

# Sub-Layer-Hopping: Optimierte Staukontrolle für IP-Multicast (Erweiterter Bericht)<sup>1</sup>

Stefan Dresler, Markus Hofmann<sup>2</sup>, Frank Müller

Institut für Telematik, Universität Karlsruhe  
Zirkel 2, 76128 Karlsruhe, Germany  
Phone: +49 721 608-6397, Fax: +49 721 388097  
E-Mail: [dresler, hofmann, fmueller]@telematik.informatik.uni-karlsruhe.de

**Zusammenfassung.** Die Multicast-Kommunikation im Internet ist derzeit Gegenstand zahlreicher Forschungsaktivitäten. Konzentrierte sich in jüngster Vergangenheit die überwiegende Zahl der Arbeiten auf die Bereitstellung skalierbarer Mechanismen zur Fehlerbehebung, so rückt derzeit die Entwicklung geeigneter Algorithmen zur Staukontrolle in den Mittelpunkt des Interesses. Eine besondere Herausforderung besteht dabei in der Unterstützung heterogener Kommunikationsgruppen, innerhalb derer die Empfänger unterschiedliche Stausituationen erfahren. Ein vom Konzept her vielversprechender Ansatz ist der sogenannte *Layered-Multicast*, bei dem der eigentliche Datenstrom in mehrere Unterströme (*Layers*) aufgeteilt wird. Jedes Gruppenmitglied empfängt durch Beitritt zu den entsprechenden Untergruppen eine seiner aktuellen Stau- und Netzcharakteristik angepaßte Menge von Teilströmen. Dadurch wird trotz Nutzung der effizienten Multicast-Technologie eine gewisse Individualität erreicht. Bisher scheiterte die praktische Umsetzung dieses vielversprechenden Ansatzes jedoch an der Trägheit heutiger Routingprotokolle. Diese sind nicht in der Lage, auf die notwendigen Wechsel in der Gruppenmitgliedschaft in einer akzeptablen Zeit zu reagieren. Insbesondere verhindern die langen Austrittszeiten des *Internet Group Management Protocol (IGMP)* eine effiziente und schnelle Regelung des Datenvolumens. Vor diesem Hintergrund wird im vorliegenden Artikel ein Verfahren entwickelt und bewertet, welches trotz der genannten Beschränkungen erstmals die praktische Umsetzung des Layered Multicast in heutigen Netzen ermöglicht. Der entwickelte Ansatz, welcher als *Sub-Layer-Hopping (SLH)* bezeichnet wird, kommt ohne Modifikationen an netzinternen Komponenten oder Routingprotokollen aus. Er basiert auf einer geschickten Nutzung des durch IP bereitgestellten Multicast-Dienstes. Analytische Betrachtungen und experimentelle Messungen zeigen, daß der Verwaltungs- und Kontrollaufwand innerhalb der Router durch das Verfahren kaum beeinflußt wird. Die erstellte Implementierung wurde schließlich für Messungen im Mbone unter Beteiligung internationaler Partner eingesetzt. Es wird belegt, daß bei Einsatz des entwickelten Verfahrens eine höhere Empfangsrate und eine gewisse Fairneß gegenüber dem restlichen TCP-Verkehr erreicht wird.

## 1 Einführung

Bedingt durch das exponentielle Wachstum der Benutzerzahlen steigt die Belastung des Internets dramatisch an. Verstärkt wird dieser Aspekt durch den zunehmenden Einsatz datenintensiver Multimedia-Anwendungen und die Nutzung des Internets für die klassische Telefonie. Unter diesen Bedingungen wird ein vollständiger Netzzusammenbruch vor allem durch den Einsatz leistungsfähiger und fairer Mechanismen zur Staukontrolle vermieden. Sie stellen praktisch den Garanten für das Funktionieren des heutigen Internets dar. Ohne solche Mechanismen könnten Stausituationen im Netz nicht aufgelöst werden, was zu einem Kollabieren der Kommunikationsinfrastruktur führen würde. Aus diesem Grunde mißt die *Internet Engineering Task Force (IETF)* bei der Bewertung neuer Internet-Protokolle der vorgesehenen Staukontrolle eine außerordentliche Bedeutung bei. Dies trifft insbesondere auch für die Entwicklung und Standardisierung von Multicast-Protokollen zu [MRB98]. Gerade bei der Multicast-Kommunikation, welche Daten oftmals an hunderte oder tausende weltweit verteilter Empfänger übermittelt, stellt die Staukontrolle eine äußerst kritische Protokollkomponente dar. Mangelnde Fairneß oder ineffiziente Algorithmen zeigen hier sogleich globale Auswirkungen.

Leider ist der seit Jahren bewährte und etablierte *Slow-Start* Mechanismus von TCP [Jac88] in unveränderten Form nicht geeignet für die Multicast-Kommunikation. Die Staukontrolle von TCP basiert auf einer Anpassung

---

<sup>1</sup> Bei dem vorliegenden Bericht handelt es sich um eine erweiterte Fassung der im Konferenzband der Konferenz "Kommunikation in Verteilten Systemen (KiVS) '99" erscheinenden Version [DrH99].

<sup>2</sup> Markus Hofmann arbeitet derzeit im Networking Software Research Department der Bell Laboratories in Holmdel, New Jersey, USA.

des Sendefensters entsprechend der Rückmeldungen des Empfängers. Im Falle großer Kommunikationsgruppen führen solche Rückmeldungen jedoch zur sogenannten *Sender-Impllosion* [Hof98]. Es wurden daher frühzeitig Methoden entwickelt, die den Sender bei der Bearbeitung von Rückmeldungen entlasten und dennoch eine TCP-ähnliche Staukontrolle ermöglichen<sup>3</sup> [Hof96, PSL97]. Jedoch ist die Anpassung der Senderate bei der Multicast-Kommunikation ungleich schwieriger. In diesem Falle müssen mehrere Empfänger berücksichtigt werden, welche unter Umständen unterschiedliche Stausituationen erfahren und verschiedene Ansprüche an die Leistungsfähigkeit des Multicast-Dienstes haben. Multicast-Protokolle mit einer TCP-ähnlichen Staukontrolle passen ihre Senderate in einem solchen Fall stets dem jeweils schlechtesten Empfänger an. Obwohl ein solches Verhalten in vielen Anwendungsszenarien aus Fairneßgründen erwünscht ist, so reduziert es doch deutlich den insgesamt erreichbaren Durchsatz [Hof98].

Wünschenswert ist demnach ein Verfahren, welches unter Berücksichtigung der individuellen Anforderungen und Eigenschaften einzelner Empfänger sowie der augenblicklichen Netzauslastung eine effiziente und zugleich faire Staukontrolle für Multicast-Übertragungen ermöglicht. Ein vielversprechender Ansatz dazu wurde mit dem *Layered Multicast* entwickelt [MJV96, CRV98, Cla97], der sich für die Übertragung von sowohl Multimediaströmen als auch von Dateien eignet. Videoströme lassen sich beispielsweise häufig in mehrere Einzelströme aufteilen, von denen einer eine in Auflösung, Frequenz oder Farbinformation reduzierte Version darstellt, welche durch Hinzunahme jeweils weiterer Ströme sukzessive verbessert werden kann [Lan98]. Beim Dateitransfer, dem Schwerpunkt dieses Dokuments, werden die zu sendenden Daten segmentiert und mittels eines geeigneten Kodierungsverfahrens (meist ein Reed-Solomon Code [Lin82, Hen97, Riz97]) transformiert. Die so gebildete Matrix aus Nutz- und Redundanzdaten wird entsprechend einem vorgegebenen Schema an  $n$  verschiedene Multicast-Gruppen gesendet (siehe [Vic97]). Diese Multicast-Gruppen bilden jeweils einen sogenannten *Layer* und werden eindeutig durch eine Multicast-Adresse beschrieben. Je mehr Layern ein Empfänger beitrifft, desto schneller kann er die Datenübertragung bei gleichzeitig höherer Netzbelastung beenden. Es ist jedoch auch möglich, alle Daten erfolgreich über eine einzelne Multicast-Adresse zu erhalten. Diese Flexibilität ermöglicht den einzelnen Empfängern je nach Paketverlustrate den Layers  $L_0, \dots, L_m$  (mit  $0 \leq m \leq n-1$ ) beizutreten und somit eine Ratenanpassung an die aktuelle Stausituation im Netz vorzunehmen.

Abschnitt 2 erläutert dieses Verfahren etwas detaillierter und zeigt zugleich die Probleme bei einer praktischen Umsetzung auf. Das in Abschnitt 3 entwickelte Verfahren zeigt eine praktikable Lösung für diese Probleme auf und ermöglicht eine Steigerung der Durchsatzrate ohne nennenswerten Mehraufwand durch Austausch von Routinginformationen. Dies wird durch Messungen in Abschnitt 4 belegt. In Abschnitt 5 werden die Ergebnisse schließlich zusammengefaßt.

## 2 Grundlagen des Layered Multicast

Wie in der Einleitung bereits angedeutet, basiert die Staukontrolle durch Layered Multicast auf dem dynamischen Beitreten und Austreten individueller Empfänger zu den verschiedenen Multicast-Strömen. Der exakte Ablauf ist durch folgenden Algorithmus gegeben:

1. Betrachtet wird ein Empfänger, welcher gerade die Layer  $L_0, \dots, L_i$  empfängt (d.h. er befindet sich in *Session*  $S_i$ ). Sobald die Paketverlustrate des Empfängers eine bestimmte Schwelle überschreitet, verläßt er den Layer  $L_i$ . Bei entsprechender Einteilung der Layer wird die Empfangsrate auf diese Weise halbiert, was einer *multiplikativen Reduktion* gleichkommt.
2. Bleibt die Paketverlustrate für einen Zeitraum  $t_p(i)$  unter einem Schwellenwert, so tritt der Empfänger in Layer  $L_{i+1}$  ein.  $t_p(i)$  ist proportional zum Durchsatz von  $L_i$ . Man erhält so eine *additive Anhebung* der Empfangsrate.

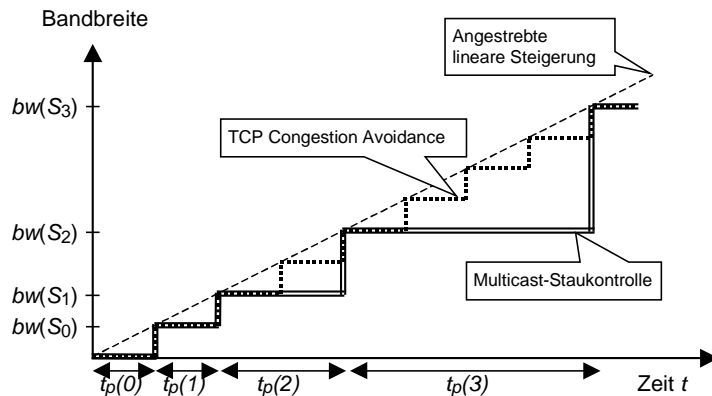
Aufgrund der exponentiellen Abstufung der Layer-Raten ist zunächst nicht offensichtlich, wie im Punkt zwei eine lineare Zunahme erreicht werden kann. Im Moment der Hinzunahme eines weiteren Layers verdoppelt sich natürlich die Empfangsrate, was einer multiplikativen Anhebung entspricht. Durch den Proportionalitätsfaktor  $t_p(i)$  ist jedoch über die Zeit gesehen der Charakter einer additiven Anhebung erreichbar (Abbildung 2). Dafür sorgt die Wahl von  $t_p(i)$  proportional zur Rate von  $L_i$ . Diese Proportionalität ist wieder unmittelbar von TCP übernommen.

Um mit diesem Verfahren einen überlasteten Pfad im Netz zu entlasten, müssen alle betroffenen Empfänger<sup>4</sup> möglichst schnell auf die Stausituation reagieren und geschlossen aus den entsprechenden Layern austreten. Nur

<sup>3</sup> Dies ist bei sogenannten empfängerorientierten Protokollen, wie beispielsweise SRM [FJM95], *nicht* möglich.

<sup>4</sup> Dies sind genau die im Multicast-Routing-Baum downstream liegenden Empfänger.

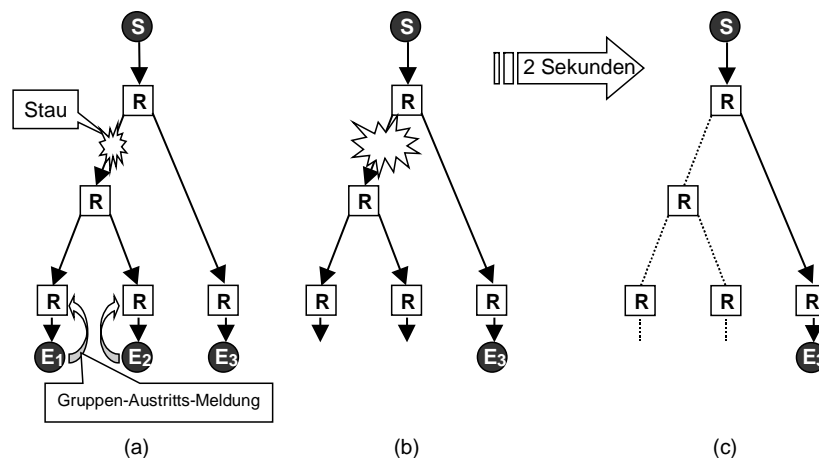
dadurch wird eine Reduzierung des Datenaufkommens auf dem überlasteten Link erreicht. Eine solche Synchronisation der Empfänger kann beispielsweise durch Hinzufügen von Kontrollinformationen an die Nutzdaten erreicht werden [CRV98]. Unter der Annahme, daß alle Empfänger den gleichen Algorithmus für die Staukontrolle verwenden und Paketverluste hauptsächlich zwischen Routern auftreten, wird eine einheitliche Sicht der Empfänger auf die Stausituation und damit die Mitgliedschaft in den gleichen Layern gewährleistet.



**Abbildung 1: Additive Anhebung des Durchsatzes**

Von entscheidender Bedeutung für das Funktionieren einer Staukontrolle durch Layered Multicast ist die schnelle Reaktion der Routingprotokolle auf das Austreten einzelner Empfänger. Nur wenn diese Information netzintern schnell genug propagiert wird, ist ein rechtzeitiges Reagieren auf Stausituationen möglich. Wie Messungen gezeigt haben, sind heutige Routingprotokolle, wie beispielsweise das *Distance Vector Multicast Routing Protocol (DVMRP)* [DPW88, Pus98], dazu durchaus in der Lage [Mül98]. In den lokalen Netzen, welche typischerweise die Blätter des Multicast-Routingbaums bilden, vergeht bei Nutzung von *IGMP (Internet Group Management Protocol)* [Dee89] jedoch eine Wartezeit von bis zu 125s bzw. bei *IGMPv2* [Fen98] von bis zu 2s, ehe der Austritt in diesem Subnetz wirklich vollzogen ist. Erst dann kann das Routingprotokoll reagieren und die Information über den Austritt in das Netz propagieren. In dieser Zeit kann sich die Stausituation im Netz jedoch bereits verändert haben. Die Wartezeit ist nötig, da der Multicast-Router — bedingt durch die ungesicherte Übertragung von IGMP-Nachrichten sowie die Unterdrückung überflüssiger Antworten der Teilnehmer auf die Anfrage des Multicast-Routers nach Mitgliedern der Gruppe — keine genauen Informationen über die Gruppenmitgliedschaft im Subnetz hat. Um die Verzögerung von 2s bei IGMPv2 weiter zu verringern, wird in [Riz98] die Modifikation des *Multicast-Routing-Daemon (mrouted)* vorgeschlagen. Die vorgeschlagenen Änderungen integrieren eine Heuristik, welche die Antwort einer Anfrage nach noch existierenden Gruppenmitgliedern vorhersagt. Durch diesen Indeterminismus entsteht jedoch eine erhöhte Wahrscheinlichkeit von Austritten aus Gruppen, die noch Mitglieder im Subnetz haben, was zu negativen Auswirkungen auf andere Anwendungen führen kann.

Die beschriebene Austrittsverzögerung führt zum Problem der sogenannten scheinbaren Paketverluste, welche im folgenden am Beispiel in Abbildung 2 erläutert werden.



**Abbildung 2: Entstehung von scheinbaren Paketverlusten**

Die vom Stau betroffenen Empfänger  $E_1$  und  $E_2$  führen einen Gruppen-Austritt, in der Regel für ihren oberen Layer, durch. Damit stoßen sie die Bearbeitung des Austritts im lokalen Netz an (Abbildung 2a). Durch die sofortige Reaktion des lokalen Interface werden ihnen von da an keine Pakete  $G$  mehr aus dem Netz in den Protokollstack hochgereicht.  $E_1$  und  $E_2$  sind keine Mitglieder der Empfänger-Gruppe von  $G$  mehr. Während IGMP seinen Austritts-Prozeß durchführt und nach verbliebenen Gruppenmitgliedern sucht, ändert sich am Routing-Baum noch nichts. Pakete für  $G$  werden also weiterhin bis in die lokalen Subnetze von  $E_1$  und  $E_2$  befördert, obwohl diese nicht mehr an  $E_1$  und  $E_2$  ausgeliefert werden (Abbildung 2b). Für  $E_1$  und  $E_2$  sind diese Pakete verloren, sie müssen später erneut übertragen werden (*scheinbare Paketverluste*). Dieser Zustand dauert für 2 Sekunden an. Sobald der IGMP Austritts-Prozeß abgeschlossen ist, kann das Routing-Protokoll über die Änderungen informiert werden. Daraufhin kommt es zur Anpassung des Routing-Baumes, die nun endlich zur Entlastung auf den überlasteten Teilstrecken führt (Abbildung 2c).

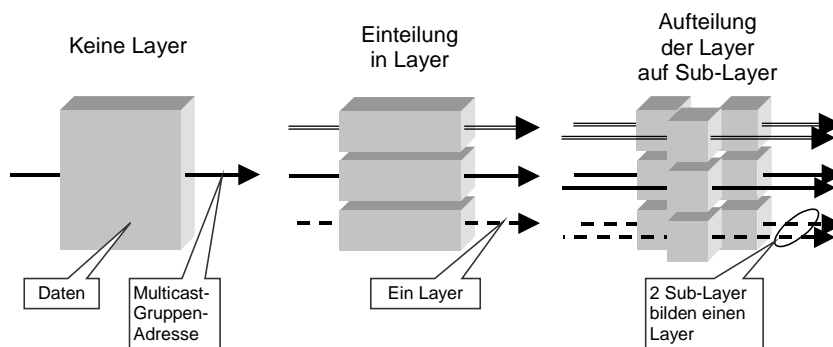
Im folgenden wird eine Technik entwickelt, welche bei Verwendung des Layered Multicast zu einer Verringerung der Austrittsverzögerung aus einem Layer führt, ohne Veränderungen an existierenden Systemen oder Protokollen durchführen zu müssen. Dabei wird auf den obersten zwei Layers einer Übertragung mit jeweils zwei Sub-Layers gearbeitet, welche alternierend verwendet werden. Diese Technik wird deshalb als Sub-Layer-Hopping bezeichnet.

### 3 Die Technik des Sub-Layer-Hoppings

Im folgenden wird zuerst das Konzept des Sub-Layer-Hoppings (SLH) eingeführt, das eine Verringerung der Reaktionszeit des Rückkanals — gebildet durch das Routing im Netzinnern und IGMP im lokalen Bereich — verspricht.

#### 3.1 Vorstellung des Konzepts

Bereits bei Einführung der Layer wurde die 1:1-Zuordnung zwischen einer Multicast-Datenübertragung und der dafür benutzten Multicast-Gruppenadresse aufgehoben. Jeder Layer bekam seine eigene Gruppenadresse. Nun wird eine weitere Aufteilung vorgenommen. Aus jedem Layer entstehen zwei *Sub-Layer*. In Abbildung 3 ist die Folge dieser sukzessiven Aufteilungen zusammenfassend dargestellt. Die grauen Quader stehen für die Daten und die Pfeile für die Multicast-Gruppenadressen, auf denen sie versendet werden.



**Abbildung 3: Aufteilung der Daten auf verschiedene Multicast-Adressen**

Die bisherige Aufteilung der Daten auf verschiedene Layer wird beibehalten. Zusätzlich erfolgt die Übertragung jedes Layers  $i$  nun aber über zwei Multicast-Gruppenadressen  $A_i$  und  $B_i$ . Sie bilden die *Sub-Layer*. Bei der Adressierung der Pakete des Layers  $i$  wechselt der Sender in periodischen Abständen zwischen den Ziel-Adressen  $A_i$  und  $B_i$  (Abbildung 4). Darin besteht die senderseitige Komponente des *Sub-Layer-Hopping*. Die beiden Sub-Layer werden also nie gleichzeitig benutzt. Als *aktiver* Sub-Layer wird immer derjenige bezeichnet, auf dem gerade übertragen wird. Zu diesem Zeitpunkt ist der andere Sub-Layer *passiv*. Der Zeitraum zwischen den Sub-Layer-Wechseln ist mit 2 Sekunden gerade so lang wie die Gruppenaustrittsverzögerung von IGMP.

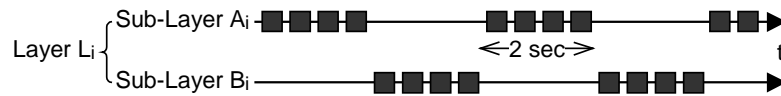


Abbildung 4: Periodischer Wechsel zwischen den Sub-Layern

### 3.2 Durchführung der Staukontrolle

Für die Empfänger sind durch die Einführung der Sub-Layer besondere Punkte während des Datenempfangs entstanden. An diesen sogenannten *Entscheidungspunkten* (EP) findet für einen bestimmten Layer der Sub-Layer-Wechsel statt. Die EP werden mittels in den Datenpaketen enthaltenen Informationen nahezu zeitgleich an alle Empfänger in einem Subnetz verteilt und dienen wie schon bei [CRV98] der Synchronisation. Um unterbrechungsfrei am Empfang eines Layers teilnehmen zu können, müssen die Empfänger hier den Wechsel auf die andere Gruppenadresse mitmachen. Dies ist die empfängerseitige Komponente des *Sub-Layer-Hopping*.

Ein Empfänger führt seine Staukontrolle zunächst nach dem bereits vom Layered Multicast bekannten Prinzip durch: bei Paketverlust wird ein Layer verlassen; bleibt der Paketverlust aus, nimmt man dagegen einen Layer hinzu. Die Entscheidung über Hinzunahme oder Entfernen eines Layers wird nun aber erst am folgenden Entscheidungspunkt in Form von Gruppenein- oder -austritt versandt.

Für den Umgang mit jedem Layer  $L_i$ , der Sub-Layer-Hopping benutzt, ergeben sich drei Fälle. Sie sind hier aus der Sicht eines beliebigen Empfängers  $E$ , der auf  $L_i$  empfängt, aufgeführt:

1. Ist gerade kein Layer-Wechsel nötig, dann wird beim Erreichen des nächsten Entscheidungspunktes ein Sub-Layer-Hopping durchgeführt. Damit ist sichergestellt, daß der Empfang auf  $L_i$  über den Wechsellpunkt hinaus nahtlos fortschreitet:
  - Für den Sub-Layer, der vor  $EP$  aktiv war, erfolgt ein Gruppenaustritt.
  - Für den Sub-Layer, der zuletzt inaktiv war, erfolgt ein Gruppenbeitritt.
2. Falls  $L_i$  der oberste Layer ist und die Paketverlustrate eine bestimmte Schwelle überschritten hat, muß  $L_i$  verlassen werden. Am nächsten Entscheidungspunkt werden folgende Aktionen durchgeführt:
  - Für den Sub-Layer, der vor  $EP$  aktiv war, erfolgt ein Gruppenaustritt.
  - Für den Sub-Layer, der zuletzt inaktiv war, erfolgt *kein* Gruppenbeitritt.
3. Bleibt die Paketverlustrate, während  $L_i$  der oberste Layer ist, solange unter einem Schwellenwert, daß  $L_{i+1}$  hinzugefügt werden kann, dann erfolgt am nächsten Entscheidungspunkt ein Gruppenbeitritt in den ab dort aktiven Sub-Layer von  $L_{i+1}$ .

Die Motivation hinter der Einführung des Sub-Layer-Hopping war eine Verbesserung der Leistung des Rückkanals. Deshalb muß man von diesen drei Fällen insbesondere den zweiten näher untersuchen. Dazu soll das Beispiel in Abbildung 5 betrachtet werden.

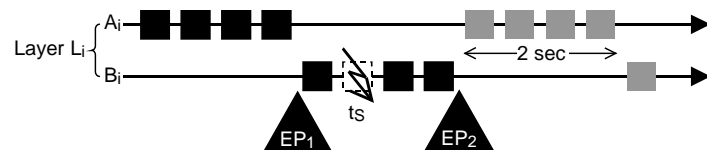


Abbildung 5: Staubehandlung beim Einsatz von Sub-Layer-Hopping

In der Abbildung sind Paketankünfte auf dem höchsten Layer  $L_i$  aus der Sicht eines Empfängers  $E$  dargestellt. Wir nehmen an, daß vor dem Entscheidungspunkt  $EP_1$  kein Stau aufgetreten sei, während zwischen  $EP_1$  und  $EP_2$  Stau entsteht. An den beiden Entscheidungspunkten kommt es in diesem Beispiel zu folgenden Abläufen:

- Bei  $EP_1$  wird ein Sub-Layer-Wechsel gemäß dem ersten Punkt von oben durchgeführt. Für die Gruppe  $A_i$  wird ein Gruppenaustritt, für  $B_i$  ein Gruppenbeitritt veranlaßt.
- Zum Zeitpunkt  $t_s$  wird nun Stau erkannt. Der Empfänger merkt sich, daß er aus  $L_i$  austreten muß, führt den Austritt aber noch nicht durch. Dadurch wird verhindert, daß aus den Paketen, die zwischen  $t_s$  und  $EP_2$  noch ankommen, scheinbare Paketverluste werden. Es ist zu beachten, daß diese Verzögerung des Gruppenaustritts keinerlei Verzögerung der Reduzierung des Datenaufkommens nach sich zieht. Insbesondere wird dadurch die Reaktionszeit für die Staubehebung *nicht* erhöht. Dies liegt darin begründet, da selbst bei einem sofortigen Austritt zum Zeitpunkt  $t_s$  wegen der Austrittsverzögerung von IGMP noch bis mindestens  $EP_2$  Daten in das Subnetz weitergeleitet werden.

- Bei  $EP_2$  schließlich greift der Empfänger auf die Information zurück, daß Stau aufgetreten ist, und entscheidet sich dafür, den Layer  $L_i$  zu verlassen. Er führt zunächst den Austritt aus  $B_i$  durch, den er auf jeden Fall gemacht hätte. Danach tritt er jedoch nicht in  $A_i$  ein. Damit ist der Austritt aus  $L_i$  vollzogen. Es kommt hier nach  $EP_2$  zu keiner zusätzlichen Austrittsverzögerung. Von den grau dargestellten Paketen sieht  $E$  nichts mehr.

Bemerkenswert daran ist, daß der Routing-Baum von  $L_i$  direkt nach dem Entscheidungspunkt, an dem der Layer verlassen wird — in diesem Beispiel war das  $EP_2$  —, an die neue Situation angepaßt ist und keinen Ast für  $A_i$  in Richtung  $E$  mehr hat. Das liegt daran, daß  $E$  — den Regeln des Sub-Layer-Hopping folgend — bei  $EP_1$  bereits aus der Gruppe  $A_i$  ausgetreten ist. Da der Abstand zwischen  $EP_1$  und  $EP_2$  gerade 2 Sekunden beträgt, wird dadurch die Austritts-Verzögerung überbrückt, und bis nach  $EP_2$  wieder die Gruppe  $A_i$  aktiv wird, ist ihr Routing-Baum bereits beschnitten.

Die Zeit, die das Routing braucht, um Gruppenaustritte für einen Routing-Baum zu verarbeiten und den Baum daraufhin entsprechend anzupassen, wird hier also dadurch verdeckt, daß während ihrer Dauer keine Pakete vom Sender auf der betreffenden Gruppen-Adresse gesendet werden. Bis dort wieder gesendet wird, ist der Routing-Baum an die Änderungen angeglichen.

Nach jedem Entscheidungspunkt wird der Routing-Baum für die danach aktive Multicast-Gruppe eines Sub-Layers teilweise neu aufgebaut, wobei nur zu jenen Empfängern Äste führen, die dies explizit durch einen Gruppen-Eintritt am Entscheidungspunkt angefordert haben. Man nutzt hier die Eigenschaft des Multicast-Routing aus, daß es auf die Bearbeitung von Gruppen-Eintritten hin optimiert wurde. Diese führen innerhalb weniger Millisekunden zur Anpassung des Routing-Baumes [Cla97, DrH97]<sup>5</sup>. Für Gruppen-Austritte entsteht dagegen die schon mehrfach beschriebene Verzögerung mit all ihren negativen Nebenwirkungen.

Zwischen benachbarten Layern führt der Sender eine Verschachtelung der Sub-Layer-Wechsel durch. Auf jedem Layer mit ungeradem Index sind die Sub-Layer-Wechselpunkte um ein halbes Sendefenster versetzt. Dadurch kann die Reaktion auf heftige Stausituationen, für deren Beseitigung der Austritt aus dem obersten Layer alleine nicht ausreicht, weiter optimiert werden. Abbildung 6 zeigt die Vorteile der versetzten Anordnung von Entscheidungspunkten (unten) im Vergleich zu einer möglichen Anordnung, bei der die Entscheidungspunkte auf jedem Layer gleichzeitig vorkommen (oben). Zusätzlich sorgt diese Verschachtelung noch für die gleichmäßigere Verteilung der Gruppenein- und -austrittsmeldungen auf dem Netz.

## 4 Bewertung des Sub-Layer-Hoppings

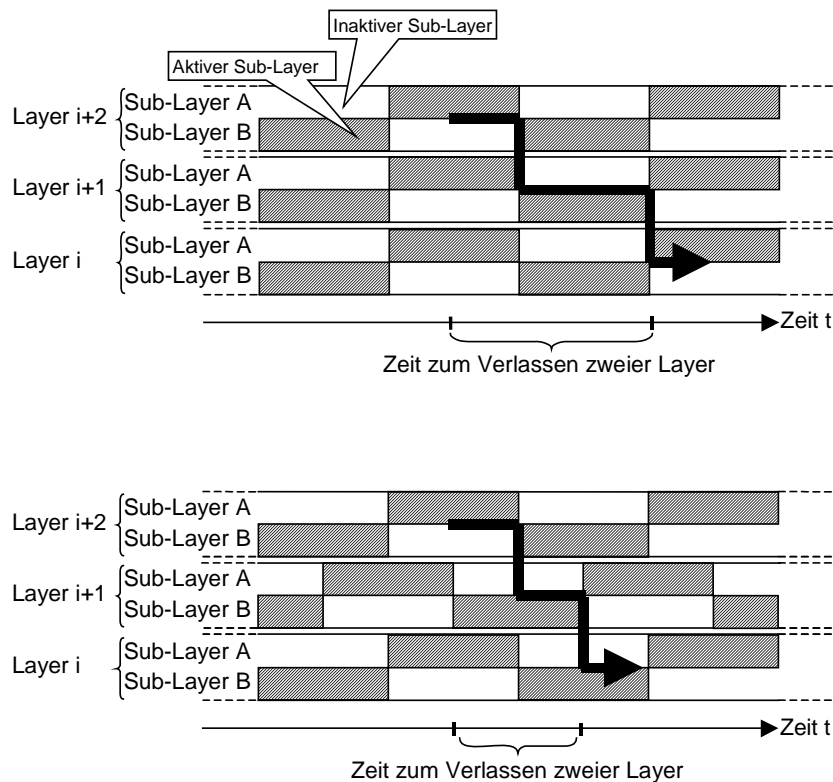
Zum Nachweis der Leistungsfähigkeit des Sub-Layer-Hoppings wurde das Protokoll zum einen einer analytischen Untersuchung unterzogen, zum anderen implementiert und vermessen. Die Staukontrolle ohne SLH wird auch hier wieder als „Basis-Staukontrolle“ bezeichnet. Die Staukontrolle mit SLH heißt entsprechend „erweiterte Staukontrolle“. Es ist unmittelbar einsichtig, daß die TCP-Freundlichkeit durch Einführung des Sub-Layer-Hoppings erhalten bleibt.

### 4.1 Reaktionszeit und Paketverluste

Der wesentliche Unterschied von SLH gegenüber dem herkömmlichen Layered Multicast ergibt sich beim Verringern der Empfangsrate. Für diese Aufgabe, die zum Zwecke der Staubehandlung durchgeführt wird, muß bei SLH nicht explizit ein langsamer Gruppen-Austritt angestoßen werden. Es genügt, beim nächsten Entscheidungspunkt keinen Gruppenbeitritt auszuführen. Verzögerungen ergeben sich hier nur noch durch eine eventuelle Reaktionszeit der Routing-Mechanismen. Während man es beim Verlassen eines Layers nicht mehr mit der Austritts-Verzögerung des Routing zu tun, muß man aber auf den nächsten Entscheidungspunkt warten.

Im schlechtesten Fall beträgt diese Wartezeit auch wieder 2 Sekunden. Dieser Fall tritt ein, falls Stau ganz kurz nach einem Entscheidungspunkt auftritt und man dann den gesamten Abstand zwischen den Entscheidungspunkten als Wartezeit hat. Im besten Fall tritt der Stau kurz vor dem nächsten Entscheidungspunkt auf. Dann bekommt man eine sehr kurze Verzögerung nahe bei Null. Im statistischen Mittel sollte man schließlich eine Wartezeit von einer Sekunde erwarten. Das bedeutet, daß sich die Reaktionszeit der Staukontrolle auf die Hälfte reduziert hat.

<sup>5</sup> Es wurde gezeigt, daß bei Tests mit Endsystemen in Karlsruhe, Braunschweig und Hannover (TTL-Entfernungen von 1, 18 und 38) die mittlere Join-Zeit zwischen 27 ms auf kürzeren Strecken und 54 ms auf längeren Strecken – mit starker Streuung – betrug. Dabei lag der jeweils letzte Gruppenaustritt (Leave) in dieser Gruppe zwischen 0 und 6 sec zurück. Die maximal gemessenen Zeiten für einen Join lagen bei ca. 400 ms.



**Abbildung 6: Beschleunigung kaskadierter Austritte**

Durch diese Vorgehensweise verfügt das Layered-Multicast mit Sub-Layer-Hopping im Vergleich zum Layered-Multicast ohne Sub-Layer-Hopping über folgende Vorteile:

- Für die Staukontrolle ergibt sich eine bessere Reaktionszeit. Daraus kann eine wirksamere Staubeseitigung resultieren, die zu einem gesteigerten Durchsatz bei der Multicast-Übertragung führt.
- Es können keine scheinbaren Paketverluste auftreten. Das liegt daran, daß die Staubehandlung hier nicht dadurch ausgelöst wird, daß nach Auftritt eines Staus sofort ein Gruppen-Austritt angestoßen wird. Stattdessen wartet man bis zum nächsten Entscheidungspunkt, wo dann zur Staubehandlung kein Gruppen-Eintritt durchgeführt wird. Dazwischen werden alle Pakete, die trotz des Staus den Empfänger noch erreichen, auch zugestellt.

## 4.2 Routerbelastung

In diesem Abschnitt wird die Belastung der Router im Rückkanal, welcher bei SLH (und allgemein bei Layered Multicast) im Netzzinnern durch ein Routingprotokoll realisiert wird, auf Basis der Gruppen-Mitgliedschaftswechsel untersucht.

### 4.2.1 Anzahl von Gruppen-Mitgliedschaftswechseln

Gruppenwechsel entstehen dadurch, daß jeder Empfänger für jeden Layer, auf dem er Sub-Layer-Hopping benutzt, beim Erreichen eines Entscheidungspunktes einen Gruppeneintritt und einen Gruppenaustritt durchführt.

Ist davor Stau aufgetreten, dann entfällt der Gruppeneintritt. Auf diesen Fall hin ist das Sub-Layer-Hopping optimiert, hier kommt es nicht zu unnötigen Gruppenwechseln. Entsteht jedoch zwischen aufeinanderfolgenden Sub-Layerwechseln kein Stau, dann entsteht ein gegenüber dem herkömmlichen Layered Multicast zusätzlicher Aus- und Eintritt. Dabei führt jeder Empfänger das Sub-Layer-Hopping nur für seine obersten beiden Layer durch. Dadurch können bei auftretendem Stau diese beiden Layer schnell verlassen werden. Auf den Layern ohne Sub-Layer-Hopping wird der Empfänger einmal Mitglied beider Multicast-Gruppen, an den Entscheidungspunkten finden dann keine Aktionen statt.

## 4.2.2 Lokale Ebene

Auf lokaler Ebene werden — unabhängig von der Anzahl der Layer — innerhalb von 2 Sekunden von jedem Empfänger zwei Gruppen-Austritte und zwei Gruppen-Eintritte durchgeführt. Diese Aktionen sind jeweils an einem Entscheidungspunkt konzentriert, finden also in unmittelbarer zeitlicher Nähe statt, wodurch eine deutlich effizientere Verarbeitung durch IGMP möglich ist.

Für den lokalen Bereich mit  $N_E$  Empfängern entstehen also alle 2 Sekunden  $2N_E$  Gruppen-Eintritts-Meldungen und genau zwei weitere Gruppen-Austritts-Meldungen. Jeder einzelne Empfänger verursacht im Durchschnitt demnach  $1+(1/N_E)$  Kontrollmeldungen pro Sekunde. Da sich diese zeitlich auf die Entscheidungspunkte konzentrieren, wird ein Großteil der Kontrollmeldungen wegen der Redundanzunterdrückung in IGMP gar nicht gesendet.

Zum Vergleich sei an das Rückmelde-Aufkommen einer TCP-Verbindung erinnert. Tritt hierbei gerade kein Stau auf, entstehen beispielsweise RTT-Werte zwischen 10 Millisekunden und 50 Millisekunden. Es wird dann zwischen 20 und 100 Mal pro Sekunde eine Rückmeldung in Form eines Acknowledgement gesendet. Da die TCP-Rückmeldungen bis zum Sender fließen, betrifft dieses Argument auch den Verarbeitungsaufwand der Router, an denen Multicast-Empfänger nicht direkt angeschlossen sind.

## 4.2.3 Router-Ebene

Zwischen den Routern entlang des Routing-Baumes sind Gruppenein- und -austritte gleichbedeutend mit dem Austausch von Graft- und Prune-Meldungen.

Den größten Mehraufwand verursacht SLH, falls im Netz gerade keine Laständerungen beobachtet werden können, auf die schnell mit der Staubehandlung reagiert werden müßte. SLH führt dann trotzdem seine periodischen Layer-Wechsel durch, die immer mit Gruppen-Ein- und -austritten verbunden sind.

Die Annahme, unter der die folgenden Betrachtungen durchgeführt werden, stellt für SLH einen *Worst Case* dar: Die Datenübertragung befinde sich in einem stabilen Zustand, in dem keine Laständerungen im Netz zu plötzlichen Stausituationen führen.

Erfahrungen aus eigenen Tests sowie Berichte in [CrV97, CRV98] zeigen, daß die Empfänger dann zwischen genau zwei Sessions oszillieren. Sie heißen im folgenden  $S_m$  und  $S_{m+1}$ . Die maximal mögliche Empfangsrate ist in Session  $S_m$  erreicht, gemäß den Aufgaben der Staukontrolle führt der Empfänger aber in periodischen Abständen einen Wechsel in  $S_{m+1}$  durch, um sicherzustellen, daß keine ungenutzte Bandbreite zur Verfügung steht. Dadurch ergibt sich folgender Ablauf:

1. Der Empfang erfolgt auf  $S_m$ .
2. Nach  $t_p(m)$  findet ein Wechsel nach  $S_{m+1}$  statt.
3. Die Empfangsrate ist nun zu hoch und es entsteht nach  $t_s$  Sekunden Stau.
4. Von  $S_{m+1}$  wird zurück zu  $S_m$  gewechselt; der Ablauf beginnt wieder bei Punkt 1.

Über die Zeitspanne  $t_p(m)$  wird die Proportionalität zwischen momentanem Durchsatz und Zeit bis zu dessen Verdopplung herbeigeführt. Dazu kann eine Proportionalitätskonstante  $C$  derart festgelegt werden, daß gilt:

$$t_p(m) = bw(S_m) C \quad (1)$$

Dabei ist  $bw(S_m) = 2^m b_0$  mit der Einheit Pakete/Sekunde und  $b_0$  die Basisdatenrate auf den untersten beiden Layern. Für  $C$  ist eine Analogie zum Quadrat der Round-Trip-Time bei TCP angedacht. Da die niedrigste Datenrate, die einer TCP-Übertragung zur Verfügung steht, aus einem Paket pro RTT besteht und  $b_0$  die niedrigste Rate der Multicast-Übertragung ist, kann für  $C$  etwa folgende Festlegung getroffen werden:

$$C = \frac{1}{b_0^2} \quad (2)$$



Aus (1) und (2) folgt zusammen mit der Definition von  $bw$ :

$$t_p(m) = \frac{2^m}{b_0} \quad (3)$$

Für  $t_s$  muß noch beachtet werden, daß es kleiner sein muß als die Zeit, die auf  $S_{m+1}$  gewartet wird, bis ein Wechsel in  $S_{m+2}$  erfolgt. Daraus ergibt sich für  $t_s$  folgende Bedingung:

$$t_s < t_p(m+1) \Leftrightarrow t_s < \frac{2^{m+1}}{b_0} \quad (4)$$

Mit diesen Größen kann der Verarbeitungsaufwand abgeschätzt werden, der während der Multicast-Übertragung mit den genannten Voraussetzungen in den Routern entlang des Routing-Baumes entsteht. Zur Vereinfachung der folgenden Betrachtungen wird angenommen, daß die Weiterleitung eines Daten-Pakets und die Verarbeitung einer Kontroll-Meldung den gleichen Aufwand für den Router verursachen. Diese Annahme kann für Software-Multicastrouter, die im heutigen MBone noch häufig anzutreffen sind, getroffen werden. Eine genauere Abschätzung zum Verhältnis zwischen dem Aufwand zur Weiterleitung eines Multicast-Pakets und jenem zur Bearbeitung einer Routing-Nachricht lag nicht vor.

Im ersten Schritt werden sowohl für die Basis-Staukontrolle als auch für die erweiterte Staukontrolle Formeln hergeleitet, welche die Anzahl der Verarbeitungsaufträge pro Sekunde angeben, die durch das Wechseln zwischen  $S_m$  und  $S_{m+1}$  in den Routern entstehen. Danach wird das Verhältnis dieser Ausdrücke untersucht.

Für die Basis-Staukontrolle entsteht ein Ablauf, wie er in Abschnitt 2 beschrieben wurde:

1. Wechsel in  $S_{m+1}$  zum Zeitpunkt 0
2. Stau nach  $t_s$ ; durch sofortigen Gruppenaustritt wird das Zurückwechseln in  $S_m$  eingeleitet
3. 2 Sekunden warten, bis nach der Austritts-Verzögerung der Gruppenaustritt wirksam wird<sup>6</sup>.
4. Nach  $t_p(m)$  Sekunden erfolgt der nächste Versuch

Für den Verarbeitungs-Aufwand in den Routern entlang des Routing-Baumes ergibt sich:

- Dauer:  $t_s + 2 + t_p(m) = t_s + 2 + 2^m/b_0$  [Sekunden]
- Kontrollmeldungen: insgesamt 2 (ein Gruppen-Eintritt und ein Gruppen-Austritt)
- Paket-Weiterleitungen:  
 $bw(L_{m+1})(t_s + 2) + bw(S_m)(t_s + 2 + 2^m/b_0) = b_0 2^{m+1}(t_s + 2) + 2^{2m}$  [Pakete]
- Durch den sofortigen Layer-Austritt kommt es zu bis zu  $bw(L_{m+1})2 = b_0 2^{m+1}$  scheinbaren Paketverlusten, die spätere Übertragungswiederholungen verursachen. Für die vorliegende Worst Case-Analyse wurde der für SLH ungünstigste Fall angenommen, daß bei der Basis-Staukontrolle keine scheinbaren Paketverluste auftreten.

Insgesamt verursacht die Basis-Staukontrolle durch den betrachteten Ablauf in den Routern  $V_B(t_s)$  Verarbeitungs-Schritte pro Sekunde, mit

$$V_B(t_s) = \frac{2 + b_0 2^{m+1} (t_s + 2) + 2^{2m}}{\left( t_s + 2 + \frac{2^m}{b_0} \right)} \quad (5)$$

Entsprechend ergibt sich für die erweiterte Staukontrolle mit Sub-Layer-Hopping:

1. Wechsel in  $S_{m+1}$  zum Zeitpunkt 0
2. Stau nach  $t_s$  Sekunden
3. Wechsel zurück nach  $S_m$  nach  $\lceil t_s/2 \rceil 2$  Sekunden

Damit ergibt sich:

- Dauer:  $\lceil t_s/2 \rceil 2 + t_p(m) = \lceil t_s/2 \rceil 2 + 2^m/b_0$  [Sekunden]
- Kontrollmeldungen: 2 pro Sekunde

<sup>6</sup> Eine solche Wartezeit wird zum Beispiel in [CRV98] unter der Bezeichnung „*deaf-period*“ eingeführt. Sie soll das Protokoll stabiler machen, indem kaskadierte Layer-Austritte als Reaktion auf den gleichen Stau verhindert werden.

- Paket-Weiterleitungen:  $bw(L_{m+1}) 2 \lceil t_s/2 \rceil + bw(S_m)(2 \lceil t_s/2 \rceil + 2^m/b_0) = b_0 2^{m+2} \lceil t_s/2 \rceil + 2^{2m}$  [Pakete]  
Insgesamt werden dadurch pro Sekunde  $V_E(t_s)$  Verarbeitungsschritte in den Routern ausgelöst:

$$V_E(t_s) = \frac{b_0 2^{m+2} \left\lceil \frac{t_s}{2} \right\rceil + 2^{2m}}{2 \left\lceil \frac{t_s}{2} \right\rceil + \frac{2^m}{b_0}} + 2 \quad (6)$$

Mit diesen Formeln ist nun ein Vergleich der beiden Verfahren möglich. Wir haben dazu das Verhältnis

$$\frac{V_B(t_s)}{V_E(t_s)} \quad (7)$$

gebildet, wobei für  $V_B(t_s)$  die Abschätzung nach unten gewählt wurde, und für unterschiedliche Werte von  $m$  und  $t_s$  aufgetragen (Abbildung 7). Für Ergebnisse kleiner 1 geht von der Basis-Staukontrolle ohne Sub-Layer-Hopping eine geringere Router-Belastung aus. Für Werte größer 1 ist dagegen die erweiterte Staukontrolle mit Sub-Layer-Hopping im Vorteil.

Die Wahl von  $b_0$  fiel immer auf den Wert 2. Dadurch stehen einerseits Sessions mit sehr geringer Rate zur Verfügung, wodurch die Empfangsrate zur wirksamen Staubebehandlung deutlich gesenkt werden kann, andererseits hält sich die Zahl der insgesamt benötigten Layer in Grenzen.

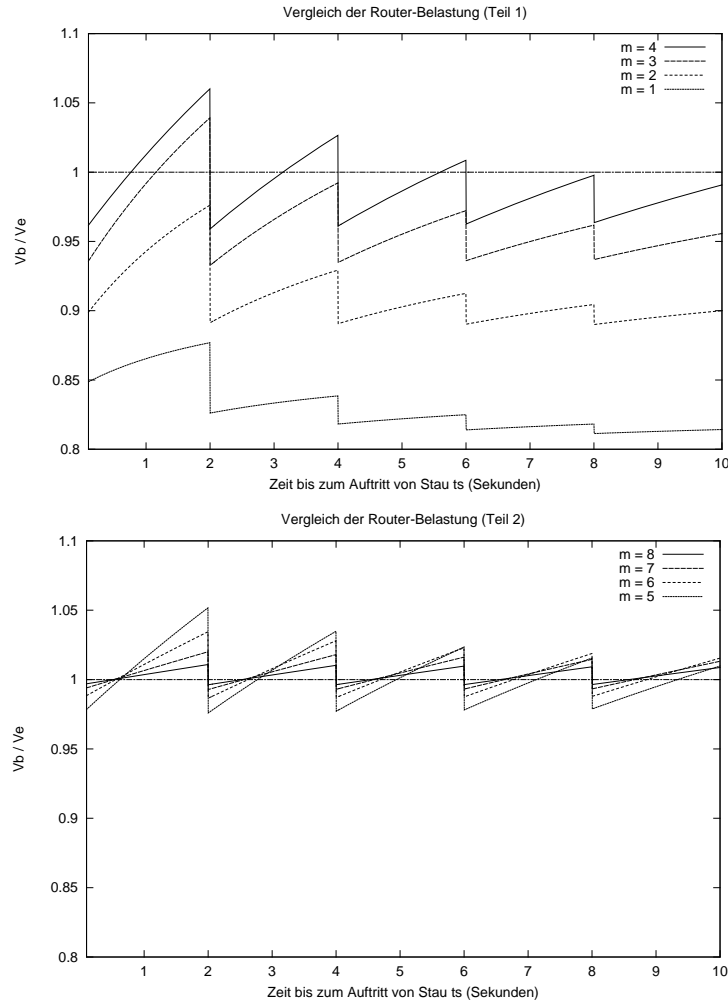


Abbildung 7: Ergebnisse der Analyse der Router-Belastung

Bei den gängigen Paketgrößen von 128, 256, 512 Byte kommen bis zum Erreichen der häufig auf maximal 500 kbit/sec festgesetzten Multicast-Übertragungsrate zwischen 7 und 9 Layer zum Einsatz. Der Parameter  $m$  wird dementsprechend höchstens mit 8 belegt. Die Kurven für die einzelnen  $m$ -Werte wurden zur besseren Übersichtlichkeit auf zwei Grafiken verteilt. Für  $t_S$  gilt weiterhin die Bedingung (4). Deshalb finden manche Kurven nur im Bereich von 0 bis zu dem durch (4) gegebenen Wert eine Entsprechung im realen System.

An den Kurven fallen besonders die Stufen auf, die jedesmal auftreten, wo der Wert für  $t_S$  wieder ein Vielfaches von 2 übersteigt. Das bedeutet für die erweiterte Staukontrolle jedesmal die Fortsetzung des Empfangs über einen weiteren Entscheidungspunkt hinaus, für höchstens 2 Sekunden. Man erkennt auch deutlich die Auswirkung der konstanten Gruppenein- und -austrittsrate der erweiterten Staukontrolle durch das Sub-Layer-Hopping. Sie ist dafür verantwortlich, daß mit zunehmendem  $t_S$  (also seltener auftretenden Staus) der untersuchte Quotient kleiner wird. Gerade bei Übertragungen mit mehr als vier Layern ist der Unterschied zwischen den beiden Verfahren gering.

#### 4.3 Nachweis der Durchsatzhöhung durch Tests

Um die Leistungsfähigkeit des Verfahrens ermitteln zu können, wurden in der Programmiersprache C unter Digital UNIX 4.0 und Sun Solaris 2.5.1 und 2.6 sowohl die Basis-Staukontrolle aus [CRV98] realisiert als auch die erweiterte Staukontrolle mittels SLH. Anschließend wurden Tests im weltweiten Mbone durchgeführt mit Rechnern in Braunschweig, Mannheim, Karlsruhe und bei Eurecom in Sophia Antipolis, Südfrankreich. Einzelne Tests wurden außerdem auf den Strecken nach Amherst, Montreal und London durchgeführt. Die beiden Staukontroll-Protokolle traten jeweils abwechselnd gegeneinander an. Durch die zeitliche Nähe und wiederholte Durchführung der Messungen wurde sichergestellt, daß beide Protokolle unter den gleichen Rahmenbedingungen im Netz operieren konnten. Die Tests wurden an verschiedenen Tagen und zu unterschiedlichen Tageszeiten wiederholt. Durch geeignete Wahl der zu übertragenden Datenmenge lag die Dauer der Testläufe stets im Bereich einiger Minuten bis zu einer Viertelstunde.

Übertragen wurden leere Pakete der Größe 512 Byte. Paketgröße, Anzahl der Layer und Basisrate auf Layer 0 sind konfigurierbar. Die Steigerung der Layer-Raten wurde auf den Faktor 2 festgelegt. Informationen über die Rate werden in jedem Sende-Thread durch die Variablen `sndwnd_time` und `sndwnd_packets` repräsentiert. In einem Sendefenster der Länge `sndwnd_time` Millisekunden werden `sndwnd_packets` Pakete übertragen. Wegen Beschränkungen der Timer-Auflösung auf 10ms werden dabei jeweils `sndwnd_packets` Pakete als Burst gesendet.

Durch Mitführen einer Bitmaske in jedem Paket des Layer 0 ist auf der Seite des Empfängers garantiert, daß ständig ein Paketstrom empfangen wird, der durch die `sublayer_mask` darüber informiert, welcher Sub-Layer auf jedem Layer gerade aktiv ist.

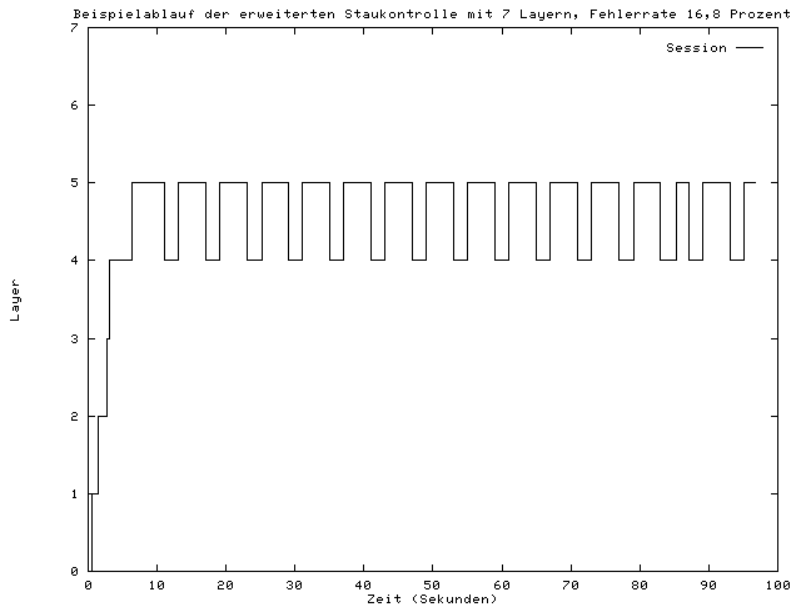
Bei den Tests standen drei Fragestellungen im Mittelpunkt:

1. Erreicht das Protokoll einen stabilen Arbeitspunkt?
2. Wie verbessert sich die Reaktionszeit durch SLH?
3. Wie verbessert sich der Durchsatz?

*Stabilität des Arbeitspunktes:* Die Stabilität der SLH-Implementierung zeigt sich vor allem dort, wo das Protokoll einen hohen Layer erreicht und halten kann. Bei einer hohen Empfangsrate kann dieser Zustand nur dann stabil sein, falls der Empfänger sich richtig auf die Sub-Layer-Wechselpunkte synchronisiert hat und dort keine Paketverluste durch verspätete Reaktionen entstehen. Diese Verluste würde der Empfänger zwangsläufig als Stausignale interpretieren. Ein stabiler Zustand könnte dann nicht gehalten werden.

Abbildung 8 zeigt das Ergebnis eines Testlaufs, bei dem die SLH-Staukontrolle unter den genannten Bedingungen einen stabilen Zustand auf Layer 4 sicher erreicht und dann halten kann. Der Empfänger prüft in periodischen Abständen auf ungenutzte Bandbreite, indem er den Layer 5 hinzufügt. Dies übersteigt jedesmal die maximale Empfangsrate, und es findet wieder ein Rücksprung auf Layer 4 statt. Da Layer 5 nicht sofort beim ersten Paketverlust verlassen wird, kann es wie in dieser Abbildung durchaus vorkommen, daß die Staukontrolle länger auf Layer 5 bleibt als auf Layer 4.

Dieses Beispiel stellt genau die Situation dar, die in der vorangegangenen Analyse als Voraussetzung der Betrachtungen diente. Die Präsentation dieses Ablaufs anhand von Testergebnissen beweist, daß die Analyse auf einer realistischen Annahme beruhte. Anhand dieser Meßwerte kann sogar ein Wertebereich für die Zeit  $t_S$  angegeben werden, die in den Ergebnissen der Analyse als Variable enthalten war. Sie betrug hier im Mittel zwischen 2 und 3 Sekunden.



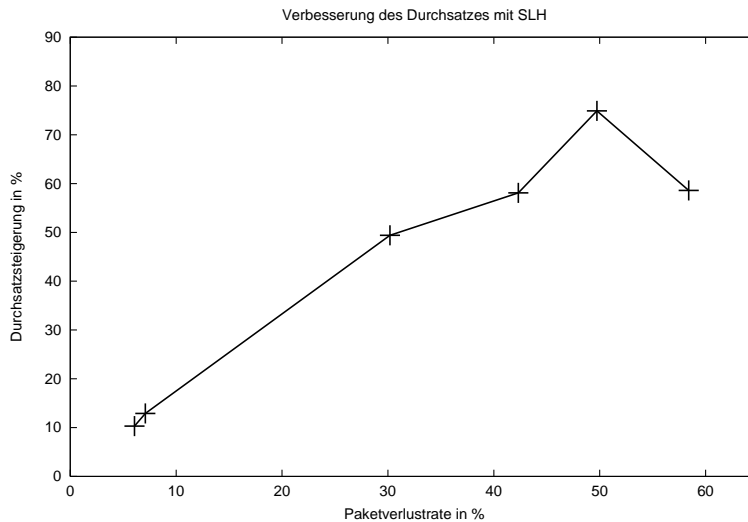
**Abbildung 8: Die SLH-Staukontrolle erreicht einen stabilen Zustand**

*Reaktionszeit:* Für jeden Austritt aus einem Layer wurde die Zeit berechnet, die zwischen dem Auftritt des Staus und dem Erreichen des nächsten Entscheidungspunktes verging. Von diesen Werten wurde anschließend der Mittelwert, die Varianz und des Minimum berechnet. Die Ergebnisse sind in der folgenden Tabelle dargestellt:

Austrittsverzögerung bei SLH		
Mittelwert	Standardabweichung	Minimum
970 msec	390 msec	31 msec

Die Messungen bestätigen also, daß sich die Austrittsverzögerung des Layered-Multicast durch SLH auf die Hälfte reduzieren läßt, mit geringer Standardabweichung.

*Durchsatzsteigerung:* Ergebnisse der vergleichenden Messungen zwischen der Basis-Staukontrolle und der erweiterten Staukontrolle sind in Abbildung 9 dargestellt. Grundlage sind 6 Testreihen mit insgesamt 70 Messungen. Pro Testreihe wurden dabei die Mittelwerte der Paketverlustraten und der Durchsatzsteigerungen ermittelt und gegeneinander aufgetragen.



**Abbildung 9: Vergleichsmessungen zwischen der Basis-Staukontrolle und der erweiterten Staukontrolle mit SLH**

Bereits bei niedrigen Paketverlustraten um 5 Prozent bringt SLH eine deutliche Steigerung des Durchsatzes. Mit zunehmender Anzahl von Stausituationen und der damit verbundenen Steigerung der Paketverlustrate nimmt die Verbesserung durch SLH noch zu. Bei Verlustraten um 50 Prozent sind zur Staubehandlung oft Gruppenaustritte über mehrere Layer hinweg in kurzer Zeit nötig. Hier erzielt SLH deshalb eine besonders hohe Durchsatzsteigerung gegenüber der Basis-Staukontrolle, im Test bis zu ca. 75%. Eine überproportionale Steigerung des Durchsatzes gegenüber der Fehlerrate (75% mehr Durchsatz bei 50% Paketverlustrate) ist deshalb möglich, weil bei SLH keine scheinbaren Paketverluste auftreten. Bei sehr hohen Verlustraten (und damit geringen Durchsatzraten) fällt der Effizienzgewinn auf einen geringeren Wert zurück, weil mitunter während der Zeit, in der das SLH-Verfahren noch Mitglied einer Multicast-Gruppe ist, während das Basisverfahren bereits ausgetreten ist, keine oder nur wenige Pakete gesendet werden und damit auch nur relativ wenige scheinbare Paketverluste auftreten.

#### 4.4 Zusammenfassung der Bewertung

Als Ergebnis der Analysen ergibt sich ein gutes Bild für die erweiterte Staukontrolle mit Sub-Layer-Hopping. Nur in Fällen, in denen die Empfänger lediglich den unteren Layern beigetreten sind, ergibt sich eine gewisse Mehrbelastung der Router durch die erweiterte gegenüber der Basis-Staukontrolle. Ansonsten sind die Unterschiede vernachlässigbar. Zwar gibt es in der erweiterten Staukontrolle mehr Prune- und Graft-Meldungen als ohne SLH, aber die dadurch schnellere Entlastung der Router von der Paket-Weiterleitung im Falle von Stau wiegt diesen Nachteil auf. Da diese Ergebnisse in einem Worst-Case-Szenario hergeleitet wurden, liegt die Situation für Verhältnisse, die den hier gemachten Annahmen nicht entsprechen, noch günstiger. Die Gegenüberstellung der Verfahren in Tests bestätigen, daß eine gemessen an der Fehlerrate überproportionale Steigerung der Durchsatzrate möglich ist.

## 5 Zusammenfassung

Die durch das herkömmliche Layered Multicast realisierte Staukontrolle ist nur eingeschränkt in der Lage, schnell auf Staus zu reagieren. Sub-Layer-Hopping reduziert die beim Layered Multicast sonst vorhandene, durch IGMP verursachte Austrittsverzögerung aus Layern, ohne Modifikationen an den Systemen zu erfordern. Außerdem werden scheinbare Paketverluste vermieden, welche die erzielbare Effizienz bei der Basisvariante des Layered Multicast unnötig verringern. Durch diese Maßnahmen wird die Effizienz des Layered Multicast erhöht, wie Analysen und Tests belegen. SLH reagiert schneller auf Stausituationen als das Basisverfahren und sorgt besonders bei hohen Paketverlustraten für eine Steigerung des Durchsatzes und damit im Mittel für eine kürzere Zeit, bis die übermittelten Nutzdaten bei den Empfängern vollständig vorliegen. Bei sich nicht stark ändernder Paketverlustrate wird ferner ein stabiler Arbeitspunkt erreicht.

Als weiterführende Arbeiten sind Messungen zur Multicast-Router-Belastung denkbar sowie zum Join-/Leave-Verhalten zwischen Multicast-Routern.

## 6 Danksagung

Den Testpartnern in Amherst, Braunschweig, Mannheim, Montreal, London und Sophia Antipolis sei an dieser Stelle recht herzlich für ihre Unterstützung gedankt.

## 7 Literatur

- [Cla97] Robert Clauß: *Fehlerbehebung in heterogenen Kommunikationsgruppen*. Diplomarbeit am Institut für Telematik, Universität Karlsruhe, Mai 1997.
- [CrV97] J. Crowcroft, L. Vicisano: *One to Many Reliable Bulk-Data Transfer in the Mbone*, Hipparch'97, Uppsala, Schweden, Juni 1997.
- [CRV98] Jon Crowcroft, Luigi Rizzo, Lorenzo Vicisano: *TCP-like congestion control for layered multicast data transfer*. IEEE Infocomm ,98, März 1998.
- [Dee89] S. Deering: *Host Extensions for IP Multicasting*, STD 5, Request for Comments (RFC) 1112, Internet Engineering Task Force (IETF), August 1989.
- [DPW88]S. Deering, C. Partridge, D. Waitzman: *Distance Vector Multicast Routing Protocol*. Internet Request for Comments RFC 1075, November 1988.

- [DrH97] Stefan Dresler, Markus Hofmann: Adaptive Error Correction in IP Multicast Communication. 6th Open Workshop on High Speed Networks, Stuttgart, 8.-9. Oktober 1997.
- [DrH99] Stefan Dresler, Markus Hofmann: Sub-Layer-Hopping: Optimierte Staukontrolle für IP-Multicast). Proceedings der KiVS ,99, Darmstadt, 2.-5. März 1999.
- [Fen98] Bill Fenner: *Internet Group Management Protocol, Version 2*. Request For Comments (RFC) 2236, Internet Engineering Task Force (IETF), November 1997.
- [FJM95] S. Floyd, V. Jacobson, S. McCanne, C.-G. Liu, L. Zhang: *A Reliable Multicast Framework for Light-weight Sessions and Application Level Framing*. Computer Communication Review, Vol. 25, No. 4, Proc. of ACM SIGCOMM'95, Boston, MA, USA, August 1995.
- [Hen97] Jean-Charles Henrion: *An Efficient Software Implementation of a Forward Error Correcting Code*. Proceedings of IDMS'97, Darmstadt, September 1997.
- [Hof96] Markus Hofmann: *A Generic Concept for Large-Scale Multicast*. International Zurich Seminar on Digital Communication, 21.-23. Februar 1996, Zürich, Schweiz, Hrsg.: B. Plattner, Lecture Notes in Computer Science, No. 1044, Springer Verlag, 1996.
- [Hof98] Markus Hofmann: *Skalierbare Multicast-Kommunikation in Weitverkehrsnetzen*. Dissertation an der Universität Karlsruhe, infix Verlag, Februar 1998.
- [Jac88] V. Jacobson: *Congestion Avoidance and Control*; Proc. of SIGCOMM'88, Stanford, CA, USA, August 1988.
- [Lan98] Hans-Peter Lang: *Unterstützung heterogener Gruppen durch hierarchisch codierte Teilströme*. Diplomarbeit am Institut für Telematik, Universität Karlsruhe, Mai 1998.
- [Lin82] J. H. van Lint: *Introduction to Coding Theory*. Springer Verlag, 1982.
- [MJV96] S. McCanne, V. Jacobson, M. Vetterli: *Receiver-Driven Layered Multicast*, SIGCOMM'96, Stanford, CA, USA, August 1996.
- [MRB98] A. Mankin, A. Romanow, S. Bradner, V. Paxson: *IETF Criteria for Evaluating Reliable Multicast Transport and Application Protocols*. Internet Request for Comments, RFC 2357, Juni 1998.
- [Mül98] Frank Müller: *Staukontrolle für heterogene Multicast-Gruppen im Internet*. Diplomarbeit am Institut für Telematik, Universität Karlsruhe, Juli 1998.
- [PSL97] S. Paul, K.K. Sabnani, J.C.-H. Lin, S. Bhattacharyya: *Reliable Multicast Transport Protocol (RMTP)*. IEEE Journal on Selected Areas in Communications, Vol. 15, No. 3, Seite 407-421, April 1997.
- [Pus98] T. Pusateri: *Distance Vector Multicast Routing Protocol*, Internet-Draft (work in progress), März 1998.
- [Riz97] Luigi Rizzo: *Effective Erasure Codes for Reliable Computer Communication Protocols*. Computer Communication Review, ACM Special Interest Group on Data Communication, Vol. 27, No. 2, April 1997.
- [Riz98] Luigi Rizzo: *Fast group management in IGMP*. Proceedings of Hipparch ,98, 15-16 Juni 1998, UCL, London, UK, außerdem auf Freenix ,98, 1998.
- [Vic97] Lorenzo Vicisano: *Notes on a Cumulative Layered Organization of Data Packets Across Multiple Streams With Variable Rate*. Technical Report, UCL, London, UK, 1997.