

LGMP/LGCP: Eine Protokoll-Suite für skalierbare Multicast-Kommunikation im Internet

Markus Hofmann

Institut für Telematik, Universität Karlsruhe¹
Zirkel 2, 76131 Karlsruhe, Germany
E-Mail: hofmann@acm.org

Zusammenfassung. Neben der klassischen Zweiparteien-Kommunikation gewinnt die Datenübertragung von einem Sender an mehrere Empfänger, die sogenannte *Multicast*-Kommunikation, zunehmend an Bedeutung. Ein großes Spektrum moderner Anwendungen aus den Bereichen verteilte Systeme, verteilte rechnergestützte Gruppenarbeit, Konferenzsysteme und Verteildienste basiert auf der Kommunikationsform des Multicast. Um den Qualitätsanforderungen dieser Anwendungen zu genügen, müssen Kommunikationssysteme über Mechanismen zur Fehlerkorrektur und zur Verkehrssteuerung verfügen. Multicast-Kommunikation erfordert im Gegensatz zur klassischen Zweiparteien-Kommunikation die Unterstützung unterschiedlich großer Empfängermengen. Die Gruppengröße kann von einigen wenigen Kommunikationsteilnehmern bis hin zu einigen tausend Empfängern variieren. Dies erfordert den Entwurf neuer Protokollmechanismen, die hinsichtlich der Empfängerzahl skalierbar ausgelegt sind und die Koordination großer Empfängermengen ermöglichen. Der vorliegende Artikel gibt einen Überblick über die Dissertationsarbeit des Autors, in welcher eine Protokoll-Suite für skalierbare Multicast-Kommunikation im Internet entwickelt, spezifiziert, simuliert, implementiert und bewertet wurde [Hof98].

Einleitung

Erste Ansätze zur Realisierung von Multicast-Diensten basieren auf erweiterten Mechanismen der klassischen Zweiparteien-Kommunikation [Str95], [KoZ96], [MRT98]. Diese sind durch eine senderorientierte Ausrichtung gekennzeichnet, wobei die Verantwortung sowohl für die Steuerung des Datenflusses als auch für die Korrektheit der Datenübertragung der sendenden Instanz obliegt. Zur Durchführung einer Verkehrssteuerung wertet der Sender die Statusmeldungen aller Empfänger aus und regelt den Datenfluß entsprechend dem Zustand der Empfängermenge. Ist diese Vorgehensweise bei der klassischen Zweiparteien-Kommunikation noch unproblematisch, so führt die Bearbeitung einer Vielzahl von Statusmeldungen im Falle großer Kommunikationsgruppen zu einer Überlastung des Senders. Zudem erfordert die Übermittlung mehrerer Statusmeldungen zusätzliche Bandbreite, was insgesamt in

¹ Markus Hofmann arbeitet derzeit im Networking Software Research Department der Bell Laboratories in Holmdel, New Jersey, USA.

einer deutlichen Verminderung der effektiven Übertragungsleistung und in einer ineffizienten Ressourcenauslastung resultiert. Auch die klassischen Verfahren zur Fehlerkorrektur, bei welchen die Empfänger fehlende Datenpakete stets direkt beim Sender anfordern, sind nicht geeignet für die Multicast-Kommunikation in Weitverkehrsnetzen. Eine solche Vorgehensweise eignet sich im Falle der Multicast-Kommunikation lediglich für den Einsatz in lokalen Netzwerken. Die Kosten zur Durchführung von Übertragungswiederholungen fallen hier vergleichsweise gering aus. Dagegen ist in Weitverkehrsnetzen die Reduzierung der Netzbelastung ein vorrangiges Ziel. Mit zunehmender Empfängerzahl wächst die Wahrscheinlichkeit eines Paketverlustes und damit die Anzahl notwendiger Übertragungswiederholungen. Messungen im Internet zeigen, daß oftmals bis zu 70% der Datenpakete aufgrund eines Paketverlustes bei mindestens einem Empfänger wiederholt zu übertragen sind [YKT96]. Eine senderorientierte Durchführung von Übertragungswiederholungen führt in diesem Fall zu einem erheblichen Mehrbedarf an Bandbreite und damit zu höheren Kosten. Dies verdeutlicht die Notwendigkeit neuer Fehlerkorrekturmechanismen zur Realisierung skalierbarer Multicast-Dienste, wie sie beispielsweise auch in [NBT97] diskutiert werden.

Aus der geschilderten Problematik ergab sich als Zielsetzung der Dissertation die Entwicklung und die Realisierung eines fortschrittlichen Kommunikationssystems, welches unter Berücksichtigung anwendungsspezifischer Anforderungen einen skalierbaren Multicast-Dienst bereitstellt. Das entwickelte Kommunikationssystem übernimmt dabei Funktionen zum Gruppenmanagement und zur dynamischen Strukturierung der Gruppenmitglieder. Dies ermöglicht erstmals die Berücksichtigung der Gruppentopologie zur Optimierung sowohl der Fehlerkorrektur als auch der Quittungsbearbeitung. Im Gegensatz zu klassischen Ansätzen löst sich das im Rahmen der Arbeit entwickelte Konzept vom senderorientierten Paradigma und bezieht statt dessen alle beteiligten Komponenten aktiv in die Fehlerkorrektur und in die Steuerung der Datenübertragung ein. Dies ermöglicht eine faire Verteilung der Bearbeitungslast auf unterschiedliche Kommunikationsteilnehmer, wodurch der Sender im Vergleich zu klassischen Verfahren deutlich entlastet wird. Ein wichtiger Entwurfsaspekt besteht dabei in der Strukturierung der Gesamtgruppe. Hierfür wurde ein Verfahren entwickelt, das unter Berücksichtigung der Anwendungsanforderungen und der aktuellen Netzauslastung die Kommunikationsgruppe in mehrere funktionale Einheiten aufteilt und deren Koordination übernimmt.

Konzeptentwurf

Ausgehend von einer Analyse existierender Multicast-Protokolle wurden mehrere Algorithmen zur Unterstützung einer skalierbaren Gruppenkommunikation erarbeitet [Hof98]. Die entwickelten Konzepte resultierten in einem generischen Rahmenwerk für skalierbare Multicast-Kommunikation in Weitverkehrsnetzen, dem sogenannten *Local Group Concept (LGC)* [Hof96]. Die grundlegende Idee besteht in einer logischen Strukturierung der globalen Kommunikationsgruppe. Diese wird in mehrere lokale Untergruppen aufgeteilt, welche ihrerseits in einer baumartigen Hierarchie angeordnet werden. Dabei faßt jede lokale Gruppe nahe beieinander gelegene

Kommunikationsteilnehmer zusammen. Als Entfernungsmaß werden mehrere Metriken herangezogen, die in Abhängigkeit vom jeweiligen Anwendungskontext unterschiedlich gewichtet werden. Im Gegensatz zu klassischen Kommunikationsprotokollen können beim Local Group Concept auch Empfänger Übertragungswiederholungen ausführen. Durch Kooperation der Mitglieder einer lokalen Untergruppe werden Übertragungsfehler zunächst ohne Beteiligung des Senders korrigiert. Haben alle Mitglieder einer lokalen Gruppe bestimmte Datenpakete nicht korrekt erhalten, so ist eine lokale Fehlerkorrektur nicht möglich. In diesem Fall werden die fehlenden Daten bei der übergeordneten Teilgruppe bzw. direkt beim Sender angefordert. Das entwickelte Verfahren beschränkt den überwiegenden Teil der notwendigen Übertragungswiederholungen auf den lokalen Bereich einer Untergruppe, was sich in einer deutlichen Entlastung der Weitverkehrsverbindungen niederschlägt. Lokale Übertragungswiederholungen werden durch einen geeigneten Empfänger, den sogenannten *Repräsentanten* der lokalen Gruppe, koordiniert. Dieser wertet zudem die Statusmeldungen der angeschlossenen Empfänger aus, faßt sie in einer einzelnen Nachricht zusammen und übermittelt diese dem Repräsentanten der übergeordneten lokalen Gruppe bzw. dem Multicast-Sender. Diese Vorgehensweise ermöglicht zusammen mit der Dekomposition der Gesamtgruppe in voneinander unabhängige Teilbereiche eine verteilte und damit parallele Steuerung des Datenflusses, was die Skalierbarkeit hinsichtlich großer Empfängerzahlen verbessert.

Protokollspezifikation und simulative Bewertung

Ausgehend von den entwickelten Konzepten wurden zwei voneinander unabhängige Protokolle entworfen, deren Gesamtfunktionalität dem Benutzer über eine erweiterte Anwendungsschnittstelle bereitgestellt wird [Hof98]. Das *Local Group based Multicast Protocol (LGMP)* ist für den eigentlichen Datentransfer verantwortlich. Es operiert auf einer hierarchischen Strukturierung der globalen Kommunikationsgruppe, welche vom *Local Group Configuration Protocol (LGCP)* erstellt wird. Die Anordnung der Untergruppen und damit der Aufbau der Hierarchie erfolgt in Abhängigkeit von den anwendungsspezifischen Anforderungen. Somit wird eine Optimierung der für die Anwendung wichtigen Kenngrößen erreicht. Dies kann beispielsweise die Minimierung der durchschnittlichen Übertragungsverzögerung oder die Reduzierung der globalen Netzlast sein. Aufgrund dynamisch veränderbarer Gruppenzugehörigkeit und schwankender Netzauslastung muß die definierte Gruppenstruktur während der Kommunikation an die aktuellen Gegebenheiten angepaßt werden. Das vom Datentransfer entkoppelte Management der Gruppenhierarchie erlaubt eine weitgehende Unabhängigkeit von Verwaltungs- und Datentransferfunktionen und damit deren parallele Bearbeitung. Der Austausch von Verwaltungs- und Statusinformation erfolgt periodisch, wodurch implizit eine hohe Robustheit und Stabilität des Gesamtsystems erreicht wird.

Simulative Untersuchungen der entworfenen Protokollmechanismen zeigen wesentlich bessere Leistungsdaten im Vergleich zu Verfahren mit ähnlicher Funktionalität. Um die Leistungsfähigkeit und die Skalierbarkeit des entwickelten Konzeptes zu belegen, wurden ausgehend von Meßergebnissen im Internet mehrere

Simulationsmodelle erstellt. Diese ermöglichen eine realitätsnahe Untersuchung von Szenarien mit mehreren tausend Empfängern, was durch Messungen mit einer realen Implementierung kaum zu erreichen ist. Die Ergebnisse zeigen, daß durch die Methode der lokalen Übertragungswiederholungen die globale Netzbelastung deutlich unter derjenigen klassischer Ansätze bleibt, wobei zugleich die durchschnittliche Übertragungsverzögerung merklich reduziert wird.

Implementierung und Bewertung

Zur Umsetzung des Entwurfs wurden die Protokolle LGMP und LGCP als erweiterte endliche Automaten modelliert, die über einen asynchronen Nachrichtenaustausch miteinander kommunizieren [Hof98]. Die Implementierung bildet die Automatenmodelle auf eigenständige Prozesse ab, welche über einen gemeinsamen Speicherbereich miteinander kommunizieren. Beide Prozesse zerfallen intern in mehrere Leichtgewichtsprozesse (Threads), was einen hohen Grad an Parallelität ermöglicht. Die derzeitige Implementierung ist ausgehend von einer anwendungsorientierten Parametrisierung in der Lage, selbständig und ohne manuelles Eingreifen eine geeignete Gruppenhierarchie zu etablieren und diese gemäß den Anforderungen der Anwendung dynamisch an Änderungen im Systemumfeld anzupassen.

Als Implementierungsplattform standen Arbeitsplatzrechner der Firma Digital unter dem Betriebssystem Digital Unix 3.2 und 4.0 sowie SUN Workstations mit dem Betriebssystem Solaris 2.5.1 und 2.6 zur Verfügung. Neben diesen Plattformen wurden die erstellten Implementierungen auch unter Linux 2.0.30 und SCO UnixWare 7.01 erfolgreich getestet. Eine Implementierung von LGCP unter Windows NT 4.0 ist bereits verfügbar, die Portierung von LGMP befindet sich derzeit in Arbeit. Während der Programmerstellung wurde besonders auf eine hohe Stabilität und eine gute Leistungsfähigkeit der Implementierung geachtet. Der erstellte Programmcode ist öffentlich verfügbar und kann über das Internet kostenlos bezogen werden [LGC98].

Neben dem eigentlichen Kommunikationssystem wurde zu Demonstrationszwecken ein Monitorsystem entworfen und implementiert, welches über eine grafische Schnittstelle die jeweils aktuelle Gruppenstruktur visualisiert und ausführliche Informationen zu den einzelnen Kommunikationsteilnehmern bereitstellt. Ebenso wurde ein Java-basiertes Meßsystem entwickelt und implementiert, das eine automatisierte und zentral gesteuerte Durchführung von Messungen zur Bewertung von Multicast-Protokollen erlaubt. Durch seine einfache Handhabung und die hohe Portabilität eignet es sich für den Einsatz in großen globalen Kommunikationsgruppen.

Zur Überprüfung der Korrektheit und der Leistungsfähigkeit der Protokollimplementierungen wurden intensive Tests im Mbone unter Beteiligung von Testpartnern in Deutschland, England, Frankreich, Kanada und den USA durchgeführt. Neben der eigentlichen Leistungsuntersuchung stand dabei auch die Überprüfung der Stabilität und der Ausfallsicherheit im Blickpunkt des Interesses. Die zahlreichen Experimente zeigen zugleich, daß die entwickelten Konzepte nicht nur auf dem Papier Bestand

Protokollen die Übertragungszeit der 1 MByte großen Datei auf weniger als die Hälfte und das Datenaufkommen auf etwa ein Viertel reduziert werden. Der Anwender profitiert bei Verwendung von LGMP von einer merklich schnelleren Kommunikation und einem wesentlich kostengünstigeren Datenaustausch.

Zusammenfassung

Zusammenfassend ist zu sagen, daß durch die Popularität und die zunehmende Kommerzialisierung des Internet die skalierbare Multicast-Kommunikation in Weitverkehrsnetzen weiterhin an Wichtigkeit gewinnt. Unterstützt wird diese Tendenz durch die zunehmende Akzeptanz rechnergestützter Informations- und Verteildienste. Die im Rahmen dieser Arbeit entwickelten Konzepte und die erstellten Implementierungen stellen Grundbausteine zur Realisierung skalierbarer Multicast-Anwendungen im Internet dar. Im Gegensatz zu den bisher bekannten Ansätzen im Bereich der Multicast-Kommunikation legt das Local Group Concept (LGC) besonderen Wert auf die automatische Konfiguration und Etablierung der Gruppenstruktur. So ist zur Einteilung der Empfänger in lokale Gruppen keinerlei manuelle Administration erforderlich. Zudem zeichnen sich die definierten Verfahren durch eine hohe Robustheit gegenüber Systemausfällen und Netzwerkstörungen aus. Die Automatisierung der Gruppeneinteilung ermöglicht die dynamische Anpassung der Struktur an sich ändernde Randbedingungen. Damit wird erstmals der praktische Einsatz eines hierarchischen Multicast-Protokolls in hochdynamischen Netzwerken, wie dem Internet, ohne den manuellen Eingriff durch Systemadministratoren ermöglicht.

Literatur

- [Hof96] M. Hofmann: *A Generic Concept for Large-Scale Multicast*; International Zurich Seminar on Digital Communication, Februar 21-23, 1996, Zürich, Switzerland, In: B. Plattner (Hrsg.), *Lecture Notes in Computer Science*, No. 1044, Springer Verlag, 1996.
- [Hof98] M. Hofmann: *Skalierbare Multicast-Kommunikation in Weitverkehrsnetzen*. Dissertation an der Fakultät für Informatik, Universität Karlsruhe, erschienen im Infix Verlag, Dissertationen zu Datenbanken und Informationssystemen (DISDBIS), Band 42, ISBN 3-89601-442-0, Februar 1998.
- [KoZ96] A. Koifman, S. Zabele: *RAMP: A Reliable Adaptive Multicast Protocol*; Proc. of IEEE INFOCOM'96, San Francisco, CA., USA, März 1996.
- [LGC98] Local Group Concept: *Homepage im WWW*. <http://www.telematik.informatik.uni-karlsruhe.de/~hofmann/lgc/>, Dezember 1998.
- [NBT97] J. Nonnenmacher, E. W. Biersack, Don Towsley: *Parity-based loss recovery for reliable multicast transmission*. ACM SIGCOMM'97, Cannes, Frankreich, September 1997.
- [MRT98] K. Miller, K. Robertson, A. Tweedly, M. White: *StarBurst Multicast File Transfer Protocol (MFTP) Specification*; Work in Progress, Internet Draft (draft-miller-mftp-spec-03.txt), April 1998.
- [Str95] W.T. Strayer (Hrsg.): *Xpress Transport Protocol Specification, Revision 4.0*; Erhältlich vom XTP Forum, Santa Barbara, USA, März 1995.
- [YKT96] M. Yajnik, J. Kurose, D. Towsley: *Packet Loss Correlation in the Mbone Multicast Network*; Technical Report 96(32), University of Massachusetts at Amherst, 1996.