

Multicast-Protokolle und Algorithmen: Ein vergleichender Überblick

Andreas Fey, Maik Hofmann

I. EINLEITUNG

Unsere heutige Gesellschaft zeichnet sich durch das Sammeln, Verarbeiten und Speichern der Ressource Information aus. Eine wachsende Herausforderung bei dem Umgang, mit dieser Ressource besteht in der Übertragung, da sich der Umfang und die Anforderungen an Geschwindigkeit und Verfügbarkeit stetig erhöhen. So begnügte man sich vor 10 Jahren noch mit der zeitunkritischen Übertragung von Texten, während man heute Videos in Echtzeit, z.B. Live-Konferenzen mit Bild und Ton, abrufen möchte.

Als Haupttransportmedium hat sich das Internet etabliert, welches einen weltweiten Verbund von unterschiedlichsten lokalen Netzwerken (Domänen) darstellt, die über mehrere Datenleitungen mit besonders hohen Kapazitäten (Backbones) miteinander verbunden sind. Einheitliche Übertragungsprotokolle in Verbindung mit speziellen Netzwerkkomponenten wie Router und Gateways sorgen für den Austausch der Daten zwischen den Netzwerken. Bei der Vermittlung von Daten (Routing) innerhalb eines lokalen Netzwerks spricht man von „Intra-Domain Routing“, eine Vermittlung über lokalen Netzwerkgrenzen (Domänengrenzen) hinweg bezeichnet man als „Inter-Domain Routing“.

Möchte ein Rechner, der Client, nun Informationen aus diesem Netzwerk abrufen, verbindet er sich im Allgemeinen mit dem Server, der die Informationsquelle darstellt und fordert die gewünschten Daten an. Client und Server bilden die Endpunkte dieser Verbindung, die Übertragung findet direkt zwischen ihnen statt, wobei der Übertragungsweg durch das zwischenliegende Netz über die verschiedensten Netzwerkkomponenten führt. Die zwei Endpunkte charakterisieren diese Punkt zu Punkt – Übertragung, welche auch Unicast genannt wird.

Bei einer Kommunikation zwischen einem Server und einer Gruppe, die für das eingangs erwähnte Beispiel einer Live-Videokonferenz

benötigt wird, müssen die abgerufenen Informationen zu der gesamten Gruppe verteilt werden. Eine Realisierung über Unicast würde hier bedeuten, dass der Server zu jedem Mitglied der Gruppe eine Verbindung aufbauen und eine Kopie der Informationen übertragen müsste. Durch das Duplizieren der Daten entstünde eine sehr hohe Last für das Kommunikationsnetz.

Ein Ansatz diese Mehrpunktverbindung zu verbessern ist, statt mehrere gleicher Datenpakete soweit wie möglich nur eine Ausprägung der Pakete zu übertragen, und erst an den Stellen, wo sich der Weg zu den Empfängern teilt, zu duplizieren. Diesen Verbindungstyp bezeichnet man als Multicast, wobei hier wiederum zwei Topologien zu unterscheiden sind. Zum einen Intra-Domain Multicast, welches für lokale (flache) Netze optimiert wurde, zum anderen Inter-Domain Multicast, eine Lösung für Mehrpunktverbindungen über einen Verbund von lokalen Netzen, wie dem Internet, hinweg.

In dieser Ausarbeitung soll ein Überblick über Multicast-Protokolle und deren Algorithmen gegeben werden.

Der zweite Teil behandelt zunächst die Gruppenverwaltung im Internet, die von allen erwähnten Protokollen benutzt wird, um Gruppenmitglieder zu identifizieren und adressieren.

Im dritten Abschnitt wird auf die Algorithmen eingegangen, die die grundlegende Mechanismen zur Informationsübertragung darstellen.

Die Protokolle, die auf diesen Algorithmen aufsetzen und Multicast implementieren, werden im Abschnitt vier behandelt.

Im fünften und letzten Abschnitt soll eine Zusammenfassung dieser Arbeit gegeben werden.

II. GRUPPENVERWALTUNG IM INTERNET

Bei einer Multicast – Kommunikation werden die Daten nicht an einzelne Empfänger, sondern an Gruppen verteilt. Das Beitreten oder Verlassen von Teilnehmern zu dieser Gruppe muss, ebenso wie Informationen über weitere existierende Gruppen, durch ein Protokoll verwaltet werden. Im Internet wurde zu diesem Zweck das Internet Group Management Protocol (IGMP) in Verbindung mit dem dazugehörigen Protokoll zur Fehlerkontrolle (Internet Control Message Protokoll, ICMP), eingeführt. Durch die Implementation im IP Modul des Internet – Hosts oder Gateway – Rechners können Multicast – Router Informationen über die Gruppenzugehörigkeit der angeschlossenen Systeme erhalten.

Gruppenmitglieder werden über eine gemeinsame Klasse - D IP-Adresse (224.x.x.x) adressiert, sodass der Sender auf IP – Ebene weder die Menge der Empfänger noch deren Adressen kennen muss.

Die Mitgliedschaften in einer Gruppe können von einem Multicast – Router über zwei Nachrichtentypen kontrolliert werden: Queries und Reports. Zum einen werden periodisch so genannte „host membership queries“ an alle Systeme (Adresse 224.0.0.1) des Subnetzes gesendet, um Mitglieder von Gruppen im lokalen Netz zu entdecken. Die Systeme senden ihrerseits einen „host membership report“ als Antwort an den Multicast – Router, in dem sich ihre Gruppenzugehörigkeiten befinden. Wenn sich ein neues Mitglied der Gruppe anschließt, sendet es sofort einen „join-group“ - Report, ohne auf eine Query zu warten; beim Verlassen erfolgt analog dazu einen „leave-group“ – Report.

Diese periodischen Abfragen werden zur Vermeidung von erhöhtem Datenverkehr und Verwaltungsaufwand nur von einem einzigen Multicast – Router im entsprechenden Netz getätigt; die Auswahl erfolgt durch die Ermittlung des Multicast – Routers mit der niedrigsten IP – Adresse.

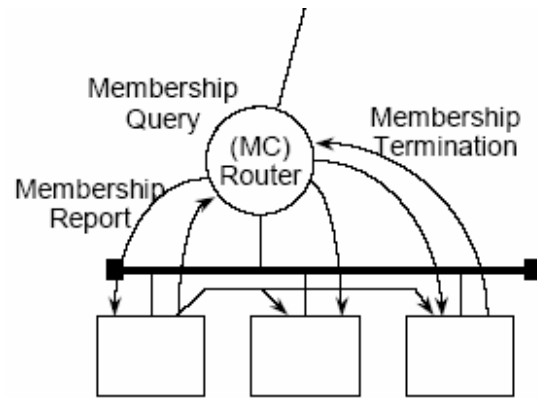


Abbildung 1: IGMP Reports und Queries

Da ein Mitlauschen der Gruppennachrichten durch ein System möglich ist, welches einfach keine Antwort auf eine Query sendet und daher unbekannt bleibt, ist IGMP nicht sicher und kann keine „bekannten“ Gruppen realisieren.

III. MULTICAST ROUTING – ALGORITHMEN

Da der Übertragungsweg einer Multicast-Verbindung ebenso wie der einer Unicast-Verbindung über zahlreiche Vermittlungsstellen führt, ist es notwendig einen optimalen Weg zwischen Kommunikationspartnern zu ermitteln.

Hierzu werden sowohl bei Uni- wie auch bei Multicast so genannte Verteilerbäume aufgebaut. Die Aufgabe der Routing-Algorithmen ist nun, einen optimierten Baum zu konstruieren. In diesen Kaptitel soll ein Überblick der Multicast Routing-Algorithmen gegeben werden. Bevor wir jedoch auf die Multicast Routing-Algorithmen im Einzelnen eingehen, sollen zunächst einige Grundlagen zum Routing erläutert werden.

Unicast Routing

Die Routing-Algorithmen lassen sich wie folgt untergliedern:

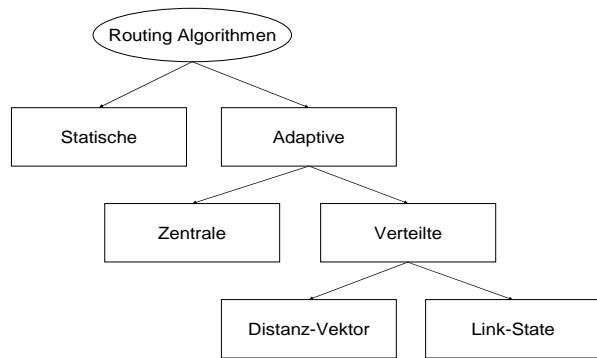


Abbildung 2: Gliederung der Routing-Algorithmen

1. Statische Routing-Algorithmen

Bei statischen Routing-Algorithmen wird die Routingtabelle bei Start des Systems initialisiert und ändert sich während des Betriebs nicht mehr. Diese Algorithmen sind daher nicht in der Lage, sich an ändernde Bedingungen (Verkehrstaus, Gruppenänderungen) im Netz anzupassen und sind somit nicht praxisrelevant.

2. Adaptive Routing-Algorithmen

Die adaptiven Routing-Algorithmen passen ihre Routinginformationen dynamisch an die aktuelle Netzsituation an und sind somit sowohl im Unicast wie auch im Multicastbereich für die Vermittlung geeignet.

Adaptive Routing-Algorithmen lassen sich nochmals wie folgt untergliedern:

2.1. Zentrale Routing-Algorithmen

Zentrale Routing-Algorithmen treffen ihre Routingentscheidung an einer Stelle im Netz und verteilen diese an die involvierten Router weiter. Diese zentrale Stelle besitzt die Information über das gesamte Netz und kann dadurch gute Entscheidungen treffen, jedoch ergeben sich durch das zentralisieren drei große Probleme:

- es ist sehr schwierig die Informationen über das Netz aktuell zu halten.
- der Speicher und Verarbeitungsaufwand der Zentrale in großen Netzen ist sehr hoch.
- fällt die Zentrale aus, bricht die gesamte Netzvermittlung zusammen (Single-Point-of-Failure).

2.2. Verteilte Routing-Algorithmen

Die heute eingesetzten Algorithmen sind in der Regel die Verteilten, hierbei trifft jeder Router mit dem ihm zur verfügungstehenden Informationen seine Routingentscheidungen selbst.

Diese Algorithmen sind auch für große Netze geeignet.

Hier kann wiederum in folgende zwei Verfahren untergliedert werden:

2.2.1. Distanz-Vektor Verfahren

Hierbei wird versucht den kürzesten Weg zwischen zwei Kommunikationspartnern zu finden, dazu ist es Notwendig, das jedes System im Netz die Distanz zu allen anderen kennt.

Die Distanzen tauschen benachbarte Router über Distanzvektoren miteinander aus.

Das Verfahren berücksichtigt jedoch nur die Entfernungen und nicht Qualität des jeweiligen Abschnitts wodurch schlechte Routingentscheidungen getroffen werden können.

Ein weiterer Nachteil ist, dass sich Veränderungen in der Netztopologie nur sehr langsam ausbreiten was wiederum dazu führt, das dieses Verfahren in großen Netzen nur Bedingt geeignet ist.

2.2.2 Link-State Verfahren

Bei den Link-State Algorithmus wird das Netz als gerichteter Graph betrachtet; der kürzeste Weg zum Zielsystem berechnet sich meist nach dem Dijkstra – Algorithmus, auch „Shortest Path First“ genannt. Diese Berechnung erfolgt individuell an jedem Router.

Jeder Router kennt seinen Nachbarn und beobachtet den dazwischen liegenden Übertragungsabschnitt durch periodisches Versenden von Dateneinheiten, auf diese Art können Netzänderungen schnell erkannt werden.

Außerdem kennt jeder Router die Topologie des Netzes und hält sie in einer so genannten Link-State-Database gespeichert.

Bei Änderungen der Netz-Topologie werden die Informationen im Netz geflutet.

Dadurch führt dieser Algorithmus zu einer schnelleren Konvergenz.

Multicast-Routing

Multicast-Routing soll nun nicht mehr nur ein einzelnen, sondern eine Gruppe von Empfängern erreichen. Dabei muss der Algorithmus berücksichtigen, dass sich die Gruppen verän-

den können. Dies kann je nach Anwendung sehr häufig eintreten und darum ist ein verteilter Ansatz unumgänglich, da ein monolithisches Verfahren bei jeder Änderung eine komplette Neuberechnung zur Folge hätte.

Die nachfolgenden Verfahren sind für die Konstruktion so genannter Verteilerbäume zuständig. Wobei sich diese wiederum in nachfolgende Gruppen untergliedern lassen.

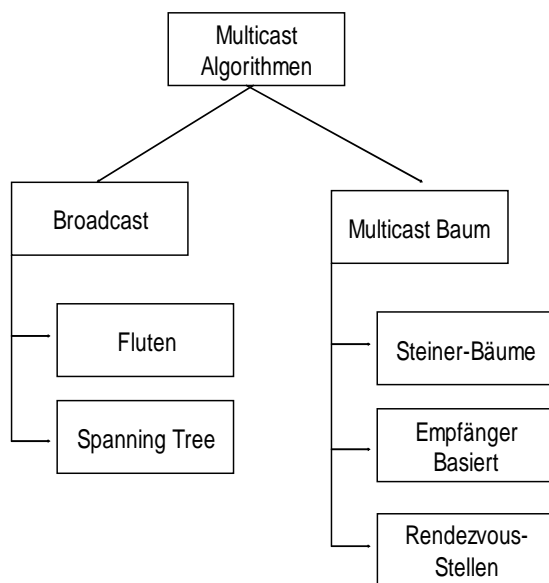


Abbildung 3: Gliederung der Multicast Algorithmen

1. Broadcast Verfahren

Bei diesen Verfahren werden die Datenpakete ohne Beachtung der Gruppenzusammensetzung, im kompletten Netz verteilt. Dadurch wird garantiert, dass jedes Gruppenmitglied die Datenpakete empfängt, was allerdings auch dazu führt, dass die Daten von allen weiteren Systemen erreicht werden, wodurch eine geschlossene Gruppe nicht realisierbar ist und dazu führt, dass unnötig Netzteile belastet werden, in denen sich keine Mitglieder befinden.

1.1. Fluten

Das einfachste Verfahren eine Verteilung der Daten an alle Gruppenmitglieder zu ermöglichen ist das Fluten. Dabei verteilt ein Router ein einkommendes Datenpaket einfach an alle

Schnittstellen mit Ausnahme derjenigen worüber die Daten empfangen wurden.

Es wird keine Auswertung der Gruppen vorgenommen wodurch eine unnötig hohe Netzlast entsteht und geschlossene Gruppen nicht zu realisieren sind, da jeder Router die Daten ungefiltert an alle Anliegenden Systeme verteilt. Ein weiteres Problem welches bei Fluten auftritt, sind so genannte Schleifen, welche entstehen wenn eine Dateneinheit ewig im Netz von einem Router zum anderen kreisen.

Um dem Problem der Schleifenbildung zu entgehen, werden heute die Dateneinheit mit einer begrenzten Lebenszeit (TTL) versehen. Eine weitere Möglichkeit ist es, zu prüfen das ein einkommende Dateneinheit zum ersten Mal empfangen wurde, was jedoch einen gewaltigen Verwaltungsaufwand nach sich zieht.

Fluten ist eigentlich ein Broadcast-Verfahren und ist für Multicast sehr ineffizient, dennoch wird es in einigen Protokollen wie z.B. DVMRP eingesetzt.

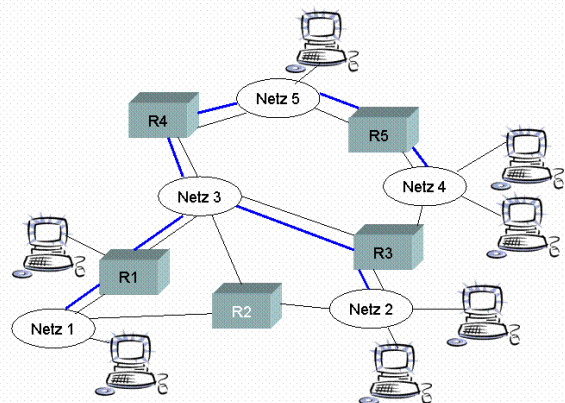


Abbildung 4: Spanning Tree

1.2. Spanning Tree Algorithmus

Der Spanning Tree-Algorithmus dient zur Vermeidung von Schleifen. Um dies zu ermöglichen wird ein Overlay-Netz aufgebaut, welches redundante Netzwerkpfade eliminiert.

Dieses Overlay-Netz bildet den Spanning Tree, welcher sich über das gesamte Netz legt und so jedes System erreicht. Ebenso wie bei dem Fluten, wird beim Spanning Tree Algorithmus keine Auswertung der Gruppen vorgenommen was wiederum dazu führt, dass keine geschlossenen Gruppen möglich sind.

Nachteile die der Spanning Tree mit sich bringt sind:

- die Verkehrskonzentration in der Wurzel und ihren Nachbarn
- die in vielen Fällen nicht optimale Pfadauswahl
- das die Wurzel ein „Single-Point-of-Failure“ darstellt.

Spanning Tree kann man als ein verbessertes Fluten werten, da hier die Schleifenbildung vermieden wird.

2. Multicast-Baum Verfahren

Die folgenden Verfahren berücksichtigen die Gruppenmitgliedschaften und bauen unter Auswertung selbiger einen so genannten Multicast-Baum auf.

2.1. Steiner-Bäume

Der Steiner-Baum ist ein monolithischer Algorithmus welcher einen global Kostenoptimierten Spanning-Tree aufbaut. Anders als beim Dijkstra wird die Optimierung global und nicht von einer Quelle ausgehend durchgeführt.

Da es sich jedoch um einen monolithischen Ansatz handelt, hat eine Gruppen- oder Topologieänderung eine komplette Neuberechnung zu Folge, was in einem großen Netz kaum tragbar ist. Außerdem kann der Steinernbaum nur in symmetrischen Netzen eingesetzt werden, da er alle Abschnitte als solche behandelt. Wegen dieser Restriktionen sind Steinerbäume nicht praxisrelevant.

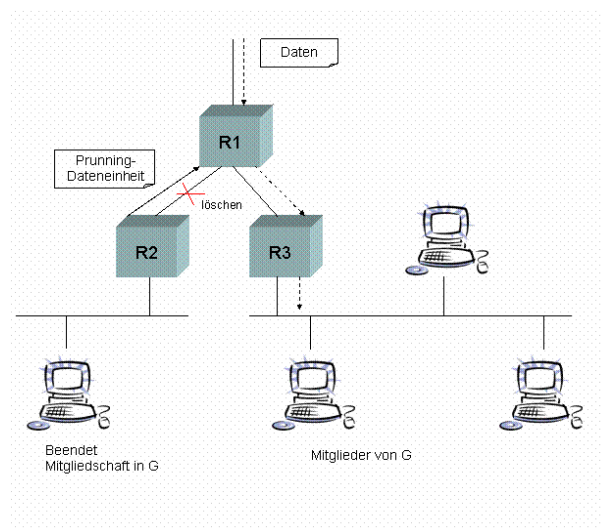


Abbildung 5: RPM

2.2. Empfängerbasiertes Routing

Bei den Empfängerbasierten Verfahren wird die Berechnung der Routing-Informationen vom Empfänger ausgelöst. Das Ergebnis der Berechnung ist ein nach Verzögerungszeiten optimierter Spanning Tree, welcher die Schleifenbildung verhindert. Das von Dala und Metcalf entwickelte Verfahren wurde hauptsächlich im Testfeld des Mbone eingesetzt, wobei Deering es durch Vermeidung von Datenduplikaten noch verbesserte. Die Basisversion des Verfahrens wurde unter dem Namen Reverse-Path-Forwarding veröffentlicht welches wie folgt funktioniert. Empfängt ein Router ein Datenpaket, bewertet er, ob über die Schnittstelle, auf der er es empfangen hat auch der kürzeste Weg zum Sender des Pakets ist. Wenn dies der Fall ist, leitet der Router das Paket an all seine Schnittstellen mit Ausnahme der weiter, auf der das Paket empfangen wurde. Ist dies nicht der Fall, wird das Datenpaket verworfen. Dadurch wird eine Schleifenbildung ausgeschlossen. Es ist eine Verbesserung zum normalen Fluten, da hier der Overhead stark gesenkt werden kann. Eine weitere Verbesserung wurde unter den Namen Reverse-Path-Broadcasting vorgeschlagen. Dabei wird, zusätzlich zu der Eingangsschnittstelle auch die Ausgangsschnittstellen bewertet. Eine Weiterleitung erfolgt nur, wenn der dort anliegende Router ebenfalls über diese Schnittstelle, den kürzesten Pfad zum Sender, findet. Um dies auch außer bei Link-State, wo jeder Router das komplette Netz kennt, auch beim Distanz-Vektor Algorithmus anwenden zu können, muss z.B. Poison-Reverse verwendet werden um die Pfade zu bewerten. Bei beiden Verfahren handelt es sich noch um Broadcasting, eine Auswertung der Gruppen wird nicht vorgenommen. Den dadurch entstehenden Nachteil, Netzteile zu überlasten in denen keine Gruppenmitglieder sind, sollte erst durch die Nachfolgenden Erweiterungen der Basisverfahren behoben werden. Unter den Namen Truncated RPB wurde erstmals die Auswertung der Gruppen unter Zuhilfenahme des IGMP vorgenommen. Ist in einem Subnetz kein Gruppenmitglied enthalten, werden auch keine Daten dorthin weitergeleitet. Die Router müssen dazu informiert sein, ob lokale Gruppenmitglieder vorhanden sind oder nicht, dies geschieht über das IGMP Protokoll. Da hierbei aber nur Subnetze ausgeschlossen werden, also trotzdem noch alle Teilnetze erreicht werden, handelt es sich immer noch um Broadcast, erst

mit dem folgenden Reverse Path Multicast wird erreicht, das Daten nur in Netzbereiche weitergeleitet werden, in den sich auch Gruppenmitglieder befinden. Um dies zu erreichen, läuft RPM wie folgt ab. Die erste Dateneinheit eines Senders wird im ganzen Netz verteilt, damit jedes potentiell Gruppenmitglied die Möglichkeit hat, dieses zu empfangen. Der vor dem Gruppenmitglied liegende Router wird als Blattrouter bezeichnet, er hat die Aufgabe zu erkennen ob lokale Gruppenmitglieder an seinen Schnittstellen vorhanden sind dies nicht der Fall, muß er den davor liegenden Router, mittels einer Dateneinheit (Pruning) diese Situation mitteilen. Diese Pruning Einheit wird mit einer Lebenszeit (TTL) von 1 ausgesendet um nur den nächsten Router zu erreichen. Hat der nachfolgende Router ein Pruning Paket erhalten, leitet er keine Daten mehr an den Sender des Pruning Pakets weiter. Jeder Router muss also die Zustandsinformation über Gruppen und Sender halten. Auf diese Art und Weise wird der Routing Baum schrittweise beschnitten bis nur noch Knoten bedient werden hinter denen sich Gruppenmitglieder befinden. Damit das Verfahren noch dynamisch arbeiten kann, werden die Zustandsinformationen mit begrenzter Lebenszeit versehen nach deren Ablauf der Baum, durch senden einer Dateneinheit in alle Teilsysteme, wieder neu aufgebaut wird. In kleineren Netzen ist diese Lösung durchaus praktikabel, jedoch bei größeren Netzen wird durch das periodische Neuaufbauen des Routingbaums, das Netz zu stark belastet. Ein weiteres Problem in großen Netzen ist die Notwendigkeit der Speicherung der Zustandsinformationen, welche bei großer Sender-, oder Gruppenanzahl nicht mehr realisierbar ist.

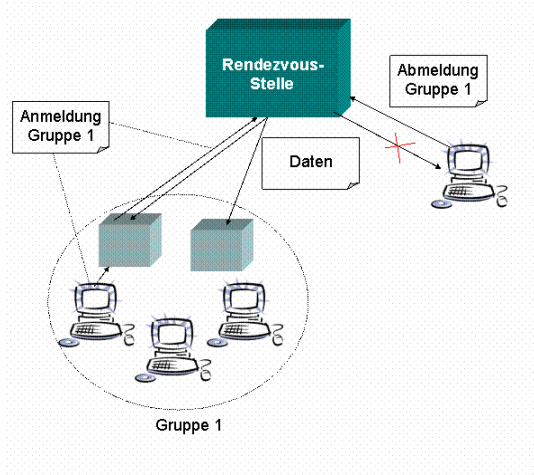


Abbildung 6: Rendezvous-Stellen

2.3. Bäume mit Rendezvous-Stellen

Bei diesen Verfahren werden im Netz so genannte Rendezvous-Stellen festgelegt, an denen sich die Gruppenmitglieder durch Versenden einer entsprechenden Dateneinheit anmelden. Bei dem Übertragen dieser Dateneinheit extrahieren die zwischen liegenden Router die Gruppenzugehörigkeit und leiten das Paket automatisch an die für diese Gruppe festgelegte Rendezvous-Stelle weiter. Es muss im Router also nur die Gruppenzugehörigkeit nachfolgender Empfänger gespeichert werden, was wiederum den Verwaltungsaufwand stark reduziert. Befindet sich hinter einem Router kein Gruppenmitglied, ist der gesamte Übertragungsabschnitt inaktiv. Die Router sind also lediglich für die Weiterleitung der Daten verantwortlich. Bei diesen Verfahren wird also ein Spanning Tree pro Gruppe etabliert. Dabei werden die Daten ausschließlich zu Gruppenmitgliedern versendet, wodurch geschlossene Gruppen möglich sind. Nachteile die die Rendezvous-Stellen mit sich bringen sind:

- die hohe Verkehrskonzentration an den RS
- das die RS ein Single-Point-of-Failure darstellen.

Außerdem besteht die neue Herausforderung darin, geeignete Rendezvous-Stellen im Netz festzulegen, was aber durch anwenden einfache Heuristiken bewältigt werden kann.

Zusammenfassung

Das Fluten stellt eine sehr einfache Methode dar, alle Mitglieder einer Gruppe mit den für sie bestimmten Informationen zu versorgen. Spanning Tree behebt das Problem der Schleifenbildung beim Fluten, kann also als eine Verbesserung gesehen werden. Beide Verfahren belasten aber das Netz sehr stark, da Teile ohne Mitglieder ebenfalls die Daten erhalten. Sie eignen sich, wenn in nahezu jedem Teilnetz, Mitglieder enthalten sind und die Gruppendichte hoch ist.

Die ersten Lösungen, welche den Charakter einer Multicastverbindung berücksichtigt, sind TRPB und RPM, da hier die Gruppenzusammensetzung zur Weiterleitung der Daten berücksichtigt wird. Da der Verwaltungsaufwand in den Routern für die Zustandshaltung relativ hoch ist und außerdem periodisch Pakete geflutet werden, sollten sie nur in kleineren Net-

zen eingesetzt werden, wo sich in nahezu allen Teilnetzen Gruppenmitglieder befinden.

Steinerbäume finden wegen ihrem monolithischen Ansatz und der damit verbundenen Statik bei Realisierung einer Multicastverbindung keinen praktischen Einsatz.

Rendezvous-Stellen sind für Weitverteilte Gruppen geeignet bei denen sich nicht in jeden Teilnetz Gruppenmitglieder befinden. Allerdings empfiehlt es sich, sie redundant einzusetzen, um der Verkehrskonzentration entgegen zu wirken und einen Ausfall einer Rendezvous-Stelle und den damit verbundenen Ausfall der Weiterleitung aller Datenpakete, vorzubeugen.

IV. MULTICAST ROUTING-PROTOKOLLE

In diesem Kapitel sollen die Routing-Protokolle für Multicast-Kommunikation erläutert werden. Allen Protokollen ist das Verwenden von IGMP für die Gruppenverwaltung gemeinsam. Die genaue Behandlung von Gruppenmitgliedern sowie die eingesetzten Routing-Algorithmen werden jedoch spezifisch in den einzelnen Protokollen geregelt. Untergliedern lassen sich Multicast Routing-Protokolle aufgrund ihrer zugrunde liegenden Netztopologie. Die früheren Protokolle basierten auf einer flachen Topologie, wie die einer abgeschlossenen Domäne („Intra-Domain Routing“). Die späteren Protokolle für hierarchische Topologien wie dem Internet ermöglichen dagegen die Gruppenkommunikation über Domänengrenzen hinweg („Inter-Domain Routing“).

1. Intra-Domain Routing Protokolle

Im Intra-Domain Routing Bereich ist es sinnvoll, die existierenden Protokolle für zwei Netztypen zu unterteilen: Protokolle für dicht besiedelte Netze („dense-mode“-Protokolle) und für dünn besiedelte Netze („sparse-mode“-Protokolle). Dense-mode Protokolle senden Multicast-Pakete an alle Router im Netz weiter, die selbst überprüfen müssen, ob sich unter ihnen Empfänger der adressierten Gruppe befinden. Sparse-mode Protokolle setzen dagegen voraus, dass sich die Empfänger an einer zentralen Stelle anmelden müssen, um die Informationen zu erhalten.

1.1. Dense-mode Protokolle

1.1.1. DVMRP

Das Distanz – Vektor Multicast – Routing Protokoll ist eines der ersten Protokolle, welches für Multicast entwickelt wurde. Als Basis wurde das für Unicast – Verbindungen stammende Routing Information Protokoll (RIP) benutzt. DVMRP wird eingesetzt, um den kürzesten Weg zur Datenquelle zu ermitteln. Im Gegensatz zu anderen Protokollen wird der Übertragungsabschnitt aber nicht zum Ziel hin, sondern zurück zur Quelle berechnet. DVMRP setzt daher auf die Pruning-Technik des RPM-Algorithmus auf und gehört daher zur Klasse der empfängerbasierten Protokolle.

Um DVMRP im Internet nutzen zu können, werden die Multicast-Pakete zwischen zwei DVMRP-fähigen Routern in Unicast-Pakete verpackt und so versendet. Diese Technik wird „Tunneling“ genannt.

Da die RPM-Technik für dicht besiedelte Netze mit vielen Gruppenmitgliedern geeignet ist, zählt DVMRP zu den so genannten „dense-mode“-Protokollen. Bei dünn besiedelten Netzen würde die Anzahl von Prune-Messages, die gesendet werden müssten, eine zu hohe Belastung und damit keinen Vorteil für das Netz darstellen

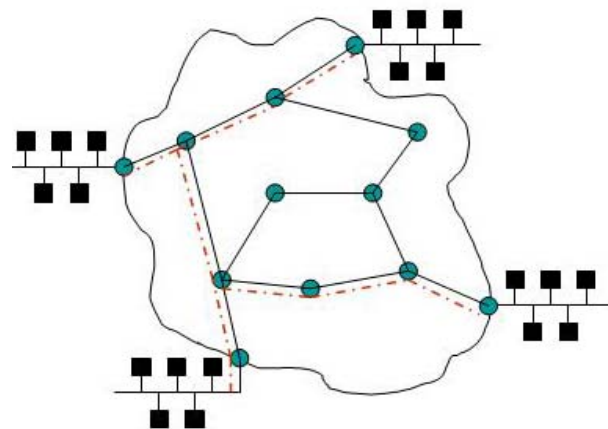


Abbildung 7: Tunnel-basiertes Routing

1.1.2. MOSPF

Das „Multicast Open Shortest Path First“-Protokoll stellt eine Erweiterung des Unicast OSPF-Protokolls dar. Es gehört zur Klasse der Link State Routing-Protokolle. Die Erweiterungen der Multicast – Fähigkeit des OSPF –

Protokolls bestehen im Wesentlichen aus zwei Teilen:

- Ergänzung der Link-State Datenbank durch die Beschreibung der Gruppenmitgliedschaft
- Für jedes Kommunikationspaar (Sender und Empfänger) muss ein eigener Multicast – Baum berechnet werden

Jeder Router hat nun eine lokale Gruppendatenbank zu führen, mit der er die Gruppenmitgliedschaft der angeschlossenen Netzwerke verwaltet. Das Vorhandensein der lokalen Gruppenmitglieder wird den anderen Routern mitgeteilt.

Auf dem gleichen Weg wie OSPF-Router unabhängig voneinander eine Unicast-Routing Topologie erzeugen, erfolgt das Routing der Multicast - Nachrichten bei MOSPF. Auf Basis der Sender- und Gruppenadresse wird ein Shortest Path Tree für je einen Sender und eine Gruppe erzeugt. Die Pakete können so in diesem Baum auf dem kürzesten Weg zum Ziel geroutet werden, wobei Gemeinsamkeiten der Pfade voll ausgenutzt werden, dass bedeutet die Duplizierung der Pakete erfolgt so spät wie möglich.

MOSPF ist zu OSPF voll abwärtskompatibel. Da Gruppeninformationen an alle MOSPF Router gesendet werden, gehört MOSPF zu den dense-mode Protokollen und wird den empfangerbasierten Verfahren zugeordnet, obwohl die Empfänger in den Gruppen selbst nur nach Anmeldung Daten empfangen.

Der Vorteil zu anderen Protokollen besteht darin, dass nur Gruppeninformationen zwischen den Routern verteilt werden, keine Nachrichten. Dies vermindert den Datenverkehr in größeren Netzen erheblich.

1.1.3. PIM Dense-Mode

Das Protocol independent multicast Protokoll untergliedert sich in zwei Teile: PIM Dense Mode und PIM Sparse Mode. Zunächst soll PIM-DM erwähnt werden.

PIM-DM ist DVMRP sehr ähnlich, es setzt auf die Technologie des Pruning. Der Unterschied zu DVMRP besteht darin, dass PIM unabhängig von dem eingesetzten Algorithmus und der davon erzeugten Routing-Tabelle arbeitet. Im Gegensatz dazu ist beispielsweise MOSPF auf den OSPF-Algorithmus festgelegt. Bei PIM-DM wird nur die Existenz einer Routing-Tabelle vorausgesetzt, egal von welchem Algorithmus diese erzeugt wurde.

Ein weiterer Unterschied zwischen DVMRP und PIM-DM betrifft den RPF – Test vor dem Weiterreichen von Paketen in einem Router; dieser entfällt bei PIM-DM. Die Pakete werden dadurch einfach an alle ausgehenden Interfaces eines Routers weitergeleitet, die Komplexität nimmt so stark ab. Jedoch werden auf diesem Weg auch unnötige Pakete an Router weitergegeben und viele Prune-Messages erzeugt. Das Fehlen des RPF-Tests kann also je nach Dichte des Netzes als Vor- oder Nachteil gewertet werden.

1.2. Sparse-mode Protokolle

1.2.1. PIM-SM

PIM Sparse-Mode (SM) wurde für Gruppenkommunikation, bei der Mitglieder weit voneinander entfernt und lose im Netz verteilt sind, entwickelt. Bezüglich der Freiheit zur Auswahl eines Algorithmus zum Aufbau der Routing-Tabelle gleicht es dem PIM-DM Protokoll, die Funktionsweise des PIM-SM ist jedoch grundlegend anders. Zunächst werden Gruppenbeitritte durch eine explizite Forderung (eine Join-Nachricht) von dem beitretenden Rechner gelöst. Durch diese expliziten Beitrittsmeldungen kann PIM-SM in die Kategorie der Bäume mit Rendezvous-Stellen eingeordnet werden. Des Weiteren wird die Verwaltung und Koordination von den Routern selbst auf einen zentralen Rendezvous-Punkt verlagert.

Der Ablauf der Paketvermittlung ist wie folgt:

- Der Beitritt eines Empfängers zu einer Gruppe wird durch eine IGMP-Nachricht veranlasst. Diese Nachricht wird vom Designated-Router (der Router mit der niedrigsten IP-Adresse) des Empfängers an den Rendezvous-Punkt geschickt, der für die Gruppe zuständig ist.
- Der Empfänger sendet diese Join-Nachrichten periodisch weiter, solange er Mitglied in einer Gruppe sein möchte. Falls nach einer bestimmten Zeit vom Router keine Join-Nachrichten mehr empfangen werden, meldet sich der Router, falls keine weiteren Abhängigkeiten existieren, mittels einer Pruning-Nachricht vom Multicast-Baum ab.
- Der Sender schickt die Daten per Unicast-Verbindung zu der Rendezvous-Stelle. Der Designated-Router des Senders kapselt die Multicast-Dateneinheiten in Unicast-Dateneinheiten, die an die Rendezvous-

Stelle geschickt werden. Dort werden die Multicast-Dateneinheiten entkapselt und per Multicast-Baum an die Empfänger der Gruppe verteilt.

- Der Rendezvous Punkt kann danach eine Join-Nachricht an den Sender schicken, um eine native Multicast-Verbindung mit diesem aufzubauen. Alle weiteren Pakete müssen nicht mehr in Unicast-Pakete verpackt werden, sondern können direkt vom RP weitergeleitet werden.
- Bei hohem Datenaufkommen kann PIM-SM von einem gemeinsamen Baum zu einem Speziellen schalten. Die Umschaltung kann nur von der Rendezvous-Stelle und den Routern der lokalen Gruppe veranlasst werden. Dies wird aber erst ab einer signifikanten Anzahl gesendeter Pakete gemacht, um den Verwaltungsaufwand für unnötige Umschaltungen zu vermeiden.

Ein Problem besteht in der Auswahl des Standortes für den Rendezvous Punkt. Das „Bootstrap Router Protokoll“ regelt diese Auswahl, um einen optimalen Router für diese Aufgabe auszuwählen.

Quittierungen finden bei PIM-SM generell nicht statt, da Dateneinheiten grundsätzlich periodisch gesendet werden.

1.2.2. CBT

Core based trees basieren ebenfalls auf dem Konzept der gemeinsamen Bäume mit Rendezvous – Punkten. Diese Bäume sind im Gegensatz zu denen von PIM-SM bidirektional, so dass sowohl von Sender zu Empfänger als auch von Empfänger zu Sender operiert werden kann. Durch quittieren einer Nachricht ist so kein periodisches Senden nötig. Als Sparse-Mode Protokoll wird bei CBT großen Wert auf die Skalierbarkeit gelegt. Diese wird durch eine Reduzierung der zu haltenden Zustandsinformationen und des Verwaltungsaufwandes durch einen gemeinsamen Baum für eine Multicast-Gruppe realisiert. Der Ablauf einer CBT – Kommunikation ist wie folgt:

1. Der Beitritt zu einer Gruppe G erfolgt ebenfalls wie im PIM-SM durch einen expliziten Gruppenbeitritt. Konkret wird eine IGMP-Join-Dateneinheit an den Designated-Router versendet. Dieser leitet die Nachricht weiter an den Rendezvous Punkt. Dieser oder ein Router auf dem Weg dorthin, der bereits Mitglied der Gruppe ist, muss jetzt, im Gegensatz zu PIM-SM, die Nachricht quittieren, um auf

ein periodisches Senden der Join-Nachrichten verzichten zu können.

2. Der Austritt aus einer Gruppe erfolgt auch explizit, was bedeutet, dass der Empfänger eine IGMP-Leave Dateneinheit an seinen Designated-Router sendet. Der Router untersucht, ob weitere Abhängigkeiten existieren, falls ja, wird die Leave-Dateneinheit verworfen, ansonsten wird sie zum nächsten Router geschickt, um diesen Teil aus dem Multicast-Baum zu entfernen.

Die Gefahr bei solch einer Kern-basierten Art der Kommunikation ist, wie bei PIM-SM, die Konzentration des Datenverkehrs um der Kern herum. In CBT existiert ebenfalls ein Bootstrap-Mechanismus zur Auswahl der Position. Dieser Mechanismus arbeitet wie PIM-SM mit dem periodischen Versenden von Bootstrap-Nachrichten. Es existiert weiterhin ein Verfahren der manuellen Konfiguration.

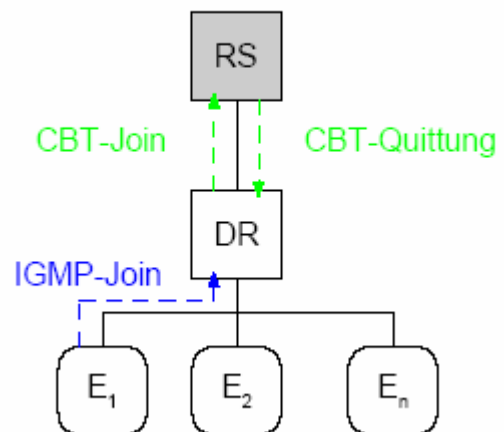


Abbildung 8: Ablauf der Paketvermittlung bei CBT (E = Endsystem, DR = Designated Router, RS = Rendezvous-Stelle)

2. Inter-Domain Routing Protokolle

Um Multicast auch im Internet nutzen zu können, werden an Inter-Domain Routing Protokolle hohe Anforderungen bezüglich Skalierbarkeit und der Einsatzmöglichkeit über Hierarchien hinweg gestellt. In der Entwicklung dieser Protokolle ergaben sich zwei Lösungswege. Die ersten Ergebnisse zielen auf eine schnelle „kurzzeit“-Lösung ab. Die daraus entstandenen Protokolle setzen auf Intra-Domain Routing Protokolle auf und haben dadurch Probleme, zwischen Domänen zu routen. Ein weiterer Weg befasst sich mit der Entwicklung von „langzeit“-Lösungen, aus

denen im Laufe der Zeit endgültige Lösungen entstehen sollen.

Aus diesem Grund sind die nun folgenden Protokolle in Kurzzeit- und Langzeit-Lösungen unterteilt.

2.1. Kurzzeit-Lösungen

2.1.1. MBGP

Als mit der Entwicklung der ersten kurzzeitigen Lösungen begonnen wurde, bestand eine der ersten Aufgaben darin, Multicast-Routing hierarchisch nutzbar zu machen. Ein Ansatz bestand darin, das Unicast Border-Gateway Protokoll (BGP) als Basis zu benutzen, da es Kontrollmöglichkeiten und Abstraktionen über Domänengrenzen hinweg integriert hat. Die inter-domain Lösung im Unicast-Bereich bestand darin, in einer Domäne ein beliebiges Routing-Protokoll zu wählen, während zwischen den Domänen BGP mittels eines Distanzvektorverfahrens die beste Route berechnet wurde.

Die Multicast-Erweiterung zum MBGP Protokoll bestand darin, dass ein „Subsequent Adress Family Identifier“ (SAFI) mit in das ursprüngliche Protokoll aufgenommen wurde. Dieser Identifier gibt an, ob es sich um eine Multicast-, Unicast- oder Multicast/Unicast-Nachricht handelt. Durch die Angabe dieses Wertes braucht nun jeder Router nur die Topologie seiner eigenen Domäne und die Pfade zu anderen Domänen anstatt die gesamten Multicast-Topologie zu kennen.

MBGP ist jedoch keine komplette Lösung, da mit diesem Verfahren keine Multicast-Bäume aufgebaut werden und Anmeldungen und Bestätigungen nicht behandelt werden.

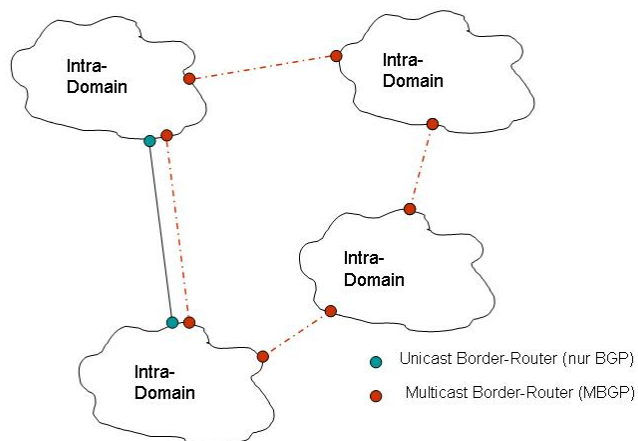


Abbildung 9: Beispiel einer Inter-Domain Multicast Topologie

2.1.2. MSDP

Das Multicast Source Delivery Protokoll stellt eine Erweiterung von PIM-SM dar. PIM-SM alleine ist es nicht möglich, einen Rendezvous-Punkt zu informieren, dass es Sender in anderen Domänen gibt. Dieses Problem tritt auf, wenn Gruppenmitglieder über mehrere Domänen verstreut sind. Die Empfänger senden in solch einem System wie gewohnt Join-Nachrichten an den Rendezvous-Punkt, Sender senden ebenfalls Nachrichten an ihren Rendezvous-Punkt. Die Rendezvous-Punkte haben keine Möglichkeit, untereinander Informationen auszutauschen, so kann kein Sender an einen Empfänger außerhalb seiner Domäne senden.

Die Idee von MSDP ist es, in jeder Domäne einen Repräsentanten zu benutzen, der die Existenz von Sendern mit Repräsentanten anderer Domänen austauscht. Diese Repräsentanten laufen auf dem Rendezvous-Punkt-Router mit.

Der Ablauf dieses Protokolls ist folgender:

1. Neue Sender registrieren sich am Rendezvous-Punkt ihrer Domäne
1. Der MSDP – Repräsentant bemerkt den Sender und schickt eine Source-Active Nachricht an alle direkt verbundenen MSDP Repräsentanten
2. Der MSDP-Empfänger-Repräsentant macht einen RPF-Check, falls dieser erfolgreich war, verteilt er die Nachricht an alle anderen MSDP (außer dem Sender) weiter
3. wenn ein MSPD Repräsentant Gruppenmitglieder in der Domäne hat, die Empfänger für diese Nachricht sind, sendet er eine Join-Nachricht an den Sender der Source-Active-Nachricht
4. wenn Daten in der Nachricht waren, werden diese vom Rendezvous-Punkt an die Empfänger in der Domäne weitergeleitet
5. Schritte 3-5 werden wiederholt, bis alle MSDP Repräsentanten die Source-Active Nachricht erhalten haben und die Gruppenmitglieder die Daten erhalten haben

Dieses Protokoll ist nur bei einer begrenzten Anzahl von Sendern im Netz anwendbar, da die Source-Active Nachrichten, die für jeden Sender erzeugt werden und mit den dazugehö-

rigen Daten in das Netz geflutet werden, sehr viel Verkehr erzeugen. Trotzdem stellt es eine schnelle Lösung für ein eiliges Problem dar, die bis zu Ablösung durch bessere Protokolle angewendet wird.

2.2. Langfristige Lösungen

2.2.1. BGMP

Das Border Gateway Multicast Protocol ist eine auf lange Sicht anzustrebende Lösung. BGMP basiert auf den Protokollen CBT und PIM-SM und ist eine Weiterentwicklung von BGP. Da ein Fluten des Netzes über Domänen hinaus unter keinen Umständen möglich wäre, benutzt BGMP wie CBT oder PIM-SM einen Rendezvous-Punkt. Gruppenmitglieder müssen explizit dem gemeinsamen Baum beitreten, um in die Multicast – Struktur mit aufgenommen zu werden.

Die grundlegenden Konzepte von BGMP sind:

- Aufbau eines gemeinsamen bidirektionalen Baumes über Domänen anstatt Router, so dass in Domänen selbst wieder intra-domain Routingprotokolle eingesetzt werden können
- Interoperabel mit anderen Protokollen
- Wahl der Wurzel domäne in Abhängigkeit des Präfixes der Multicast – Adresse

BGMP wird in den Grenzroutern einer Domäne eingesetzt, die über Domänen hinweg untereinander kommunizieren oder direkt mit Empfängern oder Sendern in der eigenen Domäne.

Die Wurzel stellt bei BGMP eine gesamte Domäne dar, die auf Basis der Multicast-Adressen ausgewählt wird. Diejenige Domäne, welche die Gruppenadresse in ihrer Multicast-Allokation zugeordnet bekommen hat, wird automatisch zur Wurzel domäne.

Damit BGMP ohne Fehler funktionieren kann, müssen die Multicast-Adressen klar definiert und nicht mehr durch Zufall bestimmt werden. Dafür wurden zwei Adress-Allokations-Schemen entwickelt: MASC, das Multicast address-set claim Protokoll und GLOP.

2.2.2. RAMA

Aufgrund der Komplexität der bisher besprochenen Protokolle und deren fehlenden Multicast-Komponenten wie Sicherheit oder Management wird an einer langfristigen Lösung namens Root Addressed Multicast Architecture gearbeitet. Die Voraussetzung dieser RAMA-

Protokolle ist, dass es nur einen Sender im Netz oder zumindest einen primären Sender gibt. Dieser (primär) Sender wird als Wurzel eines Multicast – Baumes eingesetzt, um die Komplexität, die bei anderen Protokollen entstehen würde, zu vermindern.

Zwei Protokolle implementieren diesen Ansatz: Express Multicast und Simple Multicast. Diese sollen nun erwähnt werden.

- Express Multicast
Express Multicast ist eine Single-Source Protokoll, also ein Protokoll mit nur einem Sender. Alle Gruppenmitglieder senden Join-Nachrichten über RPF an den Sender. EM sammelt Informationen über diese Teilnehmer und ist daher für Anmeldebasierte Applikationen wie Video-Broadcast oder Datei-Distributionen gedacht. Der Vorteil liegt in der geringen Komplexität und der Möglichkeit von geschlossenen Gruppen.
- Simple MC
Simple Multicast ist dem EM bis auf die Möglichkeit von mehreren Quellen ähnlich. Hier wird einer der Sender als primär ausgewählt. Nachdem auch die Empfänger ein Join an diesen Knoten gesendet haben, wird ein bidirektionaler Baum erzeugt, an dem ankommende Pakete an Routern sowohl an die Empfänger weitergeleitet werden, als auch zurück an den Wurzelknoten. Die Vorteile dieser Technik sind umstritten, jedoch sollen Probleme bei der Adress-Allokation durch dieses Verfahren vermieden werden, da die Adresse der Wurzel und die Multicast-Adresse benutzt werden, um eine Gruppe zu identifizieren.

Die Zukunft der RAMA – Technik ist ungewiss, da die eingesetzten Protokolle nach ihrer Standardisierung entweder zusammen mit alten Protokollen laufen, was allerdings die Vorteile nicht ausnutzen würde, oder alle alten ersetzen könnte, wobei allerdings nicht sichergestellt ist, ob RAMA genug Flexibilität für alle Anwendungen bietet.

V. ZUSAMMENFASSUNG

In dieser Belegarbeit wurde eine Übersicht über Multicast-Algorithmen und den darauf aufsetzenden Multicast-Protokollen geboten. Die Beschreibung der Algorithmen wurde aufgrund ihres vielfältigen Einsatzes in den Protokollen sehr allgemein gehalten, während bei den Protokollen selbst zum besseren Ver-

ständnis meist die Paketvermittlung detaillierter dargestellt wurde. Bei allen Techniken wurde zudem versucht, anhand von früheren, meist Unicast-Ansätzen, die Grundsteine und Ideen der Multicast-Verfahren zu beschreiben. Die folgende Abbildung soll das Zusammenwirken verschiedener Protokolltypen verdeutlichen.

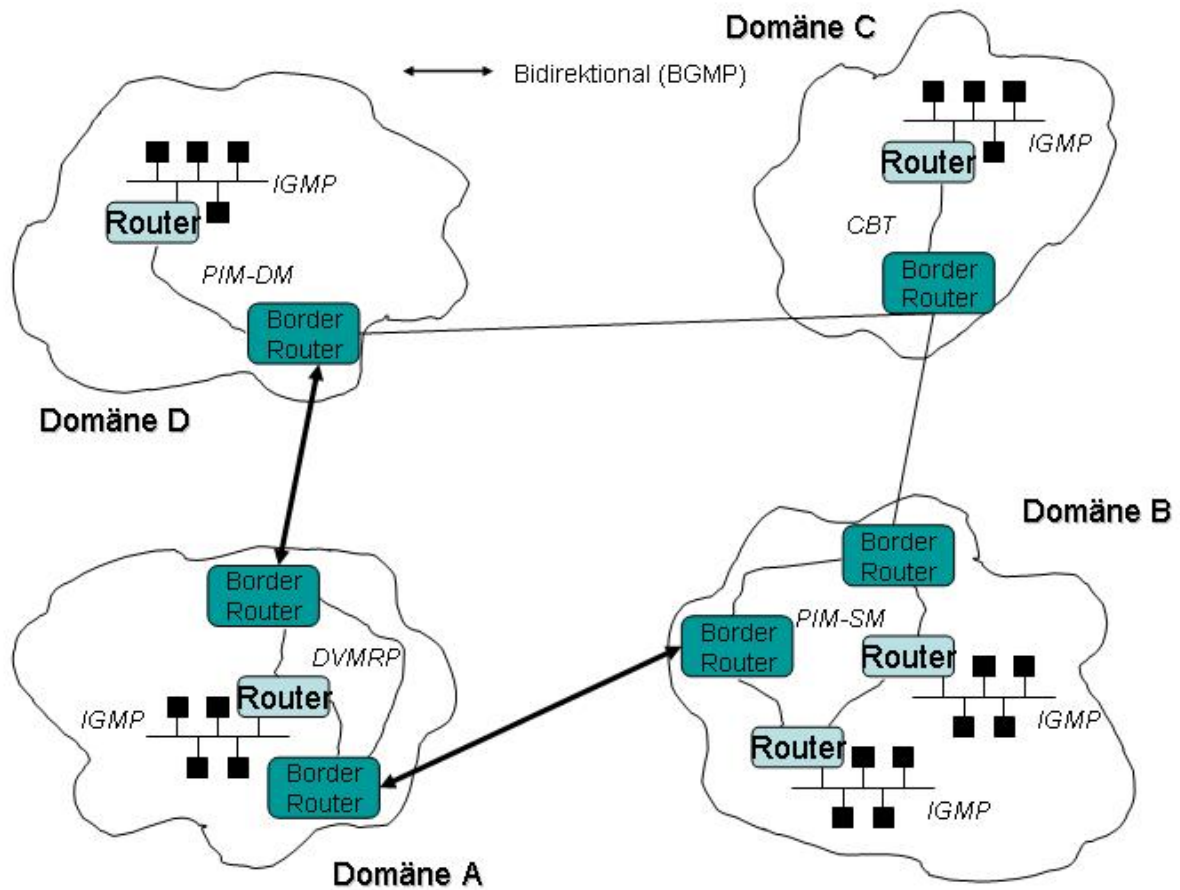


Abbildung 10: Verschiedene Multicast-Protokolle

LITERATUR

K.C.ALMEROTH: "The Evolution of Multicast: From the Mbone to Inter-Domain Multicast to Internet2 Deployment", *Department of Computer Science, University of California-Santa Barbara (1999)*

C.DIOT, W.DABBOUS, J.CROWCROFT: "Multipoint Communication: A Survey of Protocols, Functions, and Mechanisms", *IEEE Journal on selected areas in cummunication, Vol.15 No.3 (1997), 277-290*

J.DUFFNER: "Core-Based Trees und Protocol Independent Multicast (Sparse Mode)", *Seminararbeit am Lehrstuhl für Praktische Informatik IV, Prof. Dr. Wolfgang Effelsberg, Universität Mannheim (1999)*

M.FROMME, "IP Multicast", *Universität Hannover (1998)*

L.H.SAHASRABUDDHE, B.MUKHERJEE: "Multicast Routing Algorithms and Protocols: A Tutorial", *IEEE Network Jan/Feb 2000, 90-102*

D.SCHRÖDER: "Distance Vector Multicast Routing Protokoll", *Seminararbeit am Lehrstuhl für Praktische Informatik IV, Prof. Dr. Effelsberg, Fakultät für Mathematik und Informatik Universität Mannheim (1999)*

C.SHIELDS, J.J.GRACIA-LUNA-ACEVES: "The core based tree protocol", *Department of Computer Engineering, University of California-Santa Cruz*

R.WITTMANN, M.ZITTERBART: "Multicast Protokolle und Anwendungen", *dpunkt-Verlag (1999), ISBN 3-920993-40-3*