einer Videokonferenz hohe Anforderungen an die Übertragungsqualität stellen, müssen Ressourcen in größerem Umfang reserviert werden. Stehen diese zum Zeitpunkt der Konferenz mittels Sofortreservation nicht im ausreichenden Maß zur Verfügung, müssen die Teilnehmer entweder auf bessere Netzwerkbedingungen warten oder sogar die Konferenz verlegen. Das bedeutet, daß zwischen den Teilnehmern ein neuer Termin ausgehandelt werden muß, jedoch die Verfügbarkeit der Ressourcen zu diesem neuen Zeitpunkt auch nicht gesichert ist.

- Video on Demand (VoD)

Ein Fernsehnutzer hat heutzutage die Wahl zwischen verschiedenen Fernsehkanälen, kann jedoch nicht den Inhalt dieser Sender bestimmen. In der Zukunft wird es möglich sein, ausgewählte Spielfilme, Nachrichten, Spiele und anderes von einem Server für jeden beliebigen Zeitpunkt zu beziehen. Wird darüber hinaus von Nutzer statt eines konkreten Zeitpunktes ein Zeitintervall angegeben, so ist es mit flexiblen VoD-Systemen möglich, Zugriffe auf gleiche Medien zu bündeln und das Senden der Videos somit zeitlich zu koordinieren. Das System ist somit in der Lage, die Ressourcen bestmöglich auszunutzen.

- Verteilte Interaktive Systeme

Vorstellbar sind große Online - Enzyklopädien, Lehr- oder Expertensysteme bei denen der Nutzer interaktiv Abfragen durchführt, Vorträge ansieht oder eine Präsentation verfolgt. Gibt es bestimmte Nutzungsmuster, z. B. eine Folge von Videoclips, die von jedem Nutzer abgerufen werden, so kann das System dafür schon Ressourcen im voraus reservieren.

## 1.2 Zielsetzung

Vorabreservierungen werden bereits bei der Konzeption des Internet 2 mit einbezogen. So wird z. B. bei der Bereitstellung der Dienste im QBone [THD<sup>+</sup>99] bzw. im G-WIN [Pat00] die Möglichkeit für das Initiieren von zukünftigen Reservierungsanforderungen von Anwendungen berücksichtigt. Dieser Dienst wird jedoch zunächst durch die manuelle Konfiguration von separaten Verbindungen (*leased lines*) im Netz realisiert. Dies kann jedoch bei einem hohen Aufkommen an (zukünftigen) Reservierungsanforderungen zum Beispiel dazu führen, daß der zeitliche Aufwand zur Konfiguration der einzelnen Knoten

sehr hoch wird und somit die Behandlung einer kurzfristigen Reservierung nicht mehr möglich ist.

Das Ziel dieser Arbeit besteht darin, wesentliche Grundlagen für die Realisierung eines Vorabreservierungssystem zu schaffen, das durch die gemeinsame Nutzung von Sofort- und Vorabreservierungen charakterisiert ist und im folgenden daher auch als **ImRe/ReRA-System** bzw. **IR-System** bezeichnet wird. Dabei wird insbesondere Wert darauf gelegt, Vorabreservierungen in einem dynamischen Umfeld<sup>1</sup> zu realisieren, um den manuellen Konfigurationsaufwand zu umgehen und diesen Dienst einer breiten Masse von Anwendern zugänglich zu machen. Dabei ist die Anwendung von Vorabreservierungen zur Reservierung von Ressourcen für dedizierte Anwendungen beispielsweise in LAN-Umgebungen als auch für aggregierte Verbindungen in Form von *Leased Lines* bzw. *Virtual Private Networks* im WAN-Bereich denkbar.

Die Realisierung von Vorabreservierungen auf der Basis dynamischer Verbindungen erfordert grundlegende Untersuchungen der in einen solchen Vorgang involvierten Komponenten eines Reservierungssystems. So werden diesbezüglich Erweiterungen in dem Aushandlungsprotokoll, der Datenhaltung, der Zugangskontrolle und dem Routingprotokoll notwendig. Um gezielt Anforderungen an die Funktionalität dieser Komponenten ableiten zu können, werden im Rahmen dieser Arbeit zunächst umfangreiche theoretische Betrachtungen zu einem Vorabreservierungssystem durchgeführt. Diese betreffen beispielsweise:

1. den Entwurf eines Dienstmodells unter Berücksichtigung der neuen Zeitparameter, als Grundlage für die Erweiterung bereits existierender und standardisierter Signalisierungsprotokolle,
2. die Untersuchung geeigneter Verwaltungsstrukturen als wesentliche Basis für ein Vorabreservierungssystem sowie
3. die numerische und exakte Analyse des Systemverhaltens als Ausgangspunkt für weiterführende Betrachtungen beispielsweise für die Entwicklung effektiver und flexibler Zugangskontroll- und Routingalgorithmen.

Im Gegensatz zu anderen Arbeiten, in denen allgemeine Ansätze bzw. Vorschläge über die Erweiterung von Signalisierungsprotokollen in bezug auf Vorabreservierungen dargestellt werden [BLB98, FGV97, WDS<sup>+</sup>95, FKL<sup>+</sup>99, Rei95, Sch98], besteht der Anspruch dieser Arbeit darin, eine detaillierte Beschreibung der notwendigen Protokollerweiterungen zu geben. Die Konzeption und prototypische Realisierung der Aushandlung von Vorabreservierungen mit Hilfe von Reservierungsprotokollen orientiert sich dabei an konkreten Implementationen verteilter Reservierungssysteme. Die Validierung des Konzeptes erfolgte in dieser Arbeit zunächst im Signalisierungsprotokoll von ATM (UNI - User Network Interface) und einhergehend mit der Standardisierung der IntServ-Architektur in RSVP [Bre01]. Beide Signalisierungsprotokolle unterscheiden sich maßgeblich in der Art der Verwaltung der verbindungspezifischen Parameter: während in der ATM-Signalisierung

---

<sup>1</sup>Unter einem dynamischen Umfeld wird hier verstanden, daß mit Hilfe entsprechender Protokolle (z. B. Signalisierungsprotokolle) der Nutzer zu jedem Zeitpunkt Verbindungen initiieren und abbauen kann, ohne beispielsweise auf die Hilfe eines *Call Centers* angewiesen zu sein.

sogenannte *Hard States* verwaltet werden, müssen in RSVP die *Soft States* zyklisch aktualisiert werden. Aus der Sicht der Fehlerbehandlung war daher die zusätzliche Implementierung in RSVP von Interesse. Dabei stellt aus heutiger Sicht ATM eine QoS-fähige Breitbandtechnologie dar, die sich insbesondere im WAN-Bereich etabliert hat, während RSVP in Verbindung mit der IntServ-Architektur eine QoS-fähige IP-Erweiterung repräsentiert, die hauptsächlich im LAN-Bereich zum Einsatz kommen wird.

In Hinblick auf die Skalierbarkeit eines IR-Systems ist die Wahl der Verwaltung ein entscheidender Aspekt. Die Verwaltung sollte einfach zu implementieren und leistungsfähig sein und möglichst zu einer geringen Belegung von Speicherplatz führen. In diesem Zusammenhang wird eine einfache Möglichkeit der Verwaltung von Vorabreservierungen aufgezeigt, die in Form einer Intervalltabelle organisiert ist. Es werden weiterhin Größen aufgezeigt, die auf den Verwaltungsaufwand maßgeblichen Einfluß haben und geeignete Datenstrukturen untersucht und verglichen, ob sie den oben genannten Anforderungen entsprechen. Im Gegensatz zu [SNNP99], werden hier einfache dynamische Datenstrukturen mit effizienten Zugriffsoperationen (Suchen, Einfügen, Löschen) favorisiert, um den mit einer statischen Struktur verbundenen Aufwand zu vermeiden.

Ziel der numerischen Analyse ist es, grundsätzliche Aussagen über das Leistungsverhalten eines Vorabreservierungssystems treffen zu können. Bei der Entwicklung eines stochastischen Modells ist es von besonderem Interesse, einen Einblick in die Zusammenhänge eines von Sofort- und Vorabreservierungen gemeinsam genutzten Systems zu erhalten. Auf der Basis dieses Modells soll das neue Blockierungsverhalten im Detail untersucht werden. Dabei soll ebenfalls dem Unterbrechungsverhalten von Reservierungen Rechnung getragen werden, wenn die Dauer einer Reservierung zum Zeitpunkt der Reservierungsanforderung nicht bekannt ist. Unter der Voraussetzung eines vorgegebenen, mit der Zeit gleitenden, Reservierungszeitraums, wird mit Hilfe der stationären Analyse eines solchen Modells die Ermittlung umfangreicher Leistungsparameter angestrebt. Die hiermit festgelegten Rahmenbedingungen unterscheiden sich grundlegend von Modellen, die bereits in den 80er Jahren im Zusammenhang mit dem Einsatz von ISDN bzw. der Nutzung von Satellitenkanälen entwickelt wurden [Lus77, Lus80, RLLS88, Vir92, VA91]. In [RLLS88] wurde beispielsweise gezeigt, daß, ausgehend von einem allgemeinen Warteschlangenmodell für Sofortreservierungen eine Berechnung der Blockierungswahrscheinlichkeit für Vorabreservierungen bei der Betrachtung eines unbegrenzten Reservierungsbuches nicht möglich ist. In [Vir92] beschreibt Virtamo durch die Zusammenfassung von freien Zeitabschnitten zu sogenannten "Inseln" ein Modell zur Berechnung der Leistungsparameter von Vorabreservierungen in einem 1-Kanal-System. Die Anwendung dieses Modells in einem Mehrkanal-System ist lediglich für eine stark eingeschränkte Anzahl von Zeitabschnitten bzw. Kanälen geeignet.

Die Analysen und Bewertungen der Strategien und Mechanismen bzgl. der Zugangskontrolle und des QoS-Routings sind eher von allgemeiner Natur und somit in jeder QoS-Architektur einsetzbar. Im Mittelpunkt dieser Arbeit steht dabei eine detaillierte Untersuchung der Zugangskontrolle mit unterschiedlichen Gesichtspunkten, die sich aus der Koexistenz von Sofort- und Vorabreservierungen ergeben. Die entsprechenden Ausführungen zum QoS-Routing werden in [Bre01] vorgenommen.

### 1.3 Gliederung der Arbeit

Zunächst erfolgt eine Einführung in die wesentlichsten Grundlagen zum Dienstgüte-Management (QoS-Management) in verteilten Systemen. Dabei wird insbesondere auf Schwerpunkte hinsichtlich der Bereitstellung von Dienstgüte und auf konkrete QoS-Architekturen des Internets eingegangen.

Im Anschluß daran erfolgen Ausführliche Betrachtungen, die hinsichtlich der Erweiterung bestehender Reservierungssysteme für Vorabreservierungen durchgeführt wurden. In diesem Zusammenhang wird ein allgemeines Kommunikationsmodell vorgestellt und es erfolgt eine Diskussion hinsichtlich der Zeitparameter von Vorabreservierungen bzw. verwaltungsspezifischer Aspekte sowie die Beschreibung eines analytischen Modells zur Bestimmung der Leistungsparameter eines Vorabreservierungssystems. Des weiteren werden basierend auf dem analytischen Modell bzw. Simulationen umfangreiche Untersuchungen zum Systemverhalten in einem reinen Vorabreservierungssystem bzw. einem IR-System durchgeführt.

Als weiterer wesentlicher Bestandteil dieser Arbeit erfolgt eine detaillierte Untersuchung der Zugangskontrolle mit unterschiedlichen Gesichtspunkten, die sich aus der Koexistenz von Sofort- und Vorabreservierungen ergeben. Dabei werden insbesondere die Auswirkungen der Zeitparameter, die sowohl als konstante als auch variable Größen betrachtet werden, untersucht. Des weiteren werden unterschiedliche Möglichkeiten der Ressourcenaufteilung unter dem Aspekt der Effizienz und Fairneß diskutiert und bewertet. Ein weiterer wesentlicher Kernpunkt besteht darin, Strategien für die Zugangskontrolle zu entwickeln und zu bewerten, die unter Einbezug spezifischer Charakteristika von Vorabreservierungen zu Systemverbesserungen führen (Stornierungen bzw. Nichtinanspruchnahme von Vorabreservierungen, on/off-line Scheduling-Strategien). Die in diesem Zusammenhang entwickelten Ansätze werden entweder mit Hilfe von stochastischen Modellen oder mit Hilfe von Simulationen validiert.

Anschließend erfolgt die Beschreibung der Aushandlung von Vorabreservierungen im Kontext einer konkreten Implementierung. Der Entwurf und die Realisierung eines Konzeptes zur funktionalen Erweiterung eines Signalisierungsprotokolls wurde am Beispiel der ATM UNI-Signalisierung vorgenommen.

Zum Abschluß dieser Arbeit erfolgt eine Zusammenfassung der Ergebnisse, die in dieser Arbeit gewonnen wurden, und es werden Vorschläge für weiterführende Untersuchungen vorgestellt.

## 2 Zusammenfassung und Ausblick

Diese Dissertation befaßte sich mit der Problematik der mittel- und langfristigen Reservierung von Netzwerkressourcen und insbesondere mit Algorithmen der Zugangskontrolle zur Verbesserung des Leistungsverhaltens unter Berücksichtigung der neuen Zeitparameter und Charakteristika in einem erweiterten Reservierungssystem.



Zunächst wurden in den einführenden Kapiteln aktuelle Entwicklungstendenzen im Internet und die damit verbundenen Veränderungen hinsichtlich der Netzinfrastruktur, Dienstbereitstellung und Anforderungen der Anwendungen bzw. Nutzer dargelegt und diskutiert. Dabei wurde deutlich, daß unter bestimmten Bedingungen<sup>2</sup> eine Erweiterung von (Sofort-) Reservierungssystemen notwendig und sinnvoll ist, um auch Anwendungen, für deren Datenübertragung mit Dienstgüteanforderungen bereits langfristig im voraus eine Planung erfolgt, zu unterstützen.

Ausgehend von einer kurzen Beschreibung des QoS-Managements in verteilten Systemen und der Darstellung von QoS-Architekturen (ATM, IntServ, DiffServ) wurden diejenigen Komponenten eines Reservierungssystems bestimmt, die hinsichtlich der Erweiterung für die Unterstützung von Vorabreservierungen notwendig sind. Darauf aufbauend wurde das Modell eines erweiterten Reservierungssystems eingeführt, welches die Basis für alle durchgeführten Untersuchungen bildet. Dabei wurde ein Schwerpunkt auf die Beschreibung der Aushandlung und Verwaltung von Vorabreservierungen gelegt. Darüber hinaus wurde ein Modell zur analytischen Beschreibung eines erweiterten kontinuierlichen bzw. diskreten Reservierungssystems entwickelt. Das Ziel bestand darin, durch das vereinfachte analytische Modell exakte Aussagen über das Verhalten der Leistungsgrößen zu gewinnen.

Für die sich daran anschließenden Untersuchungen und Erweiterungen der Zugangskontrolle eines erweiterten Reservierungssystems war es jedoch nicht immer möglich, die komplexen Sachverhalte durch analytische Methoden zu beschreiben. Aus diesem Grund wurde ein Simulationsmodell entwickelt, mit dem eine Bewertung der neuen Strategien und Mechanismen auch bezogen auf ein Netzwerk möglich ist. Die durchgeführten Untersuchungen und Erweiterungen der Zugangskontrolle betrafen dabei insbesondere die neuen Systembedingungen, die bei der Einführung von Vorabreservierungen auftreten:

1. Untersuchungen von unterschiedlichen Strategien, die die Zuteilung der Ressource bezogen auf die Reservierungsarten ImRe und ReRA regeln,
2. eine Diskussion über die Auswirkungen der Verbindungsdauer auf die Zugangskontrolle unter Berücksichtigung der Zuteilungsstrategien,
3. Algorithmen zur Behandlung der Stornierungen von Vorabreservierungen,
4. Algorithmen zur Suche eines optimalen Reservierungsstartzeitpunktes von Vorabreservierungen innerhalb eines vorgegebenen Gültigkeitsbereiches und
5. ein Ansatz zur optimalen Auswahl von Vorabreservierungen unter Einbeziehung verzögerter Quittierungen.

Bei der Untersuchung von Möglichkeiten der Ressourcenzuteilung zu den Reservierungsarten wurden die Strategien Sharing, Partitioning und Trunking gegenübergestellt. Das Ziel bestand darin eine Bewertung der Strategien hinsichtlich ihrer Effizienz und Fairneß auf der Basis einer Kostenfunktion vorzunehmen. Da bei den Strategien der Partitionierung und des Trunkings eine Aufteilung der Ressource auf die Reservierungsarten (ImRe und

---

<sup>2</sup>Notwendigkeit der Reservierung von Netzwerkressourcen zur Einhaltung von Dienstgüteanforderungen der Anwendungen und das begrenzte Vorhandensein von Netzwerkressourcen

ReRA) zugrundeliegt, war es im Hinblick auf die Anforderungen an die Zugangskontrolle notwendig, einen Algorithmus zu entwickeln, der eine optimale Bestimmung der Partitionsgröße bzw. des Trunk-Parameters erlaubt. Die Berechnung der Blockierungswahrscheinlichkeiten für Sofort- und Vorabreservierungen geht dabei als wesentliche Größe ein. In diesem Zusammenhang wurde das mathematische Grundmodell zur Berechnung der Blockierungswahrscheinlichkeiten für die Strategie Trunking erweitert und detailliert beschrieben. Auf der Basis eines Vergleichs der Gewinnverhältnisse der einzelnen Strategien konnte festgestellt werden, daß Trunking sowohl bei bekannter als auch unbekannter Dauer von Sofortreservierungen das beste Systemverhalten aufweist, wenn der Schwerpunkt auf der Fairneß, d.h. möglichst gleicher Blockierungswahrscheinlichkeiten bezogen auf Sofort- und Vorabreservierungen liegt.

Im Bezug auf die Auswirkungen einer Verbindungsdauer bei Sofortreservierungen konnte festgestellt werden, daß bei einer unbekanntem Dauer der Reservierungsanforderung die Verwendung von Algorithmen, die es ermöglichen die Unterbrechungswahrscheinlichkeit von Sofortreservierungen zu reduzieren, in der Zugangskontrolle zu einem besseren Systemverhalten führen. In diesem Zusammenhang wurden unterschiedliche Ansätze, die in der Literatur beschrieben sind, vorgestellt und diskutiert.

Ein wesentlicher Schwerpunkt dieser Arbeit war, Leistungseinbußen infolge der Einführung von Vorabreservierungen zu reduzieren. Dies betraf Einbußen zum einen hervorgerufen durch Stornierungen und zum anderen infolge der zeitlichen Fragmentierung der Ressourcen. Das Ziel bestand darin, geeignete Algorithmen zu entwerfen und zu untersuchen, wobei die Zeitparameter als variable Größen bzw. die Charakteristiken von Vorabreservierungen bei der Zugangskontrolle berücksichtigt werden.

Ausgehend von der Annahme, daß Nutzer bzw. Anwendungen während der Intermediale Phase eine Stornierung von Vorabreservierungen vornehmen (Cancellation) bzw. zum geplanten Startzeitpunkt der Reservierung diese nicht in Anspruch nehmen (No-Show), wurden zunächst Algorithmen entwickelt, die auf der Basis von Schätzmethoden eine kontrollierte Überbuchung der Ressourcen durch Sofortreservierungen ermöglichen. Durch die Auswertung von Simulationsergebnissen konnte daraufhin gezeigt werden, daß unter bestimmten Randbedingungen (tatsächliche Überbuchungswahrscheinlichkeit, Stornierungswahrscheinlichkeit, und Systemparameter) eine wesentliche Verbesserung des Systemverhaltens (Blockierungswahrscheinlichkeit, Auslastung) erreicht werden kann. Darüber hinaus wurde anhand eines Beispiels ebenfalls eine Überbuchung von Vorabreservierungen diskutiert.

Mit dem Ziel die zeitliche Fragmentierung der Ressourcen zu reduzieren, wurden Algorithmen vorgestellt, die es ermöglichen unter verschiedenen Gesichtspunkten einen optimalen Reservierungsstartzeitpunkt von Vorabreservierungen innerhalb eines vorgegebenen Gültigkeitsbereiches zu suchen. Auf der Basis eines stochastischen Modells wurde der Algorithmus *earliest* beschrieben, der frühest mögliche Startzeitpunkt bestimmt, zu dem die Ressourcen verfügbar sind. Der Algorithmus *widest* hingegen ermittelt denjenigen Startzeitpunkt innerhalb des Gültigkeitsbereiches, der eine effektive Auslastung der verfügbaren Bandbreite ermöglicht. Es konnte am Beispiel eines Single Links unter Verwendung von Simulationsergebnissen gezeigt werden, daß unter Berücksichtigung der Ressourcenauslastung innerhalb des Gültigkeitsbereiches ein günstigeres Systemver-

ten erzielt werden kann, als mit der Einordnung der Anforderung zum frühestmöglichen Startzeitpunkt. Des weiteren konnte gezeigt werden, daß die Einführung von Gültigkeitsbereichen in einem verteilten System ebenfalls möglich ist, wenn eine entsprechende Anpassung der Routing-Algorithmen, d. h. insbesondere der *Generic Admission Control* eines dienstgütebasierten Routings vorgenommen wird. In diesem Zusammenhang wurden zwei Routing-Algorithmen vorgestellt und diskutiert. Dabei zeigt die Anwendung der Heuristik *widest shorstest* die deutlichste Reduzierung der Blockierungswahrscheinlichkeit auf, die sich ebenfalls in einer erhöhten Auslastung des Netzes widerspiegelt. Dabei wurde in Zusammenhang mit einem dienstgütebasierten Routing davon ausgegangen, daß vollständige Informationen über die Link-Zustände auf jedem Knoten der Routing-Domäne bekannt sind. Dies ist mit einem erheblichen Aufwand an Verwaltung und Aushandlung der geänderten Link-Zustände verbunden, wie in der Arbeit von [Bre01] aufgezeigt wurde. In Analogie zu dem in [Bre01] vorgestellten Ansatz wäre eine deutliche Reduzierung dieses Aufwandes denkbar, wenn die Bestimmung der optimalen Startzeiten basierend auf Belegungswahrscheinlichkeiten ermittelt werden würden. In diesem Zusammenhang sind weitere Untersuchungen notwendig, die Aussagen über die Leistungsfähigkeit derart modifizierter Algorithmen erlauben.

Die in der Arbeit dargelegten Untersuchungen zur Möglichkeit verzögerter Entscheidungen durch die Zugangskontrolle auf der Basis gesammelter sich zeitlich überlagernder Vorabreservierungen haben einen ersten Einblick in die Problematik gegeben. Es wurde ein Lösungsansatz diskutiert, der jedoch starken Einschränkungen unterworfen ist. Ausgehend von einem einfachen Beispiel wurde jedoch deutlich, daß durch verzögerte Quittierungen die Blockierungswahrscheinlichkeit der Reservierungen reduziert werden kann. Darüber hinaus können bei sehr hoher Belastung deutliche Verbesserungen erzielt werden, da hier die Auswahl der Anforderungen zugunsten einer geringen Blockierung eher gegeben ist. Dabei führt die Anwendung einer solchen Strategie dazu, daß es durch die verzögerte Entscheidung zu einer ungleichen Behandlung von bereits lange im voraus getätigten Reservierungen und relativ kurzfristigen Reservierungen kommen kann. Das bedeutet, daß die Priorisierung der Reservierungen, die sich bisher aus der zeitlichen Reihenfolge ihres Eintreffens ergab, in diesem Ansatz keine Berücksichtigung findet.

Als weiterer Schwerpunkt der Arbeit wurde ausgehend von dem allgemeinen Modell eines erweiterten Reservierungssystems eine praktische Validierung innerhalb von ATM vorgenommen. Als Grundlage diente dafür die frei verfügbare HARP-Implementierung. Die notwendigen Erweiterungen zur Behandlung von Vorabreservierungen bezogen sich auf die Spezifikation eines neuen Informationselementes und einer neuen Reservierungsnachricht zur Übermittlung der Zeitparameter und zur Steuerung spezifischer Aktivitäten, die Abbildung der ReRA-Dienstelemente und des Kommunikationsmodells auf ATM-Nachrichten und Nachrichtenbehandlungsregeln (State Machine), die Erweiterung der Verwaltungsstrukturen sowie auf die Anpassung der Schnittstellen für die Integration des neuen Signalisierungsmoduls in das Betriebssystem Digital UNIX.

In diesem Zusammenhang wurden Funktionalitätstests (Endsystem zu Endsystem) durchgeführt, die ebenfalls die Funktionalität des ReRA-aging Mechanismus sicherstellten, für den Fall, das reservierte Ressourcen nicht genutzt bzw. aktiviert wurden. Da eine Erweiterung der Switch-Software nicht realisiert werden konnte, sind Aussagen über die Ska-

lierbarkeit in erweiterten Topologien nicht möglich. Für eine Menge von etwa zehn vorab reservierten Verbindungen konnten keine Unterschiede zum Leistungsverhalten eines herkömmlichen Systems ausgemacht werden. Im Hinblick auf den praktischen Einsatz einer solchen Signalisierungssoftware beispielsweise zum dynamischen Auf- und Abbau vorab reservierter Leased Lines in Form von SVPs sind weiterführende Leistungsmessungen notwendig.

## Literaturverzeichnis

- [BBC<sup>+</sup>98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. *An Architecture for Differentiated Services*. IETF, 1998. RFC2475.
- [BCS94] R. Braden, D. Clark, and S. Shenker. *Integrated Services in the Internet Architecture: an Overview*. IETF, 1994. RFC1633.
- [BGS<sup>+</sup>01] L. Berger, D. Gan, G. Swallow, P. Pan, F. Tommasi, and S. Molendini. *RSVP Refresh Overhead Reduction Extensions*. IETF, 2001. RFC2961.
- [BIFD01] F. Baker, C. Iturralde, F. Le Faucheur, and B. Davie. *Aggregation of RSVP for IPv4 and IPv6 Reservations*. IETF, 2001. Internet Draft.
- [BLB98] Steven Berson, Robert Lindell, and Robert Braden. An Architecture for Advance Reservations in the Internet. Technical report, ISI, 1998.
- [Bre01] F. Breiter. *Dienstgütebasiertes Routing für Vorabreservierungen im Internet*. PhD thesis, Computer Science, Dresden University of Technology, 2001.
- [BZB<sup>+</sup>97] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. *Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification*. IETF, 1997. RFC2205.
- [DV00] Pressemitteilung DFN-Verein. Gigabit-Wissenschaftsnetz: Etappen für den Aufbau des Internet2 in Deutschland. <http://www.dfn.de/presse/dfn-presse/pm00-06-30d.html>, Juni 2000.
- [FGV97] Domenico Ferrari, Amit Gupta, and Giorgio Ventre. Distributed Advance Reservation of Real-Time Connections. *Multimedia Systems*, 5(3):187–198, 1997.
- [FKL<sup>+</sup>99] Ian Foster, Carl Kesselmann, Craig Lee, Bob Lindell, Klara Nahrstedt, and Alain Roy. A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation. In *IFIP Seventh International Workshop on Quality of Service, IWQoS*, pages 27–36, London, 1999. IEEE.
- [HMB<sup>+</sup>00] Hans-Christian Hege, André Merzky, Werner Bengler, Friedbert Kasper, Thomas Radke, and Edward Seidel. Schwarze Löcher in Sicht. *DFN Mitteilungen*, 52:4–6, Februar 2000.
- [Lus77] Hanan Luss. A Model for Advanced Reservations for Intercity Visual Conferencing Services. *Operational Research Quarterly*, 28(2):275–284, 1977.
- [Lus80] Hanan Luss. A Model for Advanced Reservations for Large Scale Conferencing Services. *Journal of the Operational Research Society*, 31(3):239–245, 1980.

- [MBB<sup>+</sup>97] A. Mankin, F. Baker, B. Braden, S. Bradner, M. O'Dell, A. Romanow, A. Weinrib, and L. Zhang. *Resource ReSerVation Protocol (RSVP) Version 1 Applicability Statement, Some Guidelines on Deployment*. IETF, 1997. RFC2208.
- [NBBB98] K. Nichols, S. Blake, F. Baker, and D. Black. *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*. IETF, 1998. RFC2474.
- [Pat00] Jochem Pattloch. *DFNConnect-Flexible Bandbreiten auf den Punkt gebracht*. *DFN Mitteilungen*, 52:16–17, Februar 2000.
- [Rei95] Wilko Reinhardt. *Advance Resource Reservation and its Impact on Reservation Protocols*. In *Broadband Island*, 1995.
- [RLLS88] J.W. Roberts, Y. Liang, K. Liao, and A. Simonian. *Queueing models for reserved set up telecommunications services*. *12th International Teletraffic Congress, Torino*, june 1988.
- [RVM96] James W. Roberts, Jorma Virtamo, and Ugo Mocci. *Broadband Network Teletraffic, Final Report of Action COST 242*. Springer Verlag, 1996.
- [Sch98] Olov Schelén. *Quality of Service Agents in the Internet*. PhD thesis, Luleå University of Technology, 1998.
- [Sie97] G. Siegmund. *ATM - Die Technik; Grundlagen, Netze, Schnittstellen, Protokolle*. Hüthig Verlag Heidelberg, 1997.
- [SNNP99] Olov Schelén, Andreas Nilsson, Joakim Norrgård, and Stephen Pink. *Performance of QoS Agents for Provisioning Network Resources*. In *IFIP Seventh International Workshop on Quality of Service, IWQoS*, pages 17–26, London, 1999. IEEE.
- [Ste99] R. Steinmetz. *Multimedia-Technologie*. Springer-Verlag, 1999.
- [THD<sup>+</sup>99] B. Teitelbaum, S. Hares, L. Dunn, R. Neilson, R. Narayan, and F. Reichmeyer. *Internet2 qbone: Building a Testbed for Differentiated Services*. *IEEE Network*, 13(5):8–16, September /October 1999.
- [VA91] J. T. Virtamo and S. Aalto. *Stochastic optimization of reservation systems*. *European Journal of Operational Research*, 51:327–337, 1991.
- [Vir92] J. T. Virtamo. *A Model of Reservation Systems*. *IEEE Transactions on Communications*, 40(1):109–118, Januar 1992.
- [WDS<sup>+</sup>95] Lars C. Wolf, Luca Delgrossi, Ralf Steinmetz, Sibylle Schaller, and Hartmut Wittig. *Issues of Reserving Resources in Advance*. In *Fifth International Workshop on Network and Operating System Support for Digital Audio and Video*, 1995.
- [ZS00] Florian Zeilhofer and Robert Sader. *Netze im virtuellen OP*. *DFN Mitteilungen*, 52:7–8, Februar 2000.

**Sabine Kühn** studierte Informatik an der Technischen Universität Dresden. Ihre Diplomarbeit zum Thema *Integration von IPv6 und ATM* erstellte sie 1995 am European Applied Research Center in Karlsruhe (SAP). Im Anschluß daran wurde sie wissenschaftliche Mitarbeiterin am Lehrstuhl Rechnernetze der TU-Dresden. Im Rahmen des DFG-Projektes *Anwendungs- und Protokollmanagement in Hochleistungsnetzen* war sie maßgeblich an

der Integration und Implementierung RSVP-fähiger IP/IPv6 Netzwerke und ATM beteiligt. Von 1997 bis 2001 bearbeitete sie das DFG-Projekt Ressourcenreservierung im Voraus". Ihre Dissertation Ressourcenmanagement und Zugangskontrolle eines Vorabreservierungssystems schloß sie im Jahr 2001 mit Auszeichnung ab. Seit Oktober 2001 arbeitet sie bei DANTE (Delivery of Advanced Network Technology to Europe Ltd) in Cambridge (GB) und beschäftigt sich im Rahmen des pan-europäischen Forschungsnetzes GÉANT unter anderem mit den Projekten 6NET und SEQUIN.